МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ТАВРІЙСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ В.І. ВЕРНАДСЬКОГО

Журнал заснований у 1918 році

ВЧЕНІ ЗАПИСКИ

ТАВРІЙСЬКОГО НАЦІОНАЛЬНОГО УНІВЕРСИТЕТУ ІМЕНІ В.І. ВЕРНАДСЬКОГО

Серія: Технічні науки

Том 36 (75) № 3 2025

Частина 2

Видавничий дім
«Гельветика»
2025

Головний редактор:

Кисельов Володимир Борисович – доктор технічних наук, професор, директор Навчально-наукового інституту муніципального управління та міського господарства Таврійського національного університету імені В.І. Вернадського.

Члени редакційної колегії:

Медведєв Микола Георгійович (відповідальний секретар) – доктор технічних наук, професор, завідувач кафедри загальноінженерних дисциплін та теплоенергетики Таврійського національного університету імені В.І. Вернадського;
Бронін Сергій Вадимович – кандидат технічних наук, доцент, доцент кафедри інформаційних систем та технологій Київського національного університету імені Тараса Шевченка;
Домніч Володимир Іванович – кандидат технічних наук, професор, завідувач кафедри автоматизованого управління технологічними процесами Таврійського національного університету імені В.І. Вернадського;
Дехтяр Анатолій Соломонович – доктор технічних наук, професор, професор кафедри архітектурних конструкцій Національної академії образотворчого мистецтва і архітектури;
Дичко Аліна Олегівна – доктор технічних наук, професор, професор кафедри загальноінженерних дисциплін та теплоенергетики Таврійського національного університету імені В.І. Вернадського;
Дубко Валерій Олексійович – доктор фізико-математичних наук, професор, професор кафедри вищої математики Київського національного університету технології та дизайну;
Єремєєв Ігор Семенович – доктор технічних наук, професор, професор кафедри автоматизованого управління технологічними процесами Таврійського національного університету імені В.І. Вернадського;
Лисенко Олександр Іванович – доктор технічних наук, професор, професор кафедри телекомунікацій Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»
Огородник Станіслав Станіславович – доктор технічних наук, старший науковий співробітник, професор кафедри загальноінженерних дисциплін та теплоенергетики Таврійського національного університету імені В.І. Вернадського;
Сегай Олександр Михайлович – кандидат технічних наук, доцент, доцент кафедри загально-інженерних дисциплін та теплоенергетики Таврійського національного університету імені В.І. Вернадського;
Чумаченко Сергій Миколайович – доктор технічних наук, старший науковий співробітник, завідувач кафедри інформаційних систем Національного університету харчових технологій;
Цомко Олена – доктор філософії по спеціальності «Безпека і управління інформацією», відділення комп'ютерної інженерії, Інститут Міжнародної освіти, Університет Донгсо, Республіка Корея.

Статті у виданні перевірені на наявність плагіату за допомогою програмного забезпечення StrikePlagiarism.com від польської компанії Plagiat.pl.

Рекомендовано до друку та поширення через мережу Internet
Вченою радою Таврійського національного університету імені В.І. Вернадського
(протокол № 11 від 27 червня 2025 року)

Науковий журнал «Вчені записки ТНУ імені В.І. Вернадського. Серія: Технічні науки»
зареєстровано відповідно до Рішення Національної ради України
з питань телебачення і радіомовлення № 1136 від 11.04.2024 року.

Ідентифікатор медіа: R30-04000.

Суб'єкт у сфері друкованих медіа – Таврійський національний університет імені В. І. Вернадського (просп. Берестейський, 10, м. Київ, 01135, e-mail: crimea.tnu@gmail.com, тел.: +38 (044) 529-05-16).

Мови видання: українська, англійська, польська, німецька,
французька, словацька, румунська, італійська.

Періодичність: 6 разів на рік.

Журнал включено до Переліку наукових фахових видань України (категорія «Б») з технічних наук (спеціальності: G1 Хімічні технології та інженерія; G4 Енерговиробництво (за спеціалізацією); G5 Електроніка, електронні комунікації, приладобудування та радіотехніка) відповідно до Наказу МОН України від 17.03.2020 № 409 (додаток 1), F2 Інженерія програмного забезпечення; F6 Інформаційні системи і технології; F7 Комп'ютерна інженерія; G7 Автоматизація, комп'ютерно-інтегровані технології та робототехніка; D3 Менеджмент; G11 Машинобудування (за спеціалізаціями); J5 Морський та внутрішній водний транспорт; J6 Авіаційний транспорт; J7 Залізничний транспорт; J8 Автомобільний транспорт відповідно до Наказу МОН України від 02.07.2020 № 886 (додаток 4)

Журнал включено до міжнародної наукометричної бази Index Copernicus International (Республіка Польща)

Сторінка журналу: www.tech.vernadskyjournals.in.ua

ISSN 2663-5941 (Print)
ISSN 2663-595X (Online)

© Таврійський національний університет ім. В.І. Вернадського, 2025

ЗМІСТ

ГАЛУЗЕВЕ МАШИНОБУДУВАННЯ

Сташевський В.М.
СТРАТЕГІЧНІ ОРІЄНТИРИ ІННОВАЦІЙНОГО РОЗВИТКУ ТЕХНІКО-ТЕХНОЛОГІЧНОГО ЗАБЕЗПЕЧЕННЯ АГРОПРОМИСЛОВИХ ПІДПРИЄМСТВ.....1

Топчій Н.В.
ПРОБЛЕМА ВПРОВАДЖЕННЯ КОНЦЕПЦІЇ НЕВИЗНАЧЕНОСТІ ВИМІРЮВАНЬ В МЕТОДИ КОНТРОЛЮ.....7

Fomin O.V., Burlutskyi O.V., Kubrak A.V., Holub H.M., Fursina A.D.
NON-STATIONARY FACTORS OF FATIGUE RESISTANCE AND THEIR INFLUENCE ON THE SERVICE LIFE OF FREIGHT CAR JOINTS: MATHEMATICAL MODELING AND ANALYSIS.....14

Чайковський С.Ю.
ОСОБЛИВОСТІ КОНТРОЛЮ ЯКОСТІ ПРОДУКЦІЇ В УМОВАХ СТРІМКОГО РОЗВИТКУ РИНКОВОЇ ЕКОНОМІКИ.....23

Шорнікова С.В.
ТЕХНІЧНЕ ЗАВДАННЯ – ПЕРШИЙ КРОК ДЛЯ ОТРИМАННЯ ЯКІСНОГО ЕЛЕКТРОННОГО ВИРОБУ.....30

ІНФОРМАТИКА, ОБЧИСЛЮВАЛЬНА ТЕХНІКА ТА АВТОМАТИЗАЦІЯ

Арпентій С.П.
МЕТОДОЛОГІЧНА ПАРАДИГМА АРХІТЕКТУРНОЇ ОРГАНІЗАЦІЇ АВТОНОМНИХ ТРАНСПОРТНИХ СИСТЕМ.....36

Астістова Т.І., Дроменко В.Б., Калашник В.Ю.
РОЗРОБЛЕННЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ІНФОРМАЦІЙНОГО РЕСУРСУ БІБЛІОТЕКИ ЗАКЛАДІВ ОСВІТИ.....43

Бараннік В.В., Берчанов А.А.
МЕТОД ОБРОБКИ ІНФРАЧЕРВОНИХ ЗОБРАЖЕНЬ У СПЕКТРАЛЬНОМУ ДОМЕНІ.....55

Бараннік Д.В., Сідченко Є.С., Лиходєєв О.С., Помазуєв В.В., Федоровський О.В.
МЕТОД ЗАСТОСУВАННЯ АПЕРТУР ДО АСИМЕТРИЧНИХ СИСТЕМ КОДУВАННЯ ВІДЕОДАННИХ.....61

Батюк С.Г., Воронюк М.В.
ЗАДАЧІ КОМПЛЕКСНОЇ АВТОМАТИЗАЦІЇ ЖИТЛОВОГО БУДИНКУ ЯК АВТОНОМНОЇ КІБЕР-ФІЗИЧНОЇ СИСТЕМИ.....67

Батюк С.Г., Воронюк М.В.
ЗАСТОСУВАННЯ ЦИФРОВИХ ДВІЙНИКІВ І СИСТЕМ ШТУЧНОГО ІНТЕЛЕКТУ В АВТОМАТИЗАЦІЇ ТЕПЛОЕНЕРГЕТИЧНИХ ПРОЦЕСІВ.....77

Безкровний М.М.
МЕТОДИ І ЗАСОБИ АНАЛІЗУ АСТРОНОМІЧНИХ ЗОБРАЖЕНЬ ДЛЯ ВИЗНАЧЕННЯ КООРДИНАТ НЕБЕСНИХ ОБ'ЄКТІВ.....92

Богдан Г.А., Феоклістов Д.О.
ПІДВИЩЕННЯ ТОЧНОСТІ ІНЕРЦІАЛЬНИХ НАВІГАЦІЙНИХ СИСТЕМ ЗА ДОПОМОГОЮ НЕЙРОМЕРЕЖЕВИХ ТЕХНОЛОГІЙ.....101

Бойко О.В., Ульяницька К.О., Коломієць У.Д.
ДО ПИТАННЯ ОПТИМІЗАЦІЇ ЗАПИТІВ РЕЛЯЦІЙНОЇ БАЗИ ДАНИХ ДЛЯ ВИРІШЕННЯ АНАЛІТИЧНИХ ЗАВДАНЬ.....107

Бочок В.О., Федорова Н.В.
ОБМІН ЗНАННЯМИ В INDEPENDENT DQN БАГАТОАГЕНТНІЙ СИСТЕМІ ТА ОСОБЛИВОСТІ ЙОГО РЕАЛІЗАЦІЇ.....113

Вакалюк Т.А., Антонюк Д.С., Мінтій І.С., Новіцька І.В., Довгалюк І.С.
МЕТОДИ ПОБУДОВИ РЕКОМЕНДАЦІЙНИХ СИСТЕМ ТА ІНТЕГРАЦІЯ ШТУЧНОГО ІНТЕЛЕКТУ У ФІНАНСОВИХ ДОДАТКАХ.....120

Vernik M.O., Oleshchenko L.M.
INTELLIGENT SOFTWARE FRAMEWORK FOR BIG DATA AUDIO PROCESSING AND COGNITIVE ENERGY ESTIMATION.....125
Войтко В.В., Позур М.Ю.
МЕТОД ДИНАМІЧНИХ ВИКЛИКІВ В .NET ШЛЯХОМ ГЕНЕРАЦІЇ IL КОДУ В ПРОЦЕСІ ВИКОНАННЯ ЗАСТОСУНКУ.....132
Glukhova N.V.
COMPUTER-INTEGRATED TECHNOLOGY FOR STUDYING THE SPATIAL STRUCTURE OF A CORONA GAS DISCHARGE BASED ON FRACTAL DIMENSION ESTIMATION.....138
Holod I.V.
DEVELOPMENT OF A PARAMETRIC MODEL AND FORMALIZATION OF PARAMETERS FOR MICROCLIMATE CONTROL IN INDUSTRIAL PREMISES BASED ON CYBER-PHYSICAL SYSTEMS.....144
Григор'єва Т.І., Розенвассер Д.М., Хнюнін С.Г., Крохмаль М.В.
АНАЛІЗ СКЛАДНИХ СИСТЕМ ЩОДО МАШИННОГО НАВЧАННЯ ЗА ДОПОМОГОЮ ТЕНЗОРНИХ МЕТОДІВ.....151
Гуржій С.В.
ROS 2: ОСОБЛИВОСТІ АРХІТЕКТУРИ ТА ПРАКТИЧНОГО ВПРОВАДЖЕННЯ.....157
Дем'янчук Т.М., Губар В.Г.
АНАЛІЗ ЗАСТОСУВАННЯ МІР ПОДІБНОСТІ З МЕТОЮ ПОРІВНЯННЯ СИГНАТУР СИГНАЛІВ БПЛА.....165
Заварін В.О., Терейковський І.А.
АНАЛІЗ СУЧАСНИХ РІШЕНЬ У КОМП'ЮТЕРНІЙ ДІАГНОСТИЦІ РОЗЛАДІВ ЗДОРОВ'Я ЗА РАЙДУЖКОЮ ОКА.....173
Здоренко С.С., Черепанська І.Ю.
СИСТЕМА ЕНЕРГОЕФЕКТИВНОГО ПРОГНОЗНОГО КЕРУВАННЯ ТЕМПЕРАТУРОЮ ПОВІТРЯ В БУДІВЛЯХ З АВТОНОМНИМ ОПАЛЕННЯМ.....180
Зилевіч М.О., Олійник Р.С.
КОНТРОЛЕР ПОЛЬОТУ З ВЕКТОРНИМ КЕРУВАННЯМ ТЯГОЮ НА БАЗІ STM32.....190
Зіменков Д.А., Терейковський І.А.
АНАЛІЗ ПРОБЛЕМ МЕТОДІВ РОЗПІЗНАВАННЯ МОВЛЕННЯ.....198
Зубенко В.О., Березюк І.А.
ОПТИМІЗАЦІЯ ПРОЦЕСУ КЕРУВАННЯ ПАРАМЕТРАМИ МІКРОКЛІМАТУ ТЕПЛИЦІ З ВИКОРИСТАННЯМ МАТЕМАТИЧНОГО МОДЕЛЮВАННЯ.....206
Іванишин В.В., Мошенский А.О.
АПАРАТНЕ ЗАБЕЗПЕЧЕННЯ БІО-ТЕХНІЧНОЇ СИСТЕМИ МОНІТОРИНГУ І КЕРУВАННЯ ПАСІКОЮ. МІКРОПРОЦЕСОРНІ ПЛАТФОРМИ.....214
Карпенко М.І., Мошенський А.О.
КОНСТРУЮВАННЯ ТА ПОЛЬОВА АПРОБАЦІЯ ПРИСТРОЮ ДЛЯ ДОСЛІДЖЕННЯ ПОКРИТТЯ БСМ МЕРЕЖ.....221
Кордюк Р.І.
ДЕРЕВОВИДНА МОДЕЛЬ КОЛАЖІВ ІНФОРМАЦІЙНОЇ СИСТЕМИ.....229
Кропивницька В.Б., Кропивницький Д.Р.
МОДЕЛЮВАННЯ РОБОТИ МУРАШИНИХ АЛГОРИТМІВ У ЗАДАЧАХ МАРШРУТИЗАЦІЇ.....241
Левчук І.Л., Дубовик Т.М., Гузь Г.М., Романчук О.О.
РОЗРОБКА ЛЮДИНО-МАШИННОГО ІНТЕРФЕЙСУ СИСТЕМИ КЕРУВАННЯ НА БАЗІ SCADA ПРОЦЕСОМ ОТРИМАННЯ МІДНОГО КУПОРОСУ.....248
Майданюк В.П., Складанюк О.О.
ГРАФІЧНИЙ РЕДАКТОР ДЛЯ РЕДАГУВАННЯ ДЕСКТОПНИХ ВІДЕОІГОР.....255
Мацуй А.М., Сербул О.М., Хлєбніков М.В., Ізовіта О.Л.
ОБГРУНТУВАННЯ РОЗРОБКИ КОМП'ЮТЕРИЗОВАНОЇ СИСТЕМИ СТАБІЛІЗАЦІЇ ОПТИМАЛЬНОГО РІЗНОРОЗМІРНОГО КУЛЬОВОГО ЗАВАНТАЖЕННЯ БАРАБАННИХ МЛИНІВ.....261

Мойсеєнко О.В., Заячук Я.І.
ПРОГНОЗУВАННЯ РИЗИКІВ КІБЕРАТАК НА ОСНОВІ МОНІТОРИНГУ ТРАФІКУ В КОМП'ЮТЕРНИХ МЕРЕЖАХ З ВИКОРИСТАННЯМ АЛГОРИТМІВ АНАЛІЗУ ЧАСОВИХ РЯДІВ.....268
Олещенко Л.М.
ДОСЛІДЖЕННЯ АПАРАТНИХ ТА ПРОГРАМНИХ ЗАСОБІВ ДЛЯ ПАРАЛЕЛЬНОЇ ОБРОБКИ ВЕЛИКИХ ДАНИХ І ЗАДАЧ МАШИННОГО НАВЧАННЯ У ХМАРНОМУ СЕРЕДОВИЩІ.....276
Парінов І.С.
ІНТЕГРАЦІЯ НЕЙРОМЕРЕЖЕВИХ СИСТЕМ У МЕДИЧНІ ІТ-ІНФРАСТРУКТУРИ: ТЕХНІЧНИЙ ПІДХІД ДО РОЗРОБКИ МОДУЛЯ ПРОГРЕСИВНОЇ СУМАРИЗАЦІЇ ДАНИХ ПАЦІЄНТА.....287
Pasichnyk V.V., Dobrovolskyi H.A.
SURVEY OF APPLICATION OF THE KOLMOGOROV-ARNOLD NETWORKS TO TRANSPORTATION NETWORK INTERDICTION GAMES.....297
Положаєнко С.А., Прокофьєв А.Ю.
АДАПТИВНИЙ МЕТОД КОНТРОЛЮ ТА КРИТЕРІЇ ПЕРЕВІРКИ ДОСТОВІРНОСТІ ОБЧИСЛЮВАЛЬНОЇ РЕАЛІЗАЦІЇ МОДЕЛЕЙ ДИНАМІЧНИХ СИСТЕМ.....305
Полтавський Д.А., Мироненко О.В., Снєосіков О.А.
МЕТОДИ ВИЯВЛЕННЯ ТА ЗАПОБІГАННЯ ФІШИНГОВИМ АТАКАМ У ВЕБСЕРЕДОВИЩІ.....312
Поліщук С.М.
ОГЛЯД МЕТОДІВ ДЕТЕКТУВАННЯ ЗАМАСКОВАНИХ ОБ'ЄКТІВ У КОНТЕКСТІ ВІЗУАЛЬНОЇ СХОЖОСТІ З ФОНОМ.....320
Регенель Т.Ю., Фант М.О., Савіцький Р.С.
РОЗРОБКА СИСТЕМИ КОНТРОЛЮ ВЕРСІЙ ДЛЯ МУЗИЧНИХ ПРОЄКТІВ.....328
Рихальський О.Ю.
МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ СЕМАНТИЧНОЇ СУМІСНОСТІ ПАРАМЕТРИЧНИХ КОНФІГУРАЦІЙ У CAD-СИСТЕМАХ.....334
Семенець Д.А., Васильчук Д.П., Арестова Т.В.
ЦИФРОВИЙ ПРОПОРЦІЙНО-ІНТЕГРАЛЬНИЙ-РЕГУЛЯТОР ДЛЯ ЕЛЕКТРОМЕХАНІЧНОГО ОБ'ЄКТУ ДРУГОГО ПОРЯДКУ З ЗАПІЗНЮВАННЯМ 343
Ситніков О.В., Лукінюк М.В., Фтемов А.О.
СИНТЕЗ СИСТЕМИ КЕРУВАННЯ ВОДООЧИЩЕННЯ ЗА ДОПОМОГОЮ МОДУЛЬНОГО КОНТЕЙНЕРА СИСТЕМИ ЗВОРОТНОГО ОСМОСУ.....350
Слабінога М.О., Піх В.Я., Стрілецький Ю.Й., Шекета В.І.
РОЗПІЗНАВАННЯ ТЕХНІЧНОГО СТАНУ ОБ'ЄКТІВ НА ОСНОВІ ОПРАЦЮВАННЯ ВІБРАЦІЙНИХ СИГНАЛІВ БАГАТОШАРОВИМ ПЕРЦЕПТРОНОМ.....356
Соловій П.С., Назаренко В.В.
ВПРОВАДЖЕННЯ ШІ У СПЕЦІАЛІЗОВАНІ КОМП'ЮТЕРНІ СИСТЕМИ ДЛЯ ПІДТРИМКИ КРИТИЧНОЇ ІНФРАСТРУКТУРИ.....361
Степанов Д.С., Сенів М.М.
ЗАСТОСУВАННЯ ТЕХНОЛОГІЙ НЕЗМІННОЇ ІНФРАСТРУКТУРИ ДЛЯ ЗАБЕЗПЕЧЕННЯ ПРОЗОРОГО МАСШТАБУВАННЯ БЕЗ ПРОСТОЇВ У ВЕБПОРТАЛАХ.....369
Стьопкін А.В.
РОЗПІЗНАВАННЯ НЕОРІЄНТОВАНИХ ГРАФІВ МУЛЬТИАГЕНТНОЮ СИСТЕМОЮ.....377
Танчук В.С., Колобродов В.Г.
МЕТОД МОДЕЛЮВАННЯ ТЕПЛОВІЗІЙНИХ ЗОБРАЖЕНЬ НА ОСНОВІ СИСТЕМНОГО ПІДХОДУ.....383
Терейковський І.А., Шевченко О.В.
АНАЛІЗ НЕЙРОМЕРЕЖЕВИХ РІШЕНЬ УПРАВЛІННЯ БЕЗПІЛОТНИМ ПРИСТРОЄМ.....393
Тиханська А.М., Тиханський М.П.
КЕРУВАННЯ ВАЖКИМ СТРІЧКОВИМ КОНВЕЄРОМ З УРАХУВАННЯМ ТЕХНІЧНОГО СТАНУ ЙОГО ОБЛАДНАННЯ.....400

Усов О.С.
МЕТОДИ ВІЗУАЛЬНОЇ SLAM ДЛЯ НАВІГАЦІЇ У ВНУТРІШНЬОМУ СЕРЕДОВИЩІ БЕЗ ДОСТУПУ ДО ГЛОБАЛЬНОГО ПОЗИЦІОНУВАННЯ.....411
Ushakova I.O., Bondarenko D.O., Chirva Yu.Ye., Znakhur L.V.
COMPUTER MODELING OF COMPANY EMPLOYEE CHURN USING MACHINE LEARNING AND PREDICTIVE ANALYTICS METHODS.....419
Філімончук Т.В., Колтун Ю.М., Климова І.М., Холобок В.І.
МОДЕЛЬ ПІДХОДУ РОЗРОБКИ МОБІЛЬНИХ ЗАСТОСУНКІВ.....429
Tsybulnyk S.O., Shelemakha V.V.
OVERVIEW OF METHODS AND MEANS OF AUTOMATED DETERMINATION OF THE CURRENT STATE OF BIOLOGICAL OBJECTS.....437
Чигур І.І., Лагойда А.І., Чигур Л.Я.
ІНФОРМАТИВНІ ОЗНАКИ В ЗАВДАННЯХ АВТОМАТИЗОВАНОГО РОЗПІЗНАВАННЯ ТЕХНОЛОГІЧНИХ СИТУАЦІЙ ПРИ БУРІННІ СВЕРДЛОВИН.....445
Чижмотря О.В., Вакалюк Т.А., Клименко Д.О., Чижмотря О.Г., Гордієнко І.В.
ДОСЛІДЖЕННЯ СТРУКТУРИ ВЕБ-ТРАФІКУ ТА ВИЗНАЧЕННЯ ВИРІШАЛЬНИХ ХАРАКТЕРИСТИК ЗАПИТІВ.....451
Ярецька Н.О., Рамський А.О., Поплавська О.А.
ОСОБЛИВОСТІ ВИКЛАДАННЯ ВИЩОЇ МАТЕМАТИКИ СТУДЕНТАМ ГАЛУЗІ ЗНАНЬ «ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ» В УМОВАХ ЦИФРОВІЗАЦІЇ ОСВІТИ.....457

БУДІВНИЦТВО

Деревягіна Н.І., Садовенко І.О.
ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ ЗМІН ПАРАМЕТРІВ СТІЙКОСТІ ЛЬОСОВИХ МАСИВІВ ПРИ ВИБУХОВИХ НАВАНТАЖЕННЯХ.....464

ВІДОМОСТІ ПРО АВТОРІВ.....470

CONTENTS

INDUSTRIAL ENGINEERING

Stashevskyi V.M.
STRATEGIC GUIDELINES FOR INNOVATIVE DEVELOPMENT OF TECHNICAL AND TECHNOLOGICAL SUPPORT OF AGRO-INDUSTRIAL ENTERPRISES 1

Topchii N.V.
IMPLEMENTATION OF THE CONCEPT OF MEASUREMENT UNCERTAINTY INTO CONTROL METHODS 7

Fomin O.V., Burlutskyi O.V., Kubrak A.V., Holub H.M., Fursina A.D.
NON-STATIONARY FACTORS OF FATIGUE RESISTANCE AND THEIR INFLUENCE ON THE SERVICE LIFE OF FREIGHT CAR JOINTS: MATHEMATICAL MODELING AND ANALYSIS 14

Chaikovskyi S.Yu.
FEATURES OF PRODUCT QUALITY CONTROL IN THE CONDITIONS OF RAPID DEVELOPMENT OF MARKET ECONOMY 23

Shornikova S.V.
THE TECHNICAL SPECIFICATION IS THE FIRST STEP IN OBTAINING A HIGH-QUALITY ELECTRONIC PRODUCT 30

INFORMATICS, COMPUTER ENGINEERING AND AUTOMATION

Arpentii S.P.
METHODOLOGICAL PARADIGM OF ARCHITECTURAL ORGANISATION OF AUTONOMOUS TRANSPORT SYSTEMS 36

Astistova T.I., Dromenko V.B., Kalashnyk V.Yu.
DEVELOPMENT OF SOFTWARE FOR INFORMATION RESOURCES OF EDUCATIONAL INSTITUTIONS LIBRARY 43

Barannik V.V., Berchanov A.A.
METHOD OF PROCESSING INFRARED IMAGES IN THE SPECTRAL DOMAIN 55

Barannik D.V., Sidchenko Ye.S., Lykhodieiev O.S., Pomazuiev V.V., Fedorovskiy O.V.
METHOD OF APPLYING APERTURES FOR ASYMMETRIC CODING VIDEODATA SYSTEMS 61

Batiuk S.G., Voroniuk M.V.
TASK OF COMPREHENSIVE AUTOMATION OF A RESIDENTIAL BUILDING AS AN AUTONOMOUS CYBER-PHYSICAL SYSTEM 67

Batiuk S.G., Voroniuk M.V.
APPLICATION OF DIGITAL TWINS AND ARTIFICIAL INTELLIGENCE SYSTEMS IN THE AUTOMATION OF THERMAL ENERGY PROCESSES 77

Bezkrovnyi M.M.
METHODS AND TOOLS FOR ANALYZING ASTRONOMICAL IMAGES TO DETERMINE THE COORDINATES OF CELESTIAL OBJECTS 92

Bohdan H.A., Feoklistov D.A.
IMPROVING THE ACCURACY OF INERTIAL NAVIGATION SYSTEMS USING NEURAL NETWORK TECHNOLOGIES 101

Boiko O.V., Ulianytska K.O., Kolomiiets U.D.
ON THE QUESTION OF OPTIMIZATION OF RELATIONAL DATABASE QUERIES FOR SOLVING ANALYTICAL PROBLEMS 107

Bochok V.O., Fedorova N.V.
KNOWLEDGE SHARING IN INDEPENDENT DQN MULTI-AGENT SYSTEM AND FEATURES OF ITS IMPLEMENTATION 113

Vakaliuk T.A., Antoniuk D.S., Mintii I.S., Novitska I.V., Dovgaliuk I.S.
METHODS FOR BUILDING RECOMMENDATION SYSTEMS AND INTEGRATING ARTIFICIAL INTELLIGENCE INTO FINANCIAL APPLICATIONS 120

Vernik M.O., Oleshchenko L.M.
INTELLIGENT SOFTWARE FRAMEWORK FOR BIG DATA AUDIO PROCESSING AND COGNITIVE ENERGY ESTIMATION.....125
Voitko V.V., Pozur M.Yu.
METHOD OF EXECUTING DYNAMIC CALLS IN .NET USING IL GENERATION IN RUNTIME...132
Glukhova N.V.
COMPUTER-INTEGRATED TECHNOLOGY FOR STUDYING THE SPATIAL STRUCTURE OF A CORONA GAS DISCHARGE BASED ON FRACTAL DIMENSION ESTIMATION.....138
Holod I.V.
DEVELOPMENT OF A PARAMETRIC MODEL AND FORMALIZATION OF PARAMETERS FOR MICROCLIMATE CONTROL IN INDUSTRIAL PREMISES BASED ON CYBER-PHYSICAL SYSTEMS.....144
Hryhorieva T.I., Rozenvasser D.M., Khniunin S.H., Krokhmal M.V.
ANALYSIS OF COMPLEX SYSTEMS WITH MACHINE LEARNING USING TENSOR METHODS.....151
Hurzhii S.V.
ROS 2: ARCHITECTURE AND PRACTICAL IMPLEMENTATION FEATURES.....157
Demianchuk T.M., Gubar S.G.
ANALYSIS OF THE USE OF SIMILARITY MEASURES FOR COMPARING UAV SIGNAL SIGNATURE.....165
Zavarin V.O., Tereikovskyi I.A.
ANALYSIS OF MODERN SOLUTIONS IN COMPUTERIZED DIAGNOSTICS OF HEALTH DISORDERS BY THE IRIS OF THE EYE.....173
Zdorenko S.S., Cherepanska I.Yu.
SYSTEM FOR ENERGY-EFFICIENT PREDICTIVE CONTROL OF INDOOR AIR TEMPERATURE IN BUILDINGS WITH AUTONOMOUS HEATING.....180
Zylevich M.O., Oliinyk R.M.
FLIGHT CONTROLLER WITH THRUST VECTORING BASED ON STM32.....190
Zimenkov D.A., Tereikovskyi I.A.
ANALYSIS OF PROBLEMS IN SPEECH RECOGNITION METHODS.....198
Zubenko V.O., Bereziuk I.A.
OPTIMIZATION OF THE GREENHOUSE MICROCLIMATE PARAMETERS CONTROL PROCESS USING MATHEMATICAL MODELING.....206
Ivanyshyn V.V., Moshensky A.O.
HARDWARE FOR THE BIO-TECHNICAL SYSTEM FOR MONITORING AND CONTROLLING THE APIARY. MICROPROCESSOR PLATFORMS.....214
Karpenko M.I., Moshenskyi A.O.
DESIGN AND FIELD TESTING OF A DEVICE FOR RESEARCHING WSN COVERAGE.....221
Kordiuk R.I.
TREE-LIKE COLLAGE MODEL OF INFORMATION SYSTEM.....229
Kropyvnytska V.B., Kropyvnytskyi D.R.
MODELING THE PERFORMANCE OF ANT COLONY ALGORITHMS FOR NETWORK ROUTING TASKS.....241
Levchuk I.L., Dubovyk T.M., Huz H.M., Romanchuk O.O.
DEVELOPMENT OF A HUMAN-MACHINE INTERFACE FOR A CONTROL SYSTEM BASED ON SCADA FOR THE COPPER SULFATE PRODUCTION PROCESS.....248
Maidaniuk V.P., Skladaniuk O.O.
GRAPHICAL EDITOR FOR DESKTOP VIDEO GAME SCENE EDITING.....255
Matsui A.M., Serbul O.M., Khlebnikov M.V., Izovita O.L.
JUSTIFICATION OF THE DEVELOPMENT OF A COMPUTERISED SYSTEM FOR STABILISATION OF OPTIMAL MULTI-SIZE BALL LOADING OF DRUM MILLS.....261

Moyseenko O.V., Zaiachuk Ya.I.
PREDICTION OF CYBER ATTACKS BASED ON TRAFFIC INTENSITY MONITORING IN COMPUTER NETWORKS....268
Oleshchenko L.M.
INVESTIGATION OF HARDWARE AND SOFTWARE TOOLS FOR PARALLEL PROCESSING OF BIG DATA AND MACHINE LEARNING TASKS IN CLOUD ENVIRONMENT....276
Parinov I.S.
INTEGRATION OF NEURAL NETWORK SYSTEMS INTO MEDICAL IT INFRASTRUCTURES: A TECHNICAL APPROACH TO THE DEVELOPMENT OF A MODULE FOR PROGRESSIVE SUMMARIZATION OF PATIENT DATA....287
Pasichnyk V.V., Dobrovolskyi H.A.
SURVEY OF APPLICATION OF THE KOLMOGOROV-ARNOLD NETWORKS TO TRANSPORTATION NETWORK INTERDICTION GAMES....297
Polozhaenko S.A., Prokofiev A.Ju.
ADAPTIVE CONTROL METHOD AND CRITERIA FOR VERIFYING THE RELIABILITY OF COMPUTER IMPLEMENTATION OF DYNAMIC SYSTEMS MODELS....305
Poltavskyi D.A., Myronenko O.V., Snieosikov O.A.
METHODS OF DETECTING AND PREVENTING PHISHING ATTACKS IN THE WEB ENVIRONMENT....312
Polischuk S.M.
REVIEW OF METHODS FOR DETECTING MASKED OBJECTS IN THE CONTEXT OF VISUAL SIMILARITY TO THE BACKGROUND....320
Rehenel T.Yu., Fant M.O., Savitskyi R.S.
DEVELOPMENT OF A VERSION CONTROL SYSTEM FOR MUSIC PROJECTS....328
Rykhalskyy O.Yu.
MATHEMATICAL MODELING OF SEMANTIC COMPATIBILITY OF PARAMETRIC CONFIGURATIONS IN CAD SYSTEMS....334
Semenets D.A., Vasylchuk D.P., Arestova T.V.
DIGITAL PROPORTIONAL-INTEGRAL CONTROLLER FOR A SECOND-ORDER ELECTROMECHANICAL OBJECT WITH A DELAY....343
Sytnikov A.V., Lukiniuk M.V., Ftemov A.O.
SYNTHESIS OF A WATER TREATMENT CONTROL SYSTEM USING A MODULAR CONTAINER OF A REVERSE OSMOSIS SYSTEM....350
Slabinoha M.O., Pikh V.Ya., Striletskyi Yu.Y., Sheketa V.I.
TECHNICAL CONDITION RECOGNITION BASED ON VIBRATION SIGNAL PROCESSING USING MULTI-LAYER PERCEPTRON....356
Solovii P.S., Nazarenko V.V.
IMPLEMENTATION OF ARTIFICIAL INTELLIGENCE INTO SPECIALIZED COMPUTER SYSTEMS FOR CRITICAL INFRASTRUCTURE SUPPORT....361
Stepanov D.S., Seniv M.M.
APPLICATION OF IMMUTABLE INFRASTRUCTURE TECHNOLOGIES TO ENSURE TRANSPARENT SCALING WITHOUT DOWNTIME IN WEB PORTALS....369
Stopkin A.V.
UNDIRECTED GRAPH EXPLORATION USING A MULTI-AGENT SYSTEM....377
Tanchuk V.S., Kolobrodov V.G.
THERMAL IMAGE MODELING METHOD BASED ON A SYSTEMATIC APPROACH....383
Tereikovskiy I.A., Shevchenko O.V.
ANALYSIS OF NEURAL NETWORK SOLUTIONS FOR UNMANNED VEHICLE CONTROL....393
Tykhanska A.M., Tykhanskyi M.P.
CONTROL OF A HEAVY BELT CONVEYOR TAKING INTO ACCOUNT THE TECHNICAL CONDITION OF ITS EQUIPMENT....400

Usov O.S.
VISUAL SLAM METHODS FOR NAVIGATION IN INDOOR ENVIRONMENTS WITHOUT ACCESS TO GLOBAL POSITIONING.....411
Ushakova I.O., Bondarenko D.O., Chirva Yu.Ye., Znakhur L.V.
COMPUTER MODELING OF COMPANY EMPLOYEE CHURN USING MACHINE LEARNING AND PREDICTIVE ANALYTICS METHODS.....419
Filimonchuk T.V., Koltun Yu.M., Klymova I.M., Kholobok V.I.
MODEL OF THE MOBILE APPLICATION DEVELOPMENT APPROACH.....429
Tsybulnyk S.O., Shelemakha V.V.
OVERVIEW OF METHODS AND MEANS OF AUTOMATED DETERMINATION OF THE CURRENT STATE OF BIOLOGICAL OBJECTS.....437
Chyhur I.I., Lahoida A.I., Chyhur L.Ya.
INFORMATIVE SIGNS IN THE TASKS OF AUTOMATED RECOGNITION OF TECHNOLOGICAL SITUATIONS DURING WELL DRILLING.....445
Chizhmotria O.V., Vakaliuk T.A., Klymenko D.O., Chizhmotria O.H., Hordienko I.V.
RESEARCH OF WEB TRAFFIC STRUCTURE AND DETERMINATION OF DECISIVE CHARACTERISTICS OF REQUESTS.....451
Yaretska N.O., Ramskyi A.O., Poplavska O.A.
FEATURES OF TEACHING HIGHER MATHEMATICS TO STUDENTS OF THE FIELD OF KNOWLEDGE “INFORMATION TECHNOLOGY” IN THE CONDITIONS OF DIGITALIZATION OF EDUCATION.....457

CONSTRUCTION

Dereviahina N.I., Sadovenko I.O.
EXPERIMENTAL RESEARCH OF THE CHANGES IN THE STABILITY PARAMETERS OF LOESS MASSIFS UNDER EXPLOSIVE LOADS.....464

INFORMATION ABOUT AUTHORS.....470

ГАЛУЗЕВЕ МАШИНОБУДУВАННЯ

УДК 631.3
DOI https://doi.org/10.32782/2663-5941/2025.3.2/01

Сташевський В.М.
Інститут механіки та автоматики агропромислового виробництва
Національної академії аграрних наук України

СТРАТЕГІЧНІ ОРІЄНТИРИ ІННОВАЦІЙНОГО РОЗВИТКУ ТЕХНІКО-ТЕХНОЛОГІЧНОГО ЗАБЕЗПЕЧЕННЯ АГРОПРОМИСЛОВИХ ПІДПРИЄМСТВ

У статті обґрунтовано стратегічні орієнтири та проаналізовано ключові проблеми інноваційного розвитку техніко-технологічного забезпечення виробництва продукції рослинництва в умовах переходу до сталого аграрного розвитку. Теоретичною та методологічною основою дослідження стали наукові напрацювання вітчизняних і зарубіжних учених, які займалися проблемами інноватики, модернізації технічної бази сільськогосподарських підприємств, а також було використано низку наукових методів: монографічний, абстрактно-логічний, методи аналізу й синтезу, SWOT та TOWS аналіз.

На основі SWOT аналізу визначено слабкі і сильні сторони мікросередовища модернізації техніко-технологічної бази сільськогосподарських підприємств для переходу до сталої моделі розвитку агровиробництва, окреслено можливості та загрози впровадження інновацій в рослинництві. За допомогою TOWS аналізу запропоновано шляхи подолання слабких сторін та ризиків освоєння інновацій агровиробниками. Встановлено, що питання оцінки ефективності технологічних та технічних нововведень в умовах переходу на модель сталого розвитку є багатокритеріальним і потребує розроблення теоретико-методичних підходів, які врахують сучасні вимоги до ведення сільськогосподарського виробництва, а саме: ресурсо- та енергоощадність, екологічність, людиноцентричність, прибутковість. Виділено проблемні фактори макросередовища та визначено державницькі стратегічні орієнтири стимулювання інноваційного розвитку техніко-технологічної бази сільськогосподарських підприємств для забезпечення їх конкурентоспроможності на світових ринках продовольства. Результати дослідження підтверджують, що формування сприятливого інституційного та економічного середовища через державну підтримку агроінноваційної діяльності є ключовою передумовою сталого розвитку сільського господарства України, підвищення продуктивності, ефективності та конкурентоспроможності продукції рослинництва.

Обґрунтовано необхідність оновлення машинно-тракторних парків агропідприємств і розроблення багатокритеріальних моделей оцінки впливу інноваційної техніки на енергетичні, екологічні, соціальні й економічні показники агровиробництва.

Ключові слова: *сільське господарство, машино-тракторний парк, оновлення, модернізація, інноваційний розвиток, теоретико-методичні підходи, стратегія, SWOT та TOWS аналіз.*

Постановка проблеми. Перспективи входження України в європейський простір зумовлюють необхідність розвитку вітчизняного сільського господарства високого технологічного укладу на принципах сталого розвитку, що забезпечить конкурентоспроможність та ефективне ведення аграрного бізнесу на основі інноваційних технологій виробництва продукції та сучасного машинно-тракторного парку для їх реалізації [1].

Аналітичні дослідження динаміки забезпеченості сільськогосподарською технікою та її оновлення в аграрних підприємствах України, що лише агрохолдинги із земельним банком більше 50 тис. га мають належний інвестиційний потенціал для імплементації досягнень науково-технічного прогресу у виробництво продукції рослинництва та оновлення машино-тракторного парку на інноваційній основі.

Статистика придбання техніки демонструє стрімкий спад купівельної спроможності вітчизняних агропідприємств із земельним банком менше 10 тис. га. Несвоєчасне оновлення та осучаснення технічного забезпечення вирощування сільськогосподарських культур призводить до збільшення рівня зносу засобів виробництва і витрат на їх поточний та капітальний ремонт.

Крім того, моральне старіння технічних засобів обмежує можливість реалізації інноваційних технологій землеробства, яка відповідають принципам сталого розвитку. Відтак, проблема ресурсо-інвестиційного забезпечення оновлення техніко-технологічної бази рослинництва на інноваційній основі є безперечно актуальною.

Аналіз останніх досліджень і публікацій. Питанню інноваційного розвитку матеріально-технічної бази сільського господарства в Україні присвячено роботи таких вчених, як: В. В. Адамчука, М. І. Грицишина, Ю. О. Лупенка, Я. К. Білоуська, Ю. О. Лупенка, В. В. Іванишина, О. В. Захарчука, М. Г. Михайлова та ін. [2–7].

Однак, незважаючи на глибину та різноплановість здійснених досліджень, наукові пошуки потребують узагальнення й урахування нових інструментів інноваційного розвитку техніко-технологічного забезпечення сільськогосподарських підприємств в сучасних умовах господарювання.

Постановка завдання. Метою статті є обґрунтування стратегічних орієнтирів та проблематики інноваційного розвитку техніко-технологічного забезпечення виробництва продукції рослинництва.

Виклад основного матеріалу. Теоретичною і методологічною основою досліджень були положення і розробки вітчизняних і зарубіжних вчених, що займались дослідженнями проблем інноватики, формування та використання техніко-технологічної бази сільськогосподарських підприємств. Також використано монографічний, абстрактно-логічний, метод аналізу і синтезу, SWOT та TOWS аналіз.

Перехід на якісно нову модель агробізнесу в Україні вимагає всебічного аналізу переваг і недоліків, можливостей і ризиків інноваційно-активного розвитку сільськогосподарських підприємств для пошуку внутрішніх резервів та розроблення державної стратегії для інтенсифікації оновлення техніко-технологічної бази аграрного виробництва.

Впровадження інноваційних техніко-технологічних рішень в рослинництві має ряд переваг та недоліків, які нами було систематизовано за допомогою SWOT аналізу (табл. 1).

Таблиця 1

SWOT-аналіз інноваційно-орієнтованого розвитку техніко-технологічного забезпечення сільськогосподарських підприємств

Внутрішні фактори	
S – сильні сторони	**W – слабкі сторони**
– Підвищення продуктивності машино-тракторних агрегатів – Зменшення тривалості виконання технологічних операцій у рамках агротехнічних вимог – Зменшення втрат урожаю – Оперативний контроль та аналіз витрат палива і якості виконання технологічних операцій – Зменшення витрат технологічних матеріалів – Оптимізація складу машинно-тракторного парку комфортні і безпечні умови праці персоналу	– Високий рівень зношеності матеріально-технічної бази господарств – Низький рівень інвестиційних можливостей агропідприємств – Відсутність кадрового забезпечення відповідної кваліфікації – Ризик збільшення експлуатаційних витрат при застосуванні дороговартісної техніки за її недостатнього навантаження – Необхідність укрупнення посівних площ, що негативно вплине на структуру сівозміни
Зовнішні фактори	
O – можливості	**T – загрози**
– Біологізація виробництва, застосування ґрунтозахисних технологій, перехід до екологічно-безпечного виробництва – Підвищення родючості ґрунту – Збільшення обсягів виробництва продукції рослинництва – Зменшення собівартості продукції – Підвищення енергоефективності виробництва – Отримання конкурентних переваг на ринку продовольства – Економічна оптимізація виробництва – Підвищення прибутковості агробізнесу – Державна програма часткового відшкодування вартості сільськогосподарської техніки вітчизняного виробництва – Стимулювання персоналу до підвищення кваліфікації	– Відсутність на ринку техніки вітчизняного виробництва для забезпечення інноваційних технологій вирощування сільськогосподарських культур – Висока вартість нових технічних засобів іноземного виробництва – Обмеженість кредитних, лізингових програм – Наявність аграрних та фінансових ризиків пов’язаних із війною – Низький рівень технічного сервісу – Утруднена логістика та низькі ціни реалізації сільськогосподарської продукції на зовнішні ринки продовольства через війну – Обмежена можливість довготривалого зберігання чи глибокої переробки агропродукції – Плинність кадрів

Впровадження інноваційних технічних рішень забезпечує ряд оперативних переваг, зокрема:

– Скорочення термінів виконання технологічних операцій. Завдяки більшій продуктивності сучасних машинно-тракторних агрегатів виробничий цикл значно пришвидшується, що сприяє своєчасному проведенню агротехнічних заходів і дозволяє уникнути потенційної втрати врожаю в умовах тісного сезонного календаря.

– Оперативний контроль та моніторинг. Впровадження автоматизованих систем дозволяє в режимі реального часу відслідковувати витрати технологічних матеріалів і палива, що підвищує ефективність управління виробничими процесами та сприяє оптимізації бюджету господарства.

– Оптимізація номенклатури техніки. Використання багатоопераційних сільськогосподарських агрегатів зменшує потребу у великому парку спеціалізованої техніки, знижуючи експлуатаційні витрати та спрощуючи процес обслуговування і ремонту.

– Покращення умов праці операторів. Автоматизація і роботизація виробничих процесів забезпечує безпечні та комфортні умови праці, що позитивно впливає на продуктивність та знижує ризик помилок, пов'язаних із людським фактором.

У довгостроковій перспективі інноваційно-орієнтований розвиток техніко-технологічного забезпечення агровиробництва відкриває можливості для впровадження сучасних ґрунтозахисних та еколого-орієнтованих технологій, а також систем точного землеробства:

– Збереження та відновлення родючості ґрунтів. Точне землеробство допомагає раціонально використовувати насіннєвий матеріал і добрива, що сприяє збереженню природного балансу та запобігає виснаженню ґрунту.

– Економія ресурсів та зниження негативного впливу на довкілля. Впровадження новітніх технологій дозволяє зменшити питомі витрати палива та знизити антропогенне навантаження, сприяючи екологічно безпечному виробництву.

– Підвищення врожайності та прибутковості. Інноваційні методи ведення господарства сприяють стабільному приросту продуктивності сільськогосподарських культур, що безпосередньо відображається на фінансовій ефективності агробізнесу.

З іншого боку інноваційний розвиток техніко-технологічної бази сільськогосподарського виробництва стримують такі внутрішні фактори як

– Високий рівень зносу існуючої техніки. Застаріла матеріально-технічна база не здатна забезпечити ефективне впровадження новітніх технологій.

– Обмежені інвестиційні можливості та низька купівельна спроможність. Недостатність фінансових ресурсів перешкоджає закупівлі сучасної техніки та впровадженню нових технологій.

– Кадровий дефіцит. Бойові дії в Україні спричинили «кадровий голод», що ускладнює формування кваліфікованих команд для роботи з інноваційною технікою.

– Ризик збільшення експлуатаційних витрат. Використання дороговартісної техніки при недостатньому навантаженні може призвести до збільшення витрат на експлуатацію, знижуючи рентабельність виробництва.

– Необхідність укрупнення посівних площ. Це може негативно вплинути на структуру сівозміни та збільшити навантаження на ґрунти, що обмежує можливості впровадження деяких інновацій.

Успішна реалізація інноваційного проєкту також залежить від низки зовнішніх факторів, які можуть як стимулювати впровадження нових технологій, так і створювати додаткові виклики:

– Макроекономічна нестабільність і валютні коливання. Світові коливання цін, інфляція та непередбачувані зміни курсу валют можуть призвести до різкого зростання витрат на закупівлю сучасної техніки та матеріальних ресурсів.

– Політичні та регуляторні ризики. Нестабільність у законодавстві, недостатня підтримка державних програм стимулювання агроінновацій та бюрократичні перешкоди створюють невизначеність для господарств, які прагнуть модернізувати свою технічну базу.

– Екологічні виклики та зміни клімату. Екстремальні погодні умови, зміни кліматичних режимів і збільшення вимог до екологічної безпеки можуть вимагати додаткових витрат на адаптацію агротехнологій до нових умов виробництва.

– Міжнародна конкуренція та доступ до технологій. Конкуренція з боку транснаціональних корпорацій, що мають доступ до інноваційних технологій і фінансових ресурсів, змушує національні аграрні підприємства постійно шукати шляхи для підвищення своєї конкурентоспроможності.

– Наслідки глобальних криз. Економічні, політичні або соціальні кризи можуть спричи-

нити зниження інвестиційної активності та обмеження ресурсних потоків, негативно позначаючись на можливості впровадження інновацій.

Комплексний підхід до впровадження інноваційних технічних рішень у сільському господарстві забезпечує як оперативні вигоди, так і перспективу сталого розвитку галузі. Проте для успішної реалізації інноваційного проєкту необхідно враховувати не лише внутрішні виклики агропідприємств, а й зовнішні чинники, що впливають на інвестиційний клімат і операційну стабільність. Збалансована політика державної підтримки, розвиток регуляторного середовища та стратегічне партнерство між науково-дослідними установами і виробничими підприємствами дозволять не тільки мінімізувати ризики, але і створити передумови для інтеграції сучасних агротехнологій у виробничий процес, що сприятиме підвищенню конкурентоспроможності українського сільського господарства на світових ринках.

У матриці TOWS розглянуто комбінації факторів, виявлених під час SWOT-аналізу та визначено стратегічні орієнтири для підсилення слабких сторін і нівелювання загроз (табл. 2).

За результатами здійсненого аналізу визначено, що для інноваційно-орієнтованого розвитку техніко-технологічного забезпечення рослинництва на рівні сільськогосподарських підприємств необхідно вирішити першочергові завдання:

– З допомогою консалтингових чи дорадчих організацій здійснити аналіз перспективних технології вирощування сільськогосподарських культур, які відповідають ґрунтово-кліматичним та виробничим умовам підприємства і принципам сталого розвитку.

– Для переходу на новітні технології вирощування продукції рослинництва модернізувати та оптимізувати склад машинно-тракторного парку на основі комплексної оцінки економічної, енергетичної, екологічної та соціальної ефективності впровадження нових техніко-технологічних рішень.

– Підвищити кадровий потенціал підприємства через профільне навчання спеціалістів, фахівців та керівників з метою формування у них знань, умінь і навичок, необхідних для успішного освоєння техніко-технологічних інновацій.

– Диверсифікувати джерела інвестиційних ресурсів на оновлення машинно-тракторного парку завдяки участі в державній програмі часткової компенсації вартості вітчизняної техніки, в лізингових програмах іноземних фірм-

Таблиця 2

Розробка стратегічних орієнтирів за допомогою TOWS-аналізу техніко-технологічної бази з позицій інноваційного розвитку

TOWS	O- Можливості	T – Загрози
S – Сильні сторони	**SO** – Перехід до систем точного землеробства та екологічно-безпечного виробництва – Оптимізація складу машинно-тракторного парку – Впровадження ІТ-технологій управління виробничим процесом – Економічна оптимізація виробництва – Підвищення рівня умов життя та праці персоналу	**SжT** – Аналіз та впровадження перспективних інноваційних технологій вирощування сільськогосподарських культур – Комплексна оцінка ефективності новітніх технічних засобів з метою оновлення МТП на інноваційній основі – Вертикальна інтеграція виробництва у напрямку поглиблення переробки продукції рослинництва
W – Слабкі сторони	**WO** – Участь в державних програмах стимулювання агробізнесу – Підвищення рівня кваліфікації інженерно-технічного персоналу та керівників середньої ланки – Впровадження комп'ютеризованих систем моніторингу, аналізу та управління виробництвом	**WT** – Організація ремонтних ділянок та впровадження інноваційних технологій технічного сервісу, відновлення робочих органів та вузлів сільськогосподарської техніки, для підвищення її надійності і довговічності – Залучення дорадчих служб для впровадження інновацій у виробництво Надання послуг механізованих робіт високовартісною сільськогосподарською технікою – Страхування посівів і техніки

виробників техніки, залучення кредитних коштів за державною програмою здешевлення кредитів для АПК тощо.

В даному переліку найбільш багатокритеріальним і дискусійним є питання оцінки ефективності технологічних та технічних нововведень в умовах переходу на модель сталого розвитку, тому потребує розроблення теоретико-методичних підходів, які врахують сучасні вимоги до ведення сільськогосподарського виробництва, а саме: ресурсо- та енергоощадність, екологічність, людиноцентричність, прибутковість. На кінцевий вибір новітнього технічного засобу суттєво впливають його вартість, рівень технологій конструювання, продуктивність, енергоощадність, екологічність, адаптованість до виробничих умов, надійність.

Стратегічними орієнтирами державного рівня є усунення зовнішніх загроз інноваційного оновлення техніко-технологічної бази сільськогосподарських підприємств, а саме:

– Стимулювання відродження українського машинобудування для АПК на інноваційній основі.

– Розширення фінансових механізмів підтримки аграріїв для оновлення машинно-тракторних парків.

– Започаткування державних програм стимулювання впровадження новітніх технологій в землеробстві: ґрунтозахисних, еколого- та кліматично-орієнтованих, ресурсо- та енергоощадних і т.д.

– Розроблення та упровадження регіональних програм розвитку підприємств переробки та зберігання продукції рослинництва.

Висновки. Формування сприятливого макроекономічного середовища для активізації інноваційного розвитку техніко-технологічного забезпечення агропромислових підприємств через системну державну підтримку агроінноваційної діяльності в Україні забезпечить сталий розвиток сільського господарства, збільшення обсягів виробництва та конкурентоспроможність продукції рослинництва в довгостроковій перспективі.

Вирішення тактичних та стратегічних завдань прискорення інвестиційно-інноваційних процесів в агропідприємствах дають можливість мобілізувати внутрішні ресурси для оновлення технологій землеробства, машинно-тракторних парків для їх реалізації, та переходу на принципово нову модель господарювання.

Питання багатокритеріальної оцінки ефективності оновлення машинно-тракторного парку сільськогосподарських підприємств передбачає визначення впливу впровадження інноваційної техніки на рівень енергетичної, екологічної, соціальної та економічної ефективності агровиробництва в сучасних умовах господарювання та потребує розроблення нових теоретико-методичних підходів.

Список літератури:

1. Перепелиця Н. М., Сташевський В. М. Техніко-технологічний рівень вітчизняного аграрного виробництва в контексті євроінтеграційних процесів. / *Механіка та автоматика агропромислового виробництва* – 2024. № 4 (118). С. 178–186. https://doi.org/10.37204/2786-7765-2024-1-20

2. Адамчук В. В. Матеріально-технічна база галузі рослинництва України: стан та перспективи розвитку/ В. В. Адамчук, М.І. Грицишин, Н. М. Перепелиця / Механізація та електрифікація сільського господарства. 2015. № 2 (101). С. 246–254.

3. Адамчук В. В., Грицишин М. І., Перепелиця Н. М. Техніко-технологічне забезпечення як основний фактор сталого розвитку агропромислового виробництва. Механізація та електрифікація сільського господарства. 2018. Вип. 8 (107). С. 198–208.

4. Матеріально-технічне забезпечення сільського господарства України : посібник / Ю. О. Лупенко, О. В. Захарчук, О. В. Вишневецька та ін. К. : ННЦ «ІАЕ», 2015. 144 с.

5. Білоусько Я. К. Удосконалення техніко-технологічного оснащення аграрного виробництва / Я. К. Білоусько, В. Л. Товстопят. – К.: ННЦ ІАЕ, 2012, 59 с.

6. Михайлов М. Г. Інноваційно-інвестиційний розвиток матеріально-технічної бази аграрних підприємств. К. : ННЦ «ІАЕ», 2018. 312 с.

7. Іванишин В. В. Інноваційне забезпечення оновлення техніко-технологічної бази підприємств АПК. Економіка АПК. 2010. № 1. С. 128–133.

Stashevskyi V.M. STRATEGIC GUIDELINES FOR INNOVATIVE DEVELOPMENT OF TECHNICAL AND TECHNOLOGICAL SUPPORT OF AGRO-INDUSTRIAL ENTERPRISES

The article substantiates strategic guidelines and analyses key problems of innovative development of technical and technological support of crop production in the context of transition to sustainable agricultural development. The theoretical and methodological basis of the study are the scientific achievements of domestic and foreign scientists who have dealt with the problems of innovation, modernisation of the technical base of agricultural enterprises, and a number of scientific methods were used: monographic, abstract and logical, methods of analysis and synthesis, SWOT and TOWS analysis.

On the basis of the SWOT analysis, the article identifies the strengths and weaknesses of the microenvironment of the modernisation of the technical and technological base of agricultural enterprises for the transition to a sustainable model of agricultural production development, and outlines the opportunities and threats for the introduction of innovations in crop production. The TOWS analysis is used to suggest ways of overcoming the weaknesses and risks of innovation development by agricultural producers. It has been established that the issue of assessing the effectiveness of technological and technical innovations in the context of transition to a sustainable development model is multi-criteria and requires the development of theoretical and methodological approaches that take into account modern requirements for agricultural production, namely: resource and energy efficiency, environmental friendliness, human centricity and profitability. The problematic factors of the macro-environment are allocated and the state strategic guidelines for stimulating the innovative development of the technical and technological base of agricultural enterprises to ensure their competitiveness in the world food markets are defined.

The results of the study confirm that the formation of a favourable institutional and economic environment through state support for agro-innovation is a key prerequisite for the sustainable development of agriculture in Ukraine, increasing the productivity, efficiency and competitiveness of crop production. The article substantiates the need to update the machine and tractor fleets of agricultural enterprises and to develop multi-criteria models for assessing the impact of innovative machinery on energy, environmental, social and economic indicators of agricultural production.

Key words: *agriculture, machine and tractor fleet, renewal, modernisation, innovative development, theoretical and methodological approaches, strategy, SWOT and TOWS analysis.*

УДК 681.2.08
DOI https://doi.org/10.32782/2663-5941/2025.3.2/02

Топчій Н.В.
Український науково-дослідний інститут спеціальної
техніки та судових експертиз Служби безпеки України

ПРОБЛЕМА ВПРОВАДЖЕННЯ КОНЦЕПЦІЇ НЕВИЗНАЧЕНОСТІ ВИМІРЮВАНЬ В МЕТОДИ КОНТРОЛЮ

Станом на сьогодні концепція невизначеності вимірювань прийнята міжнародними метрологічними і випробувальними організаціями. Невизначеність вимірювань – параметр, який відноситься до результатів вимірювання і характеризує розкид значень, які можна обґрунтовано приписати вимірюваній величині. Загальні правила оцінювання і вираження невизначеності вимірювань встановлює ДСТУ ISO/IEC Guide 98-3:2018. Оскільки при виконанні вимірювань і випробувань в різних галузях діють свої специфічні вимоги (особливості) випробувальні і калібрувальні лабораторії (акредитовані в національній системі) повинні відповідати вимогам стандарту ДСТУ EN ISO/IEC 17025:2019, в якому згідно п. 7.6 «лабораторії випробувальні/калібрувальні повинні оцінювати невизначеність вимірювань». Слід звернути увагу, що вищевказаний стандарт застосовується не тільки з метою акредитації випробувальних лабораторій національними агентствами з акредитації, а, і для організації роботи в самій лабораторії. Тому на міжнародному рівні випробувальні лабораторії часто декларують відповідність вимогам цього стандарту і на практиці дійсно постійно виконують його вказівки з метою підтвердження якості виконання вимірювань/випробувань у лабораторії [7].

Стаття присвячена обґрунтуванню необхідності оцінювання невизначеності вимірювання при контролюванні якості продукції. Встановлено паралель між «похибкою вимірювання» і «невизначеністю вимірювання». Проведено аналіз нормативних документів на методи контролю і показано наявність або відсутність в цих документах методики оцінювання невизначеності. Окреслено необхідність розробки методики вимірювання із урахуванням вимог ДСТУ ISO/IEC Guide 98-3:2018 на прикладі визначення твердості металів за методом Брінелля.

Ключові слова: *невизначеність вимірювань, похибка вимірювань, оцінювання невизначеності, засоби вимірювальної техніки, методика вимірювання.*

Постановка проблеми. В умовах сьогодення вираження результатів вимірювань і контролю з урахуванням невизначеності стало нормою на міжнародному рівні. Відомо, що результат будь-яких вимірювань є тільки апроксимацією або оцінкою значень вимірюваної величини і не може бути істинним без визначення невизначеності. Всі вимірювання недосконалі і мають багато потенційних джерел і варіацій. Саме оцінка невизначеності надає інтервал значень, в якому істинне значення, як вважають, визначене, із вказаною вірогідністю, а отже і є кількісним показником надійності вимірювань. Тому виникає актуальна проблема впровадження всеосяжної концепції невизначеності вимірювань при контролі якості продукції та матеріалів в основі оцінювання якої лежать фундаментальні твердження теоретичної метрології, які полягають в поетапному сходженні від абстрактного до конкретного.

Аналіз останніх досліджень і публікацій. У останні роки в теоретичній та практичній метрології виникають зміни, які пов’язані із введенням поняття «невизначеність» для оцінювання достовірності результатів вимірювань та засобів вимірювальної техніки на міжнародному рівні. Цими питаннями займалися і займаються вітчизняні вчені І.П. Захаров, Д.М. Нестерчук, С.О. Квітка, С.В. Галько, які розглядають теорію невизначеності вимірювання у відповідності з вимогами «Керівництва з оцінки невизначеності», необхідними поясненнями, математичними викладами [2, 3]. А. Коцюба підходить до цієї проблеми з точки зору фізичних явищ і пропонує не проводити повну ідентифікацію усіх помітних джерел невизначеності, що значно спрощує обчислення кінцевого результату, тобто розширеної невизначеності [4]. На жаль, концепція невизначеності має досить повільне впровадження у метрологічну практику, тому що при проведенні деяких метрологічних робіт зберігається використання поняття «похибка». І.Ф. Василенко стверджує, що така подвійність оцінки гальмує впровадження концепції невизначеності і розтлумачує

розуміння різниці в поняттях похибки та невизначеності [6].

Постановка завдання. У 2003 році Україна приєдналася до «Угоди про взаємне визнання національних еталонів і сертифікатів калібрування та вимірювання, які видаються національними метрологічними організаціями (CIPM MRA)», що зумовило перехід до нової системи забезпечення єдності вимірювання. Без оцінки невизначеності неможливо порівнювати результати вимірювань, а отже і приймати рішення щодо ствердження того, що вимірювана вами продукція (деталь, матеріал або речовина) знаходиться в межах допуску [12].

Метою статті є особливості переходу до вимірювань в концепції невизначеності та обґрунтування необхідності розробки методики оцінювання невизначеності вимірювання у кожному конкретному випадку.

Виклад основного матеріалу. Проблема забезпечення точності вимірювань, контролю і випробувань існувала завжди. Необхідно відмітити, що у 1978 році була створена робоча група в Міжнародному бюро мір і ваг для вироблення позиції рішення питань щодо неузгодженості результатів вимірювань в різних країнах [1]. За результатами роботи цієї групи була розроблена Рекомендація INC-1 (1980 р.) «Вираження експериментальних невизначеностей» основною ідеєю якої була відмова від поняття «похибка». А згодом, у 1993 році рядом міжнародних організацій було розроблено «Керівництво з оцінки невизначеності» (GUM) ДСТУ РМГ 43:2006 [11], застосування якого відразу ж після видання отримало статус неформального міжнародного стандарту і внесло узгодженість в оцінювання точності результатів у вигляді концепції «невизначеності вимірювань». У нас в країні концепція невизначеності впроваджується недостатніми темпами в усіх галузях промисловості, незважаючи на існування ДСТУ ISO/IEC Guide 98-3:2018 [8]. На жаль у багатьох організаціях продовжує використовуватись концепція «похибки вимірювання». Для розуміння необхідно провести паралель між «невизначеністю» та «похибкою». Обидві концепції мають загальну мету – кількісно охарактеризувати результат вимірювання з точки зору його точності. В обох концепціях простежується єдина схема оцінки характеристик «похибки» і «невизначеності» вимірювання, а саме: аналіз вимірювальної задачі і рівняння вимірювання; виявлення усіх джерел «похибок» («невизначеності») результату вимірювання; введення поправок на всі відомі систематичні ефекти; оцінювання характеристик складових і нарешті підрахунок результату вимірювання (розширена невизначеність).

В «Керівництві з оцінки невизначеності» замість поняття «похибка» вводиться поняття «невизначеність» вимірювання і трактується у двох сенсах:

– в широкому сенсі як «сумнів» відносно достовірності результату вимірювання (наприклад сумнів в тому, наскільки точно після внесення усіх поправок результат вимірювання представляє значення вимірюваної величини);

– у вузькому сенсі невизначеність вимірювання розуміється як параметр, пов'язаний з результатом вимірювання, котрий характеризує розкид значень, які можуть бути обґрунтовано приписані вимірюваній величині.

В концепції невизначеності цей параметр розуміється саме у вузькому сенсі і несе відомості щодо:

– забезпечення повної інформації про те, як складати звіти, що стосуються невизначеності вимірювання;

– надання підстави для міжнародного зіставлення результатів вимірювання;

– надання універсального методу для вираження і оцінювання невизначеності вимірювання застосовного до усіх видів вимірювання і до усіх типів даних, які при цьому використовуються.

Стандарт РМГ 43:2006 [11] розкриває відповідність між формами надання результатів вимірювання з використанням «похибки» і «невизначеності» вимірювання і рекомендує виражати характеристики точності вимірювань в показниках невизначеності, а не в показниках похибки, прийнятої у вітчизняній метрологічній практиці.

Причин введення концепції невизначеності вимірювання досить багато, основні з них наступні:

– поява нових (нетрадиційних) областей вимірювання (психологія, соціологія, медицина, тощо), де постулати традиційної метрології (фізична величина, одиниці вимірювання, міра, еталон, «похибка») не працюють;

– вплив нових наукових напрямків кібернетичного характеру (кібернетика, теорія інформації, математична статистика та ін.);

– відхід від поняття істинного значення вимірюваної величини як непізнаваної, через що поняття «похибка» втрачає сенс і її неможливо обчислювати, так як вона містить нікому не відоме справжнє значення;

– роздільна оцінка систематичних і випадкових «похибок» і використання для них різних характеристик (довірчий інтервал або середньо квадратичне відхилення).

Окрім того, застосування обох характеристик «похибок» (точкових та інтервальних) при визначенні результатів незручне, особливо при його подальшому застосуванні.

З метою впровадження концепції невизначеності вимірювання при контролі якості проаналізовано нормативні документи (НД) на методи контролю і випробувань металевих матеріалів і на вимоги до точності вимірювання, а саме: похибки; метрологічної простежуваності; відтворюваності, невизначеності. Результат аналізу показав, що у більшості діючих НД не прописані вимоги та методики оцінювання невизначеності вимірювання. Фрагмент аналізу НД приведений в таблиці 1.

Необхідно зауважити, що будь-які методи та методики вимірювання повинні включати вимоги до точності вимірювання або процедури контролю точності результатів.

Процес оцінювання невизначеності вимірювання складається з наступних етапів:

– *формування вимірювальної задачі*, яка включає: завдання вимірюваної величини *y*; визначення вхідних величин, які впливають на

Таблиця 1

№пп	ДСТУ	Назва ДСТУ	Вимоги до точності вимірювання			
			Похибка методу	Метрологічна простежуваність	Відтворюваність	Невизначеність
5.	ДСТУ ISO 6507-1:2007	Матеріали металеві. Визначення твердості за Віккерсом. Частина 1. Метод випробування (ISO 6507-1:2005, IDT)	+	-	-	-
6.	ДСТУ ISO 6508-1:2013	Металеві матеріали. Визначення твердості за Роквеллом. Частина 1. Метод випробування (шкали A, B, C, D, E, F, G, H, K, N, T) (ISO 6508-1:2005, IDT	+	-	-	-
7.	ДСТУ 8966:2019	Сталь. Металографічні методи визначення неметалевих включень- Вперше (зі скасуванням в Україні ГОСТ 1778-70)	-	-	-	-
8.	ДСТУ 8967:2019	Сталь. Методи визначення глибини зневуглецьованого шару	-	-	-	-
9.	ДСТУ 8968:2019	Прокат з конструкційної сталі високої оброблюваності різанням. Технічні умови-	-	-	-	-
10.	ДСТУ 8969:2019	Прокат тонколистовий зі сталі підвищеної міцності. Технічні умови.	-	-	-	-
11.	ДСТУ 8970:2019	Стрічка різана з холоднокатаного прокату. Технічні умови.	-	-	-	-
12.	ДСТУ 8971:2019	Прокат листовий холоднокатаний. Основні параметри і розміри.	-	-	-	-
13.	ДСТУ 8972:2019	Сталі та сплави. Методи виявлення та визначення величини зерна.	-	-	-	-
14.	ДСТУ 8974:2019	Сталь. Металографічний метод оцінювання мікроструктури листів та стрічки.	-	-	-	-
15.	ДСТУ 8975:2019	Сталь. Методи випробування та оцінювання макроструктури	-	-	-	-
16.	ДСТУ ISO/TS 21749:2013	Невизначеність вимірювання в метрологічній практиці. Повторні вимірювання та ієрархічні експерименти (ISO/TS 21749:2005, IDT)	-	+	+	+
17.	ДСТУ ISO 10012:2005	Системи керування вимірюванням. Вимоги до процесів вимірювання та вимірювального обладнання	-	+	+	+
18.	ДСТУ EN ISO/IEC 17025:2019	Загальні вимоги до компетентності випробувальних та калібрувальних лабораторій	-	+	+	+

величину *y*; створення моделі вимірювання; визначення розподілу ймовірностей вхідних величин;

– *визначення джерел невизначеності,* яке включає необхідність розгляду моделі вимірювання, та усіх параметрів, які входять в цю модель [2]. Також джерелами невизначеності можуть бути величини, які у явному вигляді не входять до моделі вимірювання, але впливають на результат (температура, вібрація, клас точності засобів вимірювальної техніки (ЗВТ), тощо).

Спосіб оцінювання невизначеності для отримання значення оцінки вихідної величини *y* і стандартної невизначеності *u(y)* полягає в наступному:

– визначення значень оцінок x_i вхідних величин X_i;

– визначення стандартних невизначеностей вхідних величин $u(x_i)$;

– визначення коефіцієнтів чутливості (впливу) c_i;

– визначення сумарної невизначеності;

– визначення розширеної невизначеності.

Розглянемо процес оцінювання невизначеності вимірювання на прикладі визначення твердості методом Брінелля на твердомірі NOVOTEST ТС-Б-Ц2. Твердість по Брінеллю обчислюють за формулою (1):

$$HB = \frac{2F}{\pi D(D - \sqrt{D^2 - d^2})} \qquad (1)$$

де *F* – прикладена сила Н; *D* – діаметр кульки, мм; *d* – діаметр відбитка, мм.

Для виявлення додаткових джерел невизначеності розглянемо вимоги до виконання вимірювань згідно метрологічних і технічних характеристик використовуваного твердоміра. Однією з основних вимог є проведення вимірювання при температурі 20^{+15}_{-10} °С і відсутність вібрації [10]. Отже джерелами невизначеності у нашому випадку є: прикладена сила натискання *F*, діаметри кульки та відбитку *D* та *d*, відхилення температури, наявність вібрації, похибка ЗВТ (твердоміра і мікроскопа).

Наступним етапом оцінювання є визначення стандартної невизначеності вхідних величин. Границя допустимої відносної похибки навантаження вказана в паспорті твердоміра NOVOTEST ТС-Б-Ц2 та складає ±1,0% від навантаження (будь-яка додаткова інформація відсутня). Як правило, оцінювання стандартної невизначеності зводиться до застосування апріорного закону розподілу. Тому використовуємо нормальний (рівномірний) розподіл і визначаємо стандартну невизначеність по типу *B*. В разі нормального розподілу стандартна невизначеність визначається за формулою (2)

$$u(F) = u(x_1) = \frac{a}{\sqrt{3}} \qquad (2)$$

де *a* – 1/2 ширини довірчого інтервалу, яка у нашому випадку дорівнює значенню абсолютної похибки навантажень.

При навантаження 3000Н абсолютна похибка навантажень складає 30Н. Відповідно, стандартна невизначеність навантаження дорівнює:

$$u(F) = u(x_1) = \frac{30}{\sqrt{3}} = 17{,}3205$$

Границі допустимих відхилень кульки від номінальних значень також вказані в паспорті на твердомір (в описанні типу приладу). Відповідно, також застосовуємо рівномірний розподіл і обчислюємо стандартну невизначеність відхилення кульки від номінальної за формулою (2), де в якості 1/2 ширини довірчого інтервалу *a* приймаємо граничне відхилення кульки від номінального значення. При застосуванні кульки діаметром 10 мм з граничним відхиленням ±0,005 мм, стандартна невизначеність складає

$$u(D) = \frac{0{,}005}{\sqrt{3}} = 0{,}0028.$$

Діаметр відбитка *d* вимірюють не менше трьох разів за допомогою мікроскопа з граничною похибкою ±5% (при застосуванні кульок діаметром 1.0; 2,0; 2,5 мм) і ±0,25% (при застосуванні кульок діаметром 5,0 і 10,0 мм) від діаметра кульки [8].

У випадку, якщо є масив експериментальних даних стандартну невизначеність визначаємо по типу А, тобто стандартна невизначеність вимірювання діаметра відбитка визначається за відомою формулою (3):

$$u(d) = u(x_3) = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^{n} \left(d_i - \bar{d}\right)^2} \qquad (3)$$

В ході експерименту нами були отримані наступні дані вимірювання відбитка: (4,08; 4,12; 4,12; 4,12;. 4,125) мм, середнє значення яких дорівнює

$\bar{d}$ = 4,113. D В результаті підрахунку отримуємо: $\underline{u}(d)$ = 0,0083

$$HB = \frac{2 \cdot 3000}{3{,}14 \cdot 10(10 - \sqrt{10^2 - 4{,}11^2})} = 215{,}9128$$

В разі багаторазового вимірювання відбитка стандартна невизначеність обумовлена похибкою відлікового мікроскопа по Брінеллю не враховується, так як вона увійшла в стандартну невизначеність типу *A*. Якщо вимірювання проводяться в нормальних умовах і за відсутністю вібрації, то стандартні невизначеності, які виникають внаслідок відхилення температури від номінального значення і наявності вібрації не обчислюються. В разі відхилень умов проведення вимірювань від номінальних необхідно провести додаткові дослідження із урахуванням температури та наявності вібрації.

Для визначення сумарної невизначеності враховують наявність кореляційного зв'язку між вхідними величинами, у нашому випадку існує кореляційний зв'язок між силою *F* і діаметром відбитка *d*.

Сумарна стандартна невизначеність визначається за формулою (4)

$$u_c^2(\text{HB}) = \sum_{i=1}^{N} c_i^2 u^2(x_i) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} C_F c_d\, u(x_F)\, u(x_d) r(x_{d,} x_F) \qquad (4)$$

де $u(x_i)$ – стандартна невизначеність від і-ї вхідної величини; $r(x_{d,} x_F)$ – ступінь кореляції між x_d і x_F; c_F і c_d – коефіцієнти чутливості; *N* – кількість вхідних величин.

Коефіцієнти чутливості, які визначаються як $c_i = \frac{\partial f}{\partial x_i}$, показують як змінюється вхідна оцінка через зміну вхідних оцінок $x_1, ..., x_N$ і дорівнюють:

$$C_F = \frac{\partial\left(\frac{2 \cdot F}{\pi D(D - \sqrt{D^2 - d^2})}\right)}{\partial F} = \frac{2}{\pi D(D - \sqrt{D^2 - d^2})},$$

$$C_D = \frac{\partial\left(\frac{2 \cdot F}{\pi D(D - \sqrt{D^2 - d^2})}\right)}{\partial D} = \frac{2F}{3{,}14 D^2 \sqrt{D^2 - d^2}},$$

$$C_d = \frac{\partial\left(\frac{2 \cdot F}{\pi D(D - \sqrt{D^2 - d^2})}\right)}{\partial d} = \frac{-2Fd}{3{,}14 D(D - \sqrt{(D^2 - d^2)^2}\sqrt{D^2 - d^2})}.$$

При обчисленні коефіцієнтів чутливості в якості d приймаємо середнє значення $\bar{d} = 4{,}113$, яке було розраховано раніше. В результаті обчислення отримуємо: $C_F = 0{,}0719$, $C_D = 2{,}0963$, $C_d = -110{,}0873$. Ступінь кореляції між $x_{d,}$ та x_F характеризується оцінкою кореляції і визначається як:

$$r(x_{d,} x_F) = \frac{u(x_d\, x_F)}{u(x_d)\, u(x_F)} \qquad (5)$$

Оцінку коваріації (міри спільної мінливості двох випадкових змінних) вхідних d і F з оцінками $\bar{d}$ і $\bar{F}$, отриманими повторними спостереженнями розраховуємо за формулою (6):

$$u(x_d, x_F) = \text{S}(\bar{d}, \bar{F}), \qquad (6)$$

де s($\bar{d}$, $\bar{F}$) знаходимо за формулою (7)

$$\text{S}(\bar{d}, \bar{F}) = \frac{1}{n(n-1)} \sum_{\kappa=1}^{n} (d_\kappa - \bar{d})(F_\kappa - \bar{F}) \qquad (7)$$

де d_κ і F_κ – результати вимірювання величин *d* і *F* відповідно, а $\bar{d}$ і $\bar{F}$ їх середнє значення.

В нашому випадку, при обчисленні $S(\bar{d}, \bar{F})$ за різність ($F_\kappa - \bar{F}$) приймаємо абсолютну похибку прикладених навантажень, тобто 30Н. Отримуємо: $u(x_d, x_F) = s(\bar{d}, \bar{F}) = 2{,}6645 \cdot 10^{-15}$,

$$r(x_{d,} x_F) = \frac{2{,}6645 \cdot 10^{-15}}{0{,}0083 \cdot 17{,}3205} = 1{,}8519 \cdot 10^{-14}.$$

Отже сумарна стандартна невизначеність при визначенні твердості з навантаженням 3000Н, кулькою діаметром 10 мм за методом Брінелля складає:

$$u_C^2(HB) = C_F^2 u^2(F) + C_D^2 u^2(D) + C_d^2 u^2(d) + 2\, c_F c_d\, u(x_F)\, u(x_d) r(x_{d,} x_F) = 2{,}4788,$$

відповідно:

$$u_\text{c}(\text{HB}) = \sqrt{2{,}4788} = 1{,}57$$

Розширена невизначеність вимірювання визначається за формулою (8):

$$U = k \cdot u_c, \qquad (8)$$

де *k* – коефіцієнт охоплення.

Коефіцієнт охоплення знаходять як коефіцієнт Стьюдента [2], з рівнем довіри, який приблизно дорівнює 95% при допустимому нормальному розподілі. В границях veff = ∞, k = 2. Тому U = 3,14. Результат вимірювання буде дорівнювати 215,9128 ± 3,14 при довірчій вірогідності P = 0,95.

Таким чином невизначеність вимірювання при визначенні твердості металу по Брінеллю твердоміра NOVOTEST ТС-Б-Ц2 при навантаженні 3000Н і діаметрі індентора 10 мм складає 3,13.

Висновки. Отже питання, які пов'язані з визначенням видів та методів оцінювання результатів вимірювання засобами вимірювальної техніки, відносяться до важливих науково-технічних задач, які є актуальними при гармонізації вітчизняних нормативних документів в галузі метрології та метрологічного забезпечення. Вищезазначене оцінювання невизначеності вимірювання при визначенні твердості металу за методом Брінелля показує, що існуючі алгоритми такого оцінювання є складним і трудомістким процесом. Вони передбачають виявлення джерел невизначеності, наявності кореляційної залежності між вхідним і вихідними величинами, визначення законів розподілу вхідних величин, коефіцієнтів чутливості, стандартної, сумарної та розширеної невизначеностей. В деяких випадках можуть виникати задачі дослідження стандартних невизначеностей від впливу додаткових факторів [5]. Відповідно, існує актуальна задача розробки методик оцінювання невизначеності вимірювань при контролі якості продукції і внесення процедури оцінювання невизначеності вимірювання в нормативні документи на методи контролю.

Список літератури:

1. Міжнародне бюро мір і ваг / Астрономічний енциклопедичний словник / за заг. ред. І. А. Климишина та А. О. Корсунь. Львів : Голов. астроном. обсерваторія НАН України : Львів. нац. ун-т ім. Івана Франка, 2003. С. 291. ISBN 966-613-263-X.
2. Захаров І.П. Теорія невизначеності у вимірюваннях – Харків: Консум, 2002.
3. Нестерчук Д.М., Квітка С.О., Галько С.В. Основи метрології та засобів вимірювання. Мелітополь: Видавничо-поліграфічний центр «Люкс», 2017.
4. Коцюба А.М. Оцінювання невизначеності вимірювання за результатами контрольних вимірювань з використанням стандартних зразків. URL: https://surl.li/texcbu (дата звернення 04.04.2025).
5. Оцінювання невизначеності вимірювання твердості ...Полтавська політехніка імені Юрія Кондратюка. URL: https://journals.nupp.edu.ua 'article' download (дата звернення 04.04.2025).
6. Василенко І.Ф. Вплив концепції невизначеності вимірювань на метрологічну діяльність Центральноукраїнський національний технічний університет. URL: https://surli.cc/fzcblh (дата звернення 24.04.2025).
7. Аналіз вимог. ДСТУ EN ISO/IEC 17025:2019 Вимоги до методик вимірювання та випробування. URL: https://www.ipkm.org.ua ' analiz-vimog-2-dstu-iso-iec-1 (дата звернення 15.04.2025).
8. ДСТУ ISO/IEC Guide 98–3:2018 Невизначеність вимірювань. Частина 3. Настанова щодо подання невизначеності у вимірюванні.
9. ДСТУ EN ISO/IEC 17025:2019 Загальні вимоги до компетентності випробувальних та калібрувальних лабораторій.
10. ДСТУ ISO 10012:2005 Системи керування вимірюванням. Вимоги до процесів вимірювання та вимірювального обладнання.
11. ДСТУ-Н РМГ 43:2006 Метрологія. Застосування. Посібники з вираження невизначеності вимірювань (РМГ 43–2001, IDT).
12. Інформація про участь у міжнародних, європейських та регіональних організацій з метрології стандартизації, сертифікації. Міністерство економіки України. URL: https://me.gov.ua › file (дата звернення 24.04.2025).

Topchii N.V. IMPLEMENTATION OF THE CONCEPT OF MEASUREMENT UNCERTAINTY INTO CONTROL METHODS

Nowadays, the concept of measurement uncertainty has been adopted by international metrological and testing organizations. Uncertainty of measurement is a parameter that relates to measurement results and characterizes the spread of values that can be reasonably attributed to the measured quantity. General rules for the evaluation and expression of measurement uncertainty are established by DSTU ISO/IEC Guide 98-3:2018. Since different industries have their own specific requirements (features) when performing measurements and tests, testing and calibration laboratories (accredited in the national system) must comply with the requirements of the DSTU EN ISO/IEC 17025:2019 standard. General rules for the evaluation and expression of measurement uncertainty are established by DSTU ISO/IEC Guide 98-3:2018. Since different industries have their own specific requirements (features) when performing measurements and tests, testing and calibration laboratories (accredited in the national system) must comply with the requirements of the

DSTU EN ISO/IEC 17025:2019 standard. Therefore, at the international level, testing laboratories often declare compliance with the requirements of this standard and in practice, they actually constantly follow its instructions in order to confirm the quality of measurements/tests performed in the laboratory. The article is devoted to the justification of the need to estimate measurement uncertainty when controlling product quality. A parallel is established between "measurement error" and "measurement uncertainty". An analysis of regulatory documents on control methods is carried out and the presence or absence of uncertainty estimation methods in these documents is shown. The need to develop a measurement methodology taking into account the requirements of DSTU ISO/IEC Guide 98-3:2018 is outlined using the example of determining the hardness of metals using the Brunel method.

Key words: *measurement uncertainty, measurement error, uncertainty assessment, measuring equipment, measurement methodology.*

UDC 629.463.65
DOI https://doi.org/10.32782/2663-5941/2025.3.2/03

Fomin O.V.
State University of Infrastructure and Technologies

Burlutskyi O.V.
Ukrainian State University of Railway Transport

Kubrak A.V.
National Transport University

Holub H.M.
National transport university

Fursina A.D.
National University Zaporizhzhia Polytechnic

NON-STATIONARY FACTORS OF FATIGUE RESISTANCE AND THEIR INFLUENCE ON THE SERVICE LIFE OF FREIGHT CAR JOINTS: MATHEMATICAL MODELING AND ANALYSIS

The article is devoted to mathematical modeling and analysis of the proposed modified version of the linear hypothesis of fatigue damage accumulation, this modification of the classical linear hypothesis (Palmgren-Mayer) allows more accurate prediction of the residual life of a structure. which uses a conditional value of the endurance limit determined by the coefficient of the lower limit of damaging stresses.

The linear hypothesis of fatigue damage accumulation, proposed by A. Palmgren (1924) and developed by M. Mayer (1939), is the basis for calculating the durability of materials under cyclic loading. Based on the analysis of the indirect characteristic of accumulated fatigue damage – the fracture energy of local models of freight car frame assemblies, the dependence of the endurance limit decrease on the proportion of the produced resource, which takes into account the dynamic decrease in the endurance limit during operation, is determined. The most significant disadvantages of the linear theory are that it does not describe the effect of the alternation of stresses of different levels.

Thus, for the considered design variants of freight car bearing assemblies, the application of the modified linear hypothesis using the calculated fatigue curves can provide a high accuracy of the predictive calculation of durability at a value of K = 0.735. In addition, a positive property of the proposed method is the direction of the error into the safety margin.

A computational justification of the lower limit of damaging stress is given. A modification of the linear hypothesis is proposed. It takes into account the nonlinearity due to the reduction of σ_{-1} and the introduction of σ_{min}. It is confirmed by energy and experimental data. Reduces the risk of service life overestimation by 30–50% compared to the classical approach.

Key words: *railway transport, rolling stock, car fleet, machine-building structures, fatigue damage, durability of bearing units of freight cars, mathematical description, mathematical models.*

Statement of the problem. Wear and damage to freight car joints remain a critical problem in rail transport, directly affecting traffic safety and economic efficiency of transportation. Increased requirements for freight turnover and speed of movement enhance the effect of non-stationary fatigue resistance factors, which leads to accelerated accumulation of damage in metal car structures.

Non-stationary factors of fatigue resistance include variable loads, vibrations, shocks, as well as corrosion processes, which can change their intensity and nature during the operation of the car. Their complex impact on the resource of connections is complex and requires detailed study. Traditional methods of fatigue assessment, based on constant loads, often turn out to be insufficiently accurate for predicting the durability of structural elements in real operating conditions.

Analysis of information sources on the topic under study. The article [1] investigates the

prediction of wheel wear of heavy freight cars based on tests on a roller stand and on the line. The authors consider the factors affecting wheel wear and develop models for its prediction. Experimental data obtained in laboratory and field conditions are used. The results of the study can be used to optimize wheel maintenance and improve the safety of railway transport. The article is relevant for specialists in the field of railway engineering and operation.

Article [2] presents a comparative analysis of the fatigue strength of a freight car frame. The authors investigate the influence of various design solutions on the fatigue strength of the frame. Methods for calculating stresses and strains are considered. The results of experimental studies and numerical modeling are analyzed. The results of the study can be used to optimize the frame design.

The scientific and applied work [3] investigates the determination of the durability of the supporting structure of the body of an open wagon made of round pipes during its transportation on a railway ferry. The authors analyze the influence of dynamic loads on the design of the wagon during sea transportation. Methods for calculating strength and fatigue durability are considered. The results obtained can be used to optimize the design of wagons and ensure the safety of their transportation. The article is important for design engineers and specialists in the field of transport engineering.

The authors of the publication [4] present a study of dynamic weighing to improve the safety of rail freight transport using the example of the Czech Republic. The authors analyze the effectiveness of using dynamic weighing systems to control the load on moving wagons. The advantages and disadvantages of this technology are considered. The impact of dynamic weighing on traffic safety and accident prevention is assessed. The results of the study emphasize the importance of implementing modern monitoring systems to ensure the safety of rail transport.

The article [5] investigates the determination of the dynamic load on a gondola car when it is attached to the ferry deck by a viscous coupling. The authors analyze the forces acting on the car during transportation on a ferry, taking into account the characteristics of the viscous coupling. Mathematical models and methods for calculating dynamic loads are considered. The results of the study can be used to design car fastening systems and ensure transportation safety. The article is important for specialists in the field of transport engineering.

In the article [6] the influence of the load from the axle of a gondola car on its dynamic performance and the railway track is investigated. The authors analyze the interaction of the car with the track at different load levels. The vertical and horizontal forces acting between the wheel and the rail are considered. The influence of stresses on the track and car elements is assessed. The results of the study can be used to optimize the design of cars and reduce track wear.

Document [7] is a UIC (International Union of Railways) standard and contains a test programme for wagons with steel frame and body construction and their cast steel bogies. The standard specifies test procedures and requirements to ensure the safety and reliability of wagons. The document is important for manufacturers, operators and regulatory authorities in the railway industry.

In the article [8] the dynamic calibration of the Weigh in Motion (WIM) system is considered. The authors investigate the methods of calibrating WIM systems to ensure the accuracy of measurements of the weight of vehicles in motion. Factors affecting the accuracy of measurements and methods of their compensation are considered. The results of the study contribute to improving the reliability and accuracy of WIM systems.

The article [9] considers the mathematical modeling of the thermal straightening process of the backbone beam of freight platform cars. The authors investigate the modeling of the thermal straightening process from the perspective of the non-stationarity of various factors.

The article [10] considers mathematical modeling of technological transitions in the production of bent profiles. The results of the study can be used to optimize production technologies.

The conducted literature analysis clearly demonstrates the significant attention of researchers to the problem of determining non-stationary factors of fatigue resistance in freight car joints. Mathematical modeling is a key tool for the quantitative assessment of these factors. Studies emphasize the complexity of predicting the resource of joints due to the variability of operational loads. The influence of non-stationary factors, such as dynamic shocks and vibrations, significantly reduces the service life of elements.

Research methods. To study non-stationary fatigue resistance factors, a mathematical modeling method will be used, which will allow creating a virtual model of a freight car joint and simulating the influence of various operating conditions.

Analysis of the stress-strain state of freight car joint elements under the action of non-stationary loads will be carried out using numerical methods, in particular the finite element method. In order to assess the influence of random load fluctuations on the accumulation of damage in the joint material, probabilistic and statistical methods and the theory of random processes will be used. Experimental studies of the fatigue strength of joint material samples under non-stationary loading conditions will be carried out to verify the results of mathematical modeling.

Object and subject of research. The object of research is the processes of damage accumulation and destruction of joints of load-bearing elements of freight car structures under conditions of non-stationary fatigue resistance factors.

The subject of the study is the regularities of the influence of non-stationary fatigue resistance factors (variable stress amplitudes, load frequencies, average stress levels) on the fatigue strength characteristics and predicted resource of freight car joints, determined on the basis of mathematical modeling and analysis.

Statement of the problem. The increasing intensity of operation of freight cars places increased demands on the reliability of their connections. One of the key problems is the accumulation of fatigue damage under the action of cyclic loads. Traditional methods of calculating the resource are often based on stationary assumptions about fatigue resistance factors. However, in real operating conditions, these factors, such as temperature, humidity and corrosive environment, are non-stationary and can change significantly. This leads to inaccuracies in predicting the residual resource of connections.

The non-stationarity of fatigue resistance factors determines a complex relationship between applied loads and the rate of damage accumulation. Changes in external conditions can both accelerate and slow down the fatigue failure process. Insufficient consideration of this non-stationarity can lead to premature failure of critical structural elements. Therefore, there is an urgent need to develop adequate methods for assessing the impact of non-stationary factors on the fatigue life of freight car joints.

Purpose and objectives of the study. The aim of this work is to develop a mathematical model that takes into account non-stationary fatigue resistance factors for predicting the service life of freight car joints. The study is aimed at quantitatively assessing the impact of variable loads and operating conditions on the accumulation of damage in key structural elements. The results obtained will allow to increase the accuracy of predicting the residual service life and optimize the maintenance strategies of the car fleet.

To achieve the set goal, the following tasks were solved:

- identify and classify the main non-stationary fatigue resistance factors affecting freight car joints during their operation;
- to develop mathematical models that describe the dynamics of changes in non-stationary fatigue resistance factors over time;
- to investigate the impact of each of the identified non-stationary factors on the accumulation of damage in key joints of freight cars;
- to create a mathematical model that comprehensively takes into account the influence of non-stationary fatigue resistance factors on predicting the service life of freight car joints;
- to conduct a numerical analysis of the developed model for different operating modes and parameters of freight cars;
- verify the adequacy of the developed mathematical model by comparing the modeling results with available experimental data or the results of other studies.

Outline of the main material of the study. The linear hypothesis of fatigue damage accumulation, proposed by A. Palmgren (1924) and developed by M. Mayer (1939), is the basis for calculating the durability of materials under cyclic loading. The hypothesis states:

$$\sum_{i=1}^{r} \frac{n_i}{N_i} = 1 \qquad (1)$$

$$\left(\sigma_{ai} \geq \sigma_{-1\partial}\right)$$

where n_i – number of cycles at voltage σ_i;

N_i – number of cycles to failure under stress σ_i.

For practical calculations, the Baskin equation [11] is used, which relates the durability N with the amplitude of stresses σ:

$$\sigma_a = \sigma_{f'} \cdot (2N)^b, \qquad (2)$$

σ_a – stress amplitude (MPa);

$\sigma_{f'}$ – endurance coefficient (MPa);

N – number of cycles to failure;

b is the fatigue strength index (-0.05 to -0.12 for metals).

Alternative form (for SN curve)

$$\sigma^m \cdot N = C, \qquad (3)$$

where m – power index (depends on the material, for steels $m \approx 3...8$);
C is the material constant.

For variable load with different amplitudes σ_i Baskin's equation integrates with the Palmgren-Mayer hypothesis:

$$\sum_{i=1}^{k} \frac{n_i}{N_i} = 1 \text{ або } \sum_{i=1}^{k} \frac{n_i \cdot \sigma_i^m}{C} = 1. \quad (4)$$

However, the real process of fatigue damage accumulation is nonlinear, due to the change in fatigue resistance characteristics, primarily the endurance limit, as the design life develops. Baskin equationaccording to the Palmgren-Mayer hypothesis: has a number of disadvantages, such as:

1. Linear damage accumulation – does not take into account the interaction of cycles.

2. Not suitable for low-cycle fatigue (large plastic deformations).

3. Does not take into account environmental influences (corrosion, temperature).

Taking this fact into account is the basis of the kinetic theory of fatigue by E.K. Pochtenny [12]. The kinetic theory of Pochechny provides a more realistic assessment of the resource, especially for critical structures, where traditional methods give non-conservative results. Its implementation requires additional experimental data, but allows to significantly increase the reliability of calculations. The generalization of experimental data of a number of researchers made it possible to establish the following:

– under regular loading, higher voltage has a greater damaging effect, as a result of which the intensity of the decrease in the endurance limit is higher in cases where higher voltage was applied;

– with irregular block loading, depending on the nature of the load alternation, "training" or strengthening effects appear, which reduce or increase the intensity;

– at the moment of failure of the structure, the ultimate value of the endurance limit is different from zero;

– the greatest damaging effect is caused by short-term overloads.

The nonlinearity of fatigue damage accumulation can be described by the Palmgren-Mayer hypothesis taking into account the kinetics of the endurance limit.

Henry's theory addresses the mechanisms of fatigue failure of materials, in particular the role of local plastic deformations in the process of damage accumulation. It complements the classical approaches (Baskin, Palmgren-Mayer) and the kinetic theory of Pochechny, emphasizing the micromechanical aspects of fatigue. The theory uses the plastic strain accumulation criterion [13]:

$$\Delta \in_p \cdot N_f^{\alpha} = C, \quad (5)$$

$\Delta\epsilon_p$ – amplitude of plastic deformation per cycle;
N_f – number of cycles to failure;
α, C are material constants (for steels $\alpha \approx 0.5...0.7$).

For stresses, the formula is similar to the Baskin equation, but taking into account plasticity:

$$\sigma_{Ri} = \sigma \frac{1 - \frac{n_i}{N_i}}{\frac{\sigma}{\sigma_R} - \frac{n_i}{N_i}} \quad (6)$$

Henry's theory fills the gap between macroscopic (Baskin) and microscopic (dislocation) fatigue models.

S.V. Serensens hypothesis is a development of classical approaches to the assessment of fatigue failure of materials, in particular the linear Palmgren-Mayer hypothesis. It takes into account the nonlinear nature of damage accumulation and the influence of additional factors, such as stress concentration, surface quality and residual stresses [14].

$$\frac{\sigma_{Ri}}{\sigma_R} = 1 - \left(\frac{n_i}{N_i}\right)\left(\frac{\sigma}{\sigma_R} - 1\right) \frac{1}{\frac{\sigma}{\sigma_R} - \left(\frac{n_i}{N_i}\right)^2} \quad (7)$$

where σ_{Ri} σ_R – endurance limits of the damaged and original materials, respectively; $\frac{\sigma}{\sigma_R}$ degree of overload; m, mi – characteristics of the slope of the left branch of the primary and secondary fatigue curves.

Hypothesisnonlinear accumulation of fatigue damage Marko – Starkey [15].

One of the first hypotheses of nonlinear damage accumulation was proposed by Marko and Starkey in 1954. It is based on the following propositions.

Damage curves for any amplitude value of symmetrical sinusoidal stresses can be described by the relations:

$$\sum \frac{n}{N} \approx \int_0^1 \frac{1 + \frac{N_1}{N_2} + \frac{N_1}{N_3} + ... + \frac{N_1}{N_i}}{1 + \frac{N_1}{N_2} r_2 D\left(\frac{r_2 - 1}{r_2}\right) + ... + \frac{N_1}{N_i} r_i D\left(\frac{r_i - 1}{r_i}\right)} dD, \quad (8)$$

In the article by E.K. Pochtenny [16], a modified approach to the assessment of fatigue

life of materials is proposed, which takes into account the decrease in the endurance limit in the process of damage accumulation. This method is a development of the classical linear Palmgren-Mayer hypothesis, but includes corrections for material degradation under the action of cyclic loads.

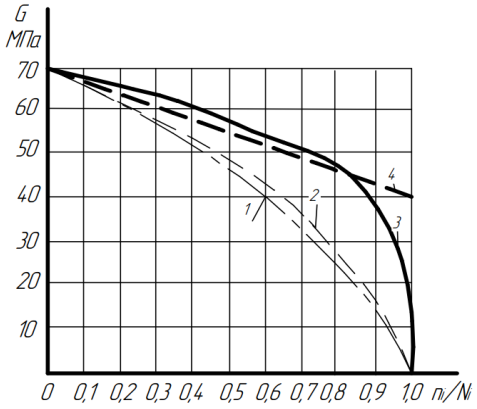
$$\sigma_{Ri} = \sigma_R \left(1 - \frac{n_i}{N_i}\right)^{1/m} \quad (9)$$

Regardless of the degree of overload, the value of the endurance limit at the moment of failure, calculated by equation (7), is zero, which is not true. Expression (8) in the range of overload degrees greater than two gives underestimated values of the endurance limit for fixed fractions of the developed resource. The parameter mi, which characterizes the accumulated fatigue damage, depends on the history of the load, and therefore the correct use of expression (9) requires substituting instead of it the functional dependence of the characteristic of the slope of the secondary curve on the fraction of the developed resource and the degree of overload, the finding of which in itself is a difficult task. The assumption in expression (10) about the invariance of the parameters NG and m makes it independent of the applied stresses, which is also unacceptable. Thus, none of the considered equations satisfies all the requirements formulated above.

In the adjusted linear hypothesis of B.P. Kogaev, to take into account the change in the endurance limit, it is recommended to adjust its initial value σR to the calculated value R (where K is the coefficient of the lower limit of damaging stresses, which is chosen a priori from the interval of empirical values of K from 0 to 1.0 [17]). The coefficient K depends on the intensity of the decrease in the endurance limit, which is determined for a specific structure by the spectrum of operational loads: the greater the proportion of the load amplitude spectrum is above the endurance limit, the smaller K.

Determining the equation of its reduction by direct experimental method requires a large amount of tests, which led to the hypothetical nature of the criteria proposed to describe the reduction of the endurance limit, the main of which are the criteria presented in (Fig. 1).

In the simplest case, it can be assumed that the abscissa of the fracture point NG and the characteristic of the angle of inclination of the left branch of the fatigue curve m are invariant. In this case, as the damage accumulates, the fatigue curve will fall down vertically, as shown in (Fig. 2).



1 – D. Henry; 2 – CB Sørensen;
3 – E.K. Pochtenny; 4 – Marko – Starkey

Fig. 1. Decrease in endurance limit during accumulation of fatigue damage, described by the criteria

Then, according to equation (1) for 0 < ni / Ni < 1, with ni / Ni = const, we can write:

$$\sigma^m \left(N_i - n_i\right) = \sigma_{Ri}{}^m N_G \quad (10)$$

Taking into account the assumption of invariance of m and NG, it follows from equation (10) that as the resource increases, the decrease in the endurance limit is described by dependence (9).

$$\sigma_{Ri} = \sigma_R \left(1 - \frac{n_i}{N_i}\right)^{1/m}$$

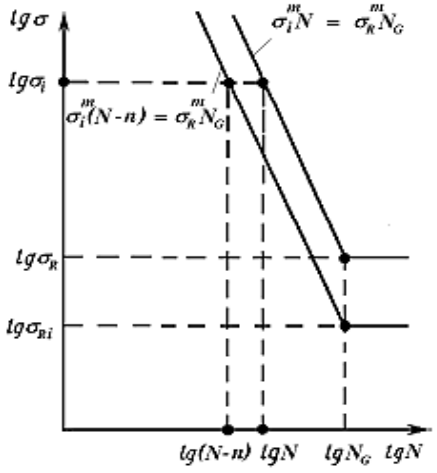


Fig. 2. Decreasing endurance limit as damage accumulates

Thus, under the above assumptions, the change in the endurance limit of the initial material is described by a decreasing function that depends only on the degree of fatigue damage (fraction of the service life), and at the moment of failure it

becomes zero. Domestic and foreign researchers have experimentally proven [12–15] that the rate of decrease in the endurance limit of steel structures depends not only on the magnitude of fatigue damage, but also on the degree of overload.

In addition, the mathematical description of the decrease in the endurance limit must be carried out in accordance with the change in the shape of the entire fatigue curve, in particular the angle of inclination of its left branch. In this regard, to describe the endurance limit, it is necessary to make appropriate adjustments to the relationship (10).

Given the large volume of tests required to obtain the dependence of the change in the endurance limit of full-scale structures by direct experimental methods, indirect methods are promising, in the opinion of the authors.

The physical justification of the shape of the curve of change in the endurance limit is based on the similar nature of the change in the damage energy of fracture. Considering that the level of damage energy in the first approximation is an indirect characteristic of the accumulated fatigue damage and is correlated with the endurance limit, it is possible to assess the nature of the change in the latter at different degrees of overload based on the results of measurements of the power supplied to the sample.

For the mathematical interpretation of the decrease in the endurance limit, taking into account the above assumptions, it is proposed to introduce into formula (9) the kinetic coefficient of fatigue damage accumulation, which depends on the relative durability and the degree of overload C, taking into account the correction, the proposed dependence (φ-criterion) has the form:

$$\sigma_{Ri} = \varphi\sigma_R\left(1-\frac{n_i}{N_i}\right)^{1/m\varphi} \qquad (11)$$

$$\varphi = \exp\left(\frac{n_i}{N_i} lg\frac{\sigma}{\sigma_R}\right) \qquad (12)$$

The proposed dependence satisfies all the requirements formulated at the beginning of the article. In addition, under irregular loading, characterized by large overload degrees and small resource fractions, a slight (from 0.5 to 1.5%) increase in the endurance limit is satisfactorily described by reducing the alternation of stresses in the program block, which can be interpreted as a manifestation of the "training effect" [16]. The reason for such an effect of large stresses is the uneven distribution of plastic deformation in the weld zone and the formation of residual compressive stresses during unloading, which contribute to the increase in the endurance limit. The obtained dependence makes it possible to take into account the change in the endurance limit in calculations of the fatigue life of load-bearing systems.

The calculation of the cyclic durability of the connection node of the pivot beam with the ridge beam of the gondola frame was made for two selected operating modes: movement on a broken dirt road (mode A) and on a dirt road in satisfactory condition (mode B).

When calculating the resource of the supporting system of the gondola frame of the highest accuracy (σ = 7%) achieved using φ-criterion. In this case, the error is not directed to the durability reserve. In general, it should be noted that the use of predictive methods for calculating the durability of load-bearing systems, based on the use of criteria that take into account the regularities of reducing the endurance limit during operation, allows you to achieve high accuracy of the calculation. In our case, the fluctuations of the total relative error of the calculated estimate are within 7% ≤ σ ≤ 27%. This confirms the effectiveness of the methods of the kinetic theory of fatigue, which is based on taking into account real physical processes of fatigue damage accumulation.

The presented computational and experimental data on the assessment of durability allow us to make an attempt to modify the method of calculating durability, based on the application of the linear hypothesis. The essence of the modification is to replace the real fatigue curve, obtained from the results of tests during regular loading, with some fictitious one, which takes into account the change in fatigue resistance factors as the resource is developed (Fig. 3).

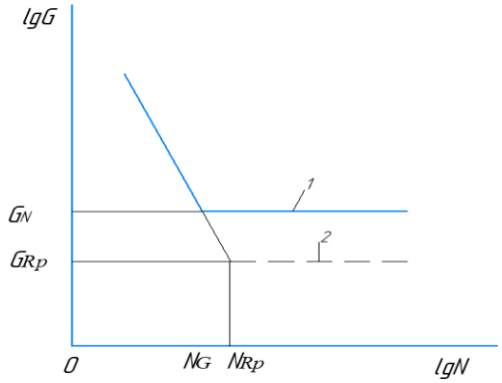


Fig. 3. Before determining the design fatigue curve

Real fatigue curve 1 with parameters σ_R, NG, m is replaced by the calculated curve 2 with parameters σ_R, NGp, m.

Calculating quantities σ_{Rp}, NGp is performed according to the formula:

$$\sigma_{Rp} = K\sigma_R \tag{13}$$

$$lgN_{Gp} = lgN_G + mlg\left(\frac{\sigma_R}{\sigma_{Rp}}\right) \tag{14}$$

For some generalized curve of the decrease in the endurance limit (Fig. 4), the weighted average value of the endurance limit.

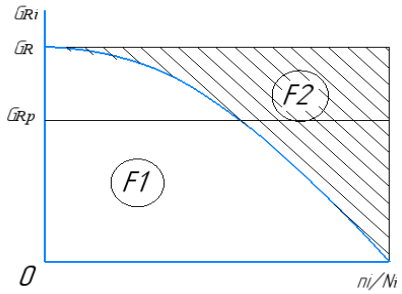


Fig. 4. Before determining the lower limit of damaging stresses

The results obtained show that the lowest accuracy is characteristic of methods based on the formulas:

$$\sigma_{Rp} = \int_0^1 \sigma_{Ri}\left(\frac{n_i}{N_i}\right)d\left(\frac{n_i}{N_i}\right) \tag{15}$$

from where we determine K according to the criteria of D. Henry, CB Sørensen and E.K. Pochtenny, taking into account that the integral in expression (15) is taken in explicit form:

– for D. Henrys criterion

$$K = \frac{\int_0^1 \sigma \frac{\left(1-\frac{n_i}{N_i}\right)}{\left(\frac{\sigma}{\sigma_R}-\frac{n_i}{N_i}\right)}d\left(\frac{n_i}{N_i}\right)}{\sigma_R} = \frac{\sigma\left(\frac{\sigma}{\sigma_R}-1\right)ln\left[\frac{\frac{\sigma}{\sigma_R}}{\frac{\sigma}{\sigma_R}-1}\right]}{\sigma_R} \tag{16}$$

– for the criterion S.V. Serensen

$$K = 1-\int_0^1\left(\frac{n_i}{N_i}\right)\left(\frac{\sigma}{\sigma_R}-1\right)\frac{1}{\frac{\sigma}{\sigma_R}-\left(\frac{n_i}{N_i}\right)^2}d\left(\frac{n_i}{N_i}\right) =$$

$$= 1+\frac{\frac{\sigma}{\sigma_R}-1}{2}ln\left[\frac{1-\frac{\sigma}{\sigma_R}}{\frac{\sigma}{\sigma_R}}\right] \tag{17}$$

– for the criterion E.K. Pochtenny

$$K = \int_0^1\left(1-\frac{n_i}{N_i}\right)^{1/m} * d\left(\frac{n_i}{N_i}\right) = \frac{m}{m+1} \tag{18}$$

Substituting the numerical values of the parameters into the obtained expressions, we obtain the value of K. The calculation results are summarized in Table 1. With sufficient accuracy, we can assume K = 0.6...0.75.

The durability of the nodes has been calculated. of freight car joints using the linear hypothesis and calculated fatigue curves. The calculation results are summarized in Table 2.

Table 1

Results of calculating the coefficient K

Name criterion	Regime	
	A	B
D. Henry	0.598	0.645
CB Serens	0.634	0.691
E.K. Pochtenny	0.757	0.757

Table 2

Results of calculating the durability of welded joints of car frames based on a modification of the linear hypothesis

No.	K	Regime	Ni,cycles	Li, km	L, km*
1	0.6	A	453258	165250	865820
		B	789520	1662510	-37.2
2	0.7	A	568952	212520	1202120
		B	1319805	1952140	-5.65

The denominator contains the calculation error (%) relative to the operational durability, the minus sign indicates that the error is directed towards the durability margin.

Conclusions. The primary analysis of the results given in the tables shows that the durability calculated by the linear hypothesis exceeds that calculated by the adjusted linear hypothesis by more than 5 times and the durability calculated using the criteria for reducing the endurance limit by 1.37 to 1.7 times. The results obtained show that the lowest accuracy is characteristic of methods based on the direct application of the linear damage summation hypothesis and the adjusted linear hypothesis.

Note that for K = 0.7 the calculated endurance limit σ_{Rp} = 48.4 MPa. Calculated abscissas of the fracture points of the fatigue curve $_{NGp}$ = 3.7*106 cycles. Thus, for the considered design variants of the nodes, the application of a modified linear hypothesis using calculated fatigue curves can provide high accuracy of predictive calculation of durability at the value of K = 0.7. In addition, a positive property of the proposed method is the direction of the error in the durability margin.

Bibliography:

1. Zhang L., Qi Y., Research on the wheel wear prediction of heavy-haul freight wagon based on the roller rig test and line test. *Industrial Lubrication and Tribology.* 2024. Vol., ahead-of-print No., ahead-of-print. P. 321–332. https://doi.org/10.1108/ILT-09-2024-0329
2. Vega B. , Pérez J., Comparative analysis of fatigue strength of a freight wagon frame. 2024. *Weld World.* № 68. P. 321–332. https://doi.org/10.1007/s40194-023-01577-5
3. Fomin O., Gerlici J., Lovska, A., Kravchenko, K., Prokopenko, P., Fomina, A., Hauser, V. Durability Determination of the Bearing Structure of an Open Freight Wagon Body Made of Round Pipes during its Transportation on the Railway Ferry. *Communications-Scientific letters of the University of Zilina*. 2019. Vol. 21. № 1. P. 28–34 https://dspace.snu.edu.ua/handle/123456789/636
4. Ližbetin, J., Vejs, P. Dynamic Weighing to Improve Rail Freight Traffic Safety. A Case Study from the Czech Republic. *Transport and Telecommunication Journal. 2022.* Vol. 23 (3). P. 220–226. https://doi.org/10.2478/ttj-2022-0018
5. Fomin O., Lovska A., Kulbovskyi I., Holub H., Kozarchuk I., Kharuta V. Determining the dynamic loading on a semi-wagon when fixing it with a viscous coupling to a ferry deck. *East European Journal of Advanced Technology*. Vol. 2 (7), P. 6–12. https://journals.uran.ua/eejet/article/view/160456
6. Shvets, A., Bolotov O. Influence of loading from the axle of a gondola car on its dynamic indicators and railway track. *Science and Transport Progress,* 2019, Vol. 1 (79). P. 151–166. https://doi.org/10.15802/stp2019/158127
7. UIC (2012) UIC B 12/RP 17 9th Edition. WAGONS. Programme of tests to be carried out on wagons with steel under frame and body structure (suitable for being fitted with the automatic buffing and draw coupler) and on their cast steel frame bogies. UIC (International Union of Railways)
8. Ding, Y, Cheng, L. Dynamic Calibration of a Weigh in Motion System. *Proceedings of the 2018 Joint Rail Conference. 2018 Joint Rail Conference. Pittsburgh, Pennsylvania*, *USA. April 18–20. 2018*. V001T03A007. ASME. https://doi.org/10.1115/JRC2018-6222
9. Фомін О.В., Логвіненко О.А., Бурлуцький О.В., Шелест Д.А., Фоміна А.М., Математичне моделювання процесу термічної правки балки хребтової вантажних вагонів-платформ , Вісник Східноукраїнського національного. 2019. № 3 (251). С. 186–190.
10. Чухліб В. Л., Губський С. О., Окунь А. О., Формалізовані підходи до визначення числа технологічних переходів при виробництві гнутих профілів , Вісник Національного технічного університету «ХПІ».2020. № 2. С. 169–173.
11. Гайдамака А. В., Підшипники кочення. Базові знання та напрямки вдосконалення. навч. посіб., *НТУ «ХПІ» /* за ред. А. В., Гайдамака. 2009. С. 248 https://repository.kpi.kharkov.ua/handle/KhPI-Press/2841
12. Почтений Е.К. Кінетична теорія механічної втоми та її застосування. Наука і техніка. Київ. 1973. 213 с.
13. Henry D. A theory of fatigue damage accumulation in steel. *Transaction ASME.* 1955. Vol. 77. № 6. P. 913–918.
14. Грибков Е. П., Динаміка та міцність металургійних машин / Навчальний посібник з дисципліни для студентів спеціальності *133 «Галузеве машинобудування».* Краматорськ, ДГМА. 2019. С. 68. https://surl.li/uycezt
15. Norman E. Dowling, Mechanical behavior of materials: engineering methods for deformation, fracture, and fatigue. Pearson. 2012, P. 977. https://edisciplinas.usp.br/mod/resource/view.php?id=2271706
16. Мороз В.І., Логвіненко О.А., Іщенко В.І., Фомін О.В., Основи розрахунків на міцність. Опорний конспект лекцій. Харків, 2012. УкрДАЗТ. С. 124.

Фомін О.В., Бурлуцький О.В., Кубрак А.В., Голуб Г.М., Фурсина А.Д. НЕСТАЦІОНАРНІ ФАКТОРИ ОПОРУ ВТОМИ ТА ЇХ ВПЛИВ НА РЕСУРС З'ЄДНАНЬ ВАНТАЖНИХ ВАГОНІВ: МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ТА АНАЛІЗ

Стаття присвячена математичному моделюванню та аналізу запропоновано модифікований варіанту лінійної гіпотези накопичення втомних ушкоджень, ця модифікація класичної лінійної гіпотези (Пальмгрена-Майера) дозволяє точніше прогнозувати залишковий ресурс конструкції. за якого використовується умовне значення межі витривалості, що визначається коефіцієнтом нижньої межі пошкоджуючих напружень.

Лінійна гіпотеза накопичення усталостних пошкоджень, запропонована А. Пальмгреном (1924) та розвинена М. Майером (1939), є основою для розрахунку довговічності матеріалів при циклічному навантаженні. На основі аналізу непрямої характеристики накопиченого втомного пошкодження – енергії руйнування локальних моделей вузлів рам вантажних вагонів визначено залежність зниження

межі витривалості від частки виробленого ресурсу, який враховує динамічне зниження межі витривалості у процесі експлуатації. До найбільш значних недоліків лінійної теорії відноситься те, що вона не описує впливу черговості напружень різних рівнів.

Таким чином, для розглянутих конструктивних варіантів вузлів несучих вузлів вантажних вагонів, застосування модифікованої лінійної гіпотези з використанням розрахункових кривих втоми може забезпечити високу точність прогнозуючого розрахунку довговічності при значенні К = 0,735. Крім того, позитивною властивістю пропонованого методу є спрямованість похибки в запас довговічності.

Наведено розрахункове обґрунтування нижньої межі пошкоджуваної напруги. Запропонована модифікація лінійної гіпотези. Враховує нелінійність через зниження σ_{-1}. та введення σ_{min}. Підтверджена енергетичними та експериментальними даними. Зменшує ризик переоцінки ресурсу на 30–50% порівняно з класичним підходом.

Ключові слова: *залізничний транспорт, рухомий склад, вагон, машинобудівні конструкцій, довговічность, комп'ютерні технології, автоматизація, комп'ютерне моделювання.*

УДК 658.652.6
DOI https://doi.org/10.32782/2663-5941/2025.3.2/04

Чайковський С.Ю.
Український науково-дослідний інститут спеціальної
техніки та судових експертиз Служби безпеки України

ОСОБЛИВОСТІ КОНТРОЛЮ ЯКОСТІ ПРОДУКЦІЇ В УМОВАХ СТРІМКОГО РОЗВИТКУ РИНКОВОЇ ЕКОНОМІКИ

У ринковій економіці проблема якості є найважливішим чинником підвищення рівня життя, економічної, соціальної й екологічної безпеки. Тому контроль якості важливий на виробництві з цілого ряду причин, найбільш очевидними з яких є висока якість продукції і правильна організація виробництва та робочих місць. Він є ключем для досягнення нормативних вимог, збільшення прибутку і оптимізації процесів, іншими словами високоякісне виробництво є вірогідним показником того, що процеси виробника працюють за планом і підприємство отримує якісну продукцію, має прибуток та бездоганну репутацію. Стрімкий розвиток економіки, обумовлений індустріальними (промисловими) революціями XX–XXI століття кардинально змінює механізми діяльності та управління промисловими підприємствами, формує нові інформаційні відносини в суспільстві та сприяє економічному зростанню та соціальному розвитку країни. Наприкінці XX – на початку XXI століття невід'ємною складовою розвитку економіки було дотримання концепції «Індустрії 3.0», яка характеризувалась переходом до застосування інформаційно-комунікаційних технологій (комп'ютер, Інтернет, електроніка). В умовах цієї концепції економічний розвиток підпорядковувався закону пріоритету науки над технологією, а технології – над промисловістю. Сучасний етап – це епоха четвертої промислової революції, яка характеризується стрімким розвитком технологій, що передбачає поширення їх на промислових підприємствах світу. Це сучасний період, який дає змогу контролювати і аналізувати дані швидше та ефективніше, стимулюючи при цьому і виробництво і зменшення витрат, що підвищує якість і продуктивність виробництва.

В статті розглянуто задачі в сфері контролю якості на українських підприємствах в умовах стрімкого розвитку ринкової економіки. Проаналізовано компоненти системи контролю якості (КСЯ) та системи управління якістю (СУЯ), задачі системи менеджменту якості (СМЯ). Визначено об'єкти контролю в СУЯ. Запропоновано організаційні заходи, направлені на рішення задач СМЯ.

Ключові слова: *промислова революція, контроль якості, управління якістю, менеджмент якості, об'єкти контролю.*

Постановка проблеми. Контроль якості продукції на виробництві здається діяльністю звичною і інтуїтивно зрозумілою. Однак все не так просто. Недостатнє розуміння суті змін, які відбулися в сучасному світі як на українських ринках так і у виробництві в 90-ті роки минулого століття, і тих змін, які відбуваються прямо зараз, не дозволяє спеціалістам багатьох українських підприємств ефективно вирішувати питання якості продукції і продуктивності виробничої системи. На початку 90-х років минулого століття у більшості промислово розвинутих країнах завершився перехід від «Індустрії 2.0» (форми організації виробництва, яка народилась в 10-ті – 30-ті роки XX століття в ході 2-ї Промислової революції) до «Індустрії 3.0» (форми організації виробництва, яка утворилась в ході 3-ї Промислової революції) [1]. Після Всесвітнього економічного форуму 2016 року, котрий традиційно відбувається в Давосі, сучасний стан суспільства почали називати «Індустрією 4.0», яка передбачає зростання споживчих очікувань, поліпшення якості товарів, групові інновації і появу нових форм організації. Приклади прояву вже спостерігаються і в Україні – основний драйвер «Індустрії 4.0» – нові фізичні кібер-системи, котрі дають змогу пов'язувати реальні об'єкти з віртуальними і отримувати конкретні економічні ефекти. Але, на жаль, в Україні такі можливості використовуються частково, так як деіндустріалізація, через яку пройшла країна протягом останніх 30 років, та відсутність нової стратегії стримують розвиток нових технологій і залишають Україну не досить привабливою для іноземних капіталів та асоціюються у світі, як економіка з дешевою робочою силою. Україні ще потрібно пройти довгий шлях і витратити велику кількість капіталовкладень, щоб «Індустрія 4.0» стала невід'ємною частиною нашого життя і промисловості [10].

Аналіз останніх досліджень і публікацій. Проблемами якості неухильно та повсякчас займаються зарубіжні та вітчизняні вчені. Е. Демінг описував спосіб забезпечення потреб споживачів і досягнення постійного покращення якості за допомогою: зосередження на процесах; визначенні внутрішніх споживачів; створенні командної організації; зниженні ієрархії і введенні департаменту з управління якістю [5]. А. Фейгенбаум, перший сформулював концепцію комплексного управління якістю [6]. Д. Джуран розробив знамениту «Спіраль якості» (спіраль Джурана) – позачасову просторову модель, яка визначає основні стадії безперервного розгортання робіт з управління якістю і яка стала прообразом багатьох моделей якості, розроблених пізніше [7]. Г. Тагуті розвинув ідею математичної статистики, що належить до статистичних методів. К.Ісікава запропонував варіант комплексного управління якістю, характерними рисами якого є загальна участь працівників в управлінні якістю [8]. В.В. Загоровська, А.А. Корнійчук досліджували фактори, які впливають на якість виготовленої продукції [9].

Але постійні зміни, які відбуваються в сучасному світі, як на ринках так і у виробництві потребують постійного рішення питань якості продукції і удосконалення виробничої системи.

Постановка завдання. В сучасних умовах беззаперечним є той факт, що конкурентоспроможність підприємства будь-якої форми власності залежить від якості продукції, яку воно виготовляє. Як показує світовий досвід провідних компаній світу, удосконалення СМЯ – надійний шлях для забезпечення успіху на міжнародному ринку. Тому треба усвідомити, що якість це не лише ґатунок продукції, а і функціонування підприємства в ринковому середовищі (це означає, що рівень розвитку СУЯ повинен бути спрямований на постійне удосконалення і підвищення ефективності виробництва). Водночас постає питання формування ефективності СУЯ, а саме: забезпечення стійкого розвитку підприємства і розширення експортного потенціалу; покращення екологічних і техніко-економічних показників виробництва, його безпеки; забезпечення взаємозв'язків усіх підрозділів підприємства; пошуків принципово нових підходів до управління соціально-економічними системами.

Метою статті є: дослідження якості технічної продукції як фактора забезпечення її конкурентної спроможності як на внутрішньому, так і на зовнішньому ринках; визначення проблеми підвищення якості та шляхи їх вирішення; всебічне вивчення проблем управління якістю в сучасних умовах і аналіз впровадження СУЯ на підприємствах України в рамках переходу від «Індустрії 2.0» до «Індустрії 3.0–4.0».

Виклад основного матеріалу. Перехід від «Індустрії 2.0» до «Індустрії 3.0» був пов'язаний з глобалізацією міжнародних ринків [1]. Логістичні ланцюжки розширились, подовжились і охопили весь світ. «Індустрія 3.0» була націлена на все людство і тому повинна була не просто виробляти продукцію в колосальних об'ємах, але і чітко враховувати різні вимоги споживачів в конкретному сегменті, в регіоні, і, що важливо, в конкретно визначений період. Скорочення часу на виготовлення замовлення і прискорення обороту капіталу на виробництві – ключові драйвери більшості змін у порівнянні з «Індустрією 2.0». Іншими словами девіз «Індустрії 2.0» – «більше, але дешевше», а девіз «Індустрії 3.0» – «чіткіше та швидше». Зміна ринкових відносин потребувала масового виробництва вже не «на склад», а «під замовлення». При такому виробництві закупка компонентів відбувалася безпосередньо для виробничої партії (відповідно об'єму замовлення) – звідси і виникла концепція постачання «точно в термін». У порівнянні з «Індустрією 2.0» все прийшлось робити по іншому: інакше планувати виробничі процеси (концепція MRP [2])[1], організовувати виробництво за видами продукції (а не за технологічним переділом[2], як в «Індустрії 2.0»), використовувати інше обладнання (яке швидше переналагоджується під інше замовлення), тощо. В Україні перехід від «Індустрії 2.0» до «Індустрії 3.0» був розпочатий в середині 80-х років минулого століття, а ближче до кінця 80-х в силу об'єктивних причин різко змінився елементарною боротьбою організацій за виживання, і тому не був усвідомлений.

Але саме цей період зачепив безліч сторін життя підприємств і організацій, як в управлінні ланцюжками постачання, так і в організаціях виробничої системи підприємств – постачальників, а саме:

1. На глобальному ринку під час цього переходу головним стало питання ефективного управ-

[1] MRP – це процес планування матеріальних потреб підприємства (material requirements planning). Він вимагає беззастережної точності, високого контролю та постійних оновлень, а також обліку величезної кількості даних. Будь-яка помилка в плануванні може призвести до збитків: простою потужностей або транспорту, перевантаження складу, втрати замовлення або контракту з клієнтом тощо.

[2] Технологічний переділ – це закінчена частина технологічного процесу або сукупність виробничих операцій, у результаті яких із сировини та матеріалів виходить проміжний продукт (напівфабрикат) або з напівфабрикату готовий виріб.

ління «довгими» (що охоплюють підприємства не одної, а багатьох країн) і «широкими» (тобто кількість постачальників при виготовленні складних видів продукції вимірюється тисячами) ланцюжками постачання. Для ефективного управління такими ланцюжками постачання необхідно забезпечити взаємозамінність організацій-постачальників в цьому ланцюжку, а не просто консолідувати взаємозамінність компонентів продукції. Для забезпечення цього був створений комплекс міжнародних управлінських стандартів серії ISO 9000 (серія стандартів Міжнародної організації зі стандартизації (ISO), які застосовуються при створенні та удосконаленні СМЯ організацій), ISO 14000 (низка міжнародних стандартів, пов'язаних з управлінням навколишнім середовищем, яке існує, щоб допомогти організаціям):, ISO 18000 (міжнародний стандарт, який описує серію різноманітних RFID-технологій, кожна з яких використовує унікальний діапазон частот) тощо, і запущений механізм підтвердження системи управління підприємствами-постачальниками вимогам цих стандартів [3, 4]

2. Змінились умови гарантій, які підприємство-постачальник зобов'язане надати замовнику на етапі контрактації. Створений і запущений механізм стандартизації таких гарантій і механізм підтвердження відповідності цих гарантій вимогам замовника [3]. В сучасних умовах постачальник зобов'язаний гарантувати замовнику наступне:

а) свою здатність забезпечити відповідність товару, що постачається, умовам замовника;

б) створити зручні для замовника процеси супутнього сервісу (процеси контрактації, оплати товару, постачання і обслуговування);

в) свою надійність як постачальника (гарантію того, що постачальник буде відповідати вимогам замовника щодо якості продукції, строкам постачання, ціні);

г) стійкість свого бізнесу як постачальника (гарантію того, що на протязі всього строку контрактних обов'язків, включаючи обслуговування після продажу, постачальник буде виконувати всі гарантійні умови із урахуванням зовнішніх для нього ризиків, включаючи повінь, снігопади, пандемії, тощо).

Така складна схема з багато-чисельними гарантіями (створення яких потребує немалих затрат) знадобилась для того, щоб замовник міг економити на сукупній вартості володіння товаром. Ще в 1951 році Е. Демінг у післявоєнній Японії пояснив японським менеджерам: «*Покінчіть з практикою оцінки і вибору постачальників лише на підставі контрактної ціни на їх продукцію. Замість цього наряду з ціною вимагайте серйозних підтверджень якості продукції...*» [5] Можна сказати, що з моменту, коли більшість японських менеджерів усвідомили цей підхід (а також інші 13 пунктів програми Е.Демінга), почалось «японське економічне диво»[3].

Сукупні затрати по Е. Демінгу, крім контрактної ціни, включають: витрати на вхідний контроль; витрати на усунення дефектів; витрати пов'язані з виявленими ризиками застосування продукції у замовника. Якщо до сукупних витрат додати затрати на планові і непланові ремонти і затрати на утилізацію виробів отримаємо сукупну вартість володіння продукцією [21, 22]. Досвідчений замовник, у повній відповідності з ідеями Е. Демінга вибере постачальника продукції не за критерієм контрактної ціни, а за критерієм сукупних витрат (для комплектувальних) або за критерієм сукупності вартості володіння виробом (для фінальних виробів) або навіть за вартістю життєвого циклу виробу (як правило, для складних технічних систем).

Але повернемось до питання гарантій. Під гарантією треба розуміти сукупність:

– документів, які підтверджують виконання робіт по зниженню відповідних ризиків до прийнятного рівня, включаючи документи оцінки таких ризиків (прикладами таких документів є протоколи технічного контролю, протоколи випробувань продукції, які виконані лабораторією, що користується довірою споживача, протоколи відділу технічного контролю (ВТК), протоколи оцінки ризиків, тощо);

– зобов'язань щодо поділу ризиків споживача з постачальником (або третьою стороною). Прикладом таких зобов'язань є гарантійні обов'язки по ремонту обладнання протягом обумовленого терміну експлуатації або зобов'язань щодо безкоштовної заміни виробу у випадку відмови споживача на протязі зазначеного терміну ДСТУ ISO 9001:2015 [14].

На сучасному виробництві, організованому у відповідності з принципами «Індустрії 3.0» СКЯ підприємства повинна формувати вказані вище документи, які підтверджують виконання робіт щодо зниження відповідних ризиків до прийнятного рівня. Концепція цієї системи була розроблена ще в 70-ті роки минулого століття і відносно успішно застосовувалась підприєм-

[3] Япо́нське економі́чне ди́во – явище стрімкого росту японської економіки з середини 1950-х років до нафтової кризи 1973 року.

ствами високотехнологічної сфери. В 90-ті роки багато підприємств вирішили, що концепція СМЯ не просто замінює, а й відмінює концепцію СКЯ. Частково виною тому став і не дуже вдалий переклад терміну «Qualiti Assurance» в стандартах серії ДСТУ ISO 9000:2015 [13] У нас його переклали як «забезпечення якості», хоча більш точний переклад слова «Assurance» – гарантія. Тобто «Qualiti Assurance» – це гарантування якості. Тоді стає зрозумілим, що якість забезпечує вся виробнича система в цілому і СМЯ як ключова частина управління виробничої системи, а система контролю якості – гарантує якість як всередині підприємства, так і для замовника.

Поява ДСТУ ISO 18091:2020 [15] фактично надає нове життя цій концепції і застосуванню підприємствами серії стандартів ISO 9000, які визначають розроблення, впровадження та функціонування систем якості. Вони не стосуються конкретного сектора промисловості чи економіки і являють собою настанови з управління якістю та загальні вимоги щодо забезпечення якості, вибору і побудови елементів систем якості. Сучасний підхід говорить, СКЯ є частиною СМЯ. Якщо головна задача СМЯ – забезпечення задоволеності споживачів і надійність постачання шляхом забезпечення прийнятного рівня операційних ризиків, то головна задача СКЯ – створення і підтвердження перерахованих вище гарантій якості шляхом постійного контролю за якістю («забезпечення якістю»). Цей термін підтверджується вимогами ДСТУ ISO 9001:2015 (розділ 9.1 «Моніторинг, вимірювання, аналіз і оцінка»).

До числа основних об’єктів контролю якості входять[11]*:*

– методи розробки; технічні умови (ТУ), конструкторська, технологічна текстова та ін. нормативно-технічна документація;

– якість сировини, матеріалів, напівфабрикатів та комплектувальних виробів як покупних так і власного виробництва [16, 17];

– технічний рівень та стан обладнання, технологічного оснащення та інструменту, прогресивність технології [12];

– кваліфікаційний рівень виконавців технологічних операцій та управлінського апарату;

– технологічна дисципліна на виробництві і якість роботи працівників;

– методи технічного контролю і випробувань продукції, наявність і технічні можливості та стан контрольно-вимірювальних приладів, пристосувань та інструменту;

– якість деталей, вузлів, складальних одиниць та готової продукції;

– якість паковання[4] і тари, засоби і правила складування, збереження і транспортування;

– правила експлуатації, технічного обслуговування і діагностики виробів споживачами, їх дотримання;

– якість ремонту і відновлення зношених деталей, вузлів і виробів в цілому, якість запасних частин;

– діяльність органів управління різних рівнів та ланок з реалізації наданих їм контрольних повноважень, процес розвитку і удосконалення СУЯ і технічного контролю на виробництвах, у галузях, тощо.

Таким чином СКЯ розповсюджується на більшість процесів виробничої системи і на, практично, усі стадії життєвого циклу продукції (рисунок 1) та включає в себе наступні елементи: об’єкти контролю якості; контрольні операції і їх послідовність; правила, методи і засоби контролю; засоби механізації і автоматизації контролю; документацію; виконавців відповідної кваліфікації.

ВТК – частина системи контролю якості підприємства, але не єдина. В цій системі повинні приймати участь і технологи і начальники виробництва і виробничі працівники.

Тому що тільки завдяки спільним зусиллям можна вирішити перераховані вище задачі СКЯ, які полягають в наступному:

1. Недопустимість потрапляння неякісної продукції (браку) замовнику (зовнішньому та внутрішньому).
2. Запобігання випуску неякісної продукції.
3. Оформлення гарантій прийнятності ризиків: застосування сировини та комплектувальних виробів при вхідному контролі; у виробничому процесі (шляхом операційного контролю); при контролі і випробуваннях готової продукції.

Зокрема при *контролі якості продукції* ВТК підтверджує:

– відповідність продукції вимогам: замовника; стандартів, обумовлених контрактом; технічних регламентів на підставі операцій контролю (випробувань);

– оформлення документів, що підтверджують належну якість (протоколи контролю та випробувань, інші підтверджувальні документи).

[4] Пако́вання (хибна калька з російської – «упаковка») – промисловий виріб, що призначається для пакування, зберігання, переміщення і реалізації товарів у сфері обігу. Паковання зберігає товар від пошкоджень, сприяє його безпечному транспортуванню, збереженню та продажу.

Рис. 1

Відносно процесів виробництва ВТК спільно з технологами підтверджує:

– що процес проходив в керованих умовах (параметри обладнання, пристосування і т.д. – в нормі; дотримувалась технологічна дисципліна);

– що процес був стабільний (тобто не було збоїв налаштувань; неконтрольованих змін технологічних режимів і т.д.).

На сучасному етапі СКЯ базується на таких принципах:

1. *При забезпечення якості закупок на:* кваліфікацію постачальників по рівню надійності (рівню ризиків постачання); зниженні поточних ризиків постачальника; перенесенні вхідного контролю на постачальника [18, 19, 20].

2. *При забезпечення якості в процесі виробництва на*: підвищенні вимог до якості продукції і стабільності процесів; зниженні ризиків процесів [18, 23].

3. *При забезпеченні якості готової продукції на:* підтвердженні стабільності продукції [23].

В реальності, функції СКЯ на підприємствах визначаються якістю СУЯ.

Часто в організаціях проводиться, в основному розбраковування продукції, а ключові причини дефектів або невідповідностей не визначаються, або визначаються примітивно (типу «брак постачальника», «брак виконавця» тощо). Приділяється недостатня увага запобіганню повторення дефектів по схожим причинам, класифікатор дефектів і типових причин відсутній. Ризики процесів, критичні технологічні елементи (КТЕ), ключові характеристики продукції і процесів не визначаються. Не проводиться статистичне управління процесами, відсутні об'єктивні свідоцтва того, що процеси стабільні і відбуваються в керованих умовах. Всі ці фактори негативно впливають на результативність контролю у виробничому процесі, знижують функціональні якості виробів і впливають на їх експлуатацію.

Висновки. Таким чином, в умовах стрімко зростаючої економіки для забезпечення контролю якості продукції підприємствам будь-якої форми власності на всіх етапах виробництва необхідно створити досконалу СУЯ, яка б відповідала міжнародним стандартам і забезпечувала гнучке управління підприємством [8], а саме:

– усі постачальники повинні мати тісні взаємовідносини із замовниками (обмін інформацією на підставі контрактів);

– технологи, начальники цехів, дільниць та ВТК повинні бути в курсі КТЕ постачальників, отримувати інформацію щодо ризиків функціонування КТЕ, і дані від постачальників щодо стабільності технологічних процесів (контрольні карти);

– ВТК спільно з конструкторами і технологами має постійно приймати участь у вихідному контролі у постачальників, знати ключові характеристики продукції і процесів встановлених конструкторами і технологами в документації; створювати класифікатори дефектів, типових їх причин;

– проводити постійний моніторинг стабільності процесів по ключовим характеристикам продукції (контрольні листки, контрольні карти);

– виробничі процеси мають бути стабільними і відбуватися в керованих умовах, що дозволяє перейти в операційному контролі на самоконтроль виконавців обмежившись летючим контролем зі сторони ВТК;

– необхідно забезпечення відсутності повторних дефектів з вихідних причин (регулярне проведення превентивних заходів);

– ВТК регулярно має співпрацювати із замовником – надавати дані моніторингу процесів і продукції, які підтверджують прийнятність ризиків замовника.

Отже, тільки подібний комплексний підхід, ефективна взаємодія підрозділів і спеціалістів (конструкторів, технологів, ВТК, виробничих працівників і менеджерів) здатні реально забезпечити якість в сучасних високотехнологічних виробництвах і допомогти українським підприємствам максимально повно використовувати можливості, які закладені у вимогах міжнародного стандарту ISO 9000.

Список літератури:

1. ІНДУСТРІЯ 1-2-3-4-5.0 – ТОВ «Укрфаворіт» URL: https://ukrfavorit.com.ua ’ industriya-1-2-3-4-5.0 (дата звернення 15.03.2025)
2. Що таке MRP система: ключові функції та переваги – Wezom. URL: https://wezom.com.ua (дата звернення 15.03.2025)
3. Етапи розвитку та міжнародний досвід управління. Луцький національний технічний університет. URL: https://surli.cc/jwbgjn (дата звернення 20.03.2025).
4. Міжнародні стандарти та принципи діяльності ISO isocert.org.ua URL: https://surl.li/roarxd (дата звернення 20.03.2025)
5. Е.Демінг Вихід із кризи. Нова парадигма управління людьми, системами і процесами bookmood. com.ua URL: https://bookmood.com.ua (дата звернення 20.03.2025)
6. Роботи А. Фейгенбаума з менеджменту якості StudFiles URL: https://surl.lu/kcvftz (дата звернення 20.03.2025)
7. Роботи Дж. Джурана з менеджменту якості info-library.com.ua. URL: https://surl.lu/wavzzb (дата звернення 20.03.2025)
8. Концепції якості. Погляди основоположників. Електронний репозитарій ДВНЗ «УжНУ» URL: https://surli.cc/wezkqy (дата звернення 20.03.2025)
9. Фактори, що впливають на якість продукції. Державний університет «Житомирська політехніка» URL: https://surl.li/nbtabn (дата звернення 20.03.2025)
10. Четверта промислова революція... Харківський національний економічний університет імені Семена Кузнеця URL: https://surl.li/ehmcej (дата звернення 20.03.2025).
11. Види і методи технічного контролю якості продукції. Луцький національний технічний університет URL: https://surl.lt/cjlgsv (дата звернення 25.03.2025)
12. Дзідока – свіжий погляд на якість.URL:. https://surl.li/yidgvy (дата звернення 30.03.2025)
13. ДСТУ ISO 9000:2015 Системи управління якістю. Основні положення та словник термінів (ISO 9000:2015, IDT).
14. ДСТУ ISO 9001:2015 Системи управління якістю. Вимоги (ISO 9001:2015, IDT).
15. ДСТУ ISO 18091:2020 Системи управління якістю. Настанови щодо застосування ISO 9001:2000 у суб’єктах місцевого самоврядування (ISO 18091:2019, IDT).
16. СОУ ДКА 0064:2011. Вхідний контроль продукції. Правила проведення.
17. ДСТУ 9027:2020 Системи управління якістю Настанови щодо вхідного контролю продукції.
18. Управління ризиками відповідно до стандарту ISO 31000-2018. URL: https://surli.cc/wmlwat (дата звернення 30.03.2025).
19. ДСТУ ISO 28000:2008 Системи управління безпекою ланцюга постачання. Вимоги (ISO 28000:2007, IDT).
20. ДСТУ ISO 21247:2007 Статистичний контроль. Комбіновані системи відбирання зразків з повною відповідністю та процедури контролювання процесу під час приймання продукції (ISO 21247:2005, IDT).
21. ДСТУ 3021-95 Випробування і контроль якості продукції. Терміни та визначення.

22. ДСТУ ISO 14040:2013 Екологічне управління. Оцінювання життєвого циклу. Принципи та структура (ISO 14040:2006, IDT).

23. Орлов П.А. Впровадження систем управління якістю: стан, проблеми, перспективи / П.А.Орлов / Стандартизація, сертифікація, якість. 2013. № 6. С. 59–63.

Chaikovskyi S.Yu. FEATURES OF PRODUCT QUALITY CONTROL IN THE CONDITIONS OF RAPID DEVELOPMENT OF MARKET ECONOMY

In a market economy, the problem of quality is the most important factor in improving the standard of living, economic, social and environmental safety. Therefore, quality control is important in production for a number of reasons, the most obvious of which are high product quality and proper organization of production and workplaces. It is the key to achieving regulatory requirements, increasing profits and optimizing processes, in other words, high-quality production is a reliable indicator that the manufacturer's processes are working according to plan and the company is receiving quality products, has a profit and an impeccable reputation. The rapid development of the economy, caused by the industrial revolutions of the 20th and 21st centuries, radically changes the mechanisms of activity and management of industrial enterprises, forms new information relations in society, and contributes to the economic growth and social development of the country. The rapid development of the economy, caused by the industrial revolutions of the 20th and 21st centuries, radically changes the mechanisms of activity and management of industrial enterprises, forms new information relations in society, and contributes to the economic growth and social development of the country. Under this concept, economic development was subject to the law of priority of science over technology, and technology over industry. The current stage is the era of the fourth industrial revolution, which is characterized by the rapid development of technologies, which involves their distribution in industrial enterprises of the world. This is a modern era that allows you to monitor and analyze data faster and more efficiently, while stimulating both production and reducing costs, which increases the quality and productivity of production.

The article considers the tasks in the field of quality control at Ukrainian enterprises in the conditions of rapid development of the market economy. The components of the quality control system (QCS) and the quality management system (QMS), the tasks of the quality management system (QMS) are analyzed. The objects of control in the QMS are determined. Organizational measures aimed at solving the tasks of the QMS are proposed.

Key words: *industrial revolution, quality control, quality management, quality management, control objects.*

УДК 658.562
DOI https://doi.org/10.32782/2663-5941/2025.3.2/05

Шорнікова С.В.
Український науково-дослідний інститут спеціальної
техніки та судових експертиз Служби безпеки України

ТЕХНІЧНЕ ЗАВДАННЯ – ПЕРШИЙ КРОК ДЛЯ ОТРИМАННЯ ЯКІСНОГО ЕЛЕКТРОННОГО ВИРОБУ

Якісна електронна продукція – запорука успіху будь-якого виробника. Виробник забезпечує випуск якісної документації та електронного виробу при умові досконалого виконання умов замовника. Тісний взаємозв'язок замовника та виробника для отримання якісного та надійного виробу забезпечується технічним завданням.

Технічне завдання є першим документом після підписання договору на розробку документації та виробництво виробу. Всі необхідні параметричні дані, які повинен виконати виробник, повинні бути викладені в технічному завданні.

В статті розглянуто етапність розробки електронного виробу: виготовлення моделі, проведення науково-дослідних робіт, етапи дослідно-конструкторських робіт, коригування чи редизайн – супроводжується вимогами вказаними в технічному завданні, відповідно до етапу проектування як всього виробу так і кожного модуля та його складових частин.

Технічне завдання етапу розробки нової моделі електронного виробу складається з вимог до загального вигляду виробу, до вимог кожного складового модуля, параметричних та електротехнічних даних, електронних вимог тощо.

Проаналізовані етапи виготовлення документації для робочого проекту, які завершуються серійним виробництвом електронного виробу, базуються на технічному завданні з урахуванням недоліків, доопрацьованих параметрів та удосконалених технологій ескізного проекту.

В статті розглянуті практичні рекомендації щодо вимог редизайну технічного завдання та необхідності внесення змін, пов'язаних з розвитком та застосуванням новітніх матеріалів, удосконаленням виробу сучасними електронними компонентами, розвитком нових технологічних процесів.

Технічне завдання оформлюється відповідно державних стандартів та нормативних документів, галузевими стандартами, стандартами підприємств та регламентуючими документами підприємства.

Ключові слова: *технічне завдання, науково-дослідна робота, дослідно-конструкторська робота, редизайн виробу, модель виробу, модуль виробу, стандарти.*

Постановка проблеми. Відповідальність за якісно виконаний виріб несуть як замовник так і виробник. Вибір відповідального виробника є запорукою отримання якісно виготовленого виробу. Однак замовник відповідає за досконало поставлену задачу, у вигляді технічного завдання, перед виробником.

Недосконаліть даних технічного завдання збільшує час на виробництво, збільшує ймовірність помилки, отримання недосконалого виробу, збільшення фінансових витрат.

Аналіз останніх досліджень і публікацій. Важливість наявності технічного завдання при проектуванні та виробництві електронного виробу відзначали у своїх роботах вітчизняні науковці: Євсіна Н.О., Дудник О.В., Вунтесмері Ю.В. тощо.

Фірми виробники, що акцентують увагу замовників на важливості технічного завдання на всіх стадіях проектування та виготовлення:

– вітчизняні – компанія Craft Electronics, «DIAS Electronics», фірма SEA, TISON тощо;

– закордонні – Alnicko Lab (Естонія), FANUC (Японія), Festo (Німеччина) тощо.

Постановка завдання. Важливість наявності технічного завдання як на стадії розробки документації на виріб, на етапі виготовлення виробу, так і на етапах удосконалення та корегування. Необхідність технічного завдання, як документа, що деталізує опис вимог до параметричних даних пристрою, особливо до модулів з електричними параметрами, кожного етапу розробки приладу.

Метою статті є визначення важливості технічного завдання для виробництва якісної електронної техніки та її складових частин

Виклад основного матеріалу. Технічне завдання – виважений документ співпраці замовника та виробника для отримання високоякісного електронного виробу з дотриманням сучасних вимог та технологій.

Технічне завдання встановлює основні вимоги до робіт, обов'язки замовника та виробника, показники якості, терміни проведення робіт, економічні показники тощо.

Наявність технічного завдання зумовлена розподілом праці на етапах життєвого циклу виробу.

При розробці технічного завдання необхідно врахувати такі інформаційні матеріали:

– науково-технічна інформація;
– патентна інформація;
– характеристика ринку збуту;
– характеристика виробництва виготовлення виробу (кваліфікація кадрів, технологічна оснащеність, організація праці тощо).

Відповідно до [5] технічне завдання – частина технічної документації – сукупність документів, необхідних і достатніх для безпосереднього використання на усіх стадіях життєвого циклу продукції.

Технічне завдання є вихідним документом, що встановлює основне призначення, тактико-технічні і техніко-економічні характеристики пред'явлені до розроблюваного виробу [1].

Відповідно до [8] технічне завдання включає наступні розділи:

1. Найменування та область застосування.
2. Підстава для розробки і її джерела.
3. Мета і призначення розробки.
4. Технічні вимоги.
5. Економічні показники.
6. Етапи розробки.
7. Порядок контролю і приймання.

У розділі «Найменування та область застосування» вказують найменування виробу і приводять коротку характеристику області його використання.

У розділі «Підстава для розробки і її джерела» указують найменування документа, на підставі якого ведеться розробка, і організації, яка затвердила цей документ.

Технічне завдання містить вимоги основних показників:

Склад та вимоги до конструкції:

– найменування, кількість складових частин виробу та їх призначення;
– вагові та габаритні характеристики;
– спосіб установки і закріплення на об'єкті;
– вимоги за рівнем мініатюризації.

Вимоги до показників радіоелектронної апаратури – встановлює основні технічні характеристики виробу, які визначають цільову спрямованість:

– радіотехнічні характеристики (режим роботи, параметри сигналів, вихідні параметри, індикація тощо);

– експлуатаційні показники технічної ефективності, що визначають основні конструкторські рішення (потужність випромінювання, споживча потужність, діапазон прийнятих частот, швидкодія тощо);

– класифікаційна характеристика об'єкта установки, що виробляється по класах, групам і підгрупам використання.

3. Вимоги по стійкості до механічних і кліматичних впливів.
4. Вимоги до надійності.
5. Вимоги ергономіки та естетики.
6. Вимоги до технологічності та уніфікації.
7. Патентно-правові вимоги.
8. Вимоги безпеки.

При розробці технічного завдання на вироби підвищеної складності попередньо пророблюється аванпроект [5], який дозволяє попередньо поглиблено пропрацювати питання, які визначать доцільність та необхідність створення даного виробу. Цей етап повинен гарантувати створення виробу, що відповідає техніко-економічним показника світового рівня.

Технічне завдання визначає стадії розробки дослідно-конструкторських робіт. Відповідно до вимог технічного завдання дослідно-конструкторські роботи виконують із урахуванням галузевих стандартів і нормативних документів відповідно певного виду продукції [3].

Технічне завдання визначає необхідність розроблення, випробування, перелік та склад документації яку необхідно розробити з метою підтвердження розрахунків та нових технічних рішень на створення моделі або експериментального зразка майбутнього виробу у процесі виконання дослідно-конструкторських робіт.

Розробку та затвердження змісту технічного завдання визначають замовник та виконавець, а під час ініціативного розроблення – виконавець.

До розробки технічного завдання можуть залучатися інші зацікавлені організації (підприємства): виробник, головна організація з виду продукції, організація-проектувальник, монтажна організація тощо.

Доручення замовника на розробку технічного завдання визначає вихідні вимоги у формі головних завдань, показників та характеристик, яким повинні відповідати результати дослідно-конструкторських робіт. Доручення на розробку технічного завдання замовник може дати декільком потенційним виконавцям для подальшого розглядання на конкурсній основі.

Забезпечення встановлених у технічному завданні показників надійності і якості продукції виконують урахуванням показників комплектувальних виробів, які до неї входять.

Розподіл функцій виконання дослідно-конструкторських робіт визначають у технічному завданні. Обсяг, складність та характер завдань при виконанні дослідно-конструкторських робіт визначають одного або декількох організацій виконавців. Якщо роботи виконує декілька організацій-виконавців, серед них визначають головного виконавця. Інші організації є співвиконавцями та виконують складові частини дослідно-конструкторських робіт.

Відповідно до [3] визначаються функції замовника та виконавця на виконання дослідно-конструкторських робіт.

Функції замовника дослідно-конструкторських робіт:

– ініціює виконання дослідно-конструкторських робіт з метою розробки чи модернізації певної продукції та видає заявку на виконання робіт;

– видає вимоги на розробку документації та виробу;

– організовує експертизу заявок на виконання дослідно-конструкторських робіт та приймає рішення на укладання договору з виконавцем робіт;

– узгоджує та затверджує технічне завдання.

Функції головного виконавця дослідно-конструкторських робіт:

– розробляє технічне завдання на проведення робіт на підставі вихідних даних замовника;

– погоджує технічне завдання з замовником;

– визначає зміст технічного завдання на складові частини у співпраці з виконавцями складових частин дослідно-конструкторських робіт та виконує стосовно них функції замовника;

– згідно технічного завдання виконує дослідно-конструкторську роботу і координує роботу виконавців складових частин;

– забезпечує виконання вимог технічного завдання завдяки розробці та реалізації заходів, що покращують технічні показники виробу;

– відповідає за дотримання виконання визначених у технічному завданні показників виробу.

Також на основі затвердженого технічного завдання, головний виконавець готує матеріали для укладання договору на виконання основної та складових частин виробу та подає замовнику.

Виконавець складової частини дослідно-конструкторських робіт виконує відповідні роботи:

– розробляє технічне завдання на складову частину робіт і подає на затвердження головному виконавцю;

– виконує об’єм робіт відповідно до затвердженого головним виконавцем технічного завдання;

– звітує про виконання робіт перед головним виконавцем, згідно умов договору.

Головний виконавець визначає необхідність виконання дослідного зразка та його виконавця, у разі коли сам не є виконавцем.

Згідно з [4] технічну документацію на дослідний зразок розробляють на підставі технічного завдання на виконання дослідно-конструкторських робіт або документа, який його замінює – контракту, договору, протоколу, ескізу тощо, визнаного замовником і розробником, та який містить необхідні і достатні вимоги щодо створення цього зразка [7], рекомендації стосовно розроблення технічного завдання на дослідно-конструкторські роботи викладені в [3], [4] та [7].

Затверджене технічне завдання на дослідно-конструкторські роботи визначає вирішення таких питань:

– конструкторську документацію – графічну, текстову чи аудіовізуальну;

– конструкторську документацію розробляють і оформляють відповідно до національних стандартів, інших нормативних документів за темою розробки, зазначених у технічному завданні;

– формалізацію виробничих стосунків між замовником, розробником, співвиконавцем, виробником, постачальником;

– у номенклатуру конструкторських документів включають стадії та терміни виконання робіт, відомості про учасників розробки тощо;

– встановлено номенклатуру підписів на конструкторську документацію;

– не розробляються технічні умови на виріб одиночного виробництва;

– розробка, виготовлення, приймання та постачання виробу здійснюються відповідно технічного завдання;

– рівень стандартизації та уніфікації виробу визначають за допомогою кількісних показників з урахуванням виконання якісних вимог зі стандартизації та уніфікації;

– оцінювання рівня стандартизації та уніфікації виробу здійснюється порівнянням одержаних кількісних показників із заданими та дотримання якісних вимог до виробу відповідно запланованих;

– приймальні випробування дослідного зразка відповідно до затвердженого документа.

Технічне завдання кожного етапу дослідно-конструкторських робіт закривається після отримання якісного дослідницького зразка чи готового виробу та затверджується визначеними відповідними підписами.

Для електронного виробу одним з показників якісного виготовлення є відповідність електричним параметрам заявлених замовником, досконало прописаним в технічному завданні та виконаних виконавцем робіт. Ці вимоги виконуються на етапі схемотехнічного та конструкторського проектування.

Первинне завдання до проектування виробу подається у вигляді сформованих вимог замовника та мають бути виконані за допомогою розробленого пристрою. Вимоги замовника повинні визначити користувацькі вимоги: функціональні, нефункціональні, критичні тощо.

Необхідно сформувати технічне завдання де визначені основні вимоги та параметри пристрою, такі як вид сигналів, які використовуються для передачі та перетворення інформації та можливість поєднання їх з ергономічним зовнішнім виглядом. Ці питання вирішує етап схемотехнічного проектування.

Схемотехнічний етап проектування, або проектування на рівні принципових схем, починається з аналізу технічного завдання, розбиття на окремі вузли розроблювального виробу й завершується розробкою технічного завдання на створення конструкторської та технологічної документації [2].

Основне завдання етапу – розробка принципових схем вузлів виробу, які забезпечують задані вимоги до їхніх електричних характеристик прописаних у технічному завданні.

Проектування принципових схем виконується кваліфікованим інженером проектувальником з використанням розрахунків за допомогою сучасних систем автоматизованого проектування відповідно до сучасної бази електронних компонентів. Розрахунки проводять використовуючи бази даних, системи експертних оцінок, оптимізаційні процедури та визначають найкращий з представлених варіантів.

Виконання вимог технічного завдання, пропонованих до окремих каскадів і до проектного пристрою в цілому, дозволяє використання вузлів, побудованих по різних принципових схемах. Враховуються електронно-технічна елементна база та технічні можливості виробника.

Конструкторський етап проектування, відповідно до технічного завдання, включає:

– розробка друкованих плат;

– міжблочні та міжвузлові з’єднання;

– конструкції блоків та виробу в цілому;

– передньої панелі та корпусу в цілому тощо.

Для отримання необхідних електричних показників, визначених технічним завданням, розробка друкованих плат виконується з використанням сучасних систем автоматизованого проектування. При конструюванні враховуються обмеження, які виникають з тепловими режимами, паразитними та розподіленими параметрами конструктивних елементів схем тощо. Обов’язкове використання новітніх електронних елементів, основ для друкованих плат та конструкторсько-технологічних вимог, що забезпечать виконання параметричних вимог технічного завдання.

Детальне технічне завдання зменшує ймовірність помилки в процесі проектування. Тому важливо вказати:

– розміри друкованих плат, їх конфігурація, наявність монтажних отворів, вимоги до зон трасування провідників;

– елементну базу у вигляді переліку елементів з посиланням на типорозмірів корпусів;

– установка нестандартних елементів, вказати висотні обмеження, місця обмеження трасування;

– тип друкованої плати: – одно-, дво-, або багатошарова;

– вид друкованої плати – на металевій основі, жорстка, гнучка, комбінована тощо;

– наявність ділянок екранізації, вимоги до трас живлення;

– технологічні особливості виготовлення;

– будь-яку необхідну важливу інформацію про проектовану друковану плату.

Наступним важливою етапом в розробці електронних виробів є виготовлення друкованих плат. Виробники друкованих плат зацікавлені отримати детальну вхідну інформацію – технічне завдання з описом вимог замовника та пакет розробленої конструкторської документації. Доку-

ментація визначає матеріали деталей, основ плат, пайки, покриття та використання технологічних процесів, що забезпечать вимоги замовника.

Тісна взаємодія замовника та виробника впродовж процесу проектування гарантує якість друкованої плати. Однак випробування та експлуатація виробу виявляють недоліки та помилки, які потребують коригування або редизайну.

Додаткове технічне завдання дозволить провести коригування або редизайн в таких випадках:

– заміна застарілих, знятих з виробництва елементів плати;

– оптимізація плати для серійного виробництва;

– адаптація плати до сучасних систем проектування;

– відновлення втраченої та доопрацювання наявної документації.

Усунення експлуатаційних недоліків всіх етапів проведення проектування та розробки виробу, відповідно визначених у технічному завданні, гарантує серійне виробництво та високих показників реалізації.

Висновки. Технічне завдання – результат плідної співпраці замовника та виробника електронних виробів. Це документ порозуміння зацікавлених сторін процесу задуму виробу, розробки комплекту документації та виготовлення електронного виробу. Досконало виписані в технічному завданні вимоги до дизайну, розмірів, механічних та електричних параметрів забезпечать випуск якісного, ергономічного, сучасного приладу. Чіткі вимоги технічного завдання до кожного етапу розробки та виробництва, дозволять уникнути помилок, ефективно використати час та фінансову складову електронного виробу.

Недостатньо ретельно обґрунтовані початкові вимоги замовника чи призначені довільно, призводять до невиправданих надлишків в параметрах виробу та його вартості. Також можливість виконання покладених на нього функцій або використання приладу в цілому.

Підписи в технічному завданні – підтвердження поставлених та виконаних вимог та зобов'язань замовника та виробника. Технічне завдання, складене відповідно до сучасних державних стандартів, з застосуванням сучасних матеріалів, новітніх електричних елементів, залученням сучасних методів виробництва тощо – перший крок до отримання якісного електронного виробу.

Список літератури:

1. Євсіна Н.О., Дудник О.В. Конспект лекцій з курсу «Конструювання та технологія виробництва систем телекомунікацій», Харків НТУ «ХПІ», 2022 р. URL https://surl.li/ywcwso (дата звернення: 15.05.2025).

2. Вунтесмері Ю.В. Проектування та конструювання мікроелектронної апаратури, Київ НТУУ «КПІ», 2012 р. URL https://ee.kpi.ua/~yv/edu/pma/pma_kp.pdf (дата звернення: 15.05.2025).

3. ДСТУ 3974-2000. Система розроблення та поставлення продукції на виробництво. Правила виконання дослідно-конструкторських робіт. Загальні положення.

4. СОУ НАН 72.1-003.2:2011. Організація і проведення дослідно-конструкторських робіт, Київ. НАН України. 2025 р.

5. ДСТУ 3278-95. Система розроблення та поставлення продукції на виробництво. Основні терміни та визначення.

6. Проєктування друкованих плат URL https://surli.cc/kxhfwo (дата звернення: 15.05.2025).

7. ДСТУ 8634:2016. Система розроблення та поставлення продукції на виробництво. Настанови щодо розроблення та поставлення на виробництво нехарчової продукції.

8. ДСТУ В 15.001:2023. Система керування життєвим циклом озброєння та військової техніки. Основні положення.

Shornikova S.V. THE TECHNICAL SPECIFICATION IS THE FIRST STEP IN OBTAINING A HIGH-QUALITY ELECTRONIC PRODUCT

High-quality electronic products are the key to the success of any manufacturer. The manufacturer ensures the production of high-quality documentation and electronic products, provided that the customer's requirements are met perfectly. The close relationship between the customer and the manufacturer to obtain a high-quality and reliable product is ensured by the technical specifications.

The technical specification is the first document after signing the contract for the development of documentation and production of the product. All the necessary parametric data that the manufacturer must fulfill must be set out in the technical task.

The article discusses the stages of development of an electronic product: model manufacturing, research and development, developmental stages, adjustment or redesign – accompanied by the requirements specified

in the technical specifications, in accordance with the design stage of the entire product and each module and its components.

The technical task of the stage of developing a new model of an electronic product consists of requirements for the general appearance of the product, requirements for each component module, parametric and electrical data, electronic requirements, etc.

The analyzed stages of preparation of documentation for a working design, which culminate in the serial production of an electronic product, are based on the technical task, taking into account the shortcomings, revised parameters and improved technologies of the preliminary design.

The article discusses practical recommendations regarding the requirements for redesigning the technical task and the need to make changes related to the development and use of the latest materials, improvement of the product with modern electronic components, and development of new technological processes.

The terms of reference are drawn up in accordance with state standards and regulations, industry standards, enterprise standards, and enterprise regulations.

Key words: *terms of reference, research and development work, development work, product redesign, product model, product module, standards.*

ІНФОРМАТИКА, ОБЧИСЛЮВАЛЬНА ТЕХНІКА ТА АВТОМАТИЗАЦІЯ

УДК 004.415.2
DOI https://doi.org/10.32782/2663-5941/2025.3.2/06

Арпентій С.П.
Український науково-дослідний інститут спеціальної техніки та судових експертиз Служби безпеки України

МЕТОДОЛОГІЧНА ПАРАДИГМА АРХІТЕКТУРНОЇ ОРГАНІЗАЦІЇ АВТОНОМНИХ ТРАНСПОРТНИХ СИСТЕМ

Протягом останніх років значний прогрес у сенсорних технологіях та все більш вдосконалених обчислювальних платформах прискорив розвиток автономних транспортних засобів. У статті розглядається комплексний аналіз архітектурних моделей систем автономного керування з акцентом на інтеграцію апаратних і програмних компонентів. Складність, притаманна цим системам, суттєво зросла, що зумовлено необхідністю побудови точних уявлень як про транспортний засіб, так і про його оточення.

Однак підкреслюється, що мінливість операційних областей – від структурованих міських до розлогих сільських ландшафтів – потребує індивідуального та специфічного для кожної області підходу до архітектурного проектування. Аналіз наукової літератури показав, що класифікація архітектури на технічні та функціональні системи не тільки впорядковує процес розробки, але й підвищує адаптивність самокерованих рішень. У статті розкрито ключові аспекти етапу сприйняття, детально розглянуто роль сенсорних підсистем та застосування методів інтеграції даних з метою досягнення надійної локалізації, точного картографування та ефективного виявлення об'єктів.

Окрему увагу приділено оцінці методологій інтеграції сенсорів, підкреслюючи, що інтеграція гетерогенних джерел даних є критично важливою для побудови надійної ситуаційної інформованості. Наголошується, що модульний дизайн і масштабованість цих архітектур є ключовими для пристосування до майбутніх технологічних досягнень, тоді як підкреслюється, що безперервне вдосконалення алгоритмів обробки є важливим для покращення прийняття рішень у реальному часі. Визначено, що нові тенденції в автономному керуванні, ймовірно, будуть зосереджені на подальшій інтеграції передових методів зондування з адаптивною обробкою даних, що обіцяє більш надійні та ефективні рішення.

Ключові слова: *автономні транспортні системи, ROS, рівень зондування та сприйняття, технічна архітектура, функціональна архітектура.*

Постановка проблеми. Попри інтенсивне збільшення наукових досліджень та промислових розробок у сфері автономних транспортних систем, формування єдиної архітектурної бази для автономних систем керування залишається складним завданням, та потребують додаткового опрацювання. Неоднорідність операційних середовищ – від міських доріг до бездоріжжя сільськогосподарських полів – обумовлює потребу у високоспеціалізованих рішеннях. Як наслідок, сфера автономного керування стала джерелом появи різноманітних архітектурних парадигм, пристосованих до конкретних сфер застосування. Однак, незважаючи на це розмаїття, певні базові компоненти та етапи обробки даних повторюються в більшості реалізацій автономних транспортних засобів. До них відносяться сенсорні підсистеми, обчислювальні платформи, системи управління і програмні стеки, які спільно забезпечують сприйняття автономною системою, прийняття рішень і виконання дій в динамічних середовищах [1].

З точки зору системної інженерії, архітектури автономного керування можна умовно розділити на дві основні категорії: технічні та функціональні. Технічні архітектури визначаються конфігурацією апаратного і програмного забезпечення, включаючи набори датчиків (наприклад, LiDAR,

радари, системи технічного зору, GNSS, IMU), вбудовані обчислювальні блоки, комунікаційні мережі та операційні системи реального часу [2]. Функціональні архітектури, навпаки, абстрагують систему на високорівневі процеси і завдання, такі як сприйняття, планування, управління і моніторинг системи. Цей поділ відповідає сучасним підходам системної інженерії на основі моделей (MBSE – model-based systems engineering), прийнятим у розробці автономних транспортних засобів, і пропонує структуровану методологію для декомпозиції складних функціональних можливостей безпілотних транспортних засобів на керовані та інтероперабельні підсистеми [3].

Критично важливим компонентом як технічної, так і функціональної архітектурної структури є рівень зондування та сприйняття [4]. Цей етап відповідає за збір необроблених даних про навколишнє середовище та сприйняття, які згодом обробляються для генерування практичних даних про положення транспортного засобу, навколишнє середовище та поведінку об'єктів або агентів, що знаходяться поблизу. Враховуючи багатофункціональну природу даних, які надходять з різних датчиків з різними просторовими, часовими і спектральними характеристиками, об'єднання даних стає незамінним методом підвищення надійності, точності і надлишковості в системі обробки даних [5].

Об'єднання датчиків або обробка даних – це багатоетапний процес, який включає узгодження даних, часову синхронізацію, асоціацію на рівні ознак і ймовірнісний висновок для інтеграції вимірювань з гетерогенних джерел [6]. Сучасні системи відеоспостереження часто використовують передові алгоритми об'єднання, включаючи байєсівські фільтри (наприклад фільтр Калмана), імовірнісні сітки заповнення та мережі обробки на основі глибокого навчання [7]. Ці методи дозволяють створювати високоточні зображення середовища, що полегшує виконання таких завдань, як одночасна локалізація і картографування (SLAM – simultaneous localization and mapping), семантичне виявлення об’єктів, динамічне відстеження перешкод і прогнозування траєкторій.

Ефективна інтеграція механізмів зондування і обробки даних має першорядне значення для безпечної і надійної автономної навігації, особливо в умовах погіршення роботи датчиків, невизначеності навколишнього середовища або в умовах відсутності сигналу GPS. Крім того, ці механізми повинні бути адаптовані до сфери оперативного проектування (ODD – Operational Design Domain) транспортного засобу, яка визначає екологічні, часові та контекстуальні обмеження, в яких повинна функціонувати автономна система [8]. До цих обмежень входять такі змінні, як тип дороги, погодні умови, рівень освітленості та географічні обмеження, які суттєво впливають на ефективність сенсорів і точність сприйняття інформації.

Незважаючи на те, що в літературі запропоновані еталонні архітектурні моделі для систем автономного керування, ці моделі часто відрізняються за обсягом, рівнем абстракції та специфікою предметної області. Проте, залишається актуальним завдання: створення архітектури, яка б не лише забезпечувала модульність і масштабованість, але й систематично впроваджувала надійні стратегії сприйняття і обробки даних для підтримки основних функцій автономного керування в широкому діапазоні ODD.

Аналіз останніх досліджень і публікацій. В останні роки значна кількість досліджень була спрямована на розробку та класифікацію архітектур систем автономного керування. Ці архітектури, як правило, розмежовуються відповідно до сфери абстракції – або як технічні архітектури, які описують фізичну та програмну реалізацію системи, або як функціональні архітектури, які визначають логічну організацію ключових можливостей, таких як сприйняття, прийняття рішень та управління [9]. Ця відмінність має вирішальне значення для проектування масштабованих, безпечних і багаторазових систем і слугує основою як для концептуального моделювання, так і для реалізації на конкретній платформі.

Фахова література у галузі технічних досліджень містить ряд публікацій, орієнтованих на розробку стратегій інтегрованої взаємодії апаратного комплексу й програмних засобів у автономних транспортних системах. Наприклад, в деяких джерелах представлені традиційні технічні архітектурні концепції для автомобільних застосувань, які акцентують увагу на просторовому і функціональному розподілі процесорів, комунікаційних шин і модулів обробки даних по всьому транспортному засобу [10]. На ці концепції все більше впливають архітектури на основі контролерів доменів, моделі зонних обчислень і парадигми централізованих обчислень, зумовлені вимогами високого рівня автономності (SAE рівні 4–5) [11]. Зокрема, в одному

з прикладів представлено три основні варіанти таких архітектур, проаналізовано їхню здатність обробляти пропускну здатність, надлишковість і відмовостійкість даних у різних операційних сценаріях.

Вимір архітектури програмного забезпечення також зазнав значного розвитку, особливо завдяки уніфікації платформ програмного забезпечення, таких як Robot Operating System 2 (ROS2) та Adaptive AUTOSAR [12]. Ці платформи використовують комунікаційну магістраль Data Distribution Service (DDS), що забезпечує модульний, чутливий до часу і незалежний від платформи зв'язок між програмними вузлами. Така інтеграція полегшує реалізацію повного автономного стеку, зберігаючи при цьому сумісність з критично важливими для безпеки автомобільними стандартами. Нові реалізації підкреслюють оптимізацію затримок обміну даними між процесами, планування повідомлень на основі QoS і розширену підтримку моніторингу безпеки в реальному часі.

Водночас практичні рішення і експериментальні платформи, як промислові, так і науково-дослідні, показують реальне втілення цих технічних планів. Наприклад, промислові тематичні дослідження детально описують комплексні проекти програмних систем, вбудованих у комерційні транспортні засоби, виділяючи такі аспекти, як обробка надлишкової інформації, можливості оновлення програмного забезпечення через «повітря» (OTA – Over-the-Air) і конвеєри попередньої обробки даних з датчиків [13]. З іншого боку, наукові платформи зосереджуються на відкритих дослідницьких платформах, пристосованих до зовнішнього середовища, з акцентом на модульні системи датчиків, конфігурацію обчислювального стеку і конвеєри обробки даних для автономних навігаційних завдань.

З функціонального боку, архітектура системного рівня, як правило, будується за триєдиною парадигмою: сприйняття, прийняття рішень (або планування) і виконання (керування транспортним засобом). Ця модель розвивається шляхом введення двох ієрархічних рівнів – когнітивного інтелекту оператора та інтерфейсу платформи транспортного засобу, пропонуючи структурований підхід до відокремлення алгоритмічних міркувань від апаратного виконання. У недавніх публікаціях запропоновано вдосконалені функціональні архітектури, в яких кожен блок систематично декомпозується на модульні підсистеми для підтримки забезпечення відстежуваності, масштабованості та відповідності стандартам, зокрема, ISO 26262 і SOTIF (Safety of the Intended Functionality – безпека передбачуваної функціональності) [14].

В одному з підходів запропоновано архітектуру, побудовану на основі V-моделі, що складається з п'яти рівнів абстракції: сприйняття навколишнього середовища, оцінка ситуації, планування поведінки, генерація траєкторії та виконання керування. Ця багаторівнева структура підтримує формальну верифікацію, уможливлює модульні процедури сертифікації та полегшує послідовну інтеграцію з середовищами цифрових копій, які використовуються для симуляції та валідації.

Крім того, були розроблені фреймворки для формалізації функцій контролю і моніторингу як невід'ємних компонентів автономної системи. За допомогою формальних методів моделювання ці фреймворки виводять поведінкові моделі з функціональних вимог і вимог безпеки, підтримуючи виявлення аномалій під час виконання і стійкість до збоїв за допомогою формальних механізмів верифікації, таких як часові логічні висновки і перевірка моделі. У низці комплексних досліджень робиться спроба подолати техніко-функціональну розбіжність, пропонуючи різносторонній аналіз існуючих архітектур систем автономного керування транспортними засобами. Ці порівняльні дослідження включають як комерційні рішення, так і обмежені в ресурсах дослідницькі прототипи. Зокрема, у дослідженні порівнюються чотири архітектури на основі спільного проектування стеку програмно-апаратних засобів, оцінюється їхня сумісність, механізми безпеки та ефективність роботи в реальному часі. В іншому дослідженні розглядаються три функціонально керовані архітектури з точки зору надійності та модульності, порівнюючи їх масштабованість за різних умов трафіку та навколишнього середовища.

Особливий інтерес викликає сенсорний компонент, який становить основу модуля сприйняття в будь-якій автономній системі. Останні дослідження акцентують увагу на інтеграції гетерогенних сенсорів – LiDAR, радарів, камер, GNSS-IMU – за допомогою складних алгоритмів обробки даних [15]. Такі алгоритми працюють на кількох рівнях абстракції: обробка первинних даних, обробка на рівні ознак і обробка на рівні рішень. Інтеграція сенсорів не лише підвищує надійність виявлення та класифікації об'єктів,

але й має вирішальне значення для точної локалізації та побудови карт у реальному часі. Досягнення останніх років у методах одночасної локалізації і картографування (SLAM – simultaneous localization and mapping), включаючи візуально-інерційну SLAM з глибоким навчанням і семантичну SLAM, продемонстрували значне підвищення точності і обчислювальної ефективності, особливо в умовах відсутності GPS або в умовах перешкод [16].

В цілому, у наукових публікаціях відображається перехід до високомодульних, функціональних і орієнтованих на безпеку архітектур автономного керування. Інтеграція вдосконаленого зондування, надійного програмного забезпечення для зв'язку та формалізованих моделей поведінки позиціонує автономні системи наступного покоління як такі, що будуть відповідати жорстким вимогам безпеки, ефективності та регуляторним вимогам у різних сценаріях застосування.

Постановка завдання. Незважаючи на значний прогрес в галузі автономних транспортних засобів за останнє десятиліття, методологічні засади систематичного проектування архітектур автономного керування залишаються недостатньо формалізованими. Сучасні дослідження зосереджуються або на окремих технічних рішеннях, або на абстрактних функціональних моделях, не поєднуючи теоретичні підходи з фактичними експлуатаційними обмеженнями.

Особлива увага приділяється проблемі інтеграції сенсорних даних. При широкому застосуванні LiDAR, радарів, камер та інерційних вимірювальних приладів існує фрагментарність у комбінуванні їхніх сигналів для забезпечення точного позиціонування, семантичного картографування та динамічного виявлення об’єктів, що призводить до розбіжностей у ефективності автономного керування в складних середовищах.

Метою даної статті є систематичний аналіз сучасних підходів до формування архітектури керування автономними транспортними засобами, з особливою увагою до сенсорних підсистем, огляд класифікації архітектур за функціональними та технічними критеріями, які визначають їх доцільність залежно від сфери застосування та проведення оцінки ролі алгоритмів обробки даних у забезпеченні ключових автономних функцій, включаючи прийняття рішень, адаптивність до змін зовнішнього середовища та забезпечення безпеки руху.

Виклад основного матеріалу. Архітектурний дизайн систем автономного керування зазвичай поділяється на технічну та функціональну складові, залежно від рівня абстракції та сфери застосування. Побудову архітектури систем автономного керування можна розглядати з двох взаємодоповнюючих точок зору: технічної та функціональної. Технічна концепція підкреслює взаємодію між апаратним і програмним забезпеченням, інтеграція яких є основою сучасних платформ безпілотного керування. На апаратному рівні сучасні датчики, зокрема камери, LiDAR, радари та ультразвукові пристрої, поєднуються з гетерогенними обчислювальними платформами. Багатоядерні процесори, графічні процесори і FPGA замінили собою традиційну залежність від обмеженої кількості електронних блоків управління, що надало змогу обробляти величезні потоки даних, які генеруються як з бортових, так і з зовнішніх джерел. Для забезпечення ефективної передачі даних між різними підсистемами використовуються високошвидкісні з'єднання (наприклад, Gigabit Ethernet і USB 3.x), які поєднуються із застарілими мережами, такими як CAN і LIN. Ця архітектурна тенденція ще більше посилюється завдяки технологіям зв’язку «транспортний засіб – будь-що» (V2X), які інтегрують дані з датчиків транспортного засобу, інтернет-джерел та інфраструктурних даних, що сприяє підвищенню рівня інформованості про навколишню обстановку [17]. Що стосується програмного забезпечення, то ці рівні еволюціонували від вбудованого програмного забезпечення, що працює в режимі реального часу на спеціальних платформах RTOS, до високорівневих фреймворків, призначених для завдань машинного навчання (ML), штучного інтелекту (AI) і глибокого навчання (DL). Сучасні програмні архітектури, такі як AUTOSAR та екосистема ROS2 встановлюють стандартизовані рекомендації та компонентні конструкції, які полегшують одночасну роботу в розподіленій системі. Вони не лише підтримують збір і обробку даних з датчиків, але й дозволяють безперешкодно інтегрувати критичні для безпеки функції, необхідні для автономних операцій. Поєднання апаратного і програмного забезпечення призвело до того, що їх часто називають «суперкомп'ютерами на колесах», де кожен компонент взаємодіє в рамках добре організованої мережі, яка відповідає сучасним вимогам безпеки в реальному часі і функціональній безпеці.

З функціональної точки зору, системи автономного керування поділяються на логічні блоки, у відповідності до потоку інформації з датчиків,

до його активації та контролю. Модуль обробки інформації є особливо важливим, оскільки він узагальнює дані як від пропріоцептивних датчиків (наприклад, GNSS, IMU, енкодерів), так і від екстероцептивних датчиків (наприклад, LiDAR, радарів, камер), формуючи цілісну модель навколишнього середовища. Для забезпечення надійного, точного і контекстно-орієнтованого прийняття рішень, інтеграція даних, яку також називають мультисенсорною інтеграцією, стає фундаментальною обчислювальною методологією. У цьому контексті інтеграція сенсорів відіграє незамінну роль, узгоджуючи кілька потоків даних з різною роздільною здатністю і часовими характеристиками для виконання точної локалізації та картографування. Застосування удосконалених алгоритмів інтеграції, які часто ґрунтуються на імовірнісних методах і глибокому навчанні, дає змогу виявляти об'єкти з високою точністю, зменшуючи невизначеність, притаманну окремим вимірюванням датчиків. Такі методи, як одночасна локалізація і картографування (SLAM), ілюструють, як ітеративні корекції даних можуть врахувати дрейф, компенсуючи динамічну природу міських і неструктурованих середовищ. Процес інтеграції об'єднує просторові та часові сигнали, де високочастотні інерційні дані об'єднуються з щільними хмарами точок з LiDAR або насиченими візуальними сигналами з камер. Такий синтез не лише підвищує надійність представлення навколишнього середовища, але й безпосередньо інформує модулі планування та прийняття рішень. Тут динамічні карти керують стратегією навігації, безперервно перекалібровуючи траєкторію транспортного засобу у відповідь на виявлені перешкоди і зміни маршруту. Системи керування згодом перетворюють ці високорівневі плани на точні дії, модулюючи команди виконавчих механізмів на основі динаміки транспортного засобу та поточного операційного контексту [18]. Процес інтеграції даних може відбуватися на різних рівнях – на рівні необроблених даних, на рівні функцій або на рівні рішень – залежно від вимог програми і архітектури системи. Крім того, для контролю за роботою апаратних і програмних компонентів використовуються системи моніторингу та спостереження, які гарантують, що апаратні і програмні компоненти працюють в рамках встановлених обмежень безпеки, тим самим зменшуючи ризики, пов'язані з несправністю компонентів або невідповідністю даних.

Таким чином, подвійна класифікація архітектур автономного керування – технічна і функціональна – забезпечує цілісну основу для вирішення складних питань автономності транспортних засобів. Можливо констатувати і той факт, що інтеграція даних з GNSS, IMU, сенсорів LiDAR, камер та радарів, при використанні алгоритмів багатосенсорного об’єднання даних, суттєво знижує похибки вимірювань, адже завдяки застосуванню, зокрема, Калманівського фільтру, помилка позиціонування може бути зведена до рівня близько 5 см [19]. Завдяки узгодженому підходу до калібрування та синхронізації даних усіх сенсорів забезпечується референсна точка, що дозволяє об’єднувати дані з різних джерел у єдиний часово-просторовий каркас.

На практиці інтеграція даних з LiDAR та камер за допомогою алгоритмів SLAM дозволяє формувати карти з роздільною здатністю до 1 см, що є критичним для розуміння навколишнього середовища та прийняття рішень в режимі реального часу. Крім того, використання глибоких нейронних мереж для обробки даних з різних сенсорів забезпечує підвищення точності розпізнавання об'єктів на 30% у порівнянні з підходами, заснованими на окремих сенсорах. Як наслідок, алгоритми, що поєднують дані з камер, радарів та LiDAR, сприяють значnove зниженню ризику аварійних ситуацій, що підтверджується недавніми експериментальними випробуваннями, де показано зменшення аварійності на 40%.

Слід зазначити, що система, яка інтегрує дані із внутрішніх та зовнішніх сенсорів, показує суттєву перевагу в обробці даних. Наприклад, у випадку збоїв GNSS, система може безперебійно використовувати INS та дані LiDAR для підтримання стабільної навігації, що гарантує підвищену надійність руху. Використання гібридних обчислювальних платформ із GPU та FPGA дозволяє зменшити затримки обробки сенсорних даних до менше 10 мс, що критично для своєчасної реакції системи на динамічні зміни у середовищі руху. Така інтеграція забезпечує не лише високий рівень точності, але і стійкість до можливих втрат або дефектів окремих даних від певних сенсорів.

Впровадження методики інтеграції в системах керування відкриває нові горизонти для оптимізації алгоритмів автономного керування, оскільки кожен сенсор компенсує недоліки іншого, а інтеграція даних сприяє досягненню синергетичного ефекту. Подібний підхід довів свою ефективність у дослідженнях 2023 року, де комбіноване використання даних від камер,

LiDAR і радарів сприяло на 25% швидшому реагуванню транспортних засобів на перешкоди. Завдяки цьому, сучасні автономні системи набувають характеристик вимірювань параметрів з високою точністю і швидкого адаптивного планування маршруту, що фізично підтверджується зменшенням похибок локалізації на 50% у складних навколишніх умовах [20].

Висновки. Еволюція архітектур автономного керування, які поділяються на технічні та функціональні області, має вирішальне значення для досягнення надійної автономної навігації в реальному часі. За останнє десятиліття значний прогрес у сенсорних технологіях, мережевих можливостях і високопродуктивних обчисленнях перетворив традиційні архітектури на інтегровані «суперкомп'ютери на колесах», де технічна основа забезпечує ефективний збір, обробку і передачу величезних обсягів сенсорних даних. Водночас, функціональна складова вдосконалює модуль обробки даних, зосереджуючись на об'єднанні різнорідних потоків даних для створення точного, динамічного уявлення про транспортний засіб та його оточення.

Досягнення в галузі LiDAR, радарів, стереокамер і камер глибини дозволяють встановлювати кілька датчиків одного типу, що, в свою чергу, дає можливість більш детального відображення зовнішніх сценаріїв. Так за рахунок інтеграції GNSS, IMU і одометрії за допомогою таких методів, як фільтрація Калмана, точність локалізації може бути покращена до кількох сантиметрів. Аналогічно, застосування алгоритмів SLAM призвело до значного підвищення роздільної здатності картографування довкілля, з помітним покращенням вилучення об’єктів і точності карт. Крім того, механізми виявлення об’єктів на основі глибокого навчання, які об’єднують сигнали з камер, LiDAR і радарів, дозволили підвищити точність розпізнавання об'єктів на 30% порівняно з системами, що використовують дані з одного датчика. Двоїстість цих архітектур – коли надійна технічна платформа доповнюється оптимізованим функціональним дизайном – забезпечує синергетичну взаємодію компонентів системи, долаючи таким чином такі проблеми, як синхронізація датчиків, пропускна здатність даних і обмеження обробки в реальному часі.

Важливо зазначити, що синергія технічної надійності та функціонального інтелекту, досягнута шляхом розширеного злиття даних, є суттєвим елементом у досягненні автономії на рівнях 4 та 5 відповідно до стандарту SAE.

Список літератури:

1. Buehler K., Iagnemma K., Singh S. The DARPA Urban Challenge: Autonomous Vehicles in City Traffic, Springer Tracts in Advanced Robotics. Vol. 56. 2009. P. 3–30.
2. Smith A., Johnson B. Impact of Systems Engineering on Autonomous Vehicle Architecture, IEEE Systems Journal, Vol. 15. No. 3. 2023. P. 155–165.
3. Evans D.R., Lewis P. Model-Based Systems Engineering Approaches in Automated Vehicle Design, Systems Engineering. Vol. 27. No. 5. 2020. P. 389–402.
4. Nguyen T., Hernandez F., Becker R. Sensing and Perception in Autonomous Driving: Challenges and Advancements, IEEE Sensors Journal. Vol. 20. 2021. P. 1022–1035.
5. Liu S., O’Neill P. Data Fusion Methods for Robust Perception in Autonomous Vehicles, IEEE Trans. on Intelligent Transportation Systems. Vol. 21. No. 2. 2020. P. 876–888.
6. Peterson R., Gupta A. Multi-Sensor Data Integration Techniques for Autonomous Navigation, Robotics and Autonomous Systems. Vol. 95. 2021. P. 76–89.
7. Cheng Y., Wang F., Li J. Advancements in Bayesian Filtering and Deep Learning for Sensor Fusion in Autonomous Vehicles, IEEE Access. Vol. 9. 2021. P. 1024–1026.
8. Kim J., Park S. Adaptive Operational Design Domain Specification for Autonomous Vehicles, IEEE Intelligent Systems. Vol. 28. No. 4. 2023. P. 55–63.
9. Garcia M., Zhang L., Brown T. Functional Architectures for Autonomous Vehicles: A Comparative Study, IEEE Trans. on Intelligent Transportation Systems. Vol. 23. 2023. P. 210–225.
10. Martinez O., Roberts S. Traditional Technical Architecture Concepts for Automotive Applications, Proc. IEEE Int. Conf. Intelligent Transportation Systems. 2021. P. 123–130.
11. Thrun S., Montemerlo M. Domain Controller–Based Architectures for High-Autonomy Vehicles: Zonal vs. Centralized Approaches, IEEE Trans. on Intelligent Vehicles. Vol. 5. No. 2. 2022. P. 50–67.
12. Guo Q. Software System of Autonomous Vehicles: Architecture, Network and OS, University of Pennsylvania. 2020. P. 1–50.
13. Gyagenda N., Hatilima J.V., Roth H., Zhmud V. A Review of Over-the-Air Update and Data Pre–Processing Techniques in Autonomous Vehicles, Robot. Auton. Syst. Vol. 152, 2022, P. 104069–104073.

14. Garcia M., Lee K., Patel S. Modular Decomposition and Standardization in Functional Architectures for Autonomous Systems, Robot. Auton. Syst. Vol. 168. 2022. P. 45–57.

15. Cheng Y., Wang F. Advanced Algorithms for Multi-Sensor Data Fusion in Autonomous Driving, Robot. Auton. Syst. Vol. 157. 2022. P. 78–90.

16. Gyagenda N., Hatilima J.V., Roth H., Zhmud V. A Review of SLAM Techniques for Autonomous Navigation under Adverse Conditions, Robot. Auton. Syst. Vol. 152. 2022. P. 104069–104073.

17. Zhukov B., Kovalchuk P. V2X communication for enhanced situational awareness in autonomous vehicles, IEEE Commun. Mag. Vol. 60. No. 8. 2022. P. 36–44.

18. Panov Y., Sokolov E. Iterative correction in SLAM for dynamic urban environments, Robot. Auton. Syst. Vol. 155. 2022. P. 90–100.

19. Kovalenko P., Smirnov I. Sensor Fusion Using the Kalman Filter in Autonomous Navigation Systems, Robot. Auton. Syst. Vol. 157. 2022. P. 78–86.

20. Baranov I., Melnik A. Advanced Sensor Integration Techniques for Reducing Localization Errors in Challenging Environments, Robot. Auton. Syst. Vol. 158. 2023. P. 210–219.

Arpentii S.P. METHODOLOGICAL PARADIGM OF ARCHITECTURAL ORGANISATION OF AUTONOMOUS TRANSPORT SYSTEMS

In recent years, significant advances in sensor technologies and increasingly sophisticated computing platforms have accelerated the development of autonomous vehicles. This paper presents a comprehensive analysis of architectural models of autonomous driving systems with a focus on the integration of hardware and software components. The complexity inherent in these systems has increased significantly due to the need to build accurate representations of both the vehicle and its environment.

However, it is emphasised that the variability of operational areas – from structured urban to sprawling rural landscapes – requires an individual and area-specific approach to architectural design. An analysis of the scientific literature has shown that the classification of architecture into technical and functional systems not only streamlines the development process, but also increases the adaptability of self-managed solutions. The article reveals the key aspects of the perception stage, discusses in detail the role of sensor subsystems and the use of data integration methods to achieve reliable localisation, accurate mapping and efficient object detection.

Particular attention is paid to the evaluation of sensor integration methodologies, emphasising that the integration of heterogeneous data sources is critical to building reliable situational awareness. It is noted that the modular design and scalability of these architectures are key to adapting to future technological advances, while it is emphasised that continuous improvement of processing algorithms is essential to improve real-time decision-making. It is determined that new trends in autonomous control are likely to focus on the further integration of advanced sensing techniques with adaptive data processing, which promises more reliable and efficient solutions.

Key words: *autonomous transport systems, ROS, sensing and perception level, technical architecture, functional architecture.*

УДК 004.42
DOI https://doi.org/10.32782/2663-5941/2025.3.2/07

Астістова Т.І.
Київський національний університет технологій та дизайну

Дроменко В.Б.
Таврійський національний університет імені В.І. Вернадського

Калашник В.Ю.
Київський національний університет технологій та дизайну

РОЗРОБЛЕННЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ІНФОРМАЦІЙНОГО РЕСУРСУ БІБЛІОТЕКИ ЗАКЛАДІВ ОСВІТИ

В статті наведено аналіз використання цифрових технологій в інформаційних системах наукових бібліотек провідних університетів України. На основі проведеного аналізу визначені функціональні вимоги та завдання для розроблення інтерактивного додатку при взаємодії з бібліотекою закладу освіти за сучасними стандартами.

Архітектура розробленого web-додатку містить три основні компоненти: клієнтський компонент, серверний компонент і сервер бази даних, яка реалізована за допомогою використання шаблону MVC (Model-View-Controller).

Клієнтська частина реалізована за допомогою інтегрованого середовища розробки програмного забезпечення Intellij Idea та системи для побудови шаблонів і сторінок мовою розмітки HTML та стилізована за допомогою фреймворку Bootstrap та CSS та JS. Для динамічної генерації HTML-шаблонів було використано сучасний серверний шаблонізатор для Java Thymeleaf. Для реалізації серверної частини використано мову програмування Java та фреймворк SpringBoot з використанням Java бібліотек. Для управління базою даних обрано PostgreSQL. Для ORM (Object-Relational Mapping) був обраний Hibernate, який забезпечує зручний спосіб взаємодії з базою даних через об'єкти Java Hibernate, Spring Data JPA.

Розроблений web-додаток є закінченим програмним продуктом, який пройшов апробацію та використовується в Київському фаховому коледжу прикладних наук, і забезпечує підтримку мультимедійних ресурсів з субтитрами та жестовою мовою для здобувачів з вадами слуху. В проєкті реалізована генерація QR-коду для кожної книги, за допомогою якого читач може отримати завчасно додану до своїх улюблених книжку. Запроваджена система зменшує часові витрати користувача та навантаження на бібліотекаря, ліквідує паперову документацію та робить процедуру отримання книг легкою та прозорою.

Ключові слова: *архітектура, сервіси, методи, класи, контролери, технологічний стек, MVC, PostgreSQL, інтегроване середовище Intellij Idea, Framework java Spring Boot , HTML, CSS.*

Постановка проблеми. Сучасний навчальний процес у закладах освіти та сфери, які пов›язані з цим, все більше потребують використання цифрових технологій з метою підвищення якості освіти [1].

Бібліотека, як частина закладу освіти, також потребує інновацій та оновлення застарілої системи інформаційного ресурсу. Сучасні технології змінили вимоги і до організації діяльності бібліотек. Зміни відбувалися у напрямку доступності та модернізації роботи бібліотеки, у використанні сучасних технологій та фреймворків, інформаційних функціях, розробленні баз даних, поліпшені масштабованості, дизайну, інтерактивності. Сучасна бібліотека не лише забезпечує доступ до різноманітних джерел інформації, але й створює середовище, сприятливе для навчання, розвитку та надає здобувачам та викладачам доступ до книг, наукових журналів, електронних ресурсів та інших навчальних матеріалів, що є важливими для навчання та досліджень [2]. З розвитком хмарних технологій, штучного інтелекту та інших інноваційних напрямків, потреба в програмному забезпеченні лише зростає. Для вирішення поставлених проблем, потрібно побудувати гнучку систему, що легко адаптується на основі нових принципів з використанням найсучасніших комп›ютерних та інформаційних технологій, які мають в своєму розпорядженні засоби для організації будь-яких процесів, у тому числі і в освітній сфері.

Електронні бібліотеки використовують електронний формат видань та забезпечують доступ до віддаленого ресурсу за допомогою телекомунікаційних технологій. Читач, який звертається до бібліотеки, повинен мати можливість отримати необхідну інформацію, як в електронному вигляді, так і в надрукованому [3]. Система особистого кабінету читача повинна мати користувацький інтерфейс з сучасним дизайном для зручного перегляду, замовлення, бронювання та сортування книг, наявних в бібліотеці. Найбільшим недоліком в інтерфейсі вебдодатків багатьох бібліотек закладів освіти є ігнорування потреб людей з обмеженими можливостями, що значно обмежує доступність і зручність використання, що може призвести до виключення великої кількості потенційних користувачів.

Аналіз останніх досліджень і публікацій. Існуючі платформи для розробки онлайн електронно-інформаційних ресурсів пропонують цінні інформаційні ресурси, але обмеження з точки зору доступності, функціональності та технологій перешкоджають їхній ефективності для користувачів з різноманітними потребами.

Щоб окреслити функціональні вимоги до розробки онлайн бібліотеки для закладу освіти, проведемо аналіз існуючих інформаційних систем наукових бібліотек провідних університетів України, зокрема: Київського національного університету імені Тараса Шевченка (КНУ), Львівського національного університету імені Івана Франка (ЛНУ), Національного університету «Києво-Могилянська академія» (НаУКМА), Національного університету біоресурсів і природокористування України (НУБіП), Київського національного університету технологій та дизайну (КНУТД).

1. Наукова бібліотека КНУ ім. Тараса Шевченка [4]. Електронна бібліотека одного з провідних київських університетів має великий книжковий фонд та багату кількість параметрів пошуку інформації. Інтерфейс каталогу має загальну продуманість для швидкого і легкого пошуку інформації, максимальну зрозумілість всіх елементів управління. До недоліків системи можна віднести використання застарілих версій мов програмування (використання php версії 4.4.9, яка була випущена ще у 2008 році) (див. рис. 1). Актуальна версія бібліотеки потребує оновлення інтерфейсу та застосування більш новітніх технологій. Користувацький інтерфейс не відповідає вимогам UI/UX дизайну, які передбачають ознайомлення з інтерфейсом майже не витрачаючи часу на те, щоб розібратись, як все працює.

2. Наукова бібліотека ЛНУ ім. Івана Франка [5]. Вебсайт наукової бібліотеки пропонує користувачам широкий спектр функцій, зокрема: пошук книги за назвою, автором, темою, ISBN/ISSN та іншими критеріями. Система пошуку з автозаповненням та розширеним пошуком дає змогу користувачам знайти шукану інформацію, але в кінцевому результаті можна отримати лише інформацію про наявність книги у бібліотеці (рис. 2). Електронних джерел інформації система не пропонує, що є великим мінусом у сучасному світі, також сайт не адаптовано під мобільні телефони.

3. Наукова бібліотека Національного університету «Києво-Могилянська академія» [6]. Система онлайн бібліотеки на перший погляд має найбільш привабливий інтерфейс та функціонал на сьогоднішній день (рис. 3).

Але при роботі з функціоналом (натисканні клавіш) відбувається переадресація на застарілу інформаційну систему (рис. 4) у якій інтерфейс не відповідає сучасним вимогам.

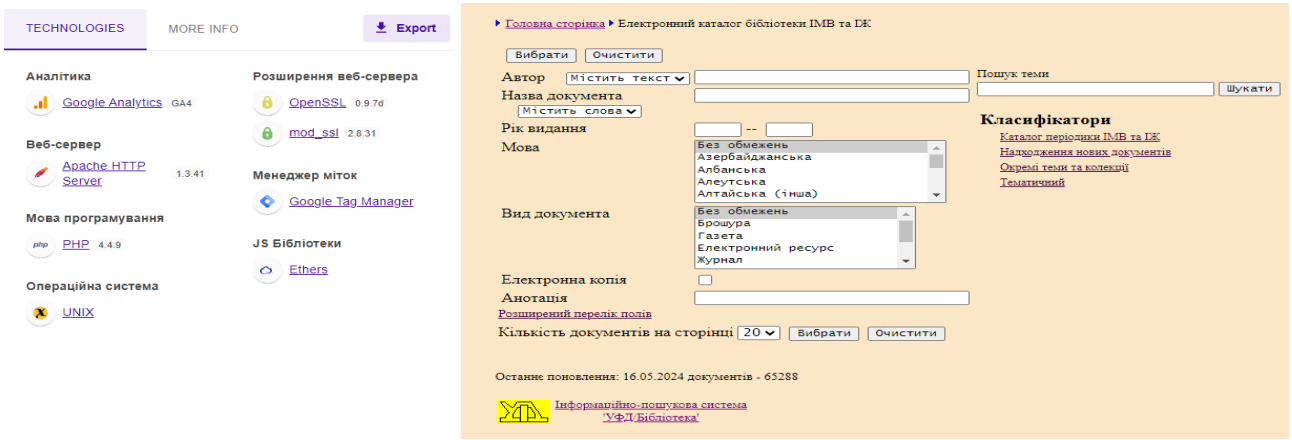


Рис. 1. Аналіз технологій та інтерфейс каталогу бібліотеки КНУ

II 308008
Шевченко, Артем.
Слов'янськ. Початок війни [Текст] / Артем Шевченко ; [вступна стаття А.Авакова]. — Харків : "Фоліо", 2020. — 432 с., [24] с. кольор. фотоіл. : фотоіл. — (Хроніка). — Дар Bookforum 28.
ISBN 978-966-03-9102-4

- **Серія**

Хроніка.

- **Теми документа**

  - УДК // Українська література

Примірники		
Місцезнаходження	**Кількість**	**В наявності**
Книгозберігання	1	1

Рис. 2. Результат пошуку у бібліотеці ЛНУ

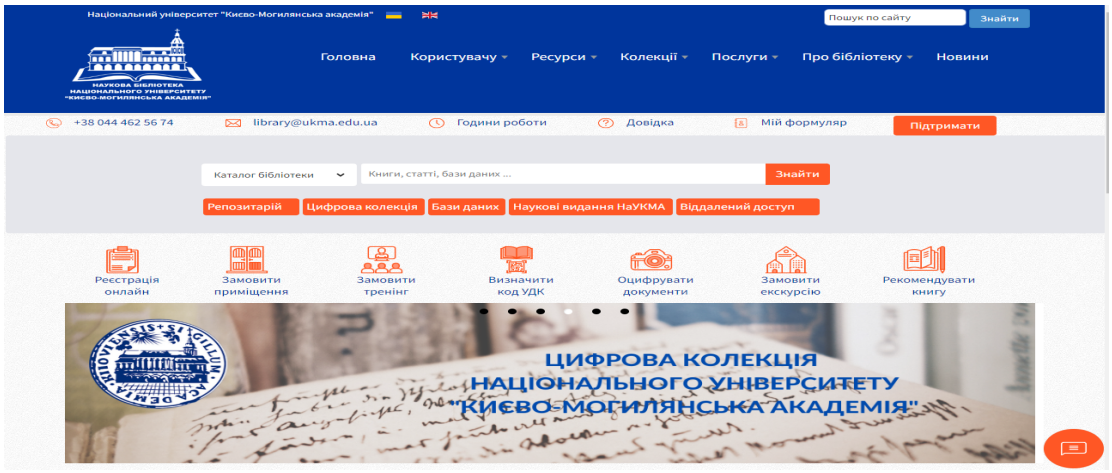


Рис. 3. Інтерфейс наукової бібліотеки НаУКМА

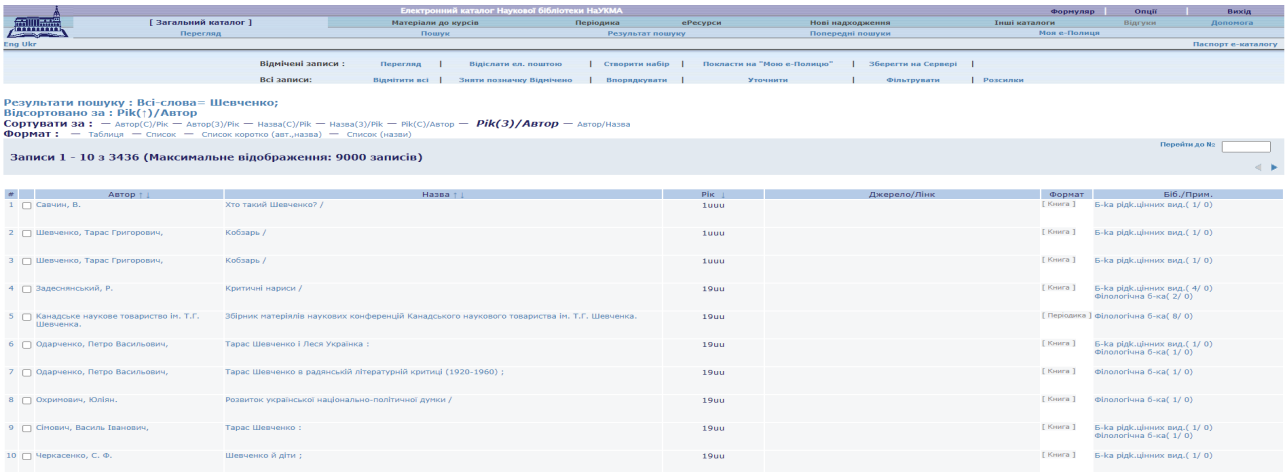


Рис. 4. Застарілий інтерфейс бібліотеки НаУКМА

4. Наукова бібліотека Національного університету біоресурсів і природокористування України [7]. Сайт має достатньо привабливий, стриманий інтерфейс та наявність корисних функцій для користуванням електронним ресурсом (рис. 5).

Інтерфейс відображає весь наявний каталог, що спрощує взаємодію користувача з відображеною літературою, а панелі сортування та вибору категорій наукової літератури додають інтерактивності.

До мінусів в наявній системі можна віднести відсутність кабінету читача, що робить неможливим зберігати свої уподобання, історію пошуку та персональні налаштування, що покращило б зручність використання бібліотеки. Читач не може переглядати свої позики, замовляти та резервувати книги.

5. Наукова бібліотека Київського національного університету технологій та дизайну [8]. Бібліотека пропонує доступ до електронних каталогів, академічних баз даних та різних подій, таких як вебінари і наукові конференції тощо (рис. 6).

Головне меню дає змогу знайти основні розділи бібліотеки, що робить навігацію зрозумілою, але деякі складові підменю не інтуїтивні для нових користувачів і потребують додаткових підказок або інтерактивних елементів, які могли б покращити взаємодію з каталогом. Візуальний дизайн інтерфейсу не відволікає від основного контенту, але потребує оновлення до відповідності сучасним стандартам.

Проведений аналіз використання цифрових технологій в інформаційних системах наукових бібліотек українських університетів показав, що кожна з них має свої переваги та недоліки, які перешкоджають ефективному використанню цих ресурсів здобувачами та викладачами. Основні проблеми стосуються застарілих технологій (системи використовують застарілі версії мов програмування та ін'єкцій залежностей, що робить їх вразливими до безпекових загроз і ускладнює їхнє подальше оновлення та масштабування), недосконалого користувацького інтерфейсу (інтерфейси бібліотек не відповідають сучасним вимогам UI/UX дизайну [9], що ускладнює навігацію та зменшує зручність використання), обмеження функціональних можливостей (вебсайти наукових бібліотек не надають доступ до електронних версій

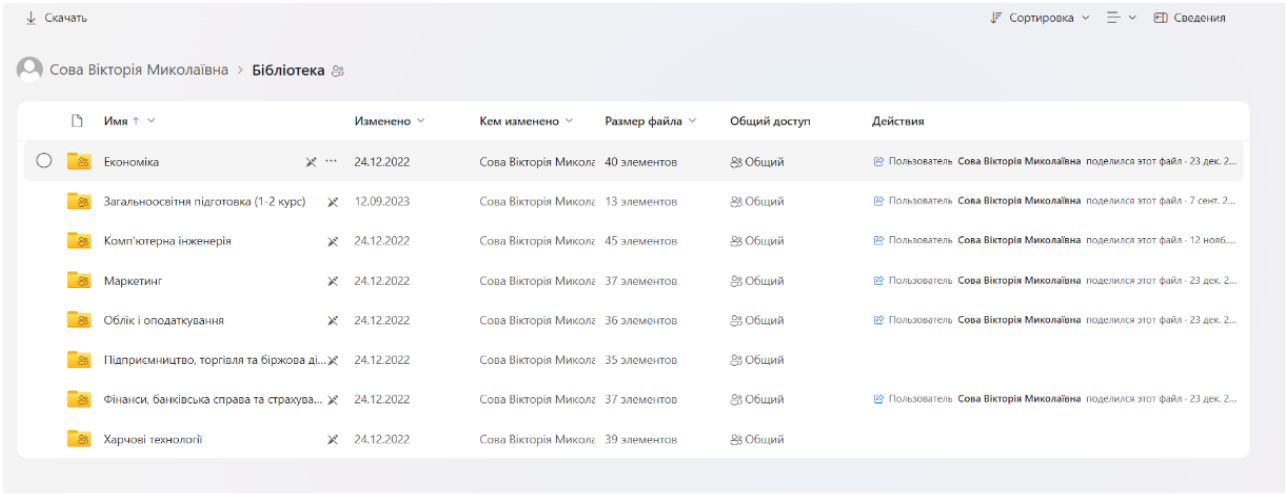


Рис. 5. Інтерфейс електронного каталогу НУБіП

Рис. 6. Інтерфейс електронного каталогу КНУТД

книг), використовуються CMS, які не дозволяють гнучко налаштовувати та масштабувати системи в майбутньому, відсутність адаптації під мобільні пристрої (бібліотеки не надають доступ до електронних версій книг, що є суттєвим недоліком у сучасному цифровому світі) і головне – у жодній з цих наукових бібліотек немає функцій, які б сприяли доступності людей з обмеженими можливостями, що є суттєвим недоліком на сьогодні.

Дослідження існуючих рішень показало, що жодне з них не в повній мірі відповідає критеріям повноцінної функціональності та інформаційної адекватності та виявило кілька проблем, які можна ефективно вирішити за допомогою розробки програмного забезпечення [10].

Постановка завдання. Метою роботи є розроблення програмного забезпечення для віддаленої взаємодії з бібліотекою у вигляді *web*-додатку, який дає змогу ознайомитись з наявними літературними джерелами та можливостями бібліотеки, підвищить швидкість опрацювання матеріалу, зробить систему кросплатформну, ліквідує паперову документацію, що зробить процедуру отримання книг легким та прозорим для всіх читачів.

Створений сервіс у вигляді *web*-додатку, дозволить перебудувати застарілу систему роботи бібліотеки завдяки інтерфейсам бібліотекар-комп'ютер та відвідувач-комп'ютер, привабить більшу кількість відвідувачів та заохотить здобувачів проводити більше часу за книгами. Увага в розробці була приділена інклюзивності та доступності до інформаційних джерел всіх категорій користувачів, забезпечуючи широке охоплення аудиторії, що буде сприяти подальшому розвитку, а можливо й інвестицій в цей проєкт [3].

Для досягнення мети, були вирішені такі завдання:

- аналіз алгоритму взаємодії між читачем та працівниками бібліотеки;
- розробка інтуїтивного та зручного інтерфейсу користувача програмного продукту для людей з обмеженими можливостями;
- вибір оптимального типу архітектури системи;
- вибір технологій та інструментів;
- розробка модулів для забезпечення функціональності бібліотеки, налагодження та тестування системи;
- тестування по зручності користування розробленої інформаційної системи.

Виклад основного матеріалу. Архітектура *web*-додатку включає три основні компоненти: клієнтський компонент, серверний компонент і сервер бази даних. Архітектурна частина додатка реалізована за допомогою використання шаблону *MVC* (*Model-View-Controller*), що дозволяє розділити окремо інтерфейс користувача, базу даних (моделі) та логіку [11]. Завдяки цьому гнучкому патернові подальші зміни та розширення проєкту можуть бути здійснені без жодних проблем, а структура додатку стає більш зрозумілою як окремо для *frontend* (розробник клієнтської частини) так і для *backend* (розробник серверної частини додатку) [12].

Інтеграція з серверною частиною додатку є важливим етапом розробки *frontend*. Клієнтська частина дає можливість взаємодіяти з базою даних, обробляти запити користувачів та надавати функціональні можливості, якими користуються на серверному боці. Правильна взаємодія між клієнтської та серверною частинами гарантує ефективне і надійне функціонування всього *web*-додатку.

Крім того, розробка інтерфейсу для *web*-ресурсу включає в себе важливий етап тестування, який дозволяє виявляти та усувати помилки, адаптувати дизайн до різних величин екранів та переконатися, що користувачі можуть легко взаємодіяти з усіма елементами сторінки чи додатку. Загалом, розробка інтерфейсу для *web*-ресурсу є ключовим компонентом сучасних технологій та *web*-розробки. Вона впливає на враження користувачів від використання *web*-додатку [12–13].

Розроблення клієнтської частини *web*-додатку було виконано для Київського фахового коледжу прикладних наук. Технологічний стек обраних технологій представлено на рис. 7.

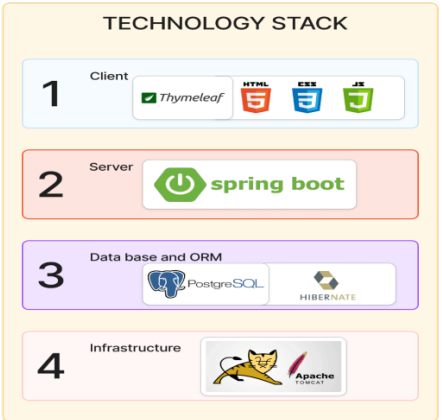


Рис. 7. Технологічний стек обраних технологій

В закладі навчаються здобувачі з вадами слуху, яким слід приділити особливу увагу, адже бібліотека забезпечує їм доступ до інформації, яка може бути недоступною з інших джерел, слід врахувати їхні потреби, що до дизайну та доступності сприйняття жестової мови. Наприклад, бібліотека може надавати послуги з перекладу інформації мовою жестів та використовувати сторонні сервіси для допомоги у користуванні бібліотекою. Використання сучасних цифрових сервісів та мультимедійного контенту для людей з порушенням слуху є невід'ємною складовою при розробці клієнтської частини. На рис. 8 наведено головну сторінку клієнтської частини розробленого додатку.

Забезпечити підтримку мультимедійних ресурсів з субтитрами та жестовою мовою для здобувачів з вадами слуху вдалося за допомогою сервісу УТОГ (Українське товариство глухих), яке надає послуги для людей із вадами слуху.

В розробленні клієнтської частини особливу увагу приділено створенню якісного функціонування та взаємодії між читачем та бібліотекарем. Для цього клієнтську частину *web*-додатку було розділено на дві частини: інтерфейс читача та *Admin*-панель бібліотекаря, який здійснює керування системою (рис. 9).

Клієнтська частина була реалізована за допомогою системи для побудови шаблонів та сторінок мовою розмітки *HTML* та стилізована за допомогою фреймворку *Bootstrap* та *CSS* та *JS* (*JavaScript*). *Bootstrap* – це клієнтський фреймворк з безкоштовним набором інструментів з відкритим кодом, призначений для створення вебсайтів та вебдодатків, який містить шаблони *CSS* та *HTML* для типографіки, форм, кнопок, навігації та інших компонентів інтерфейсу, а також додаткові розширення *JavaScript*. Він спрощує розробку динамічних вебсайтів і вебдодатків [13].

Для динамічної генерації *HTML*-шаблонів було використано *Thymeleaf* – сучасний серверний шаблонізатор для *Java*, який дозволяє зручно створювати *HTML*-сторінки, використовуючи природний синтаксис *HTML*. Він легко інтегрується з *Spring Boot*, що робить його ідеальним вибором для створення вебдодатків [14].

Система має 2 типи користувачів, а отже і 2 види доступу до даних *admin* та *simple user*. До категорії *simple user* також відносяться і викладачі, оскільки в них немає привілей порівняно зі здобувачами, це зроблено лише для ідентифікування користувачів, як викладачів.

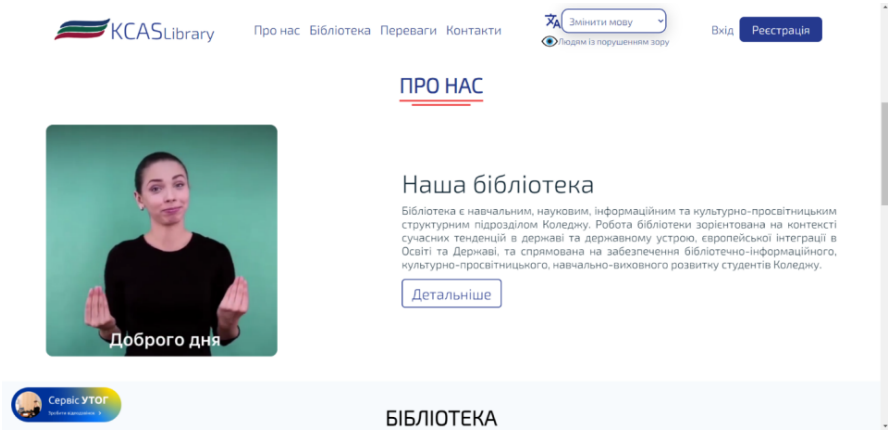


Рис. 8. Мультимедійний контент для людей з порушенням слуху

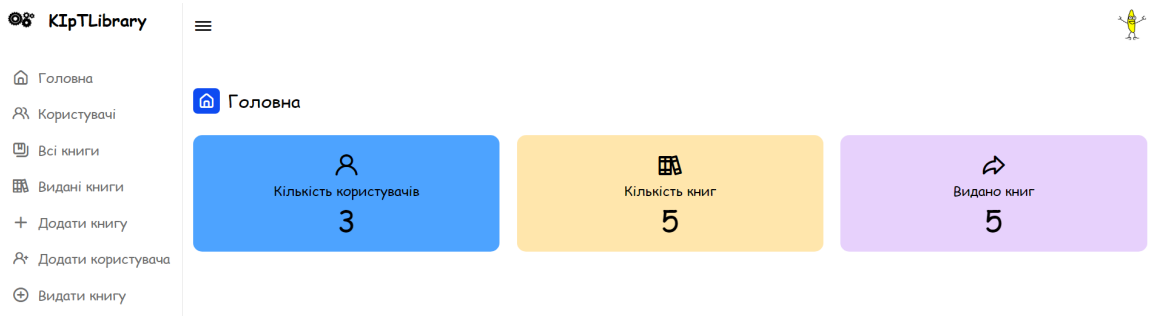


Рис. 9. *Admin*-панель бібліотекаря

В проєкті реалізована генерація *QR*-коду для кожної книги, за допомогою якого читач може отримати завчасно додану до своїх улюблених книжку (рис. 10).

Інтерфейс дає змогу адміністратору за допомогою розроблених форм (рис. 11) додавати в базу електронного каталогу нову літературу. Слід зауважити, що система має 2 основні категорії книг: паперові та електронні.

Адміністратор бібліотеки отримавши замовлення у *Admin*-панелі (рис. 9) може заздалегідь підготувати книгу для видачі. Для видачі книги без замовлення бібліотекарю потрібно обрати користувача та обрати книгу для видачі (рис. 12). Для спрощення пошуку було додано поле для пошуку користувачів та книги за ключовими словами, а також за *id QR*-коду.

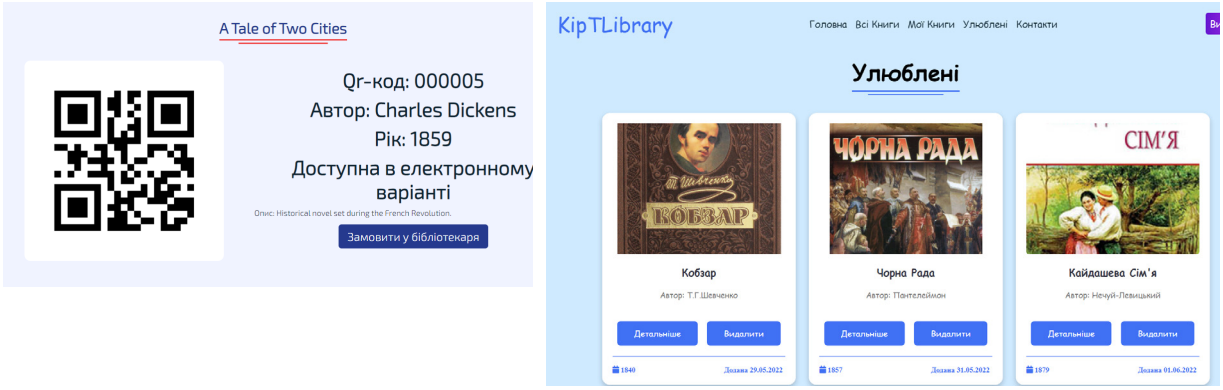


Рис. 10. Обрана книга для замовлення за допомогою QR-коду та улюблені книжки користувача

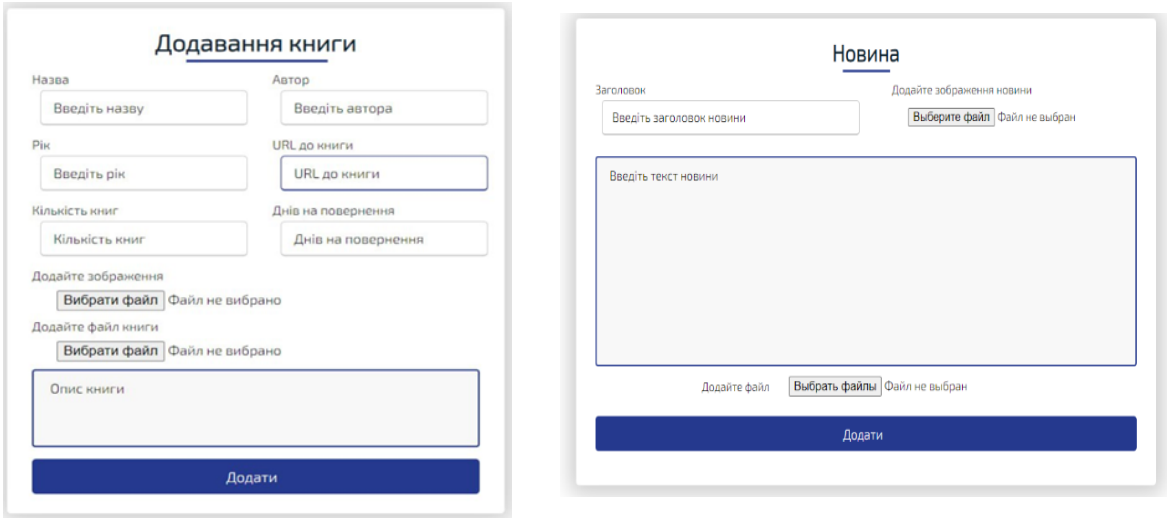


Рис. 11. Форми додавання в базу електронного каталогу нової літератури

Рис. 12. Вибір книги для користувача

Окрім адміністрування та управління основним функціоналом електронного каталогу *admin* має змогу створювати розважальний та науково-дослідницький контент розміщений на головній сторінці *web*-додатку.

Створення інтерфейсу для *Admin*-панелі забезпечує бібліотекарям інтуїтивно зрозумілий та функціональний інструмент для щоденної роботи і може повністю замінити паперові формуляри у бібліотеках навчальних закладів.

Для реалізації серверної частини нашого проєкту було використано мову програмування *Java* та фреймворк *Spring Boot* з використанням *Java* бібліотек [15]. В якості середовища розробки було проаналізовано та обрано інтегроване середовище розробки (*IDE*)-*IntelliJ IDEA*.

Специфікування предметної галузі проєкту наведено на рис. 13, де зображено розроблювані класи і зв'язки між ними.

Використання зв'язку ManyToMany дає змогу встановити зв'язок «багато-до-багатьох» між двома об'єктами, як у випадку класів «Книга» та «Категорії книг», оскільки одна книга може бути у декількох категоріях одночасно, а також одна категорія може мати декілька книг. Зв›язок ManyToOne вказує на те, що між двома об'єктами існує відношення «багато-до-одного», як у випадку класів «Книга» та «Автор», бо кожна книга може мати лише одного автора, але у одного автора може бути багато книг. Зв'язок OneToOne вказує на те, що об'єкт може бути пов›язаний лише з одним екземпляром іншого об'єкта, і навпаки.

Для управління базою даних було обрано *PostgreSQL*, що є потужною системою управління реляційними базами даних з відкритим вихідним кодом, яка підтримує стандарт *SQL*. Для *ORM* (*Object-Relational Mapping*) був обраний *Hibernate*, який забезпечує зручний спосіб взаємодії з базою даних через об'єкти *Java Hibernate*, *Spring Data JPA*, які є тісно пов›язаними технологіями, що використовуються для роботи з базами даних у *Java*-додатках [16].

Серверна частина в нашому проєкті містить 14 класів, 8 сервісів та 14 репозиторіїв. Кожен з класів має анотацію *@Entity*, яка є частиною пакету *javax.persistence* та використовується на рівні класу і позначає клас як персистентну сутність. Вона сигналізує провайдеру *JPA*, що клас слід розглядати як таблицю в базі даних.

Для коректної роботи *JPA* додана анотація *@Id*, яка використовується для позначення пер-

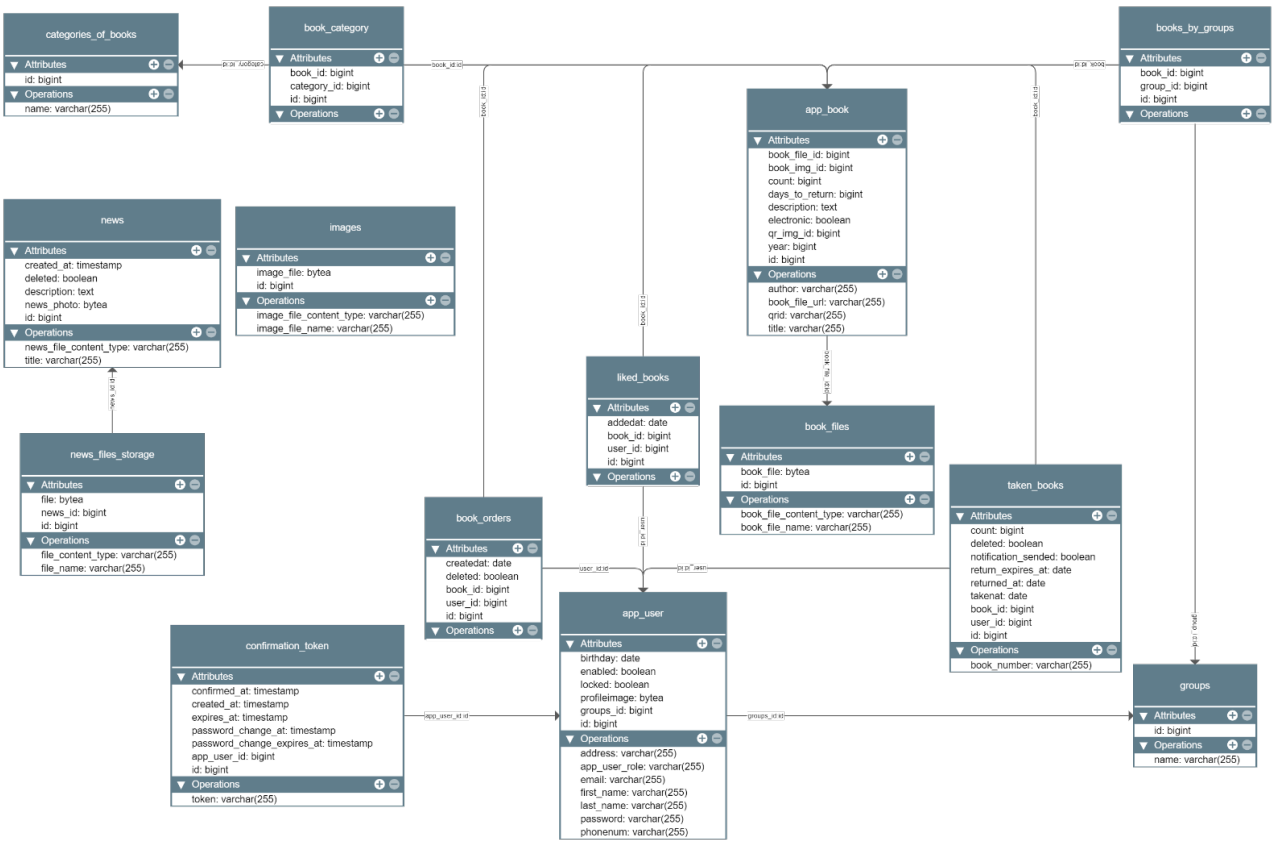


Рис. 13. Діаграма класів

винного ключа, який однозначно ідентифікує кожен екземпляр в базі даних та застосовується до поля або властивості в класі.

Точки доступу «*endpoint*», через які клієнти можуть взаємодіяти з сервером, допомагають серверу зрозуміти куди клієнт відправив запит, а контролер відповідає за обробку цього запиту та за відправку відповіді. Над кожним методом, що обробляє запит буде анотація з методом запиту, наприклад *@GetMapping* для *Get*-запиту, який може повертати різні типи даних, включаючи рядки, об›єкти, тощо та *@PostMappring*, який зазвичай приймає дані (можуть бути в різних форматах) з тіла запиту.

Наведемо приклад, де використовується «endpoint» за шляхом «/» для розроблення головної сторінки. В контролері створюємо *Get*-метод, та передаємо в нього дані, які міститимуться на головній сторінці.

```
@GetMapping(«/»)
public String home(Model model) {
Page<NewsRepository.NewsNoFile> news =
newsRepository.findAllByDeletedIsFalse(PageRequest.of
(0, 5, Sort.Direction.DESC,»createdAt»));
model.addAttribute(«news», news);
return «home»;
}
```

В результаті отримуємо *html* сторінку (рис. 14) з атрибутом новин, який оброблює *Thymeleaf*.

За аналогією з головною сторінкою будуємо метод у контролері для сторінки реєстрації:

```
@GetMapping(«/register»)
public String showRegistrationForm(Model model) {
List<Groups> groups = groupsRepository.findAll();
model.addAttribute(«groups», groups);
model.addAttribute(«user», new AppUser());
return «signup_form»;
}
```

В результаті отримуємо *html* сторінку (рис. 15) з атрибутами груп та атрибутом нового користувача, які також будуть оброблені за допомогою *Thymeleaf*.

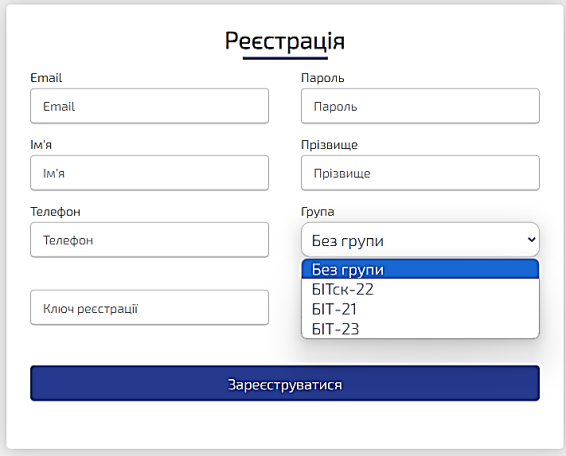


Рис. 15. Реєстрація нового користувача

Після заповнення форми система відправляє ці дані на «*endpoint*» «*/process_register*», сервер отримавши запит обробляє його, перевіряє дані та додає їх до бази даних, після чого надсилає *email* для підтвердження пошти.

Підтвердивши пошту користувач має змогу зайти в свій електронний кабінет який знаходиться на «*endpoint*» «*/editprofile*» (рис. 16).

За аналогією будуються також інші контролери, наведені на рис. 17. Контролери, які починаються з слова */admin* містять методи з «*endpoints*», що допомагає у розробці безпеки.

Функціонал доступний користувачам: перегляд всіх наявних книг в бібліотеці та доступ до електронних книг, можливість вибирати книги за різними фільтрами, позначати «улюблені» та бронювати бажані. Крім цього, є можливість брати

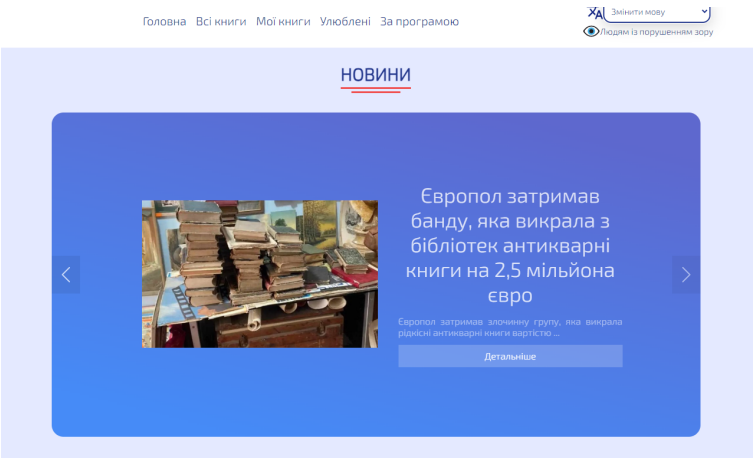


Рис. 14. Сторінка новин

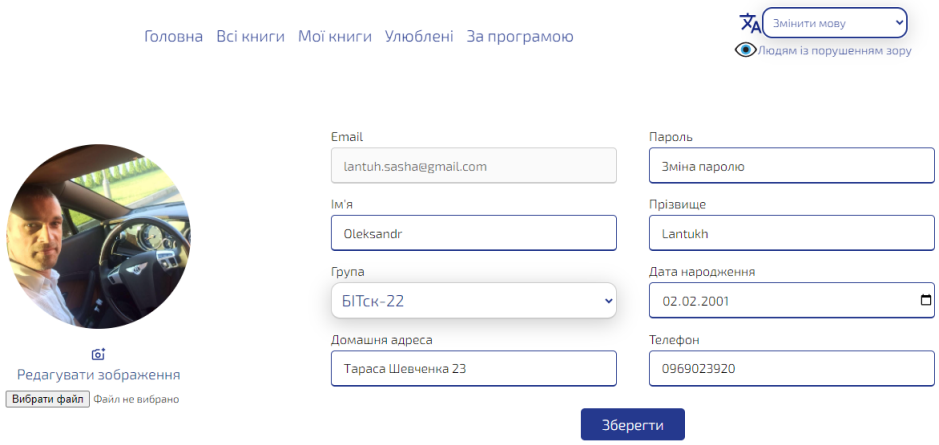


Рис. 16. Профіль користувача (електронний кабінет)

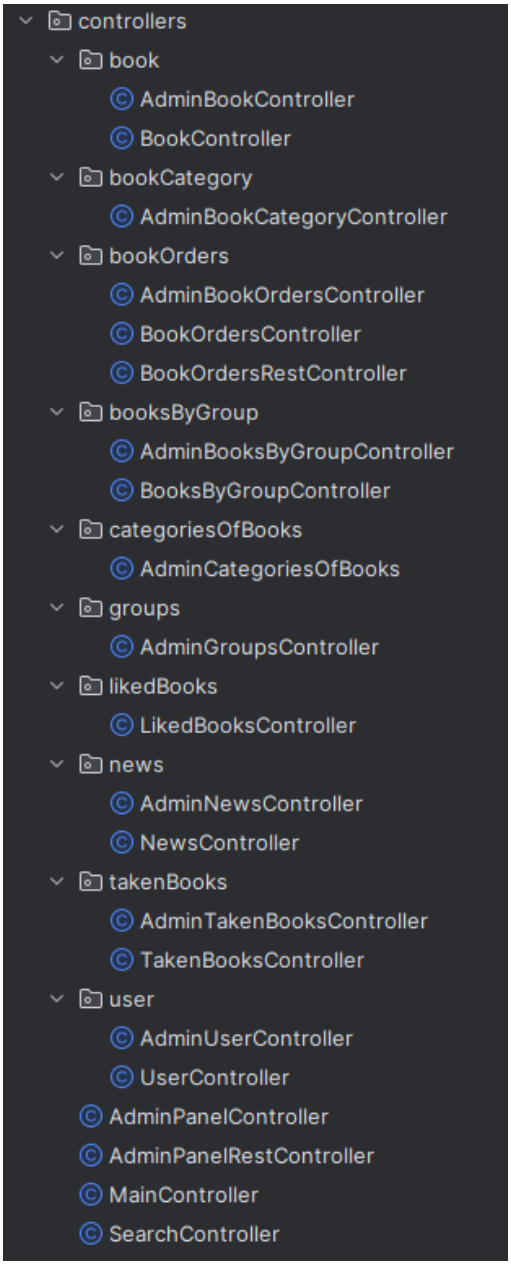


Рис. 17. Контролери проєкту

участь в онлайн марафонах та заходах, які організовує бібліотека разом з партнерами, виступах тренерів, та гостей. *Web*-додаток розміщений на хмаровому сервісі *Microsoft Azure*, що дозволяє забезпечити якісну роботу сайту в будь-якому місці та в будь-який час.

Висновки. Розроблений програмний продукт є кроком у напрямку модернізації освітнього процесу в Україні. Ця система забезпечує зручний та доступний інструмент для отримання навчальних матеріалів як здобувачам, так і викладачам.

Розроблений програмний продукт дозволяє збільшити доступність навчальних ресурсів, підвищити якість освіти через різноманітність форматів матеріалів та стимулювати самостійне навчання. План розроблення програмного продукту, який включав етапи проєктування, розробки, тестування та впровадження, відображає системний та професійний підхід до створення інтерактивного додатку для взаємодії з бібліотекою закладу освіти за всіма сучасними стандартами. Заплановані мікросервіси всередині проєкту дозволяють створити різноманітне та комфортне програмне забезпечення для здобувачів та викладачів, що дозволить навчатися їм більш продуктивно та комфортно. Аналітична робота дозволила визначити функціональні вимоги та особливості, необхідні для створення web-додатку для онлайн бібліотеки закладів освіти. Вибір правильних технологій та інструментів дозволить легко підтримувати та розширювати функціонал в майбутньому, вдосконалити архітектуру до багаторівневої, сприяє ефективній розробці, а наведені етапи розробки дозволили досягти успішного завершення проєкту. Програмний продукт може бути використаний для автоматизації роботи бібліотеки будь-якого закладу освіти або іншої організації.

Розроблений проєкт пройшов апробацію та використовується в Київському фаховому коледжу прикладних наук. Використання даної системи зменшує часові витрати користувача та зменшує навантаження на бібліотекаря, забезпечує підтримку мультимедійних ресурсів з субтитрами та жестовою мовою для здобувачів з вадами слуху, ліквідує паперову документацію та робить процедуру отримання книг легкою та прозорою.

Список літератури:

1. Моргулець О.Б., Астістова Т.І. Сучасні комп'ютерні технології в організації освітньої діяльності навчального закладу. *Вісник КНУТД*. Київ, 2016. № 5 (101), спецвипуск. С. 351–359.
2. Астістова Т.І. Розробка системи керування розкладом занять у вищому навчальному закладі. *Вісник КНУТД*. Київ, 2016. № 4 (100). С. 13–18.
3. Smorzhevskyi N.V, Astistova T.I. Development of software for interactive interaction with the library. *Інноватика в освіті, науці та бізнесі: виклики та можливості* : збірник наукових праць 1 Всеукраїнської конференції, м. Київ, 17 листопада 2020 р. Київ, 2020. С. 111–119.
4. Наукова бібліотека ім. М. Максимовича. URL: https://library.knu.ua/
5. Наукова бібліотека ЛНУ імені Івана Франка. URL: https://www.lnulibrary.lviv.ua/
6. Наукова бібліотека Національного університету «Києво-Могилянська академія». URL: https://library.ukma.edu.ua/
7. Наукова бібліотека Національного університету біоресурсів і природокористування України. URL: https://nubip.edu.ua/structure/library
8. Наукова бібліотека Київського національного університету технологій та дизайну. URL: http://lib.knutd.edu.ua/
9. Веб-дизайн, UI/UX, ефективні інтерфейси. URL: https://goldwebsolutions.com/uk/uiux-uk/
10. Дроменко В.Б., Лісовець С.М. Особливості інженерії програмного забезпечення при розробленні систем з елементами штучного інтелекту. *Вчені записки Таврійського Національного Університету імені В.І. Вернадського. Серія: Технічні науки.* 2024. Т. 35 (74), № 4. С. 74–82. DOI: 10.32782/2663-5941/2024.4/12
11. Клієнт-серверна архітектура. URL: https://surl.li/cppxry
12. Frontend. Інтерфейс для взаємодії між користувачем і backend. URL: https://goldwebsolutions.com/uk/blog/frontend/
13. Bootstrap. URL: https://surli.cc/uivmrr
14. Thymeleaf: Документація Thymeleaf. URL: https://www.thymeleaf.org/doc/tutorials/3.0/usingthymeleaf.html
15. Лантух О.М., Астістова Т.І Розробка серверної частини для веб-додатку *Мехатронні системи: інновації та інжиніринг* : зб. тез доп. VII Міжнародної науково-практичної конференції, м. Київ, 23 листопада 2023 р. Київ, 2023. С. 165.
16. Java: Офіційна документація Oracle. URL: https://docs.oracle.com/javase/tutorial/

Astistova T.I., Dromenko V.B., Kalashnyk V.Yu. DEVELOPMENT OF SOFTWARE FOR INFORMATION RESOURCES OF EDUCATIONAL INSTITUTIONS LIBRARY

The article presents an analysis of the use of digital technologies in the information systems of scientific libraries of leading Ukrainian universities. On the basis of the analysis, the functional requirements and tasks for developing an interactive application for interaction with the library of an educational institution according to modern standards are defined.

The architecture of the developed web application contains three main components: a client component, a server component, and a database server, which is implemented using the MVC (Model-View-Controller) pattern.

The client side was implemented using the integrated software development environment Intellij Idea and a system for building templates and pages in the HTML markup language and stylised using the Bootstrap framework and CSS and JS. For the dynamic generation of HTML templates, we used a modern server-side templating engine for Java, Thymeleaf. To implement the server side, we used the Java programming language and the SpringBoot framework with Java libraries. PostgreSQL was chosen for database management. For ORM (Object-Relational Mapping), Hibernate was chosen, which provides a convenient way to interact with the database through Java Hibernate objects, Spring Data JPA.

The developed web application is a complete software product that has been tested and is used at the Kyiv Professional College of Applied Sciences, and provides support for multimedia resources with subtitles and sign language for students with hearing impairments. The project implemented the generation of a QR code for each book, with which the reader can receive a book added to their favorites in advance. The implemented system reduces user time and librarian workload, eliminates paperwork, and makes the procedure for obtaining books easy and transparent.

Key words: *architecture, services, methods, classes, controllers, technology stack, MVC, PostgreSQL, integrated environment, Integrated IDE, java Spring Boot framework, HTML, CSS.*

УДК 004.622: 517.927
DOI https://doi.org/10.32782/2663-5941/2025.3.2/08

Бараннік В.В.
Харківський національний університет імені В.Н. Каразіна

Берчанов А.А.
Харківський національний університет імені В.Н. Каразіна

МЕТОД ОБРОБКИ ІНФРАЧЕРВОНИХ ЗОБРАЖЕНЬ У СПЕКТРАЛЬНОМУ ДОМЕНІ

Стаття присвячена підвищенню ефективності обробки та кодування інфрачервоних (ІЧ) зображень. У статті розкрито зростаючу потребу у швидкому і достовірному обміні ІЧ зображеннями, зокрема у оборонних та безпекових галузях, що зумовлює актуальність дослідження. Сучасні методи (JPEG 2000, H.265) усувають статистичну, просторову та психовізуальну надмірність, завдяки чому зменшують бітовий об'єм видової інформації. Проте дані кодеки не враховують специфіки ІЧ зображень, та як наслідок вносять спотворення, що змінюють температурну семантику кадрів і знижують ефективність обробки та інтелектуального аналізу. Звідси, науково-прикладна задача дослідження – підвищити ефективність обробки й кодування ІЧ зображень. З'ясовано, що специфіка ІЧ зображень полягає у тому, що теплові сигнатури об'єктів мають підвищену контрастність. Відповідно області із підвищеною контрастністю мають більші високочастотні показники у спектральному домені. Що пропонується врахувати при кодуванні ІЧ зображень. Мета дослідження – розробити метод обробки інфрачервоних зображень у спектральному домені. Для локалізації областей інформативності пропонується застосувати ієрархічну сегментацію кадру. ІЧ зображення можна розділити на сегменти 8×8. Кожен сегмент у свою чергу розділити на мінісегменти 4×4. До кожного мінісегменту можна застосувати перетворюються у різницевий простір. Це дозволить зменшити динамічним діапазоном значень та при цьому зберегти їх семантичну структуру. Для перетворення у спектральний домен пропонується використовувати вейвлет-перетворення на базисі Хаара. На основі високочастотних коефіцієнтів обчислюється нормована метрика інформативності. Завдяки ній можна класифікувати сегменти на інформаційно важливі та другорядні. Експерименти показали підвищення коефіцієнта стиснення на 2,63 порівняно зі спектрально-груповим кодуванням. Отримані результати доводять ефективність розробленого методу й можуть бути використані для вдосконалення стандартів стиснення ІЧ даних і створення інтелектуальних сервісів обробки зображень.

Ключові слова: *інфрачервоні зображення, обробка інфрачервоних зображень, стиснення інфрачервоних зображень, зображення із широким динамічним діапазон, бітовий простір, бітові площини, біфуркація бітових площин.*

Постановка проблеми. Системи спостереження із використанням інфрачервоних (ІЧ) джерел інформації еволюціонують та набувають ширшого застосування. Вони мають критичну роль у оборонній та безпековій галузях, зокрема із застосуванням у безпілотних системах. На сьогоднішній день, потокові видові джерела інформації все частіше обробляють із застосуванням інтелектуальних сервісів. Це дозволяє отримати мета відомості: ідентифікація об'єктів та їх стану, прогнозування динамічного стану об'єктів інтересу. На основі таких мета відомостей працюють системи: автоматичного сповіщення сповіщення щодо кризових ситуацій, автономного керування роботизованими комплексами. ІЧ джерела інформації дозволяють покращити роботу таких систем, створюючи можливості для: виявлення замаскованих об'єктів, виявлення об'єктів в умовах обмеженої видимості (нічний час, погані погодні умови). Для підвищення ефективності застосування інфрачервоних камер, виробники покращують їх характеристики: розмір матриці, кількість кадрів, чутливість до теплового випромінення. Це у свою чергу збільшує об'єм інформації, що створює додаткове навантаження на канали передачі даних. У наслідок цього збільшуються затримки при доставці даних. Для вирішення цієї проблеми застосовують методи кодування видової інформації, що дозволяють зменшити бітовий об'єм даних.

Серед таких методів популярні JPEG, WebP, HEIF або потокові H.264, H.265, AV1. Ці методи

дозволяють зменшити статистичну, просторову та психовізуальну надмірність. В результаті дані передаються у закодованому вигляді та займають на багато менше бітового об'єму. Однак недолік зазначених методів полягає у тому, що вони вносять спотворення у семантику видової інформації. В результаті чого, декодованні кадри можуть бути інтерпретовані некоректно людиною або невірно опрацьовані інтелектуальним сервісом. Отже, виникає протиріччя: з одної сторони важлива якість та цілісність ІЧ зображень для забезпечення повноти інформації, з іншої сторони важлива швидкість доставки даних.

Аналіз останніх досліджень і публікацій. Для підвищення ефективності обробки ІЧ зображень необхідно враховувати їх специфіку, яка полягає в тому, що фіксується теплове випромінювання об'єктів, а не відбите світло, як у видимому діапазоні [1, с. 12; 2, с. 16; 3, с. 19]. Це призводить до низки особливостей, які впливають на їхню структуру та характеристики: представлення інформації: ІЧ зображення відображають розподіл температури поверхні, а не візуальні деталі. Це може призводити до зображень з менш вираженими текстурами, розмитими межами об'єктів та іншими шумами, відмінними від шуму на оптичних зображеннях [4, с. 7; 5, с. 19]; динамічний діапазон та контраст: Розподіл інтенсивностей в ІЧ зображеннях часто відрізняється від оптичних. Контраст може бути низьким або зосередженим у певних діапазонах температур, що ускладнює застосування стандартних методів покращення якості [6, с. 21; 7, с. 11]; шуми: ІЧ сенсори мають свої специфічні джерела шуму (наприклад, температурний дрейф, нерівномірність відгуку пікселів), які потребують спеціалізованих методів фільтрації [8, с. 11; 9, с. 13]; просторова роздільна здатність: ІЧ зображення часто мають нижчу просторову роздільну здатність порівняно з оптичними зображеннями, що впливає на деталізацію та можливість розпізнавання дрібних об'єктів [10, с. 18; 11, с. 5].

Крім цього, важливо забезпечити низьку обчислювальну складність при кодуванні ІЧ зображень. Адже висока складність обробки призводить до таких недоліків [12, с. 21]:

– підвищення часу на обробку, що призводить до затримок на доведення інформації;

– збільшене використання енергії, що критично для автономних комплексів.

Для забезпечення низької обчислювальної складності та врахування специфіки ІЧ зображень, пропонується враховувати семантичну складову кадрів у спектральному домені.

Постановка завдання. На підставі визначеного протиріччя постає актуальна науково-прикладна задача: підвищення ефективності обробки та кодування інфрачервоних зображень. Отже, мета статті полягає у розробці методу обробки інфрачервоних зображень у спектральному домені для підвищення ефективності їх обробки та кодування.

Виклад основного матеріалу. Необхідно враховувати структурні особливості інфрачервоних зображень. Так як зображення, що містять об'єкти інтересу є гетерогенним, необхідно забезпечити локалізацію більш інформативних сегментів зображення, та менш інформативних. Для цього розділимо зображення A, що необхідно закодувати, на сегменти $A^{(k,l)}$, де (k,l) є координатами сегменту.

Для кожного сегменту $A^{(k,l)}$ необхідно оцінити метрику інформативності $A_M^{(k,l)}$:

$$A_M^{(k,l)} = Metr\left(A^{(k,l)}\right),$$

де $Metr\left(A^{(k,l)}\right)$ – функція визначення метрики інформативності.

Спираючись на значення $A_M^{(k,l)}$ можна диференціювати сегменти за рівнем інформативності. Для цього використаємо порогове значення T, що може змінюватись у залежності від характеристик та динамічної насиченості інформаційного каналу. Отже, якщо $A_M^{(k,l)} > T$, то сегмент $A^{(k,l)}$ вважається інформативним, інакше малоінформативним. У залежності від рівня інформативності сегменту, при подальшій обробці можуть використовуватись різні методи кодування чи різні конфігурації одного методу. Назвемо це семантично-орієнтованою обробкою та позначимо як $SAP\left(A^{(k,l)}\right)$. Описаний метод визначення інформативності сегменту $A^{(k,l)}$ зображено на рисунку 1.

При визначенні метрики інформативності $A_M^{(k,l)}$ для покращення локалізованості пропонується розділити сегмент $A^{(k,l)}$ на чотири мінісегменти $A_{r,t}^{(k,l)}$, де: r,t – координати мінісегменту у середині сегменту. Тобто:

$$A^{(k,l)} \in \{A_{1,1}^{(k,l)}, A_{1,2}^{(k,l)}, A_{2,1}^{(k,l)}, A_{2,2}^{(k,l)}\}.$$

Перетворимо мінісегменти $A_{r,t}^{(k,l)}$ у різницевий простір. Це дозволить зменшити динамічний діапазон значень, та при цьому зберегти семантичну структуру мінісегментів.

$$A'^{(k,l)}_{r,t} = subtr\left(A_{r,t}^{(k,l)}\right),$$

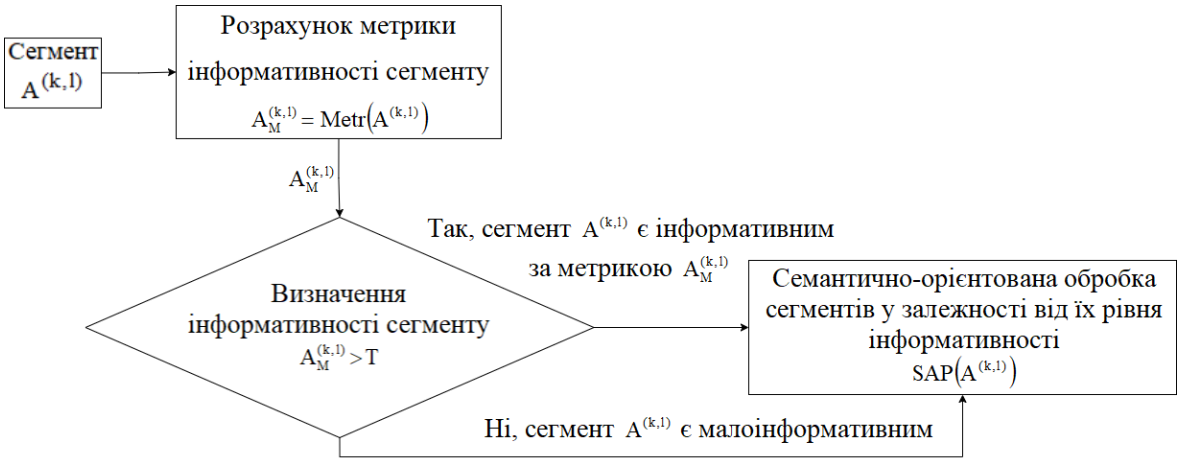


Рис. 1. Метод визначення інформативності сегменту $A^{(k,l)}$

де:

– $subtr\left(A_{r,t}^{(k,l)}\right)$ – функція перетворення <<Eqn023.wmf>> у різницевий простір;

– $A'^{(k,l)}_{r,t}$ – утворений різницевий різницевий мінісегмент.

В отриманому мінісегменті $A'^{(k,l)}_{r,t}$ необхідно визначити наявність теплових сигнатур. Оскільки специфіка ІЧ зображень полягає в тому, що теплові сигнатури мають підвищений контраст, пропонується трансформувати дані у спектральний домен. Адже висококонтрастні області мають підвищені високочастотні характеристики. Тож у спектральному домені можна їх виявити.

Для трансформації у спектральний домен пропонується використати вейвлет перетворення на базисі Хаара. Цей метод відомий тим, що має низьку обчислювальну складність та дозволяє отримати високочастотні значення окремо від низькочастотних. Позначимо таку трансформацію як:

$$\left\{ L_{r,t}^{(k,l)}(I), H_{r,t}^{(k,l)} \right\} = WH\left(A'^{(k,l)}_{r,t} \right),$$

де:

– $WH\left(A'^{(k,l)}_{r,t}\right)$ – функція вейвлет перетворення;

– $L_{r,t}^{(k,l)}(I)$ – кінцеве низькочастотне значення;

– $H_{r,t}^{(k,l)}$ – група високочастотних значень.

При розрахунку метрики інформативності необхідно врахувати високочастотні значення із кожного мінісегменту $A'^{(k,l)}_{r,t}$, тому визначимо їх суму. Оскільки високочастотні значення можуть бути від'ємними, використаємо їх абсолютні значення:

$$HSum_{r,t}^{(k,l)} = \sum \left| H \right|, H \in H_{r,t}^{(k,l)},$$

де:

– $\left| H \right|$ – абсолютне високочастотне значення;

– $HSum_{r,t}^{(k,l)}$ – сума абсолютних високочастотних значень для мінісегменту $A'^{(k,l)}_{r,t}$.

У свою чергу розрахуємо суму високочастотних значень $HSum^{(k,l)}$ для сегменту $A^{(k,l)}$:

$$HSum^{(k,l)} = HSum_{1,1}^{(k,l)} + HSum_{1,2}^{(k,l)} + HSum_{2,1}^{(k,l)} + HSum_{2,2}^{(k,l)}.$$

Значення $HSum^{(k,l)}$ необхідно привести до загального вигляду, діапазону від 0 до 1. Для цього виконаємо нормалізацію $Norm\left(HSum^{(k,l)}\right)$. Результатом нормалізації є метрика інформативності $A_M^{(k,l)}$:

$$A_M^{(k,l)} = Metr\left(A^{(k,l)}\right) = Norm\left(HSum^{(k,l)}\right).$$

На рисунку 2 зображена схема розрахунку метрики інформативності $A_M^{(k,l)}$ для сегменту $A^{(k,l)}$.

Експериментальна оцінка методу обробки інфрачервоних зображень у спектральному домені

Експериментальна оцінка розробленого методу проводилась на ІЧ зображеннях із глибиною пікселів 16 біт. Набір ІЧ зображень було закодовано методом спектрально-групового кодування (СГК). Для порівняння ці ж зображення було закодовано розробленим методом обробки (SAP). Порівняльні показники коефіцієнтів стиснення зображено на рисунку 3. Конфігурація методів кодування була встановлена таким чином, щоб показники PSNR мали однакові значення, вони зображені на рисунку. Можна побачити, що коефіцієнти стиснення більші для методу SAP. У середньому вони збільшились на 2.63.

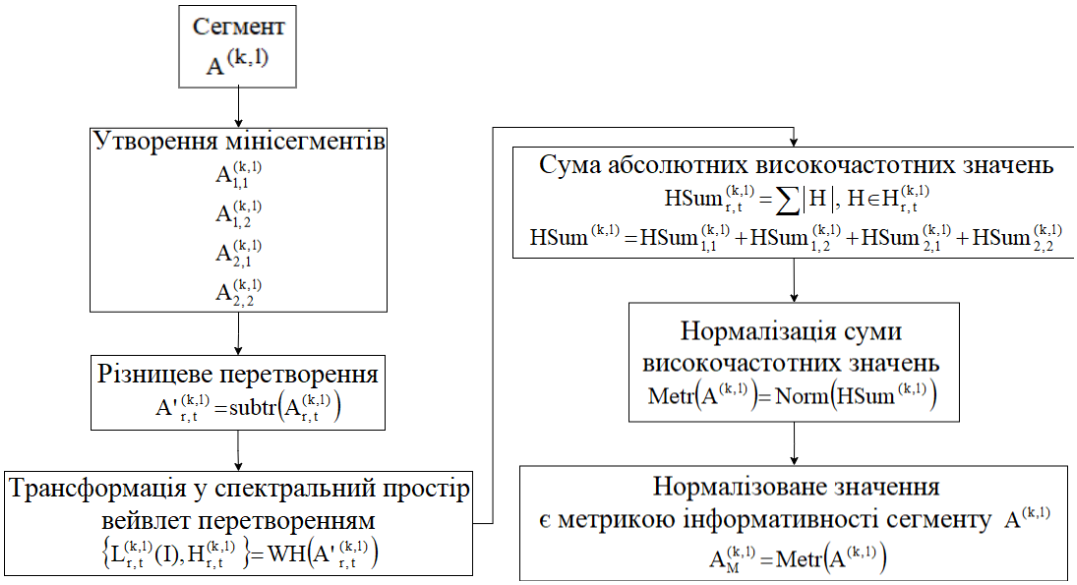


Рис. 2. Метод розрахунку метрики інформативності $A^{(k,l)}_M$

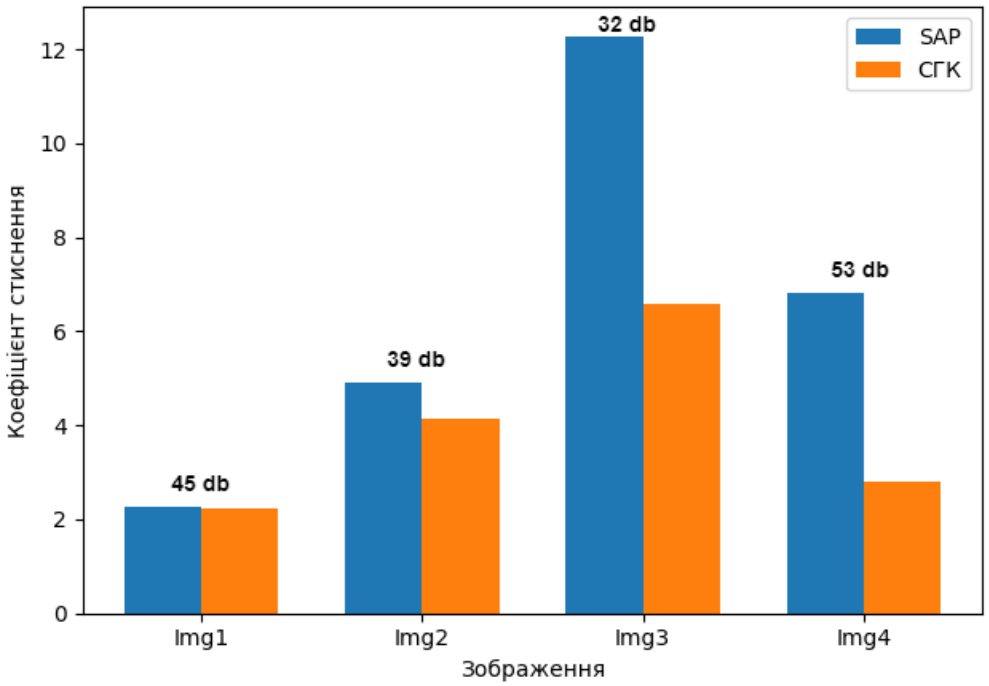


Рис. 3. Коефіцієнти стиснення після кодування методами СГК та SAP

Висновки. 1. У статті обгрунтована актуальність розвитку систем спостереження із використанням ІЧ джерел інформації. Окреслено існуюче протиріччя між збереженням необхідної якості кадрів та вимогою зменшити затримки у їх передачі. У зв'язку з цим зазначена науково-прикладна задача: підвищити ефективність обробки та кодування інфрачервоних зображень.

2. Зазначені недоліки існуючих методів кодування видової інформації. Як наслідок, обгрунтована необхідність врахування семантичних особливостей ІЧ кадрів при їх кодуванні та обробці.

3. Розроблено метод обробки інфрачервоних зображень у спектральному домені, який базується на:

1) визначенні метрики інформативності для сегментів кадру;

2) трансформації сегментів у різницевий простір та спектральний домен;

3) врахуванні високочастотної закономірності у семантиці ІЧ кадрів.

4. Була проведена експериментальна оцінка, яка показала підвищення коефіцієнту стиснення у середньому на 2.63 при застосуванні розробленого методу.

Список літератури:

1. Barannik V., Shulgin S., Babenko Yu., Onyshchenko R., Revva K., Belikova T., Ignatyev O. Technology of Sliding Coding of Uneven Diagonal Sequences in Two-Dimensional Spectral Space of Transformants. *Visnyk NTUU KPI Seriia – Radiotekhnika Radioaparatobuduvannia.* 2023. № 94. P. 13–23. DOI: 10.20535/RADAP.2023.94.13-23.

2. Putra A., Supriadi S., Wibawa A., Pranolo A., Gaffar A. Modification of a gray-level dynamic range based on a number of binary bit representation for image compression. *Science in Information Technology Letters.* 2020. Vol. 1, No. 1. P. 9–16. DOI: 10.31763/sitech.v1i1.17.

3. Omar H. M., Morsli M., Yaichi S. Image Compression using Principal Component Analysis. *Proceedings of the 2020 2nd International Conference on Mathematics and Information Technology (ICMIT).* 2020. P. 226–231. DOI: 10.1109/ICMIT47780.2020.9047014.

4. Barannik V., Sidchenko S., Barannik V., Chornomaz I., Hurzhii P., Grygorian M. Saving Elements Methods for Service Components of Images Cryptocompression Codograms. *Visnyk NTUU KPI Seriia – Radiotekhnika Radioaparatobuduvannia.* 2023. № 92. P. 28–40. DOI: 10.20535/RADAP.2023.92.28-40.

5. Odarchenko R., Gnatyuk S., Gnatyuk V., Abakumova A. Security key indicators assessment for modern cellular networks. *System Analysis & Intelligent Computing (SAIC): proceedings of the IEEE First International Conference.* 2018. P. 1–7. DOI: https://doi.org/10.1109/SAIC.2018.8516889.

6. Shah T. J., Banday M. T. A Review of Contemporary Image Compression Techniques and Standards. *Examining Fractal Image Processing and Analysis.* 2020. P. 121–157. DOI: 10.4018/978-1-7998-0066-8.ch006.

7. Козловський В., Толстікова О., Савченко А., Клобукова Л. Критерії вибору спектрально-ефективних сигналів у бездротових інформаційних мережах. *Наукоємні технології.* 2022. № 4 (56). С. 273–286. DOI: 10.18372/2310-5461.56.17125.

8. Krasnorutsky A., Barannik V., Shulgin S., Kolesnyk V., Kharchenko N., Revva K. Integration of Video Image Decryption Coding into a Remote Video Information Service. *2023 IEEE 5th International Conference on Advanced Information and Communication Technologies (AICT)*. 2023. P. 1–6. DOI: 10.1109/AICT61584.2023.10452679.

9. Garg G., Kumar R. Analysis of image types, compression techniques and performance assessment metrics: A review. *Journal of Information and Optimization Sciences.* 2022. Vol. 43, № 3. P. 429–436. DOI: 10.1080/02522667.2022.2037282.

10. Uthayakumar J., Elhoseny M., Shankar K. Highly Reliable and Low-Complexity Image Compression Scheme Using Neighborhood Correlation Sequence Algorithm in WSN. *IEEE Transactions on Reliability.* 2020. Vol. 69, № 4. P. 1398–1423. DOI: 10.1109/TR.2020.2972567.

11. Resdiansyah R., Adrian J. D., Wijaya A. H., Hakim L., Tannady H. Comparing Freeman Chain Code 4 Adjacency Algorithm and LZMA Algorithm in Binary Image Compression. *Journal of Physics: Conference Series.* 2021. Vol. 1783. Article ID: 012045. DOI: 10.1088/1742-6596/1783/1/012045.

12. Одарченко Р., Рябенко М., Іванова М., Аль-Мудхафар Акіл Абдулхусейн М. Метод аналізу взаємодії параметрів QOE та QOS на основі алгоритмів керування машинами. *Наукоємні технології.* 2022. № 4 (56). С. 305–316. DOI: 10.18372/2310-5461.56.17130.

Barannik V.V., Berchanov A.A. METHOD OF PROCESSING INFRARED IMAGES IN THE SPECTRAL DOMAIN

The article is devoted to increasing the efficiency of processing and encoding infrared (IR) images. The article reveals the growing need for fast and reliable exchange of IR images, particular in the defense and security industries, which determines the relevance of the study. Modern methods (JPEG2000, H.265) eliminate statistical, spatial and psychovisual redundancy, thereby reducing the bit volume of species information. However, these codecs do not take into account the specifics of IR images, and as a result introduce distortions that change the temperature semantics of frames and reduce the efficiency of processing and intellectual analysis. Hence, the scientific and applied task of the study is to increase the efficiency of processing and encoding IR images. It was found that the specificity of IR images is that the thermal signatures of objects have increased contrast. Accordingly, areas with increased contrast have higher high-frequency indicators in the spectral domain. What is proposed to be taken into account when encoding IR images. The purpose of the study is to develop a method for processing infrared images in the spectral domain. To localize the areas of informativeness, it is proposed to apply hierarchical frame segmentation. The IR image can be divided into 8×8 segments. Each segment, in turn, can be divided into 4×4 minisegments. Each minisegment can be

transformed into a difference space. This will allow reducing the dynamic range of values while preserving their semantic structure. For transformation into the spectral domain, it is proposed to use a wavelet transform based on the Haar basis. Based on the high-frequency coefficients, a normalized informativeness metric is calculated. Thanks to it, it is possible to classify segments into informationally important and secondary. Experiments have shown an increase in the compression ratio by 2.63 compared to spectral group coding. The results obtained prove the effectiveness of the developed method and can be used to improve IR data compression standards and create intelligent image processing services.

Key words: *infrared images, infrared image processing, infrared image compression, wide dynamic range images, bit space, bit planes, bit plane bifurcation.*

УДК 004.622:517.927
DOI https://doi.org/10.32782/2663-5941/2025.3.2/09

Бараннік Д.В.
Харківський національний університет радіоелектроніки

Сідченко Є.С.
Харківський національний університет імені В.Н. Каразіна

Лиходєєв О.С.
Харківський національний університет Повітряних Сил

Помазуєв В.В.
Харківський національний університет Повітряних Сил

Федоровський О.В.
Харківський національний університет Повітряних Сил

МЕТОД ЗАСТОСУВАННЯ АПЕРТУР ДО АСИМЕТРИЧНИХ СИСТЕМ КОДУВАННЯ ВІДЕОДАННИХ

У статті розглянуто актуальні проблеми стиснення відеозображень для інформаційних систем, що функціонують в умовах обмежених або нестабільних каналів зв'язку. Проаналізовано недоліки класичних і сучасних форматів кодування, зокрема їхню низьку стійкість до помилок та неадаптивність до структурної складності зображень, зокрема показано на практичному прикладі слабкість класичної сегментації 8x8, яка не враховує особливості зображення. Обґрунтовано доцільність використання асиметричних систем числення (ANS), зокрема варіанта rANS, як ефективного інструменту ентропійного кодування, що поєднує високу швидкодію з наближенням до теоретичної межі стиснення. Разглянуто недоліки ANS, а саме завадостійкість та залежність ефективності методів від вхідних дани та алфавітів. Для розв'язання цієї проблеми, далі у статті розроблено метод сегментації зображень на основі виявлення локальних областей когерентності, що характеризуються подібними структурно-візуальними властивостям. Запропоновано механізм апертурного поділу з урахуванням амплітудних варіацій пікселів та адаптивної апроксимації значень усередині апертур. Показано, що такий підхід дозволяє значно підвищити завадостійкість до помилок при передачі даних, локалізуючи їхній вплив у межах окремих фрагментів зображення. Проведено порівняльне експериментальне дослідження ефективності запропонованого методу у поєднанні з rANS-кодуванням. Отримані результати демонструють зниження обсягу закодованих даних без значущих втрат візуальної якості та суттєве зменшення масштабів спотворень при виникненні помилок у передаваному потоці. Завдяки врахуванню особливостей зображення, при внесенні канальних помилок, об'єкти на зображенні залишаються гарно видимими для людського ока. Надані експерементальні результати дослідження на прикладах, які підтверджують ефективність та запропонованого підходу у порівнянні до классичних методів сегментації, що підтвержу перспективність методів для використання в реальному часі в системах із підвищеними вимогами до надійності відеоінформації.

Ключові слова: *апертура, області когерентності, асиметричні системи числення, стиснення з втратами, шум, завадостійкість.*

Постановка проблеми. Сучасні інформаційні системи, що працюють з великомасштабними зображеннями та відеопотоками, стикаються з викликами, пов'язаними із забезпеченням високої якості візуального сприйняття в умовах обмежених ресурсів – зокрема, пропускної здатності каналів зв'язку, обчислювальної потужності та енергоспоживання [1]. Ці виклики особливо актуальні для інтелектуальних сервісів реального часу, таких як автоматизоване відеоспостереження, смарт-аналітика міського середовища, автономне керування транспортом, оперативне реагування у кризових ситуаціях. У подібних середовищах, де часто мають місце нестабільні або обмежені канали зв'язку (наприклад, у бездротових сенсорних мережах або мобільних пристроях), критичною стає потреба в ефективному стисненні зображень і відео [2]. Ключові вимоги до таких методів

включають зменшення обсягу переданих даних, зниження затримок і навантаження на мережу, швидке декодування та мінімізацію енергоспоживання. Водночас важливо зберігати не лише візуальну якість, але й семантичну інформативність даних для подальшої інтелектуальної обробки. Однак традиційні формати стиснення (JPEG, H.264/AVC, HEVC, AV1) демонструють обмеження в адаптації до складної структурної організації зображень [3, 4]. Особливо проблематичною є ситуація, коли через втрату навіть одного байта кодувального потоку відбувається порушення декодування всієї секції – що характерно, зокрема, для форматів JPEG XL та Jpegli [5]. Попри їхню високу ефективність і використання сучасних механізмів обробки (адаптивне квантування, модульна архітектура, ентропійне кодування на базі ANS) [6], ці формати залишаються вразливими до помилок, не маючи вбудованих засобів структурної локалізації або самовідновлення. Окремою проблемою є недостатня ефективність традиційного рівномірного поділу простору зображення (наприклад, блоки 8×8), який не враховує нерівномірну структурну насиченість. Унаслідок цього ресурси витрачаються нераціонально, а помилки в критично важливих зонах можуть спричинити каскадне погіршення якості [7].

Аналіз останніх досліджень і публікацій. Основною метрикою для областей когерентності є нормалізований коефіцієнт кореляції $\bar{\mu}(x_{\xi,y};x_{\xi,y+\tau})$, який визначається наступною формулою:

$$\bar{\mu}(x_{\xi,y};x_{\xi,y+\tau})=\frac{\sum_{\xi=1}^{M}\sum_{y=1}^{N}(x_{\xi,y}-x_{\xi,y+\tau})(x_{\xi,y}-x_{\xi,y+\tau})}{R_{kov}},$$

де $x_{\xi,y}$ та $x_{\xi,y+\tau}$ відповідно τ-ий елемент фрагмента зображення, R_{kov} – довжина фрагмента зображення, протягом якої спостерігається статистична залежність між елементами [8].

Водночас, нормалізований коефіцієнт кореляції неефективно враховує наявність психовізуальної надмірності в зображеннях. Тому для опису областей когерентності пропонується використовувати міру амплітуди незначних варіацій δ. Значення δ визначає послідовність (сегмент) зображення, для якої, з точки зору візуального сприйняття, прийнятно апроксимувати значення пікселів таким чином, що збільшує нормалізований коефіцієнт кореляції. В узагальненому вигляді це можна описати наступним чином. Якщо для послідовності X^{ξ} елементи $x_{\xi,y}$ мають нерівні значення та виконується умова

$$x_{\xi,y+r}\in[a_{\xi}-\frac{\delta}{2};a_{\xi}+\frac{\delta}{2}],r=\overline{0,R_{\xi}-1}$$

тоді послідовність X^{ξ} буде областю когерентності (апертурою). Тут, R_{ξ} – довжина послідовності, для якої виконується умова, a_{ξ} – базовий параметр області когерентності, фізичний сенс якого полягає в тому, що при апроксимації опису елементів апертури X^{ξ} зі значенням a_{ξ}, це не призведе до втрат у якості візуального сприйняття зображення. Водночас, перший елемент, для якого умова перестане виконуватися, буде першим елементом наступної сусідньої апертури. Стиснення даних переважно базується на двох підходах до ентропійного кодування: кодуванні Хаффмана та арифметичному/діапазонному кодуванні. Перший підхід набагато швидший, але апроксимує ймовірності степенями 2, що зазвичай призводить до відносно низьких коефіцієнтів стиснення. Другий підхід використовує точні ймовірності, тим самим практично наближаючись до теоретичної межі ефективності стиснення (ентропія Шеннона), але ціною значно вищих обчислювальних витрат [9]. Асиметричні системи числення (ANS) представляють собою новий крок у точному ентропійному кодуванні, поєднуючи переваги обох підходів. Таким чином, для розміру алфавіту 256 швидкість кодування АНС приблизно на 50% вища, ніж кодування Хаффмана, наближаючись до ефективності стиснення арифметичного кодування [10]. Цього спрощення було досягнуто завдяки тому, що замість двох чисел для визначення діапазону використовується одне просте число як поточний стан. Окрім спрощення перенормування, це також дозволяє представити всю поведінку для заданого розподілу ймовірностей у відносно невеликій таблиці, що представляє автомат ентропійного кодуваннях [11].

Постановка завдання. На підставі визначеного протиріччя постає актуальна науково-прикладна задача: підвищення цілісності інформації в умовах внесення канальних помилок, завдяки подальшого вдосконалення методів сегментації та стиснення відеоданих. Отже, мета статті полягає у розробці методу стиснення відеозображень на основі асиметричного кодування та впровадженням механізмів локалізації впливу канальних помилок на цілісність відновлених відеоданих.

Виклад основного матеріалу. Як математичну модель було обрано діапазонно-орієнто-

ване ANS (rANS). Хоча воно має дещо нижчу ефективність стиснення, ніж табличний варіант (tANS), воно менш ресурсоємне, дозволяє працювати з більшими алфавітами та є прямою альтернативою класичному арифметичному кодуванню [12]. Стан на наступному кроці кодування обчислюється наступним чином:

$$blockId = \frac{x}{l_s};\ slot = b_s + \mathrm{mod}(s, l_s);$$

$$C(s,x) = blockId \cdot m + slot,$$

де x – попередній стан, s – позиція поточного символу, який потрібно закодувати, в алфавіті, l_s – частота появи символу в алфавіті, $m = \sum_{i=1}^{s} l_i$, b_s – кумулятивна частота появи символу, розрахована за формулою $b_s = \sum_{i=1}^{s-1} l_i$, що враховує позицію s у поточному циклі. Крок декодування визначається наступним чином: $blockId = \frac{x}{m}$; $s = (\mathrm{mod}(x,m));\quad D(x) = (s, l_s \cdot blockId + slot - b_s)$.

Якщо ми виберемо m як ступень числа 2, множення та ділення на нього можна виконати за допомогою побітових зсувів, і $mod(x,m)$ можна обчислити за допомогою побітового «І» з маскою. У цьому випадку для декодування потрібне лише одне множення на крок. Хоча rANS забезпечує відносно високий коефіцієнт стиснення та працює з високою швидкістю, були виявлені проблеми із завадостійкістю. У наступних експериментах у файл записувався лише результат кодування, тоді як вся службова інформація, така як алфавіт із частотами символів, зберігалася окремо. Навіть у цьому випадку внесення помилок у результат кодування суттєво вплинуло на вихідне зображення після декодування. Нижче наведено приклади зображень після декодування, де в декодовані дані було внесено різний ступінь помилок. Як показано на рис. 1–5, при рівномірній сегментації 8x8 без поділу апертури помилки, внесені в сегменти з високою нерівномірністю, викликають значне спотворення зображення. Це відбувається тому, що rANS має сувору залежність поточного стану кроку кодування від попереднього стану кроку. Отже, коли помилки вносяться в сегмент рівномірної сегментації, весь сегмент спотворюється на випадкові пікселі з алфавіту поточного сегмента. Щоб вирішити цю проблему, зображення пропонується розділити на апертури. В якості першого елемента апертури береться опорний елемент апертури a_ξ. Поріг для незначних змін амплітуди δ було встановлено на 40% значення, отже, умова поділу на апертури у вигляді $x_{\xi,y+r} \in [0.8 \cdot a_\xi; 1.2 \cdot a_\xi]$. В результаті, для зображення 1, представленого на рис. 1, з похибкою введення 10e-1, було отримано наступний результат, який представлено на рис. 6.

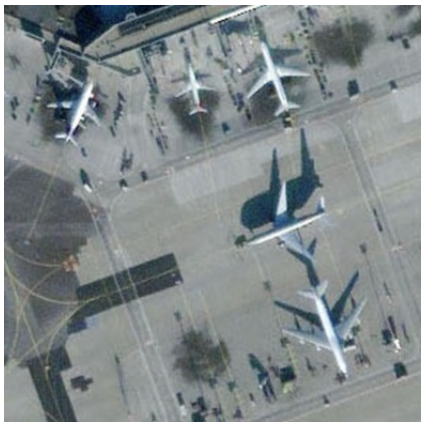


Рис. 1. Вихідне зображення 1, рівномірна сегментація 8x8

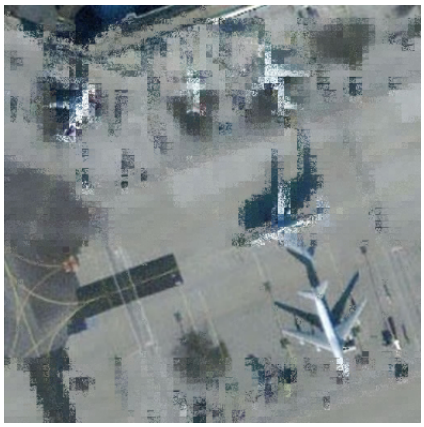


Рис. 2. Зображення 1 з введенням помилок 10e-4, рівномірна сегментація 8x8

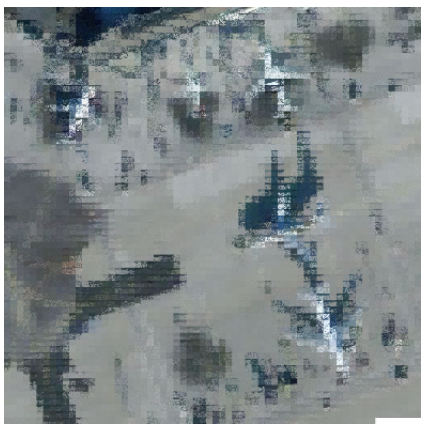


Рис. 3. Зображення 1 із введенням помилок 10e-3, рівномірна сегментація 8x8

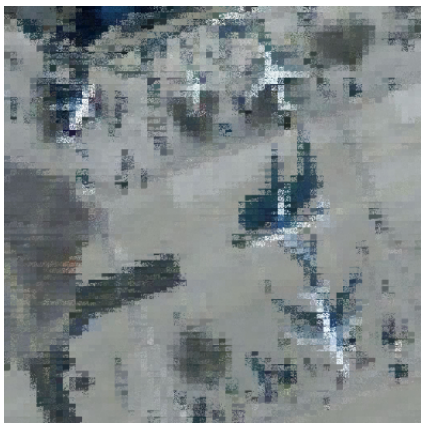


Рис. 4. Зображення 1 з введенням помилок 10e-2, рівномірна сегментація 8x8

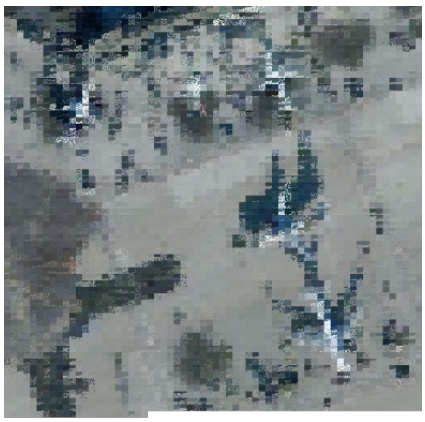


Рис. 5. Зображення 1 з введенням помилок 10e-1, рівномірна сегментація 8x8

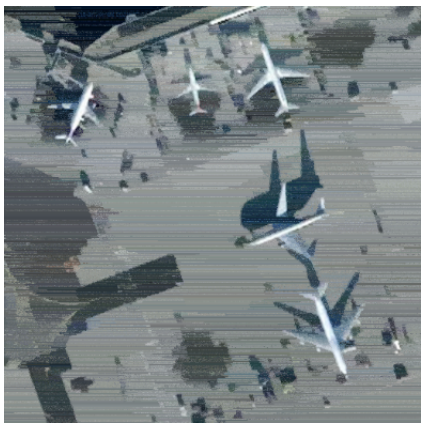


Рис. 6. Зображення 1 з введенням помилки 10e-1, поділ апертури

Очевидно, що об'єкти на зображенні чітко розрізняються, що забезпечує практичне підтвердження ефективності використання аперртурного поділу замість рівномірної сегментації. Цей ефект досягається тому, що алфавіт кожної апертури містить лише схожі пікселі. Навіть якщо в межах апертури виникає помилка та відновлення зображення після декодування rANS може бути порушено, але значення пікселів залишаться в межах алфавітного діапазону, в результаті чого об'єкти на зображенні залишатимуться розрізняними.

Таким чином, хоча rANS забезпечує високу ефективність стиснення даних, вона все ще залежить від правильного поділу зображення на фрагменти та побудови алфавітів з ймовірностями.

Розглянемо приклад зображення 1, представленого на рис. 1. Вихідний масив пікселів зображення, представлений у колірному просторі ARGB, має розмір 2 097 152 байти. Результат кодування за допомогою rANS з рівномірною сегментацією 8x8 після запису у файл становить 406 181 байт. При простому поділі апертури з використанням розмір становить 395 595 байт, що суттєво не відрізняється від сегментації 8x8. Це пояснюється тим, що навіть пікселі з близькими значеннями все ще створюють свій власний символ в алфавіті, а різноманітність символів в алфавіті збільшує зростання стану на наступному кроці, особливо при великій кількості символів з низькими частотами. Однак, при виконанні апроксимації значень у межах кожної апертури:

$$x_{\xi,y} \in [0.9 \cdot a_{\xi}; 1.1 \cdot a_{\xi}]$$

ми отримуємо розмір 228 453 байти, при цьому візуальні характеристики об'єктів на зображенні залишаються незмінними, лише з незначним розмиттям в однорідних сегментах зображення. Цей результат вже перевершує результат для рівномірної сегментації 8x8 і представлений на рис. 7.

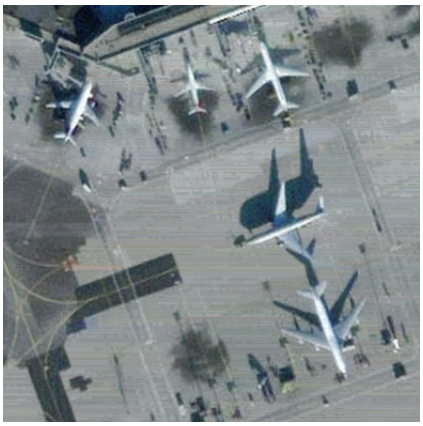


Рис. 7. Зображення 1 з поділом апертури та подальшою апроксимацією значення

Встановлення мінімальної довжини апертури також забезпечує значну перевагу в ефективності стиснення rANS, але також може призвести

до значного збільшення часу розрахунку, якщо мінімальна довжина апертури взята настільки великою. Те саме можна сказати і про великі коефіцієнти апертури. Крім того, в цьому випадку завадостійкість починає значно знижуватися.

Висновки. У цьому дослідженні запропоновано та практично підтверджено ефективність використання апертурного поділу замість рівномірної сегментації для подальшого кодування rANS. Навіть проста реалізація апертурного поділу значно покращує завадостійкість кодованих даних. Додавання апроксимацій у межах апертури підвищує ефективність подальшого кодування за допомогою ANS без шкоди для візуальної якості значних областей та об'єктів на зображенні.

Список літератури:

1. Метод групового кодування інфрачервоних зображень у спектрально-хвильовому просторі. Вісник НТУУ «КПІ». Серія Радіотехніка, Радіоапаратобудування. 2025. № 99. С. 24–34. DOI: 10.20535/RADAP.2025.99.24-34.

2. Ballé J., Laparra V., Simoncelli E.P. End-to-end Optimized Image Compression. 2016. DOI: 10.48550/arXiv.1611.01704.

3. Sun Y., Li L., Ding Y., Bai J., Xin X. Image Compression Algorithm Based On Variational Autoencoder. Journal of Physics: Conference Series. 2021. Vol. 2066. № 1. DOI: 10.1088/1742-6596/2066/1/012008.

4. Метод зберігання елементів службових складових криптокомпресійних кодограм відеозображень. Вісник НТУУ «КПІ». Серія Радіотехніка, Радіоапаратобудування. 2023. № 92. С. 28–40. DOI: 10.20535/RADAP.2023.92.28-40.

5. Rhatushnyak A., Wassenberg J., Sneyers J., Alakuijala J., Vandevenne L., Versari L., Obryk R., Szabadka Z., Kliuchnikov E., Comsa I.-M., Potempa K., Bruse M., Firsching M., Khasanova R., van Asseldonk R., Boukortt S., Gomez S., Fischbacher T. Committee Draft of JPEG XL Image Coding System. DOI: 10.48550/arXiv.1908.03565.

6. Duda J. Asymmetric numeral systems: entropy coding combining speed of Huffman coding with compression rate of arithmetic coding. DOI: 10.48550/arXiv.1311.2540.

7. Barannik V., Onyshchenko R., Krasnorutsky A., Barannik D. The Methods of Intellectual Processing of Video Frames in Coding Systems in Progress Aeromonitor to Increase Efficiency and Semantic Integrity. 2022 IEEE 4th International Conference on Advanced Trends in Information Theory (ATIT). Kyiv, Ukraine. 2022. P. 53–56. DOI: 10.1109/ATIT58178.2022.10024208.

8. Krasnorutsky A., Barannik V., Shulgin S., Kolesnyk V., Kharchenko N., Revva K. Integration of Video Image Decryption Coding into a Remote Video Information Service. 2023 IEEE 5th International Conference on Advanced Information and Communication Technologies (AICT). Lviv, Ukraine. 2023. P. 1–6. DOI: 10.1109/AICT61584.2023.10452679.

9. Duda J., Niemiec M. Lightweight Compression with Encryption Based on Asymmetric Numeral Systems. Applied Mathematics and Computer Science. 2023. DOI: 10.34768/amcs-2023-0004.

10. Mamedov R. Analysis and Enhancement of Lossless Image Compression in JPEG-XL. 2024. DOI: 10.48550/arXiv.2404.19755.

11. Duda J. The Use of Asymmetric Numeral Systems Entropy Encoding in Video Compression. In: Distributed Computer and Communication Networks. 2019. Vol. 862. P. 125–139. DOI: 10.1007/978-3-030-36614-8_10.

12. Технологія ковзного кодування нерівномірних діагональних послідовностей в двовимірному спектральному просторі трансформанти. Вісник НТУУ «КПІ». Серія Радіотехніка, Радіоапаратобудування. 2023. № 94. С. 13–23. DOI: 10.20535/RADAP.2023.94.13-23.

Barannik D.V., Sidchenko Ye.S., Lykhodieiev O.S., Pomazuiev V.V., Fedorovskiy O.V. **METHOD OF APPLYING APERTURES FOR ASYMMETRIC CODING VIDEODATA SYSTEMS**

The article considers the current problems of video image compression for information systems operating in conditions of limited or unstable communication channels. The shortcomings of classical and modern coding formats are analyzed, in particular, their low error resistance and non-adaptability to the structural complexity of images, in particular, the weakness of classical 8x8 segmentation, which does not take into account the features of the image, is shown on a practical example. The feasibility of using asymmetric number systems (ANS), in particular the rANS variant, as an effective entropy coding tool that combines high speed with approaching the theoretical compression limit is substantiated. The shortcomings of ANS are considered, namely, noise immunity and the dependence of the effectiveness of the methods on the input data and alphabets. To solve this problem, further in the article a method of image segmentation based on the detection of local coherence regions characterized by similar structural and visual properties is developed. An aperture division

mechanism is proposed, taking into account the amplitude variations of pixels and adaptive approximation of values inside the apertures. It is shown that such an approach allows to significantly increase the noise immunity to errors during data transmission, localizing their influence within individual image fragments. A comparative experimental study of the effectiveness of the proposed method in combination with rANS coding is carried out. The results obtained demonstrate a reduction in the volume of encoded data without significant loss of visual quality and a significant reduction in the scale of distortions when errors occur in the transmitted stream. Due to the consideration of the features of the image, when introducing channel errors, objects in the image remain clearly visible to the human eye. Experimental results of the study are provided on examples that confirm the effectiveness of the proposed approach in comparison with classical segmentation methods, which confirms the prospects of the methods for use in real time in systems with increased requirements for the reliability of video information.

Key words: *aperture, coherence regions, asymmetric numeral systems, lossy compression, noise, noise immunity.*

УДК 62-533.66
DOI https://doi.org/10.32782/2663-5941/2025.3.2/10

Батюк С.Г.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Воронюк М.В.
Київський національний університет будівництва та архітектури

ЗАДАЧІ КОМПЛЕКСНОЇ АВТОМАТИЗАЦІЇ ЖИТЛОВОГО БУДИНКУ ЯК АВТОНОМНОЇ КІБЕР-ФІЗИЧНОЇ СИСТЕМИ

Стаття присвячена розробці структури кібер-фізичної системи автономного житлового будинку. Автономність житлового будинку забезпечується автономністю системи електропостачання і автономністю системи життєзабезпечення житлового будинку.

Продемонстровано, що ефективним рішенням є впровадження автоматизованої системи автономного електропостачання і життєзабезпечення, яка дозволяє ефективно використовувати альтернативні джерела енергії, накопичувати та розподіляти електроенергію за допомогою сучасних технологій.

Визначено, що задачі ефективої автоматизованої системи управління інженерними системами електропостачання і життєзабезпечення в складі кібер-фізичної системи житлового будинку мають бути задачами комплексної автоматизаці житлового будинку.

Сформульоване визначення комлексої автоматизації як інтегроваої автоматизація технологічних процесів і бізнес-процесів підприємства. Комплексна автоматизована система управління – інтегрована автоматизована система управління масштабу підприємства. Синергетичний ефект комплексної автоматизації – інтегральна (системна) ефективність більша за суму ефектів компонентів (консолідація і агрегація даних). Кумулятивний ефект комплексної автоматизації – цільова ефективність – реалізація бізнес-тріада «звітність – аналітика – планування», зорієнтованої на обгрунтування прийняття рішень власником. Платформою комлексної автоматизації є цифровізація підприємства, що поєднує інформаційні технології і операційні технологі (технології автоматизації). Визначені маркери комплексної автоматизації: 1) інтегрованість (автоматизація усіх технологічних і бізнес процесів); 2) консолідованість (реалізація бізнес-тріади «звітність – аналітика – планування»); 3) адпативність (системи екстремального управління і автотюнінг регуляторів систем автоматичного регулювання).

Розроблена структура автоматизованої системи автономного електропостачання і життєзабезпечення житлового будинку, яка забезпечує стабільне електропостачання і безлюдне функціонування інженерних систем життєзабезпечення житлового будинку. Розроблена структура автоматизованої системи управління інженерними системами житлового будинку реалізує сформульовану авторами ідеологію комплексної автоматизації технологічних процесів.

Ключові слова: *теплова енергетика, інженерні системи будівель, автономне електропостачання житлового будинку, автономне життєзабезпечення житлового будинку, автоматизація технологічних процесів, автоматизація теплоене ргетичних процесів, автоматизація інженерних систем житлового будинку, комплексна автоматизація, кібер-фізична система, автоматизована система управління, система автоматичного регулюваннння, інформаційно-вимірювальна система, система програмно-логічного управління, програмно-технічні засоби автоматизації, контролерна функціональність, супервізорна функціональність, цифровий двійник, граничний девайс, інтернет речей, імітаційне моделювання, функціональне моделювання, структурне моделювання, параметрична адаптація.*

Постановка проблеми. У сучасних умовах питання енергетичної незалежності житлових будинків набуває особливої актуальності. Перебої в електропостачанні можуть призвести до відключення критично важливих систем життєзабезпечення, таких як опалення, водопостачання, вентиляція, зв'язок тощо. Використання традиційних джерел енергії збільшує фінансові витрати та негативно впливає на екологію.

Одним із перспективних рішень є впровадження автоматизованої системи автономного електропостачання і життєзабезпечення

Автономне електрозабезпечення дозволяє ефективно використовувати альтернативні джерела енергії [1], накопичувати та розподіляти електроенергію за допомогою сучасних технологій [2].

Однак в той же час важливо розуміти, що сонячні панелі та вітряки не можуть повністю забезпечити автономність будинку без фідеру енергосистеми як основного джерела живлення і мобільного паливного електрогенератору як резервного джерела живлення.

Аналіз останніх досліджень і публікацій. Сучасні тенденції у сфері автономного електропостачання дедалі частіше фокусуються на інтеграції кількох джерел енергії з метою досягнення максимальної ефективності та надійності [3]. Останні дослідження демонструють, що використання сонячних панелей є ефективним рішенням для забезпечення базових енергетичних потреб будинку в літній період або в режимі мінімального навантаження (stand-by), коли працюють лише критично важливі прилади – освітлення, холодильник, система охорони тощо.

Втім, у зимовий період, коли світловий день коротший, а споживання електроенергії зростає через використання опалення та інших енергомістких приладів, виникає потреба в додаткових джерелах живлення. У таких випадках важливу роль відіграють акумуляторні батареї, які зберігають надлишки енергії, накопичені вдень, або резервні генератори, що забезпечують стабільне постачання електроенергії у разі її нестачі.

Крім того, інтеграція систем життєзабезпечення, таких як опалення, кондиціонування, вентиляція та гаряче водопостачання, у єдину систему енергоменеджменту дозволяє оптимізувати енергоспоживання. Завдяки інтелектуальному управлінню енергією можна досягти вищого рівня автономності та зменшити витрати як на енергію, так і на обслуговування системи в цілому.

Постановка завдання. Метою даного дослідження є розробка структури автоматизованої системи автономного електропостачання та життєзабезпечення житлового будинку. Така система, на додаток до сучасних базових задач автоматизації житлових будтнків [4], має забезпечити стабільне постачання електроенергії незалежно від зовнішніх джерел живлення, а також гарантувати безперебійне й безлюдне функціонування ключових інженерних систем – опалення, вентиляції, кондиціонування, водопостачання та системи безпеки.

Завдання полягає у створенні комплексного рішення, яке поєднує різні джерела енергії (сонячні панелі, акумулятори, резервні генератори) з інтелектуальною системою управління. Остання повинна автоматично реагувати на зміну умов – споживання енергії, погодні умови, рівень заряду акумуляторів – і відповідно адаптувати роботу енергосистеми для досягнення максимальної ефективності та автономності. У результаті система повинна функціонувати з мінімальним залученням людини та забезпечувати комфорт і безпеку у будинку за будь-яких умов.

Комлексна автоматизація інженерних систем електропостачання і життєзабезпечення житлового будинку має враховувати сучасний стан теорії автоматичного управління [5] і досвід промислової автоматизації [6] в частині адаптації САР під параметричні збурення технологічного об’єкту управління [7]. Автори віддають перевагу класичному підходу [8] в частині використання жорсткої адаптації САР (тобто, використанню САР, малочутливих до параметричних збурень) і періодичної активної адаптації САР (тобто, екстремальному управлінню і автотюнінгу САР).

САР в системі комплексній автоматизації житлового будинку мають бути малочутливими до параметричних збурень від інженерних систем (тобто мають мати достатньо великий запа сталості). Для активної адаптації САР мають використовуватись екстремальне управління продуктивністю технологічних агрегатів і автотюнінг ПІД-регуляторів.

Виклад основного матеріалу. У межах дослідження були проаналізовані актуальні підходи до автономного електропостачання житлових будинків та створено функціональну модель кіберфізичної системи, здатної забезпечити повну або часткову енергетичну незалежність житлового об’єкту.

Розроблена система автономного електропостачання житлового будинку ґрунтується на поєднанні кількох взаємопов’язаних модулів, які забезпечують не лише стабільне постачання електроенергії, а й повноцінне управління інженерними системами життєзабезпечення. Такий підхід дозволяє досягти високого рівня автономності, ефективності та гнучкості у використанні ресурсів.

Джерела енергії в системі автономного електроживлення

Автономна система електропостачання використовує кілька типів генерації енергії для забезпечення надійності в різних умовах:

• Сонячні панелі – основне джерело енергії в літній період та в умовах достатньої соняч-

ної активності. Ефективні для підтримки роботи базових систем у денний час, але не здатні забезпечити повну енергонезалежність у хмарні дні або вночі.

• Вітрові генератори – доцільні у вітряних регіонах. Можуть ефективно доповнювати сонячну енергетику вночі або в похмуру погоду, забезпечуючи безперервність генерації.

• Резервні джерела енергії – бензинові, дизельні або газові генератори використовуються у разі високих навантажень або довготривалої відсутності вироблення з основних джерел. Забезпечують надійний резерв у зимовий період або при аварійних ситуаціях.

Накопичення енергії в системі автономного електроживлення

Збереження надлишкової енергії – критично важливий компонент автономності. Система накопичення базується на акумуляторах, які забезпечують живлення в нічний час або при несприятливих погодних умовах.

У системі можуть використовуватись:

• Гелеві акумулятори – стійкі до низьких температур, мають відносно тривалий строк служби, проте менш ефективні при високих навантаженнях.

• Літій-іонні (Li-ion) – характеризуються високою енергоємністю, швидкою зарядкою, тривалим терміном експлуатації (до 15 років), компактністю та мінімальним саморозрядом. Чутливі до перегріву, потребують системи терморегулювання.

• Літій-залізо-фосфатні (LiFePO4) – найефективніші за співвідношенням енергії та надійності. Мають ККД до 98%, термін служби до 15 років, низький рівень деградації та високу пожежобезпечність. Ідеально підходять для сучасних домашніх систем, хоча мають вищу вартість.

Управління енергією в системі автономного електроживлення

Центральною ланкою автономної енергосистеми є Smart Grid – розумна електромережа, яка забезпечує оптимальне керування усіма джерелами та споживачами енергії.

Склад SmartGrid-енергосистеми:

• Інтелектуальні контролери – здійснюють моніторинг вироблення та споживання електроенергії, керують розподілом навантажень, враховуючи пріоритетність підключених систем.

• Алгоритми прогнозування – аналізують дані про погоду, типові шаблони споживання та стан акумуляторів. На основі цього система приймає рішення про доцільність використання збереженої енергії або активацію резервних джерел.

• Система дистанційного моніторингу – користувач може контролювати та налаштовувати роботу енергосистеми через мобільний додаток або веб-інтерфейс. Це дозволяє адаптувати систему до змін у способі життя, тривалих відряджень або сезонних коливань споживання.

Інтеграція із системами життєзабезпечення

Автоматизована система електропостачання тісно взаємодіє з інженерними системами житлового будинку, оптимізуючи їх роботу:

Опалення:

• Використання теплових насосів типу повітря-вода або вода-вода.

• Інтелектуальне управління котлом дозволяє знижувати витрати енергії в періоди відсутності мешканців або вночі.

Водопостачання та водовідведення:

• Автоматизовані насосні станції з регулюванням тиску забезпечують сталість подачі води.

• Системи фільтрації та очищення води покращують якість питної води та підвищують загальний комфорт.

Вентиляція та кондиціонування:

• Системи з рекуперацією тепла дозволяють утилізувати тепло з повітря, що виходить, знижуючи потребу в додатковому обігріванні.

• Інтелектуальне управління кондиціонерами, що враховує температуру, вологість, час доби та присутність людей, зменшує споживання енергії без втрати комфорту.

Ці модулі працюють як єдина система, здатна адаптуватися до змін навантаження, погодних умов, поведінки мешканців та цін на енергію (у разі підключення до «розумної» міської мережі).

Структура автономної кібер-фізичної системи житлового будинку

Кібер-фізична система поєднує фізичні об’єкти (джерела енергії, споживачі, системи накопичення) з операційними технологіями (датчики, програмні алгоритми, комунікаційні інтерфейси).

Структура кібер-фізичної системи

В кібер-фізичній системі вирізняються наступні рівні прийняття рішень з автоматизації і управління:

1. Фізичний рівень

• Генерація: сонячні та вітрові установки.

• Накопичення: акумуляторні системи.

• Споживання: системи опалення, освітлення, побутова техніка.

2. Кібернетичний рівень

• Сенсори: температури, вологості, руху, напруги, струму.

• Актори: реле, електронні клапани, контролери.

• Комунікація: Wi-Fi, Zigbee, Modbus, Bluetooth, Ethernet.

3. Управлінський рівень

• Програмовані логічні контролери (PLC).

• Інтелектуальні модулі прийняття рішень.

• Прогнозно-аналітичні алгоритми, що адаптують поведінку системи в реальному часі.

4. Інтерфейсний рівень

• Мобільний додаток або веб-інтерфейс.

• Голосове управління (опціонально).

• Візуалізація даних у режимі реального часу (споживання, заряд, статуси пристроїв).

У результаті кібер-фізична система функціонує як єдиний «організм», що здатен до самодіагностики, саморегулювання і, за потреби, повідомлення власника або технічну службу про необхідність втручання. Така система створює основу для повністю безлюдного функціонування житла, з одночасною енергонезалежністю та високим рівнем комфорту.

Принципи роботи кібер-фізичної системи житлового будинку

Автоматизована система автономного електропостачання житлового будинку побудована на основі адаптивних алгоритмів управління [7], що дозволяють забезпечити ефективне використання енергії, зниження витрат та підтримання стабільної роботи інженерних систем навіть у непередбачуваних умовах. Основні принципи функціонування системи проаналізовані нижче.

Пріоритетне використання відновлюваних джерел енергії

Система зорієнтована на максимальне використання сонячної та вітрової енергії як основних джерел живлення. У сприятливих умовах генерації (сонячна погода, наявність вітру) вона забезпечує повне або часткове покриття потреб будинку за рахунок цих джерел. Нижче на рис. 1 наведена діаграма зміни рівня автономності електропостачання житлового будинку залежно від сезону. Надлишок електроенергії автоматично направляється на заряд акумуляторів, що дозволяє використовувати його в нічний час або при несприятливій погоді.

Найвища автономність досягається влітку, коли сонячні панелі працюють максимально ефективно, а найнижча – взимку через короткий світловий день і високе енергоспоживання

Автоматична оптимізація споживання

Інтелектуальні контролери постійно відстежують рівень споживання енергії і порівнюють його з доступними ресурсами.

У разі перевищення критичних значень або швидкого розряду акумуляторів система автоматично:

• Визначає непріоритетні електроприлади (наприклад, декоративне освітлення, підігрів підлоги, зарядні пристрої) і тимчасово відключає їх.

• Оптимізує роботу життєво важливих систем, наприклад, знижує потужність обігрівачів або переходить на економний режим вентиляції.

Це дозволяє знизити пікові навантаження та уникнути аварійного відключення систем.

Автоматичне перемикання між джерелами енергії

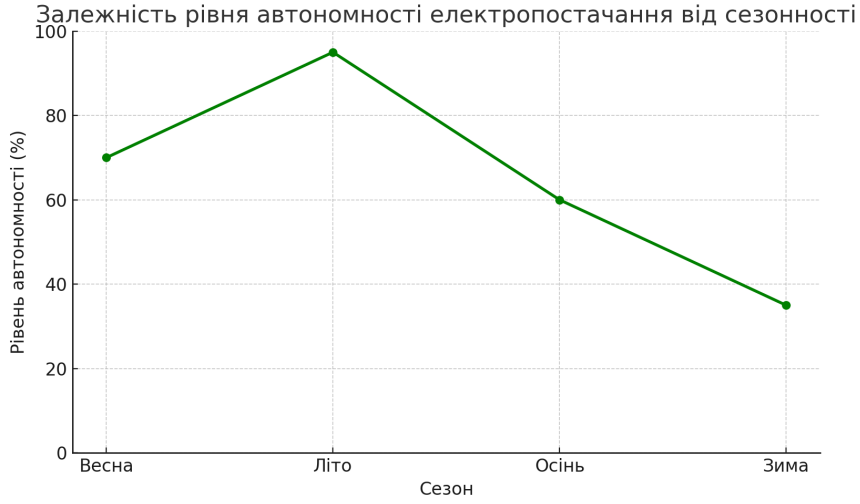


Рис. 1. Діаграма зміни рівня автономності електропостачання житлового будинку залежно від сезону

Система автоматично керує джерелами живлення залежно від наявності енергії в акумуляторах і потреб споживання. Розрізняють три основні режими роботи:

• Stand-by режим: мінімальне енергоспоживання (робота лише базових систем – охорона, холодильник, освітлення). Використовується в разі тривалої відсутності мешканців або енергетичної кризи.

• Стандартний режим: використовується накопичена енергія з акумуляторів, а за потреби – підключаються поновлювані джерела.

• Екстрений режим: при критичному зниженні заряду або несприятливих погодних умовах система автоматично вмикає резервний генератор (дизельний або газовий) або, за наявності, перемикається на зовнішнє електропостачання (наприклад, міську мережу).

Ці сценарії реалізуються без участі користувача, завдяки вбудованим алгоритмам реагування.

Дистанційний контроль та моніторинг

Система обладнана інтерфейсом дистанційного керування, який дозволяє власнику:

• Моніторити рівень генерації, заряд акумуляторів, стан споживання в реальному часі.

• Змінювати пріоритетність споживачів або режими роботи систем (наприклад, увімкнути економний режим під час відпустки).

• Отримувати повідомлення про несправності, перевантаження або відключення живлення.

Доступ до системи здійснюється через мобільний додаток або веб-інтерфейс, що працює в захищеному середовищі та підтримує авторизацію користувача.

Опис структури автономної кібер-фізичної системи житлового будинку

Автономна кібер-фізична система житлового будинку (КФС ЖБ) реалізована у вигляді автономного автоматизованого технологічного комплексу (ААТК ЖБ), який забезпечує енергонезалежне функціонування як систем електропостачання, так і систем життєзабезпечення.

Структура ААТК ЖБ

ААТК ЖБ складається з двох основних підсистем:

• АТК системи автономного електропостачання (САЕ).

• АТК системи автономного життєзабезпечення (САЖ).

Система автономного електропостачання (САЕ)

Задача САЕ – забезпечити постійне електропостачання будівлі (постійне електроживлення будівлі) незалежно від зовнішніх нештатних ситуацій і загроз.

Структура САЕ

• ФЦЕ – фідер центральної енергосистеми (резервне підключення до мережі).

• АСЕ – автономна сонячна електростанція.

• АВЕ – автономна вітряна електростанція.

• МПЕ – мобільний паливний електрогенератор (дизель/газ).

• СРЕ – система резервування електропостачання (акумулятори, стабілізатори).

Ці компоненти інтегруються в автоматизований технологічний комплекс (АТК) САЕ, який складається із самої САЕ та системи програмно-логічного управління (СПЛУ).

СПЛУ САЕ реалізована на базі автономних програмованих логічних контролерів (ПЛК), кожен з яких виконує роль IIoT-пристрою.

До складу СПЛУ входять:

• СПЛУ/ПЛК ФЦЕ.
• СПЛУ/ПЛК АСЕ.
• СПЛУ/ПЛК АВЕ.
• СПЛУ/ПЛК МПЕ.
• СПЛУ/ПЛК СРЕ.

Загалом – 5 ПЛК, які взаємодіють через Industrial Ethernet і підтримують протоколи OPC-UA та MQTT.

Система автономного життєзабезпечення (САЖ)

Здача САЖ – забезпечити постійне життєзабезпечення будівлі (постійну роботу інженерних систем будівлі) незалежно від зовнішнії нештатних ситуацій і загроз.

Структура САЖ

• ЦЕК – центральний електричний котел.
• СПП – система підігріву підлоги.
• СРО – система радіаційного опалення.
• СГВП – система гарячого водопостачання.
• СВК – система водопостачання і каналізації.

Як і у випадку з САЕ, САЖ реалізується в складі АТК САЖ, що містить відповідну систему автоматичного управління (САУ) з 5 ПЛК, кожен із яких є автономним IIoT-вузлом.

До складу САУ входять:

• САР/ПЛК ЦЕК.
• САР/ПЛК СПП.
• САР/ПЛК СРО.
• САР/ПЛК СГВП.
• САР/ПЛК СВК.

Інтегрована структура ААТК ЖБ

Загальна архітектура ААТК ЖБ передбачає поєднання технологічних об’єктів управління

(ТОУ ЖБ) та автоматизованої системи управління (АСУ ЖБ).

До складу АСУ ЖБ входять:

- СПЛУ САЕ.
- САУ САЖ.
- Систему технічної діагностики (СТД).
- Систему адаптації параметрів (САП).

Система технічної діагностики (СТД ТОУ ЖБ)

Реалізована у вигляді хмарної платформи з використанням цифрового двійника (ЦД ЖБ).

Містить експертну систему управління (ЕСУ), яка використовує дедуктивне машинне навчання для прогнозування несправностей, оптимізації ресурсів та передачі керуючих впливів.

Система адаптації параметрів (САП АСУ ЖБ)

САП забезпечує самоналаштування локальних підсистем на основі індуктивного машинного навчання.

Структури САП

- Системи екстремального управління (СЕУ) – реалізують градієнтний пошук оптимальних значень техніко-економічних параметрів (ТЕП), наприклад, за допомогою:
- Позиціювання сонячних панелей (АСЕ) по відношенню до сонця.
- Адаптивне регулювання кута атаки лопатей вітрових установок (АВЕ).
- Системи автотюнінгу регуляторів (САТ) – здійснюють автоматичне налаштування регуляторів у процесі роботи об'єкта управління (ОУ).

Центральний ПЛК – граничний пристрій

У системі використовується центральний граничний ПЛК, який є концентратором даних і координатором локальних ПЛК.

Функціональність центрального ПЛК

- Реалізований як промисловий ПК з CoDeSys 3.5 (софтПЛК).
- Підтримує HIL-моделювання АТК, OPC-UA, MQTT, Industrial Ethernet (до 1024 м через маршрутизатори).
- Інтегрований із хмарною HMI/SCADA-системою AVEVA System Platform та хмарним архіватором даних.

Режими управління в КФС ЖБ

Система підтримує декілька рівнів та режимів керування.

Для експертної системи (ЕСУ):

- ПЦУ (пряме цифрове управління) – граничний ПЛК автоматично передає уставки до локальних ПЛК.
- ПОР (режим порадника) – оператор перевіряє та дозволяє передавання уставок від ЦД.

Для систем автотюнінгу (САР):

- АВТ (автоматичний) – ПЛК автономно керують виконавчими механізмами (ВМ) згідно з уставками.
- НАВ (напівавтоматичний) – оператор задає уставки, ПЛК керують ВМ.
- РУЧН (ручний) – оператор напряму керує ВМ, минаючи ПЛК.

Цей опис демонструє складність, гнучкість і масштабованість автономної КФС житлового будинку як кібер-фізичної платформи нового покоління для енергоефективного та стабільного проживання.

Аналіз ефективності КФС ЖБ

Ефективність запропонованої кібер-фізичної системи житлового будинку (КФС ЖБ) оцінюється за кількома ключовими критеріями: енергетична автономність, економічна доцільність, технічна надійність, масштабованість та адаптивність до змін середовища.

Енергетична автономність КФС ЖБ

Завдяки використанню відновлюваних джерел енергії (сонячні панелі, вітрові генератори) та систем накопичення енергії:

- Рівень автономності у літній період сягає 80–100% (за сприятливих погодних умов).
- У зимовий період або при пікових навантаженнях автономність зберігається на рівні 60–75% завдяки підключенню резервних джерел живлення.
- Інтелектуальна система управління забезпечує оптимальний баланс між генерацією, споживанням і накопиченням енергії, мінімізуючи втрати.

Економічна ефективність КФС ЖБ

Техніко-економічні показники інсталяції АСУ ЖБ враховують інвестиції в обладнання і експлуатацію та очікуваний термін його служби (від 10 до 20 років):

- Термін окупності системи: від 5 до 8 років залежно від конфігурації та цін на енергоносії. Нижче на рис. 2 наведений графік окупності для трьох різних типів систем електропостачання.
- Скорочення витрат на електроенергію: до 70% на рік у порівнянні з традиційним підключенням до мережі.
- Вартість обслуговування значно зменшується завдяки системам самодіагностики та віддаленого моніторингу.

Автономна КФС, попри вищі початкові витрати, окупається приблизно за 7–8 років, що робить її вигідною інвестицією в довгостроковій перспективі порівняно з іншими системами.

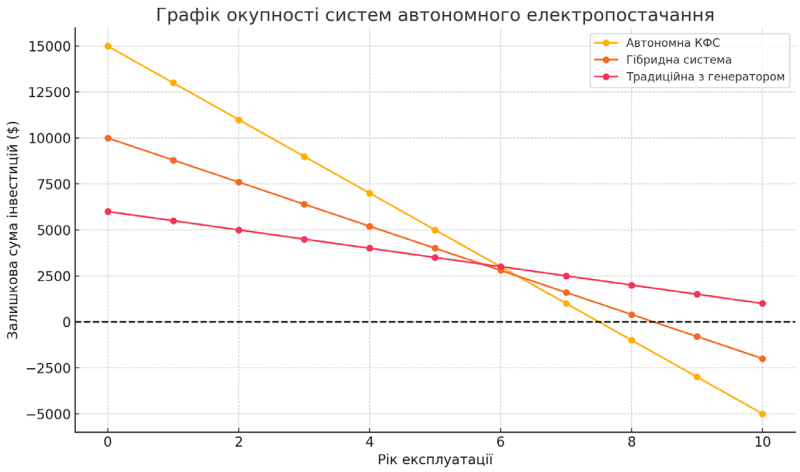


Рис. 2. Графік окупності для трьох типів систем електропостачання

Технічна надійність КФС ЖБ

• Кожен функціональний модуль системи має резервування критичних елементів (джерела живлення, мережеві вузли, сенсори).

• Вбудовані алгоритми діагностики (СТД + ЦД + ЕСУ) дозволяють заздалегідь виявляти потенційні збої та планувати технічне обслуговування.

• Використання IIoT-архітектури та модульного підходу забезпечує стійкість до локальних збоїв і можливість гарячої заміни компонентів.

Адаптивність і масштабованість КФС ЖБ

• Система підтримує адаптацію до змінних кліматичних, експлуатаційних і поведінкових умов, завдяки САП (СЕУ + САТ), які реалізують індуктивне машинне навчання.

• Архітектура дозволяє масштабування – як горизонтальне (додавання нових зон чи модулів), так і вертикальне (інтеграція з міськими або регіональними енергомережами).

• Протоколи обміну (OPC-UA, MQTT) та хмарна SCADA-платформа дозволяють інтегрувати систему в єдине цифрове середовище «розумного міста».

Екологічна ефективність КФСЖБ

Екологічна ефективність є невід'ємною складовою профілю автономного житлового будинку.

Джерела екологічної ефективності:

• Скорочення споживання викопного палива та зменшення викидів CO_2 – до 2–4 тонн на рік для одного будинку середнього розміру.

• Використання чистих джерел енергії відповідає принципам сталого розвитку та екологічної відповідальності.

КФС система автономного електропостачання та життєзабезпечення демонструє високий рівень техніко-економічної ефективності, забезпечуючи:

• Енергетичну незалежність користувача.

• Зниження експлуатаційних витрат.

• Надійність та безперервність життєзабезпечення навіть у позаштатних умовах.

• Гнучкість та адаптивність до змін зовнішнього середовища.

Нижче в табл. 1 представлена порівняльна характеристика ефективності систем електропостачання. Таблиця демонструє переваги автономної кібер-фізичної системи (КФС), зокрема високий рівень автономності, екологічність, економічність та гнучкість управління в порівнянні з іншими варіантами.

Таким чином, впровадження подібних систем у житловому будівництві є доцільним і стратегічно важливим кроком до підвищення енер-

Таблиця 1

Порівняльна таблиця ефективності систем електропостачання

Показник	Автономна КФС	Гібридна система	Традиційна система
Інтегрованість	10	7	4
Консолідованість	10	6	3
Адаптивність	9	6	3
Окупність (років)	7	8	5
Рівень автономності (%)	85	60	40
Екологічність (1–10)	9	7	3

гетичної безпеки та цифрової трансформації інфраструктури.

Нижче на рис. 3 наведена структура КФС ЖБ (яка реалізує принцип комплексної автоматизації житлового будинку). Структура КФС ЖБ є універсальною і забезпечує автономність ЖБ від зовнішніх нештатних ситуацій і загроз техногенного і воєнного характерів.

В структурній схемі КФС ЖБ використані наступні позначення:

• САЕ – Система автономного електропостачання, до складу якої входять: ФЦЕ – Фідер центральної енергосистеми. АСЕ – Автономна сонячна електростанція. АВЕ – Автономна вітряна електростанція. МПЕ – Мобільний паливний електрогенератор. СРЕ – Система резервування електропостачання.

• САЖ – Система автономного життєзабезпечення, до складу якої входять: ЦЕК – Центральний електричний котел. СПП – Система підігріву підлоги. СРО – Система радіаційного опалення. СГВП – Система гарячого водопостачання. СВК – Система вентиляції і кондиціювання.

Розроблена структура КФС ЖБ забезпечує стабільне електропостачання і безлюдне функціонування інженерних систем життєзабезпечення житлового будинку.

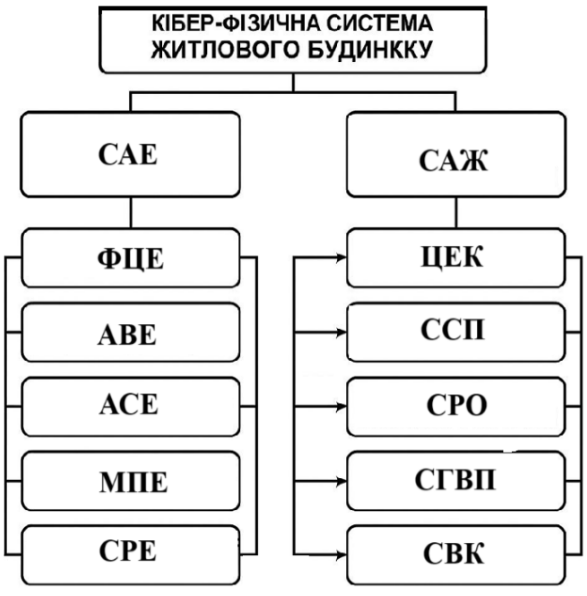


Рис. 3. Схема структурна автномної КФС ЖБ

Висновки. У ході розробки було створено модель автономної кібер-фізичної системи житлового будинку (КФС ЖБ), що реалізує принципи комплексної автоматизації – глибоко інтегрованої автоматизації як технологічних процесів (систем електропостачання та життєзабезпечення), так і функціональних бізнес-процесів, спрямованих на управління ресурсами, звітністю та прогнозуванням.

Комплексна автоматизована система управління (КАСУ)

Розроблена КФС ЖБ відповідає сучасній парадигмі КАСУ – комплексної автоматизованої системи управління масштабу об'єкта.

Основні ознаки реалізованої КАСУ:

• Інтегрованість: усі підсистеми, зокрема САЕ (автономне електропостачання) та САЖ (життєзабезпечення), охоплені автоматизованим керуванням за єдиною архітектурою.

• Консолідованість: поєднання цифрового двійника (ЦД) об'єкта з експертною системою управління (ЕСУ) дозволяє реалізувати активну систему технічної діагностики (СТД ТОУ) [9].

• Адаптивність: автономна система адаптації параметрів (САП), що містить системи екстремального управління (СЕУ) та системи автотюнінгу (САТ), забезпечує динамічну оптимізацію параметрів у реальному часі.

Досягнутий ефект від інсталяції КАСУ

Синергетичний ефект досягається за рахунок об'єднання даних усіх підсистем в єдине інформаційно-аналітичне середовище, що дозволяє:

• підвищити енергоефективність роботи будинку (через погоджене керування САЕ і САЖ);

• покращити точність діагностики стану обладнання;

• зменшити потребу в людському втручанні (режими ПЦУ та АВТ).

Кумулятивний ефект проявляється у забезпеченні бізнес-тріади «звітність – аналітика – планування» для власника будівлі, що досягається завдяки:

• автоматизованому збору, зберіганню і візуалізації даних через хмарну SCADA-платформу;

• використанню алгоритмів машинного навчання для прогнозування навантажень і адаптації систем;

• створенню цифрового близнюка як бази для симуляцій, аналізу та прийняття рішень.

ААТК ЖБ як платформа цифровізації ЖБ

Автоматизована архітектура ААТК ЖБ реалізує платформу комплексної автоматизації, яка об'єднує:

• Інформаційні технології (ІТ): хмарна SCADA, ЦД, мобільний інтерфейс управління, машинне навчання;

• Операційні технології (ОТ): IIoT-пристрої, ПЛК, автономні об'єкти управління, інтелектуальні регулятори.

Загальний висновок щодо комплексної автоматизації автономного ЖБ

Комлексна автоматизація – інтегрована автоматизація технологічних процесів і бізнес-процесів підприємства. Комплексна АСУ (КАСУ) – інтегрована АСУ масштабу підприємства. Синергетичний ефект – інтегральна (системна) ефективність більша за суму ефектів компонентів (консолідація і агрегація даних). Кумулятивний ефект – цільова ефективність – бізнес-тріада «звітність – аналітика – планування», зорієнтована на власника. Платформа комлексної автоматизації – цифровізація підприємства = інформаційні технології + операційні технології. Маркери КАСУ: 1) інтегрованість (автоматизація усіх технологічних і бізнес процесів); 2) консолідованість (СТД ТОУ = ЦД ТОУ + ЕСУ); 3) адпативність (СМН = СЕУ + САТ).

Запропонована структура кібер-фізична системи дозволяє забезпечити стабільне, автономне, енергоефективне функціонування житлового будинку, орієнтоване на довгострокову надійність, зниження витрат і цифрову трансформацію інфраструктури.

Таким чином, модель КФС ЖБ може бути використана як референтна архітектура для проектування «розумних будинків» та об'єктів критичної інфраструктури, здатних функціонувати у відриві від центральних мереж і людського втручання.

На стадії проєктування функціональність КАСУ моделюється і тестується в програмно-технічному симуляторі промислових АТК [10]. Для апроксимації динаміки ОУ використовуються інваріантні моделі об'єктів управління і універсальні структури ПІД-регуляторів [11].

Список літератури:

1. Ткаченко П.М., Іванов Ю.В. Енергетичні системи з відновлюваними джерелами енергії. Харків : ХНУРЕ, 2020. 278 с.

2. Khoa, N.T., & Lee, M. IoT-based Energy Management for Smart Buildings with Renewable Sources and Battery Storage. *Renewable Energy*. 2021. Vol. 170, P. 203–216.

3. Khalil, H., & Zaytoon, J. Smart Home Energy Management Systems: A Review. *IEEE Transactions on Industrial Informatics.* 2019. Vol. 15 (2), P. 1221.

4. Ковтун В.І. Системи автоматизації житлових будинків: сучасні рішення та перспективи розвитку. Львів : ЛНУ, 2022. 198 с.

5. Попович М.Г., Ковальчук О.В. Теорія автоматичного керування : підручник. Київ : Либідь, 2007. 656 с.

6. Баранов С.А., Ковальов Д.О. Автоматизовані системи управління: принципи побудови та приклади реалізації. Київ : Наукова думка, 2021. 312 с.

7. Novikov, P., Shtifzon, O., Bunke, O., & Batiuk, S. Selecting a method for the parametric adaptation of pi-controller in the control systems of boiler assemblies at thermal power stations with supercritical parameters. *Eastern-European Journal of Enterprise Technologies*. 2022. 2, 2(116). P. 61-68. DOI: https://doi.org/10.15587/1729-4061.2022.254116.

8. Батюк С.Г., Ковриго Ю.М., Бунь В.П. Порівняльний аналіз класичного і некласичного підходів до створення АСУ ТП в енергетиці на основі сучасних ПТК. *Наукові вісті НТУУ КПІ*. 1999. № 2. С. 40–48.

9. Пастухов А.В., Литвиненко В.С. Системи технічної діагностики в енергетиці. Дніпро : НГУ, 2021. 245 с.

10. Batiuk S.G. Simulation modeling and digital twinning of energy cyber-physical systems (cyber-energy systems). *Achievements of Ukraine and EU countries in technological innovations and invention* : collective monograph. Riga : Baltija Publishing, 2022. P. 44–109. DOI: https://doi.org/10.30525/978-9934-26-254-8. DOI: https://doi.org/10.30525/978-9934-26-254-8-2.

11. Батюк С.Г., Васянович В.М., Воронюк М.В., Ворошилов А.І. Інваріантні моделі об'єктів управління і універсальні структури ПІД-регуляторів в імітаційному моделюванні теплоенергетичних систем автоматичного регулювання. *Вчені записки Таврійського національного університету імені В.І. Вернадського. Серія: Технічні науки.* 2024. Том 35 (74). № 6. Частина 2. С. 1–12. DOI: https://doi.org/10.32782/2663-5941/2024.6.2/01.

Batiuk S.G., Voroniuk M.V. TASK OF COMPREHENSIVE AUTOMATION OF A RESIDENTIAL BUILDING AS AN AUTONOMOUS CYBER-PHYSICAL SYSTEM

The article is devoted to the development of the structure of the cyber-physical system of an autonomous residential building. The autonomy of a residential building is ensured by the autonomy of the power supply system and the autonomy of the life support system of the residential building.

It is demonstrated that an effective solution is the implementation of an automated system of autonomous power supply and life support, which allows for the effective use of alternative energy sources, accumulation and distribution of electricity using modern technologies.

It is determined that the tasks of an effective automated system for managing engineering systems of power supply and life support as part of the cyber-physical system of a residential building should be the tasks of complex automation of a residential building.

The definition of complex automation is formulated as an integrated automation of technological processes and business processes of an enterprise. A complex automated control system is an integrated automated management system of the enterprise scale. The synergistic effect of complex automation is an integral (system) efficiency greater than the sum of the effects of the components (data consolidation and aggregation). The cumulative effect of complex automation is target efficiency – implementation of the business triad "reporting – analytics – planning", focused on substantiating decision-making by the owner. The platform of complex automation is the digitalization of the enterprise, which combines information technologies and operational technologies (automation technologies). The markers of complex automation are determined: 1) integration (automation of all technological and business processes); 2) consolidation (implementation of the business triad "reporting – analytics – planning"); 3) adaptability (extreme control systems and auto-tuning of regulators of automatic regulation systems).

The structure of the automated system of autonomous power supply and life support of a residential building has been developed, which ensures stable power supply and unmanned operation of engineering systems of life support of a residential building. The developed structure of the automated system of management of engineering systems of a residential building implements the ideology of complex automation of technological processes formulated by the authors.

Key words: *thermal energy, engineering systems of buildings, autonomous power supply of a residential building, autonomous life support of a residential building, automation of technological processes, automation of thermal energy processes, automation of engineering systems of a residential building, complex automation, cyber-physical system, automated control system, automatic regulation system, information and measuring system, system of program-logic control, software-technical means of automation, controller functionality, supervisory functionality, digital twin, edge device, Internet of Things, simulation modeling, functional modeling, structural modeling, parametric adaptation.*

УДК 62-533.66
DOI https://doi.org/10.32782/2663-5941/2025.3.2/11

Батюк С.Г.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Воронюк М.В.
Київський національний університет будівництва та архітектури

ЗАСТОСУВАННЯ ЦИФРОВИХ ДВІЙНИКІВ І СИСТЕМ ШТУЧНОГО ІНТЕЛЕКТУ В АВТОМАТИЗАЦІЇ ТЕПЛОЕНЕРГЕТИЧНИХ ПРОЦЕСІВ

Стаття присвячена використанню цифрових двійників в класичних задачах автоматизації теплоенергетичних процесів.

Сформульовані робочі поняття цифрових двійників і систем штучного інтелекту для задач автоматизації. Продемонстровано, що технології цифрових двійників і систем штучного інтелекту широко використовувались і використовуються в автоматизації технологічних процесів взагалі і автоматизації теплоенергетичних процесів зокрема.

Проаналізовані класичні приклади використання цифрових двійників і систем штучного інтелекту в автоматизації теплоенергетичних процесів: 1) цифрові двійники в задачах верхнього (супервізорного) рівня автоматизації методичних печей прокатних станів – розрахунок в реальному часі температурних епюр металевих слябів; 2) системи штучного інтелекту в задачах нижнього (контролерного) рівня автоматизації промислових топок – градієнтний пошук оптимального значення коефіцієнту надлишку повітря; 3) цифрові двійники в задачах верхнього (супервізорного) рівня автоматизації енергоблоку елетростанції – технічна діагностика тепло-механічного стану електрогенератору; 4) цифрові двійники в задачах нижнього (контролерного) рівня автоматизації енергоблоку елетростанції – фільтрація виміряних температур корпусу електрогенератору; 5) цифрові двійники в задачах верхнього (супервізорного) рівня автоматизації скловарної печі – розрахунок температури і рівня скломаси в басейні печі; 6) цифрові двійники в задачах нижнього (контролерного) рівня автоматизації скловарної печі – формування команди на перекидання факелу в регенеративній топці; 7) системи штучного інтелекту в задачах верхнього (супервізорного) рівня автоматизації лінії формування скляних труб методом горизонтального витягування – градієнтний пошук мінімуму собівартості виробів; 8) цифрові двійники в задачах нижнього (контролерного) рівня автоматизації лінії формування скляних труб методом горизонтального витягування – формування уставки регулятору аеродинамічного режиму формування склотрубок; 9) системи штучного інтелекту в задачах нижнього (контролерного) рівня автоматизації теплоенергетичних процесів – автотюнінг ПІД-регуляторів; 10) системи штучного інтелекту в задачах комплексної автоматизації інженерних систем – екстремальне управління позиціюванням панелей батарей сонячних електростанцій і лопатей турбін вітряних електростанцій.

Ключові слова: *четверта промислова революція, кібер-фзична система, інтернет речей, граничний девайс, інтелектуальні системи управління, цифровий двійник, штучний інтелект, агент штучного інтелекту, система штучного інтелекту, комплексна автоматизація, автоматизація теплоенергетичних процесів, автоматизація інженерних систем, автоматизована система управління, система автоматичного регулюваннння, контролерна функціональність, супервізорна функціональність, адаптивні системи управління, експертні системи управління, системи екстремального управління, автотюнінг ПІД-регуляторів, системи з навчанням, системи з самонавчанням, моделювання в реальному часі.*

Постановка проблеми. Цифрові двійники (ЦД) і системи штучного інтелекту (СШІ; наприклад, ШІ-агенти) активно використовуються в промисловості і будівництві. Основні напрями застосування ЦД – технічна діагностика і попереджувальне обслуговування технологічних агрегатів. Основні напрями використання ШІ-агентів – сервісні функції для побутових і промислових гаджетів. Актуальним є питання – яким чином ЦД і СШІ можуть бути використані в автоматизації технологічних об’єктів управління (ТОУ) взагалі і в автоматизації інженерних систем зокрема?

Аналіз останніх досліджень і публікацій. ЦД і СШІ – надновітні цифрові технології (особливо ШІ-агенти). Традиційно ЦД і СШІ відносять до інформаційних технологій (ІТ). Використання ЦД і СШІ як операційних технологій (ОТ), тобто як технологій автоматизації, є дискусійним. Виникають наступні питання.

1. Чи були раніше вже реалізовані ЦД і СШІ в автоматизації технологічних процесів?

2. Якщо були, то що це були за рішення і чи були вони ефективними? І чи не є терміни «ЦД» і «СШІ» новими назвами відомих і ефективних (чи неефективних) рішень з автоматизації? Як, наприклад, новітній термін «кібер-фізична система (КФС)» замість класичного терміну «автоматизований технологічний комплекс (АТК)».

3. Якщо ЦД і СШІ вже імпліцитно використовувались в класичних системах автоматичного регулювання (САР), то які перспективи їх використання в сучасних автоматизованих системах управління технологічними процесами (АСУ ТП) ТОУ, зокрема АСУ технологічними процесами?

Постановка завдання. Для змістовних відповідей на перераховані вище актуальні питання щодо використання ЦД і СШІ в промисловій автоматизації потрібно проаналізувати і визначити наступне.

1. Дати робочі визначення ЦД і СШІ для ОТ.

2. Вказати аналоги ЦД і СШІ в задачах класичної промислової автоматизації, проаналізувати їх ефективність і перспективність їх застосування в автоматизації технологічних процесів технологічних агрегатів.

3. Надати приклади застосування ЦД і СШІ в автоматизації теплоенергетичних процесів.

Виклад основного матеріалу. ЦД – це «дитина» 4-ої промислової революції. Промислова революція почалась приблизно від 1800-го року і неперервно продовжується вже понад 200 років, докорінно змінюючи умови життя і праці людства. Класична теорія автоматичного управління (ТАУ) [1] і, відповідно, класична промислова автоматизація [2] – це «діти» попередніх етапів (1-го, 2-го і 3-го) перманентної промислової революції. ШІ і СШІ разом з роботизацією – вже «діти» 5-ої промислової революції (яка, власне, і розпочалось з впровадження СШІ в ІТ).

Що спільного у ЦД/СШІ з класичною автоматизацією і чи, взагалі, релевантні аналогії між сучасними ЦД/СШІ і класичними рішеннями з автоматизації кінця 20-го століття? Аналіз показує, що так і що деякі добре відомі класичні ОТ [3] є аналогами сучасних ЦД і СШІ. Більше того, відносна рідкість застосування цих рішень в промисловій автоматизації ретроспективно демонструє, що вони випередили свій час, бо не могли бути реалізовані тоді з причини обчислювальних обмежень тогочасних програмно-технічних засобів автоматизації.

Визначення ЦД для ОТ

ЦД – це цифрова копія (імітаційна модель) ТОУ, яка керується виміряними параметрами і використовується для обчислення параметрів, які не вимірюються. Модель може бути точною або приблизною, динамічною або статичною, формульною, табличною або алгоритмічною. ЦД не визначається способом свого застосування; засадничим для визначення моделі ТОУ як ЦД ТОУ є використання в моделі виміряних параметрів для обчислення невиміряних параметрів. Обчислені параметри можуть бути як технологічні (температури, тиски, рівні, витрати, тощо) так і техніко-економічні (ТЕП; КПЕ – ключові показники ефективності; собівартість, випрацювання ресурсу, надійність функціонування, показники якості перехідних процесів). Класичне застосування ЦД ТОУ – технічна діагностика обладнання ТОУ з метою попереджувального обслуговування (predictive maintenance). Як правило, ЦД має свою цифрову тінь (історію, тобто історичні дані в архіваторах), але наявність цифрової тіні не є засадничою для дефініції ЦД ТОУ.

Визначення СШІ для ОТ

СШІ – це система машинного навчання (СМН). Взагалі, ШІ – це не пристрій, не програма і не алгоритм. ШІ – це галузь кібернетики, яка вивчає (аналізує і синтезує) системи, здатні до навчання. Термін «штучний інтелект» є неточним перекладом з англійського «artificial intelligence (AI)» (порівняння: CIA – Central Intelligence Agency – ЦРУ). Intelligence – це здатність досліджувати, здобувати знання та навички. Intellect – це здатність міркувати та об›єктивно розуміти. Відповідно, англомовний термін AI точніше перекласти як «штучне дослідження / штучний дослідник». Алгоритми ШІ не імітують роботу людського мозку (хоча нейронні мережі, на основі яких працює людський мозок, є найбільш поширеним механізмом реалізації алгоритмів ШІ, це не єдиний такий алгоритм). Алгоритми ШІ імітують процес навчання.

СШІ – це пристрій або програма для реалізації алгоритмів ШІ. Інакше, СШІ – це пристрій або програма, які можуть навчатись. Терміни

ШІ і «машинне навчання» (МН) практично тотожні за значенням (хоча МН в ШІ розглядається як одна з технологій ШІ). Вирізняють дедуктивне машинне навчання (ДМН) та індуктивне машинне навчання (ІМН). ДМН – формалізація знань експертів у вигляді баз знань (БЗ). ІМН – пошук емпіричних залежностей в експериментальних даних. Алгоритми ДМН реалізуються експертними системами управління (ЕСУ). Алгоритми ІМН реалізуються пошуковими системами управління (ПСУ). ЕСУ і ПСУ відомі як інтелектуальні системи управління (ІСУ). ПСУ, за визначенням і фактично, є адаптивними системами управління – системами адаптації параметрів (САП) САР режимних параметрів ТОУ.

Застосування ЦД в промисловій автоматизації

Можна виділити всього п'ять застосувань ЦД в промисловій автоматизації, які і перераховані нижче.

1. Технічна діагностика ТОУ [4]. Це класичне і актуальне застосування для попереджувальнго обслуговування ТОУ.

2. Розрахунок оптимальних значень уставок режимних параметрів САР. Це все ж таки більш високорівнева функція інтегрованої АСУ масштабу підприємства; відповідно, реалізується в АСУ виробництвом (АСУВ), а не в АСУ ТП.

3. Компенсація виміряних сигнальних збурень розрахованими компенсаторами в прямому зв'язку комбінованих САР. Проте розраховані компенсатори (з виходом як на вхід, так і на вихід регулятору) є фізично нереалізованими, так як мають містити ланки ідеального або реального диференціювання.

4. Розрахунок невиміряного параметру з проміжної точки або невиміряного вихідного параметру каскадної САР. Це класичне і актуальне застосування ЦД в САР реального часу.

5. Розрахунок параметрів ТОУ і компенсація дрейфу параметрів ТОУ (добовий; сезонний; віковий) в САП. Адаптація САР під розраховані параметричні збурення є класичним і актуальним застосуванням ЦД в САР реального часу.

З п'яти вказаних застосувань ЦД в промисловій автоматизації реально актуальні тільки три застосування, а саме 1), 4) і 5). Застосування 2) і 3) – неактуальні.

Застосування СШІ в промисловій автоматизації

Можна виділити всього три застосування СШІ (інакше – ІСУ) в промисловій автоматизації, які і перераховані нижче.

1. ЕСУ і ЦД ТОУ в складі системи технічної діагнстики (СТД) ТОУ. Реалізується режим порадника з метою попереджувального обслуговування. Це класичний і актуальний приклад сумісного використання ЦД і СШІ в автоматизації.

2. В САП як системи екстремального управління (СЕУ). СЕУ реалізують алгоритми градієнтного пошуку екстремальних значень техніко-економічних параметрів (ТЕП; ключових показників ефективності КПЕ).

3. В САП як системи автотюнінгу (САТ) регуляторів САР. САТ реалізують алгоритми автоматичного пошуку оптимальних параметрів регуляторів пробними діями на ОУ в розімкненій САР.

З трьох вказаних застосувань СШІ в промисловій автоматизації реально актуальні усі три застосування.

Висновки щодо застосування ЦД/СШІ в промисловій автоматизації

Всього, як вказано вище, актуальними є три застосування ЦД і три застосування СШІ в промисловій автоматизації (всього разом шість). Застосування 1-СШІ є сумісним з застосуванням 1-ЦД. Застосування 2-3-СШІ замінюють застосування 5-ЦД, як більш ефективні, бо є активною адаптацією, а не пасивною (як застосування 5-ЦД).

Таким чином, можна виділити всього чотири застосування ЦД/СШІ в промисловій автоматизації, які і перераховані нижче.

1. Розрахунок невиміряного параметру з проміжної точки або невиміряного вихідного параметру каскадної САР. Це класичне і актуальне застосування ЦД в САР реального часу.

2. ЕСУ з ЦД ТОУ в складі системи технічної діагнстики (СТД) ТОУ. Реалізується режим порадника з метою попереджувального обслуговування. Це класичний і актуальний приклад сумісного використання ЦД і СШІ в автоматизації інженерних систем.

3. В САП як СЕУ. Реалізація СЕУ електропостачанням в сонячній електростанції (поворотом мобільної електростанції за сонцем) і у вітряній електростанції (позиціюванням кута атаки лопастей вітряку).

4. В САП як САТ регуляторів САР. САТ реалізують алгоритми автоматичного пошуку оптимальних параметрів регуляторів температури в приміщеннях пробними діями на ОУ в розімкнених САР.

Застосування ЦД/СШІ в автоматизації інженерних систем

В автономних системах електропостачання і життєзабезпечення [5] внутрішні збурення зі

сторони регулюючого органу (РО) застабілізовані в процесі підготовки інженерних систем до автоматизації, тому вони виключені. Це означає, що в автономних автоматизованих інженерних системах [6] використовуються переважно одноконтурні САР, а каскадні САР переважно не використовуються (основна перевага каскадних САР над одноконтурними – кращий перехідний процес при збуреннях зі сторони РО – нівелюється принциповою відсутністю збурень зі сторони РО).

Таким чином, в автоматизації автономних інженерних систем з перерахованих вище чотирьох застосувань ЦД/СШІ використовуються тільки три застосування, які перераховані нижче.

1. ЕСУ з ЦД ТОУ в складі системи технічної діагностики (СТД) ТОУ. Реалізується режим порадника з метою попереджувального обслуговування. Це класичний і актуальний приклад сумісного використання ЦД і СШІ в автоматизації інженерних систем.

2. В САП як СЕУ. Реалізація СЕУ електропостачанням в сонячній електростанції (поворотом мобільної електростанції за сонцем) і у вітряній електростанції (позиціюванням кута атаки лопастей вітряку).

3. В САП як САТ регуляторів САР. САТ реалізують алгоритми автоматичного пошуку оптимальних параметрів регуляторів температури в приміщеннях пробними діями на ОУ в розімкнених САР.

Всі ці вказані три застосування ЦД/СШІ є застосуваннями СШІ. Застосування ЦД для технічної діагностики інженерних систем входить в застосування СШІ в якості ЕСУ, що є яскравим прикладом сумісного застосування ЦД і СШІ в автоматизації інженерних систем.

Примітка. Стандартна класифікація адаптивних САР [1]

1. САР з жорсткою (малочутливі і робастні) і гнучкою адаптацією (активні і пасивні). Малочутлива (груба або робастна) САР є лінійною.

2. САР з активною адаптацією – СЕУ і САТ. Такі САР є лінійними.

3. САР з пасивною адаптацією – обчислення параметрів ОУ шляхом пасивного спостереження за входами і вихлдами ОУ. Якщо обчислення параметрів ОУ і коригування оптимальних параметів регулятора здійснюється в реальному чсі – така пасивна адптивна САР є нелінійною.

Приклади застосувань ЦД і СШІ в автоматизації теплоенергетичних процесів

ЦД і СШІ активно використовуються в класичній промисловій автоматизації. Різноманітність і нестандартність класичних рішень з ЦД і СШІ достатньо вражаючі. Нижче наведені приклади застосувань ЦД і СШІ в різних класичних задачах автоматизації теплоенергетичних процесів.

Автоматизація печей – класичний полігон промислової автоматизації. Автоматизація методичних печей, які прогрівають металеві сляби перед прокаткою на прокатних станах, – класичний полігон ЦД, які тоді (до появи терміну ЦД) називались «моделями реального часу» (МРЧ). ЦД (= МРЧ) використовувались також як в автоматизації енергоблоків теплових і атомних електростанцій, так і в автоматизації скловарних печей. СШІ використовувались в автоматизації теплоенергетичних процесів в якості СЕУ і САТ, які раніше (до появи термінів ШІ і ШІ-агент) називались «системами з навчанням (або з самонавчанням)».

Нижче в табл. 1 перераховані типові задачі нижнього (контролерного) рівня дворівневої (класична ієрархія автоматизації) АСУТП методичної печі. Виділені типові задачі регулювання теплового режиму промислової топки (регулювання температурного режиму, режиму економічності згоряння палива, аеродинамічного режиму), актуальні задачі регулювання продуктивності теплоенергетичного агрегатів і спеціальні задачі регулювання перегріву пари.

Задачі контролерного рівня автоматизації є типовими для теплоенергетичних ТОУ (промислові печі; енергетичні котли; інженерні системи малої енергетики), які містять топки з пальниками для спалення палива і системи нагрівання робочих тіл (води, пари, повітря, поверхонь тощо). Ці задачі реалізуються типовими САР – одноконтурними (більшість) і двоконтурними (здебільшого – каскаднми з двома регуляторами, значно рідше – з одним двоімпульсним регулятором). ПІД-регулятотри в САР розрховуються, як правило, на аперіодичний або з 20%-им перерегулюванням перехідні процеси. Використання таких відносно «повільних» регуляторів (відносно малий коефіцієнт передачі і відносно велика стала часу інтегрування) синтезує САР з високим запасом сталості. Такі САР є малочутливими до параметричних збурень ОУ. Інакше кажучи, вони є САР з жорсткою адаптацією (тобто заздалегідь адаптовані під параметричні збурення ОУ своєю функціональною малочутливістю до цих збурень).

Нижче в табл. 2 перераховані типові задачі верхнього (супервізорного) рівня дворівневої

Таблиця 1

Типові задачі регулювання нижнього (контролерного) рівня теплоенергетичної АСУТП

№ п/п	Функція	Задача	САР режимного параметру	Тип САР
1	2	3	4	5
1	Регулювання температурного режиму	1.1. Стабілізація температури газового середовища вздовж прохідної топки (позонна стабілізація температури). 1.2. Програмне регулювання (в часі) температури газового середовища в камерній топці	САР температури газового середовища (димових газів)	Одноконтурна САР «температура – витрата палива». Каскадна САР «температура – задана витрата палива – переміщення РО»
2	Регулювання економічності згоряння палива	2.1. Стабілізація витрати повітря на значенні $\alpha^* \cdot L_0 \cdot B_n$, де B_n – поточна витрата палива, α^* – заданий коефіцієнт надлишку (витрати) повітря, L_0 – стехіометричне співвідношення "повітря/паливо" (для газу – ~ 10 м³/м³). 2.2. Стабілізація вмісту кисню в димових газах	2.1. САР співвідношення повітря/паливо (= економічності згоряння палива = коефіцієнту надлишку повітря). 2.2. САР вмісту кисню в димових газах	2.1. Одноконтурна САР «витрата палива – витрата повітря». 2.2. Каскадна САР «вміст кисню в димових газах – заданий тиск повітря – переміщення РО».
3	Регулювання аеродинамічного (гідравлічного) режиму топки	Стабілізація розрідження в топці	САР розрідження в топці	Одноконтурна САР. Комбінована САР (з компенсацією збурення витратою повітря)
4	Регулювання продуктивності	4.1. Стабілізація рівня рідини в баці (барабані котла). 4.2. Стабілізація швидкості обертання ротору турбіни. 4.3. Стабілізація тиску в колекторі	4.1. САР рівня. 4.2. САР частоти обертів ротору турбіни 4.3. САР тиску в колекторі	4.1. Одноконтурна САР. САР з триімпульсним регулятором. 4.2. Каскадна САР «частота змінного струму – електрична потужність – частота обертів ротору турбіни" 4.3. Одноконтурна САР. Система зв'язаного регулювання
5	Регулювання перегріву пари	5.1. Стабілізація температури первинної пари. 5.2. Стабілізація температури вторинної пари	5.1. САР температури первинної пари. 5.2. САР температури вторинної пари	Двоімпульсна САР (з корекцією за температурою з проміжної точки пароперегрівача)

(класична ієрархія автоматизації) АСУТП методичної печі. Задачі супервізорного рівня автоматизації включають 1) задачі SCADA-автоматизації (задачі HMI-Standard і MES-Lite) і 2) задачі моделювання і оптимізації реального часу (які, як вказано вище, і є ЕСУ з ЦД).

Типова HMI/SCADA-функціональність:

- Обмін даними з контролерами.
- Передача уставок в контролери.
- Архівування даних від контролерів.
- Візуалізація даних мнемосхем (mimics).
- Ідентифікація і відображення алармів (нештатних ситуацій).
- Побудова і відображення трендів (реального часу та історичних).
- Реалізація систем звітності (локальної і консолідованої аналітичної).

Сучасна SCADA-система не реалізує оптимізацію технологічного процесу в реальному часі, вона в реальному часі реалізує людино-машинний інтерфейс HMI (мнемосхеми, аларми, тренди тощо). Задачі моделювання і оптимізації реаль-

ного часу (тобто, функціональність ЕСУ і ЦД) є додатковими до HMI/SCADA-функціональності і реалізуються спеціальними програмними продуктами, як правило, пропрієтарними.

Реалізація промислової АСУТП «великих» ТОУ без онлайнових задач моделювання і оптимізації додатково до стандартної SCADA-функціональності є практично неможливою, відповідно, обов'язковою.

Вказані в табл. 2 супервізорні функції управління з використанням ЕСУ і ЦД, хоча і сформульовані для АСУТП методичної печі, є типовими для теплоенергетичних ТОУ, так як включають типові генеалогію, інформаційне супроводження і модельні розрахунки в реальному часі невимірюваних параметрів на основі вимірюваних параметрів.

Приклад № 1. ЦД в задачах верхнього (супервізорного) рівня автоматизації методичних печей прокатних станів – розрахунок в реальному часі температурних епюр металевих слябів

Нижче на рис. 1 наведені діаграми режимів нагрівання слябів в методичній печі.

Виділяють режим нагрівання слябів (з різних партій, які одночасно знаходяться в печі) за балансовим методом і за середньо-інтегральною температурою слябу (по його висоті).

Нижче на рис. 2 наведений графік гарантованого нагрівання слябів в методичній печі. Гарантований режим нагрівання слябів є компромісним режимом нагрівання слябів в умовах неконтрольованого змінного темпу роботи прокатного стану (темпу прокатки).

Таблиця 2

Типові задачі управління верхнього (суперізорного) рівня АСУТП методичної печі

№ п/п	Функція / задача	Система автоматичного управління (САУ) / комплекс задач
1	Інформаційне супроводження слябів в методичних печах	САУ транспортуванням слябів
2	Прогноз темпу прокатки прокатного стану (тривалості нагріву слябів в печах)	Комплекс задач прогнозування темпу прокатки прокатного стану
3	Цифровий двійник топки печі (моделювання внутрішнього і зовнішнього теплообмінів топки в реальному часі і розрахунок додаткових технологічних параметрів, які неможливо виміряти сенсорами – температури печі (?), температури поверхні кожного слябу, температури центру кожного слябу, середньомасової (середньоінтегральної) температури по висоті кожного слябу)	Комплекс задач розрахунку теплообміну
4	Реалізация гарантованого режиму нагріву слябів з фіксуванням моменту виходу в режим витомлення	САУ нагріванням слябів
5	Реалізація режиму нагріву слябів за середньомасовою (середньоінтегральною) температурою слябів	САУ нагріванням слябів
6	Реалізація режиму нагріву слябів за балансовим методом	САУ нагріванням слябів
7	Діагностування стану автоматизованого технологічного комплексу, розрахунок техніко-економічних показників (КПЕ/KPI)	Комплекс задач діагностики і розрахунку техніко-економічних показників (КПЕ/KPI)

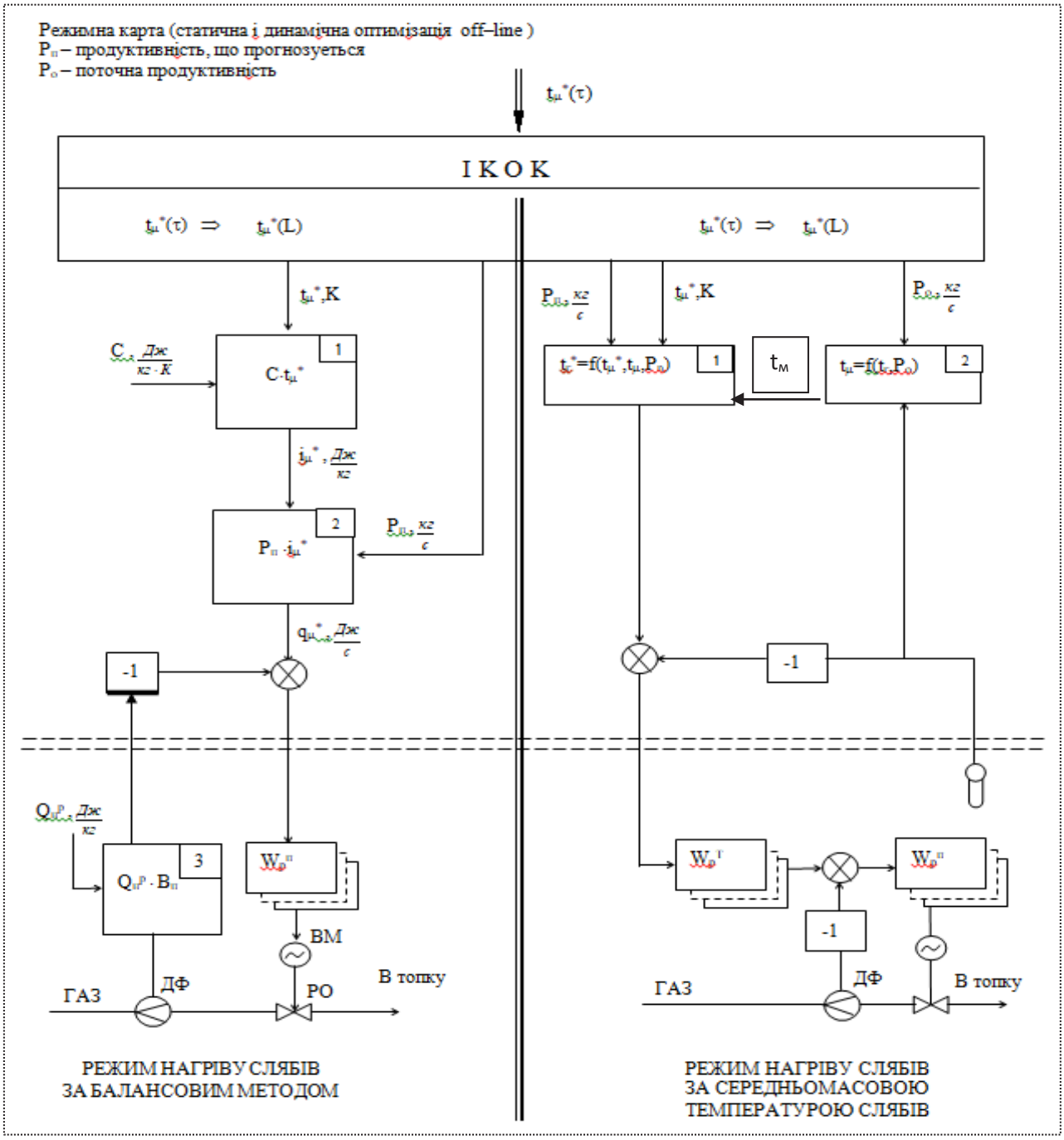
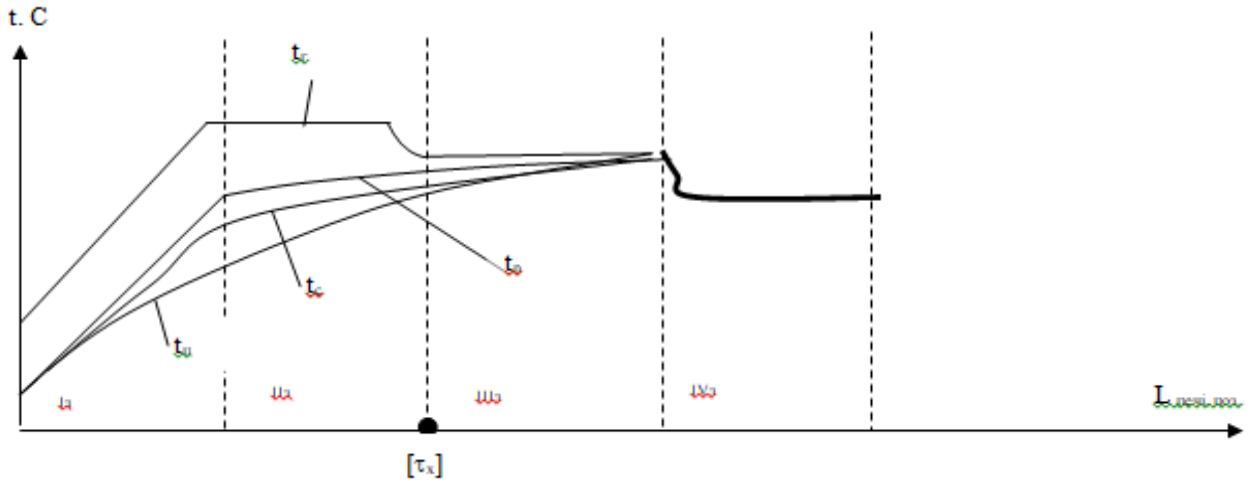


Рис. 1. Діаграми нагрівання слябів в методичній печі

Загальним збуренням для АСУТП методичної печі є зміна темпу прокатки металевих слябів на прокатному стані. Це є класичне сигнальне зовнішнє збурення для ТОУ – збурення за навантаженням. Головна задача методичної печі – прогріти великий металевий сляб (розміром десь 2 м * 2 м * 2 м) до температури приблизно 723 К по ширині, довжині і висоті. В ідеальному випадку – температури поверхонь слябів і цетру слябу є однаковими. Така рівномірна внутрішня температурна епюра слябу забезпечує його легку, штатну і безаварійну прокатку валками прокатного стану (стискання слябу висотою 2 метри в профілі або в листи товщиною 2 мм). Від темпу прокатки залежить час нагрівння (перебування) слябів в методичній печі – звідси мають бути різні швидкості нагрівання, витримки і охолодження слябів на позиціях зон методичної печі. Ситуація ускладнюється тим фактом, що в методичній печі можуть одночасно нагріватись (перебувати) сляби з різних партій. Партія слябів – це сляби однакового



I_3 – зона швидкого нагріву при q=const
II_3 – зона витрати при $t_г^*$=const
III_3 – зона витомлення при $t_г^{**}$=const, $t_г^{**} \le t_г^*$
IV_3 – зона охолодження при $t_г^{***}$=const, $t_г^{***} \le t_г^{**}$

$t_г$ – температура газового середовища
$t_п$ – температура поверхні слябу
$t_ц$ – температура центру слябу
t_c – середньомассова температура слябу
τ_x – момент фіксації виходу слябів в режимі витомлення (момент τ_x кореспондується з відповідної позицією нагріву в печі. Вірогідність правильної фіксації

$$P=\prod_{i=1}^{n}(1-p_i),$$

де pi – вірогідність правильної фіксації моменту виходу в режим витомлення за окремою ознакою,
n – кількість вибраних ознак, при цьому P>min{Pi}).

Рис. 2. Графік гарантованого нагрівання слябів в методичній печі

типорозміру (однакових теплофізичних характеристик), відлиті з металу в ливарних печах за однаковою технологією. Сляби різних партій, як сляби різних типорозмірів, потребують різних режимів нагрівання слябів в методичних печах. Режим нагрівання слябів в печі – це температурний режим печі, тобто температурний графік по довжині печі (тобто в часі).

В загальному випадку важливим є температурний графік як графік температури центру слябу, яка в кінці процесу нагрівання має вирівнятись з температурою поверхні сляба і ці температури мають дорівнювати заданій температурі прокатки – ідеальний варіант. На практиці, температура центру слябу ніколи не досягає температури поверхні слябу, тому кажуть про середньо-інтегральну температуру слябу (по перерізу слябу). Є дві проблеми управління нагріванням слябів. Перша – температуру центру слябу фізично виміряти неможливо (можливе тільки періодичне вимірювання температури поверхні слябу пірометром у фурмі). Рішення проблеми – замість температурного графіку центру слябу (графіку середньо-інтегральної температури слябу) використовувати графік температури димових газів по довжині печі. Але виникає друга проблема – температури димових газів вимірюються термопарами, але що насправді вимірює термопара? В теплотехніці замість температури димових газів використовується термін «температура печі»о – термопара на позиції в печі вимірює як температуру самих димових газів, так і факелу. і слябу, і кладки печі. Рішення проблеми – на основі виміряних «температур печі» в зонах печі розрахувати як температури дивових газів, так і середньо-інтегральні температури слябів. Виміряні температури печі (або розраховані температури димових газів) є параметрами з проміжної точки для ОУ як каналу передачі регулювальної дії «зміна витрати палива – зміна температури печі – зміна середньо-інтегральної температури слябу». Відпо-

відно, САР температури в табл. 1 (задачі контролерні автоматизації) можуть бути побудовані або як одноконтурні САР температур печі (які вимірюються), або як одноконтурні САР температур димових газів (які розраховуються), або як одноконтурні САР середньо-інтегральних температур слябів (які розраховуються), або як каскадні САР середньо-інтегральних температур слябів (головні праметри – середньо-інтегральні температури слябів – розраховуються для кожного слябу; проміжні параметри – це або виміряні термопарами температури печі, або розраховані температури димових газів).

Таким чином, для оптимального вирішення задач контролерного рівня автоматизації (табл. 1) в задачах супервізорної автоматизації (табл. 2) виникає комплексна задача онлайнового (тобто, в режимі реального часу) моделювання процесів тепломасообміну в печі, яка наразі вирішує дві задачі: 1) на основі виміряних температур печі розраховуються невимірювані температури поверхні і центру слябу (середньо-інтегральної температури слябу); 2) на основі розрахованих середньо-інтегральних температур слябів розраховуються задані значення температур димових газів (температур печі) в зонах печі або задані значення витрат палива в зонах печі такі, які забезпечують своєю стабілізацієюстабілізацію заданого графіку середньо-інтегральної температури слябів.

Відповідно, можуть бути реалізовані три режими нагрівання слібів в методичній печі (рисунки 1 і 2):

1. режим нагрівання за середньо-інтегральною температурою слябів (справа на рис. 1): 1) онлайн прогноз темпу прокатки слябів; 2) онлайн розрахунок заданого графіку середньо-інтегральної температури слябів по довжині печі на основі темпу прокатки; 3) одноконтурна або каскадна стабілізація зміною зональних витрат палива розрахованих в реальному часі середньо-інтегральних температур слябів в зонах печі (задані і поточні середньо-інтегральні температури різних слябв в одній зоні печі усереднюються) температури;

2. режим нагрівання за балансовим методом (зліва на рис. 1): 1) онлайн прогноз темпу прокатки слябів; 2) офлайн розрахунок заданого графіку зональних витрат палива по довжині печі на основі теплового балансу печі з урахуванням темпу прокатки; 3) одноконтурна або каскадна стабілізація виміряних зональних витрат палива переміщенням РО на зональних газопроводах;

3. гарантований режим нагрівання (рис. 2): 1) офлайн розрахунок заданого графіку зональних температур печі по довжині печі, якийй гарантує нагрівання і прогрівання слябів за найшвидшого темпу прокатки, тобто за найменший час нагрівання (пербування) слябів в печі (режим не оптимальний – сляб спочатку швидко нагрівається, а потім довго витомлюється (прогрівається), за рахунок чого є перевитрата палива); 2) одноконтурна або каскадна стабілізація виміряних зональних температур печі зміною зональних витрат палива (газу).

Основним режимом нагрівання слябів в методичній печі є режим нагрівання за середньо-інтегральною температурою слябів. Режим нагрівання за балансовим методом використовується як альтернативний, якщо темп прокатки змінюється епізлдично і заздалегідь відомо. Гарантований режим нагрівання слябів використовується у випадках неможливості прогнозу темпу прокатки слябів або неможливості розрахунку поточних середньо-інгтегральних температур слябів в печі.

Для реалізації вказаних режимів нагрівання слябів в методичній печі потрібно знати генеалогію слябів (з якої партії сляби – тобто з якими типорозмірами, тобто теплофізичними характеристиками), потрібно знати місцезнаходження кожного слябу від ливарної печі до подачі на прокатний стан після методичної печі (інформаційне супроводження слябів), потрібно знати історію нагрівання слябів (архівування виміряних і розрахованих температур). Вказані задачі, які є задачами ЦД, реалізуються САУ в складі супервізорного рівня АСУТП методичної печі.

ЦД (онлайнові моделі, тобто моделі в реальному часі, теплообміну в промислових топках) в АСУТП методичної печі використовуються як в автоматичному режимі, так і в складі ЕСУ в режимі порадника. Саме АСУТП методичної печі є класичним і найбільш ефективним прикладом використання ЦД в промислові автоматизації.

Приклад № 2. СШІ в задачах нижнього (контролерного) рівня автоматизації промислових топок – градієнтний пошук оптимального значення коефіцієнту надлишку повітря

Класична задача контролерної автоматизації – СЕУ коефіцієнтом надлишку повітря в топці печі. Газ, який подається в пальники печі, в залежності від своїх температури, тиску і хімічного складу має різну теплотворну здатність. Крім того, якість спалення газу (кількість

виділеного при спаленні газу тепла) залежить від конструкції пальників, температури і вологості повітря, яке подається в пальники. З цих причин заздалегідь визначити і оперативно (в реальному часі) розрахувати оптимальне значення коефіцієнту надлишку повітря практично неможливо. Значення коефіцієнту надлишку повітря, яке забезпечує заданий виіст кисню в димових газах (тобто, потрібну якість згоряння палива), може бути знайдено активним експериментом – екстремальним управлінням. Пробна дія – зміна коефіцієнту надлишку повітря (для постійного стехіометричного співвідношення повітря/паливо – фактично, зміна витрати повітря). Регульований параметр – вміст кисню в димових газах (вимірюється оксиметром). Відсутність хибних локальних екстремумів.

Приклад № 3. ЦД в задачах верхнього (супервізорного) рівня автоматизації енергоблоку елетростанції – технічна діагностика тепло-механічного стану електрогенератору [7]

Класична задача автоматизації – технічна діагностика тепломеханічного стану електрогенератору енергоблоку. ЦД – тепломеханічна модель електрогенератору, керована виміряними температурами різних елементів електрогенератору (всього 128 параметрів). ЦД розраховує механічні навантаження електрогенератору і прогнозує його ресурс роботи.

Приклад № 4. ЦД в задачах нижнього (контролерного) рівня автоматизації енергоблоку елетростанції – фільтрація виміряних температур корпусу електрогенератору [7]

Класична задача автоматизації – фільтрація вимірюваних контрольних параметрів. Некласичний фільтр – реалізується двоступенева фільтрація. ЦД – модель двоступеневого фільтру з нелінійним видаленням граничних амплітуд викидів і лінійним фільтром ковзаючого середнього. Виміряні температури імпортуються в реальному часі в модель фільтру. Модель розраховує нові параметри реального двоступеневого фільтру.

Приклад № 5. ЦД в задачах верхнього (супервізорного) рівня автоматизації скловарної печі – розрахунок температури і рівня скломаси в басейні печі

Класична задача автоматизації – регулювання температури скломаси і рівня скломаси в басейні скловарної печі. Температуру скломаси в басейні виміряти фізично неможливо, а рівень – проблематично. ЦД – модель тепломасообміну в скловарній печі. ЦД розраховує температуру скломаси в басейні і рівень скломаси в басейні. Реалізується каскадна САР температури скломаси в басейні зі стабілізуючим регулятором температури печі (димових газів). Реалізується каскадна САР рівня скломаси зі стабілізуючим регулятором витрати шихти (регулятором рівня шихти в бункері шихти).

Приклад № 6. ЦД в задачах нижнього (контролерного) рівня автоматизації скловарної печі – формування команди на перекидання факелу в регенеративній топці

Класична задача автоматизації – реалізація перекиданння факелу в регенеративній скловарній печі. Реалізується СПЛУ перекиданням факелу. Нестандартною є задача формування команди на перекидання факелу (після чого послідовність перекидання факелу реалізується відповідною СПЛУ).

Формування команди на перекидання факелу виконується за такими ознаками:

1) за максимальним інтервалом часу між перекиданнями:

$$\tau \geq \tau^*,$$

де τ^* – заданий – максимальний інтервал часу між перекиданнями;

2) за максимальною температурою верхніх частин регенераторів, які гріються (через які під час поточного інтервалу між перекиданнями проходять димові гази):

$$t_{вр} \geq t^*_{вр},$$

де $t^*_{вр}$ – задана максимальна температура;

$t_{вр}$ – середня поточна температура верхніх частин регенераторів;

3) за максимальним теплом, яке передається димовими газами до верхніх частин регенераторів, які гріються:

$$\int_0^{\tau} [t_{вр}(\tau) - t_o] d\tau \geq q^*,$$

де q^* – задане максимальне значення інтегралу;

$t_{вр}$ – середня поточна температура регенераторів після закінчення перекидання (див. п. 2);

t_0 – середня початкова температура регенераторів після закінчення перекидання;

4) за мінімальною температурою нижніх частин регенераторів, які охолоджуються (через які під час поточного інтервалу між перекиданнями проходить повітря):

$$t_{нр} < t^*_{нр},$$

де $t^*_{нр}$ – задана мінімальна температура;

$t_{нр}$ – середня поточна температура нижніх частин регенераторів;

5) за максимальним теплом, яке відбирається повітрям від нижніх частин регенераторів, які охолоджуються:

$$\int_0^{\tau} [t_o - t_{np}(\tau)]d\tau \geq q^*,$$

де q^* – задане максимальне значення інтегралу;

t_{np} – середня поточна температура регенераторів після закінчення перекидання (див. п. 4);

t_0 – середня початкова температура регенераторів після закінчення перекидання.

Вимірювання температури насадок рекуператорів є складною технічною задачею. Одноточкові вимірювання температур насадок не є репрезентативними. Рішення проблеми – розрахувати температурні епюри регенераторів або кількість підведеного до холодного регенератора чи відведеного від гарячого регенератора тепла.

ЦД – модель теплообміну системи регенераторів скловарної печі. На основі виміряних температур печі і одноточкових вимірювать температури насадок регенераторів розраховуються температурні епюри регенераторів, кідькість підведеного чи відведеного телпа і формується команда на перекидання факелу.

Нижче на рис. 3 наведений графік перекидання факелу в регенеративній скловарній печі.

Перекидання факелу зліва направо (вихідна позиція – факел зліва; кінцева позиція – факел зправа) реалізуєьбся в послідовності: 1) зупин подачі палива зліва; 2) вентилювання печі повітрям зліва; 3) зупин подачі повітря зліва; 4) перенаправлення засувками потоків енергоносіїв зліва направо; 5) подача повітря зправа і вентилювання печі; 6) подача палива зправа.

Аналогічно реалізується перекидання факелу зправа наліво (вихідна позиція – факел зправа; кінцева позиція – факел зліва).

Наведена циклограма є типовою і багаторазово реалізована як для скловарних печей, так і для інших регенеративних печей.

Графік перекидання факелу реалізується типовою СПЛУ перекиданням факелу. Таким чином, ЦД системи регенераторів печі (модель реального часу теплообміну в системі регенераторів печі) фактично є складовою частиною типової СПЛУ, навіть не САР!

Приклад № 7. СШІ в задачах верхнього (супервізорного) рівня автоматизації лінії фор-

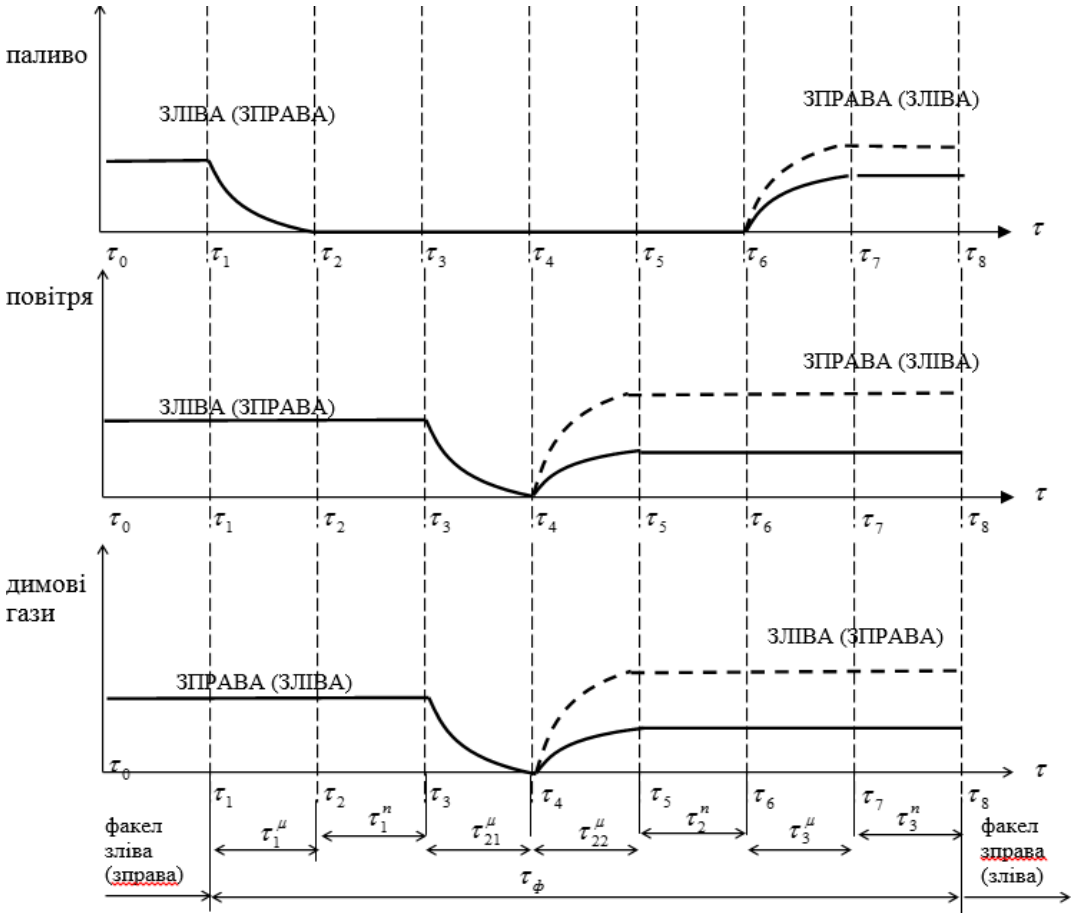


Рис. 3. Циклограма перекидання факелу в топці скловарної печі

мування скляних труб методом горизонтального витягування – градієнтний пошук мінімуму собівартості виробів [8]

Класична задача автоматизації – екстремальне управління продуктивністю технологічного агрегату. Ліня горизонтального витягування скляних труб (лінія ГВТ) – технологічна лінія для виробництва ламп денного світла. Нетривіальним і характерним є той факт, що екстремізація різних техніко-економічних показників роботи (в сучасних термінах – ключових показників ефективності КПЕ – KPI – Key Production Indicator) лінії ГВТ досягається для різних значень продуктивності лінії ГВТ. Розташування екстремумів КПЕ на осі продутивності лінії ГВТ (загальної кількості вироблених скляних труб для ламп денного світла) наступне (в напрямку зростання оптимального значення продуктивності): максимум коефіцієнту виходу придатних виробів – мінімум собівартості виробів – масимум прибутку від реалізації придатних вирбів – максимум кількості придатних виробів. СЕУ продуктивністю лінії ГВП здійснює градієнтний пошук глобального екстремуму заданого КПЕ (хибні локальні екстремуми відсутні). Задача екстермального управління продуктивністю лінії ГВП може бути ускладнена – наприклад, реалізована як градієнтний пошук мінімума собівартості виробів при обмеженні знизу на кількість виробленої продукції (обмеження вводиться, так як мінімум собівартості вироблених скляних труб досягається при відносно малих значеннях кількості придатних вироблених скляних труб, яких тоді просто не вистачатиме для виробництва кінцевої продукції – ламп денного світла).

Приклад № 8. ЦД в задачах нижнього (контролерного) рівня автоматизації лінії формування скляних труб методом горизонтального витягування – формування уставки регулятору аеродинамічного режиму формування склотрубок [8]

Використання теорії подібності (пі-теорема Бекінгема; степеневе рівняння розмірностей; критерії інваріантності) для аналізу реологічних моделей – класична (хоча непоширена і нетривіальна) задача автоматизації. Лінія ГВТ (горизонтального витягування скляних труб для ламп денного світла) – класичний приклад застосування реологічних моделей (при математичному моделювання процесу витягування) і, віідповідно, елементів теорії подібності. Типовим є наступний приклад застосування критеріїв інваріантності (подібності). ЦД – реологічна модель витягування склотрубки. На основі виміряних поточних діаметрів склотрубок обчислюється дисперсія відхилень діаметрів від заданого значення і через реологічну модель на основі критеріїв подібності розраховується уставка регулятору аеродинамічного режиму формування (сума з ваговими коефцієнтами тиску повітря в «цибулині» скла на мундштуку робочої камери і витрати повітря на роздування «цибулини»). В загальному випадку критерії подібності можуть бути використані (так як вони завжди однакові для реального фізичного процесу і його моделі) для обчислення невимірюваних параметрів (які входять в критерії подібності) на основі вимірюваних параметрів (які входять в ті ж самі критерії подібності).

Приклад № 9. СШІ в задачах нижнього (контролерного) рівня автоматизації теплоенергетичних процесів – автотюнінг ПІД-регуляторів

Класична задача автоматизації – адаптація САР під параметричні збурення ОУ. Адаптивна САР – це САР з адаптивним ПІД-регулятором. Якщо регулятор розраховується як малочутливий до параметричних збурень – така САР є САР з жорсткою (наперед сформованою за рахунок відносно «повільного» регулятору) адаптацією (тобто САР з наперед сформованим великим запасом сталості по модулю і фазі). Гнучка адаптація ПІД-регулятору до параметричних збурень – це САР з пасивною адаптацією [9] (обчислення параметрів ОУ шляхом пасивного спостереження за входами і виходами ОУ і періодичне коригування оптимальних параметрів регулятору) чи активною адаптацією (обчислення параметрів ОУ періодично реалізується шляхом активного експерименту, на основі якого коригуються оптимальні параметри регулятору) під параметричні збурення ОУ. Недоліки пасивної адаптації: 1) неможливість врахувати всі параметричні збурення ОУ; 2) нелінійсть САР (якщо обчислення параметрів ОУ і коригування оптимальних параметів регулятору здійснюється в реальному часі). Переваги активної адаптації: 1) активний експеримент сигнальним збуренням «автоматично» враховує всі параметричні збурення ОУ; 2) лінійність САР, що забезпечує відтворюваність результатів розрахунку і моделювання для усіх сигнальних збурень САР і параметричних збурень ОУ. Активний екпримент для лінійної САР – це отримання перехідної характеристики ОУ. Для апроксимації динаміки ОУ можуть бути використані інваріантні моделі об'єктів управління і універсальні структури ПІД-регуляторів [10]. Активна адап-

тація САР з ПІД-регуляторами відома як автотюнінг ПІД-регуляторів. Реалізована в багатьох ПЛК на рівні бібліотечних функціональних блоків від виробника ПЛК. Можлива реалізація користувачем власного функціонального макроблоку (на основі бібліотечних функціональних блоків від виробника) автотюнінгу. Автотюнінг ПІД-регуляторів – класичний і ефективний спосіб активної адаптації САР під параметричні збурення (і практично безальтернативний).

Приклад № 10. СШІ в задачах комплексної автоматизації інженерних систем – екстремальне управління позиціюванням панелей батарей сонячних електростанцій і лопатей турбін вітряних електростанцій

Сучасна задача автоматизації системи автономного електропостачання житлового будинку – екстремальне управління продуктивністю сонячної електростанції і вітряної електростанції.

СЕУ продуктивністю сонячної електростанції реалізує градієнтний пошук масимуму сонячного споживання позиціюванням кута повороту панелей сонячних батарей за сонцем.

СЕУ продуктивністю вітряної електростанції реалізує градієнтний пошук максимуму споживання сили вітру позиціюванням кута атаки лопатей вітряної турбіни вітром. Особливість екстремального управління сонячною батареєю і вітряною турбіною – наявність хибних локальних максимумів сонячної і вітряної активностей.

Пошук глобальних максимумів сонячного і вітряного енергопостачання потребує реалізації в СЕУ спеціальних алгоритмів ігнорування локальних максимумів, що суттєво ускладнює практичну реалізацію вказаних СЕУ Тим не менше, СЕУ можуть бути реалізовані і реалізуються стандартними (штатними) програмно-технічними засобами – в сучасних ПЛК.

СЕУ продуктивністю сонячних і вітряних електростанцій є актуальними. Можуть функціонувати в безлюдному режимі 214/7 впродовж всього терміну функціонування самих цих електростанцій.

Примітка. Моделювання теплоенергетичних СЕУ і САТ

СЕУ (градієнтний пошук екстремумів регульованих параметрів) і САТ (автотюнінг ПІД-регуляторів) можуть бути ефективно змодельовані методами функціонального і структурного імітаційного моделювання [11].

Висновки. 1. Використання ЦД і СШІ в автоматизації теплоенергетичних процесів є традиційними і класичними способами підвищення якості автоматизованого і автоматичного управління ТОУ. Широко використовувались і використовуються в АСУТП теплоенергетичними агрегатами.

2. ЦД в АСУТП теплоенергетичних агрегатів були відомі як «моделі реального часу», бо вони використовувались для розрахунків в реальному часі невимірюваних параметрів ОУ на основі вимірюваних параметрів ОУ. ВАЖЛИВО! Класичні «моделі реального часу» – історично перші ЦД, а не навпаки.

3. СШІ в АСУТП теплоенергетичних агрегатів були відомі як СЕУ (бо використовувались для градієнтного пошуку глобальних екстремумів регульованих параметрів) і САТ (бо використовувались для автотюнінгу класичних ПІД-регуляторів активними експериментами з сигнальним збуренням регулюючою дією). ВАЖЛИВО! Класичні СЕУ і САТ – історично перші СШІ, а не навпаки. Класичні екстремальне управління і автотюнінг – історично перші алгоритми навчання СШІ, а не навпаки.

4. Порівняння генеративного ШІ (ГШІ) і автотюнінгового СШІ-ПІД-регулятору

1. ГШІ і СШІ-ПІД мають алгоритми роботи. В ГШІ – це LLM-алгоритм генерації тексту. В СШІ-ПІД – це ПІД-закон регулювання.

2. ГШІ і СШІ-ПІД мають вбудовані алгоритми навчання. В ГШІ – це якийсь алгоритм з підкріпленням для навчання LLM-алгоритму (= генерації тексту-відповіді на промпт). В СШІ-ПІД – це алгоритм автотюнінгу (= пошуку оптимальних параметрів ПІД-регулятору під параметричні збурення від ОУ). ВАЖЛИВО! Тут ми чітко бачимо, що ознакою СШІ є не алгоритм роботи (LLM в ГШІ і ПІД в СШІ-ПІД), а саме алгоритм навчання (підкріплення в ГШІ і автотюнінг в СШІ-ПІД).

3. Аналогом промпту для ГШІ+Людина від людини є сигнальне збурення в САР з СШІ-ПІД від ОУ. ГШІ навчено реагує на промпт (тобто, ефективно реалізує LLM-генерацію) і генерує відповідь. СШІ-ПІД має навчені параметри ПІД-регулятору (тобто, нові, знайдені пошуком автотюнінгу, оптимальні ПІД-параметри під нові ОУ-параметри) і реалізує стабілізацію регульованого параметру з заданими показниками якості перехідного процесу.

4. ГШІ поки що система з навчанням. СШІ-ПІД з автотюнінгом – це система з самонавчанням. Тобто, зараз автотюнінговий СШІ-ПІД – більш потужний СШІ, ніж ГШІ (не з точки зору алгоритмічної складності і обчислювальної

потужності, а принципово, з точки зору самого визначення СШІ). Але ця перевага тимчасова, тому що ГШІ з самонавчанням вже якщо не працюють, то стукають у двері (а в роботах це взагалі буде основний режим роботи).

5. Принципової технічної різниці між ГШІ і СШІ-ПІД немає, бо вони обидва мають класифіційну рису ШІ – навчання. Між ГШІ і СШІ-ПІД є прицпові технологічна і соціальна різниці. Технологічна різниця – для реалізації ГШІ потрібна принципово більша обчислювальна потужність (СШІ-ПІД реалізується стандартними обчислювачами – промисловими ПЛК); тому генеративний інтелект точніше називати обчислювальним інтелектом (Computational Intelligence). Соціальна різниця – соціальні наслідки впровадження ГШІ в соціум є непередбачуваними (використання СШІ-ПІД є повністю контрольованим).

5. Порівняння перспектив використання ЦД і СШІ в автоматизації теплоенергетичних процесів

1. Питання пасивного спостереження чи активного експерименту – це водорозділ між просто теорією управління як «про управління взагалі» і ТАУ як практичною дисципліною. Активний експеримент – це засадниче питання для ТАУ. В ТАУ нас цікавить не фізика ОУ (наприклад, як впливають теплоємності на тепломасообмін), а динаміка і саме (це дуже важливо!) динаміка через канал регулювальної дії (тобто, як наша регулювальна дія впливає на регульований параметр). Якщо у нас немає регулювальної дії або динаміка її погана, то все інше нас не цікавить і потрібно або взагалі відмовитись від регулювання, або шукати іншу регулювальну дію (і це буде вже інша динаміка, хоча, можливо, та сама фізика). Саме тому в ТАУ ідентифікація ОУ розуміється не як математичне моделювання (пасивне спостереження «вхід-вихід» чорної скрині або складання диференціальних рівнянь), а як активний експеримент (зняття кривої розгону черех регулювальний канал з перебудовою її в перехідну характеристику). Так само і щодо пасивної і активної адаптації. Активна адптація краща за пасивну. Тому, наприклад, автотюнінговий СШІ-ПІД-регулятор кращий за пасивну адаптивну САР (з ЦД ОУ, який пасивно розраховує нові параметри ОУ на основі виміряних технологічних параметрів).

2. ЦД ефективний не як ЦД ОУ (тобто, не з точки зору управління замкненого чи розімкненого), а як ЦД ТОУ (тобто, з точки зору технології і стану самого технологічного агрегату). Тому роль ЦД в реальній автоматизації обмежена. ЦД не має сенсу використовувати в оперативному управлінні. ЦД – це, передусім, нативна для нього технічна діагностика. ЦД ефективні в різного роду ЕСУ в режимах порадника (попереджувальне обслуговування). Є тільки один виняток для ЦД – це розрахунок невиміряного параметру з проміжної точки або невиміряного вихідного параметру каскадної САР. Приклади: розрахунок температури поверхні і центра слябу в методичній печі; розрахунок вмісту вуглецю в розплавленому металі в конверторах; розрпхунок температури розплавленої скломаси; тощо. Чим краще офлайн спроєктований і відналагоджений технологічний процес або агрегат, тим менша потреба в ЦД (бо в ОУ «все схоплено», тобто все виміряно, а всі невимірювані сигнальні збурення «застабілізовані» і не «проходять» на ОУ) і тим більша потреба в СШІ (бо ПІД-регулятори мають періодично в реальному часі адаптуватись під параметричні збурення самого ОУ, які неможливо «застабілізувати»).

Список літератури:

1. Попович М.Г., Ковальчук О.В. Теорія автоматичного керування : підручник. Київ : Либідь, 2007. 656 с.
2. Баранов С.А., Ковальов Д.О. Автоматизовані системи управління: принципи побудови та приклади реалізації. Київ : Наукова думка, 2021. 312 с.
3. Батюк С.Г., Ковриго Ю.М., Бунь В.П. Порівняльний аналіз класичного і некласичного підходів до створення АСУ ТП в енергетиці на основі сучасних ПТК. Наукові вісті НТУУ КПІ. 1999. № 2. С. 40–48.
4. Пастухов А.В., Литвиненко В.С. Системи технічної діагностики в енергетиці. Дніпро : НГУ, 2021. 245 с.
5. Ткаченко П.М., Іванов Ю.В. Енергетичні системи з відновлюваними джерелами енергії. Харків : ХНУРЕ, 2020. 278 с.
6. Ковтун В.І. Системи автоматизації житлових будинків: сучасні рішення та перспективи розвитку. Львів : ЛНУ, 2022. 198 с.
7. Батюк С.Г., Олійник М.Ю. Методика оптимальної фільтрації даних температурного контролю турбогенераторів в умовах значних промислових перешкод. Автоматика. Автоматизация. Электротехнические комплексы и системы. 2009. № 1 (23). С. 147–151.

8. Батюк С.Г. Синтез технічно оптимальної системи управління формуванням скляних труб методом горизонтального витягування: дисертація на здобуття наукового ступеня кандидата технічних наук : спеціальність 05.13.07 «Автоматизація технологічних процесів і виробництв в склоробній промисловості». Київ : Київський інститут автоматики, 1991. 278 с.

9. Novikov, P., Shtifzon, O., Bunke, O., & Batiuk, S. Selecting a method for the parametric adaptation of pi-controller in the control systems of boiler assemblies at thermal power stations with supercritical parameters. *Eastern-European Journal of Enterprise Technologies*. 2022. 2, 2 (116). P. 61–68. DOI: https://doi.org/10.15587/1729-4061.2022.254116.

10. Батюк С.Г., Васянович В.М., Воронюк М.В., Ворошилов А.І. Інваріантні моделі об'єктів управління і універсальні структури ПІД-регуляторів в імітаційному моделюванні теплоенергетичних систем автоматичного регулювання. Вчені записки Таврійського національного університету імені В.І. Вернадського. Серія: Технічні науки. 2024. Том 35 (74). № 6. Частина 2. С. 1–12. DOI: https://doi.org/10.32782/2663-5941/2024.6.2/01.

11. Batiuk S.G. Simulation modeling and digital twinning of energy cyber-physical systems (cyber-energy systems). *Achievements of Ukraine and EU countries in technological innovations and invention* : collective monograph. Riga : Baltija Publishing, 2022. P. 44–109. DOI: https://doi.org/10.30525/978-9934-26-254-8. DOI: https://doi.org/10.30525/978-9934-26-254-8-2.

Batiuk S.G., Voroniuk M.V. APPLICATION OF DIGITAL TWINS AND ARTIFICIAL INTELLIGENCE SYSTEMS IN THE AUTOMATION OF THERMAL ENERGY PROCESSES

The article is devoted to the use of digital twins in classical tasks of automation of thermal energy processes.

The working concepts of digital twins and artificial intelligence systems for automation tasks are defined. It is demonstrated that the technologies of digital twins and artificial intelligence systems have been widely used and are used in the automation of technological processes in general and the automation of thermal energy processes in particular.

The classical examples of the use of digital twins and artificial intelligence systems in the automation of thermal energy processes are analyzed: 1) digital twins in tasks of the upper (supervisory) level of automation of methodical furnaces of rolling mills – real-time calculation of temperature diagrams of metal slabs; 2) artificial intelligence systems in tasks of the lower (controller) level of automation of industrial furnaces – gradient search for the optimal value of the excess air coefficient; 3) digital twins in the tasks of the upper (supervisory) level of automation of the power unit of the power plant – technical diagnostics of the thermal and mechanical state of the electric generator; 4) digital twins in the tasks of the lower (controller) level of automation of the power unit of the power plant – filtering of measured temperatures of the electric generator housing; 5) digital twins in the tasks of the upper (supervisory) level of automation of the glass melting furnace – calculation of the temperature and level of the glass mass in the furnace pool; 6) digital twins in the tasks of the lower (controller) level of automation of the glass melting furnace – formation of a command to flip the torch in the regenerative furnace; 7) artificial intelligence systems in the tasks of the upper (supervisory) level of automation of the glass tube forming line by the horizontal drawing method – gradient search for the minimum cost of products; 8) digital twins in the tasks of the lower (controller) level of automation of the glass tube forming line using the horizontal drawing method – forming the set-point of the regulator of the aerodynamic mode of glass tube forming; 9) artificial intelligence systems in the tasks of the lower (controller) level of automation of heat and power processes – auto-tuning of PID controllers; 10) artificial intelligence systems in the tasks of complex automation of engineering systems – extreme positioning control of solar power plants batteries panels and wind power plants turbines blades.

Key words: *fourth industrial revolution, cyber-physical system, internet of things, edge device, intelligent control systems, digital twin, artificial intelligence, artificial intelligence agent, artificial intelligence system, complex automation, automation of thermal energy processes, automation of engineering systems, automated control system, automatic regulation system, controller functionality, supervisory functionality, adaptive control systems, expert control systems, extreme control systems, auto-tuning of PID-controllers, learning systems, self-learning systems, real-time modeling.*

УДК 004.75
DOI https://doi.org/10.32782/2663-5941/2025.3.2/12

Безкровний М.М.
Тернопільський національний технічний університет імені Івана Пулюя

МЕТОДИ І ЗАСОБИ АНАЛІЗУ АСТРОНОМІЧНИХ ЗОБРАЖЕНЬ ДЛЯ ВИЗНАЧЕННЯ КООРДИНАТ НЕБЕСНИХ ОБ'ЄКТІВ

У роботі розглядається задача високоточного визначення координат небесних об'єктів на цифрових зображеннях, отриманих із ПЗЗ-матриць. Сучасна астрономія висуває підвищенні вимоги до точності визначення положення небесних об'єктів на цифрових зображеннях і має фундаментальне значення для широкого спектра задач – від супроводу малих тіл Сонячної системи до точного позиціонування далеких галактик.

Актуальність теми обумовлена необхідністю автоматизованої обробки великих обсягів астрономічних даних у сучасних оглядових проєктах (Gaia, LSST, Pan-STARRS), де значна частина об'єктів має перекриваючі зображення. Стандартні методи, засновані на максимумі інтенсивності або центрі мас, не забезпечують необхідної точності при складних умовах спостереження. У роботі запропоновано два методи фітингу – SimpleFit (з фіксованою кількістю джерел) та AdaptiveFit (з адаптивним вибором кількості компонент за критерієм Акаіке), реалізовані в програмному модулі мовою Python.

Обидва методи мінімізують функцію відхилення між експериментальними та змодельованими даними з використанням алгоритму Нелдера–Міда. Експериментальне тестування показало, що AdaptiveFit має вищу точність відновлення координат, проте гіршу збіжність, тоді як SimpleFit є більш стабільним, але менш точним. Комбінований підхід, який обирає найкращий результат з обох моделей за критерієм помилки моделі, дозволяє зменшити похибку координат до комбінований підхід зменшив середні похибки до –0.02 по X та –0.08 по Y пікселя відповідно, що свідчить про значне покращення точності у порівнянні з індивідуальними методами.

Результати спостережень, отримані з недостатньою просторовою точністю, можуть призвести до суттєвих похибок у побудові траєкторій, розрахунках швидкостей, оцінці еволюційних характеристик зоряних скупчень та у плануванні подальших місій. Отримані результати підтверджують доцільність поєднання декількох методів фітингу в єдиній системі для підвищення точності й надійності визначення координат у щільних астрономічних полях. Перспективами подальших досліджень є врахування реальних PSF-профілів, фону, застосування змішаних моделей та нейромережевих методів попередньої ініціалізації.

Ключові слова: *астрометрія, фітинг, ПЗЗ-матриця, координати небесних об'єктів, адаптивна модель, інформаційний критерій Акаіке, субпіксельна точність, моделювання зображень, автоматична локалізація, астрономічне зображення.*

Постановка проблеми. Однією з головних задач аналізу астрономічних зображень є точне визначення координат небесних об'єктів, що перекриваються на кадрах, отриманих за допомогою пристроїв із зарядовим зв'язком (ПЗЗ-матриць (англійською: CCD – Charge-Coupled Device)), які використовується в астрономії (цифрові телескопи), медичній візуалізації, цифровій фотографії, оптичному скануванні тощо) [1].

Стандартні методи оцінки положення об'єкта, засновані на знаходженні максимуму яскравості або визначенні центру мас, показують задовільні результати лише у випадку ізольованих об'єктів. При перекритті профілів яскравості від кількох джерел центр мас зміщується, що призводить до систематичних помилок у визначенні координат [2].

Особливістю аналізу таких зображень є необхідність роботи з групованими вибірками: замість спостереження координати кожного окремого фотона, доступні лише інтегральні значення яскравостей пікселів. Це обмежує обсяг доступної інформації і ускладнює задачу розділення джерел. Проблема також ускладнюється через неоднорідність профілів яскравості об'єктів, які можуть мати різні амплітуди та ширини, а також через наявність шумів, зокрема фонового космічного випромінювання та шумів самої ПЗЗ-матриці [3].

Тому для коректного визначення координат необхідно застосовувати методи фітингу інтен-

сивностей зображення моделями [3], що враховують суперпозицію декількох об'єктів, адаптивне оцінювання кількості джерел та їх параметрів, а також оптимізаційні процедури, стійкі до локальних мінімумів і варіацій форми об'єктів.

Астрономічні зображення, отримані з ПЗЗ-матриці, являють собою матрицю дискретних значень яскравості, де кожен піксель фіксує сумарну кількість фотонів, що потрапили на відповідну ділянку сенсора. При цьому реальний профіль світлорозподілу є неперервним, і центр мас об'єкта зазвичай не збігається з центром жодного з пікселів. Це унеможливлює точну локалізацію небесного тіла без використання математичних моделей фітингу.

На рис. 1 зображено приклад такого переходу: зліва – вихідне піксельне зображення об'єкта на CCD-матриці, справа – згладжена поверхня інтенсивностей, апроксимована двовимірною гауссовою функцією.

Аналіз останніх досліджень і публікацій. Проблема точного визначення координат небесних об'єктів за результатами цифрових астрономічних спостережень є предметом активного наукового інтересу протягом останніх десятиліть. У багатьох сучасних дослідженнях наголошується, що прості методи, засновані на знаходженні пікселя з максимальною яскравістю або обчисленні центру мас, є недостатньо точними, особливо в умовах перекриття зображень джерел. Так у роботі Саневич та ін. запропоновано субпіксельну гауссову модель, реалізовану в системі CoLiTec, що дозволяє досягати точності до 0.05 пікселя, хоча її ефективність обмежена простими сценами без перекриттів [3].

Загальноприйнятим підходом у задач астрометрії є моделювання інтенсивностей зображень за допомогою функцій розсіювання світла – PSF. У класичних роботах, зокрема Стетсона П. [4, 5] здійснено фітинг зоряних профілів у щільних полях із використанням фіксованої кількості компонент. Системи, що подібні до SExtractor [1], автоматизують виявлення об'єктів і визначення координат, але зазвичай не враховують адаптивне коригування кількості джерел у перекритті.

З метою зниження похибок при автоматичному визначенні положень об'єктів запропоновано застосування статистичних критеріїв для вибору кількості джерел. Зокрема, в роботі [6] розглядають підходи з адаптивним вибором кількості компонент за допомогою інформаційних критеріїв (AIC, BIC), що дає змогу автоматично моделювати складні сцени з перекриттями.

Питання зміщення положень джерел через їхню морфологічну складність розглядалися в роботі Макарова та ін. [7, 8], де показано, що внутрішня структура об'єкта (наприклад, джети у квазарах) може спричиняти систематичне зміщення координат.

Також перспективним напрямом є застосування моделей Gaussian Process Regression (GPR). У роботі Лукемайра М. [9] описано побудову емулюючих моделей для складних обчислювальних задач в астрономії, включно з інтерполяцією інтенсивності сигналу в субпіксельному масштабі.

Більшість сучасних підходів вказують на необхідність переходу від фіксованих евристичних алгоритмів до адаптивних моделей фітингу, здатних точно описувати розподіл світла у склад-

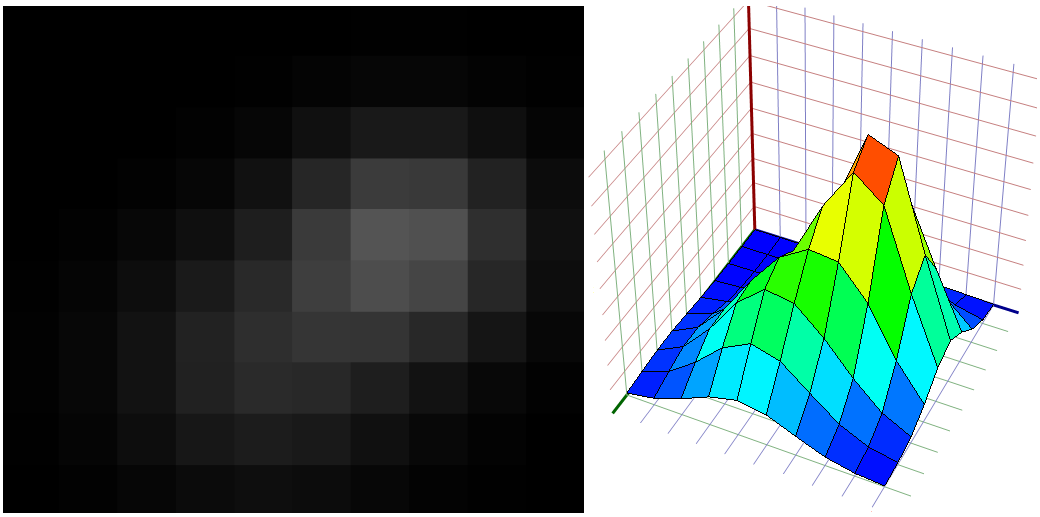


Рис. 1. Зображення небесного об'єкта на CCD-матриці (ліворуч) та побудована поверхня інтенсивності за допомогою фітингу (праворуч)

них спостереженнях. Наявні рішення потребують доопрацювання щодо автоматизації оцінки кількості джерел, корекції фону та врахування статистичних шумів ПЗЗ-матриці.

В напрямку аналізу астрономічних зображень для визначення координат небесних об'єктів, враховуючи сучасний стан досліджень, можна розробити нові фітингові моделі, методи автоматичної оцінки кількості об'єктів, процедури субпіксельної локалізації, враховувати інструментальні та атмосферні спотворення, оптимізувати обробку для великих масивів даних, а також будувати системи контролю якості автоматичного визначення координат.

Розробкою методів визначення координат небесних об'єктів на астрономічних зображеннях займалися численні науковці, які внесли значний вклад у цю галузь: досліджували методи багатошарового аналізу астрономічних зображень із використанням вейвлет-перетворень, що дозволяють виявляти слабкі сигнали на фоні шуму [10]; розробили програми, які стала стандартом для автоматизованого виявлення та фітингу джерел на великих наборах даних [1, 2]; запропонував критерії (наприклад, AIC) для вибору оптимальної моделі при фітингу, що знайшло застосування в обробці зображень із перекритими об'єктами [11, 12]; розробили субпіксельну модель фітингу для точної локалізації астероїдів [3]; розглянули питання точного відновлення положень та орбітальних характеристик небесних тіл, де запропоновано генеративну модель формування каталогу Gaia з урахуванням селекційних функцій та ймовірнісного моделювання виявлення джерел [13].

На сьогодні значна кількість досліджень у цій області сфокусована як на підвищенні точності виявлення координат, так і на оптимізації алгоритмів для роботи з великими обсягами даних.

Постановка завдання. Розглядається задача визначення координат небесних об'єктів на фрагменті цифрового астрономічного зображення, отриманого з ПЗЗ-матриці. Нехай у кадрі спостерігається компактна група з Q небесних джерел, зображення яких частково перекриваються, утворюючи складний розподіл яскравості на ділянці зображення. Нехай Ω – обмежена область на зображенні, у межах якої з певною ймовірністю очікується наявність зазначених Q об'єктів. Область Ω можна визначити як область внутрікадрової корекції об'єктів (ОВКО) – множину пікселів, що містить інтегральну яскравість, пов'язану із сумарним сигналом від компактної групи джерел. У загальному випадку область Ω задається як прямокутник розміром $M \times N$ пікселів з відомими координатами меж. Мета – відновити параметри всіх Q об'єктів за наявними пікселями значеннями. Передбачається, що кожне джерело випромінювання можна апроксимувати двовимірною гауссовою функцією. Альтернативно, ОВКО може бути задана у вигляді прямокутної області, визначеної координатами її меж та розмірами.

Виклад основного матеріалу. Астрономічні знімки, особливо отримані у щільно населених областях неба, можуть містити велику кількість об'єктів, чиї зображення частково перекриваються внаслідок обмежень просторової роздільної здатності телескопів та впливу атмосферної турбулентності. У таких умовах стандартні методи знаходження максимуму яскравості або використання одномодельного фітування дають систематичні похибки у визначенні координат [1].

Актуальність проблеми також зростає у зв'язку із сучасними тенденціями створення великих оглядових проєктів (таких як LSST, Pan-STARRS), де обробка мільярдів об'єктів вимагає високої автоматизації та стійкості алгоритмів до різноманітних типів спостережних ситуацій.

Розробка методів точного визначення координат небесних об'єктів, що перекриваються, з урахуванням статистичних особливостей групованих вибірок фотонів на ПЗЗ-матрицях, є важливою задачею як для прикладної астрометрії, так і для подальших теоретичних досліджень у галузі аналізу космічних даних.

Фітинг (від англ. fitting – «підгонка») у контексті аналізу зображень – це математичний процес підбору параметрів певної моделі так, щоб вона найкраще наближала експериментальні дані, що зводиться до задачі мінімізації функції помилки (критерію невідповідності) між реальною експериментальною інформацією та теоретичною моделлю, шляхом оптимального вибору параметрів моделі.

У випадку астрономічних зображень:

– експериментальні дані – це матриця яскравостей пікселів ПЗЗ-кадру;

– модель – це математична функція, яка описує, як має виглядати яскравість від ідеального небесного об'єкта (наприклад, двовимірна гаусіана, профіль Мофата тощо);

– завдання фітингу – підібрати параметри (амплітуду, координати центру, ширину профілю), щоб модель максимально наближала реальний розподіл яскравостей на зображенні.

Зображення небесних об'єктів на ПЗЗ-кадрі формується за рахунок потрапляння сигнальних фотонів від цих об'єктів у пікселі ПЗЗ-матриці. При цьому межі *ik*-го пікселя ПЗЗ-матриці можна виразити через координати прив'язки центру x_{it}, y_{kt} *ik* -го пікселя *t*-го кадру ПЗЗ-матриці:

$$x_{bi} = x_{it} - \frac{\Delta_{CCD}}{2},\ x_{ei} = x_{it} + \frac{\Delta_{CCD}}{2}; \qquad (1)$$

$$y_{bk} = y_{kt} - \frac{\Delta_{CCD}}{2},\ y_{ek} = y_{kt} + \frac{\Delta_{CCD}}{2}, \qquad (2)$$

де

x_{bi}, x_{ei} – межі (початок і кінець відповідно) *ik*-го пікселя ПЗЗ-матриці за координатою x;

y_{bk}, y_{ek} – межі *ik* -го пікселя ПЗЗ-матриці за координатою y

Спостереженню та реєстрації доступні яскравості A^*_{ikt} пікселів ОВКО (пікселів із множини Ω_{IPS}) на виході ПЗЗ-матриці. Доступні до спостереження напруги на виході N_{CCD} пікселів ПЗЗ-матриці можна привести до експериментальних відносних частот потрапляння фотонів у *ik*-й піксель ПЗЗ-матриці в *t*-му кадрі υ^*_{ikt} (надалі у роботі – відносні яскравості відповідних пікселів):

$$\upsilon^*_{ikt} = \frac{A^*_{ikt}}{\sum\limits_{i,k}^{N_{IPS}} A^*_{ikt}}, \qquad (3)$$

де A^*_{ikt} – експериментальна яскравість *ik*-го пікселя ПЗЗ-матриці на t-му кадрі, що належить досліджуваній ОВКО.

Зображення небесних об'єктів на кадрі мають форму точки, а з урахуванням атмосферної турбулентності стають розмитими. Основними характеристиками для кожного з Q гіпотетичних небесних об'єктів на зображенні кадру є: положення X_{jt}, Y_{jt}; A_G теоретична амплітуда; параметр форми зображення об'єкта σ_G.

Сукупність зазначених параметрів є унікальною для кожного об'єкта і підлягає оцінюванню. Час експозиції кадру вважається настільки малим, що переміщення об'єктів за цей час можна знехтувати. Іншими словами, під час оцінки положень об'єктів компактної групи використання моделі їхнього руху не вимагається.

Вважається, що за час експозиції в пікселі ПЗЗ-матриці потрапляє дискретна кількість фотонів. При цьому точні координати падіння фотонів усередині *ik*-го пікселя принципово недоступні для спостереження. Таким чином, усі події типу «фотон потрапив у точку з певними координатами всередині *ik*-го пікселя» реєстраційною системою об'єднуються в подію «фотон потрапив у *ik*-й піксель». Спостереженню підлягають лише об'єднані (груповані) події. Такі вибірки в математичній статистиці називають груповаными вибірками.

Урахування факту роботи з груповаными вибірками дає можливість витягнути максимально доступну кількість інформації. Таким чином, при оцінці положень елементів компактної групи об'єктів за їхнім ПЗЗ-зображенням доцільно враховувати, що на вхід процедури оцінювання надходить зображення (сукупність потенціалів пікселів ПЗЗ-матриці), яке є групованою вибіркою у зазначеному сенсі.

Передбачається, що істинні положення об'єктів локалізуються в околицях пікселів-піків (локальних максимумів яскравості зображення). У межах цієї роботи під піком розуміють піксель ПЗЗ-матриці, яскравість якого перевищує яскравість усіх суміжних пікселів.

Початкові наближення координат небесних об'єктів на t-му кадрі x_{jt}, y_{jt}, де $j = \overline{1,Q}$ приймаються рівними координатам прив'язки пікселів-піків зображення, що гіпотетично відповідають положенню об'єктів.

Запропоновані методи. Для розв'язання поставленої задачі розглянуто два підходи до фітування інтенсивності як суму декількох гаусіан (Сума 2D-Гаусів, фітинг MSE):

1. SimpleFit (фітинг суми 2D-Гаусіан з фіксованою кількістю компонент і мінімізацією MSE без складної оптимізації).

Вважається, що зображення (матриця інтенсивностей) складається з суми N двовимірних гаусіан:

$$I(x,y) = \sum_{i=1}^{N} A_i \cdot exp\left(-\frac{(x-x_i)^2 + (y-y_i)^2}{2\sigma_i^2} \right)$$

де:

A_i – амплітуда і-го піку;

(x_i, y_i) – координати центру;

σ_i^2 – стандартне відхилення (ширина піку).

У якості цільової функції використовується проста середньоквадратична похибка (MSE) між реальною матрицею і змодельованою:

$$\mathcal{L}_{MSE} = \frac{1}{HW} \sum_{x=1}^{W} \sum_{y=1}^{H} \left[D(x,y) - I(x,y) \right]^2$$

Оптимізація ведеться по всіх A_i, x_i, y_i, σ_i.

2. AdaptiveFit (фітинг суми 2D-Гаусіан з адаптивним вибором кількості компонент на основі Akaike Information Criterion (AIC) та попередньою оцінкою центрів компонент [12, 14, 15]).

Більш адаптивна модель з AIC (Akaike Information Criterion) – це метрика [14], яка використовується для вибору найкращої моделі з набору альтернатив, беручи до уваги як якість фітингу, так і складність моделі.

Критерій базується на принципі, що модель має не лише добре описувати дані, але й бути якомога простішою, щоб уникнути переобучення.

$$\mathcal{L}_{AIC}(\theta) = 2k + n \cdot \ln\left(\frac{RSS(\theta)}{n}\right)$$

де:

k – кількість параметрів у моделі

n – кількість спостережень (пікселів, у випадку зображень: $n = H \cdot W$);

$RSS(\theta)$ – сума квадратів залишків (Residual Sum of Squares):

$$RSS(\theta) = \sum_{x,y}\left[D(x,y) - I(x,y;\theta)\right]^2$$

$\theta \in \mathbb{R}^{4N}$ – вектор параметрів Гауссіан:

$$\theta = (A_1, x_1, y_1, \sigma_1, ..., A_N, x_N, y_N, \sigma_N)$$

У якості моделі використовується також сумма 2D-Гаусів, але додатково застосовано наступні модифікації:

– центри оцінюються спочатку за допомогою функції для позначення зв'язаних компонент на зображенні, зазвичай бінарному – це дозволяє ініціалізувати центри ближче до «справжніх» піків;

– параметри σ_i оцінюються за емпіричними правилами на основі локального контрасту.

На відміну від попереднього методу кількість компонент автоматично підбирається на основі Akaike Information Criterion (AIC). Для розрахунку обох моделей використовувся метод Нелдера-Міда. Цей алгоритм допомагає мінімізувати функції втрат наведені вище. Має наступні переваги:

– простота реалізації;

– працює без градієнтів;

– підходить для неаналітичних функцій;

– добре справляється з задачами малої та середньої розмірності (до 20 змінних).

Алгоритм передбачає формування початкового симплексу у просторі параметрів:

– кожна вершина симплексу – це набір параметрів $\theta_i \in \mathbb{R}^{4N}$;

– початкова точку θ_1 ініціалізуєтьсяз rough-fit (наприклад, пікселем з максимумом інтенсивності + EstimateSigma);

– інші N точок генеруються шляхом зсуву по кожній координат.

В кожній ітерації:

1. Обчислюється втрата $\mathcal{L}(\theta_i)$ для кожної вершини симплексу.

2. Вершини сортуються за значенням втрати:

$$\mathcal{L}(\theta_1) \le \mathcal{L}(\theta_2) \le ... \le \mathcal{L}(\theta_{N+1})$$

3. Обчислюється центр мас (без найгіршої):

$$\overline{\theta} = \frac{1}{N}\sum_{i=1}^{N}\theta_i$$

4. Проводиться рефлексія через відбиття найгіршої точки:

$$\theta_r = \overline{\theta} + \alpha\left(\overline{\theta} - \theta_{N+1}\right)$$

де α – коефіцієнт рефлексії.

5. Проводиться розтягування (опціональний крок). Якщо $\mathcal{L}(\theta_r) < \mathcal{L}(\theta_1)$:

$$\theta_e = \overline{\theta} + \gamma\left(\theta_r - \overline{\theta}\right)$$

6. Якщо $\mathcal{L}(\theta_r) > \mathcal{L}(\theta_N)$ застосовується стискання:

– зовнішнє стискання: $\theta_{oc} = \overline{\theta} + \beta\left(\theta_r - \overline{\theta}\right)$;

– внутрішнє стискання: $\theta_{ic} = \overline{\theta} - \beta\left(\overline{\theta} - \theta_{N+1}\right)$.

7. Проводиться редукція, якщо жоден з варіантів не покращує ситуацію:

$$\theta_i = \theta_1 + \delta\left(\theta_i - \theta_1\right), i = 2, ..., N+1.$$

Мінімізація функції данним методом є більш простішою ніж наприклад метод Левенберга-Маквардта чи градієнтного спуску. Має більше ітерації але меншу обчислювальну складність та потребує менше аналітичної підготовки (не треба будувати матрицю Гессе).

Аналіз якості. Для реалізації задачі фітингу астрономічних зображень із метою точного визначення координат перекритих об'єктів було розроблено спеціалізований програмний модуль мовою програмування Python, із використанням бібліотек NumPy (лінійна алгебра та масиви), SciPy (оптимізація та інтерполяція [16]), matplotlib (візуалізація) та scikit-image (обробка зображень).

Архітектура модуля побудована на модульно-компонентному підході. Основні структурні блоки реалізують два незалежні методи фітингу – SimpleFit і AdaptiveFit, які можуть бути викликані окремо або в паралельному режимі для порівняння результатів. Управління виконанням здійснюється через основний координуючий інтерфейс, який також відповідає за підготовку даних, запуск обчислювальних процедур, вибір параметрів, контроль за збіжністю, збереження та візуалізацію результатів.

Вхідним сигналом для модуля є двовимірна матриця інтенсивностей – або експериментальні ПЗЗ-зображення. Результатом роботи модуля є набір параметрів кожного виявленого об'єкта:

– координати центру (з точністю до 0.01 пікселя);
– амплітуда (максимальна яскравість);
– ширина профілю (σ);
– залишкова похибка (сумарне відхилення моделі від зображення).

У разі запуску обох методів (SimpleFit і AdaptiveFit), система обирає результат із меншим значенням критерію якості (наприклад, середньоквадратичної похибки або AIC). Такий підхід забезпечує як надійність (простий метод завжди має fallback), так і адаптивність (складні випадки обробляє AdaptiveFit).

Для порівняння результатів кожної з моделей було використано набір зображень з перекриттям (для спрощення використовувались зображення з двома об'єктами). У якості критерію точності використовувались середнє значення похибок координат (в пікселях) та параметру сердньоквадратичного відхилення в моделі: $\overline{\Delta}_x, \overline{\Delta}_y, \overline{\Delta}_\sigma$, а також середньоквадратичне значення цих похибок (табл. 1).

Таблиця 1

Отримані результати дослідження якості запропонованих методів

Параметр	SimpleFit	AdaptiveFit
$\overline{\Delta}_x$	0.09	-0.04
$\overline{\Delta}_y$	-0.22	0.02
$\overline{\Delta}_\sigma$	-0.2	-0.24
$\sigma_{\Delta x}$	1.36	1.41
$\sigma_{\Delta y}$	1.19	1.18
$\sigma_{\Delta\sigma}$	0.82	0.7

З данних видно що методи мають приблизно однакову точність. На графіку (рис. 2) зображено некоректно обраховані гауссіани (алгоритм не зійшовся) в залежності від значень амлітуди та сігми.

SimpleFit має набагато менше проблем зі сходженням обрахунку, але має меншу точність. Те що алгоритми мають помилки для різних

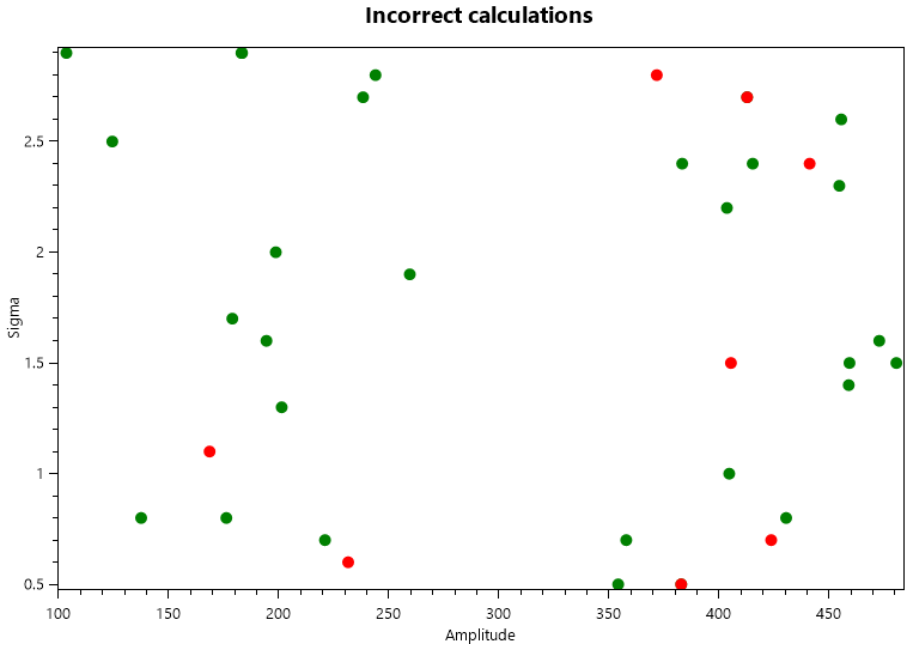


Рис. 2. Некоректно обраховані гауссіани: SimpleFit – червоні маркери, AdaptiveFit – зелені маркери

типів об'єктів дозволяє застосувати обидва ці алгоритми використовуючи найкращій результат. Критерієм вибору може бути використана сумма похибки між реальною матрицею і змодельованою. В цьому випадку точність та якість обрахунків значно зростає.

Таблиця 2

Отримані результати дослідження якості при комбінованому підході

Параметр	Best calculation result
$\overline{\Delta}_x$	-0.02
$\overline{\Delta}_y$	-0.08
$\overline{\Delta}_\sigma$	-0.12
$\sigma_{\Delta x}$	0.87
$\sigma_{\Delta y}$	0.88
$\sigma_{\Delta\sigma}$	0.5

Порівняльне тестування показало, що SimpleFit має кращу сходимість – швидше та стабільніше знаходить оптимальні параметри. Водночас його точність обмежена через фіксовану ширину піків, що призводить до систематичних зсувів у визначенні координат при перекритті джерел.

Метод AdatptiveFit, навпаки, забезпечує вищу точність та краще розділення перекритих об'єктів, особливо при великій щільності джерел, але потребує гарної початкової оцінки для стабільної роботи.

Висновки. Досліджено задачу високоточного визначення координат небесних об'єктів на ПЗЗ-зображеннях, зокрема у випадках часткового перекриття профілів інтенсивності. Це завдання має особливу актуальність у контексті сучасних оглядових програм, таких як Gaia, LSST та Pan-STARRS, які потребують автоматизованих і стійких алгоритмів обробки мільйонів зображень джерел.

Запропоновано два методи фітингу зображень на основі моделі суми двовимірних гаусіан: Fixed Multi-Gaussian Fitting (SimpleFit), який передбачає фіксовану кількість джерел та параметрів, Adaptive Multi-Gaussian Fitting with AIC (AdaptiveFit), який дозволяє автоматично визначати кількість джерел і здійснювати оптимізацію моделі з урахуванням інформаційного критерію Акаіке.

Запропоновано архітектуру програмного модуля, який реалізує обидва методи з використанням Python-бібліотек (NumPy, SciPy, matplotlib). Оптимізація параметрів здійснювалась за допомогою методу Нелдера–Міда, що дозволило уникнути необхідності в обчисленні градієнтів, забезпечивши при цьому збіжність у більшості випадків.

На основі серії тестів встановлено, що: метод SimpleFit має вищу стабільність з точки зору збіжності алгоритму, але дещо поступається в точності оцінки координат; метод AdaptiveFit демонструє кращу точність, однак є більш чутливим до варіацій форми об'єктів і частіше не сходиться при екстремальних значеннях параметрів; комбінований підхід, що полягає у виборі результату з мінімальною похибкою між змодельованим і реальним зображенням, дозволяє значно зменшити як середню похибку координат, так і середньоквадратичне відхилення.

Отримані результати підтверджують доцільність поєднання кількох методів фітингу у єдиній системі. Запропонований підхід забезпечує субпіксельну точність визначення координат навіть у випадках перекриття об'єктів і флуктуацій фону, що робить його придатним для практичного застосування в астрометрії, автоматичному виявленні нових джерел та попередній обробці даних великих оглядових проєктів.

Перспективними напрямами подальших досліджень є:

– узагальнення моделі на випадок реального PSF-профілю (не гауссового);

– використання змішаних моделей типу Gaussian Mixture Models (GMM);

– адаптація методів фітингу до даних зі змінним фоном та нерегулярними шумами;

– впровадження нейромережевих підходів до попередньої ініціалізації кількості джерел.

Список літератури:

1. Bertin, E., Arnouts, S. SExtractor: Software for source extraction. *Astronomy and Astrophysics Supplement Series.* 1996. Vol. 117. No. 2. P. 393–404. URL: https://doi:10.1051/aas:1996164.

2. Rubin Observatory. NSF's National Optical Infrared Astronomy Research Laboratory. URL: https://rubinobservatory.org/explore/how-rubin-works/lsst (date of access:20.04.2025)

3. Savanevych, V. E., Briukhovetskyi, O. B., Sokovikova, N. S., Bezkrovny, M. M., Vavilova, I. B., Ivashchenko, Yu. M., Elenin, L. V., Khlamov, S. V., Movsesian, Ia. S., Dashkova, A. M., Pogorelov, A. V.

A new method based on the subpixel Gaussian model for accurate estimation of asteroid coordinates. *Eastern-European Journal of Enterprise Technologies*. 2016. Vol. 6/9. No. 84. P. 41–49. URL: https://doi.org/10.48550/arXiv.1505.04247.

4. Cigan, P., Makarov, V. V., Secrest, N., Gordon, D., Johnson, M., Lambert, S. Metrics of Astrometric Variability in the International Celestial Reference Frame: Statistical analysis and selection of the most variable sources. *The Astrophysical Journal Supplement*. 2024. Vol. 274. No. 2. P. 1–18. URL: https://doi.org/10.3847/1538-4365/ad6772.

5. Stetson, P. B. DAOPHOT: A computer program for crowded-field stellar photometry. *Publications of the Astronomical Society of the Pacific.* 1987. Vol. 99. No. 613. P. 191.

6. Araya-Hernandez, L., Silva, J. F., Osses, A., Tobar, F. A Bayesian mixture-of-Gaussians model for astronomical observations in interferometry. *2017 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*, Pucon, Chile. 2017. P. 1–5. URL: https://doi.org/10.1109/CHILECON.2017.8229662.

7. Makarov, V. V., Lambert, S., Gordon, D., Johnson, M., DiLullo, C., Lambert, S. Partial Alignment of Astrometric Position Excursions of International Celestial Reference Frame Quasars with Radio Jet Structures. *The Astrophysical Journal Letters.* 2024. Vol. 977. No. 1. URL: https://www.researchgate.net/publication/386143973.

8. Lukemire, M., Rasmussen, C., Scargle, J. D. Statistical analysis of complex computer models in astronomy. *Astrostatistics Journal*. 2021. Vol. 18. No. 2. P. 45–67.

9. Starck, J.-L., Murtagh, F., Bijaoui, A. Image Processing and Data Analysis: The Multiscale Approach. Cambridge University Press, 1998.

10. Deng, Q. A Combined OCBA–AIC Method for Stochastic Variable Selection in Data Envelopment Analysis. *Mathematics*. 2024. Vol. 12. No. 18. URL: https://doi.org/10.3390/math12182913.

11. Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P. *Numerical Recipes: The Art of Scientific Computing*. 3rd ed. Cambridge: Cambridge University Press, 2007. ISBN: 9780521880688. URL: https://www.cambridge.org/0521880688.

12. El-Badry, K., Lam, C., Holl, B., Halbwachs, J.-L., Rix, H.-W., Mazeh, T., Shahaf, S. A generative model for Gaia astrometric orbit catalogs: Selection functions for binary stars, giant planets, and compact object companions. 2024. *Open Journal of Astrophysics*. Vol. 7. URL: https://doi.org/10.48550/arXiv.2411.00088.

13. Akaike, H. A New Look at the Statistical Model Identification. In: Parzen, E., Tanabe, K., Kitagawa, G. (Eds.). *Selected Papers of Hirotugu Akaike*. Springer Series in Statistics. New York: Springer, 1974. P. 215–222. URL: https://doi.org/10.1007/978-1-4612-1694-0_16.

14. Ma, G., Yang, J., Song, X., Zhao, H**.** An Adaptive Clustering Algorithm of Gaussian Curve Fitting the Histogram Based on K-Means. *2019 IEEE International Conferences on Ubiquitous Computing & Communications (IUCC), Data Science and Computational Intelligence (DSCI), and Smart Computing, Networking and Services (SmartCNS)*, Shenyang, China. 2019. P. 141–145. URL: https://doi.org/10.1109/IUCC/DSCI/SmartCNS.2019.00052.

15. SciPy. scipy.optimize.minimize – Minimization of scalar function of one or more variables. URL: https://surl.li/wnpcqm (date of access: 20.04.2025)

16. Virtanen, P., Gommers, R., Oliphant, T. E. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*. 2020. Vol. 17. P. 261–272. URL: https://doi.org/10.1038/s41592-019-0686-2.

Bezkrovnyi M.M. METHODS AND TOOLS FOR ANALYZING ASTRONOMICAL IMAGES TO DETERMINE THE COORDINATES OF CELESTIAL OBJECTS

The article considers the issue of high-precision determination of the coordinates of celestial objects in digital images obtained from CCD matrices. Modern astronomy imposes increasing demands on the accuracy of determining the position of celestial objects in digital images and is fundamental to a wide range of tasks, from tracking small bodies in the solar system to precise positioning of distant galaxies.

The relevance of the topic is due to the need for automated processing of large amounts of astronomical data in modern survey projects (Gaia, LSST, Pan-STARRS), where a significant number of objects have overlapping images. Standard methods based on the maximum intensity or center of mass do not provide the required accuracy under difficult observational conditions.

Different methods of fitting are proposed in this work: SimpleFit (with a fixed number of sources) and AdaptiveFit (with an adaptive selection of the number of components according to the Akaike criterion), implemented in a Python program module. Both methods minimize the deviation function between experimental and modeled data using the Nelder-Mead algorithm.

In experimental testing, AdaptiveFit has a higher accuracy of coordinate recovery but worse convergence, while SimpleFit is more stable but less accurate. The combined approach, which selects the best result from both models according to the model error criterion, reduces the coordinate error. The combined approach reduced the average errors to -0.02 in X and -0.08 in Y pixels, respectively, which indicates a significant improvement in accuracy compared to the individual methods.

Observational results obtained with insufficient spatial accuracy can lead to significant errors in trajectory construction, velocity calculations, assessment of the evolutionary characteristics of star clusters, and in planning future missions. The obtained results confirm the feasibility of combining several fitting methods in a single system to improve the accuracy and reliability of coordinate determination in dense astronomical fields. Prospects for further research include taking into account real PSF profiles, background, the use of mixed models, and neural network methods of preliminary initialization.

Key words: *astrometry, fitting, CCD sensor, coordinates of celestial objects, adaptive model, Akaike information criterion, subpixel accuracy, image modeling, automatic localization, astronomical image.*

УДК 004.75;004.8
DOI https://doi.org/10.32782/2663-5941/2025.3.2/13

Богдан Г.А.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Феоклістов Д.О.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

ПІДВИЩЕННЯ ТОЧНОСТІ ІНЕРЦІАЛЬНИХ НАВІГАЦІЙНИХ СИСТЕМ ЗА ДОПОМОГОЮ НЕЙРОМЕРЕЖЕВИХ ТЕХНОЛОГІЙ

Датчики MEMS, які використовуються в системах орієнтації та навігації, є невід'ємною частиною сучасного технологічного прогресу. Неочевидно, але вони присутні практично у всіх сферах техніки і технологій, починаючи від смартфонів і комп'ютерних технологій, продовжуючи автоматизованими складальними лініями на заводах, обладнанням для точної обробки металу, дерева, пластику та інших матеріалів, закінчуючи дронами і роботами-помічниками.

Ця стаття написана з метою розкриття теми підвищення точності датчиків орієнтації MEMS, в складі інерційних навігаційних блоків, за допомогою нейромереж та методів машинного навчання. В ній розглянуто ключові причини виникнення похибок в цих датчиках. Було описано як зміна температури впливає на МЕМС датчики в інерційних системах навігації до яких належать гіроскопи та акселерометри.

В статті було розглянуто методи боротьби з температурною похибкою МЕМС датчиків. До них можна віднести механічні, такі які боряться з похибкою шляхом температурної ізоляції датчика, та створення його з матеріалів, що відрізняються малою чутливістю до зміни температури. Було визначено основні недоліки притаманні цим методам. Також розглянуто інші методи, що звуться температурною компенсацією, або ж алгоритмічною компенсацією. Їх дія заснована на математичному моделюванні сигналу похибки з ціллю його виділення із загального сигналу виходу датчика.

Основна увага була приділена методам, які в собі містять застосування нейромережевих технологій та методів машинного навчання для побудови моделі похибки. Проаналізовано сучасні дослідження які ведуться в даному напрямку, що дало змогу визначити те, що методи компенсації температурної похибки МЕМС датчиків, шо використовують нейронні мережі, здатні значно покращити точність цих датчиків порівняно з іншими методами.

Ключові слова: *МЕМС, навігаційні системи, нейронна мережа, машинне навчання, компенсація похибок.*

Постановка проблеми. Системи глобальної супутникової навігації (СГСН), часто використовуються в парі з інерційними навігаційними системами (ІНС). Така комбінація дозволяє використовувати преваги обох систем. Система СГСН забезпечує регулярну корекцію навігаційних даних, що надаються ІНС. Остання в свою чергу бере на себе роль навігаційного інструменту в ситуаціях коли сигнал СГСН є недоступним або ненадійним. Це можливо в таких місцях як: гірські долини, густі міські райони та лісова місцевість. Проте, інерційні навігаційні системи, мають властивість накопичувати свою помилку з часом, що може призвести до серйозних відхилень в процесі навігації.

На сьогоднішній день більшість інерційних навігаційних систем створюється на основі мікроелектромеханічних систем (МЕМС) гіроскопів та акселерометрів.

МЕМС гіроскоп це пристрій що складається з мікроскопічних механічних елементів (таких як: балки, мембрани, гіроскопічні маси, інше), що є інтегрованими разом з електронними компонентами. В їх основі роботи лежить ефект Коріоліса. Згідно з ним, якщо об'єкт рухається круговою траєкторією на нього починає впливати сила Коріоліса, яка змінює напрямок його руху в залеж-

ності від кута обертання. У випадку з МЕМС гіроскопом цим об'єктом виступає певна маса, прикріплена до пружного підвісу. Під час роботи гіроскопа, сила Коріоліса впливає на підвішену масу і викликає її зміщення в певному напрямку. Це зміщення можна виміряти через зміни в електричних параметрах, інтегрованих в датчик електричних елементів, які виникають через рух маси. Сигнал, отриманий від зміщення маси, перетворюється в електричний сигнал, який далі аналізується мікропроцесором. За допомогою цього аналізу можна визначити кутову швидкість обертання об'єкта, на якому встановлений гіроскоп.

МЕМС акселерометр, є пристроєм для вимірювання прискорення. Вимірюване прискорення може бути як лінійним так і гравітаційним. Значення прискорення датчик перетворює в електричний сигнал. Основою конструкції даного типу датчиків є маса, що закріплена на пружному підвісі. Під дією зовнішнього прискорення маса переміщується пропорційно цьому прискоренню відносно «жорсткої» частини датчика. Зміщення маси фіксуються ємнісними або резистивними вимірювачами, та в подальшому перетворюється в електричний сигнал який можна інтерпретувати як величину прискорення в напрямку вимірювання.

До переваг МЕМС датчиків відносять: малий розмір, низьке енергоспоживання, інтеграцію датчиків та електроніки на одному пристрої, низька вартість в порівнянні з аналогами [1]. Найбільший недолік датчиків MEMS є їх чутливість до температури.

Вплив температури на точність МЕМС сенсорів, можна описати через декілька ключових факторів, що пов'язані з фізичними властивостями датчиків та їх роботою. До таких факторів належать:

1. Зміни в матеріалі датчика. Датчики МЕМС складаються з мікроскопічних елементів, та виготовлені з кремнію або інших матеріалів, що можуть змінювати свої фізичні властивості при зміні температури. До таких властивостей належать, модуль пружності та фізичні розміри, що під впливом теплового розширення може викликати деформацію датчика, що безпосередньо впливає на його точність.

2. Зміни електричних властивостей. Таких як резистивність чи ємність, що впливає на вихідний сигнал датчика та призводить до зниження точності вимірювань.

3. Шум та термічна стабільність. Термічний шум збільшується при високих або низьких температурах, що значно впливає на чутливість та як наслідок точність датчика. Результатом його дії стає те, що датчик починає давати більші помилки при критичних температурах.

Усі зазначені фактори впливають на роботу МЕМС датчиків, що робить важливим проведення тестів та калібрування цих датчиків для забезпечення їх точної роботи в умовах зміни температури. Таким чином актуальним питанням є пошук оптимального методу, який дозволяє зменшити похибку МЕМС сенсорів викликаної зміною температури.

Аналіз останніх досліджень і публікацій. Аналіз сучасних публікацій дозволив виділити основні методи, що використовуються для подолання похибок МЕМС датчиків [2], [3], [4].

Зазвичай, зменшення похибки викликаної зміною температури досягається за рахунок зниженні чутливості матеріалу датчика до температури. Для цього використовують два методи:

1. Виготовлення чутливого елемента датчика з матеріалу зі зниженою чутливістю до змін температури.

2. Теплова ізоляція датчика.

До недоліків першого методу відносять, значні часові та фінансові затрати, пов'язані з пошуком нових матеріалів. Також, зрозуміло, що зі сторони фізики є неможливим створення матеріалу який би зовсім не піддавався змінам під дією температур.

Недоліком другого методу є те, що він не захищає датчик від внутрішнього нагріву, який виникає під час його роботи. Також, такий метод накладає обмеження на сам температурний ізолятор, що значно ускладнює пошук необхідного матеріалу для його побудови. Шар ізоляції не має бути громістким, оскільки це б нівелювало перевали МЕМС датчиків викликані їх малим розміром.

Інший підхід, який дозволяє зменшити температурну залежність МЕМС сенсорів називається температурною компенсацією. В основі підходу лежить математичне моделювання похибки за допомогою програмних технологій, та фільтрацію вихідного сигналу датчика з подальшим виокремленням сигналу похибки. Існують різні методи моделювання сигналу похибку, які можуть бути засновані на принципах математичної апроксимації та створенню багатопараметричних моделей, а також методів машинного навчання, або штучних нейронних мереж.

Теоретично штучні нейронні мережі є ідеальними з точки зору архітектури та принципу

роботи для вирішення проблеми компенсації таких помилок [5]. Нейронна мережа складається з певного набору «нейронів», з'єднаних між собою зв'язками з певним ваговим коефіцієнтом. Кількість нейронів і зв'язків між ними може бути абсолютно довільним, головне, щоб їх було достатньо для виконання завдання. Така гнучка архітектура нейронних мереж є великим плюсом для вирішення проблем корекції порівняно з константністю математичних конструкцій. Усі дослідження в області застосування нейронних мереж для компенсації помилок MEMS-сенсорів зводяться до пошуку оптимальної структури мережі, що означає визначення необхідної кількості нейронів, їх розташування в шарах мережі та знаходження оптимальних значень коефіцієнтів зв'язків між нейронами.

Постановка завдання. Мета цієї статті – огляд можливостей використання методів машинного навчання для зменшення похибки МЕМС сенсорів в системах орієнтації та навігації

Виклад основного матеріалу. Автори даної статті [6], розглядають проблеми, пов'язані з стохастичною характеристикою та невизначеністю кількісної оцінки MEMS акселерометрів. Запропонований підхід передбачає навчання сурогатних моделей з даними, отриманими в результаті початкового високоточного аналізу кінцевих елементів, що забезпечує багаті набори даних, які генеруються в автономному режимі. Застосування даної системи продемонстровано на прикладі оцінки чутливості MEMS акселерометра з урахуванням невідомих механічних зсувів, надмірного травлення та варіацій товщини. Також авторами продемонстровано використання підходу до апостеріорного розподілу невідомих виробничих параметрів пристрою на основі його реакції на перехідні сигнали напруги. Даний метод значно покращує зворотну та пряму кількісну оцінку невизначеності, даючи точні результати, водночас значно покращуючи ефективність і результативність процесу визначення характеристик. Також використання цього методу дозволяє реконструювати чутливість пристрою, використовуючи тільки сигнали напруги.

В статті [7], пропонується нова модель температурної компенсації для інерційного вимірювального блоку, що заснована на радіальнобазисній нейронній мережі. Дана мережа здатна компенсувати значні детерміновані похибки як акселерометра та і гіроскопічних тріад, в широкому діапазоні температур. Автори даної статті розглядають температурну компенсацію MEMS датчиків як задачу апроксимації функції, а похибки цих датчиків моделюється на основі функції температури та вихідних вимірювань датчика. Автори статті акцентують свою увагу на тому, що запропонований ними метод дозволяє калібрувати як тріади акселерометрів так і гіроскопів.

В [8], запропоновано метод усунення шумів у вихідних сигналах MEMS датчиків у складі інерційних навігаційних блоків, із застосуванням методів штучного інтелекту. Зокрема розглядаючи сигнали MEMS гіроскопа як часовий ряд та застосовуючи для фільтрації вихідного сигналу, рекурентну нейронну мережу із маніпуляціями в довготривалій та короткотривалій пам'яті. Автори зазначають, що даний метод дозволив зменшити стандартне відхилення гіроскопа на 60,3%, 37% і 44,6% для гіроскопів у тріаді відповідно, у порівнянні з необробленим сигналом. Також помилка орієнтації по трьох осях вдалося зменшити на 15,8%, 18,3% і 51,3% відповідно до осі. Також автори даної статті порівнюють свій метод з методом заснованим на моделі автоматичної регресії із фіксованими параметрами, та зазначають, що похибка орієнтації зменшилась на 47,6%, 42,3% та 52,0%, по відповідним осям.

Автори статті [9], використовують нейронну мережу зворотного поширення для прогнозування фазових помилок на основі резонансної частоти квадратурної корекції напруги. Як зазначається, фазова помилка демодуляції призводить до поєднання квадратурної помилки з вихідним сигналом, що призводить до погіршення продуктивності MEMS гіроскопа. Основою метода є, застосування вихідного сигналу кутової швидкості, вигляду прямокутної хвилі, до гіроскопа з автоматичним визначенням значення помилки фази демодуляції за допомогою розробленого алгоритму автоматичної ідентифікації. В рамках дослідження було проведено ряд тестів в діапазоні температур від -40 до 60 °C. Тести показали, що при нормальній температурі 25 °C, похибка зсуву гіроскопа зменшується у вісім разів, а у крайніх випадках температурного діапазону до чотирьох разів, до 0,1 °/год після компенсації під час роботи. Автори зазначають, що перевагами запропонованого методу є точна ідентифікація та компенсація фазової помилки під час роботи.

В статті [10], автори розглядають феномен високої похибки MEMS систем, що виникає через погане співвідношення сигнал/шум, та пропонують для вирішення застосувати нейронну

мережу зворотного поширення вейвлет-порогу алгоритму онлайн-компенсації. Основною ціллю алгоритму є придушення випадкової помилки масивів MEMS датчиків. Для розробки методу було створено матрицю з 10 MEMS гіроскопів. Для злиття даних масиву використовувався метод опорної матриці. Як зазначають автори компенсація випадкової похибки одного гіроскопа стала головним пріоритетом дослідження. Алгоритм запропонований ними, це гібридний алгоритм фільтрації який безпосередньо поєднує вейвлет-порогове зменшення шуму та нейронну мережу зворотного поширення. Завдяки техніці вейвлет-порогового шумозаглушення для оброблення необроблених даних вдалося успішно зменшити ступінь дискретності вихідних даних, кроку ми успішно зменшили ступінь дискретності вихідних даних, що забезпечило більш стабільну та надійну базу даних для подальшого навчання нейронної мережі. Оптимальна мережева модель була отримана шляхом багаторазового моделювання та навчання.

В статті [11], автори концентрують свою увагу на похибках MEMS резонаторів з нітриду алюмінію. Як зазначається даний тип резонаторів є одним із кандидатів на MEMS-резонатори для подолання недоліків громіздких та неінтегрованих кристалів і резонаторів на поверхневих акустичних хвилях. Також автори акцентують увагу на їх компактні розміри, низькій вартості, високій частоті, сумісності з процесами CMOS і здатності охоплювати кілька частот на одному чіпі. Використання штучної нейронної мережі для реалізації цього методу, дозволив покращити частотно-температурну стабільність в діапазоні температур від -40 до 80 °C.

Автори статті [12], зазначають, що Традиційна модель компенсації температурного дрейфу MEMS датчиків не може роз'єднати температурну залежність матеріалів на основі кремнію, оскільки температурно корельовані величини (TCQ) не були отримані всебічно, а адаптивність MEMS системи гіроскопів знижується в різноманітних, складних умовах. Авторами статті було представлено модифіковану модель компенсації MEMS гіроскопів на основі аналізу мікроструктурного теплового ефекту. Для початку, було вивчено температурну залежність матеріалів на основі кремнію, з урахуванням мікроструктури, ефекту теплового розширення та структурної деформації. Параметри модифікованої моделі було визначено на основі радикальної базової функції штучної нейронної мережі та оцінено її продуктивність. Проведені тести показали, що стабільність зміщення MEMS-гіроскопів становила до 10% від традиційної моделі, температурна залежність матеріалів на основі кремнію була краще відокремлена модифікованим, а адаптивність MEMS-гіроскопів до навколишнього середовища була покращена для розширення їх застосування в різноманітних складних умовах.

У статті [13], було запропоновано відразу три алгоритми для боротьби з температурним дрейфом MEMS датчиків. До них входять метод з використанням рекурентної нейронної мережі з глибоким стрибком, також метод із застосуванням методу метелика монарха та його оптимізованої версії. Як зазначають автори, кожен із цих трьох алгоритмів поєднується з моделлю компенсації в реальному часі для компенсації температурного дрейфу акселерометра MEMS. Проведення експериментів підтвердило ефективність використання даних методів для компенсації температурного дрейфу акселерометрів. Також експерименти показали, що у відсотковому значенні поліпшення досягає 96,5%. Температурний дрейф акселерометра розглядають як випадковий часовий ряд, що гарно підходить для обробки рекурентною нейронною мережею, оскільки її контекстні вузли можуть запам'ятовувати попередню інформацію. Нейромережа запропонована в даній статті є різновидом звичайної рекурентної мережі, і високою швидкістю і хорошою точністю.

Авторами статті [14], Запропоновано модель температурної компенсації MEMS гіроскопа на основі нейронної мережі зворотного поширення. Як вказується в статті температурна похибка складає близько 80% від загальної помилки гіроскопа MEMS. Автори звертають увагу на нелінійності температурної похибки. Традиційні методи компенсації включають у себе застосування кусково-лінійної функції, щоб відповідати моделі помилки температури, вони можуть отримувати точні результати підгонки в точках сегментації. Але такий метод не може давати ефективних результатів компенсації в інших точках. Саме це сприяло вибору нейронної мережі зворотного поширення як основу запропонованого методу. Через її потужну здатність обробляти нелінійні моделі. Результатом цього дослідження стало створення моделі температурної похибки MEMS гіроскопа на основі нейромережі зворотного поширення, та підтвердження її ефективності в порівнянні з класичними методами.

В статті [15], проаналізовано ефект повзучості та тепловий ефект, притаманний кварцевим

MEMS акселерометрам. Авторами даної статті пропонується нова температурна модель кварцевого акселерометра на основі нелінійної авторегресії із зовнішнім виходом та вдосконаленим вейвлет-перетворенням. Також метод сприяє вирішенню проблеми затримки спричиненої тепловим ефектом. Вказується, що помилка повзання, що спричиняє несправність деформації структури акселерометра відокремлена від загальних похибок і лише теплова похибка, вплив якої є сильно залежним від температури компенсується запропонованою моделлю. Ще дві температурні моделі було встановлено як моделі порівняння для перевірки ефективності запропонованого методу компенсації. Проведені експерименти показують, що ефективність запропонованого методу є найкращою серед порівняльних моделей.

В [16], авторами запропоновано три алгоритми для компенсації дрейфу MEMS акселерометра, спричиненого дією температури. Вони включають в себе алгоритм заснований на застосуванні рекурентної нейронної мережі глибокої довготривалої та короткотривалої пам'яті. Також метод на основі алгоритму пошуку горобця, та його вдосконаленої версії. Поміж того, для порівняння ефективності запропонованих методів досліджується метод заснований на кусково-лінійній апроксимації. Дослідження запропонованих методів, починається з проведення температурного експерименту. Далі з отриманих даних формується модель похибки заснована на вище перелічених інструментах. Як висновок, автори зазначають, що всі три методи дали значно кращі результати в порівнянні з класичним методом заснованим на використанні кусково-лінійної апроксимації. Найкращим із запропонованих виявився метод на основі рекурентної нейронної мережі глибокої довготривалої та короткотривалої пам'яті з додаванням алгоритму пошуку горобця. Він дозволив покращити показники MEMS акселерометра, в середньому, на 96,68%.

Висновки. В даній статті було розглянуто сучасні методи підвищення точності мікроелектромеханічних датчиків в інерційних системах навігації. Розглянуті методи були вибрані, тому що містять в собі застосування нейронних мереж та методів машинного навчання для вирішення проблеми компенсації похибок викликаних в основному зміною температури в діапазоні від -40 до 80 °C та іншими додатковими чинниками. Дані методи досліджують не тільки компенсацію похибок окремого датчика, та збільшення його точності, а й компенсацію похибок системи датчиків, такі як тріади акселерометрів та гіроскопів, що використовуються в інерційних навігаційних блоках. Після розгляду обраних методів стало зрозуміло, що тематика підвищення точності MEMS датчиків, за допомогою нейромереж та методів машинного навчання є актуальною. Також стало зрозуміло, що такі методи здатні забезпечити значне покращення точності датчиків датчиків в порівнянні з класичними методами, до яких входять методи засновані на побудові моделі температурної похибки за допомогою різних математичних функцій, включаючи кусково-лінійну апроксимацію, апроксимацію поліномом та інше. В порівнянні з класичними методами, методи розглянуті в даній статті давали покращення точності MEMS датчиків до 96,68%.

Список літератури:

1. Sohn H., Wang M. L., Lynch J. P. Sensor technologies for civil infrastructures : volume 1: sensing hardware and data collection methods for performance assessment. Woodhead Publishing, 2022. 825 p.

2. A micro oven-control system for inertial sensors / D. Yang et al. *Journal of microelectromechanical systems*. 2017. Vol. 26. No. 3. P. 507–518. URL: https://doi.org/10.1109/jmems.2017.2692770 (date of access: 09.04.2025).

3. Shi Qiang Liu, Rong Zhu, Heng Gao Ding. A temperature compensation method for micromachined thermal gas gyroscope. *2015 IEEE Sensors*, Busan, 1–4 November 2015. 2015. URL: https://doi.org/10.1109/icsens.2015.7370336 (date of access: 09.04.2025).

4. Feng Y., Li X., Zhang X. An adaptive compensation algorithm for temperature drift of micro-electro-mechanical systems gyroscopes using a strong tracking kalman filter. *Sensors*. 2015. Vol. 15. No. 5. P. 11222–11238. URL: https://doi.org/10.3390/s150511222 (date of access: 09.04.2025).

5. Zhang Q., Tan Z., Guo L. Compensation of temperature drift of MEMS gyroscope using BP neural network. *2009 international conference on information engineering and computer science*, Wuhan, China, 19–20 December 2009. 2009. URL: https://doi.org/10.1109/iciecs.2009.5365140 (date of access: 09.04.2025).

6. Neural networks based surrogate modeling for efficient uncertainty quantification and calibration of MEMS accelerometers / F. Zacchei et al. *International journal of non-linear mechanics*. 2024. P. 104902. URL: https://doi.org/10.1016/j.ijnonlinmec.2024.104902 (date of access: 09.04.2025).

7. Araghi G., Landry R. J. Temperature compensation model of MEMS inertial sensors based on neural network. *2018 IEEE/ION position, location and navigation symposium (PLANS)*, Monterey, CA, 23–26 April 2018. 2018. URL: https://doi.org/10.1109/plans.2018.8373395 (date of access: 09.04.2025).

8. A MEMS IMU de-noising method using long short term memory recurrent neural networks (LSTM-RNN) / C. Jiang et al. *Sensors*. 2018. Vol. 18. No. 10. P. 3470. URL: https://doi.org/10.3390/s18103470 (date of access: 09.04.2025).

9. An in-run automatic demodulation phase error compensation method for MEMS gyroscope in full temperature range / J. Wang et al. *Micromachines*. 2024. Vol. 15. No. 7. P. 825. URL: https://doi.org/10.3390/mi15070825 (date of access: 09.04.2025).

10. Hybrid filtering compensation algorithm for suppressing random errors in MEMS arrays / S. Liang et al. *Micromachines*. 2024. Vol. 15. No. 5. P. 558. URL: https://doi.org/10.3390/mi15050558 (date of access: 09.04.2025).

11. Xu C., Piazza G. Artificial neural network based digital temperature compensation method for aluminum nitride MEMS resonators. *2017 IEEE international ultrasonics symposium (IUS)*, Washington, DC, 6–9 September 2017. 2017. URL: https://doi.org/10.1109/ultsym.2017.8091532 (date of access: 09.04.2025).

12. Modification of mtea-based temperature drift error compensation model for mems-gyros / B. Qi et al. *Sensors*. 2020. Vol. 20. No. 10. P. 2906. URL: https://doi.org/10.3390/s20102906 (date of access: 09.04.2025).

13. Real-Time temperature drift compensation method of a MEMS accelerometer based on deep GRU and optimized monarch butterfly algorithm / G. Gang-Qiang et al. *IEEE access*. 2023. Vol. 11. P. 10355–10365. URL: https://doi.org/10.1109/access.2023.3240766 (date of access: 09.04.2025).

14. Jiang F., Zhang M., Shao T. Temperature compensating model of MEMS gyro based on BP neural network. *IOP conference series: materials science and engineering*. 2018. Vol. 382. P. 052034. URL: https://doi.org/10.1088/1757-899x/382/5/052034 (date of access: 09.04.2025).

15. Temperature compensation for quartz flexible accelerometer based on nonlinear auto-regressive improved model / X. Wang et al. *Measurement and control*. 2023. P. 002029402210892. URL: https://doi.org/10.1177/00202940221089264 (date of access: 09.04.2025).

16. Temperature drift compensation of a MEMS accelerometer based on DLSTM and ISSA / G. Guo et al. *Sensors*. 2023. Vol. 23. No. 4. P. 1809. URL: https://doi.org/10.3390/s23041809 (date of access: 09.04.2025).

Bohdan H.A., Feoklistov D.A. IMPROVING THE ACCURACY OF INERTIAL NAVIGATION SYSTEMS USING NEURAL NETWORK TECHNOLOGIES

MEMS sensors used in orientation and navigation systems are an integral part of modern technological progress. It is not obvious, but they are present in almost all areas of technology and engineering, starting from smartphones and computer technology, continuing with automated assembly lines in factories, equipment for precision processing of metal, wood, plastic and other materials, ending with drones and assistant robots.

This article was written to reveal the topic of increasing the accuracy of MEMS orientation sensors, as part of inertial navigation units, using neural networks and machine learning methods. It considers the key causes of errors in these sensors. It was described how temperature changes affect MEMS sensors in inertial navigation systems, which include gyroscopes and accelerometers.

The article considered methods for combating the temperature error of MEMS sensors. These include mechanical ones, which combat error by thermally insulating the sensor and making it from materials that are insensitive to temperature changes. The main disadvantages inherent in these methods were identified. Other methods, called temperature compensation or algorithmic compensation, were also considered. Their action is based on mathematical modeling of the error signal in order to isolate it from the general output signal of the sensor.

The main attention was paid to methods that include the use of neural network technologies and machine learning methods to build an error model. Modern research conducted in this direction was analyzed, which made it possible to determine that methods for compensating for the temperature error of MEMS sensors using neural networks can significantly improve the accuracy of these sensors compared to other methods.

Key words: *MEMS, navigation systems, neural network, machine learning, error compensation.*

УДК 528.852
DOI https://doi.org/10.32782/2663-5941/2025.3.2/14

Бойко О.В.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Ульяницька К.О.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Коломієць У.Д.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

ДО ПИТАННЯ ОПТИМІЗАЦІЇ ЗАПИТІВ РЕЛЯЦІЙНОЇ БАЗИ ДАНИХ ДЛЯ ВИРІШЕННЯ АНАЛІТИЧНИХ ЗАВДАНЬ

У роботі проаналізовано методи оптимізації запитів до реляційної бази даних з метою підвищення їхньої продуктивності під час вирішення повсякденних завдань аналітиками в умовах постійно зростаючих обсягів інформації для подальшого аналітичного аналізу.

Було визначено, що хоча сучасні оптимізатори запитів для реляційних баз даних мають значний потенціал і надають багато корисних функцій, вони можуть не відповідати очікуванням аналітиків. Оптимізатори запитів мають обмежені можливості для ефективної оптимізації складних або дуже складних запитів, які зазвичай використовуються в роботі аналітиків. При наявності великої кількості таблиць, складних умов, підзапитів та функцій можуть виникнути труднощі під час оптимізації запитів. Тобто, оптимізатори не можуть автоматично адаптуватися до змін навантаження або до конкретних умов виконання запиту. Оптимізатори часто зосереджуються на оптимізації самого запиту, але не забезпечують глибокого аналізу структури даних.

Крім того, наведено практичні рекомендації щодо ручної оптимізації запитів, тобто переписування існуючого запиту з урахуванням поточної ситуації на основі глибокого розуміння особливостей структурованої мови запитів, що, як правило, залишається поза досвідом звичайного аналітика та які не враховуються існуючими оптимізаторами. Отже, створення нового оптимізатора може бути виправдане необхідністю врахування цих недоліків та забезпечення оптимального та адаптивного підходу до оптимізації запитів. Новий оптимізатор може зосередити свої ресурси на забезпеченні ефективного реагування на зміни в даних, використанні обчислювальних ресурсів для генерації швидких та оптимальних планів виконання, а також максимальному використанні паралельного та розподіленого виконання для забезпечення високої продуктивності в різноманітних середовищах даних.

Ключові слова: *оптимізація запитів, реляційна база даних, структурована мова запитів, аналітик, аналітична робота.*

Постановка проблеми. Відомо [1], що однієї з основних умов прийняття єдино правильного, ефективного в умовах непередбаченості і кризових явищ управлінського рішення є забезпечення особи, яка приймає рішення (управлінця), необхідною і достатньою кількістю аналітичної інформації. Таким чином, інформаційно-аналітична діяльність певною мірою убезпечує, захищає керівників, управлінців від ризиків, небезпек і викликів сьогодення, рекомендуючи те чи інше ефективне управлінське рішення, прогнозуючи наперед наслідки його прийняття чи неприйняття, чи бездіяльності. На думку вчених [1], інформаційно-аналітична діяльність – це сукупність дій на основі концепцій, методів, засобів, нормативно-методичних матеріалів для збору, накопичення, обробки та аналізу даних з метою обґрунтування та прийняття рішень. Однак в останні роки спостерігається тенденція лавинообразного збільшення обсягів інформації, яку аналітику треба своєчасно опрацювати та проаналізувати. Відомо, що з кожним роком значно зростає кількість оброблюваних цифрових даних у світі. За 13 років цей показник зріс у 60 разів, починаючи з 2 зетабайтів у 2010 році. Так очікується [2], що 120 зетабайт, згенерованих у 2023 році, у 2025 році зростуть більш ніж на 150% і досягнуть 181 зетабайт.

Для зберігання, зміни та обробки взаємозалежної інформації, переважно великих обсягів, застосовуються бази даних, зокрема реляційні, запити на аналітику до яких стають більш складними, а користувачі очікують оперативних відповідей на запити у реальному часі. На сьогодні однієї з найпопулярніших мов програмування щодо оперування даними, що зберігаються в реляційній базі даних, є SQL [3] та 78% професійних аналітиків даних регулярно використовують його у своїй повсякденній роботі [3]. Тому оптимізація складних запитів до реляційних баз даних (РБД) може допомогти зменшити час та ресурси, витрачені на аналітичну обробку, та стає ключовим чинником для забезпечення швидкості обробки даних, ефективного використання ресурсів та підтримання конкурентоспроможності в сучасному бізнес-середовищі.

Аналіз останніх досліджень і публікацій. Теоретичні засади оптимізації запитів до баз даних з метою підвищення продуктивності баз даних висвітлено у [4–7]. В дослідженнях А. Бари, І. Лунгу та інших [4–5] підкреслюють важливість аналізу вартості виконання SQL запитів у базах даних Oracle. Роботи Белоуса Р., Крилова Є., Анікіна В. [7] присвячено підходам до оптимізації продуктивності запитів у розподілених базах даних, для зниження витрат на опрацювання множинних запитів. Ці роботи демонструють важливість розвитку в галузі методів оптимізації баз даних, з врахуванням особливостей збільшення обсягів даних, однак питання оптимізації складних запитів, які сьогодні здебільшого використовуються у аналітичної діяльності залишаються поза увагою цих науковців.

Постановка завдання. Метою статті є аналіз методів та можливостей сучасних оптимізаторів запитів до реляційної бази даних з метою підвищення продуктивності під час вирішення аналітиками повсякденних завдань в умовах сьогодення, а саме постійного збільшення обсягів інформації для аналітичного аналізу.

Виклад основного матеріалу. Згідно з дослідженням McKinsey & Company, проведеним у 2024 році, компанії, які ефективно використовують передову аналітику даних, в середньому мають на 23% вищий приріст доходу, ніж їхні конкуренти. Автоматизація рутинних аналітичних завдань може заощадити аналітикам до 40% їхнього часу, звільняючи їх для зосередження на більш стратегічних завданнях [3]. Саме тому оптимізація запитів SQL є важливою навичкою для аналітиків, оскільки вона може значно підвищити ефективність використання баз даних під час інформаційно-аналітичної діяльності. Тому однією з головних вимог, яким сьогодні має відповідати аналітик, є навички написання оптимальних запитів до РБД, якими не завжди володіють відповідні фахівці з аналітичної діяльності.

В той же час сучасні РСУБД в своєму складі мають оптимізатори. Оптимізатор SQL – це частина механізму баз даних, призначена для створення найефективнішого плану виконання SQL-запитів. Тобто це алгоритм, який аналізує запит і вибирає найефективніший спосіб його виконання, наприклад шляхом вибору відповідних індексів, способу об'єднання таблиць або порядку операцій. Як правило, оптимізатор аналізує SQL-запит, враховуючи його структуру, використані оператори та доступні індекси. Після цього оптимізатор генерує різні плани виконання, тобто послідовності операцій, які можна використовувати для виконання запиту. Далі здійснюється оцінка продуктивності кожного згенерованого плану (на основі статистики бази даних, часу виконання, споживання ресурсів) та вибір найефективнішого. Вибраний план виконання потім використовується для виконання запиту. Оптимізатори SQL призначені, як правило, для адміністраторів баз даних та розробників, які створюють та підтримують запити. Розуміючи, як працює оптимізатор, користувачі, до яких можна віднести аналітиків, можуть оптимізувати свої запити, покращуючи продуктивність використання програми.

Розглянемо оптимізатори, що використовують популярні РБД.

1. SQL Server Query Optimizer – це оптимізатор, що працює на основі витрат. Кожен можливий план виконання має пов'язані з ним витрати з точки зору обсягу використаних обчислювальних ресурсів. Оптимізатор запитів повинен проаналізувати можливі плани та вибрати той, що має найнижчу розрахункову вартість. Оптимізатор спирається на статистику розподілу, коли оцінює витрати ресурсів різних методів для вилучення інформації з таблиці або індексу. Статистика розподілу зберігається для стовпців та індексів і містить інформацію про щільність базових даних. Щільність визначає розподіл унікальних значень, що існують у даних, або середню кількість дублікатів значень для заданого стовпця. Зі зменшенням щільності селективність значення зростає.

2. Oracle SQL Optimizer – інструмент розроблений для оптимізації запитів до баз даних Oracle. Його головна мета – мінімізувати витрати (час), необхідні для отримання результатів запиту. Оптимізатор працює на основі багатьох факторів, до яких можна віднести.

Статистика бази даних. Оптимізатор використовує статистику, зібрану з бази даних, таку як розподіл значень у стовпцях, кількість рядків у таблицях та розмір блоків даних. Якщо статистика застаріла або неточна, оптимізатор може приймати неоптимальні рішення.

Структура запиту. Складні запити з багатьма реченнями JOIN, підзапитами або функціями може бути складніше оптимізувати. Структура запиту та оператори, що використовуються в ньому, впливають на вибір плану оптимізатором.

Доступність індексів. Індекси є ключовим інструментом для прискорення запитів, що дозволяє оптимизатору швидко отримувати доступ до необхідних даних. Відсутність належних індексів або їх неправильне використання може призвести до уповільнення виконання запитів.

Параметри конфігурації:. Параметри конфігурації бази даних, такі як розмір буферної пам'яті та параметри паралельної обробки, можуть впливати на продуктивність запитів.

Історія виконання запитів. Оптимізатор може використовувати інформацію про попередні виконання запитів, щоб краще адаптувати плани до поточних потреб.

3. PostgreSQL Query Optimizer – оптимізатор запитів для PostgreSQL. Він використовує генетичні алгоритми та здійснює планування запитів, використовуючи евристичний пошук. Це дозволяє скоротити час планування для складних запитів (у яких з'єднуються багато відносин), за рахунок того, що іноді отримані плани поступаються за якістю планам, що вибираються при переборі. Тип оптимізатора – вартісний. Особливості: ліцензійна плата за використання; складний у використанні для початківців; вимагає налаштування та підтримки спеціаліста, який має глибокі знання у даній сфері, якими, нажаль володіють не всі аналітики.

Для оцінки ефективності оптимізації запиту відповідними оптимизаторами застосовують наступні показники.

Час виконання запиту – загальний час, який потрібно для виконання запиту до бази даних. Чим менше час виконання, тим ефективніше запит.

План виконання запиту – візуалізація запропонованого після оптимізації плану виконання запиту, тобто порядку виконання операцій, використання індексів, методів з'єднання таблиць тощо.

Кількість прочитаних/записаних рядків – інформація про кількість прочитаних або записаних рядків під час виконання запиту. Менша кількість операцій зчитування/запису може свідчити про більш ефективне виконання запиту.

Використання індексів – які індекси були використані під час виконання запиту. Використання індексів може покращити продуктивність запиту.

Обсяг пам'яті та ресурсів. Це може допомогти виявити проблеми з використанням ресурсів та встановити оптимальні значення.

Статистика запиту. Це може допомогти виявити слабкі місця та виконати точкову оптимізацію, тощо

Проведемо експеримент, чи здатна система, де застосовується існуючий оптимізатор, оптимізувати запит, який використовують аналітики у реальній повсякденної діяльності так, як це виконує людина. Усі дослідження проводилися в тому ж тестовому середовищі, тобто на комп’ютері (операційна система Windows 10 Home 64bit, процесор AMD Ryzen 5 2600, Ram 16 GB 3000MHz CL15, SSD Samsung 970 1TB). Щоб забезпечити точніші вимірювання на момент дослідження на комп’ютері працювали лише операційна система, тестова програма, та відповідна РСУБД. Комп’ютер було відключено від мережі, а решта програм, особливо антивірусне програмне забезпечення і брандмауер було вимкнено. Для дослідження було використано базу даних кол-центру, яка містить 10 млн. записів. Треба визначити кількість дзвінків від клієнтів за місяць. Було створено два різні запити (запит 1, запит 2) в результаті яких було розв’язано одне й теж саме завдання та отримано одну й ту ж інформацію, але під час написання запитів використовувалися різні оператори та логіка написання у зв’язку з різним практичним досвідом аналітиків у створенні запитів (рис. 1, рис. 2). Причому запит 2 написано більш досвідченим фахівцем у написанні запитів SQL. Також визначено план виконання запиту 1 та запиту 2 за допомогою оптимізатора PostgreSQL Query Optimizer (рис. 3, рис. 4).

```
explain(buffers,analyze)
select count(*), min_login_date::date
from
(
  select min(login_date) min_login_date
  from users u
    left join account a on u.id = a.user_id
    left join account_log al on al.account_id = a.id
  group by u.id
) t
where min_login_date between '2023-04-01' and '2023-04-30'
group by min_login_date::date
order by min_login_date::date
```

Рис. 1. Запит 1 для розрахунку кількості клієнтів за місяць

postgres/postgres@PostgreSQL 11

Query Editor | Query History

Data Output | Explain | Messages | Notifications

	QUERY PLAN text
15	-> Merge Left Join (cost=0.99..550060.76 rows=10000020 width=16) (actual time=0.032..5334.003 rows=10048024 loops=1)
16	Merge Cond: (u.id = a.user_id)
17	Buffers: shared hit=970813 read=223089 written=428
18	-> Index Only Scan using users_pkey on users u (cost=0.43..423617.73 rows=10000020 width=8) (actual time=0.013..2697.031 rows=10000000 loops=1)
19	Heap Fetches: 10000000
20	Buffers: shared hit=4 read=191251 written=370
21	-> Materialize (cost=0.55..88942.97 rows=1000000 width=16) (actual time=0.016..1485.136 rows=1000000 loops=1)
22	Buffers: shared hit=970809 read=31838 written=58
23	-> Nested Loop Left Join (cost=0.55..86442.97 rows=1000000 width=16) (actual time=0.012..1370.956 rows=1000000 loops=1)
24	Join Filter: (al.account_id = a.id)
25	Buffers: shared hit=970809 read=31838 written=58
26	-> Index Scan using ix__account__user_id on account a (cost=0.42..71434.83 rows=1000000 width=16) (actual time=0.006..1201.071 rows=1000000...
27	Buffers: shared hit=970808 read=31838 written=58
28	-> Materialize (cost=0.13..8.14 rows=1 width=16) (actual time=0.000..0.000 rows=0 loops=1000000)
29	Buffers: shared hit=1
30	-> Index Scan using ix__account_log__account_id on account_log al (cost=0.13..8.14 rows=1 width=16) (actual time=0.002..0.002 rows=0 loops=1)
31	Buffers: shared hit=1
32	Planning Time: 0.622 ms
33	Execution Time: 7168.892 ms

Рис. 2. План виконання запиту 1

postgres/postgres@PostgreSQL 11

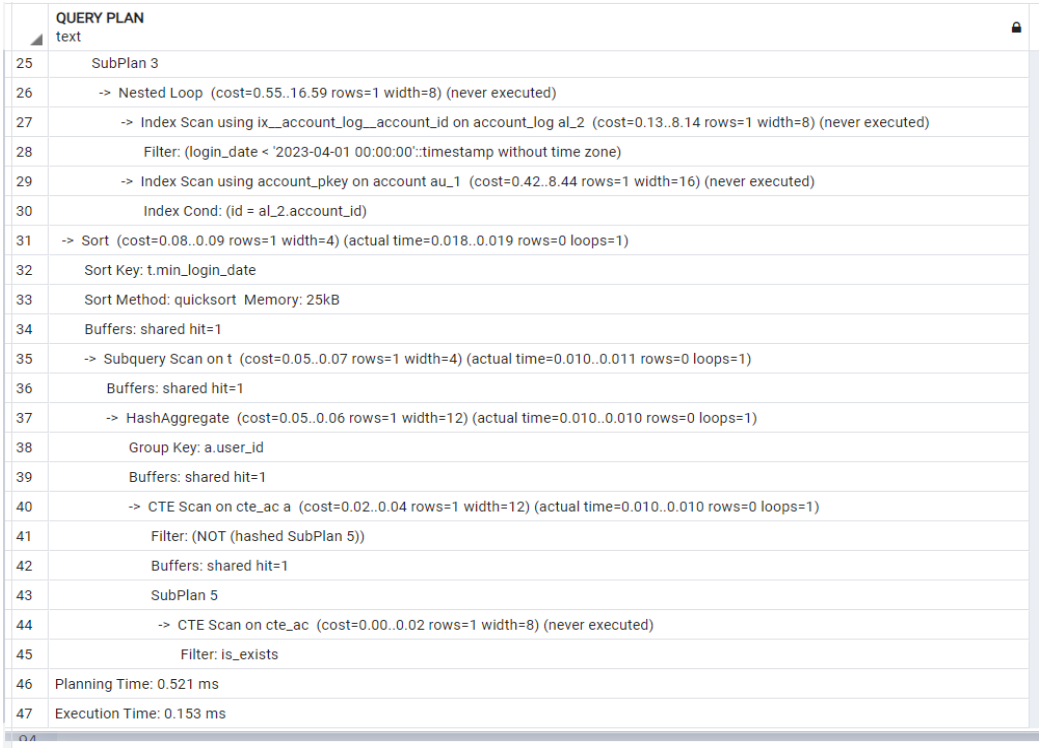
Query Editor | Query History

```
with cte
as
(
 select min(login_date :: date) min_login_date, al.account_id
  from  account_log al
  where al.login_date between '2023-04-01' and '2023-04-30'
  group by al.account_id
),
cte_ac
as
(
  select min_login_date, a.user_id,
           exists(select * from account_log al
           join account au on au.id = al.account_id where au.user_id = a.user_id and al.login_date < '2023-04-01' ) is_exists
  from cte
    join account a on a. id = cte.account_id
)
select count(*), min_login_date
from
(
  select a.user_id, min(min_login_date) min_login_date
  from cte_ac a
  where a.user_id not in (select user_id from cte_ac where is_exists)
  group by a.user_id
) t
group by min_login_date
order by min_login_date
```

Data Output | Explain | Messages | Notifications

	QUERY PLAN text
1	GroupAggregate (cost=33.30..33.32 rows=1 width=12) (actual time=0.009..0.011 rows=0

Рис. 3. Запит 2 для розрахунку кількості клієнтів за місяць



	QUERY PLAN text
25	SubPlan 3
26	-> Nested Loop (cost=0.55..16.59 rows=1 width=8) (never executed)
27	-> Index Scan using ix__account_log__account_id on account_log al_2 (cost=0.13..8.14 rows=1 width=8) (never executed)
28	Filter: (login_date < '2023-04-01 00:00:00'::timestamp without time zone)
29	-> Index Scan using account_pkey on account au_1 (cost=0.42..8.44 rows=1 width=16) (never executed)
30	Index Cond: (id = al_2.account_id)
31	-> Sort (cost=0.08..0.09 rows=1 width=4) (actual time=0.018..0.019 rows=0 loops=1)
32	Sort Key: t.min_login_date
33	Sort Method: quicksort Memory: 25kB
34	Buffers: shared hit=1
35	-> Subquery Scan on t (cost=0.05..0.07 rows=1 width=4) (actual time=0.010..0.011 rows=0 loops=1)
36	Buffers: shared hit=1
37	-> HashAggregate (cost=0.05..0.06 rows=1 width=12) (actual time=0.010..0.010 rows=0 loops=1)
38	Group Key: a.user_id
39	Buffers: shared hit=1
40	-> CTE Scan on cte_ac a (cost=0.02..0.04 rows=1 width=12) (actual time=0.010..0.010 rows=0 loops=1)
41	Filter: (NOT (hashed SubPlan 5))
42	Buffers: shared hit=1
43	SubPlan 5
44	-> CTE Scan on cte_ac (cost=0.00..0.02 rows=1 width=8) (never executed)
45	Filter: is_exists
46	Planning Time: 0.521 ms
47	Execution Time: 0.153 ms

Рис. 4. План виконання запиту 2

Відповідно у запиті 1 Planning Time = 0.6 мс та Execution Time = 7168.1 мс, а у запиті 2 Planning Time = 0.5 мс та Execution Time = 0.153 мс. Отже продуктивність запиту SQL залежить від досвіду фахівця в створенні запитів SQL а використання сучасних засобів оптимізації поки не взмозі повною мірою замінити цей досвід.

Тому існують проблеми [7–9], які не не враховують сучасні оптимізатори, але їх треба враховувати досвідченим фахівцям під час створення запитів, але які доступні для розуміння не кожному аналітику з урахуванням його професійної спрямованості. А саме:

Синтаксичні помилки, які можуть призвести до неправильних або неефективних запитів, тому запити не інтерпретуються обробником баз даних належним чином.

Надто складні запити, які важко зрозуміти оптимізувати. Використання CTE замість вкладених підзапитів може покращити як читабельність, так і продуктивність.

Правильне використання агрегатних функцій, таких як COUNT, SUM, AVG, є важливим для ефективності операцій з даними. Важливо використовувати їх у контексті для отримання оптимальних результатів.

Уникнення надмірної кількості об’єднань: Занадто багато об’єднань між таблицями призводить до збільшення складності запитів та збільшення часу їх виконання. Зосередження на мінімізації кількості підключень допомагає покращити продуктивність.

Неправильний вибір типів даних призводить до неефективного зберігання інформації та проблем із продуктивністю.

Хаотичне використання функцій JOIN, занадто багато вкладених запитів та відсутність належних індексів.

Відсутність індексів для стовпців, що використовуються в операторі WHERE.

Уникати IN та використовувати EXISTS, щоб перевірити чи існують запитувані дані в таблиці.

Уникати HAVING та, де можливо, використовувати WHERE.

Обовязкове використання параметрів у SQL-запитах. Загрози безпеці SQL, включаючи атаки SQL-ін’єкцій, можуть призвести до серйозних порушень даних та втрати довіри користувачів. Тому цей метод запобігає вставці шкідливого коду в запити неавторизованими користувачами, тим самим захищаючи базу даних від атак та покращити читабельність коду, що полегшуєйого підтримку.

Отже, оптимізація запитів – це складний процес, і багато факторів, включаючи згадану раніше статистику бази даних, можуть призвести до повільнішого виконання запитів, ніж очікувалося.

Висновки. Оптимізація аналітичних запитів вимагає від аналітиків глибокого розуміння мови SQL та її принципів, тому що без розуміння і того, як працює механізм SQL, як він організовує дані та як інтерпретує команди неможливо створити оптимальний запит в сучасних умовах, що значно підвищує вимоги до компетентностей аналітиків.

Хоча сучасні оптимізатори запитів до баз даних мають значний потенціал і надають багато корисних функцій, все ж існують певні аспекти, в яких вони можуть не відповідати очікуванням аналітиків, а саме, комплексність запитів, тобто оптимізатори запитів мають обмежену здатність до ефективної оптимізації складних або дуже складних запитів, які зазвичай використовують у роботі аналітики. При наявності багатьох таблиць, складних умов, підзапитів і функцій можуть виникати виклики результативної оптимізації. Це означає, що вони не можуть автоматично адаптуватись до змін у навантаженні або конкретних умов виконання запиту. Як результат, вони можуть не завжди найкращим чином пристосовуватись до змінних умов та динамічних вимог. Отже, створення нового оптимізатора може бути обгрунтовано необхідністю вирішення цих недоліків та надання оптимального та адаптивного підходу до оптимізації запитів.

Список літератури:

1. Варенко В.М. Інформаційно-аналітична діяльність: Навч. посіб. / К.: Університет «Україна», 2014. 417 с.
2. Fabio Duarte. Amount of Data Created Daily. 2025. URL: https://explodingtopics.com/blog/data-generated-per-day
3. Kamila Ostrowska. The Most Popular Databases in 2024. URL: https://learnsql.com/blog/most-popular-databases-2024/
4. Rupley, M. L. Introduction to query processing and optimization. Indiana University, South Bend, South Bend, IN, USA, TechReport TR–20080105–1.2.Bhajipale, R.,Bisen, P., Meshram, A., & Thakur, S. S. 2016.
5. Ahmad K. Query Performance in Database Operation. URL: https://www.ftsm.ukm.my/v5/file/research/technicalreport/PS-FTSM-2020-045.pdf
6. Tuning the SQLquery in order to reduce time consumption. International Journalof Computer Science Issues, 9 (4/3), 418–423.4.Habimana, J. 2015.
7. Query optimization techniques – tips for writing efficient and faster SQL queries. International Journal of Scientific &Technology Research, 4 (10), 22–26.5.Sahal, R., Nihad, M., Khafagy, M.H., & Omara, F. A. 2018.
8. Белоус Р., Крилов Є., Анікін В. Методи оптимізації запитів розподілених БД. Адаптивні системи автоматичного управління. 2021. Т. 2, № 39. С. 3–11. URL: https://doi.org/10.20535/1560-8956.39.2021.247364
9. Query optimization techniques in SQL Server: Database Design and Architecture July 13, 2018 by Ed Pollack URL: https://surli.cc/nsnbgq

Boiko O.V., Ulianytska K.O., Kolomiiets U.D. ON THE QUESTION OF OPTIMIZATION OF RELATIONAL DATABASE QUERIES FOR SOLVING ANALYTICAL PROBLEMS

The work considers analyzed methodsfor optimizing queries to a relational database in order to increase their productivity when solving everyday tasks by analysts in conditions of constantly increasing amounts of information for further analytical analysis.

It was determined that although modern query optimizers for relational databases have significant potential and provide many useful functions, they may not meet the expectations of analysts. Query optimizers have a limited ability to effectively optimize complex or very complex queries, which are usually used in the work of analysts. When there are many tables, complex conditions, subqueries and functions, difficulties may arise when optimizing queries. That is, optimizers cannot automatically adapt to changes in the load or to specific conditions for executing the query. Optimizers often focus on optimizing the query itself, but do not provide a deep analysis of the data structure.

In addition, there are practical recommendations for manual query optimization, i.e. rewriting an existing query taking into account the current situation based on a deep understanding of the features of the structured query language, which, as a rule, remains beyond the experience of an ordinary analyst and which are not taken into account by existing optimizers. Therefore, the creation of a new optimizer can be justified by the need to take these shortcomings into account and provide an optimal and adaptive approach to query optimization. The new optimizer can focus its resources on ensuring effective response to changes in data, using computational resources to generate fast and optimal execution plans, and maximizing the use of parallel and distributed execution to ensure high performance in a variety of data environments.

Key words: *query optimization, relational database, structured query language, analyst, analytical work.*

УДК 004.8+004.02+004.9
DOI https://doi.org/10.32782/2663-5941/2025.3.2/15

Бочок В.О.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Федорова Н.В.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

ОБМІН ЗНАННЯМИ В INDEPENDENT DQN БАГАТОАГЕНТНІЙ СИСТЕМІ ТА ОСОБЛИВОСТІ ЙОГО РЕАЛІЗАЦІЇ

У статті розглянуто проблему підвищення ефективності навчання агентів у багатоагентних системах шляхом впровадження механізму обміну знаннями. Зазначено, що традиційні підходи, засновані на алгоритмах Deep Q-Network (DQN), переважно орієнтовані на використання власного досвіду агентів, що обмежує швидкість збіжності, знижує якість отриманих стратегій та не дозволяє повною мірою використовувати потенціал системи. Водночас у багатоагентних системах існує значний потенціал покращення результатів завдяки обміну знаннями між агентами, що дозволяє ефективніше використовувати наявні ресурси досвіду.

Запропоновано механізм обміну знаннями за моделлю «вчитель–учень», що базується на дистиляції політики. Агент із вищими результатами визначається як вчитель і передає знання агенту-учню з урахуванням різниці в сумарно отриманих за епізод винагородах. Передбачено регулювання інтенсивності навчання на основі співвідношення нагород, що дозволяє зберігати індивідуальність агентів, поступово адаптувати їхні стратегії, а також застосовувати більш інтенсивне навчання для агентів зі значною різницею в отриманих винагородах, яка використовується як об'єктивний показник якості стратегії.

Особливу увагу приділено реалізації асинхронного обміну знаннями без блокування агентів. Запропонований алгоритм дає змогу агентам самостійно накопичувати досвід, публікувати повідомлення з узагальненою інформацією про політики та обирати відповідного вчителя без необхідності синхронізації процесів навчання або зупинки виконання інших агентів. Це забезпечує безперервність навчання, покращує гнучкість системи та підвищує ефективність її роботи в розподілених обчислювальних середовищах.

Експериментальні дослідження, проведені в середовищі OpenAI Gym CartPole-v1 з використанням набору DQN-агентів, підтвердили ефективність запропонованого підходу. Встановлено, що обмін знаннями за моделлю «вчитель–учень» сприяє прискоренню збіжності навчання, підвищенню стабільності стратегій агентів, покращенню загальної продуктивності системи та зменшенню часу досягнення заданих цільових показників якості.

Ключові слова: *Deep Q-Network (DQN), Обмін знаннями, Асинхронна комунікація, Багатоагентні системи, Аналітика даних, Навчання з підкріпленням.*

Постановка проблеми. Багатоагентні системи (Multi-Agent Systems, MAS) є одним із ключових напрямів сучасної інформатики та штучного інтелекту, що знаходять все ширше застосування – від комп'ютерних ігор до розподілених систем управління і робототехніки. Особливістю MAS є взаємодія декількох інтелектуальних агентів, які спільно вирішують завдання, часто недоступні для монолітних підходів [1]. Така парадигма забезпечує низку переваг: декларативність постановки цілей, автономність кожного агента, стійкість системи до відмов окремих компонентів та відсутність централізованого керування [1; 2]. Завдяки децентралізованому характеру MAS є гнучкими та масштабованими: нові агенти або функції можуть додаватися без суттєвої переробки всієї системи, а вихід з ладу окремого агента не призводить до відмови всієї системи [2]. Таким чином, багатоагентний підхід дозволяє створювати розподілені і стійкі до збоїв рішення для складних проблем.

Одним із методів, що надає агентам можливість ефективно навчатися поведінки на основі взаємодії з середовищем, є навчання з підкрі-

пленням (reinforcement learning, RL). У рамках RL агент вивчає оптимальну стратегію дій шляхом багаторазового проб і помилок із отриманням зворотного зв'язку у вигляді винагород чи штрафів за дії [3]. Цей підхід особливо корисний для автономних агентів, оскільки не потребує нагляду або попередньо розмічених даних: агент самостійно вдосконалює свою політику, поступово максимізуючи сумарну винагороду [3]. Завдяки навчанню з підкріпленням інтелектуальні агенти здатні адаптуватися до динаміки середовища та опановувати навички, не закладені явно в їх програмі, що робить цей підхід фундаментальним для розробки багатоагентних систем.

Для розв'язання задач навчання агентів у середовищах з великими або неперервними просторами станів одним із класичних підходів став алгоритм Deep Q-Network (DQN). DQN поєднав метод Q-навчання з глибинними нейронними мережами, що дозволило агентам досягти рівня людини в комп'ютерних іграх, навчаючись безпосередньо за сирими піксельними спостереженнями (наприклад, у середовищі Atari) [4]. Втім, базовий алгоритм DQN має низку обмежень, виявлених подальшими дослідженнями. По-перше, навчання DQN може бути нестабільним: оцінки функції цінності здатні коливатися або навіть дивергувати за неналежних параметрів, тому для стабілізації застосовуються спеціальні прийоми (наприклад, фіксована цільова мережа) [5]. По-друге, збіжність DQN зазвичай вимагає дуже великої кількості взаємодій з середовищем, тобто алгоритм навчається відносно повільно [6]. Частково це зумовлено тим, що DQN неефективно використовує отриманий досвід: кожна взаємодія зі середовищем робить лише обмежений внесок у навчання, тому агенту потрібні численні повторні епізоди для наближення до оптимальної стратегії [6]. У результаті, незважаючи на успіх DQN, його класична версія вважається алгоритмом із низькою стабільністю, повільною збіжністю та недостатнім використанням досвіду [5; 6]. Зазначені проблеми стимулювали появу низки удосконалень, зокрема методів зменшення переоцінювання Q-значень дій (Double DQN) [5] та пріоритетного повторного відтворення досвіду (Prioritized Experience Replay) [6].

Окрім технічних недоліків алгоритму, у середовищах, що змінюються з часом або містять кілька активних агентів, постає проблема адаптації: агентам необхідно безперервно навчатися протягом всього циклу їх функціонування. Система може змінюватися з часом, а отже фіксованої моделі, навченої один раз, може виявитися недостатньо – потрібна здатність алгоритму навчання підлаштовуватися до нових умов і продовжувати навчання без повторного перезапуску. Безперервне навчання агентів дозволяє підтримувати ефективність системи у довгостроковій перспективі та враховувати появу нових ситуацій, відсутніх на етапі початкового тренування. Таким чином, розробка методів, які забезпечують тривале поступове навчання (continual learning) агентів у динамічних багатоагентних середовищах, є актуальним науковим завданням [7].

Більшість існуючих підходів до покращення ефективності DQN-агентів зосереджуються на вдосконаленні процесу навчання на основі власного досвіду. Проте в умовах багатоагентних систем, де одночасно функціонує кілька незалежних агентів з різними траєкторіями навчання, обмін досвідом між агентами може стати додатковим джерелом інформації, що сприяє прискоренню збіжності та покращенню якості навчання. Такий підхід не заперечує наявні модифікації DQN, а доповнює їх, дозволяючи ефективніше використовувати наявні дані в системі.

Особливої уваги потребує реалізація механізмів обміну знаннями у системах, де агенти працюють незалежно та асинхронно. У подібних умовах важливо уникати блокування чи затримок, спричинених очікуванням синхронізації з іншими агентами. Відтак, виникає науково-практичне завдання: розробити ефективний механізм обміну знаннями між агентами, який враховує гетерогенність їх досвіду та забезпечує асинхронність навчального процесу.

Аналіз останніх досліджень і публікацій. Автори [8] розглядали сценарій, у якому агент, натренований на виконання окремої дії, передає свої знання іншому агентові, який має вирішувати багатозадачні проблеми. Передача здійснювалася за допомогою дистиляції політики – методу, який дозволяє переносити стратегію одного агента до іншого через навчання за м'якими мітками. Результати експериментів продемонстрували ефективність такого підходу навіть у випадках, коли агенти не були гетерогенними, що свідчить про широку застосовність методів дистиляції для обміну знаннями у багатоагентних середовищах. Крім того, у роботі коротко обговорено можливість організації процесу «онлайн» дистиляції, тобто безперервної передачі знань під час навчання, однак без поглибленого опрацювання цього питання [8].

Ще одним прикладом сучасного підходу до обміну знаннями є метод KnowSR, запропонований у [9]. У цій роботі було розроблено механізм кооперації однорідних агентів у рамках багатоагентного навчання з підкріпленням. KnowSR базується на концепції дистиляції знань, що дозволяє агентам не обмежуватися лише власним досвідом, а також інтегрувати інформацію, отриману іншими агентами. Це забезпечує скорочення часу, необхідного для досягнення прийнятної політики, та підвищує ефективність навчання загалом. З технічної точки зору, метод був реалізований на базі алгоритму MADDPG (Multi-Agent Deep Deterministic Policy Gradient), який передбачає централізоване навчання із використанням критика, що має доступ до глобальної інформації, та децентралізоване виконання дій кожним агентом окремо. Така архітектура дозволяє ефективно поєднати індивідуальне та колективне навчання, що є важливим у контексті складних багатоагентних середовищ [9].

Таким чином, результати останніх досліджень підтверджують перспективність підходів до обміну знаннями для прискорення та покращення процесів навчання в багатоагентних системах. Надалі важливим завданням є подальший розвиток методів безперервного обміну знаннями в динамічних та нестаціонарних середовищах.

Постановка завдання. Середовище, в якому функціонує багато агентів, часто є нестаціонарним, що значно ускладнює процес навчання та призводить до зниження ефективності класичних DQN-агентів. Це зумовлено тим, що алгоритм DQN спочатку розроблявся для розв'язання завдань в умовах стаціонарних середовищ і не враховує впливу дій інших агентів, які змінюють динаміку середовища та, відповідно, впливають на формування винагород. Проте існують середовища, які залишаються стаціонарними або можуть розглядатися як такі за певних припущень, навіть за наявності декількох агентів. Прикладом таких систем є фінансові торгові середовища, у яких дії окремого агента мають незначний вплив на загальний стан середовища і, в більшості випадків, не змінюють поведінку інших агентів.

У зв'язку з цим, завданням даного дослідження є розробка механізму обміну знаннями між агентами у стаціонарному середовищі, аналіз його впливу на ефективність системи, а також вивчення особливостей програмної реалізації такого підходу в умовах безперервного навчання агентів.

Виклад основного матеріалу. Було проведено серію експериментів з дослідження підходу до обміну знаннями між агентами. Експеримент проводився у класичному середовищі OpenAI Gym CartPole-v1 [10]. Завдання полягає в тому, щоб утримувати жердину (Pole) у вертикальному положенні на рухомій платформі (Cart), балансуючи нею шляхом застосування сил вліво або вправо. Агент отримував нагороду +1 за кожну ітерацію, доки жердина залишалася у межах допустимого кута нахилу (±24 градусів) та діапазону положення (± 4.8 одиниці). Додатково було обмежено кількість кроків у 500 максимум. Для представлення Q-функції використовувалася глибока нейромережа з двома прихованими шарами (24 нейрони, функція активації ReLU, 12 нейронів, функція активації ReLU). Відповідно, вхідний шар – 4 нейрони (відповідають 4-елементному вектору стану середовища), і вихідний шар – 2 нейрони (Q-значення для кожної з двох можливих дій: рух платформи ліворуч або праворуч). Функція втрат – середньоквадратична похибка (MSE), яка використовується для оновлення оцінок Q-функції. Оптимізація параметрів здійснювалася за допомогою алгоритму Adam (Adaptive Moment Estimation) із коефіцієнтом навчання 0.001 [11]. Batch size дорівнює 128. Значення discount factor у рівнянні Беллмана було встановлено на 0.95, що дозволяє агенту враховувати майбутні винагороди при прийнятті рішень [12]. Для розв'язання дилеми дослідження-експлуатації використовувався Adaptive ε-greedy exploration підхід [13], де параметр ε зменшувався з 1 (повністю випадкові дії агента) на 0.5% з кожним епізодом. Для покращення стабільності навчання застосовувався непріорітезований буфер досвіду, що зберігав до 10 тисяч останніх переходів (стан, дія, винагорода, наступний стан, завершення епізоду). У кожній імітації навчалися 3 агенти, що діяли у власних середовищах та обмінювалися досвідом за одним з запропонованих нижче методів.

Обмін знаннями між вчителем і учнем. Запропонований підхід умовно названий «Обмін знаннями між вчителем і учнем» (рис. 1), базується на ідеї дистиляції політики. У межах цього підходу агент, обраний системою як вчитель, передає свої знання агенту-учню, який сприймає їх як еталон.

Основною ідеєю методу є припущення, що наближення розподілу політики учня до політики вчителя позитивно вплине на його показники. При цьому повне узгодження розподілів

не є метою; передбачається часткова дистиляція, яка зберігає індивідуальні відмінності агентів та дозволяє змінювати домінуючу стратегію у різних агентів у процесі навчання.

Регулювання інтенсивності навчання здійснювалося через кількість епох оптимізації. Базове значення становило 3 епохи. Залежно від співвідношення сумарної винагороди вчителя та учня, кількість епох коригувалася шляхом множення на відношення їх нагород, однак не перевищувала 5 епох. Таким чином, якщо агент-учень демонстрував суттєво гірші результати, сила його узгодження із вчителем збільшувалася. Для агентів із близькими показниками вплив навчання залишався мінімальним. Варто зауважити, що обрані параметри є типовими для проведеного експерименту, однак можуть змінюватися залежно від характеристик середовища та архітектури нейронних мереж.

Вибір вчителя здійснювався на основі порівняння сумарних нагород, отриманих агентами за попередній епізод: агент із вищою нагородою вважався краще адаптованим до середовища та визначався вчителем.

Для перевірки ефективності методу «Учень–вчитель» було створено спрощену експериментальну систему, у якій після завершення епізоду кожен агент очікував завершення епізодів усіх інших агентів. Після цього визначалися ролі вчителя та учня, а також сила навчального впливу. За результатами експериментів встановлено, що запропонований підхід сприяє прискоренню процесу навчання агентів (рис. 2).

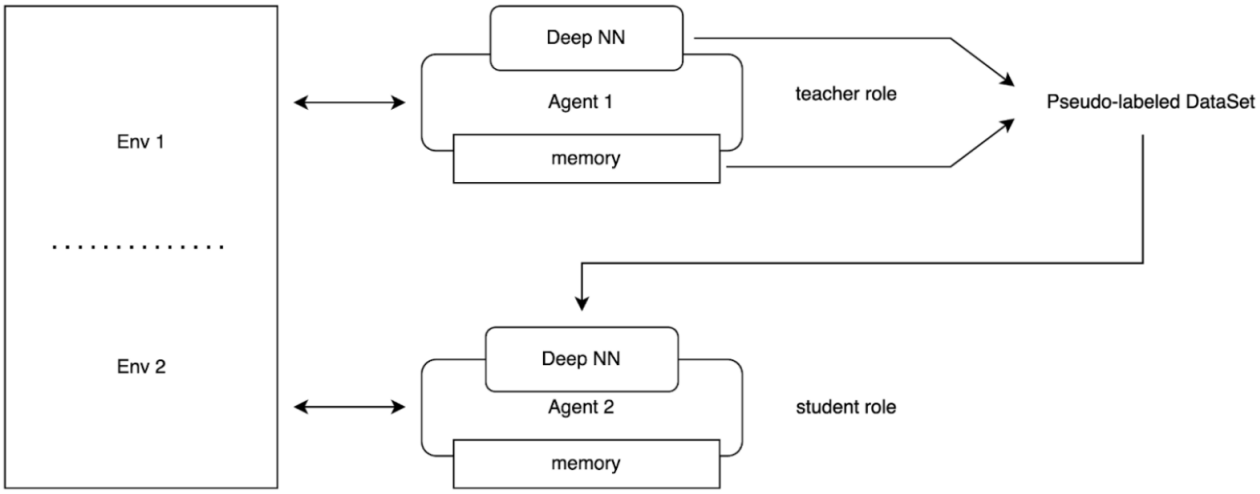


Рис. 1. Схема обміну знаннями між учнем і вчителем

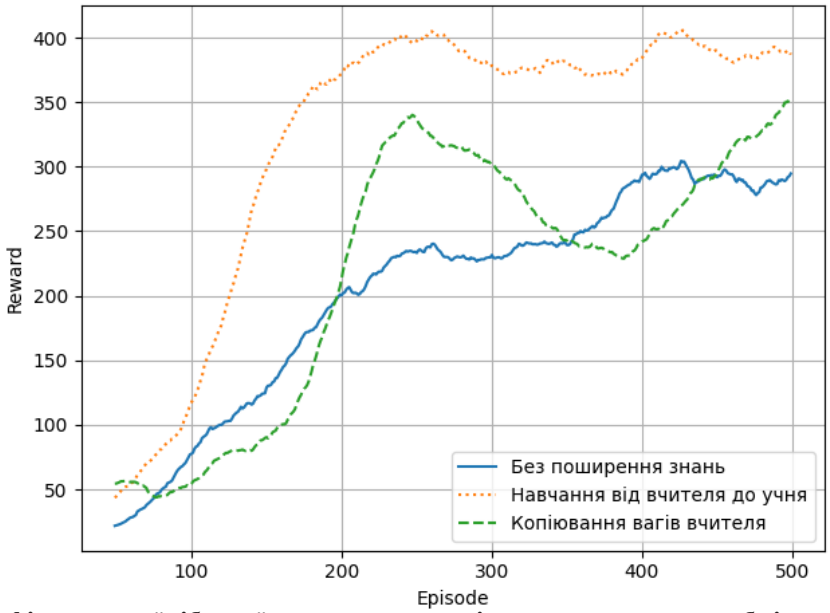


Рис. 2. Графік сумарної зібраної винагороди за епізод з застосуванням обміну знаннями між учнем та вчителем (графік зглажений скользячим середнім з вікном у 50)

Важливо зазначити, що запропонований метод використовує показник зібраної винагороди для визначення більш впливових даних, а отже навчання на власному досвіді має відбуватися після поширення знань, для збереження оціненої моделі. Також, досліди показують, що навчання на власному досвіді має бути так само присутнім, як і обмін знаннями. Більше того, навчання на власному досвіді має бути більше, ніж поширення знань. Інакше, це приводить до сповільненого росту показників, нестабільності або ж деградації.

Особливості неблокуючого обміну знаннями. У реальних багатоагентних системах тривалість епізодів для різних агентів може істотно відрізнятися. Це створює проблему синхронізації: примусове очікування завершення епізодів іншими агентами може призвести до блокування роботи системи, що є неприйнятним в умовах реального часу або розподілених обчислень. Відтак виникає потреба в реалізації алгоритму обміну знаннями, що не вимагає синхронної взаємодії агентів і забезпечує безперервність їх роботи.

Запропонований асинхронний механізм обміну знаннями має наступний алгоритм функціонування:

1. Накопичення досвіду. Під час взаємодії з середовищем агент накопичує дані у власному буфері досвіду. Кожен запис містить інформацію про стан, обрану дію, отриману нагороду, наступний стан та ознаку завершення епізоду.

2. Обробка після завершення епізоду. Після завершення власного епізоду агент виконує наступні дії:

   1. Обчислює сумарну нагороду, отриману за весь епізод.
   2. Випадковим чином обирає певну кількість записів зі свого буфера та обчислює для них передбачення Q-функції.
   3. Формує повідомлення, яке містить обрану вибірку, передбачені Q-значення та сумарну нагороду, і публікує його для інших агентів.

3. Отримання знань від інших агентів. Агент отримує повідомлення, надіслані іншими агентами. Якщо серед цих повідомлень є такі, де сумарна нагорода перевищує власну, агент визначає відповідного агента як вчителя.

4. Навчання. Після визначення вчителя агент додатково навчається на псевдо-розміченому наборі даних, наданому вчителем. Окрім цього, продовжується навчання на власному досвіді, накопиченому у буфері.

Такий підхід (рис. 3) дозволяє організувати процес обміну знаннями без синхронізації агентів, мінімізуючи затримки у навчанні та забезпечуючи безперервність роботи багатоагентної системи.

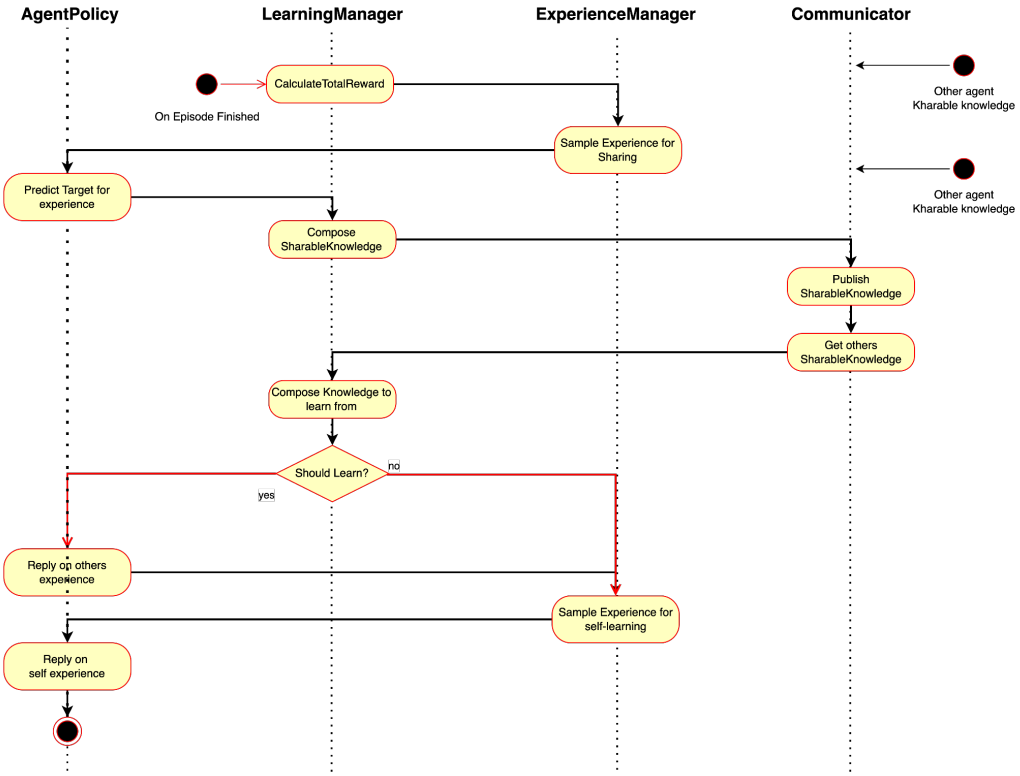


Рис. 3. Схема асинхронного обміну знаннями за схемою Учень-вчитель

Конкретна реалізація механізму обміну повідомленнями залежить від архітектури багатоагентної системи та організації середовища. Існує кілька можливих підходів до побудови комунікатора:

• Комунікатор на рівні середовища. У цьому варіанті роль комунікаційного вузла виконує саме середовище. Воно приймає повідомлення від агентів, зберігає їх і забезпечує доступ для читання іншими агентами. Такий підхід спрощує логіку кожного окремого агента, оскільки обмін знаннями відбувається через централізовану проміжну ланку.

• Спеціалізований агент-комунікатор. Альтернативним варіантом є виділення окремого агента, який відповідає за прийом, зберігання та розповсюдження повідомлень. Інші агенти передають йому записи для публікації та отримують повідомлення у відповідь. Це дозволяє відокремити функції комунікації від основної логіки агентів і забезпечити масштабованість системи.

• Прямий обмін повідомленнями між агентами. У цьому випадку агенти зв›язуються попарно напряму без використання центрального вузла. Кожен агент самостійно приймає, зберігає та оновлює актуальні версії отриманих повідомлень. Під час публікації даних агент самостійно знаходить інших доступних агентів і передає їм свої повідомлення. Можлива також організація передачі за принципом «ланцюга», коли агент передає дані іншому, який у свою чергу ретранслює їх далі.

У всіх описаних варіантах ключовою особливістю є відсутність блокування агентів для очікування синхронізації. Кожен агент продовжує навчання незалежно від інших, що забезпечує асинхронність роботи системи та високу ефективність у реальному часі.

Висновки. У цій роботі було досліджено підходи до підвищення ефективності навчання в багатоагентних системах шляхом обміну знаннями між агентами. Встановлено, що поряд із вдосконаленням індивідуальних алгоритмів навчання (наприклад, модифікацій DQN), впровадження механізмів обміну знаннями є перспективним напрямом для прискорення збіжності та підвищення стабільності навчального процесу.

На основі аналізу існуючих підходів, зокрема методів дистиляції політики та обміну досвідом (Policy Distillation, KnowSR), запропоновано дві модифікації організації навчання:

• Модель «вчитель–учень», у якій обраний агент-вчитель передає знання агенту-учню з контролем сили навчання залежно від відставання в сумарній винагороді. Проведені експерименти продемонстрували, що навіть часткове узгодження розподілів політик сприяє прискоренню навчання.

• Асинхронний механізм обміну повідомленнями, який дозволяє агентам обмінюватися досвідом без необхідності синхронізації завершення епізодів. Це забезпечує безперервність роботи системи та робить підхід придатним для застосування в реальних розподілених середовищах.

Розглянуто також можливі варіанти реалізації системи обміну знаннями: через середовище, через окремого агента-комунікатора та за допомогою прямого обміну повідомленнями між агентами. Кожен варіант має свої переваги та обмеження залежно від архітектури середовища та вимог до масштабованості.

Список літератури:

1. Wooldridge M. An Introduction to MultiAgent Systems. John Wiley & Sons, 2002. 366 с.
2. Tichý P., Staron R.J. Multi-agent technology for fault tolerant and flexible control. In: D. Srinivasan, L.C. Jain (eds.). Innovations in Multi-Agent Systems and Applications – 1. Studies in Computational Intelligence. Vol. 310. Berlin, Heidelberg: Springer, 2010. С. 223–246.
3. Sutton R.S., Barto A.G. Reinforcement Learning: An Introduction. 2nd ed. Cambridge, MA: MIT Press, 2018. 552 с.
4. Mnih V., Kavukcuoglu K., Silver D. et al. Human-level control through deep reinforcement learning. Nature. 2015. Vol. 518. No. 7540. С. 529–533.
5. Van Hasselt H., Guez A., Silver D. Deep reinforcement learning with double Q-learning. Proc. of the 30th AAAI Conf. on Artificial Intelligence. 2016. С. 2094–2100.
6. Schaul T., Quan J., Antonoglou I., Silver D. Prioritized experience replay. Proc. of 4th Int. Conf. on Learning Representations (ICLR). 2016. 9 с.
7. Hernandez-Leal P., Kaisers M., Baarslag T., Muñoz de Cote E. A survey of learning in multiagent environments: dealing with non-stationarity. arXiv preprint arXiv:1707.09183. 2019. 64 с. URL: https://arxiv.org/abs/1707.09183.

8. Rusu A.A., Colmenarejo S.G., Gulcehre C. et al. Policy Distillation [Електронний ресурс]. arXiv preprint arXiv:1511.06295. 2015. URL: https://doi.org/10.48550/arxiv.1511.06295.

9. Gao Z., Xu K., Ding B. et al. KnowSR: Knowledge Sharing Among Homogeneous Agents in Multi-agent Reinforcement Learning. arXiv preprint arXiv:2105.11611. 2021. URL: https://doi.org/10.48550/arxiv.2105.11611.

10. Brockman G., Cheung V., Pettersson L. et al. OpenAI Gym. arXiv preprint arXiv:1606.01540. 2016. URL: https://doi.org/10.48550/arxiv.1606.01540.

11. Kingma D.P., Ba J. Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980. 2014. URL: https://doi.org/10.48550/arxiv.1412.6980.

12. Sutton R.S., Barto A.G. Reinforcement Learning: An Introduction. MIT Press, 2018. 552 c.

13. Tokic M. Adaptive ε-greedy exploration in reinforcement learning based on value differences. Annual Conf. on Artificial Intelligence. 2010.

Bochok V.O., Fedorova N.V. KNOWLEDGE SHARING IN INDEPENDENT DQN MULTI-AGENT SYSTEM AND FEATURES OF ITS IMPLEMENTATION

This paper addresses the problem of improving agent learning efficiency in multi-agent systems by introducing a knowledge-sharing mechanism. It is noted that traditional approaches based on Deep Q-Network (DQN) algorithms mainly focus on utilizing each agent's own experience, which limits the convergence rate, reduces the quality of the obtained strategies, and prevents the system from fully realizing its potential. At the same time, multi-agent systems have significant potential for performance improvement through knowledge sharing among agents, enabling more efficient utilization of available experience resources.

A knowledge-sharing mechanism based on the "teacher–student" model is proposed, which utilizes policy distillation techniques. The agent with the highest performance is designated as the teacher and shares its knowledge with the student agent, considering the difference in their cumulative episode rewards. The learning intensity is regulated based on the ratio of obtained rewards, allowing agents to preserve individuality, gradually adapt their strategies, and apply more intensive learning for agents with significant differences in performance, using the reward as an objective quality indicator.

Special attention is given to the implementation of asynchronous knowledge sharing without blocking agents. The proposed algorithm enables agents to independently accumulate experience, publish messages with generalized policy information, and select an appropriate teacher without the need for process synchronization or halting other agents' activities. This ensures continuous learning, improves system flexibility, and enhances the overall efficiency of operations in distributed computational environments.

Experimental studies conducted in the OpenAI Gym CartPole-v1 environment using a set of DQN agents confirmed the effectiveness of the proposed approach. The results demonstrate that knowledge sharing based on the "teacher–student" model accelerates convergence, improves policy stability, increases overall system performance, and reduces the time required to achieve target quality metrics.

Key words: *Deep Q-Network (DQN), Knowledge Sharing, Asynchronous Communication, Multi-Agent Systems, Data Analytics, Reinforcement Learning.*

УДК 004.89:336.7
DOI https://doi.org/10.32782/2663-5941/2025.3.2/16

Вакалюк Т.А.
Державний університет «Житомирська політехніка»

Антонюк Д.С.
Державний університет «Житомирська політехніка»

Мінтій І.С.
Лодзький університет у Польщі

Новіцька І.В.
Житомирський державний університет імені Івана Франка

Довгалюк І.С.
Державний університет «Житомирська політехніка»

МЕТОДИ ПОБУДОВИ РЕКОМЕНДАЦІЙНИХ СИСТЕМ ТА ІНТЕГРАЦІЯ ШТУЧНОГО ІНТЕЛЕКТУ У ФІНАНСОВИХ ДОДАТКАХ

У статті здійснено комплексне дослідження методів побудови рекомендаційних систем у фінансових додатках та особливостей інтеграції штучного інтелекту в сучасні фінансові платформи. Проаналізовано основні підходи до формування персоналізованих рекомендацій, зокрема колаборативну фільтрацію, яка виявляє схожі моделі споживання користувачів, контентну фільтрацію, що враховує характеристики обраних продуктів, а також гібридні методи, які поєднують переваги обох підходів для підвищення точності прогнозування фінансової поведінки. Досліджено використання нейронних мереж для врахування не лише історичних даних, але й контекстуальних факторів, що підвищує релевантність рекомендацій та довіру користувачів. Розглянуто технічні аспекти інтеграції штучного інтелекту в мобільні платформи з використанням спеціалізованих SDK (TensorFlow Lite, Core ML) та веб-додатки через хмарні сервіси (AWS AI, Google Cloud AI, Microsoft Azure AI), а також можливості застосування сторонніх API для розширення функціональності фінансових додатків. Проаналізовано різноманітні сфери застосування AI у фінансових додатках: автоматизація рутинних операцій, прогнозування витрат, виявлення аномалій у транзакціях, персоналізація інвестиційних продуктів та кредитних програм, оптимізація бюджету та поліпшення клієнтського досвіду через використання AI-асистентів. Висвітлено проблеми захисту персональних даних, необхідність забезпечення прозорості алгоритмів, уникнення упередженості та дискримінації як ключові етичні аспекти використання рекомендаційних систем. Практична значущість дослідження полягає у систематизації методів та підходів до створення інтелектуальних фінансових помічників, які здатні не лише аналізувати індивідуальні фінансові патерни, але й передбачати загальні економічні тренди, що сприяє підвищенню якості обслуговування клієнтів, зниженню витрат компаній та оптимізації процесу прийняття фінансових рішень.

Ключові слова: *рекомендаційні системи, фінансові додатки, штучний інтелект, персоналізація, мобільні платформи, веб-платформи, машинне навчання, фінансові помічники, AI-асистенти, фінансові прогнози.*

Постановка проблеми. Методи побудови рекомендаційних систем у фінансових додатках є важливим елементом сучасних цифрових рішень, що сприяють підвищенню ефективності роботи, оптимізації обробки великих обсягів даних та забезпеченню високого рівня персоналізації послуг. Рекомендаційні системи базуються на алгоритмах машинного навчання, штучного інтелекту та аналізу великих даних (Big Data), що дозволяє автоматизувати прийняття рішень і підвищити точність прогнозів. Колаборативна фільтрація аналізує поведінку користувачів і виявляє схожі моделі споживання, пропонуючи послуги чи продукти, популярні серед корис-

тувачів зі схожими звичками. Контентна фільтрація, у свою чергу, враховує характеристики продуктів або послуг, обраних користувачем, і генерує рекомендації, що відповідають його уподобанням.

Гібридні підходи, які поєднують переваги колаборативної та контентної фільтрації, забезпечують створення більш точних рекомендацій. Методи, що використовують нейронні мережі, дозволяють враховувати не тільки історичні дані, але й контекстуальні фактори, підвищуючи релевантність рекомендацій. Наприклад, аналіз транзакційних даних у реальному часі дозволяє пропонувати індивідуалізовані фінансові рішення, такі як оптимізація витрат, планування заощаджень або інвестицій. Алгоритми прогнозування забезпечують можливість передбачати майбутні фінансові дії користувачів на основі аналізу їхньої попередньої поведінки, що дозволяє ефективніше управляти бюджетом або портфелем активів [1, 2].

Аналіз останніх досліджень і публікацій. У банківському секторі та інвестиційних сервісах рекомендаційні системи можуть бути використані для персоналізації пропозицій, таких як кредитні лінії, страхові продукти або інвестиційні стратегії. Програмні засоби управління особистими фінансами можуть прогнозувати витрати, оцінювати платоспроможність користувачів або навіть попереджати про можливі фінансові ризики, ґрунтуючись на аналізі поточних даних. У [8] розглядають застосування штучного інтелекту для підвищення ефективності фінансових рішень. Вони підкреслюють важливість аналізу фінансової поведінки, виявлення аномалій та надання персоналізованих рекомендацій для оптимізації фінансової поведінки користувачів. Особливу увагу приділено етичним аспектам, таким як забезпечення конфіденційності, справедливості та прозорості, а також підтримка автономії користувача у прийнятті фінансових рішень.

Водночас використання рекомендаційних систем супроводжується низкою викликів. Захист персональних даних, прозорість алгоритмів, уникнення упередженості та дискримінації – це ключові аспекти, які необхідно врахувати для забезпечення довіри клієнтів та ефективності рішень. Інтеграція рекомендаційних систем у фінансові додатки повинна враховувати як технічні, так і етичні аспекти. У [9] досліджує методи, моделі та алгоритми персоналізації в цифровому середовищі. Вона надає загальну характеристику типів рекомендаційних систем, їх методів, моделей та алгоритмів, а також переваг та недоліків їх застосування. Авторка пропонує застосування таксономії знань для створення інтелектуального додатку персоналізації, що підтримує прийняття маркетингових рішень для підприємств у цифровому середовищі.

Загалом, рекомендаційні системи у фінансових додатках сприяють не лише зменшенню витрат компаній, але й покращенню якості обслуговування клієнтів, допомагаючи їм приймати більш обґрунтовані фінансові рішення. Такі системи відповідають сучасним тенденціям цифрової економіки, відкриваючи нові можливості для автоматизації процесів та індивідуалізації фінансових послуг. У [10] розглядають основні принципи побудови рекомендаційної системи та методи вирішення проблеми холодного старту, яка виникає внаслідок недостатньої взаємодії користувача з програмним засобом на початкових етапах роботи з ним. Вони пропонують способи збільшення ефективності роботи рекомендаційної системи за умов недостатньої вибірки даних та при появі в системі нових елементів, для яких відсутня статистика.

Постановка завдання. Метою статті є дослідження та систематизація методів побудови рекомендаційних систем у фінансових додатках, аналіз особливостей інтеграції штучного інтелекту в мобільні та веб-платформи фінансового призначення, а також визначення основних підходів і технологій, що використовуються для розробки фінансових помічників на основі штучного інтелекту.

Виклад основного матеріалу. Впровадження рекомендаційних систем у фінансові додатки відкриває можливості для прогнозування не лише фінансових рішень, але й загальних економічних трендів. Наприклад, системи можуть передбачати зміни в попиті на певні фінансові продукти, враховуючи макроекономічні фактори, зміни на ринку або глобальні економічні коливання. Цей підхід дозволяє компаніям не тільки адаптуватися до змін, але й проактивно пропонувати користувачам продукти, що можуть відповідати їхнім потребам у майбутньому. Такі прогнози можуть базуватися на поточних тенденціях, історичних даних та факторів, що впливають на ринки, що дозволяє досягти ще більшої персоналізації фінансових рішень [3].

У майбутньому з розвитком технологій та поглибленням інтеграції штучного інтелекту, рекомендаційні системи зможуть використову-

вати ще більше джерел даних – від соціальних мереж до аналізу споживчих моделей у режимі реального часу. Це дозволить створювати ще точніші моделі прогнозування та персоналізації. Проте такі можливості також піднімають питання етики, конфіденційності та безпеки даних, що вимагає ще більш ретельного регулювання та забезпечення прав користувачів на захист своїх персональних даних у цифровому середовищі [4, 5].

Інтеграція штучного інтелекту (AI) у мобільні та веб-платформи відкриває нові горизонти для фінансових установ, компаній та розробників. Вона дає можливість створювати гнучкі, персоналізовані та інноваційні сервіси, здатні задовольнити високі очікування сучасних користувачів. Завдяки AI платформи стають не лише інструментами для виконання фінансових операцій, але й розумними помічниками, що можуть передбачати потреби клієнтів, автоматизувати рутинні процеси та забезпечувати високий рівень безпеки [6].

Інтеграція AI дозволяє аналізувати поведінку користувачів у реальному часі, що дає можливість створювати індивідуалізовані пропозиції для кожного користувача. Застосовуючи алгоритми машинного навчання, платформи можуть надавати рекомендації щодо інвестиційних продуктів, орієнтуючись на історію транзакцій та фінансові цілі клієнта. Також системи можуть допомогти в плануванні бюджету та надавати пропозиції з оптимізації витрат, враховуючи щоденні витрати користувача. Крім того, AI дозволяє створювати персоналізовані кредитні програми, що ґрунтуються на кредитній історії клієнта.

Завдяки персоналізації клієнти отримують послуги, що найбільше відповідають їхнім потребам. Це сприяє підвищенню рівня лояльності, адже користувачі відчувають, що їхні фінансові цілі і потреби враховуються при наданні послуг.

Веб- і мобільні додатки з штучним інтелектом здатні автоматизувати складні фінансові операції, що значно полегшує життя користувачів. Наприклад, системи можуть автоматично генерувати податкові звіти на основі отриманих і витрачених коштів, що значно спрощує процес обліку. Також додатки можуть надавати нагадування про майбутні платежі або заплановані транзакції, що дозволяє уникнути пропущених термінів [7].

AI може автоматично сортувати фінансові операції за категоріями, що допомагає користувачам краще розуміти свої витрати та доходи. Завдяки цьому автоматизація рутинних завдань звільняє користувачів від необхідності вручну виконувати ці дії, даючи змогу зосередитись на більш важливих завданнях.

Використання AI-асистентів, таких як чат-боти, значно полегшує комунікацію з клієнтами. Ці віртуальні помічники здатні надавати відповіді на поширені запитання в будь-який час доби, що дозволяє компаніям підтримувати клієнтів 24/7. Крім того, вони можуть допомогти у вирішенні складних завдань, таких як відновлення доступу до рахунку або розрахунок кредиту, що робить процес взаємодії з банком чи фінансовою установою більш зручним для користувачів.

AI-асистенти також можуть інформувати клієнтів про актуальні акції чи нові продукти, що допомагає компаніям залишатись у полі зору споживачів та збільшувати лояльність до своїх послуг. Завдяки таким можливостям покращується рівень обслуговування клієнтів, а компанії знижують витрати на підтримку.

Технології штучного інтелекту дають змогу мобільним та веб-платформам аналізувати величезні обсяги даних в реальному часі. Це відкриває нові можливості для виявлення аномалій у фінансових транзакціях, що може вказувати на шахрайство, і дозволяє оперативно реагувати на потенційні загрози. Такі можливості є важливими для забезпечення безпеки і стабільності фінансових систем.

Крім того, штучний інтелект дозволяє створювати точні прогнози ринкових трендів для інвесторів, що допомагає приймати обґрунтовані фінансові рішення. Важливим аспектом є також можливість оперативно реагувати на зміни в поведінці користувачів, наприклад, пропонуючи персоналізовані акційні програми чи бонуси, що підвищує зацікавленість клієнтів і підтримує їхню лояльність.

Інтеграція штучного інтелекту в мобільні додатки вимагає дотримання низки технічних стандартів, щоб забезпечити ефективність і безпеку роботи програм. Одним із ключових моментів є використання спеціальних SDK (Software Development Kit), таких як TensorFlow Lite або Core ML, для того, щоб забезпечити локальну обробку даних без необхідності передавати їх на сервери. Це дозволяє зберігати конфіденційність інформації та знижує навантаження на мережу.

Крім того, важливо інтегрувати API для хмарних сервісів, що дозволяє виконувати більш

складні обчислення безпосередньо в хмарі. Це зручно, оскільки дозволяє зменшити навантаження на мобільний пристрій, що особливо важливо при роботі з великими обсягами даних. Також необхідно оптимізувати AI-моделі для мобільних пристроїв з обмеженими обчислювальними потужностями, щоб забезпечити їх швидку та ефективну роботу навіть на менш потужних гаджетах.

Що стосується веб-платформ, то тут AI найчастіше інтегрується через хмарні сервіси, такі як AWS AI, Google Cloud AI чи Microsoft Azure AI. Це дозволяє реалізувати машинне навчання для аналізу даних в реальному часі, що є важливим для фінансових і бізнес-додатків, де необхідно швидко обробляти велику кількість запитів.

Завдяки використанню алгоритмів глибокого навчання можна будувати точні фінансові прогнози, що дозволяє компаніям приймати обґрунтовані рішення на основі великих обсягів даних. Хмарні платформи також забезпечують високу продуктивність, навіть коли на веб-платформі працює велика кількість користувачів одночасно, що робить їх ідеальним рішенням для масштабованих систем.

Окрім внутрішньої інтеграції AI, можна використовувати сторонні API, що значно розширює можливості мобільних і веб-платформ. Це дозволяє інтегрувати сторонні сервіси для обробки тексту, зображень, відео або аудіо. Наприклад, використовуючи Google Vision API або IBM Watson API, можна безпечно й ефективно обробляти різноманітний мультимедійний контент без необхідності розробляти власні складні алгоритми для таких завдань.

Крім того, завдяки стороннім API, можна отримати доступ до спеціалізованих моделей для обробки природної мови (NLP) або розпізнавання емоцій. Це дозволяє використовувати готові рішення, що значно скорочує час на розробку таких функцій, оскільки немає потреби створювати і навчати ці моделі з нуля. Це дає можливість застосовувати складні алгоритми у фінансових, медичних чи інших галузях без великих витрат на дослідження і розробку.

Також сторонні API можуть забезпечити готові рішення для конкретних завдань, наприклад, для інтеграції чат-ботів або рекомендаційних систем. Це значно спрощує процес розробки і впровадження нових функцій у додатках, що дозволяє зменшити час, витрачений на розробку, і збільшити швидкість виходу продукту на ринок.

Висновки. У статті висвітлено методи побудови рекомендаційних систем, які в фінансових додатках допомагають користувачам ухвалювати обґрунтовані рішення на основі аналізу їхніх фінансових звичок, цілей та потреб. Розглянуто як фільтрація за змістом, так і колаборативні методи, що забезпечують точність рекомендацій.

Інтеграція ШІ у мобільні та веб-платформи є важливою темою, оскільки вона дозволяє фінансовим помічникам бути доступними для користувачів на різних пристроях, зберігаючи високий рівень інтерактивності та функціональності.

Технічна реалізація таких систем вимагає комплексного підходу з використанням спеціалізованих SDK для мобільних платформ та інтеграції хмарних сервісів для веб-додатків, що забезпечує оптимальну продуктивність та безпеку даних користувачів. Залучення сторонніх API суттєво розширює функціональні можливості фінансових додатків, дозволяючи інтегрувати передові технології обробки природної мови, розпізнавання образів та інші досягнення у сфері штучного інтелекту.

Узагальнюючи, дане дослідження надає глибоке розуміння основних підходів і технологій, що використовуються для розробки фінансових помічників на основі штучного інтелекту, і визначає важливі аспекти їх реалізації в реальних умовах.

Список літератури:

1. Solovii Kh., Horodyska N. Using of artificial intelligence in the field of financial services. III International Scientific and Practical Conference of Young Scientists and Students ‘Philosophical Dimensions of Technology’ (PDT-2022). P. 87–89. URL: https://elartu.tntu.edu.ua/bitstream/lib/39709/2/PDT_2022_Solovii_Kh-Using_of_artificial_intelligence_87-89.pdf
2. How to use ChatGPT API in php. URL: https://www.educative.io/answers/how-to-use-chatgpt-api-in-php
3. What is artificial intelligence (AI) in finance? URL: https://cloud.google.com/discover/finance-ai
4. How Artificial Intelligence is Transforming the Financial Services Industry. URL: https://surl.lu/qthqqk
5. 21 Examples if AI in Finance. URL: https://surl.li/pbumpq
6. Naqvi, H. A. Artificial Intelligence in Finance: Challenges, Opportunities and Regulatory Developments. 2022. Springer Publishing. URL: https://surli.cc/bvmvcf

7. Liu, Zechen. Artificial Intelligence as Personal Financial Advisor in the Future? – A Case Study Based on Algorithmic Innovation Strategies. Advances in Economics, Management and Political Sciences. 87. 2024. P. 272–278. 10.54254/2754-1169/87/20241030.

8. Shuryhin K.A., Zinovatna S.L. Recommendation system for financial decision-making using Artificial intelligence. Applied Aspects of Information Technology. 2024; Vol. 7, No. 4: 348–358. DOI: 10.15276/aait.07.2024.24.

9. Yuliia Kotliarova. Analysis of methods, models and algorithms of personalization for the recommender systems development. Scientific Journal "ScienceRise", Vol. 12, 2018. P. 19–29. DOI: 10.15587/2313-8416.2018.152949

10. Даниленко М., Колесник І. Методи розробки рекомендаційних систем. Інформаційні технології та комп'ютерна інженерія, 52 (3), 2021, 10–15. https://doi.org/10.31649/1999-9941-2021-52-3-10-15

Vakaliuk T.A., Antoniuk D.S., Mintii I.S., Novitska I.V., Dovgaliuk I.S. METHODS FOR BUILDING RECOMMENDATION SYSTEMS AND INTEGRATING ARTIFICIAL INTELLIGENCE INTO FINANCIAL APPLICATIONS

The article provides a comprehensive study of methods for building recommendation systems in financial applications and the specifics of integrating artificial intelligence into modern financial platforms. The main approaches to the formation of personalised recommendations are analysed, in particular collaborative filtering, which identifies similar user consumption patterns, content filtering, which takes into account the characteristics of selected products, as well as hybrid methods that combine the advantages of both approaches to improve the accuracy of financial behaviour prediction. The use of neural networks to take into account not only historical data but also contextual factors is investigated, which increases the relevance of recommendations and user trust. Technical aspects of integrating artificial intelligence into mobile platforms using specialised SDKs (TensorFlow Lite, Core ML) and web applications via cloud services (AWS AI, Google Cloud AI, Microsoft Azure AI) are considered, as well as the possibility of using third-party APIs to extend the functionality of financial applications. Various areas of AI application in financial applications are analysed: automation of routine operations, cost forecasting, detection of anomalies in transactions, personalisation of investment products and credit programmes, budget optimisation and improvement of customer experience through the use of AI assistants. The problems of personal data protection, the need to ensure algorithm transparency, and the avoidance of bias and discrimination are highlighted as key ethical aspects of the use of recommendation systems. The practical significance of the study lies in the systematisation of methods and approaches to creating intelligent financial assistants that are capable of not only analysing individual financial patterns but also predicting general economic trends, which contributes to improving customer service, reducing company costs and optimising the financial decision-making process.

Key words: *recommendation systems, financial applications, artificial intelligence, personalisation, mobile platforms, web platforms, machine learning, financial assistants, AI assistants, financial forecasts.*

UDC 004.4
DOI https://doi.org/10.32782/2663-5941/2025.3.2/17

Vernik M. O.
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"

Oleshchenko L. M.
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"

INTELLIGENT SOFTWARE FRAMEWORK FOR BIG DATA AUDIO PROCESSING AND COGNITIVE ENERGY ESTIMATION

The article presents an innovative software framework for the analysis of human-produced audio aimed at quantifying cognitive energy. A theoretical model has been proposed for representing daily cognitive energy (E), time (T), and actions (D1...Dn), with a distinction between productive and unproductive activities, as well as the identification of a specific "flow state." A new metric, Countable Cognitive Energy (CCE), has been introduced, derived from transcribed audio of daily tasks. The technical system architecture has been described, which includes continuous audio recording (~16 hours/day, ~1.2 GB/day per user) via wearable devices, transcription using Whisper AI, and segmentation and labeling of data for further analysis. The software system comprises a React-based interface for visualization and upload, Azure Blob Storage for scalable data storage, Azure Functions for transcription orchestration and NLP processing, a .NET API for access control, and a Python backend for speech and language processing. The architectural design of developed software has been illustrated with data flow diagrams and wearable device schematics. Key limitations of proposed software system have been examined, such as transcription delays, GPU constraints, and bandwidth limitations. The scalability of the system has been analyzed. The proposed approach has been compared to existing software solutions (LENA, TILES Audio Recorder, Friend wearable), with an emphasis on its ability to visualize user productivity patterns and detect flow states. Relevant literature in lifelogging and productivity research has been reviewed. The prospects for future research include improving NLP-based classification of activities, validating CCE through user studies, and incorporating additional contextual sensors such as computer usage and geolocation. Potential extensions of the framework to distinguish between cognitive and physical energy have been discussed.

Key words: *software framework, big data, audio, mp3, wav, cognitive energy, human-produced data, recorder, NLP, Whisper AI, Python, Azure Blob Storage.*

Statement of the problem. Continuous self-monitoring through lifelogging offers unprecedented outcomes into personal productivity, time use, and mental state. Motivated by the idea of Maximum Productivity, a conceptual framework positing that individuals have finite daily energy (*E*) and time (*T*) which must be optimally allocated among actions $D_{1...n}$ – we propose a holistic system merging theory and practice. Philosophical ideas like energy depletion and "flow state" productivity gains marketingmag.com.au are quantified via continuous audio lifelogging using a wearable device.

Prior works such as the LENA device (Language ENvironment Analysis) [1] have demonstrated the feasibility and research value of all-day audio recording in specific contexts. More recent developments, including wearable "always listening" assistants (Bee AI, Omi), underline emerging commercial interest in passive audio capture for personal analytics [2].

Analysis of recent research and publications. Early lifelogging visions like MyLifeBits and Total Recall advocated capturing one's entire life digitally. Continuous audio recording has been explored via devices like LENA (for child language environments) and research prototypes like the Electronically Activated Recorder (EAR). The TILES Audio Recorder (TAR) introduced a wearable sensor and mobile app for recording audio features in sensitive environments, demonstrating passive capture with privacy-preserving feature extraction [3]. Such systems aimed to minimize intrusiveness and maximize battery life. Recent wearable recorders (e.g., "Friend" pendant on Kickstarter, and devices by Bee AI and Omi) show the transition from lab

prototypes to consumer products. While these devices capture and transcribe conversations to provide summaries and to-do lists, our system uniquely focuses on interpreting transcripts through the lens of productivity theory to derive cognitive energy usage metrics. Psychologist Mihaly Csikszentmihalyi defined flow as a state of complete absorption in an activity where time seems to vanish [4]. Flow is associated with peak performance. Flow and productivity have been linked to optimal energy management, more so than mere time management.

The concept of ego depletion suggests cognitive energy is finite and can be exhausted by tasks requiring self-control. We draw on these ideas to categorize recorded activities and estimate how each consumes or replenishes a person's cognitive energy E [5]. Our work is informed by previous productivity literature, including Deep Work (advocating focused, flow-like work states) and personal analytics tracking efforts (e.g., lifelogging diaries on LessWrong).

Task statement. This research is structured around a set of guiding questions that emerged during the daily activity process.

1. Formalize a mathematical model of daily cognitive energy (E), time (T), and actions ($D_1 \ldots D_n$), including productive vs. unproductive categorization and the flow-state parameter (φ). Define and validate the Countable Cognitive Energy (CCE) formula by assigning energy-rate weights (w_i) to transcribed activity segments. Which features – textual, temporal, and behavioral – can be extracted efficiently and interpreted meaningfully to support future analytical tasks such as narrative detection or engagement modeling?

2. Design a low-power wearable audio recorder microscheme for continuous ~16 h/day capture. Specify an end-to-end cloud pipeline (React front-end, .NET API, Azure Blob Storage, Azure Functions, Python/Whisper NLP) for large-scale ingestion, transcription, analysis, and secure data delivery.

Outline of the main material of the study. We formalize the intuitive notion that each person wakes up with a finite reserve of cognitive energy (E) and a fixed amount of time (T = 24h) for the day. Let E be a quantitative measure (e.g., in arbitrary units in this paper) of cognitive energy available at the start of the day. Actions (tasks, activities) draw from or replenish this energy. We define an action D_i by its duration Δt_i (part of T) and its energy cost or gain ΔE_i. Productive actions are those that move a person towards their goals (work, learning, creative tasks), often requiring significant energy but yielding progress or skill improvements. Unproductive actions might be necessary but do not yield long-term benefits (e.g., idle scrolling or avoidable distractions).

At any time t, the remaining energy $E(t)$ is impacted by preceding actions:

$$E(t) = E_0 + \sum_i \Delta E_i \quad (1)$$

where each productive action typically has

$\Delta E_i < 0$ – energy consumption,

$\Delta E_i \geq 0$ – restful or enjoyable actions might have (energy maintenance or recovery).

Our model aligns with the concept of ego depletion, where tasks requiring discipline deplete a common cognitive energy pool.

By categorizing each action D_i as productive if it contributes to personal or professional objectives (writing code, studying, exercising), or unproductive if it is a time filler or procrastination (excessive social media, aimless web browsing). Importantly, an unproductive action still consumes time T and may consume E (e.g., decision fatigue from context switching). Our later sections describe how transcribed audio can be analyzed and tagged to classify segments into these categories automatically, via NLP keyword spotting or context analysis.

A flow state action is a special productive action wherein the individual experiences heightened focus, enjoyment, and often loses sense of time. In our framework, flow state actions are productive actions with a drastically reduced perceived cost to E per unit time, due to deep intrinsic motivation and focus. In some cases, being in flow can even feel energizing (ΔE_i approaches zero or turns positive, as tasks "fuel" the person). We model flow by a parameter φ_i for each action,

with $\varphi = 1$ indicating normal state,

and $\varphi << 1$ indicating flow.

The effective energy cost in flow is scaled:

$$\Delta E_{i,\ effective} = \varphi_i \cdot \Delta E_i \quad (2)$$

For a flow-state task, φ might be 0.5 or lower (halving energy drain), consistent with studies showing dramatically increased productivity and delayed onset of fatigue.

Flow states are detected post-hoc by analyzing behavioral cues: extended periods of uninterrupted activity on a single task, higher speed of speech or coding, enthusiastic tone, etc. In our audio-based

approach, evidence of flow might include sustained monologues about a work topic, few pauses or distractions, and possibly keywords expressing excitement or deep concentration.

Hypothesis of Maximum Productivity

This hypothesis posits an optimal distribution of actions $D_{1...n}$ that maximizes output given constraints of E and T. It encourages prioritizing high-impact tasks when energy peaks and minimizing energy leakage on low-value tasks.

Our model incorporates the idea that managing energy is more important than managing time. Thus, the goal is not to be busy every minute, but to strategically allocate E and T to maximize productive outcomes and, ideally, achieve flow for the most important activities.

Data Capture and Ethical Considerations

We introduce Countable Cognitive Energy (CCE) as a quantitative estimate of energy expenditure and replenishment derived from the day's audio transcripts. Conceptually, CCE is computed by assigning energy costs (or gains) to each transcribed activity and summing them.

To compute CCE the continuous audio (approx. 16 hours of awake time) is transcribed to text via an automatic speech recognition (ASR) model (OpenAI Whisper in our case). This yields a chronological sequence of utterances with timestamps.

The transcript is segmented into distinct activities or contexts using cues like silence gaps, location changes (if available), or keywords. For example, segments might be "morning routine," "coding at work," "lunch with colleagues," "scrolling Twitter," "evening workout." Techniques such as unsupervised topic modeling or supervised classification with known activity keywords are used to label segments.

We maintain a knowledge base of typical energy costs for activities.

Focused work (coding, writing) – high energy cost per minute.

Social interaction (meeting, conversation) – moderate cost (can vary if the person is extroverted vs. introverted).

Passive leisure (watching TV, browsing) – low to moderate cost, possibly slight energy recovery if relaxing.

Rest (napping, meditation) – energy gain (negative cost).

Flow state instance of any above – reduce cost by factor φi.

CCE is the net sum of energy expenditures over all activity segments. We compute it as:

$$CME = \sum_{i \in segments} w_i \cdot \Delta t_i \quad (3)$$

where:

w_i is the energy cost rate of segment i (units of energy per minute);

Δt_i is the duration of segment i (in minutes).

If w_i is negative (i.e., an energy-replenishing activity), it contributes positively to the overall sum.

A large negative CCE (e.g., $-E_0$) indicates a fully draining day.

A smaller magnitude (e.g., $-0.7 \cdot E_0$) suggests only 70% of the day's energy reserve was used.

A positive CCE would imply net energy gain – rare, but possible on particularly restful days (e.g., meditation).

By converting qualitative transcript segments into a single quantitative metric, CCE enables day-to-day or person-to-person comparisons of cognitive workload. Accuracy depends on reliable transcript segmentation (e.g., via Whisper ASR) and correct assignment of w_i values. Silent but productive activities may be inferred contextually (e.g., long quiet coding sessions), and future extensions could integrate additional sensors (e.g., EEG) for more precise flow detection.

The designed system comprises hardware for audio capture and a cloud-based software pipeline for data processing and analysis. We describe each component and their integration. The wearable device (Fig. 1) is a small module attached to the user's shirt or as a pendant, also will be a custom circuit implementation instead of built-in solution in the II-phase of the research, which should include [6]:

Fig. 1. Wearable dictaphone

– Microphone with omnidirectional electret or MEMS mic to capture ambient sound and speech.

– Microcontroller (MCU) to manage audio sampling (e.g., 16 kHz for speech fidelity), optional compression, and storage writing. A low-power MCU with integrated ADC (analog-to-digital converter) can digitize audio. We use a microSD

or flash chip for ~8–16 GB storage, sufficient for ~1 week of audio if compressed.

– A rechargeable lithium battery (~200–300 mAh) provides power. Based on similar devices (Capture recorder used a 135 mAh battery), our design supports ~16 hours recording by using low-power modes when idle and efficient encoding (e.g., MP3 compression on the fly).

– Storage is offloaded via USB or Bluetooth daily. We opt for manual offload (user uploads nightly) to avoid constant wireless transmission (which would drain battery).

Users interact with their lifelog data through a React-based web application. The front-end serves two main functions.

1. "Upload interface" allows users to upload daily audio files from the wearable (likely after transferring from the device to a computer or phone). The UI shows upload progress and ensures the ~1.2 GB file is successfully sent. Large file upload is handled via chunking and resume support, given potential network issues.

2. "Visualization Dashboard" after processing, the web app displays transcripts, CCE metrics, and interactive visualizations. For example, a timeline highlights productive vs. unproductive segments in different colors, and pie charts summarize time spent in various categories. Flow state occurrences might be marked with a special icon. The React app fetches this data via API calls (discussed below) and uses libraries (like D3.js) for dynamic charts.

Keeping the front-end decoupled ensures responsiveness; heavy processing is done server-side. We introduce a .NET (ASP.NET Core) proxy API as a secure middle layer. This service serves as the gateway to Azure resources.

– It accepts file uploads from the React app and streams them to Azure Blob Storage, authenticating using Azure AD or SAS (Shared Access Signatures) tokens. The API can enforce rate limits (one file per day per user) and file size limits, returning appropriate errors for oversized or too frequent uploads.

– It exposes endpoints for retrieving processed results (transcripts, analytics). Instead of the React app hitting storage or databases directly, the .NET layer fetches data (from Blob) and returns it after authorization checks.

– This design addresses security: user authentication tokens are validated by the .NET API, and direct access to cloud data is prevented, reducing the risk of leaks.

Cloud Storage (Azure Blob) for scalable, durable storage of large audio and text files.

– Audio blob container has raw uploaded audio files land here. Each file is named by user and date (e.g., user123/2025-04-01.wav).

– Transcript container saves a text file or JSON with timestamps (e.g., user123/2025-04-01-transcript.json) after transcription.

– Results Container with final analysis outputs, such as a JSON containing segment classifications, CCE values, etc., are stored here (e.g., user123/2025-04-01-analysis.json).

Azure Blob automatically triggers events on new uploads. Specifically, we use Blob storage events to initiate Azure Functions in our pipeline (next subsection). Blob Storage offers virtually infinite scalability and cost-effective storage for large data, crucial for lifelogging where data accumulates quickly.

The core processing is orchestrated via Azure Functions, enabling an event-driven, scalable workflow.

1. Transcription function (triggered by "Audio Upload") is using a blob trigger on the audio container, this function runs when a new file is uploaded. It loads the audio (streaming directly from Blob to avoid local temp storage) and invokes the Whisper ASR model (which we package with the function or call a containerized service with GPU if available). The function generates a transcript and stores it as JSON in the transcript blob container. We include metadata like recognized language and a confidence score. Execution time is a bottleneck: Whisper's speed varies by model (tiny/large) and hardware (GPU vs CPU).

For ~16 h of audio, even a fast model on CPU might take multiple hours. To mitigate, we can use an Azure Batch Transcription service or optimize by splitting audio into chunks processed in parallel (Azure Functions can process chunks concurrently if we split the file and signal multiple triggers).

2. NLP analysis function is triggered by "Transcript Upload" when the transcript JSON is saved, another blob trigger fires. This function performs several NLP tasks:

– Detects segment boundaries via silence timestamps or topic shifts.

– Tags each segment as productive/unproductive using keywords or a classifier. For example, "code", "write", "design" suggest productive, whereas "scroll", "TV", "idle" suggest unproductive. If uncertain, default to unproductive to avoid overstating productivity.

– Flow detection via searches for long monologues or exclamations indicating high

engagement. Perhaps if the user is humming or talking excitedly about work for >30 minutes uninterrupted, mark flow.

– CCE computation using the classifications and known segment durations. Also computes subtotals (e.g., CCE from morning vs afternoon).

– Optionally generate a summary of the day (for user readability).

– Save a structured output (JSON with segments list and metrics) to the results container.

Software architecture diagram on Fig. 2 depicts the end-to-end system flow from wearable to cloud to user interface.

The wearable generates an audio file which the user uploads via the React app to the .NET API. The API stores it in Azure Blob (Audio File). The upload triggers the Whisper transcription function which saves a transcript to Blob (Transcript).

That triggers the NLP analysis function, which produces results saved to Blob (Results). The user can then fetch and visualize those results in the React app via the API. This loosely coupled, serverless architecture is resilient and scalable: each component can be updated independently, and Azure handles the heavy lifting of scaling storage and compute.

Azure Functions provide automatic scaling. If many users upload simultaneously, multiple instances of these functions spawn. ASR is resource

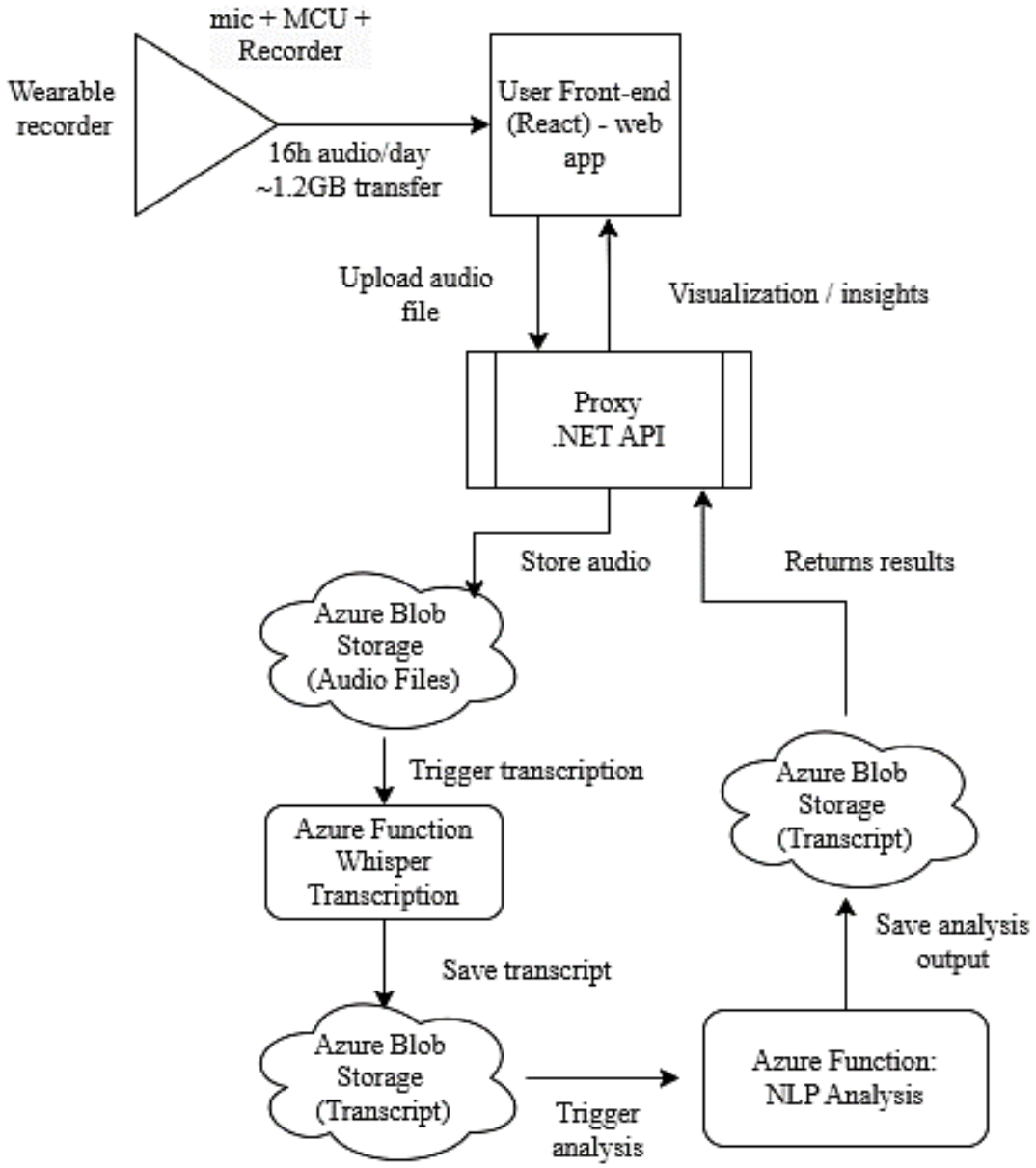


Fig. 2. Software system architecture data flow

heavy; a limitation is the availability of GPUs in Azure Functions.

If needed, we might offload transcription to Azure Cognitive Services for Speech to Text (batch transcription API), which provides a scalable managed service (at cost). Our architecture is flexible to swap in such services without changing the dataflow.

For the initial PoC dataset, the device was used for a 10-day period, ending up in 12 GB of raw audio data *.wav* and compressed into *.mp3* format respectively reducing size to ~1.6 GB and in total of 160 hours, and activity data was analyzed based on that.

The most used activities are described in Table 1, with rounded timing data for simplicity of the subjective analysis. Table presents various daily activities categorized by their duration, cognitive cost, and contribution to a cumulative cognitive effectiveness metric.

A positive CCE indicates net cognitive gain – reflecting the significant flow and recovery periods outweighing costly segments.

In a typical workday without flow or recovery breaks, we would expect a large negative CCE, signaling energy depletion.

Each row represents a specific activity segment, such as Morning Routine, Focused Coding, Deep Work, or Lunch Break.

Duration Δt_i *(min)* is time spent on the activity in minutes.

Base cost w_i *(units/min)* is a qualitative measure of cognitive load or benefit per minute, ranging from negative values (indicating recovery or relaxation) to high positive values (indicating high mental effort).

Flow φ_i is a multiplier representing how well the activity promotes a state of flow or engagement (1.0 being full flow).

Effective cost $\varphi_i \cdot w_i$ *(units/min)* is the product of flow and base cost, representing the actual mental effort or benefit during the activity.

CCE contribution (units) is the total contribution to cognitive effectiveness over the duration of the activity, calculated as duration × effective cost.

Conclusions. The importance of this research lies in its innovative integration of AI, software engineering, and human-computer interaction to create a practical and theoretical foundation for understanding and quantifying cognitive energy in daily life. By introducing Countable Cognitive Energy (CCE) as a measurable metric derived from transcribed audio, the study offers a novel lens to analyze productivity and mental effort across different activities.

The intelligent software framework enables continuous, passive data collection via wearable devices, combined with powerful backend processing pipelines using Whisper AI and natural language techniques. These components allow not only for the identification of high- and low-energy segments but also the detection of psychologically optimal «flow states.» This capability to visualize and assess daily cognitive patterns has implications for improving time management, optimizing work-rest cycles, and enhancing well-being. Furthermore, the system's modular architecture supports scalability and future extensions, such as integration with physiological data, making it a valuable foundation for both academic research and real-world productivity tools.

This research enables a form of productivity lifelogging, where not just events are recorded, but their qualitative impact on one's productive life is assessed.

The ability to visualize how one's energy ebbs and flows throughout the day, and to pinpoint pshycological flow state periods, could help users

Table 1

Activity segments

Segment	Duration Δt_i (min)	Base cost w_i (units/min)	Flow φ_i	Effective cost $\varphi_i \cdot w_i$ (units/min)	CCE contribution (units)
Morning Routine	30	–0.5 (recovery)	1.0	–0.5	–15
Focused Coding	120	1.0 (high cost)	0.6	0.6	72
Team Meeting	45	0.8 (moderate cost)	1.0	0.8	36
Email and Administrative	30	0.5 (low–moderate)	1.0	0.5	15
Social Media Break	15	0.3 (low cost)	1.0	0.3	4.5
Deep Work (Flow)	90	1.2 (very high)	0.5	0.6	54
Lunch Break	60	–0.2 (light recovery)	1.0	–0.2	–12
Commute (Podcasts)	45	0.2 (low engagement)	1.0	0.2	9
Reading (Leisure)	30	–0.1 (relaxing)	1.0	–0.1	–3

restructure their routines for maximal effectiveness and well-being.

It also opens research avenues: for example, correlating CCE with physiological signals (heart rate, cortisol) to validate the energy model, or determine how much energy is expended during an argument.

We plan to refine the NLP models for activity classification (possibly fine-tune a transformer on annotated lifelog data), implement a pilot with users to gather feedback on CCE's usefulness, and explore integration of additional sensors (like detecting computer usage or location to supplement context).

On the theory side, the framework could be expanded to account for different types of energy (cognitive vs physical), and how actions trade off between them. We believe this interdisciplinary approach – mixing theory, AI, and hardware – illustrates a path forward in personal productivity tools and lifelogging research.

Bibliography:

1. LENA Research Foundation. LENA Technology. URL: https://www.lena.org/ (date of access: 20.04.2025).
2. Shah M., Mears B., Chakrabarti C., Spanias A. Lifelogging: archival and retrieval of continuously recorded audio using wearable devices. *2012 IEEE International Conference on Emerging Signal Processing Applications*. 2012. P. 99–102. DOI: https://doi.org/10.1109/ESPA.2012.6152455.
3. Feng T., Nadarajan A., Vaz C., Booth B., Narayanan S. TILES audio recorder: an unobtrusive wearable solution to track audio activity. *Proceedings of the 4th ACM Workshop on Wearable Systems and Applications (WearSys '18)*. 2018. P. 33–38. DOI: https://doi.org/10.1145/3211960.3211975.
4. Csikszentmihalyi M. Flow: The Psychology of Optimal Experience. URL: https://surl.lu/wecjzl (date of access: 20.04.2025).
5. Baumeister R., Bratslavsky E., Muraven M., Tice D. Ego Depletion: Is the Active Self a Limited Resource? *Journal of Personality and Social Psychology*. 1998. № 74 (5). P. 1252–1265. DOI: https://pubmed.ncbi.nlm.nih.gov/9599441/.
6. AN278: Voice Recorder Reference Design. URL: https://www.silabs.com/documents/public/application-notes/AN278.pdf (date of access: 20.04.2025).

Вернік М.О., Олещенко Л.М. ІНТЕЛЕКТУАЛЬНА ПРОГРАМНА ПЛАТФОРМА ДЛЯ ОБРОБКИ ВЕЛИКИХ АУДІО ДАНИХ ТА ОЦІНЮВАННЯ КОГНІТИВНОЇ ЕНЕРГІЇ

Стаття присвячена представленню інноваційної програмної платформи для аналізу аудіо, згенерованого людиною, з метою кількісної оцінки її когнітивної енергії. Розроблено теоретичну модель, яка описує щоденну когнітивну енергію (E), час (T) та дії (D1...Dn), з виокремленням продуктивної та непродуктивної діяльності та визначенням особливого стану потоку. Запропоновано нову кількісну метрику – Countable Cognitive Energy (CCE), що обчислюється на основі транскрипцій аудіо щоденних завдань. Описано технічну архітектуру системи, яка передбачає безперервний запис аудіо за допомогою носимих пристроїв (~16 год/день, ~1,2 ГБ/день на користувача), транскрибування за допомогою Whisper AI, сегментацію та позначення для подальшого аналізу. Програмну систему реалізовано з використанням інтерфейсу на основі React для візуалізації та завантаження даних, Azure Blob Storage для масштабованого зберігання, Azure Functions для керування транскрипцією та обробкою мови, .NET API для керування доступом, а також серверної частини, написаної на Python для розпізнавання мовлення та обробки природної мови. Архітектуру розробленого програмного забезпечення зображено схемами потоків даних та мікросхемою пристрою. З'ясовано ключові обмеження запропонованої програмної системи, зокрема, тривалість транскрипції, обмежену доступність GPU і пропускну здатність мережі. Проаналізовано можливості масштабування системи у майбутньому. Проведено порівняння розробленої платформи із наявними програмними рішеннями (LENA, TILES Audio Recorder, Friend wearable) з акцентом на можливості візуалізації шаблонів продуктивності користувача та виявлення станів потоку. Проаналізовано відповідну літературу з тематики лайфлогінгу та продуктивності. У межах подальших досліджень передбачено вдосконалення NLP-класифікації дій, експериментальну перевірку корисності CCE за участі користувачів та інтеграцію додаткових контекстних сенсорів, таких як використання комп'ютера або геолокації. Також окреслено можливість розширення платформи з розділенням когнітивної та фізичної енергії для подальшого розвитку засобів продуктивності та лайфлогінгу.

Ключові слова: *програмна платформа, великі дані, аудіо, mp3, wav, когнітивна енергія, згенеровані людиною дані, диктофон, NLP, Whisper AI, Python, Azure Blob Storage.*

УДК 004.624
DOI https://doi.org/10.32782/2663-5941/2025.3.2/18

Войтко В.В.
Вінницький національний технічний університет

Позур М.Ю.
Вінницький національний технічний університет

МЕТОД ДИНАМІЧНИХ ВИКЛИКІВ В .NET ШЛЯХОМ ГЕНЕРАЦІЇ IL КОДУ В ПРОЦЕСІ ВИКОНАННЯ ЗАСТОСУНКУ

У статті розглядається розробка методу здійснення динамічних викликів в .NET шляхом генерації IL коду в процесі виконання застосунку. Було проаналізовано основний спосіб здійснення динамічних викликів в .NET з використанням рефлексії. Визначено, що ключовим недоліком рефлексії є її швидкодія, тому існує необхідність пошуку альтернативного рішення для здійснення динамічних викликів. Було розроблено метод, який дозволить здійснювати динамічні виклики аналогічно до рефлексії, але з вищою швидкодією. Запропоноване рішення опирається на використання структури метаданих аналогічно до рефлексії, де метадані членів типу даних асоціюються з відповідними статичними викликами без використання динамічного виклику рефлексії. Така структура використовує асоціацію між власними метаданими типу та метаданими типу рефлексії. Реалізована з використанням хеш-таблиць, що дозволяє отримувати необхідні метадані об'єкту аналогічно до рефлексії. Генерація метаданих відбувається в процесі виконання застосунку через використання спеціально визначеного функціоналу або автоматично при спробі отримання метаданих типу, якщо останній відсутній у системі, що дозволяє використовувати запропонований метод як для типів даних, які були доданими в процесі виконання застосунку, так і анонімних типів даних. Статичні виклики, які асоціюються з відповідними метаданими, генеруються разом з метаданими. Для генерації статичних викликів пропонується використання функціоналу платформи .NET, що дозволяє розширити байт-код додатку в процесі виконання. До цієї категорії відноситься механізм Expression, що дозволяє будувати вирази в процесі виконання додатку та компілювати їх в лямбда-функції, та Reflection.Emit, що дозволяє написання програмних компонентів безпосередньо на рівні байт-коду в процесі виконання застосунку. Запропонований метод динамічних викликів було реалізовано у вигляді бібліотеки. Функціонал генерації статичних викликів розроблено з використанням механізму Expression. Швидкодія запропонованого методу перевірена з використанням бібліотеки Benchmark.NET.

Ключові слова: *метапрограмування, .NET, рефлексія, генерація коду, Expression.*

Постановка проблеми. Сучасне програмне забезпечення є доволі складним та масштабним, що потребує значних ресурсних і часових затрат з боку розробників. У зв'язку з цим популярності набуває використання засобів метапрограмування в процесі розробки програмного забезпечення. До таких засобів входять інструменти, що дозволяють генерацію вихідного коду, і механізми мов програмування та фреймворків, що дозволяють використовувати підходи метапрограмування в процесі виконання додатку [1, с. 33]. В контексті платформи .NET одним із найбільш поширених засобів метапрограмування є механізм рефлексії [2, с. 16]. Цей механізм дозволяє отримувати метадані додатку в процесі його виконання та виконувати низку операцій з цими метаданими. Одним із найбільш типових сценаріїв використання рефлексії в .NET є здійснення динамічних викликів [3, с. 10]. Під динамічним викликом тут розуміємо виклик, що здійснюється з використанням динамічного визначення імені (dynamic name resolution). Ключовою особливістю таких викликів є те, що член типу даних, для якого відбувається виклик, визначається не на етапі компіляції, а під час виконання застосунку. Найпоширенішим сценарієм використання такого типу викликів є серіалізація та десеріалізація даних. Проте, ключовим недоліком рефлексії є швидкодія викликів у порівнянні зі статичними викликами [4], що обумовлює необхідність пошуку альтернативних рішень.

Метою роботи є підвищення швидкодії процесу створення програмних засобів шляхом розробки та впровадження методу динамічних викликів для платформи .NET з генерацією IL коду при виконанні застосунку, що дозволить

покращити швидкодію програмного забезпечення в сценаріях, які опираються на використання динамічних викликів.

Об'єктом дослідження постають процеси розробки програмного забезпечення на платформі .NET з використанням метапрограмування.

Предметом дослідження є моделі та методи розробки програмного забезпечення на платформі .NET з метапрограмуванням.

Аналіз останніх досліджень і публікацій. У напрямку оптимізації динамічних викликів на платформі .NET є низка сучасних публікацій. У публікації [5] розглянуто використання механізму Expression для реалізації функціоналу рефлексії в контексті зчитування значення властивостей. Автор описує використання механізму динамічної компіляції Expression в лямбда функції для реалізації функціоналу динамічних викликів рефлексії. У результаті було отримано аналогічний до рефлексії функціонал зчитування властивостей, що має кращу швидкодію викликів у порівнянні з рефлексією. У публікації [6] розглянуто метод оптимізації рефлексії за рахунок використання динамічної компіляції Expression у лямбда функцію та кешування вже скомпільованої функції.

Таким чином, існуючі рішення опираються на динамічну генерацію статичних викликів та повторне використання згенерованого коду для досягнення кращих показників швидкодії. Проте, варто зазначити, що розглянуті рішення описують лише процес генерації та використання лямбда-функції, але не розробку механізму, що дозволив би зберігання таких функцій та їх повторне використання.

Постановка завдання. Головною задачею є розробка методу, який дозволить здійснення динамічних викликів у .NET за рахунок генерації байт-коду (IL коду) в процесі виконання застосунку. Такий метод опиратиметься на можливість платформи .NET динамічно розширювати власний байт-код у процесі виконання, що дозволяє динамічно генерувати лямбда-функції. Використання лямбда-функції в комбінації з метаданими дозволить отримати функціонал динамічних викликів, схожий до рефлексії, але з вищими показниками швидкодії.

Виклад основного матеріалу. При розробці власного механізму здійснення динамічних викликів у .NET, у першу чергу, варто розглянути рефлексію, яка є основним механізмом платформи .NET для здійснення такого типу викликів. Розглянемо виконання динамічного виклику за допомогою рефлексії (рис. 1).

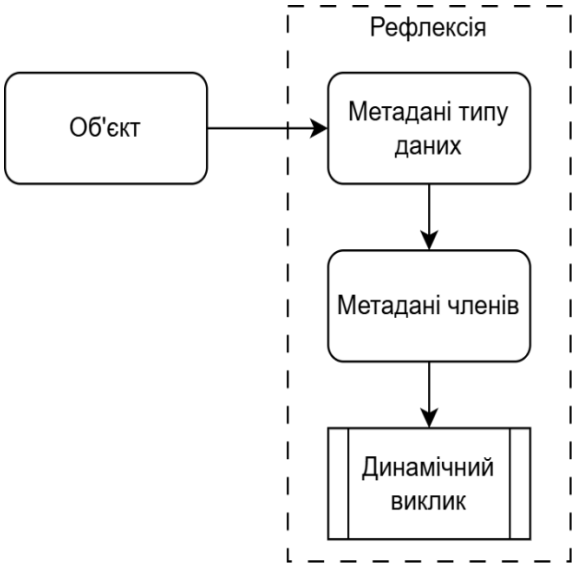


Рис. 1. Модель здійснення динамічного виклику в .NET з використанням рефлексії

Для здійснення динамічного виклику необхідно мати об'єкт, для члену якого відбуватиметься такий виклик. Далі потрібно отримати метадані типу даних (System.Type), що містить у собі метадані всіх членів відповідного типу даних. Для динамічного виклику того чи іншого члену необхідно отримати його метадані. Використовуючи об'єкт метаданих, виконується сам динамічний виклик. У публікації [4] розглядається здійснення такого виклику на рівні середовища виконання платформи .NET Common Language Runtime (CLR). Виконання динамічного виклику за допомогою рефлексії потребує великої кількості операцій на рівні CLR, що призводить до нижчої порівняно зі статичними викликами швидкодії.

Проте, якщо розглядати весь процес здійснення динамічного виклику з використанням рефлексії, то можна побачити, що визначення члену, для якого здійснюється виклик, відбувається на етапі отримання метаданих цього члену, а не на етапі здійснення самого виклику. Якщо асоціювати об'єкт метаданих члену зі статичним викликом цього члена, можна отримати аналогічний до рефлексії функціонал, але з вищою швидкодією.

Проте з таким підходом виникають деякі проблеми:

– необхідність створення власної системи метаданих, адже неможливо додати посилання на статичні виклики до існуючих метаданих;

– необхідність генерації відповідних статичних викликів.

Ключовою задачею власної системи метаданих є забезпечення швидкого доступу до метаданих. Для отримання кращих показників швидкодії у порівняні з рефлексією розроблена система метаданих має забезпечувати кращу або аналогічну швидкість знаходження необхідних метаданих. Для цього доцільно використовувати структуру метаданих, що є схожою до рефлексії. Визначимо основні елементи такої системи:

– *ExpReflection*. Статичний клас, що зберігає у собі посилання на загальну структуру даних та надає доступ до функціоналу розроблюваної системи.

– *ExpRefelctionTypeData*. Клас, що відповідає за метадані типу даних у розроблюваній системі.

– *ExpReflectionPropertyData*. Клас, що відповідає за метадані властивості типу даних у розроблюваній системі.

Така структура (рис. 2) дозволить відтворити елементи рефлексії, що є важливими в контексті здійснення динамічних викликів, та забезпечити аналогічну швидкість знаходження метаданих.

Поле *data* статичного класу *ExpReflection* містить посилання на хеш-таблицю, яка використовується для асоціації метаданих типу рефлексії з метаданими типу розробленої системи. Ключем такої хеш-таблиці є об'єкт метаданих типу рефлексії (*System.Type*), а не назва типу даних. Це пов'язано з обраховуванням хешу для посилальних типів даних у CLR. Значення хешу кожного об'єкта обраховується один раз та зберігається в заголовку цього об'єкта [7]. Таким чином, обрахунок хешу для посилального типу даних зводиться до повернення вже обрахованого результату. Оскільки метадані в CLR існують в єдиному екземплярі, то кожен виклик *object.GetType*() повертатиме об'єкт *System.Type*, що матиме вже обрахований хеш. Так можна досягти кращих показників швидкодії пошуку метаданих типу в розробленій системі.

Об'єкт, що відповідає за метадані типу (*ExpReflectionTypeData*) в розробленій системі має дві структури даних, що містять метадані членів цього типу. Одна структура – зв'язний список, інша – хеш-таблиця. Список використовується, коли необхідно отримати метадані всіх членів, тоді як хеш-таблиця – лише у випадках, коли відбувається пошук конкретного члену з використанням його імені.

Розроблена система розглядається в контексті динамічних викликів саме властивостей класів (зчитування та присвоєння їх значень). Це пов'язано з тим, що такі виклики найчастіше використовуються саме для властивостей. Отже, структура розглядає об'єкт метаданих властивості класу (*ExpReflectionPropertyData*). Такий об'єкт має посилання на відповідні функції, що здійснюють статичний виклик зчитування (*getValueFunc*) або присвоєння (*setValueFunc*) значення цієї властивості для певного об'єкту відповідного типу даних.

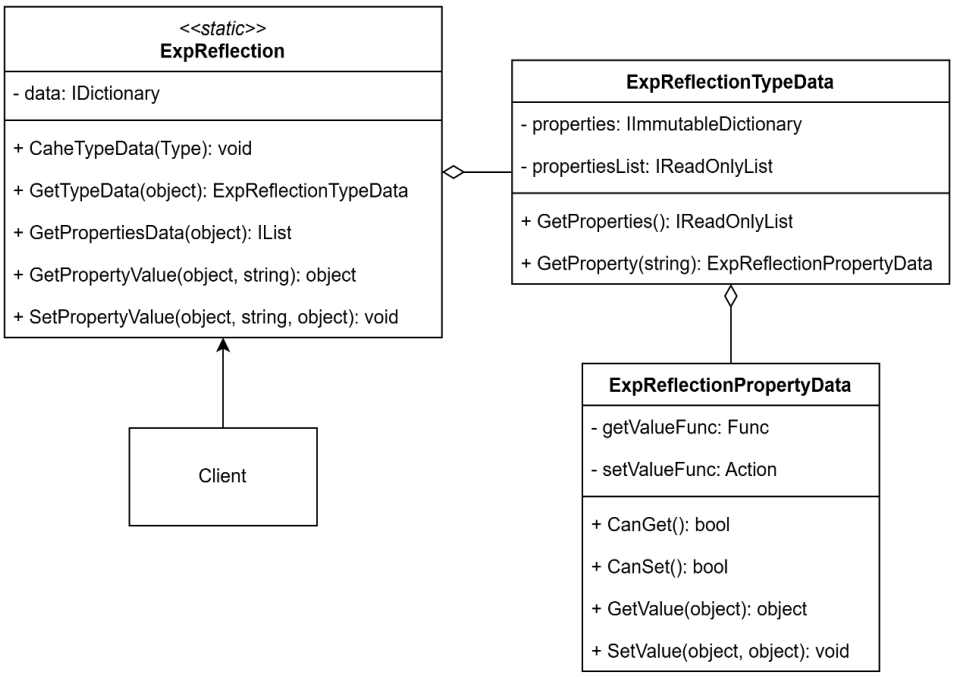


Рис. 2. UML-діаграма розроблюваної системи для здійснення динамічних викликів

Динамічний виклик з використанням розробленої системи описується моделлю (рис. 3).

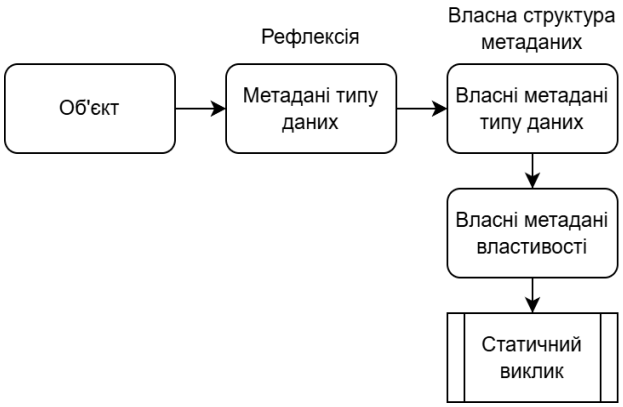


Рис. 3. Модель здійснення динамічного виклику в .NET з використанням розробленої системи

З моделі здійснення динамічного виклику та структури розробленої системи видно, що метадані властивості мають асоціюватися з відповідними статичними викликами. Проте такі статичні виклики необхідно згенерувати перед тим, як їх використовувати. Для цього можна використовувати як механізм Expression, так і Reflection.Emit. Обидва підходи дозволять згенерувати лямбда-функцію, що виконуватиме відповідний статичний виклик у процесі виконання застосунку. Такі функції повинні генеруватися разом із додаванням метаданих до системи.

Оскільки розроблена система метаданих використовується лише для здійснення динамічних викликів, то нема необхідності відразу генерувати всі метадані. Найбільш доцільною тут буде генерація метаданих лише для типів, для яких здійснюються динамічні виклики. Для цього процес отримання метаданих типу з такої системи має враховувати той факт, що тип може бути відсутнім (рис. 4).

Таким чином, якщо метадані необхідного типу відсутні в системі, то відбудеться їх генерація. Така генерація включає в себе не лише створення відповідних об’єктів, а й генерацію статичних викликів для властивостей цього типу даних.

Отже, для реалізації запропонованого методу динамічних викликів потрібно:

1) визначити власну структуру метаданих, що складається з метаданих членів типу, метаданих типів та статичного класу, що надає доступ до цих метаданих;

2) метадані членів мають містити посилання на відповідні статичні виклики;

Рис. 4. Блок-схема алгоритму процесу отримання метаданих властивостей типу даних

3) визначити операцію генерації метаданих типу у розробленій системі, яка повинна забезпечити генерацію як об’єктів метаданих, так і відповідних статичних викликів для членів цього типу;

4) при отриманні метаданих типу необхідно перевіряти, чи є такий тип у системі, якщо ні, то генерувати його метадані, використовуючи операцію, визначену кроком 3;

5) для здійснення динамічних викликів використовувати розроблену систему метаданих.

Запропонований метод динамічних викликів порівнювався за швидкодією з рефлексією. Для цього було використано бібліотеку Benchmark.NET [8]. Порівняння відбувалося як для зчитування значення властивості, так і для присвоєння. Оскільки процес отримання метаданих є частиною здійснення динамічного виклику, то його було включено до порівняння. Результати порівняння швидкодії розробленого методу динамічного виклику та швидкодії динамічного виклику з використанням рефлексії наведено у таблиці 1.

За результатами порівняння можна зробити висновок, що запропонований метод здійснення динамічних викликів є в середньому на 60% швидшим у порівняні з використанням рефлексії. Варто зазначити, що запропонована реалізація методу використовує механізм Expression для генерації лямбда-функцій у процесі виконання додатку.

Таблиця 1

Порівняльний аналіз здійснення динамічних викликів присвоєння та зчитування значень властивості з використанням розробленого методу динамічного виклику та рефлексії

Method	Mean	Error	StdDev	Ratio
ExpReflectionGet	15.79 ns	0.643 ns	0.602 ns	0.41
ReflectionGet	38.50 ns	1.182 ns	1.106 ns	1.00
ExpReflectionSet	16.22 ns	0.783 ns	0.733 ns	0.38
ReflectionSet	42.21 ns	0.279 ns	0.146 ns	1.00

Висновки. Розроблено метод здійснення динамічних викликів у .NET за рахунок генерації IL коду в процесі виконання застосунку. Розроблений метод опирається на використання окремої системи метаданих, де метадані кожного члену типу даних асоціюються з відповідними статичними викликами. Для генерації таких викликів використовується механізм Expression платформи .NET. Швидкодія розробленого методу є в середньому на 60% вищою за швидкодію рефлексії в сценаріях здійснення динамічних викликів зчитування та присвоєння значень властивостей. Однією з переваг запропонованого методу є те, що генерація метаданих та статичних викликів відбувається в процесі виконання застосунку, що дозволяє його використання для типів даних, які були динамічно додані, та для анонімних типів даних.

Список літератури:

1. Bock J., Hazzard K. Metaprogramming In . NET. Manning Publications Co. LLC, 2012.
2. Ingebrigtsen E. Metaprogramming in C#: Automate Your .NET Development and Simplify Overcomplicated Code. Packt Publishing, Limited, 2023.
3. Beaumont A., Bakhtiari Bastaki B. An Investigation into the Prevalence of Reflection Techniques in Distributed Microsoft .Net NuGet Artefacts. *ICSCA 2022: 2022 11th International Conference on Software and Computer Applications*, м. Melaka Malaysia. New York, NY, USA, 2022. URL: https://doi.org/10.1145/3524304.3524329.
4. Warren M. Why is reflection slow?. URL: https://surl.li/kxttwl (дата звернення: 17.05.2025).
5. Haytam Z. C# Expression Trees: Property Getters. URL: https://blog.zhaytam.com/2020/11/17/expression-trees-property-getter/ (дата звернення: 17.05.2025).
6. Ricardo Peres. Using Generated Methods Instead of Reflection. URL: https://weblogs.asp.net/ricardoperes/using-generated-methods-instead-of-reflection (дата звернення: 17.05.2025).
7. GetHashCode inside CLR: Reference types. URL: https://tearth.dev/posts/gethashcode-inside-clr-reference-types/ (дата звернення: 17.05.2025).
8. BenchmarkDotNet. URL: https://benchmarkdotnet.org/index.html (дата звернення: 17.05.2025).

Voitko V.V., Pozur M.Yu. METHOD OF EXECUTING DYNAMIC CALLS IN .NET USING IL GENERATION IN RUNTIME

The article discusses the development of a method for making dynamic calls in .NET by generating IL code during application execution. The main method for making dynamic calls in .NET which relies on using reflection was analyzed. It was determined that the key disadvantage of the reflection in a context of making dynamic calls is its performance, so there is a need to find an alternative solution for making dynamic calls. A method was developed that will allow making dynamic calls in a same way as by using reflection, but with a better performance. The proposed solution is based on the use of a metadata structure, similar to the reflection, where the metadata of type members are associated with corresponding static calls. This way using a reflection dynamic call can be avoided. Such a structure defines an association between the own type metadata and the reflection type metadata, implemented using hash tables. This allows obtaining the necessary metadata for the object in a same way as by using reflection. Metadata generation occurs during application execution through the use of specially defined functionality or automatically when attempting to obtain type metadata if the latter is missing in the system, which allows using the proposed method for both data types that were added during application execution and anonymous data types. Static calls associated with the corresponding metadata are generated along with the metadata. To generate static calls, it is proposed to use the functionality of the .NET platform, which allows for extending the application's bytecode

in runtime. This category includes the Expression mechanism, which allows you to build expressions during application execution and compile them into lambda functions, and Reflection.Emit, which allows you to write program components directly at the bytecode level during application execution. The proposed method of dynamic calls was implemented as a library. The functionality of generating static calls was developed using the Expression mechanism. The performance of the proposed method was tested using the Benchmark.NET library.

Key words: *metaprogramming, .NET, reflection, code generation, Expression.*

UDC 001.89:004.31
DOI https://doi.org/10.32782/2663-5941/2025.3.2/19

Glukhova N.V.
Dnipro University of Technology

COMPUTER-INTEGRATED TECHNOLOGY FOR STUDYING THE SPATIAL STRUCTURE OF A CORONA GAS DISCHARGE BASED ON FRACTAL DIMENSION ESTIMATION

The article analyzes the prospects for using the gas-discharge radiation method to study the properties of water and aqueous solutions. A review of modern achievements in the field of theoretical and experimental studies of water properties is provided. The main problems that arose in the study of water properties and the further need to improve models of the liquid phase of water are analyzed. The main ways of developing theoretical and experimental studies of the physical, chemical and biological properties of water are outlined. The relevance of studying the biological properties of water is analyzed based on modern ideas in the field of quantum electrodynamics about the presence of coherent domains in the structure of the liquid phase of water.

Experimental studies of water using the gas-discharge radiation method have revealed the regularities of the formation of individual gas discharges when an external pulsed electromagnetic field is applied to the liquid sample under study.

Analysis of the obtained images of gas-discharge radiation allowed us to establish the regularities of the formation of gas-discharge tracks for different types of water, which manifest themselves as various specific features of the geometric and photometric parameters of the images.

For the quantitative analysis of the obtained experimental data, the method of determining the fractal dimension was chosen, since the images of the formation of gas-discharge tracks have a tree-like geometric structure with pronounced signs of self-similarity.

The article presents the results and performs an analysis of experimental data and the results of their processing based on the procedure for determining the fractal dimension. The formation of a branched tree-like structure of gas-discharge tracks under the influence of an external pulsed electromagnetic field requires the presence of free charge carriers, the source of which can be coherent domains. The minimum values of the fractal dimension are characteristic of distilled water samples. The obtained result is consistent with the results of research in the field of quantum electrodynamics, since distilled water is characterized by a minimum degree of coherence.

Key words: *image analysis, computer-integrated technology, fractal dimension, gas-discharge radiation.*

Statement of the problem. Studies of various physical processes that are accompanied by the occurrence of a gas discharge have remained relevant for several decades. The study of the features of the formation of gas discharges is of interest to researchers from both a theoretical and a practical point of view. A gas discharge is one of the types of plasma, which is classified as a low-temperature plasma, in contrast to another type of plasma, which is called high-temperature plasma.

The features of the formation of a gas discharge are determined both by the external conditions during laboratory research (for example, temperature, humidity, pressure, air composition, etc.), as well as by the features of the object under study and the parameters of the electrode system used to create the electromagnetic field necessary to ensure an active measuring experiment during which a gas discharge is formed.

Analysis of recent research and publications. A gas discharge can be formed at atmospheric pressure, in which case the distance between the electrodes used to create the electromagnetic field has a significant impact on the course of this physical process [1].

Requirements for the design of devices for the formation of gas discharges can differ significantly and depend on the purpose of theoretical research or practical tasks. To date, numerous examples of the development and practical implementation of devices for the formation of a gas discharge and the registration of an image of its distribution are known.

The use of devices based on the formation of a gas discharge is known in the fields of biology and medicine [2, 3], biotechnology and agrophysics [4], the study of the properties of biological objects and liquids [5–7], the study of the properties of water and aqueous solutions [8, 9], and ecology [10–12].

Since water is one of the most common liquids on Earth and a necessary substance for the life and health of living beings, research into its properties has remained relevant for many years. Theoretical models of the existence of the liquid phase of water are being developed, and experimental methods for determining its characteristics, composition, and structure are being improved.

It should also be noted that theoretical and experimental studies of the properties of water are incomplete and require further improvement, since questions related to a number of so-called anomalous properties of water remain unresolved. Improving theoretical models of water leads to the possibility of a more detailed and comprehensive description of the properties of water, including its physical, chemical and biological properties.

The development of theoretical models of water in historical retrospect indicates that the anomalous properties of water discovered experimentally lead to attempts to further refine existing water models. Along with this, the improvement of methods and measuring instruments led to the experimental discovery of additional new anomalous manifestations of previously unknown properties of water, which necessitate new stages of the need to revise and improve water models. Thus, a continuous relationship is formed between theoretical and experimental achievements in the field of research into the properties of water, which complement each other and contribute to the further development of scientific achievements.

The tasks of studying the properties of water and building its models are not limited to establishing its physical and chemical properties in a static mode, since the movement of molecules and other dynamic phenomena in the structure of water are an integral aspect of its existence as an open system.

For example, the so-called classical model of the liquid phase of water was based on the concept that assumed the presence of hydrogen bonds. The hydrogen bond model was developed by Bernal J. D., Fowler R. H. [13]. A significant scientific achievement of this model was the idea of the structure of water in which individual molecules are connected by hydrogen bonds, which are organized in the form of a continuous three-dimensional network. However, experimental studies have revealed significant shortcomings of this model of the liquid phase of water, which consisted in the inability of the model to describe and explain non-monotonic dependences of a number of important characteristics of water, in particular such parameters as compressibility, volume, heat capacity as functions of other physical quantities, such as temperature.

The existing contradictions between the theoretical description in the form of a water model and the results of experimental studies have led to the need to improve theoretical models. In the course of further development of theoretical scientific research on the properties of water, a number of other theoretical models have been proposed. For example, models based on the percolation theory [14] have been proposed, in particular cluster and mixed cluster-fractal models, etc.

However, despite the large number of developed theoretical models, there are still many unresolved issues in this field. As noted in [15], there are problems in scientific research of microscopic structures in the liquid phase of water, which are caused by fluctuations in the structure of the network consisting of hydrogen bonds.

Along with scientific developments based on the study of clusters in the structure of water, theoretical and experimental studies of water are being carried out within the framework of the direction of quantum dynamics. These studies are based on the idea of the existence of coherent domains in the structure of the liquid phase of water.

A variant of the theoretical model of the existence of the liquid phase of water based on the idea of the presence in its structure of certain areas in the form of coherent domains turned out to be extremely important from the point of view of explaining the key role of water in the cells of living beings. The vast majority of previously developed theoretical models of the liquid phase of water were based on an analytical description of the structure of water as a set of molecules connected by H-bonds, a characteristic feature of which is their existence on the basis of short-term static forces. Within the framework of quantum electrodynamics, a model of water was proposed that took into account the presence of an electromagnetic field and dynamic interactions in the structure of water.

Considering the structure of water in this way allowed us to develop a new model based on the representation of certain areas in its structure as coherent domains, for the characterization of which a physical variable in the form of a phase is used, which is related to the electric potential and the vector of magnetic potentials [16].

In parallel with these studies, which were related to the development of a theoretical model of the liquid phase of water, scientific research

was conducted, as a result of which coherence phenomena were established in the nature of living beings [17]. As a result of the studies performed in work [16], it was concluded that the developed model of the liquid phase of water allows solving the problem of thermodynamic inconsistency based on the use of the concept of coherent domains and phase correlation between molecules in the water model. In work [18], the importance of studying phenomena caused by coherence for describing the properties of water in the cells of living beings is considered.

Task statement. Analysis of theoretical and applied achievements in the field of application of research methods based on the formation of a gas discharge around objects indicates the relevance of further improvement of measuring technologies for detecting specific geometric and photometric features of gas discharge images, which depend on the properties of the object under study.

The statement of the scientific research task is aimed at studying the properties of water and aqueous solutions based on the application of the method of forming a corona gas discharge around a liquid sample based on the use of a system of electrodes to create a pulsed electromagnetic field.

Outline of the main material of the study. Studies of water properties related to coherence phenomena in the structure of water are proving to be an important aspect in determining the properties of water in the fields of biology, medicine, and ecology. The works [6, 19] present the results of studies that indicate the prospects of using the gas-discharge radiation method to study the biological properties of water, which are due to the existence of coherent domains in its structure.

Since the gas-discharge radiation method involves the impact on the sample of the liquid under study by an external electromagnetic field, the physical processes of the emergence and maintenance of the corona gas discharge are associated with the emission of free charge carriers from the surface and upper layers of the sample of the liquid under study. As follows from the model of the liquid phase of water, built on the basis of the ideas about the existence of coherent domains in its structure [16], the emission of free charge carriers occurs with greater intensity from the surface of coherent domains, since this requires less energy consumption compared to other areas not organized into coherent domains.

The general appearance of gas discharge radiation images includes characteristic areas that correspond to the following image elements:

1) image background;

2) an inner circle with the location of the liquid sample under study, the quantity and volume of which is formed by a special device that is part of the gas discharge radiation image recording device;

3) an image area that consists of gas discharge tracks radially located around the liquid sample under study, the spatial superposition of which forms the resulting corona gas discharge picture.

When studying the physical processes of emission of free charge carriers from the surface of a liquid sample, the focus of attention is on the features of the formation of gas discharges in radial directions from the inner circle of the image, which corresponds to the area of the liquid drop.

In [19], a method for analyzing images of gas discharge radiation was proposed, which is based on the analysis of high-frequency components of the image brightness profile, which is formed only along one radial direction. Further development of computer-integrated technologies for a detailed study of the features of the emission of free charge carriers from the surface of a liquid-phase object, the corresponding geometric and photometric properties of corona gas discharge images is based on the application of fractal geometry methods to analyze the geometry of the structure of gas discharge tracks.

Fractal geometry methods for determining the characteristics of gas discharge radiation images around liquid samples were selected based on the analysis of the features of the images, which consist of a superposition of a set of gas discharge tracks.

The images obtained as a result of corona gas discharge registration have a characteristic shape of radially arranged gas discharge tracks, which resemble a tree-like structure of individual branches that can intersect each other, forming complex geometric spatially distributed fragments.

Corona gas discharge fragments for the same liquid sample, on the one hand, have a stochastic nature of geometric and photometric features, and on the other hand, have a certain repeatability of textural features in the radial direction within one corona gas discharge image.

For example, images of distilled water samples are characterized by a more uniform light color of the image fragment corresponding to the gas discharge formation area, but without the presence of clearly defined individual gas discharge tracks. For natural water samples, the corona gas discharge consists of clearly defined individual gas discharge tracks, which have a tree-like geometric structure and are characterized by greater intensity compared

to images of the corona gas discharge of distilled water.

Qualitative analysis of corona gas discharge images obtained when an external pulsed electromagnetic field is applied to liquid samples indicates the validity of the choice of fractal geometry methods for highlighting specific features of gas discharge radiation images.

From the analysis of gas-discharge radiation images of different types of water, it can be concluded that it is possible to detect self-similarity of fragments for image textures. The pattern of formation of individual gas-discharge tracks is characterized by a tree-like structure with specific patterns. At the same time, methods of fractal image analysis provide the possibility of quantitative analysis of self-similarity, i.e. certain repetitions in the geometry of patterns at different image scales [20].

Table 1 shows several examples of the results of fractal dimension calculations for different types of water.

For each type of water, three samples with identical physicochemical properties were selected. The analysis of the obtained results shows that the smallest value of fractal dimension corresponds to the distilled water sample. This result is consistent with the results of modern scientific research in the field of quantum electrodynamics, since distilled water is characterized by a minimal degree of coherence.

Conclusions. The article presents the results of the development of a computer-integrated technology for studying the spatial structure of a corona gas discharge based on the assessment of fractal dimension. The article reviews modern ideas about the structure of water and formulates the main problems and ways to improve theoretical and experimental studies of water properties.

Table 1

Determining fractal dimension for gas discharge radiation images

Water sample	Fractal dimension		
	Sample 1	Sample 2	Sample 3
Distilled	0.95098	0.94832	0.94982
Natural (concentration of dissolved impurities TDS 700 mg/l)	1.3047	1.3154	1.3208
Natural (concentration of dissolved impurities TDS 1060 mg/l)	1.4396	1.5173	1.4491

The results of an experimental study of the properties of water of various types based on the application of the gas-discharge radiation method are obtained. The computer-integrated technology includes a stage of calculating the fractal dimension of gas-discharge radiation images.

The obtained experimental data and the results of their processing with the procedure for quantitative assessment of the self-similarity of the geometry of gas-discharge tracks indicate the prospects for further improvement of methods for studying water properties using the gas-discharge radiation method.

Bibliography:

1. Bogaerts A., Neyts E., Gijbels R., Van der Mullen, J. Gas discharge plasmas and their applications. *Spectrochimica Acta Part B*: *Atomic Spectroscopy.* 2002. N. 57 (4). H. 609–658.

2. Yamasaki H., Terao K., Suzuki T., Simokawa F., Takao, H. Patternable atmospheric-pressure plasma jets with gas discharge in microfluidic channel array. *2013 IEEE 26th International Conference on Micro Electro Mechanical Systems* (MEMS). P. 205–208.

3. Laroussi M. Cold gas plasma sources and the science behind their applications in biology and medicine. *arXiv preprint arXiv:2106.01366*. 2021.

4. Malinina A. O., Aksenyuk I. I., Molnar, K. B. Optical characteristics of a gas discharge radiator of orange-red spectral range. 2020. *Uzhhorod University Scientific Herald. Series Physics*. Issue 47. P. 82–93.

5. Kulyk Y. A., Knysh B. P., Maslii R. V., Kvyetnyy R. N., Shcherba V. V., Kulyk, A. I. Method and gas discharge visualization tool for analyzing liquid-phase biological objects. *Informatyka, Automatyka, Pomiary W Gospodarce I Ochronie Środowiska*. 2021. N.11 (3). P. 22–29.

6. Mintser, O. P., Pesotskaya, L. A., Gorovaya, A. I., Glukhova, N. V., Shchukina, O. S. Influence of coherent properties of water on biological growth of plants. *Ukrainian Journal of Medicine, Biology and Sport*. 2021. N.6 (34). P. 267–274.

7. Povstyanyi V.A., Pesotskaya L.A., Glukhova N.V., Yevdokimenko N. M., Nikogosyan L. R., Koshelnik Ye.L. Peculiarities of gas-discharge luminescence of biological fluid from the achilles tendon in the late postmortem period. *Journal of Education, Health and Sport. – Licensee Open Journal Systems of Kazimierz Wielki University in Bydgoszcz, Poland*. – 2017. N. 7 (2). P. 498–508. eISSN 2391-8306. http://dx.doi.org/10.5281/zenodo.399324

8. Ignatov I., Antonov A., Neshev N. et al. Color coronal spectral analysis of bioelectrical effects of humans and water. *Contemp. Eng. Sci*. 2021. N. 14 (1) : 61–72. https://doi.org/10.12988/ces.2021.91781

9. Ignatov I., Pesotskaya L., Glukhova N. et al. Registration of different types of water with corona gas discharge effects and parameters of brightness. *Port. Electrochim. Acta. N.* 43 (2025) 217–224. https://doi.org/10.4152/pea.2025430401

10. Liang J. P., Zhao Z. L., Zhou X. F., Yang D. Z., Yuan H., Wang W. C., Qiao J. J. Comparison of gas phase discharge and gas-liquid discharge for water activation and methylene blue degradation. *Vacuum*. 2020. 181, 109644.

11. Yuchao Zhou, Yi Hong, Zhenghua Li, Zhenhua Bi, Jing Zhang, Jinhai Niu, Jing Pan, Jie Li, Yan Wu. Investigation of discharge characteristics of DBD plasma produced with multi-needle to plate electrodes in water by optical emission spectroscopy. *Vacuum*, 2019. N. 162, 121–127.

12. Glukhova N., Pesotskaya L., Krachunov H. Development of a methodology for processing experimental studies of liquid-phase objects. *Grundlagen der modernen wissenschaftlichen Forschung: der Sammlungwissenschaftlicher Arbeiten "ΛΌГОΣ" zu den Materialien der VI Internationalen wissenschaftlich-praktischen Konferenz*, Zürich, 24. Mai, 2024. Zürich-Vinnytsia: BOLESWA Publishers & UKRLOGOS Group LLC. P. 156–158.

13. Bernal J. D., Fowler R. H. A theory of water and ionic solution, with particular reference to hydrogen and hydroxyl ions. *J. chem. Phys*. 1933. 1 (8), 515–548.

14. Stanley H. E., Andrade Jr, J. S., Havlin S., Makse H. A., Suki B. Percolation phenomena: a broad-brush introduction with some recent applications to porous media, liquid water, and city growth. *Physica A.: Statistical Mechanics and its Applications*. 1999. N. 266 (1–4). P. 5–16.

15. Gao Y., Fang H., Ni K., Feng, Y. Water clusters and density fluctuations in liquid water based on extended hierarchical clustering methods. *Scientific Reports*, 2022. N. 12 (1). 8036.

16. Bono I., Del Giudice E., Gamberale L., Henry M. Emergence of the coherent structure of liquid water. *Water*. 2012. N. 4 (3). P. 510–532.

17. Collini E., Wong C.Y., Wilk K.E., Curmi P.M.G., Brumer P., Scholes G.D. Coherently wired light-harvesting in photosynthetic marine algae at ambient temperature. *Nature*. 2010. N. 463. P. 644–647.

18. Del Giudice E., Spinetti P.R., Tedeschi A. Water dynamics at the root of metamorphosis in living organisms. *Water*. N. 2010. N. 2. P. 566–586.

19. Глухова Н.В. Комп'ютерно-інтегрована технологія дослідження процесів емісії носіїв заряду з поверхні рідиннофазного об'єкту. *Вчені записки ТНУ імені В.І. Вернадського. Серія: Технічні науки*. 2023. V. 34 (73). N. 5. P. 122–127. https://doi.org/10.32782/2663-5941/2023.5/20

20. Menon A., Basu B. Uncovering the fractal nature of water vapor distribution above the surface of the earth. *Chaos, Solitons & Fractals*. 2023. N. 175. 113984.

Глухова Н.В. КОМП'ЮТЕРНО-ІНТЕГРОВАНА ТЕХНОЛОГІЯ ДОСЛІДЖЕННЯ ПРОСТОРОВОЇ СТРУКТУРИ КОРОННОГО ГАЗОВОГО РОЗРЯДУ НА БАЗІ ОЦІНКИ ФРАКТАЛЬНОЇ РОЗМІРНОСТІ

У статті виконаний аналіз перспектив застосування методу газорозрядного випромінювання для дослідження властивостей води та водних розчинів. Виконаний огляд сучасних досягнень в галузі теоретичних та експериментальних досліджень властивостей води. Проаналізовано основні проблеми, які виникали при вивченні властивостей води, та подальша необхідність удосконалення моделей рідкої фази води. Окреслено основні шляхи розвитку теоретичних та експериментальних досліджень фізичних, хімічних та біологічних властивостей води. Проаналізовано актуальність дослідження біологічних властивостей води на базі сучасних уявлень в галузі квантової електродинаміки про наявність когерентних доменів у структурі рідкої фази води.

Проведені експериментальні дослідження води методом газорозрядного випромінювання дозволили виявити закономірності формування окремих газових розрядів при впливі на досліджуваний зразок рідини зовнішнього імпульсного електромагнітного поля.

Аналіз отриманих зображень газорозрядного випромінювання дозволив встановити закономірності формування газорозрядних треків для різних типів води, які проявляються як різні специфічні ознаки геометричних та фотометричних параметрів зображень.

Для кількісного аналізу отриманих експериментальних даних був обраний метод визначення фрактальної розмірності, оскільки зображення формування газорозрядних треків володіють деревоподібною геометричною структурою з вираженими ознаками самоподібності.

У статті наведено результати та виконаний аналіз експериментальних даних і результатів їх опрацювання на базі процедури визначення фрактальної розмірності. Формування розгалуженої деревоподібної структури газорозрядних треків при впливі зовнішнього імпульсного електромагнітного поля вимагає наявності вільних носіїв заряду, джерелом яких можуть виступати когерентні домени. Мінімальні значення фрактальної розмірності є характерними для зразків дистильованої води. Отриманий результат узгоджується з результатами досліджень в галузі квантової електродинаміки, оскільки дистильована води характеризується мінімальним ступенем когерентності.

Ключові слова: *аналіз зображень, комп'ютерно-інтегрована технологія, фрактальна розмірність, газорозрядне випромінювання.*

UDC 621.3
DOI https://doi.org/10.32782/2663-5941/2025.3.2/20

Holod I.V.
Kharkiv National University of Radio Electronics

DEVELOPMENT OF A PARAMETRIC MODEL AND FORMALIZATION OF PARAMETERS FOR MICROCLIMATE CONTROL IN INDUSTRIAL PREMISES BASED ON CYBER-PHYSICAL SYSTEMS

The rapid development of various industries, its digitalization, the comprehensive spread of the Internet, its availability has led to the increasing application of the principles of the Concept of Industry 4.0 and the emergence of cyber-physical systems. Such systems, based on these sensors, are capable of making decisions and performing actions in the physical world.

The article is devoted to the development of a parametric model for microclimate control in industrial premises based on cyber-physical systems. With the automation advancement and increased demands for energy efficiency, stability of technological processes, and workplace safety, there is a need to create adaptive control systems capable of self-learning and real-time operation. The aim of the study is to enhance the effectiveness of microclimate parameter regulation through formalization of a parametric model, which considers both external influences and internal technological disturbances.

Within the scope of the research, a parametric model has been developed describing the dynamics of temperature, humidity, gas composition, and air exchange in industrial areas. The model incorporates the operational logic of actuators such as heating, ventilation, and cooling systems, and includes a decomposition graph that visually represents the cause-and-effect relationships among system parameters. Special attention is given to the model's adaptability to changing external conditions through the implementation of self-learning algorithms and the utilization of historical data.

The results of the study demonstrate that the proposed approach allows for the creation of a flexible and scalable parametric microclimate model, easily adaptable to the specifics of different industrial processes. Integration with cyber-physical systems opens up possibilities for building autonomous and intelligent solutions for industrial automation. The formalized parameter model presented serves as a universal platform for implementing advanced microclimate control strategies in industrial enterprises.

Key words: *microclimate, cyber-physical systems, adaptive model, industrial automation, actuators, energy efficiency.*

Formulation of the problem. In modern production conditions, the requirements for the quality, stability of technological processes, energy efficiency and safety of staff are increasing. One of the key factors affecting these parameters is the microclimate in production facilities. The fluctuations in temperature, humidity and concentration of harmful gases can not only adversely affect the health of workers, but also reduce the accuracy and durability of technological equipment. In view of this, the automation of microclimate control systems is becoming more relevant, especially in the context of the introduction of cyber-physical systems and Industry 4.0 and 5.0 concepts.

Despite the availability of numerous solutions aimed at controlling individual microclimate parameters, most modern systems do not provide sufficient adaptability to rapidly changing environmental conditions. Often there are no integrated models that can take into account the relationships between different parameters and manage actuators on the basis of comprehensive analysis of input data. This leads to poor use of resources, energy overruns and reducing production efficiency.

Thus, the relevance of the research lies in the need to develop a universal parametric model of intelligent microclimate control, combining physical modeling, actuator logic, and self-learning capabilities. Such a model should be capable of flexible real-time adaptation and easily scaled for various industrial environments, which is particularly important within the context of industrial digital transformation.

Analysis of recent research and publications. Research [1] examined approaches to building an automated system for measuring microclimate parameters in industrial premises. The authors emphasize the development of a reliable, cost-effective, and flexible platform for collecting temperature and humidity data, which serves as the basis for further auto-

mation of regulatory processes. Holinko, I.M. and Halytska, I.Ye. in [2] proposed methods for integrating microclimate subsystems into a unified enterprise automation management system. This approach synchronizes environmental parameters with technological processes, enhancing resource management efficiency. Antoniuk V.S. and Merezhanyi Yu.G. in their paper [3] investigated the impact of microclimate parameters on the reliability and lifespan of precision mechanisms. It was demonstrated that even minor deviations in temperature and humidity could significantly reduce the lifetime of precision equipment, highlighting the importance of continuous monitoring and adaptive control. In [4], Aliinyk Yu.V. and Skidan V.V. focused on issues of energy efficiency in climate control systems. They suggested methods for optimizing the operation of ventilation and heating equipment to reduce electricity consumption without compromising the stability of environmental parameters. In article [5], Altayeva A.B. et al. presented an IoT-oriented intelligent microclimate control system employing fuzzy logic. This approach enables remote monitoring and decision-making based on multiparametric analysis of sensor data in real-time conditions. Kostarev S.N. and Sereda T.G. in [6] developed a basic microclimate control model that accounts for interactions of temperature, humidity, and ventilation through an integrated feedback platform. In subsequent work [7], the same authors, together with Novikova (Kochetova) O.V. and Ivanova A.S., adapted their system for poultry farm conditions, achieving practical regulation of environmental conditions for optimal quail growth, thus confirming the versatility of their model. Makarenko A. and Brajon J. in [8] investigated the use of cellular automata with internal boundaries for solving intelligent control problems. This approach enables decentralized systems with local data processing, offering potential for zone-based microclimate control. In [9], Moholivets Yu.I. and Yaremchuk N.A. proposed a method using fuzzy linguistic variables to derive conclusions on microclimate parameters based on fuzzy scales. This approach allows operation with data characterized by low precision, typical in industrial conditions.

Thus, the analysis of recent studies shows significant progress both in hardware solutions for measuring microclimate parameters and in intelligent algorithms for their regulation. The integration of IoT, fuzzy logic, physical models, and adaptive strategies enables the creation of a new generation of microclimate systems capable of real-time operation with high accuracy and energy efficiency.

Task statement. In this scientific research it is necessary to develop and formalize a parametric model for microclimate control in industrial premises based on cyber-physical systems, enabling improved efficiency of temperature, humidity, air gas composition, and air exchange regulation, ensuring adaptive, energy-efficient, and reliable operation of actuators under changing industrial conditions.

Outline of the main material of the study. Modern production requires effective approaches to maintaining optimal microclimate parameters in industrial premises. This necessity is driven not only by the need to provide comfortable conditions for employees but also to comply with technological parameters, critically important for the process's stability and the final product quality. The application of intelligent microclimate control systems based on cyber-physical systems allows achieving high accuracy in control and adaptability to changes in external and internal environmental conditions. The mathematical model developed in this research takes into account fundamental physical parameters such as temperature, humidity, air gas composition, and actuators that regulate these parameters in real-time.

The mathematical model for intelligent microclimate control is based on a parametric approach, allowing the formalization of dynamic changes in microclimate parameters under the influence of both external and internal factors. The core dependency of the model is described by the equation:

$$MicroClimate =< T_{in}, T_{out}, H_{in}, H_{out}, GasMix, N_{piople}, Actuators, Disturbances > \quad (1)$$

Where T_{in} – indoor air temperature;

T_{out} – outdoor air temperature;

H_{in} – indoor air humidity;

H_{out} – outdoor air humidity;

$GasMix$ – total concentration of other gases;

N_{piople} – number of people present in the premises;

$Actuators$ – mechanisms regulating the microclimate;

$Disturbances$ – disturbances affecting the microclimate, such as heat emissions from equipment, opening of doors, and solar radiation.

The indoor air temperature (T_{in}) directly affects comfortable working conditions and the effectiveness of technological processes:

$$T_{in} = \frac{\sum_{i=1}^{Nz} T_{zonei} * Area_{zonei}}{\sum_{i=1}^{Nz} Area_{zonei}} \quad (2)$$

Where T_{zonei} – temperature in the i-th zone of the premises;

$Area_{zonei}$ – area of the i-th zone (defines the size of a specific zone where microclimate control is implemented);

N_z – total number of zones in the premises.

The temperature in a zone (T_{zone}) can vary depending on these factors:

$$T_{zone} =< T_{set}, RangeFactore, H_{in}, T_{out}, Actuators, HeatSources, Disturbances > \quad (3)$$

Where T_{set} – a setpoint temperature for the respective zone. It defines the desired temperature depending on the specifics of the technological process. For example, in areas with high technological requirements, temperature can be set at a certain level for optimal equipment operation or worker comfort.

RangeFactore – a coefficient accounting for different temperature ranges, used to adjust temperature according to specified temperature modes, such as low, normal, or high temperatures:

$$RangeFactor = \begin{cases} -1, if\ -30°C \le T_{zone} < +10°C\ (low\ mode) \\ 0, if\ +10°C \le T_{zone} < +25°C\ (normal\ mode) \\ 1, if\ +25°C \le T_{zone} \le +60°C\ (high\ mode) \end{cases} \quad (4)$$

HeatSources – additional heat sources (technological equipment, radiation, heat from personnel):

$$HeatSources = \sum_{i=1}^{Nh} P_{heat,i} \quad (5)$$

Where $P_{heat,i}$ – power of the i-th heat source, i=1,2,...,N;

Nh – total number of heat sources in the zone.

Disturbances – disturbances influencing temperature changes in a zone (e.g., door openings, ventilation speed variations, personnel movements):

$$Disturbances = \begin{cases} 1, if\ active\ disturbance \\ 0, if\ no\ disturbance \end{cases} \quad (6)$$

The outdoor air temperature (T_{out}) is a significant parameter influencing the microclimate inside industrial premises. Its formulation may be as follows:

$$T_{out} = f\left(T_{atm}, Solar_{rad}, Wind_{speed}, RangeFactor\right) \quad (7)$$

Where T_{atm} – ambient air temperature (meteorological data);

$Solar_{rad}$ – intensity of solar radiation impacting the heating of external constructions:

$$Solar_{rad} = \begin{cases} 0, if\ 18 \le Time < 6\ (night / evening) \\ \alpha \cdot (Time - 6), if\ 6 \le Time < 12\ (morning) \\ \beta \cdot (18 - Time), if\ 12 \le Time < 18\ (daytime) \end{cases} \quad (8)$$

Where – morning growth coefficient of solar radiation (6:00–12:00). Defines the rate of increase of solar radiation after sunrise. Depends on geographical latitude (larger closer to the equator), season (lower in winter than in summer), cloudiness (reduced on cloudy days). For example, in mid-latitudes, a typical summer day has $\alpha \approx 50–100$ W/m²·hr, and in winter – 10–30 W/m²·hr.

β – coefficient of solar radiation decrease after noon (12:00–18:00). Defines how quickly the intensity of radiation decreases after peak intensity. Also dependent on geographical latitude and season; generally, β is slightly lower than α in mid-latitudes because heated surfaces continue emitting heat longer. For example, a typical summer day has $\beta \approx 40–90$ W/m²·hr, and in winter – 5–25 W/m²·hr.

These coefficients can be calibrated based on actual meteorological data from a specific region to achieve greater accuracy.

$Wind_{speed}$ – wind speed affecting premises cooling. This factor is not considered since industrial premises are typically insulated from direct wind exposure, and its influence on the overall heat balance is insignificant compared to other factors;

Time – time of day determining the level of solar radiation and possible temperature variations.

Indoor air humidity (H_{in}) is also a key parameter of the microclimate, affecting technological process efficiency, equipment conditions, and personnel comfort:

$$H_{in} = f\left(H_{out}, T_{in}, \text{Actuators}, RangeFactor\right) \quad (9)$$

The external humidity level (H_{out}) can be described as a function of external factors:

$$H_{out} = f\left(T_{out}, Weather, RangeFactor\right) \quad (10)$$

Where *Weather* – current weather conditions (rain, snow, wind, etc.):

$$Weather = \begin{cases} 1, if\ precipitation\ (rain, snow, fog, etc.) \\ 0, if\ no\ precipitation\ (clear\ or\ cloudy\ without\ precipitation) \end{cases} \quad (11)$$

Air gas composition (GasMix) is a critically important microclimate parameter defining the air quality in industrial premises:

$$GasMix = f(O_2, CO_2, NH_3, NO_x, SO_2, CH_4, Actuators, RangeFactor) \quad (12)$$

Where $O_2, CO_2, NH_3, NO_x, SO_2, CH_4$ – concentrations of main gases;

$Actuators$ – executive mechanisms regulating the temperature in the zone (heating systems, cooling systems, on/off fans, variable-speed fans, and air volume control mechanisms).

Activation (13) and deactivation (14) of heating are based on the logic related to the setpoint temperature (T_{set}) and hysteresis values ($Hyst_{on}, Hyst_{off}$) to prevent fluctuations:

$$Heat_{on} = True,\ if\ \ T_{zone} < (T_{set} - Hyst_{on}) \quad (13)$$

$$Heat_{off} = True, if\ T_{zone} \geq (T_{set} + Hyst_{off}) \quad (14)$$

Cooling systems are responsible for reducing the zone temperature when it exceeds the defined threshold:

$$Cool_{on} = True, if\ \ T_{zone} > (T_{set} + Hyst_{on}) \quad (15)$$

$$Cool_{off} = True, if\ T_{zone} \leq (T_{set} - Hyst_{off}) \quad (16)$$

$Fans_{on/off}$ – fans operating in on/off mode:

$$Fans_{on} = True\ , if\ \ T_{zone} > (T_{set} + Hyst_{on}) \quad (17)$$

$$Fans_{off} = True, if\ T_{zone} \leq (T_{set} - Hyst_{off}) \quad (18)$$

Fans can also be activated in the event of insufficient air exchange in the zone:

$$Fans_{on} = True\ if\ AirV < AirV_{reg} \quad (19)$$

Where $AirV$ – actual volume of air processed by the ventilation system;

$AirV_{reg}$ – required air volume for the specific technological process, determining minimal ventilation requirements.

Fans are additionally activated when harmful gas concentrations exceed permissible limits in the zone:

$$Fans_{on} = True\ if\ GasMix > GasMix_{max} \quad (20)$$

Where $GasMix$ – total concentration of harmful gases (expression 12);

$GasMix_{max}$ – permissible threshold value of gas concentration in the zone.

Variable-speed fans allow precise adjustment of airflow intensity for accurate temperature regulation within the zone. They operate based on the defined temperature regime, helping achieve optimal indoor microclimate conditions:

$$Fans_{reg} = f\left(T_{zone}, T_{set}, T_{out}, H_{in}, H_{out}, GasMix, N_{piople}, AirV\right) \quad (21)$$

The air volume ($AirV$) processed by ventilation systems ensures necessary circulation and microclimate maintenance in the industrial premises. This parameter defines the total airflow through the system, balancing exhaust and intake air:

$$AirV = f\left(Fans_{\underline{on}}, Fans_{reg}, T_{zone}, T_{out}, N_{piople}, AirV_{reg}, FAir_{reg}\right) \quad (22)$$

Where $FAir_{reg}$ – parameter defining the degree of opening of automated fresh air dampers used for supplying fresh air to the production zone. The $FAir_{reg}$ value is based on the controlled pressure indicator in the production zone (ΔP_{zone}), indicating vacuum or excess pressure. Depending on this parameter, dampers open or close to the required level, adjusting the fresh air inflow speed and influencing the microclimate:

$$FAir_{reg} = f(T_{zone},, T_{out}, H_{in}, H_{out}, GasMix, N_{piople}, AirV, Disturbances, \Delta P_{zone}) \quad (23)$$

$Alarm$ – is a component of the control system providing prompt notification of hazardous or unstable conditions potentially leading to technological disruptions or personnel safety threats.

Unlike other actuators, the $Alarm$ does not directly influence the microclimate but serves as a monitoring tool, drawing operator attention to situations requiring intervention or special monitoring:

$$Alarm = f\left(GasMix_{cls}, T_{in}, T_{set}, \Delta T, Time, H_{in}, T_{zone}\right) \quad (24)$$

To ensure efficient microclimate management in industrial environments, having not only a mathematical model but also its visual representation is crucial. Visualization of the parametric microclimate model in the form of a decomposition graph allows for clear representation of the relationships between input parameters, intermediate calculations, and actuators.

The model's decomposition graph provides a structured depiction of each parameter's influence on the final microclimate state. Input parameters (temperature, humidity, gas concentration) form initial conditions processed by intermediate mechanisms: heating, ventilation, and cooling.

This enables real-time monitoring of the system state and forecasting necessary adjustments.

Additionally, visual representation facilitates faster diagnosis of deviations and response to critical situations, enhancing control efficiency and maintaining microclimate stability in the premises.

The first level includes all input parameters that are directly measured or externally set.

The second level involves categorization of these parameters according to defined criteria and boundary conditions, forming an intermediate layer utilized for decision-making within the model.

The third level comprises the logic for activating or deactivating actuators, which directly influence microclimate conditions within the zone.

The Alarm system is placed on a separate level and activates exclusively under critical microclimate conditions.

The implementation of the developed parametric microclimate model into industrial processes provides new opportunities for enhancing environmental management efficiency through cyber-physical systems (CPS). The integration of this model with CPS enables linking physical production processes with virtual models, thus ensuring continuous monitoring, analysis, and adaptive real-time control.

This approach allows continuous updating of microclimate sensor data, automatically adjusting temperature, humidity, ventilation, and air purification parameters. Thanks to self-learning algorithms, the model can incorporate historical data, improving forecasting accuracy and optimizing energy consumption.

Scaling this model for large industrial facilities is achievable due to the modular structure of the decomposition graph. This structure allows easy integration of new actuators or additional control zones, adapting the model to specific requirements of various enterprises.

Therefore, the parametric microclimate model integrated with cyber-physical systems not only ensures high flexibility and adaptability but also enhances overall production efficiency through effective energy resource management and technological process stability.

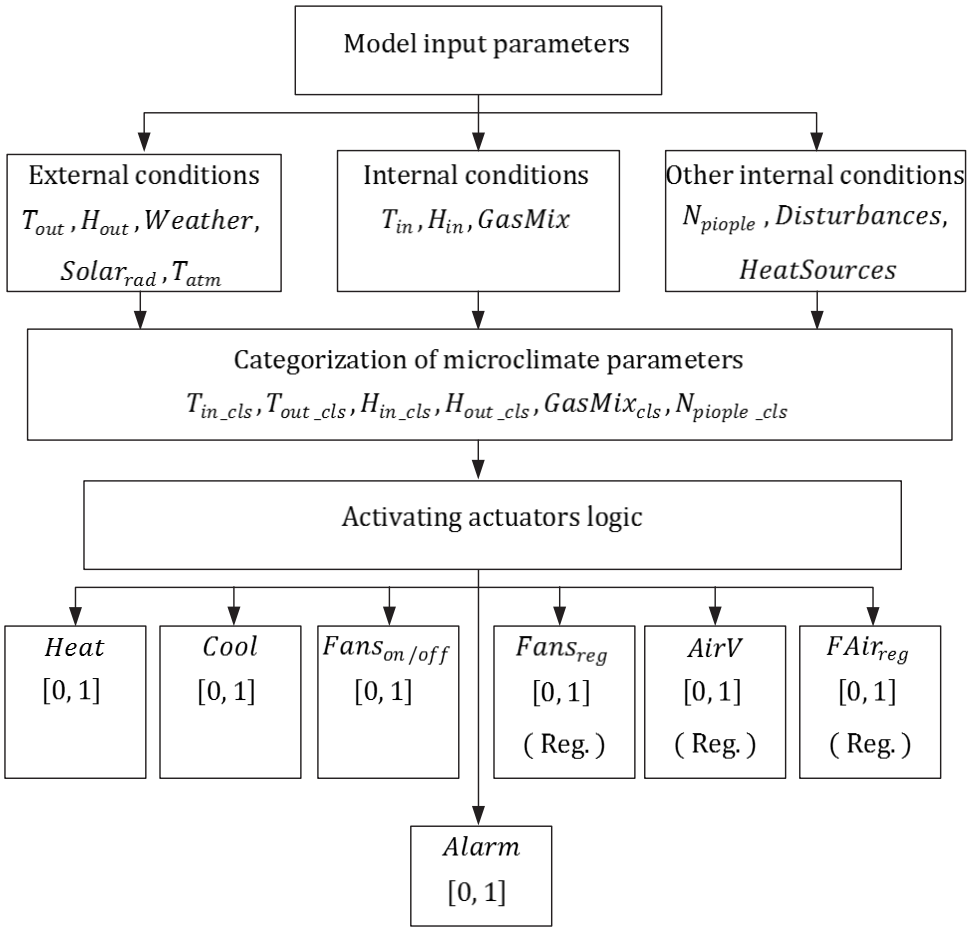


Fig. 1. General decomposition graph of the parametric microclimate model

Conclusions. The developed parametric model for microclimate control in industrial premises based on cyber-physical systems provides efficient control of environmental parameters by formalizing the relationships between input parameters, technological disturbances, and actuator logic. The model considers fundamental physical parameters – temperature, humidity, air gas composition, and air exchange – ensuring structured and adaptive real-time control. Through a decomposition graph, formalized parameters provide a clear representation of causal relationships and facilitate prompt managerial decision-making.

The proposed formalization of model parameters offers several advantages in real industrial conditions. Firstly, it provides a high degree of adaptability through self-learning algorithms. Secondly, optimal energy consumption is achieved due to clearly defined actuator control logic. Thirdly, stability of technological processes is enhanced by the precision of formalized parameters and their interdependencies.

Further research can focus on improving formalization mechanisms for adaptation, expanding self-learning capabilities, and integrating with other cyber-physical system components to develop fully autonomous microclimate management systems. Additionally, scaling the developed parametric model for large industrial complexes with heterogeneous microclimate conditions appears to be a promising research direction.

Bibliography:

1. Ігнатенко Д. В., Боцман І. В. Розробка автоматизованої системи вимірювання параметрів мікроклімату у виробничих приміщеннях. Автоматизація та комп'ютерно-інтегровані технології: матеріали міжнар. наук.-практ. конф. (ACS-2020, Харків, 26 листопада 2020 р.). Харків: ХНУРЕ, 2020. С. 93–95.

2. Голінко І. М., Галицька І. Є. Інтегрування систем керування мікрокліматом в автоматизовану систему керування підприємством. Інформаційні системи, механіка та керування. 2020. № 19–20. С. 78–87. DOI: https://doi.org/10.20535/2219-3804202019194413.

3. Антонюк В. С., Мережаний Ю. Г. Вплив параметрів мікроклімату виробничих приміщень на ресурс прецизійних механізмів приладів. Вісник двигателебудування. 2012. № 1. С. 98–102.

4. Алійник Ю. В., Скідан В. В. Методи підвищення ефективності системи автоматизованого керування клімат-контролю виробничих приміщень. Мехатронні системи: інновації та інжиніринг: матеріали VI міжнар. наук.-практ. конф. (Київ, 2022). 2022. С. 141–142.

5. Altayeva A. B., Omarov B. S., Cho Y. I. Intelligent microclimate control system based on IoT. International Journal of Fuzzy Logic and Intelligent Systems. 2016. Vol. 16, No. 4. P. 254–261. DOI: https://doi.org/10.5391/IJFIS.2016.16.4.254.

6. Kostarev S. N., Sereda T. G. Microclimate control system development. IOP Conference Series: Materials Science and Engineering. 2018. Vol. 450. P. 062013. DOI: https://doi.org/10.1088/1757-899X/450/6/062013.

7. Kostarev S. N., Sereda T. G., Novikova (Kochetova) O. V., Ivanova A. S. Development of a microclimate control system for a quail farm. IOP Conference Series: Earth and Environmental Science. 2022. Vol. 1043. P. 012004. DOI: https://doi.org/10.1088/1755-1315/1043/1/012004.

8. Makarenko A., Brajon J. New possibilities of using cellular automata in space with internal boundaries for computational intelligence problems. Computational Intelligence: Proc. of the 5th Int. Sci.-Pract. Conf. (Uzhhorod, Ukraine, 15–20 April 2019). 2019. P. 17–20.

9. Моголівець Ю. І., Яремчук Н. А. Способи отримання результату вимірювання параметрів мікроклімату промислових приміщень за шкалою з нечіткою лінгвістичною змінною. Обчислювальний інтелект. 2019. № 1 (51). С. 257–259.

Голод І.В. РОЗРОБКА ПАРАМЕТРИЧНОЇ МОДЕЛІ ТА ФОРМАЛІЗАЦІЯ ПАРАМЕТРІВ КЕРУВАННЯ МІКРОКЛІМАТОМ У ВИРОБНИЧИХ ПРИМІЩЕННЯХ НА ОСНОВІ КІБЕРФІЗИЧНИХ СИСТЕМ

Стрімкий розвиток різних галузей виробництва, його цифровізації, всеохоплюючого розповсюдження Інтернету, його доступності призвів до все більшого застосування принципів концепції Індустрія 4.0 та виникнення кіберфізичних систем. Такі системи на основі даних сенсорів здатні самостійно ухвалювати рішення та виконувати дії в фізичному світі.

Статтю присвячено розробці параметричної моделі керування мікрокліматом у виробничих приміщеннях на основі кіберфізичних систем. З розвитком автоматизації та зростанням вимог до енергоефективності, стабільності технологічних процесів і безпеки працівників виникає потреба

у створенні адаптивних систем керування, здатних до самонавчання та роботи в реальному часі. Метою дослідження є підвищення ефективності регулювання параметрів мікроклімату шляхом формалізації параметричної моделі, яка враховує як зовнішні впливи, так і внутрішні технологічні збурення.

У рамках роботи побудовано параметричну модель, що описує динаміку зміни температури, вологості, газового складу повітря та повітрообміну у виробничих зонах. До моделі включено логіку роботи виконавчих механізмів, таких як обігрів, вентиляція та охолодження, а також сформовано граф декомпозиції, що дозволяє візуалізувати причинно-наслідкові зв'язки між параметрами системи. Особливу увагу приділено адаптивності моделі до зміни зовнішніх умов завдяки реалізації алгоритмів самонавчання та використанню історичних даних.

Результати дослідження демонструють, що запропонований підхід дозволяє створити гнучку та масштабовану параметричну модель мікроклімату, яка легко адаптується до специфіки різних виробничих процесів. Інтеграція з кіберфізичними системами відкриває перспективи побудови автономних і розумних рішень для промислової автоматизації. Запропонована формалізована модель параметрів є універсальною платформою для впровадження сучасних стратегій керування мікрокліматом на підприємствах.

Ключові слова: *мікроклімат, кіберфізичні системи, адаптивна модель, промислова автоматизація, виконавчі механізми, енергоефективність.*

УДК 519.718.2
DOI https://doi.org/10.32782/2663-5941/2025.3.2/21

Григор'єва Т.І.
Міжнародний гуманітарний університет

Розенвассер Д.М.
Міжнародний гуманітарний університет

Хнюнін С.Г.
Міжнародний гуманітарний університет

Крохмаль М.В.
Міжнародний гуманітарний університет

АНАЛІЗ СКЛАДНИХ СИСТЕМ ЩОДО МАШИННОГО НАВЧАННЯ ЗА ДОПОМОГОЮ ТЕНЗОРНИХ МЕТОДІВ

У статті розглянуто застосування тензорних методів для аналізу складних систем у контексті машинного навчання. З огляду на зростаючий обсяг багатовимірних та слабо структурованих даних, що характерні для сучасних прикладних задач, виникає нагальна потреба у використанні потужних інструментів, здатних ефективно обробляти інформацію з високою розмірністю, зберігаючи внутрішню структуру даних. Одним із таких інструментів є тензорний аналіз, який уможливлює формалізацію, декомпозицію, оптимізацію та інтерпретацію складних взаємозв'язків між змінними.

В роботі розглядаються тензорні моделі складних систем. Показана можливість використання тензорів для представлення залежностей між компонентами системи в компактній і узагальненій формі. Тензорний підхід забезпечує інваріантність рівнянь під час перетворень координат. Використання тензорного формалізму для аналізу надійності складних систем відкриває можливості точнішого моделювання складних взаємодій і багатовимірних залежностей. Цей підхід відкриває перспективи для підвищення стійкості систем шляхом оптимізації на основі математичних методів, що ґрунтуються на тензорах.

В роботі наведено ключові тензорні підходи, які наразі активно впроваджуються в алгоритми машинного навчання. Зокрема, розглянуто методи тензорного розкладання та декомпозиція, які дозволяють зменшити розмірність даних, зберігаючи характеристики, що є особливо актуальним для задач класифікації, кластеризації та рекомендацій. Окрему увагу приділено тензорній факторизації, яка застосовується для побудови компактних моделей і виявлення прихованих закономірностей у багатовимірному просторі ознак. Також розглянуто використання тензорних операцій у глибокому навчанні – включаючи згорткові нейронні мережі, де тензорні представлення сприяють зниженню обчислювальної складності моделей без зменшення їх точності. Висвітлено роль тензорних методів оптимізації, таких як тензорне градієнтне спускання, у прискоренні навчання моделей шляхом врахування структурних особливостей параметричного простору.

У статті досліджено тензорні підходи до аналізу надійності складних систем, які дозволяють ефективно опрацьовувати багатовимірні та структурно складні дані. Особливу увагу приділено інтеграції тензорного аналізу з методами моделювання систем за допомогою випадкових процесів. Результати дослідження демонструють, що поєднання тензорного підходу та ймовірнісного моделювання дозволяє досягти високої точності у прогнозуванні відмов та зменшити обчислювальну складність задачі.

Отримані результати підтверджують, що тензорний аналіз є надзвичайно перспективним напрямом у розвитку машинного навчання. Його використання дозволяє не лише підвищити точність моделей, а й суттєво знизити їх складність, що є критично важливим у прикладних сценаріях з обмеженими ресурсами. Запропоновані підходи можуть бути застосовані в широкому спектрі задач – від комп'ютерного зору до прогнозування поведінки складних систем у реальному часі.

Ключові слова: *складні системи, надійність складних систем, математичне моделювання складних систем, тензорні моделі складних систем, випадкові процеси, оптимізація, машинне навчання, тензорна факторизація, методи тензорного розкладання, тензорна декомпозиція.*

Постановка проблеми. У сучасних умовах стрімкого зростання обсягів і складності даних, що виникають у процесах проєктування, аналізу та експлуатації складних систем, виникає необхідність у використанні більш ефективних інструментів математичного моделювання та обробки інформації. Традиційні методи лінійної алгебри та класичної статистики виявляються обмеженими у здатності враховувати багатовимірну структуру даних, складні нелінійні залежності між компонентами системи, а також часову динаміку процесів, що відбуваються в реальному часі.

Одним з перспективних напрямків вирішення цих проблем є застосування тензорних методів, які надають можливість компактного й інформативного представлення багатовимірних даних у вигляді тензорів – багатовимірних масивів, що розширюють поняття матриць. Ці методи особливо ефективні в поєднанні з алгоритмами машинного навчання, де необхідно аналізувати дані з високою розмірністю та складною внутрішньою структурою.

Таким чином, постає наукова проблема розробки та впровадження узагальненого підходу до аналізу складних систем, який би об'єднував тензорні методи, машинне навчання та ймовірнісно-статистичне моделювання. Це дозволить не лише ефективно аналізувати наявні дані, а й здійснювати передбачення поведінки системи у змінних умовах експлуатації, що є особливо важливим для систем критичного призначення: в оборонній, медичній, транспортній та енергетичній галузях.

Аналіз останніх досліджень і публікацій. Аналіз надійності складних систем за допомогою тензорних методів і випадкових процесів є відносно новим, але швидко зростаючим напрямом досліджень у галузі прикладної математики, обчислювальної інженерії та машинного навчання. Із зростанням складності сучасних систем виникла потреба у більш гнучких і потужних засобах моделювання багатовимірних процесів, що враховують часову, просторову та функціональну взаємозалежність змінних. В [1] були запропоновані фундаментальні підходи до розкладання тензорів і продемонстровано їхню ефективність у задачах зменшення розмірності та виділення структур у багатовимірних даних. Подальші дослідження зосереджувалися на адаптації цих методів до реальних систем з часовими залежностями та стохастичною природою. Наприклад, в [2, 3] розглядаються тензорні мережі, які ефективно використовуються для представлення динамічних даних. У галузі теорії надійності важливими є роботи, які поєднують стохастичні моделі з тензорними уявленнями. В [4] досліджуються випадкові процеси, представлені у тензорному вигляді, що значно зменшує обчислювальну складність при збереженні точності. Однак, попри стрімкий розвиток напряму, питання інтеграції тензорного формалізму з теорією випадкових процесів для моделювання надійності ще недостатньо систематизоване. Необхідне подальше вдосконалення моделей, здатних враховувати як структурну складність систем, так і динаміку в умовах стохастичної невизначеності. Це створює підґрунтя для подальших досліджень, зокрема в напрямку розробки тензорно-стохастичних моделей.

Постановка завдання. Метою статті є дослідження ефективності та доцільності використання тензорних методів для аналізу складних систем у контексті машинного навчання, а також обґрунтування переваг тензорного підходу для обробки багатовимірних, структурованих і неструктурованих даних. У роботі ставиться завдання узагальнити існуючі методи тензорного аналізу, адаптувати їх до потреб сучасних задач машинного навчання, та запропонувати підходи до оптимізації моделей з урахуванням високої розмірності вхідного простору.

Виклад основного матеріалу. Методи тензорного аналізу в машинному навчанні. Тензорний аналіз є потужним математичним інструментом, який відіграє важливу роль у сучасному машинному навчанні. Він дозволяє ефективно працювати з багатовимірними даними, моделювати складні залежності та оптимізувати обчислювальні процеси. Застосування тензорних методів у машинному навчанні охоплює широкий спектр задач, включаючи глибоке навчання, розпізнавання зображень, обробку природної мови та розв'язання задач оптимізації.

Одним із ключових аспектів використання тензорного аналізу є представлення даних у тензорному вигляді. Наприклад, у глибоких нейронних мережах ваги та активації нейронів можуть бути зображені у вигляді багатовимірних масивів, що спрощує операції над ними. Тензорні обчислення широко використовуються у фреймворках машинного навчання, що дозволяє значно прискорити виконання моделі.

Тензорна факторизація [5] використовується для автоматизованого виявлення значущих ознак у високорозмірних даних. Факторизація тензора,

тобто представлення тензора у вигляді добутку компонентів нижчої розмірності, дає змогу зменшити обчислювальну складність та виявити приховані залежності. Тензорна факторизація дозволяє виявити як лінійні, так і нелінійні залежності між змінними. Це корисно у таких завданнях, як кластеризація, побудова рекомендованих систем, обробка мовлення, біоінформатика, та медична діагностика.

Важливою категорією є тензорні обчислення в глибокому навчанні. Нейронні мережі за своєю природою працюють із багатовимірними тензорами, а операції згортки, нормалізації та поширення градієнтів виконуються за допомогою тензорної алгебри. Операція декомпозиції сприяє зменшенню кількості параметрів та підвищенню ефективності моделей без втрати точності. Такий підхід підвищує узагальнювальну здатність мережі та пришвидшує її тренування.

Тензорні методи оптимізації також відіграють важливу роль у навчанні моделей. Тензорний градієнтний спуск [6] дозволяє знаходити глобальні мінімуми функцій втрат в умовах великої кількості параметрів і складних залежностей між ознаками. Його перевага – врахування структурованості параметричного простору, що значно зменшує обчислювальні витрати.

Ще одним важливим напрямком є тензорні представлення у обробці природної мови [7]. У цьому контексті тензорні моделі застосовуються для створення багатовимірних векторних уявлень слів, речень або навіть цілих документів. Наприклад, тензорне розкладання дозволяє аналізувати частотні та семантичні взаємозв'язки між словами у корпусі, формалізуючи їх у вигляді тривимірного тензора «слово – контекст – документ».

Окрім факторизації, у рамках тензорного аналізу активно застосовуються операції тензорної регуляризації, методи зменшення розмірності, а також класифікаційні моделі, що базуються на тензорних ядрах [8]. Зокрема, тензорні ядрові функції дозволяють адаптувати класичні алгоритми до багатовимірних структур, не втрачаючи інформації про їхню внутрішню будову.

Таким чином, тензорні методи стали фундаментальним інструментом у машинному навчанні для аналізу складних, багатовимірних і слабоструктурованих даних. Тензорні методи поєднують високу точність із компактністю моделей, що є критично важливим у реальних прикладних задачах, таких як медичні дослідження, обробка мовлення, моделювання соціальних мереж і рекомендаційні системи. Подальші дослідження спрямовані на інтеграцію тензорних архітектур з байєсівськими підходами, розробку адаптивних розкладів та впровадження у квантові обчислення, що відкриває нові горизонти в теорії та практиці штучного інтелекту.

Аналіз надійності складних систем тензорними методами. Теорія надійності складних систем є важливою галуззю досліджень, що охоплює оцінку, моделювання та прогнозування відмов. Традиційні методи аналізу надійності, засновані на ймовірнісно-статистичних підходах, є доцільними для розгляду окремих компонентів, але стають менш ефективними при описі складних багатокомпонентних систем із тісними зв'язками між їх елементами. Тензорний аналіз, що розглядає багатовимірні залежності, пропонує більш гнучкий та універсальний математичний апарат для моделювання та оцінки надійності систем.

Розглянемо систему S, яка описується як сукупність n-вимірного простору компонентів. Надійність системи представимо у вигляді тензорного рівняння:

$$R_{i_1 i_2 \ldots i_n} = P(T > t \mid X),$$

де $R_{i_1 i_2 \ldots i_n}$ – тензор надійності, який визначає ймовірність безвідмовної роботи окремих компонентів за умов часу $T > t$, а $X = \left(X_{i_1}, X_{i_2}, \ldots, X_{i_n}\right)$ – вектор параметрів системи (наприклад, напруга, температура, інтенсивність навантаження). Індекси $i_1, i_2, \ldots, i_n$ відповідають станам окремих компонентів системи.

Для складних систем із взаємозалежними компонентами пропонується застосовувати тензор кореляційних коефіцієнтів, який враховує зв'язки між станами компонентів. Тензор кореляцій можна визначити як:

$$K_{i_1 i_2 \ldots i_n} = \frac{Cov\left(X_{i_1}, X_{i_2}, \ldots, X_{i_n}\right)}{\sqrt{\prod_{k=1}^{n} Var\left(X_{i_k}\right)}},$$

де $Cov\left(X_{i_1}, X_{i_2}, \ldots, X_{i_n}\right)$ – тензор коваріацій, а $Var\left(X_{i_k}\right)$ – дисперсія змінної X_{i_k}.

Для опису надійності складених систем розглянемо тензор $\Phi_{i_1 i_2 \ldots i_n}$, що визначає залежність загальної надійності системи від станів її компонентів. Наприклад, для послідовної системи ця залежність може бути записана як:

$$\Phi_{i_1 i_2 \ldots i_n} = \prod_{k=1}^{n} R_{i_k}.$$

Для паралельної системи:

$$\Phi_{i_1 i_2 \ldots i_n} = 1 - \prod_{k=1}^{n} (1 - R_{i_k}).$$

Для визначення інтенсивності відмов будемо використовувати тензор відмов $\lambda_{i_1 i_2 \ldots i_n}$. Побудуємо тензорну модель динаміки змін надійності у часі за допомогою диференційних рівнянь з використанням коваріантного диференціювання:

$$\frac{DR_{i_1 i_2 \ldots i_n}}{Dt} = \lambda_{i_1 i_2 \ldots i_n} R_{i_1 i_2 \ldots i_n},$$

де $\frac{D}{Dt}$ – оператор коваріантного диференціювання, що враховує зміни стану системи у криволінійному просторі, заданому умовами експлуатації.

Таким чином, використання тензорного аналізу в теорії надійності дозволяє моделювати складні багатовимірні системи, враховувати взаємозв'язки між компонентами та змінність параметрів у часі. Такий підхід забезпечує глибший аналіз властивостей систем, оптимізацію їхньої роботи та передбачення можливих відмов. Тензорний підхід також відкриває нові можливості для розробки інженерних систем із заданими характеристиками надійності в умовах складних експлуатаційних середовищ.

Моделювання систем за допомогою випадкових процесів тензорними методами. Математичне моделювання систем за допомогою випадкових процесів є важливим підходом для аналізу складних явищ, що відбуваються в умовах невизначеності. Інтеграція випадкових процесів із тензорним аналізом дає можливість описувати взаємозв'язки між компонентами системи в багатовимірному просторі, враховуючи їхню стохастичну природу. Такий підхід відкриває нові можливості для прогнозування поведінки систем, оцінки ризиків і оптимізації їхньої роботи.

Розглянемо систему S, яка описується множиною параметрів, кожен з яких є випадковим процесом, що змінюється у часі $t \in [0,T]$. Позначимо тензор станів системи через $X(t) \in R^{n \times m}$, де n – кількість параметрів системи, а m – кількість вимірів або станів, у яких кожен параметр може перебувати. Кожен елемент $X_j^i(t)$ тензора $X(t)$ є випадковим процесом.

Побудуємо тензорну модель динаміки системи у стохастичному середовищі за допомогою стохастичних диференціальних рівнянь у тензорному вигляді:

$$dX_j^i(t) = A_j^i(X,t)dt + B_j^i(X,t)dW_j^i(t),$$

де $A_j^i(X,t)$ – тензор визначених (детермінованих) складових зміни стану системи, $B_j^i(X,t)$ – тензор стохастичних впливів, $W_j^i(t)$ – процес Вінера, який моделює випадкові впливи.

Через $F(X,t,\Delta t)$ позначимо тензор розподілу ймовірностей станів системи. Тоді для прогнозування поведінки системи будемо використовувати тензор ймовірності:

$$P_{ij}(X(t+\Delta t) \mid X(t)) = \frac{\partial^2 F(X,t,\Delta t)}{\partial X^i \partial X^j}.$$

Застосування тензорного аналізу до моделювання систем за допомогою випадкових процесів є важливим інструментом для вирішення задач у таких галузях, як економіка, біологія, фізика, інженерія та машинне навчання. Це забезпечує більш глибоке розуміння динаміки систем, підвищує точність прогнозування і дозволяє оптимізувати управління процесами в умовах невизначеності. Розробка таких моделей відкриває перспективи для створення більш ефективних і стабільних систем у реальному світі. Тензорна форма представлення випадкових процесів дозволяє не лише компактно описувати математичні моделі, але й забезпечує інваріантність рівнянь при зміні координат. Це особливо важливо для аналізу систем, що функціонують у різноманітних середовищах або під впливом багатовимірних зовнішніх факторів.

Таким чином, тензорний аналіз надає потужний математичний інструментарій для моделювання складних систем через вивчення їх багатовимірних взаємозв'язків.

Висновки. В статті розглянуто застосування тензорних методів для моделювання та оптимізації складних систем у машинному навчанні. Показано, що тензорний підхід, зокрема тензорні розклади, факторизація, тензорні ядрові методи та регуляризація, дозволяє зберегти структурну цілісність багатовимірних даних, зменшити розмірність простору ознак і, водночас, підвищити ефективність моделей. Використання тензорних обчислень у глибокому навчанні дозволяє зменшити кількість параметрів у нейронних мережах, оптимізувати обчислювальні ресурси та забезпечити кращу здатність до узагальнення. Отже, тензорний аналіз постає не лише як потужний математичний інструмент, а й як методологічна основа для побудови більш точних і масштабованих моделей машинного навчання, особливо в умовах великої кількості ознак і складних внутрішніх зв'язків між ними.

У контексті стрімкого розвитку штучного інтелекту та машинного навчання тензорні методи набули особливої актуальності завдяки своїй здатності ефективно моделювати багатовимірні дані, що притаманні складним системам.

Складні системи характеризуються високою розмірністю, гетерогенністю, наявністю латентних структур та сильними кореляціями між компонентами. Традиційні матричні підходи виявляються обмеженими при роботі з такими даними, тоді як тензори – багатовимірні узагальнення матриць – забезпечують природну структуру для зберігання і обробки інформації з кількома аспектами (час, простір, категорія, джерело тощо).

У даному дослідженні розглянуто застосування тензорних методів для аналізу надійності складних систем, зокрема у поєднанні з підходами стохастичного моделювання. Встановлено, що використання тензорного апарату дозволяє ефективно представляти багатовимірні залежності між елементами складної системи, зокрема у випадках, коли ці елементи піддаються випадковим впливам з часом. Тензорні моделі забезпечують природний спосіб представлення стохастичних процесів у вигляді багатовимірних масивів, що змінюються за часовими або структурними координатами, зберігаючи при цьому взаємозв'язки між компонентами системи. Було показано, що стохастичні процеси, описані тензорними структурами, дозволяють враховувати часові й структурні залежності у багатокомпонентних системах, що є неможливим у рамках класичних матричних моделей. Таким чином, поєднання тензорного аналізу з методами випадкових процесів створює нову якісну платформу для аналізу та прогнозування надійності складних систем, забезпечуючи гнучкість, масштабованість та високу інтерпретованість результатів.

Отже, тензорний підхід надає унікальні інструменти для аналізу складних систем у машинному навчанні. Його головні переваги – здатність працювати з високорозмірними даними без втрати структури, виявлення закономірностей. Подальші дослідження орієнтовані на розробку адаптивних тензорних архітектур, що поєднують автоматичне навчання структури тензора та інтеграцію з байєсівськими моделями для підвищення надійності.

Тензорні методи машинного навчання дозволяють розширити можливості обробки складних багатовимірних даних, покращити продуктивність моделей та спростити їхню інтерпретацію. Завдяки своїм математичним властивостям тензори забезпечують ефективне використання ресурсів і глибокий аналіз даних, що робить їх незамінним інструментом у сучасних алгоритмах штучного інтелекту. Подальше дослідження та вдосконалення тензорних методів відкриває нові перспективи для машинного навчання та оптимізації складних систем.

Список літератури:

1. Tamara G. Kolda, Brett W. Bader. Tensor Decompositions and Applications / Siam Review 2009. Society for Industrial and Applied Mathematics. Vol. 51, No. 3. P. 455–500.

2. Stoudenmire E.M., David J. Schwab. Supervised Learning with Tensor Networks / 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

3. Glen Evenbly, A Practical Guide to the Numerical Implementation of Tensor Networks I: Contractions, Decompositions and Gauge Freedom. Front. Appl. Math. Stat., 01 June 2022. Sec. Mathematics of Computation and Data Science. Volume 8 – 2022. https://doi.org/10.3389/fams.2022.806549

4. B. Staber and J. Guilleminot, Stochastic modeling and generation of random fields of elasticity tensors: A unified information-theoretic approach, C. R. Mécanique 345 (2017). P. 399–416. DOI: 10.1016/j.crme.2017.05.001

5. Nickel, M., Tresp, V. (2013). Tensor Factorization for Multi-relational Learning. In: Blockeel, H., Kersting, K., Nijssen, S., Železný, F. (eds) Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2013. Lecture Notes in Computer Science, vol 8190. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-40994-3_40

6. Yifan Zhao, Hashim Sharif, Vikram Adve, Sasa Misailovic, Felix: Optimizing Tensor Programs with Gradient Descent / April 2024. Conference: ASPLOS '24: 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 3, 2024. https://doi.org/10.1145/3620666.3651348

7. Vajjala S., Majumder B., Gupta, A., Surana, H. (2020), Practical Natural Language Processing: A Comprehensive Guide to Building Real-world NLP Systems, O'Reilly Media.

8. Pei-Chang Guo, New regularization methods for convolutional kernel tensors / September 2023. AIMS Mathematics 8 (11):26188-26198 DOI:10.3934/math.20231335

Hryhorieva T.I., Rozenvasser D.M., Khniunin S.H., Krokhmal M.V. ANALYSIS OF COMPLEX SYSTEMS WITH MACHINE LEARNING USING TENSOR METHODS

The article discusses the application of tensor methods for the analysis of complex systems in the context of machine learning. Given the growing volume of multidimensional and poorly structured data that is characteristic of modern applied problems, there is an urgent need to use powerful tools that can effectively process high-dimensional information while preserving the internal structure of the data. One such tool is tensor analysis, which allows for the formalization, decomposition, optimization, and interpretation of complex relationships between variables.

The paper considers tensor models of complex systems. The possibility of using tensors to represent dependencies between system components in a compact and generalized form is shown. The tensor approach ensures the invariance of equations during coordinate transformations. The use of tensor formalism for analyzing the reliability of complex systems provides the possibility of more accurate modeling of complex interactions and multidimensional dependencies. This approach opens up prospects for increasing the stability of systems through optimization based on mathematical methods based on tensors.

The paper presents key tensor approaches that are currently being actively implemented in machine learning algorithms. In particular, the methods of tensor decomposition and decomposition are considered, which allow reducing the dimensionality of data while preserving characteristics, which is especially relevant for classification, clustering, and recommendation tasks. Special attention is paid to tensor factorization, which is used to build compact models and detect hidden patterns in a multidimensional feature space. The use of tensor operations in deep learning is also considered – including convolutional neural networks, where tensor representations help reduce the computational complexity of models without reducing their accuracy. The role of tensor optimization methods, such as tensor gradient descent, in accelerating model training by taking into account the structural features of the parameter space is highlighted.

The article investigates tensor approaches to the analysis of the reliability of complex systems, which allow for the effective processing of multidimensional and structurally complex data. Particular attention is paid to the integration of tensor analysis with methods for modeling systems using stochastic processes. The results of the study demonstrate that the combination of the tensor approach and probabilistic modeling allows achieving high accuracy in predicting failures and reducing the computational complexity of the problem.

The results obtained confirm that tensor analysis is an extremely promising direction in the development of machine learning. Its use allows not only to increase the accuracy of models, but also to significantly reduce their complexity, which is critically important in application scenarios with limited resources. The proposed approaches can be applied in a wide range of tasks – from computer vision to predicting the behavior of complex systems in real time.

Key words: *complex systems, reliability of complex systems, mathematical modeling of complex systems, tensor models of complex systems, random processes, optimization, machine learning, tensor factorization, tensor decomposition methods, tensor decomposition.*

УДК 004.415.2
DOI https://doi.org/10.32782/2663-5941/2025.3.2/22

Гуржій С.В.
Український науково-дослідний інститут
спеціальної техніки та судових експертиз Служби безпеки України

ROS 2: ОСОБЛИВОСТІ АРХІТЕКТУРИ ТА ПРАКТИЧНОГО ВПРОВАДЖЕННЯ

Стаття присвячена дослідженню сучасних тенденцій та особливостей застосування покращеної версії операційної системи для робототехнічних платформ – Robot Operating System 2 (ROS 2). У статті розкрито еволюцію концептуального та архітектурного підходу до побудови програмного забезпечення для автономних роботів у порівнянні з попередньою версією – ROS 1. Розкрито функціональні можливості ROS 2 у контексті забезпечення модульності, масштабованості, надійності та безпеки в умовах гетерогенних і динамічних експлуатаційних середовищ. Визначено, що ROS 2 забезпечує підтримку життєвого циклу вузлів, розширені засоби міжвузлової комунікації (через DDS), багатоплатформність (Linux, Windows, macOS), а також підвищену продуктивність і безпеку при виконанні критичних для місій обчислень. З'ясовано, що впроваджені механізми управління якістю обслуговування (QoS), підтримка функціонування у реальному часі, можливість безпечного розподілу задач між модулями, а також чітко визначена структура розробки та тестування сприяють адаптації ROS 2 до широкого спектру задач – від наукових досліджень до комерційного впровадження. У статті проаналізовано ефективність застосування ROS 2 для безпілотних систем різних типів – повітряних (UAV), наземних (UGV) та морських (USV). Установлено, що ROS 2 пропонує високу ефективність у ситуаціях, які потребують стабільної міжвузлової взаємодії в умовах нестабільного зв'язку, інтеграції великої кількості сенсорів та забезпечення надійної навігації з підтримкою відмовостійкості. Також висвітлено тенденції впровадження ROS 2 у робототехнічних системах та підкреслено її здатність до прискорення процесу переходу від дослідницького прототипу до серійного продукту. Окреслено, що архітектурні інновації ROS 2, зокрема підтримка багатопоточності, гнучких шаблонів вузлів та автоматизованих засобів діагностики, є ключовими факторами успішного застосування даної операційної системи в складних багатокомпонентних системах автономного управління. Представлені дані доводять, що ROS 2 стає стандартом де-факто для розробки автономних робототехнічних систем наступного покоління, здатних функціонувати у складних, мінливих і багатозадачних середовищах. Таким чином, ROS 2 не лише усуває архітектурні та функціональні обмеження попередньої версії, але й формує підґрунтя для впровадження робототехнічних технологій у нових прикладних доменах, включно з промисловими, військовими, пошуково-рятувальними та сервісними застосуваннями.

Ключові слова: *робот, операційна система, ROS, алгоритм, DDS.*

Постановка проблеми. Швидкий розвиток інтелектуальних автономних систем, включаючи безпілотні літальні апарати (UAV – unmanned aerial vehicles), безпілотні наземні апарати (UGV – unmanned ground vehicles) і безпілотні надводні апарати (USV – unmanned surface vessel), зумовив необхідність розробки високомодульних, масштабованих і функціонально сумісних програмних фреймворків. Програмні середовища для робототехніки, які функціонують як базові інфраструктурні платформи, створюють основу для ефективної взаємодії сенсорних систем, актуаторів, керуючих алгоритмів та комунікаційних протоколів. У цьому контексті операційна система роботів (ROS – Robot Operating System) набула статусу провідного рішення з відкритим вихідним кодом, яке визначає вектор розвитку сучасних робототехнічних досліджень та їх практичної імплементації в промисловості.

Оригінальна ROS 1, яка була представлена дослідницькою лабораторією Willow Garage, ініціювала зміну парадигми, сприяючи створенню спільнотою системи багаторазових програмних пакетів, інструментів для моделювання та візуалізації, а також стандартизованих інтерфейсів для керування роботами [1]. ROS 1 значно знизила бар'єр для входу в розробку робототехнічних систем, сприяючи швидкому створенню прототипів та впровадженню інновацій. Однак, незважаючи на широке розповсюдження і корисність, ROS 1 не була спроектована з урахуванням критично важливих вимог до системного рівня. Їй бракувало вбудованої підтримки кібербезпеки, можливості детермінованого виконання в реаль-

ному часі, надійності в умовах погіршення якості мережевого середовища та формального контролю робочого циклу – факторів, необхідних для розгортання критично важливих для безпеки та орієнтованих на виконання завдання застосунків у гетерогенних середовищах.

Оскільки автономні роботизовані системи все частіше використовуються в реальних умовах – від моніторингу цивільної інфраструктури і логістики до оборони, реагування на катастрофи і дослідження космосу – зазначені обмеження стають все більш помітними. Спеціальні архітектурні модифікації і власні розширення, запроваджені розробниками для обходу недоліків ROS 1, виявили нагальну потребу у фундаментально модифікованій системній архітектурі.

У відповідь на зростаючі вимоги робототехнічне програмне забезпечення другого покоління, ROS 2, було повністю перебудоване та адаптоване для забезпечення надійності, масштабованості та безпеки в автономних системах високого рівня продуктивності. ROS 2 побудовано на основі комунікаційного робототехнічного програмного забезпечення Data Distribution Service (DDS) – стандартизованого протоколу обміну даними, який широко використовується в аерокосмічній, оборонній, фінансовій та промисловій галузях автоматизації. DDS додає до ROS 2 набір розширених функцій, таких як налаштування політики якості сервісу (QoS) в реальному часі, децентралізоване виявлення подій, безпечний зашифрований трафік та надійну підтримку гетерогенних і обмежених в ресурсах середовищ, включаючи вбудовані системи та периферійні обчислювальні платформи.

Незважаючи на ці досягнення, на шляху до широкомасштабного впровадження та оптимізації ROS 2 для конкретних прикладних програм, що включають широкий спектр повітряних, наземних та морських автономних платформ, залишається чимало невирішених питань. Такі питання, як синхронізація в реальному часі в ройовій робототехніці, контекстно-орієнтована навігація в умовах відсутності GPS і енергоефективні конвеєри сприйняття, потребують подальших досліджень і методологічного вдосконалення. Крім того, інтеграція ROS 2 з новими технологіями, в тому числі з прийняттям рішень на основі штучного інтелекту, розподіленим периферійним інтелектом і мережевими парадигмами наступного покоління (наприклад, 5G/6G, TSN), відкриває як можливості, так і технічні перешкоди, які необхідно подолати.

Аналіз останніх досліджень і публікацій. Еволюція архітектур керування роботами та робототехнічних платформ триває понад п’ять десятиліть, відображаючи безперервне вдосконалення принципів, які лежать в основі розвитку автономних систем. Так з появою систем, таких як Shakey the Robot, були закладені основи які впровадили уявлення про символічне мислення та планування в контролі поведінки роботів. Науково-технічні досягнення в області безпілотних систем отримали подальший розвиток завдяки удосконаленню модульного дизайну, впровадженню класичних стратегій планування, створенню архітектур на основі поведінки, а також розробці багаторівневих схем управління. Значним проривом у ранніх дослідженнях стала архітектура управління завданнями (TCA – Task Control Architecture), яка базувалася на використанні системи передачі повідомлень для координації поведінкових алгоритмів роботів. Ця концепція була використана в таких системах, як CARMEN, що використовували IPC (міжпроцесорну комунікацію) для полегшення модульного управління роботами. Основою сучасних робототехнічних систем є ефективна організація обміну повідомленнями, що забезпечує інтегровану взаємодію між розподіленими компонентами. Для реалізації цього підходу використовуються протоколи, перевірені в середовищах розподілених обчислень, зокрема MQSeries від IBM, Jini від Java та легковагові рішення типу MQTT. Інтеграція даних технологій у структури робототехнічних систем дозволила підійняти на перший план питання модульності програмного забезпечення, повторного використання компонентів та масштабованості розробок. Як наслідок, перші архітектури програмного забезпечення для роботів, такі як Player, базувалися на клієнт-серверних моделях, де центральний сервер безпосередньо взаємодіяв з апаратним забезпеченням для реалізації логіки управління, а клієнтські додатки отримували дані з сенсорів та актуаторів через TCP-з’єднання. Однак, незважаючи на свій внесок у абстракцію та віддалений доступ, архітектура Player виявила критичні недоліки в плані модульності, надійності та адаптивності. Подальші розробки програмного забезпечення для робототехніки, такі як Yet Another Robot Platform (YARP), намагалися подолати ці обмеження, впроваджуючи однорангові моделі зв’язку. YARP акцентувала увагу на гнучкості завдяки підтримці декількох протоколів зв’язку і забезпечувала високу

швидкодію в сценаріях управління в реальному часі. Це дозволило повторно використовувати код і сприяло спільній розробці [2]. Інший відомий фреймворк, Lightweight Communications and Marshalling (LCM), був адаптований до середовищ з високою пропускною здатністю і низькою затримкою, використовуючи архітектуру публікації/підписки з багатомовними прив'язками. LCM виявився ефективним для швидкого обміну даними в режимі реального часу, але залишався обмеженим для обміну повідомленнями та серіалізації, залишаючи завдання координації на системному рівні розробникам [3]. Доповнюючи LCM, проект Open Robot Control Software (OROCOS) зосередився на управлінні рухом в реальному часі і включав передові обчислювальні інструменти, такі як кінематичні розв'язувачі та байєсівська фільтрація [4]. Інтегруючи програмне забезпечення на основі CORBA, OROCOS підтримував детерміноване виконання та жорсткі часові обмеження. Однак, як LCM, так і OROCOS працювали лише з окремими підсистемами робототехніки, таким чином, не маючи комплексної основи для розробки наскрізних робототехнічних додатків.

У цьому контексті поява операційної системи для роботів (ROS) ознаменувала трансформаційний прогрес в екосистемах програмного забезпечення для робототехніки. ROS 1, вперше випущена в 2010 році, представила модульну архітектуру, яка включала широкий спектр функціональних можливостей, необхідних для автономних систем, в тому числі самоаналіз повідомлень, моніторинг процесів, управління перетворенням координат і синхронізацію часу. Вона підтримувала розподілену модель виконання з різноманітними бібліотеками, які включають сприйняття, планування, активацію і SLAM. Дозволяє швидко створювати прототипи і розгортати гетерогенні роботизовані платформи [5].

Постановка завдання. Сучасні тенденції розвитку безпілотних автономних систем, зокрема повітряних (UAV), наземних (UGV) та надводних (USV) дронів, супроводжуються зростаючими вимогами до їхньої функціональної надійності, структурної живучості, адаптивності до середовищ функціонування та загальної ефективності виконання місій різного рівня складності. Умови експлуатації таких систем охоплюють широкий спектр сценаріїв – від моніторингу, картографування та екологічного спостереження до високоризикованих військових і пошуково-рятувальних операцій. Внаслідок цього особливої актуальності набуває впровадження уніфікованих, масштабованих, модульних та інтероперабельних програмних архітектур, здатних забезпечити гнучке управління апаратними ресурсами та адаптацію до динамічних умов зовнішнього середовища. У даному контексті операційна система ROS 2 є одним із найперспективніших засобів побудови програмної архітектури автономних робототехнічних платформ завдяки підтримці DDS-комунікації, QoS, безпеки, сумісності з вбудованими системами. Такий підхід забезпечує реалізацію як централізованих, так і розподілених систем управління залежно від прикладних вимог.

У зв'язку з цим, метою даного дослідження є проведення комплексного аналізу архітектурних особливостей операційних систем типу ROS 2 у контексті їхньої ефективності при розробці та експлуатації різних типів автономних дронів. Зокрема, передбачається здійснити систематизацію функціональних можливостей ROS 2, оцінити її здатність до масштабування, забезпечення відмовостійкості, підтримки модульності та забезпечення безпечного управління. Також ставиться завдання визначити оптимальні конфігурації ROS 2 для різних типів роботизованих платформ, враховуючи специфіку їх апаратної реалізації та цільових сценаріїв використання.

Виклад основного матеріалу. У контексті еволюції програмних парадигм для роботизованих систем, поява версії ROS 1, датованої 2010 роком, стала знаковою віхою, яка започаткувала інноваційний підхід до розробки операційних систем для робототехнічних застосувань. Архітектурна концепція ROS 1 базувалася на модульності, інтегруючи широкий спектр функціональних блоків, критично важливих для забезпечення автономності роботизованих систем. До ключових компонентів належали механізми аналізу обміну повідомленнями, моніторингу процесів, керування трансформаціями координатних систем та синхронізації часових міток. Реалізація розподіленої моделі виконання, доповнена розгалуженою екосистемою бібліотек, розроблених зусиллями спільноти, охоплювала широкий спектр функціональності, включаючи перцепцію, планування траєкторій, керування приводами та алгоритми одночасної локалізації та картографування (SLAM – Simultaneous Localization and Mapping) [6]. Це сприяло прискореному прототипуванню та розгортанню гетерогенних роботизованих плат-

форм, зокрема безпілотних літальних, наземних, надводних апаратів.

Попри інноваційний підхід, ROS 1 мала низку архітектурних обмежень, що знижували її ефективність у критичних застосуваннях. Зокрема, централізована обробка даних створювала єдину точку відмови, що знижувало надійність у мережах із нестабільним або переривчастим зв'язком (наприклад, Wi-Fi чи супутникові канали). Крім того, базові протоколи TCP/UDP не забезпечували підтримки механізму забезпечення якості обслуговування (QoS – Quality of Service), що ускладнювало використання ROS 1 у системах реального часу та безпеки у критичних додатках. Питання кібербезпеки також залишались відкритими: хоча ініціативи на кшталт Security-Enhanced ROS (SROS) інтегрували шифрування та автентифікацію, їм бракувало стандартизованості та масштабованості, необхідних для відповідності сучасним вимогам у сфері інформаційної безпеки [7].

У відповідь на окреслені недоліки, розробка ROS 2 стала зрушенням у напрямку створення програмного забезпечення корпоративного рівня для робототехніки. ROS 2 інтегрує в якості фундаментальної комунікаційної інфраструктури стандарт Data Distribution Service (DDS) – високопродуктивний протокол, який підтвердив свою ефективність у критично важливих напрямках, таких як аерокосмічна інженерія та охорона здоров'я. Завдяки своїй архітектурі, DDS забезпечує нативну підтримку якості обслуговування (QoS), децентралізоване виявлення учасників та гарантовану низьку латентність, що суттєво розширює застосовність ROS 2 у розподілених системах, чутливих до часових затримок. Крім того, в ROS 2 запроваджено концепцію вузлів, що забезпечує детермінований контроль над переходами станів (наприклад, конфігурація, активація, деактивація) та є критично важливим для систем з високими вимогами до надійності та передбачуваності поведінки.

Ще однією ключовою архітектурною інновацією ROS 2 є спільна клієнтська бібліотека (rcl), яка є основою для реалізацій на різних мовах програмування, підвищуючи крос-платформну консистентність та мінімізуючи фрагментацію екосистеми. Модель багатопотоковості також зазнала значних покращень завдяки впровадженню плагінів, що надає розробникам гнучкість в адаптації стратегій паралелізму до специфічних вимог реального часу. ROS 2 підтримує багатовузлове виконання в межах одного процесу та запроваджує стандартизований інтерфейс на основі сервісів для керування параметрами, замінюючи застаріле рішення, що базувалося на XMLRPC, на більш надійні API з типізованими параметрами.

Структура архітектури системи ROS 2 складається з трьох основних підсистем, кожна з яких виконує окремі, але взаємозалежні функції, необхідні для розробки розподілених та автономних робототехнічних застосунків:

– комунікаційне програмне забезпечення (міжпроцесорна та розподілена інфраструктура обміну повідомленнями) – рівень програмного забезпечення в ROS 2 базується на стандарті DDS (Data Distribution Service) [8]. Цей протокол забезпечує надійну, масштабовану і малозатратну передачу повідомлень між обчислювальними вузлами, гарантуючи підтримку параметрів якості обслуговування (QoS), що є критично важливими для безпечних і чутливих до часу роботизованих операцій. Програмне забезпечення на основі DDS абстрагує комунікаційні API, серіалізацію/десеріалізацію даних та мережеві транспортні рівні, таким чином підтримуючи операції в реальному часі в середовищах зі змінними мережевими характеристиками, наприклад, з якими стикаються автономні дрони під час польотів на великі відстані або висоті, або підводні апарати в морському середовищі з нестабільним зв'язком;

– робототехнічні алгоритми та функціональні бібліотеки – ROS 2 надає широку колекцію багаторазових, параметризованих алгоритмічних модулів, що включають основні можливості робототехніки. До них відносяться прогресивні методи одночасної локалізації і картографування (SLAM), модулі розпізнавання з використанням глибокого навчання і об'єднання сенсорів (LiDAR, RGB-D, RADAR тощо), алгоритми планування руху і траєкторії в реальному часі (наприклад, RRT, DWA і оптимізація траєкторії), а також фреймворки координації поведінки, такі як дерева поведінки і ієрархічні кінцеві автомати. Ці компоненти виступають фундаментальними будівельними блоками для реалізації автономії безпілотних систем, які працюють в умовах відсутності GPS, роїв дронів, які виконують спільну розвідку, або автономних морських систем, які проводять розвідувальні місії;

– інструментарій розробника та операційні утиліти – для підтримки повного життєвого циклу роботизованого програмного забез-

печення – від проектування, моделювання та розгортання до моніторингу під час виконання та аналізу після завершення місії. ROS 2 пропонує комплексний набір інструментів на основі командного рядка та графічного інтерфейсу. Сюди входять інструменти для конфігурації системи (ros2 param, ros2 launch), самоаналізу в реальному часі (ros2 topic, ros2 node info), візуалізації стану (Rviz2) та 3D-симуляції середовища (Gazebo, Ignition та Webots). Інструментарій для безперервної інтеграції та розгортання (CI/CD) включає автоматизовані системи збірки (colcon), робочі процеси контролю версій (vcstool, rosinstall_generator) та розповсюдження бінарних файлів за допомогою ROS Index та ROS Buildfarm. Цей набір інструментів не лише прискорює цикли розробки, але й полегшує відтворюваність та масштабованість, що є критично важливими для розгортання автономних транспортних засобів на рівні кластерів на гетерогенних платформах [9].

На додаток до цих основних компонентів, ROS 2 активно підтримує розширення за допомогою пакетів і фреймворків, які орієнтовані на специфічні вимоги конкретної галузі. Наприклад, micro-ROS забезпечує сумісність ROS 2 з платформами на базі мікроконтролерів, що значно розширює її застосування для периферійних обчислень в умовах обмежених ресурсів, таких як мініатюрні дрони або вбудовані підсистеми в модульних платформах дронів. Аналогічно, інтеграція з операційними системами реального часу (RTOS), такими як FreeRTOS і NuttX, підвищує детермінованість і безпеку в критично важливих сценаріях.

Оптимальне використання платформи ROS 2 потребує ретельної систематизації та оцінки ключових архітектурних критеріїв, які складають фундамент для розробки та інтеграції автоматизованих робототехнічних систем. При цьому особливу увагу слід приділити необхідності проведення всебічного аналізу продуктивності ROS 2, що набуває особливої актуальності при розгортанні систем на базі безпілотних систем різного типу – від легких комерційних дронів до високотехнологічних індустріальних платформ. Такий підхід дозволяє не лише виявити поточні сильні сторони технології, але й сприяє подальшій оптимізації систем автономного управління в умовах складних експлуатаційних середовищ.

Однією з ключових особливостей є повна відмова від централізованої схеми управління на користь децентралізованої комунікаційної моделі, заснованої на протоколі DDS. Завдяки цьому досягається висока стійкість систем до відмов, що особливо актуально для автономних мультиробототехнічних груп, зокрема дронів, які здійснюють операції в умовах нестабільного або переривчастого зв'язку, таких як морські або аерокосмічні місії. Додатковою перевагою є покращене абстрагування за рахунок використання кросплатформеного клієнтського інтерфейсу rcl, що сприяє ефективному переносу програмного забезпечення між різними апаратними платформами, включно з системами реального часу. Такий підхід є важливим для побудови гібридних архітектур із використанням мікроконтролерів та вбудованих систем. ROS 2 також підтримує асинхронну подієво-керовану архітектуру, що дозволяє оперативно реагувати на зміну ситуацій та ефективно розподіляти обчислювальні ресурси. Це критично для дронів, які виконують задачі навігації або ухилення від перешкод у реальному часі. Водночас модульна структура ROS 2 забезпечує високу масштабованість і повторне використання компонентів, що сприяє зниженню витрат на розробку та спрощує реалізацію розподілених функцій, таких як локалізація, планування або міжагентна взаємодія.

Удосконалена комунікаційна інфраструктура ROS 2 забезпечує низку ключових переваг для побудови безпілотних систем різного типу завдяки реалізації трьох основних парадигм взаємодії: топіки, сервіси та дії. Комунікація на основі топіків, яка використовує модель "publish-subscribe", особливо ефективна для високочастотної передачі даних, як-от у наземних UGV-платформах, де необхідна паралельна обробка інформації з низькою затримкою. Інтеграція DDS із підтримкою гнучких налаштувань QoS дозволяє дронам типу UAV зберігати надійність зв'язку навіть у складних умовах із втратами пакетів понад 40% [10]. Парадигма сервісної взаємодії дозволяє вирішувати проблеми синхронізації, що критично важливо для морських USV, де обробка запитів з налаштовуваним тайм-аутом підвищує надійність при плануванні маршруту. Компонентна реалізація дій у ROS 2 забезпечує розширену підтримку станів машин, дозволяючи реалізовувати довготривалі автономні завдання з постійним зворотним зв'язком, що особливо корисно для USV в умовах динамічного морського середовища.

З погляду архітектури міжпрограмного забезпечення ROS 2 реалізує багаторівневу модель абстракції, яка базується на універсальному

інтерфейсі (rcl – клієнтська бібліотека ROS). Цей інтерфейс підтримує міжмовну сумісність та є важливою характеристикою для мультидисциплінарних проєктів. Механізм життєвого циклу вузлів забезпечує стандартизоване управління станами, що знижує ймовірність втрат ресурсів і покращує стійкість системи при нештатних ситуаціях. Аналіз результатів тестування свідчить про зменшення відмов на 47% у порівнянні з ROS 1 завдяки покращеному управлінню помилками. Застосування удосконаленої системи API-інтерфейсів сприяє зниженню інтеграційних помилок у багатокомпонентних системах на більш ніж 30% [11].

Також значно модернізована інфраструктура тестування: підтримка симуляцій та валідації в реальному часі дозволяє виявляти потенційні дефекти ще до розгортання в реальному середовищі. Базові пакети ROS 2 мають тестове покриття, яке перевищує 90%, що суттєво відрізняє їх від інших фреймворків.

Процес розробки документується через механізм ROS Enhancement Proposal (REP), що підвищує прозорість архітектурних рішень, однак якість документації залишається нерівномірною. Слід зазначити, що для розробників критичних безпілотних систем такий підхід створює додаткові вимоги до верифікації сторонніх компонентів перед їх впровадженням.

Швидкодія та надійність ROS 2 характеризуються оптимізацією для ресурсообмежених середовищ безпілотних систем. Модернізація розподілу пам'яті, зокрема впровадження об'єднаних розподільників, мінімізує фрагментацію стеку, що за результатами тривалих стрес-тестів знижує ймовірність переповнення пам'яті орієнтовно на 53%. Розширені можливості роботи в реальному часі досягаються завдяки механізмам успадкування пріоритетів та часово-орієнтованому плануванню. Емпіричні дані свідчать про досяжність субмілісекундної точності планування (0,84 мс) у належно конфігурованих системах ROS 2, що є достатнім для більшості застосувань керування безпілотними транспортними засобами [12]. Проте, досягнення високих показників реального часу вимагає ретельної системної конфігурації та потенційних модифікацій ядра.

Відмовостійкість ROS 2 суттєво покращена завдяки багаторівневій ізоляції помилок на архітектурному рівні. Вузлова архітектура локалізує збої в межах окремих компонентів, запобігаючи каскадним системним відмовам. Засоби відстеження подій забезпечують детальну діагностичну інформацію, що прискорює аналіз першопричин несправностей приблизно на 47% порівняно з традиційними методами ведення журналів.

Архітектура безпеки ROS 2 значно вдосконалена завдяки комплексній моделі безпеки. Інфраструктура сертифікатів X.509 забезпечує криптографічну ідентифікацію, усуваючи вразливості автентифікації в розподілених роботизованих системах та ефективно протидіючи поширеним векторам атак, включаючи імітацію вузлів та несанкціоновану модифікацію параметрів. Завдяки дозволам DDS Security Domain Participant вдалося підвищити рівень деталізації контролю доступу, що забезпечує оптимальне управління обмеженнями ресурсів у межах операційних потреб. У контексті багатовузлових систем UGV ця функціональність гарантує безпечну взаємодію між різними суб'єктами контролю, підтримуючи необхідний рівень ізоляції. Перевірка дозволів здійснюється на етапі виявлення, запобігаючи встановленню неавторизованих каналів зв'язку.

Безпека та цілісність системи в ROS 2 інтегрована на рівні специфікації DDS Security. Впровадження автентифікації вузлів, криптованого зв'язку, контролю доступу та верифікації журналів забезпечує відповідність високим вимогам безпеки, критичним для оборонних та розвідувальних операцій з використанням безпілотних платформ. На відміну від фрагментарних рішень безпеки ROS 1 (наприклад, SROS), ROS 2 надає комплексні засоби наскрізного управління безпекою, що є принциповим для UAV та UGV, які передають сенситивні дані через потенційно вразливі комунікаційні канали.

Криптографічний захист реалізовано на кількох рівнях протоколу, забезпечуючи багаторівневий захист від різноманітних загроз. До того ж застосування шифрування AES-GCM-GMAC гарантує як конфіденційність, так і цілісність даних з мінімальними обчислювальними витратами (зазвичай 3–5% додаткового завантаження процесора). Аналіз показує, що для платформ з обмеженими ресурсами можливе селективне шифрування критично важливих комунікаційних каналів, зберігаючи необхідний рівень безпеки ключових підсистем.

Інтеграція ROS 2 зі стеком мікро-ROS значно розширює сферу застосування, забезпечуючи функціональність ROS 2 на мікроконтролерах з ультранизьким енергоспоживанням та обмеженими обчислювальними ресурсами. Це відкриває

перспективи для створення гетерогенних роботизованих систем, в яких периферійний інтелект може бути реалізований безпосередньо на сенсорах, актуаторах або контролерах польоту. Для UAV це означає можливість незалежного функціонування контурів керування в реальному часі (наприклад, стабілізації висоти або інерціальної навігації) від основного обчислювального блоку, підвищуючи загальну надійність та відмовостійкість. Аналогічно, UGV можуть перенести низькорівневе керування приводами або попередню обробку даних сенсорів на вузли мікроконтролерів, що сприяє побудові ієрархічних архітектур керування.

Згідно з актуальними науковими даними, ROS 2 ефективно оперує в різнорідних мережевих середовищах, які характеризуються значною втратою пакетів, нестабільною затримкою та використанням множинних транспортних протоколів (Wi-Fi, мобільні мережі 4G/5G, супутниковий зв'язок). На рівні DDS імплементовані політики якості обслуговування, які надають інженерам можливість оптимізувати комунікаційні параметри відповідно до місії, збалансовуючи надійність, часові рамки та управління даними. Ця функціональність має вирішальне значення при експлуатації безпілотних повітряних апаратів на великих відстанях та безпілотних надводних суден.

Інтеграція з операційними системами реального часу (PREEMPT-RT, NuttX) та низьколатентними профілями DDS забезпечує часовий детермінізм в ROS 2. Результати кількісних експериментів підтверджують ефективність такого підходу для створення прецизійних систем керування, необхідних для високодинамічних маневрів UAV та точного позиціонування UGV.

Дослідження показують, що ROS 2 має високу промислову готовність, що підтверджується підтримкою Open Robotics Foundation та значними технологічними гігантами, зокрема NVIDIA, Amazon, Bosch і Apex.AI. Дистрибутиви з довготривалою підтримкою, а також детерміновані обчислювальні конвеєри та відповідність стандартам ISO 26262 та MISRA-C++ підкреслюють здатність платформи до використання в критично важливих комерційних проектах. Таким чином, ROS 2 являє собою сертифіковану та валідовану базу для розробників безпілотних систем різного призначення.

Висновки. У контексті застосування ROS 2 для різних типів безпілотних платформ встановлено низку техніко-функціональних переваг, які дозволяють підвищити ефективність їх автономного функціонування. Зокрема, для UAV ключовими є можливості конфігурації параметрів QoS, які забезпечують надійний зв'язок у нестабільному бездротовому середовищі, підтримка легких клієнтських бібліотек для розгортання на польотних контролерах, механізм управління життєвим циклом вузлів для забезпечення передбачуваної поведінки під час критичних фаз польоту, а також інтегровані засоби кібербезпеки для запобігання несанкціонованому втручанню. У випадку UGV ROS 2 пропонує переваги завдяки високопродуктивним схемам обміну даними для опрацювання значної кількості сенсорної інформації, гнучкій модульній архітектурі, що спрощує інтеграцію апаратних прискорювачів, використанню стандартизованих навігаційних стеків для вирішення задач навігації, а також реалізації механізмів відмовостійкості, що є критичними для тривалих місій у складних умовах. Для USV платформ ROS 2 забезпечує ефективну підтримку розподілених операцій на кількох суднах із нестабільним каналом зв'язку, можливість реалізації довготривалих дій з моніторингом виконання завдань, захист від несанкціонованого доступу в умовах мережевої експлуатації, а також підтримку управління в реальному часі, що є необхідним для точного маневрування в динамічному морському середовищі. Таким чином, ROS 2 є універсальною та високоефективною програмною платформою для реалізації автономних функцій у безпілотних літальних, наземних і морських системах.

Список літератури:

1. Morgan Q., Brian G., Conley K., Faust J., Foote T., Leibs J., Berger E., Wheeler R. ROS: an open-source Robot Operating System, IEEE International Conference on Robotics and Automation Workshop on Open Source Software, 2009. P. 1–6.

2. Vaughan R.T., Dellaert F., O'Connor J., Kuffner J.J. Player and YARP: A Study of Robot Software Architectures, Proceedings of the IEEE International Conference on Robotics and Automation, 2006. P. 2213–2218.

3. Fox A., J. W. T. Lightweight Communications and Marshalling (LCM): A High-Performance Middleware for Robotics, Proceedings of the IEEE International Conference on Robotics and Automation, 2011. P. 133–138.

4. Bruyninckx H., Soetens P., Koninckx B. The real-time motion control core of the Orocos project, Proceedings of the 2003 IEEE International Conference on Robotics and Automation (ICRA), Taipei, IEEE; 2003. P. 2766–2771.

5. Quigley M., Conley K., Gerkey B., Faust J., Foote T., Leibs J., Wheeler R. ROS: an open-source Robot Operating System, ICRA Workshop on Open Source Software, Vol. 3, No. 3.2, 2009. P. 5–10.

6. Macenski S., Foote T., Gerkey B. Towards a Distributed and Real-Time Framework for Robots: Evaluation of ROS 2.0 Communications for Real-World Robotic Applications, Autonomous Robots, 45, 2021. P. 611–626.

7. Dieber B., Breiling B., Taurer S., Kacianka S., Rass S. Security for the Robot Operating System, Robotics and Autonomous Systems, 98, 2017. P. 192–203.

8. Maruyama Y., Kato S., Azumi T. Performance analysis of a software platform based on ROS 2 and DDS implementations, Journal of Systems Architecture, Vol. 112, 2021. P. 101861–101877.

9. Quigley M., Gerkey B., Smart W.D. Programming Robots with ROS 2: A Practical Introduction to the Robot Operating System, O'Reilly Media, 2022. P. 445.

10. Kronauer T., Pohlmann J., Matthies M., Rust T. Reliable Communication in Lossy Networks for UAVs: A Quantitative Study of ROS 2 QoS Policies, Journal of Field Robotics. Vol. 39. No. 3, 2023. P. 340–358.

11. White R., Christensen H., Quigley M. API Design Patterns for Robotics: Lessons Learned from ROS 2, International Conference on Intelligent Robots and Systems (IROS), 2022. P. 3417–3424.

12. Casini D., Biondi A., Buttazzo G. Response-Time Analysis of ROS 2 Processing Chains under Reservation-Based Scheduling, IEEE Transactions on Computers. Vol. 71. No. 7, 2022. P. 1459–1474.

Hurzhii S.V. ROS 2: ARCHITECTURE AND PRACTICAL IMPLEMENTATION FEATURES

The article is devoted to the study of current trends and peculiarities of application of the improved version of the operating system for robotic platforms – Robot Operating System 2 (ROS 2). The article reveals the evolution of the conceptual and architectural approach to building software for autonomous robots in comparison with the previous version – ROS 1. The functional capabilities of ROS 2 in the context of ensuring modularity, scalability, reliability and safety in heterogeneous and dynamic operating environments are revealed. It is determined that ROS 2 provides support for the node life cycle, advanced means of inter-node communication (via DDS), multiplatform (Linux, Windows, macOS), as well as increased performance and security in performing mission-critical computing. It has been found that the implemented mechanisms for managing the quality of service (QoS), support for real-time operation, the ability to securely distribute tasks between modules, as well as a clearly defined development and testing structure, contribute to the adaptation of ROS 2 to a wide range of tasks – from scientific research to commercial implementation. The article analyses the effectiveness of ROS 2 for unmanned systems of various types – airborne (UAV), ground (UGV) and marine (USV). It is established that ROS 2 offers high efficiency in situations requiring stable inter-node interaction in conditions of unstable communication, integration of a large number of sensors and ensuring reliable navigation with fault tolerance. The article also highlights the trends in the implementation of ROS 2 in robotic systems and emphasises its ability to accelerate the process of transition from a research prototype to a serial product. It is outlined that the architectural innovations of ROS 2, in particular support for multithreading, flexible node templates and automated diagnostic tools, are key factors in the successful application of this operating system in complex multi-component autonomous control systems. The data presented here prove that ROS 2 is becoming the de facto standard for the development of next-generation autonomous robotic systems capable of operating in complex, changing and multitasking environments. Thus, ROS 2 not only eliminates the architectural and functional limitations of the previous version, but also forms the basis for the introduction of robotic technologies in new application domains, including industrial, military, search and rescue, and service applications.

Key words: *robot, operating system, ROS, algorithm, DDS.*

УДК 004.9
DOI https://doi.org/10.32782/2663-5941/2025.3.2/23

Дем'янчук Т.М.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Губар В.Г.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

АНАЛІЗ ЗАСТОСУВАННЯ МІР ПОДІБНОСТІ З МЕТОЮ ПОРІВНЯННЯ СИГНАТУР СИГНАЛІВ БПЛА

У даній статті розглядається застосування метричних методів кількісного аналізу подібності сигналів у контексті виявлення та ідентифікації електромагнітного випромінювання безпілотних літальних апаратів (БПЛА) на основі їх сигнатур, отриманих у ході радіоелектронної розвідки (РЕР). Дослідження має на меті продемонструвати потенціал використання формалізованих математичних підходів, зокрема метричних методів подібності, для підвищення точності класифікації сигналів, які надходять із різних джерел – особливо в умовах зашумлення, фрагментації чи інших викривлень, що часто трапляються у реальному бойовому середовищі. В основі дослідження лежить концепція аналізу сигнатур сигналів БПЛА розвідників типу «Орлан», що широко використовуються противником.

Виявлення сигнатур сигналів БПЛА є складним завданням через наявність природних та штучно створених завад, динамічність радіочастотного середовища, а також змінюваність параметрів сигналів під впливом умов передачі, переадресації та обробки. Традиційні підходи до розпізнавання, як-от фіксація порогових рівнів потужності або аналіз форм сигналу вручну, не забезпечують необхідної надійності, особливо в умовах автоматичного або напівавтоматичного моніторингу. У цьому контексті актуальним є застосування кількісних методів оцінки подібності, які дають змогу порівнювати реєстровані сигнали із заздалегідь створеними еталонними сигнатурами. Це дозволяє реалізувати ефективні алгоритми, здатні ідентифікувати тип БПЛА або інше джерело сигналу навіть за наявності значного рівня шуму або за часткової відсутності сигналу.

У рамках цього дослідження була сформована методика, яка включає кілька ключових етапів: вилучення комплексного сигналу з цифрового запису у форматі WAV, побудову спектрального представлення сигналу, виділення фрагменту спектрограми (сигнатури), подальше усереднення даних для зниження впливу флуктуацій, додавання штучного шуму контрольованих параметрів, та порівняння отриманої зашумленої сигнатури з початковою (референсною). Шум моделюється у вигляді комплексного додатку з випадковою амплітудною складовою, з можливістю введення імпульсних сплесків, що моделює несприятливі умови реального радіоелектронного середовища.

Один із розглянутих підходів до аналізу подібності сигналів ґрунтується на використанні коефіцієнта кореляції Пірсона, який дозволяє оцінити лінійний взаємозв'язок між двома векторами даних – еталонною та зашумленою сигнатурою. Цей метод забезпечує числову інтерпретацію рівня подібності та може застосовуватись як індикатор впізнаваності сигналу. Водночас, через свою обмеженість – зокрема, нечутливість до нелінійних залежностей і залежність від масштабу – він розглядається у межах ширшого набору метричних підходів. У роботі також досліджуються альтернативні метрики, зокрема косинусна подібність, рангові коефіцієнти Спірмена і Кендалла, евклідова, мангеттенська та чебишевська відстані. Такий комплексний підхід дозволяє оцінити сигнатурну подібність з різних точок зору та формує універсальну платформу для метричного аналізу сигнальних даних в умовах шуму та часткової фрагментації.

Для реалізації обчислень було використано мову програмування Python, яка завдяки своїм бібліотекам (NumPy, SciPy, Matplotlib тощо) забезпечує високоефективне опрацювання масивів даних, виконання чисельних операцій, побудову графіків та спектрограм. Однією з важливих частин дослідження став візуальний аналіз, який допомагає експериментально підтвердити ступінь подібності між сигналами, а також виявити нетипові або аномальні випадки, де метрика може давати хибні оцінки. Побудовані графіки сигналів дають змогу інтерпретувати числові результати у зрозумілому вигляді, що особливо важливо при подальшій інтеграції запропонованого підходу до реальних систем аналізу.

У результаті дослідження було продемонстровано, що навіть при значному рівні зашумлення (додавання комплексного шуму), коефіцієнти кореляції можуть зберігати відносно високе значення, що свідчить про збереження основної структури сигналу. Це дозволяє зробити висновок про потенційну ефективність використання цього підходу для попереднього аналізу та фільтрації даних у системах РЕР.

Практична цінність отриманих результатів полягає в тому, що вони можуть бути використані для створення автоматизованих або напівавтоматизованих систем виявлення БПЛА, інтегрованих у складі мобільних або стаціонарних радіомоніторингових комплексів. Такі системи можуть працювати у режимі реального часу, порівнюючи нові сигнали із базою відомих сигнатур, що дозволяє оперативно реагувати на загрози. Крім того, підхід може бути масштабований для роботи з широким спектром типів сигналів, а не лише з БПЛА «Орлан», що робить його універсальним інструментом для аналізу радіоелектронної обстановки у складних умовах.

Таким чином, проведений огляд демонструє підхід до аналізу сигналів у радіоелектронній розвідці, заснований на застосуванні метричних методів подібності. Запропонована методика є ефективною, гнучкою та перспективною для подальшого вдосконалення. Вона може бути інтегрована до систем аналізу з використанням елементів штучного інтелекту або машинного навчання, де метрики подібності використовуються як вхідні ознаки або критерії подальшої класифікації. Надалі можливе також тестування алгоритмів на великих масивах емпіричних даних, що дозволить покращити точність і чутливість систем виявлення БПЛА в реальних умовах.

Ключові слова: *міри подібності, виявлення БПЛА, спектральний аналіз, SDR, класифікація сигналів, радіочастотний моніторинг.*

Постановка проблеми. У сучасних умовах ведення війни, що відзначаються високою інтенсивністю бойових дій, застосуванням інтелектуальних систем озброєння, а також активною участю високотехнологічних рішень, виникає нагальна потреба у вдосконаленні засобів виявлення, спостереження та протидії загрозам з повітря. Безпілотні літальні апарати (БПЛА) стали одним з основних інструментів ведення розвідки, коригування вогню, здійснення точкових ударів, забезпечення ретрансляції зв'язку тощо. Масштабне вторгнення Російської Федерації на територію України у 2022 році стало каталізатором стрімкого зростання кількості БПЛА, які використовуються обома сторонами конфлікту, що, у свою чергу, обумовило нові виклики для систем протидії.

БПЛА різних класів – від побутових дронів до тактичних платформ типу «Орлан» – активно використовуються для проникнення в тил, збору розвідувальної інформації, ураження цілей або виконання радіоелектронних завдань. У зв'язку з цим зростає значення технологій радіоелектронної розвідки (РЕР) та радіоелектронної боротьби (РЕБ), які повинні оперативно виявляти джерела сигналів, аналізувати їх і класифікувати. На перший план виходить завдання ідентифікації БПЛА не лише за візуальними або акустичними ознаками, але й за характеристиками електромагнітного випромінювання.

Кожен дрон або наземна станція керування випромінює унікальний за спектральними характеристиками сигнал, який можна охарактеризувати через так звану сигнатуру – частотне представлення, що відображає властивості модуляції, ширину смуги, спектральну щільність потужності, рівень шуму. У разі БПЛА типу «Орлан», що активно використовується військами РФ для розвідки і коригування артилерії, сигнатури сигналів мають чіткі повторювані ознаки, які, за умови правильного вилучення та обробки, можуть бути надійним індикатором присутності БПЛА у повітряному просторі.

Завдяки розвитку програмно-визначеного радіо (SDR) з'явилася можливість високоточних записів широкосмугових сигналів, їх збереження, постобробки і візуального аналізу спектрограм. SDR відкриває нові горизонти для РЕР, однак одночасно вимагає наявності методів, які здатні з великого масиву сигналів виділяти ті, що потенційно є сигнатурами БПЛА. Через високу динаміку умов (наявність перешкод, зміна частот, періодів мовчання тощо), сигнали часто є фрагментарними, частково зашумленими або взагалі не відповідають формальним шаблонам.

У таких умовах традиційні методи детектування, які базуються, наприклад, на пороговому аналізі амплітуд, фіксованих спектральних маркерах або передбачуваних структурах пакету, втрачають ефективність. Саме тому зростає інтерес до методів оцінки подібності, що дозволяють порівнювати новий сигнал із базою еталонних сигнатур і робити висновок щодо ймовірної відповідності між ними. Такий підхід

відкриває перспективи для реалізації інтелектуальних алгоритмів класифікації, які здатні виявляти навіть модифіковані або частково спотворені сигнали.

Одним із найпростіших і водночас ефективних підходів до такого порівняння є використання метрик подібності. Найвідоміша з них – коефіцієнт кореляції Пірсона, що дозволяє визначити ступінь лінійної залежності між двома масивами даних. У випадку аналізу сигнатур, це може означати, що, незважаючи на наявність шуму або масштабування, загальний характер сигналу зберігається, і він має високу подібність до оригіналу. Однак Пірсонова кореляція чутлива до нелінійних викривлень, тому доцільним є паралельне використання альтернативних метрик, таких як косинусна подібність (яка порівнює напрямки вектора ознак), рангові коефіцієнти (наприклад, Спірмена), а також відстані – евклідова, мангеттенська, Чебишова тощо.

Застосування цих метрик особливо актуальне в умовах, коли сигнал є неідеальним – наприклад, коли фіксується лише частина спектру, або коли на нього накладаються завади (внутрішні від SDR, зовнішні з довкілля, або навіть спрямовані атаки за допомогою генераторів завад). Щоб оцінити ефективність методів подібності в таких умовах, доцільно створювати контрольовані експерименти, у яких до еталонної сигнатури додається штучний шум із регульованими характеристиками, після чого проводиться порівняння із «чистим» сигналом. Таким чином можна емпірично оцінити, які саме метрики залишаються стійкими до шуму, і за яких умов сигнал ще можна вважати впізнаваним.

У перспективі, такий підхід дозволить не лише виявляти присутність відомого БПЛА у зоні моніторингу, але й оцінювати ступінь невизначеності в сигналі, сигналізуючи оператору про необхідність перегляду або уточнення еталонної сигнатури. Більше того, це створює підґрунтя для навчання систем виявлення на основі накопичення та аналізу сигнатур, включаючи аномальні, спотворені, або нові типи випромінювання.

Отже, огляд метрик подібності є частиною ширшого наукового напрямку – розробки інтелектуальних систем виявлення на основі спектрального аналізу та метричних оцінок подібності. Вони мають важливе значення не лише в контексті сучасного військового конфлікту, а й у сфері цивільної безпеки, моніторингу повітряного простору, а також у наукових дослідженнях електромагнітних впливів і класифікації джерел випромінювання в умовах невизначеності.

Аналіз останніх досліджень і публікацій. У контексті виявлення та класифікації сигнатур сигналів безпілотних літальних апаратів (БПЛА) спостерігається зростання інтересу до застосування метричних методів для оцінки подібності сигналів, особливо в умовах впливу шуму, фрагментації та динаміки радіочастотного середовища. У науковій літературі широке поширення отримали дослідження, присвячені застосуванню кореляційних методів для аналізу сигналів у системах зв'язку та радіоелектронної розвідки. Наприклад, коефіцієнт кореляції Пірсона традиційно використовується для оцінювання лінійного взаємозв'язку між сигналами, однак його чутливість до масштабних змін та відсутність врахування нелінійностей обмежують сферу ефективного застосування.

Попередні дослідження [1–3] демонструють доцільність використання спектральних ознак для ідентифікації джерел випромінювання, проте часто обмежуються формальним порівнянням пікових частот, ширини смуги або рівня шуму. У той час як сигнатурний підхід передбачає комплексне трактування сигналу як узагальненої часово-частотної структури, що є більш інформативним і стійким до завад. У цьому контексті все частіше використовуються методи обробки спектрограм та машинного навчання для розпізнавання сигнатур [4–6], однак ці підходи часто потребують великого обсягу навчальних даних і не завжди добре працюють у реальному часі.

Аналіз літератури свідчить, що порівняно мало уваги приділяється адаптивним методам метричного аналізу сигнатур у зашумленому середовищі. Тому виникає потреба у формуванні універсального підходу, який дозволяє здійснювати гнучке порівняння сигналів за допомогою набору метрик та методів нормалізації, адаптованих до характеристик БПЛА-сигналів.

У статті, що аналізується, автори запропонували підхід до порівняння сигнатур на основі формалізованих метричних методів, що поєднує класичні статистичні індикатори з можливістю моделювання умов зашумлення. Важливою особливістю є використання спектрального представлення сигналу (через спектрограми) та його обробка із залученням засобів мов програмування (зокрема, Python).

Таким чином, порівняно з попередніми дослідженнями, дана робота виокремлюється цілісним експериментальним підходом, в якому симулю-

ються типові умови завад, а також застосовуються одразу декілька метрик подібності. Це відкриває шлях до створення систем виявлення, здатних працювати у реальному часі й виявляти сигнатури БПЛА з високою точністю навіть за несприятливих умов. Актуальність підходу підтверджується зростанням практичної потреби у таких системах у межах сучасної гібридної війни, зокрема в умовах протидії БПЛА типу «Орлан».

Постановка завдання. У сучасних умовах ведення гібридної війни, як уже зазначалося вище, дедалі зростає роль безпілотних літальних апаратів, які широко використовуються у розвідувальних та ударних операціях. У зв'язку з цим зростає необхідність у розробці ефективних методів виявлення та класифікації таких об'єктів на основі їхніх сигнатур. Особливу цінність у таких завданнях мають алгоритми, що здатні працювати в режимі реального часу з високою точністю, в умовах сильного радіочастотного зашумлення, фрагментації сигналу та обмеженого обсягу даних. Саме тому у цій статті метою є огляд методів метричного порівняння сигнатур БПЛА «Орлан», що можуть застосовуватися у згаданих алгоритмах.

Основним об'єктом дослідження є радіочастотна сигнатура (Рисунок 1) БПЛА типу «Орлан», отримана із записаного реального сигналу у вигляді файлу .wav. Цей сигнал обробляється для виділення так званого водопаду (Рисунок 2) – двовимірного масиву, що відображає зміну частотного спектра в часі. Такий підхід дозволяє зосередитися на часо-частотній структурі сигналу, яка є унікальною для кожного типу передавача та апаратури БПЛА.

З технічного боку реалізовано декілька ключових етапів. Перший – це формування спектрограми сигналу за допомогою алгоритму швидкого перетворення Фур'є, який дозволяє отримати спектральний склад сигналу. У цьому процесі враховано як реальну, так і уявну частини сигналу, закодовані у двоканальному .wav файлі, що відповідає комплексному представленню електромагнітного коливання. Подальший аналіз проводиться над вибраним фрагментом спектрограми, із якого обчислюється усереднений профіль – репрезентативна (референсна) сигнатура, яка використовується для порівняння.

Одним з основних викликів, який постає при роботі з реальними сигналами БПЛА, є високий рівень шуму, спотворення та фрагментарність. У роботі моделюються ці умови за допомогою функції, яка додає до сигналу складову випадкового шуму з амплітудними сплесками, симулюючи типові перешкоди в бойових умовах. Це дозволяє протестувати стійкість обраних метрик до завад та оцінити їхню придатність до реального застосування.

У рамках роботи застосовано набір метричних функцій, кожна з яких по-різному вимірює схожість між «референсною» сигнатурою та її зашумленою версією.

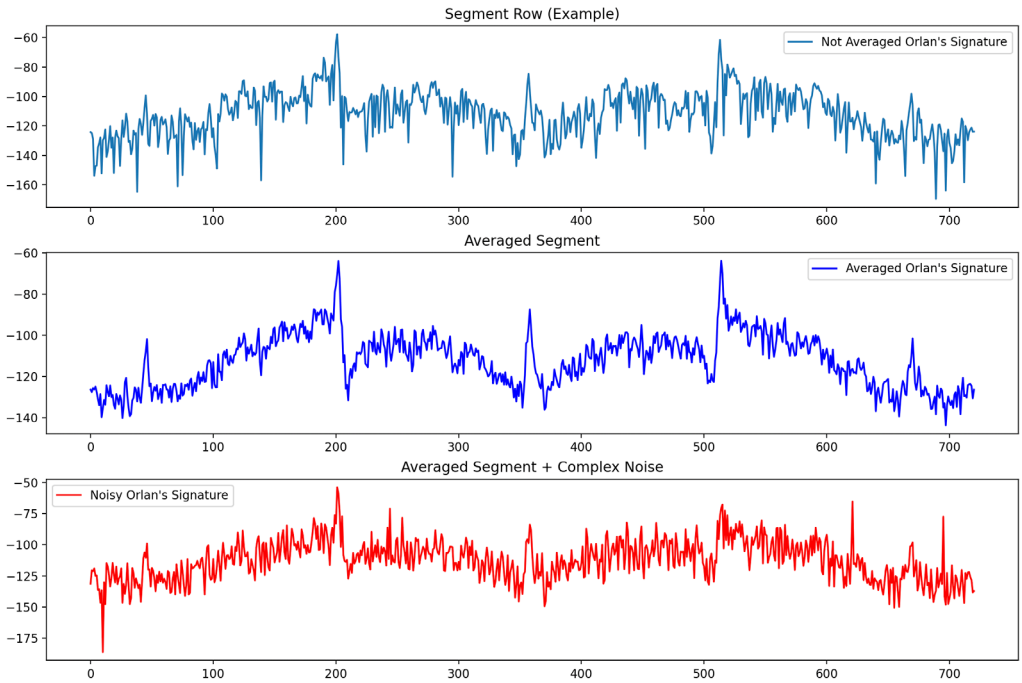


Рис. 1. Сигнатури сигналу БПЛА «Орлан»

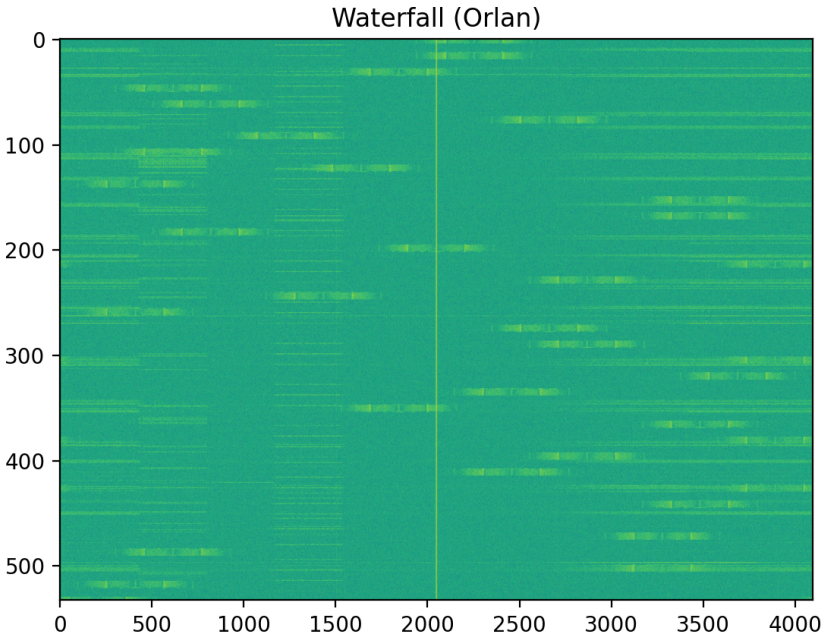


Рис. 2. «Waterfall» БПЛА «Орлан»

Виклад основного матеріалу. Застосування мір подібності для порівняння між собою референсної сигнатури (Рисунок 1, синій графік, Averaged Orlan's Signature) та її зашумленої версії (Рисунок 1, червоний графік, Noisy Orlan's Signature).

1. Кореляція Пірсона – це міра лінійного зв'язку між двома числовими величинами. Вона була запропонована британським статистиком Карлом Пірсоном у 1896 році. Це один із найстаріших інструментів у статистиці, який досі широко використовується в багатьох галузях науки: біології, соціології, економіці, машинному навчанні тощо. Кореляція Пірсона показує, наскільки сильно і в якому напрямку зміни однієї змінної пов'язані зі змінами іншої. Значення коефіцієнта Пірсона лежить у межах від -1 до 1 та обчислюється згідно формули (1).

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}, \quad (1)$$

де: x_i, y_i – значення у відповідних векторах, $\bar{x}$ та $\bar{y}$ середні значення (математичні сподівання) відповідних векторів.

Обчисливши значення кореляційного коефіцієнта Пірсона отримано результат, що становить 0,808, або ж – 80,8%.

2. Кореляція Спірмена – це рангова кореляція, яка вимірює монотонний (не обов'язково лінійний) зв'язок між двома змінними. Вона оцінює, наскільки добре одна змінна може бути описана монотонною функцією іншої. Коефіцієнт рангової кореляції був запропонований Чарльзом Спірменом (Charles Spearman) у 1904 році, британським психологом і статистиком. Він розробив цю метрику для аналізу зв'язку між оцінками людей у психологічних тестах, де точні значення не мали великого значення, а от порядок (ранг) – так. Ідея така: замість самих значень використовуються ранги (позиції в упорядкованому списку), після чого розраховується звичайна кореляція Пірсона між цими рангами. Значення кореляції Спірмена, як і у Пірсона, лежить у межах від -1 до 1 та обчислюється згідно формули (2).

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n\left(n^2 - 1\right)}, \quad (2)$$

де d_i – різниця між рангами $R(x_i)$ та $R(y_i)$, n – кількість пар значень.

Обчисливши значення кореляційного коефіцієнта Спірмена отримано результат, що становить 0,806, або ж – 80,6%.

3. Коефіцієнт Кендалла – це ще один непараметричний метод вимірювання зв'язку між двома змінними, подібний до Спірмена, але з дещо іншим підходом. Коефіцієнт був запропонований американським статистиком Морісом Кендаллом у 1938 році. Він прагнув створити більш «логічну» міру кореляції для порядкових (рангових) даних, яка прямо спирається на кіль-

кість узгоджених/неузгоджених пар. Він ґрунтується на порівнянні парних значень: наскільки часто пара значень має узгоджений або неузгоджений порядок у двох вибірках. Значення коефіцієнта Кендалла, як і у Пірсона та Спірмена, лежить у межах від -1 до 1 та обчислюється згідно формули (3).

$$\tau = \frac{C - D}{\frac{1}{2}n(n-1)}, \quad (3)$$

де C – кількість узгоджених пар, D – кількість неузгоджених пар та n – загальна кількість пар.

Обчисливши значення коефіцієнта Кендалла отримано результат, що становить 0,608, або ж – 60,8%.

4. Косинусна подібність – це міра, яка показує, наскільки два вектори спрямовані в одному напрямку. Її спочатку широко застосовували у текстовій аналітиці (порівняння документів, рекомендаційні системи). У радіоелектроніці та спектральному аналізі вона дозволяє порівнювати форму сигнатур, ігноруючи амплітуду – це особливо важливо, якщо сигнал ослаблений або спотворений. Застосовується переважно у випадках, коли важлива форма або напрямок вектора, а не його довжина чи абсолютні значення. Значення косинусної міри подібності лежить у межах від 0 до 1 та обчислюється згідно формули (4).

$$\cos(\theta) = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\| \cdot \|\vec{y}\|} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2} \cdot \sqrt{\sum_{i=1}^{n} y_i^2}}, \quad (4)$$

де $\vec{x} \cdot \vec{y}$ – скалярний добуток векторів, $\|\vec{x}\|$ – довжина (норма) вектора $\vec{x}$, $\|\vec{y}\|$ – довжина (норма) вектора $\vec{y}$.

Обчисливши значення косинусної міри подібності отримано результат, що становить 0,996, або ж – 99,6%.

5. Евклідова відстань – це одна з найпростіших і найвідоміших метрик відстані. Вона вимірює пряму відстань (лінію) між двома точками у багатовимірному просторі, подібно до того, як рулетка вимірює відстань між двома точками в реальному світі. На відміну від попередніх мір подібності, Евклідова відстань після її обрахування надає результат, що лежить в межах від нуля до нескінченності ([0, ∞)). Таким чином, про максимальну подібність порівнюваних векторів говоритиме значення результату, що максимально приближене до нуля. Обчислюється Евклідова Відстань згідно формули (5).

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}, \quad (5)$$

де (x, y) – два вектор (наприклад, сигнатури), x_i *та* y_i значення елементів векторів на позиції *i*, *n* – довжина векторів.

Обчисливши значення Евклідової відстані отримано результат, що становить 255,2.

6. Манхеттенська відстань, або також **L1-норма** – це метрика, що вимірює відстань між двома точками шляхом підсумовування абсолютних різниць їхніх координат. Вона не підносить різниці до квадрату, тому більші відхилення мають менший вплив. Аналогічно з Евклідовою відстанню Манхеттенська відстань надає результат, що лежить в межах від нуля до нескінченності ([0, ∞)). Обчислюється Манхеттенська Відстань згідно формули (6).

$$d(x,y) = \sum_{i=1}^{n} |x_i - y_i|, \quad (6)$$

де (x, y) – пара векторів (наприклад, сигнатур), x_i *та* y_i значення елементів векторів на позиції *i*, *n* – довжина векторів.

Обчисливши значення Манхеттенської відстані отримано результат, що становить 5483.

7. Відстань Чебишова – це міра відстані, яка визначає найбільшу абсолютну різницю між відповідними координатами двох векторів. Її ще називають L∞-норма або максимальна норма. Цей підхід моделює ситуацію, коли переміщення обмежується лише найповільнішим або найгіршим компонентом, і тому використовується там, де важливе найбільше можливе відхилення. Відстань Чебишова обчислюється згідно формули (7).

$$d(x,y) = \max_i |x_i - y_i|, \quad (7)$$

де (x, y) – пара векторів, x_i *та* y_i – відповідні елементи векторів.

Обчисливши значення відстані Чебишова отримано результат, що становить 57,7.

Висновки. У статті розглянуто метод метричного аналізу сигнатур безпілотних літальних апаратів (БПЛА) типу «Орлан» із метою підвищення ефективності їх виявлення в умовах реального, зашумленого радіочастотного середовища. Було проаналізовано низку класичних метрик подібності – кореляцію Пірсона, Спірмена, Кендалла,

косинусну подібність та евклідову відстань – на предмет їхньої здатності виявляти сигнатури навіть за наявності спотворень та шуму.

Результати експериментів продемонстрували, що найбільш стійкими до завад є метрики, які враховують загальну форму та напрям векторів (зокрема косинусна подібність), тоді як метрики рангового типу (Спірмен, Кендалл) також показали добру адаптивність до нелінійних змін. Евклідова відстань, хоча і є менш інформативною при високому рівні шуму, може бути корисною для доповнення загальної оцінки подібності.

Практична цінність запропонованого підходу полягає у його здатності адаптуватися до непередбачуваних умов роботи системи РЕР, що особливо актуально в умовах гібридної війни. Отримані результати можуть бути використані для побудови алгоритмів виявлення БПЛА, які поєднують спектральний аналіз із інтелектуальними методами класифікації на основі багатометричної оцінки. Перспективи подальших досліджень полягають у розширенні бази еталонних сигнатур, залученні методів машинного навчання та впровадженні виявлених підходів у реальні системи ситуаційного моніторингу.

Список літератури:

1. Ansari M.I., Hasan T. SpectNet: End-to-End Audio Signal Classification Using Learnable Spectrograms. arXiv preprint arXiv:2211.09352. 2022. URL: https://arxiv.org/abs/2211.09352 (дата звернення: 28.05.2025).
2. Harp G.R., et al. Machine Vision and Deep Learning for Classification of Radio SETI Signals. arXiv preprint arXiv:1902.02426. 2019. URL: https://arxiv.org/abs/1902.02426 (дата звернення: 28.05.2025).
3. Tekbıyık K., et al. Spectrum Sensing and Signal Identification with Deep Learning based on Spectral Correlation Function. arXiv preprint arXiv:2003.08359. 2020. URL: https://arxiv.org/abs/2003.08359 (дата звернення: 28.05.2025).
4. Kuang T., et al. Abnormal Signal Recognition with Time-Frequency Spectrogram: A Deep Learning Approach. arXiv preprint arXiv: 2205.15001. 2022. URL: https://arxiv.org/abs/2205.15001 (дата звернення: 28.05.2025).
5. Zhou Y., et al. Cancelable biometric schemes for Euclidean metric and Cosine metric. Cybersecurity. 2023. Т. 6, № 1. С. 1–16. URL: https://cybersecurity.springeropen.com/articles/10.1186/s42400-023-00137-0 (дата звернення: 28.05.2025).

Demianchuk T.M., Gubar S.G. ANALYSIS OF THE USE OF SIMILARITY MEASURES FOR COMPARING UAV SIGNAL SIGNATURE

This article explores the application of metric-based quantitative methods for analyzing signal similarity in the context of detecting and identifying electromagnetic emissions from unmanned aerial vehicles (UAVs) based on their signatures obtained through electronic intelligence (ELINT). The research aims to demonstrate the potential of using formalized mathematical approaches–particularly similarity metrics–to enhance the accuracy of signal classification from various sources, especially under conditions of noise, fragmentation, or other distortions commonly encountered in real-world combat environments. The study is centered around the analysis of UAV signal signatures, focusing specifically on the "Orlan" reconnaissance drones, which are widely used by the adversary.

Detecting UAV signal signatures poses significant challenges due to the presence of both natural and artificially generated interference, the dynamic nature of the radio frequency environment, and variability in signal parameters caused by transmission conditions, redirection, and processing. Traditional recognition approaches, such as threshold-based power detection or manual waveform analysis, often fail to deliver reliable results–particularly in automated or semi-automated monitoring systems. In this context, the application of quantitative similarity assessment methods becomes highly relevant, allowing for the comparison of recorded signals against pre-established reference signatures. This enables the implementation of effective algorithms capable of identifying the type of UAV-or other signal source–even in the presence of high noise levels or when signal information is partially missing.

Within this study, a methodology has been developed that includes several key steps: extraction of a complex signal from a digital WAV-format recording; construction of its spectral representation; selection of a spectrogram fragment (signature); subsequent averaging of the data to reduce the influence of fluctuations; addition of artificial noise with controlled parameters; and comparison of the resulting noisy signature with the original (reference) version. The noise is modeled as a complex additive component with random amplitude characteristics and the potential inclusion of impulse spikes to simulate adverse real-world electromagnetic environments.

One of the approaches considered for signal similarity analysis is the Pearson correlation coefficient, which evaluates the linear relationship between two data vectors–the reference and the noisy signature. This method offers a numerical interpretation of similarity and can be used as an indicator of signal recognizability. However, due to its limitations–namely, insensitivity to nonlinear dependencies and sensitivity to scaling–it is considered within a broader set of metric-based approaches. The study also examines alternative similarity measures, such as cosine similarity, rank-based coefficients (Spearman and Kendall), Euclidean, Manhattan, and Chebyshev distances. This multifaceted approach enables a comprehensive evaluation of signal signature similarity from different perspectives and provides a universal platform for metric-based analysis of signal data under noise and partial fragmentation conditions.

The computational framework was implemented in Python, which–through its libraries (NumPy, SciPy, Matplotlib, etc.)–provides high-performance data processing, numerical computation capabilities, and tools for plotting and spectrogram visualization. A key component of the research is visual analysis, which assists in experimentally validating the degree of similarity between signals and identifying atypical or anomalous cases where certain metrics may produce misleading results. The plotted signal graphs allow for an intuitive interpretation of numerical results, which is especially important when integrating the proposed approach into real-world analysis systems.

The results of the study demonstrate that even under significant noise levels (via the addition of complex noise), correlation coefficients can retain relatively high values, indicating the preservation of the signal's core structure. This supports the conclusion that the proposed approach has potential effectiveness for preliminary analysis and signal filtering in ELINT systems.

The practical value of the findings lies in their potential use for the development of automated or semi-automated UAV detection systems integrated into mobile or stationary radio monitoring complexes. These systems can operate in real time, comparing incoming signals with a database of known signatures, enabling timely responses to threats. Additionally, the approach can be scaled to work with a wide variety of signal types–not limited to "Orlan" UAVs–making it a universal tool for analyzing the electromagnetic environment in complex operational scenarios.

In summary, this review presents a signal analysis approach for electronic intelligence based on the application of similarity metrics. The proposed methodology is effective, flexible, and promising for further development. It can be integrated into analysis systems employing artificial intelligence or machine learning elements, where similarity metrics serve as input features or criteria for subsequent classification. Future directions include testing the algorithms on large empirical datasets, which would further enhance the accuracy and sensitivity of UAV detection systems in real-world conditions.

Key words: *similarity measures, UAV detection, spectral analysis, SDR, signal classification, radio frequency monitoring.*

УДК 004.932
DOI https://doi.org/10.32782/2663-5941/2025.3.2/24

Заварін В.О.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Терейковський І.А.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

АНАЛІЗ СУЧАСНИХ РІШЕНЬ У КОМП'ЮТЕРНІЙ ДІАГНОСТИЦІ РОЗЛАДІВ ЗДОРОВ'Я ЗА РАЙДУЖКОЮ ОКА

У статті представлено поглиблений та всебічний аналіз сучасного стану комп'ютеризованих систем і методів, призначених для діагностики різноманітних розладів здоров'я людини шляхом аналізу цифрових зображень райдужної оболонки ока. Іридодіагностика, як галузь альтернативної медицини, що історично ґрунтується на візуальній інтерпретації складних візерунків, варіацій кольору, текстурних особливостей та інших морфологічних характеристик райдужки, останніми десятиліттями отримала значний імпульс для свого розвитку, наукової об'єктивізації та потенційного клінічного застосування завдяки стрімкому прогресу в суміжних галузях комп'ютерного зору, цифрової обробки медичних зображень та, особливо, штучного інтелекту, зокрема методів машинного та глибокого навчання. Актуальність даного дослідження визначається невпинно зростаючим суспільним та медичним попитом на розробку та впровадження нових неінвазивних, доступних, економічно ефективних та, водночас, інформативних методів ранньої діагностики широкого спектра захворювань, а також для моніторингу загального стану здоров'я та оцінки ефективності терапевтичних втручань. Перехід від суб'єктивних оцінок традиційної іридології до кількісних, алгоритмізованих підходів комп'ютерної діагностики відкриває перспективи для підвищення відтворюваності, надійності та доказовості цього методу. Метою даного аналітичного огляду є проведення комплексної систематизації наявних наукових знань, детальне дослідження методологічних засад та критична оцінка ефективності існуючих підходів до комп'ютерної іридодіагностики. Важливим завданням є виявлення їхніх ключових переваг, таких як неінвазивність, відносна простота та швидкість отримання даних, а також іманентних недоліків, зокрема проблем, пов'язаних із варіабельністю якості зображень, складністю виділення стабільних діагностичних ознак, питаннями стандартизації протоколів дослідження та інтерпретації результатів, особливо отриманих за допомогою складних моделей «чорної скриньки». Аналіз спрямований на ідентифікацію технологічних обмежень сучасних систем та визначення найбільш значущих тенденцій і перспективних напрямків подальшого науково-технологічного розвитку цієї галузі, що потенційно може призвести до створення валідованих та клінічно корисних діагностичних інструментів.

У статті послідовно та детально розглядаються фундаментальні етапи, що складають типовий технологічний конвеєр сучасної комп'ютерної іридодіагностичної системи. Перший, підготовчий, етап – попередня обробка зображень – включає низку критично важливих процедур: точну автоматичну локалізацію області інтересу, тобто самої райдужної оболонки, та сегментацію її внутрішньої (зіничної) та зовнішньої (лімбальної) меж. Обговорюються різні алгоритми сегментації, від класичних, заснованих на перетвореннях Хафа або інтегро-диференціальному операторі Даугмана, до сучасних підходів на основі глибоких згорткових нейронних мереж архітектури U-Net. Далі відбувається нормалізація сегментованої кільцеподібної області райдужки, зазвичай шляхом її перетворення в полярні координати (модель «гумового листа»), що забезпечує інваріантність до змін розміру зіниці та форми райдужки, а також до невеликих зміщень та поворотів ока. Важливою складовою цього етапу є також розробка та застосування ефективних методів для виявлення та подальшої компенсації впливу різноманітних артефактів, таких як відблиски від джерел світла, оклюзії (перекриття) ділянок райдужки віями та повіками, що можуть суттєво спотворювати діагностичну інформацію. Другий ключовий етап – виділення діагностично значущих ознак з підготовленого зображення райдужки. У статті детально аналізуються методи екстракції різнотипних ознак: текстурних, що характеризують складну волокнисту структуру строми райдужки (наприклад, на основі фільтрів Габора, вейвлет-перетворення, статистичних методів, таких як матриці суміжності рівнів сірого

(GLCM), або дескрипторів типу локальних бінарних патернів (LBP)); топологічних, спрямованих на детекцію та кількісну оцінку специфічних іридологічних знаків (таких як крипти, лакуни, контракційні борозни, кільця адаптації, стан коларетки); та кольорових, що оцінюють загальний колір райдужки, наявність пігментних плям та ділянок диспігментації (шляхом статистичного аналізу в різних колірних просторах, гістограмного аналізу). Третій, вирішальний, етап – застосування різноманітних алгоритмів машинного навчання для побудови класифікаційних та прогностичних моделей стану здоров'я на основі виділених ознак. Розглядаються як класичні підходи (опорні векторні машини – SVM, дерева рішень, логістична регресія), так і більш сучасні та потужні методи, зокрема нейронні мережі (особливо згорткові нейронні мережі – CNN, що здатні автоматично вивчати ієрархічні ознаки безпосередньо з зображень) та ансамблеві методи (такі як випадковий ліс, градієнтний бустинг), що часто демонструють покращену точність та стійкість.

В огляді проаналізовано значну кількість актуальних наукових публікацій, які демонструють практичне застосування та потенційну ефективність комп'ютерної іридодіагностики для виявлення ознак широкого спектра захворювань та патологічних станів, включаючи ішемічну хворобу серця, хронічну хворобу нирок, цукровий діабет, різноманітні захворювання шлунково-кишкового тракту, обструктивні захворювання легень та навіть деякі психічні розлади, такі як шизофренія. Багато з цих досліджень звітують про досягнення високих показників діагностичної точності, чутливості та специфічності, що свідчить про наявність інформативних патернів у райдужці, корельованих зі станом здоров'я. Наводяться конкретні приклади досліджень із зазначенням використаних методів та отриманих результатів, що дозволяє оцінити поточний рівень розвитку технологій у цій сфері.

Незважаючи на очевидні успіхи та значний науковий і практичний потенціал, у статті детально висвітлюються та критично аналізуються актуальні проблеми та виклики, що стоять на шляху широкого та надійного клінічного впровадження комп'ютерної іридодіагностики. До найважливіших з них належать: нагальна необхідність розробки та впровадження єдиних стандартизованих протоколів отримання зображень райдужки та формування великих, репрезентативних, якісно анотованих наборів даних для навчання та валідації моделей; завдання підвищення робастності розроблюваних методів до неминучих варіацій якості зображень та умов їх зйомки в реальних клінічних умовах; фундаментальна проблема забезпечення належного рівня інтерпретованості та прозорості діагностичних висновків, особливо для складних моделей глибокого навчання, які часто функціонують як «чорні скриньки»; та нагальна потреба у проведенні масштабних, добре спланованих, мультицентричних клінічних валідаційних досліджень для об'єктивного підтвердження реальної діагностичної цінності та безпеки пропонованих методів. Обговорюються також питання, пов'язані з відсутністю загальноприйнятих біомаркерів в іридології та скептицизмом з боку частини медичної спільноти. Стаття узагальнює поточний стан наукових досліджень у цій динамічно розвиваючій галузі та визначає найбільш перспективні напрямки для її подальшого розвитку як ефективного, доступного та об'єктивного інструменту неінвазивної медичної діагностики.

Ключові слова: *іридодіагностика, комп'ютерна діагностика, аналіз зображень райдужки, машинне навчання, глибоке навчання, обробка зображень, сегментація райдужки, виділення ознак, класифікація методів, аналіз літератури, стан здоров'я, неінвазивна діагностика.*

Постановка проблеми. Іридодіагностика, як напрям комплементарної медицини, базується на припущенні, що дослідження візерунків, кольорів та інших характеристик райдужної оболонки ока може надати інформацію про стан здоров'я людини та її органів [15]. Традиційні методи іридодіагностики, що покладаються на візуальний огляд лікарем-іридологом, неминуче пов'язані з високим ступенем суб'єктивізму та залежністю від досвіду фахівця, що обмежує їхню надійність та наукову доказовість [15].

З розвитком комп'ютерних технологій виник новий напрямок – комп'ютеризована іридодіагностика, що має на меті автоматизувати аналіз, підвищити об'єктивність та точність діагностичних висновків. Незважаючи на значний інтерес та певні успіхи, існує потреба в систематизації накопиченого досвіду, аналізі ефективності різних підходів та визначенні невирішених проблем. Даний аналіз спрямований на всебічне дослідження існуючих комп'ютерних рішень для виявлення розладів здоров'я за райдужкою ока, їхню класифікацію та критичну оцінку поточного стану галузі.

Аналіз останніх досліджень і публікацій. Наукова література останніх років свідчить про активні дослідження у сфері комп'ютеризованої іридодіагностики. У дослідженні Ф. Озбілгіна (F. Özbilgin) та ін. (2023) було запропоновано метод прогнозування ішемічної хвороби серця (ІХС) за допомогою SVM та ознак з вейвлет-перетворення і GLCM, досягши 93% точ-

ності [1]. С. Музаміл (S. Muzamil) та ін. (2020) розробили інтелектуальну систему на основі CNN для ідентифікації хронічної хвороби нирок (ХХН) з точністю 96,8% [2]. Для діагностики цукрового діабету II типу К. Сруті (K. Sruthi) та ін. (2022) застосували AlexNet для аналізу інфрачервоних зображень райдужки, досягши 95,85% точності [3]. Е. В. П. Каррера (E. V. P. Carrera) та П. А. М. Майя (P. A. M. Maya) (2018) запропонували систему для діагностики захворювань шлунково-кишкового тракту (ШКТ) з використанням обробки зображень та SVM, досягши 96% точності [4]. А. Бансал (A. Bansal) та ін. (2018) розробили інструмент на основі вейвлет-перетворення та SVM для попередньої діагностики обструктивних захворювань легень (ОЗЛ) з точністю 89,0% [5]. Дослідження Л. Ван (L. Wang) та ін. (2022) виявило кореляцію між характеристиками райдужки (крипти, пігментні плями) та шизофренією за допомогою логістичної регресії [6]. Н. Шарма (N. Sharma) та ін. (2017) запропонували метод для раннього виявлення серцевих захворювань, використовуючи тензорний градієнт, фільтри Габора та SVM [7]. Г. Гіта (G. Geetha) та ін. (2024) представили фреймворк на основі ШІ, що інтегрує федеративне навчання, CNN та генетичні алгоритми для прогнозування діабету та серцево-судинного ризику [8]. М. Хаммад (M. Hammad) та ін. (2021) розробили систему IKDIS з використанням CNN для діагностики хвороб нирок, досягши 86,9% точності [12]. С. Кумар (S. Kumar) та ін. (2021) запропонували метод ранньої діагностики хронічних захворювань печінки з використанням інфрачервоних зображень райдужки та ансамблевого навчання з точністю 98% [9]. Ці та інші дослідження [10, 11, 13, 14] демонструють потенціал комп'ютерної іридодіагностики, але відмінності в методологіях та даних ускладнюють пряме порівняння їхньої ефективності.

Формулювання цілей статті

Постановка завдання. Метою статті є аналіз існуючих комп'ютеризованих методів і систем для виявлення розладів здоров'я за райдужкою ока.

Завдання:

1. Систематизувати основні підходи та технології.

2. Проаналізувати ключові етапи обробки даних.

3. Класифікувати методи за типом алгоритмів обробки зображень та машинного навчання.

4. Визначити досягнення, переваги та недоліки рішень.

5. Окреслити невирішені проблеми та перспективи досліджень.

Виклад основного матеріалу. Комп'ютеризована іридодіагностика, як багатоетапний процес, включає низку послідовних кроків, кожен з яких використовує специфічні методи та алгоритми, що суттєво впливають на кінцевий результат [13].

• Етапи попередньої обробки зображень райдужки, Якість попередньої обробки є фундаментальною.

◦ **Локалізація та сегментація райдужки і зіниці:** Першочерговим завданням є точне визначення меж райдужки: внутрішньої (зіничної) та зовнішньої (лімбальної).

▪ *Класичні методи:* Інтегро-диференціальний оператор Дж. Г. Даугмана (J. G. Daugman) [10] шукає кругові межі, ефективний для якісних зображень. Перетворення Хафа виявляє параметричні криві, стійке до розривів контуру. Моделі активних контурів (Snakes) ітераційно еволюціонують початковий контур, адаптуючись до довільних форм, але чутливі до ініціалізації.

▪ *Методи на основі глибокого навчання:* Архітектури CNN, такі як U-Net О. Роннебергера (O. Ronneberger), П. Фішера (P. Fischer) та Т. Брокса (T. Brox) [14] та її модифікації, демонструють високу точність семантичної сегментації навіть за наявності шумів, автоматично вивчаючи ієрархічні ознаки. Недоліком є потреба у великих анотованих наборах даних.

◦ **Нормалізація зображень:** Після сегментації кільцеподібна область райдужки нормалізується – перетворюється у прямокутне зображення фіксованого розміру для забезпечення інваріантності ознак до варіацій масштабу та форми. Найпоширенішим є модель «гумового листа» Даугмана (Daugman) [10], що «розгортає» райдужку в полярні координати. Точність залежить від якості сегментації.

◦ **Виявлення та видалення артефактів:** Зображення райдужки часто містять артефакти.

▪ *Відблиски світла:* Детектуються за високою інтенсивністю пікселів.

▪ *Оклюзії віями та повіками:* Виявляються методами детекції країв (наприклад, Кенні) та перетворенням Хафа для повік; для вій використовуються аналіз текстури та фільтри Габора.

▪ *Обробка артефактів:* Маскування (ігнорування пікселів артефактів) або відновлення текстури (inpainting) для заповнення пошкоджених областей [13].

• Виділення діагностично значущих ознак Після попередньої обробки з нормалізованого зображення виділяються діагностично значущі ознаки, що кількісно характеризують патерни, структури та кольорові особливості райдужки [13].

◦ **Текстурні ознаки:** Текстура райдужки, утворена мережею колагенових волокон, судин та пігментних клітин, є інформативною.

▪ *Фільтри Габора:* Аналіз текстури в просторово-частотній області, чутливі до орієнтації та частоти текстурних елементів [7].

▪ *Вейвлет-перетворення:* Розкладання зображення на компоненти з різною просторовою та частотною роздільною здатністю, аналіз текстурних особливостей на різних масштабах [1, 5].

▪ *Матриця суміжності рівнів сірого (GLCM):* Статистичний метод, що описує просторові взаємозв'язки між парами пікселів. Розраховуються ознаки: енергія, контраст, кореляція, гомогенність, ентропія [1].

▪ *Локальні бінарні патерни (LBP):* Ефективний дескриптор локальної мікротекстури, інваріантний до монотонних змін освітлення [13].

▪ *Фрактальний аналіз:* Кількісна оцінка складності та самоподібності текстури [13].

◦ **Топологічні ознаки:** Автоматизація виявлення та кількісної оцінки специфічних структурних «знаків».

▪ *Крипти:* Аналізуються розмір, форма, глибина, кількість, розташування [6].

▪ *Борозни (Контракційні кільця):* Аналізуються кількість, глибина, регулярність.

▪ *Коларетка (Автономне нервове кільце):* Аналізуються форма, чіткість, цілісність, ширина, положення.

▪ *Інші ознаки:* Вузлики Вольфліна, радіуси соляріс, нервові кільця.

◦ **Кольорові ознаки:** Аналіз змін кольору, пігментних плям, ділянок гіпо-/гіперпігментації.

▪ *Аналіз у різних колірних просторах:* RGB, HSV, HSL, Lab, YCbCr для відокремлення тону, насиченості, яскравості [13].

▪ *Колірні гістограми:* Одно- або багатовимірні гістограми розподілу кольорів для кількісного опису [13].

▪ *Кількісна оцінка пігментації:* Автоматичне виявлення та оцінка параметрів пігментних плям.

Класифікація існуючих методів у комп'ютеризованій іридодіагностиці Існуючі методи можна класифікувати за типом використовуваних алгоритмів машинного навчання та методів обробки зображень.

◦ **Алгоритми машинного навчання:**

▪ *Опорні векторні машини (SVM):* Часто демонструють високу точність [1, 4, 5]. Ефективні з нелінійно роздільними даними завдяки «трюку з ядром».

▪ *Нейронні мережі (НМ) та глибоке навчання (ГН):* CNN автоматично вивчають ієрархічні ознаки. Використовуються для діагностики хвороб нирок [2, 12], цукрового діабету [3]. MLP також застосовуються для класифікації виділених ознак.

▪ *Інші алгоритми машинного навчання:* Дерева рішень, випадковий ліс (використовувався для діагностики діабету П. Самантом (P. Samant) та Р. Агарвалом (R. Agarwall) [11]), метод k-найближчих сусідів, методи кластеризації та ансамблеві методи (бустинг, стекинг), які комбінують прогнози кількох моделей (наприклад, для діагностики хвороб печінки [9]).

◦ **Методи обробки зображень (як основа для виділення ознак):**

▪ *Вилучення ознак:* Вейвлет-перетворення [1, 5], GLCM [1], SURF, фільтри Габора [7], LBP.

Сегментація та нормалізація: Перетворення Хафа, модель Даугмана [10], U-Net [14].

▪ *Покращення зображення:* Методи типу CLAHE (Contrast Limited Adaptive Histogram Equalization) для покращення локального контрасту [13].

◦ **Гібридні підходи:** Поєднання різних алгоритмів машинного навчання або інтеграція машинного навчання з традиційними методами обробки зображень для досягнення кращих результатів [7].

▪ Проблеми якості даних та їхній вплив Якість, обсяг та репрезентативність даних є фундаментальними факторами.

◦ **Вплив умов отримання зображень:** Освітлення, фокусування, роздільна здатність, артефакти (відблиски, оклюзії) суттєво впливають на стабільність ознак і точність діагностичних моделей [13]. Нестандартизовані умови ускладнюють навчання робастних моделей.

◦ **Проблеми формування та стандартизації наборів даних:**

▪ *Розмір вибірки:* Часто відносно невеликий, що може призвести до перенавчання моделей.

▪ *Анотування та «Ground Truth»:* Анотування зображень райдужки для діагностичних цілей є складним та суб'єктивним. Визначення об'єктивного «ground truth» ускладнене через недостатню наукову валідацію багатьох традиційних іридологічних знаків [15].

▪ *Стандартизація анотації:* Відсутність єдиних протоколів призводить до варіабельності оцінок.

▪ *Зміщення (Bias) у даних:* Набори даних можуть містити систематичні зміщення (відбору, обладнання, анотувальника). Моделі, навчені на таких даних, можуть успадковувати ці зміщення. Незбалансованість класів також є поширеною проблемою.

Висновки. Проведений аналіз свідчить про значний прогрес та потенціал комп'ютерної іридодіагностики. Застосування сучасних методів обробки зображень та машинного навчання дозволило досягти високої точності у виявленні ознак різноманітних захворювань.

Однак, для широкого клінічного застосування необхідно вирішити низку проблем:

• **Фундаментальне наукове обґрунтування:** Багато аспектів традиційної іридодіагностики потребують підтвердження [15]. Комп'ютеризовані методи також вимагають масштабних клінічних випробувань.

• **Стандартизація та якість даних:** Відсутність стандартизованих протоколів отримання та анотування зображень, а також репрезентативних наборів даних, є серйозною перешкодою.

• **Інтерпретованість моделей:** Складні моделі глибокого навчання часто функціонують як «чорні скриньки», що ускладнює їх розуміння та довіру з боку медиків. Розвиток методів пояснювального ШІ (XAI) є важливим.

• **Подолання скептицизму медичної спільноти:** Необхідно продемонструвати не лише точність, але й зрозумілість та реальну користь систем.

Перспективними напрямками подальших досліджень є розробка робастних алгоритмів, вдосконалення методів виділення ознак, застосування новітніх архітектур глибокого навчання, активне використання XAI, дослідження генетичних кореляцій та проведення масштабних клінічних валідацій. Вирішення цих завдань дозволить реалізувати потенціал комп'ютеризованої іридодіагностики як цінного інструменту сучасної медицини.

Список літератури:

1. Özbilgin F., Kurnaz Ç., Aydın E. Prediction of Coronary Artery Disease Using Machine Learning Techniques with Iris Analysis. *Diagnostics*. 2023. Vol. 13, No. 6. P. 1081. DOI: 10.3390/diagnostics13061081.

2. Muzamil S., Hussain T., Haider A., Waraich U., Ashiq U., Ayguadé E. An Intelligent Iris Based Chronic Kidney Identification System. *Symmetry*. 2020. Vol. 12, No. 12. P. 2066. DOI: 10.3390/sym12122066.

3. Sruthi K., Vijayakumar J., Thavamani S. Deep Learning-Based Verification of Iridology in Diagnosing Type II Diabetes Mellitus. *International Journal of Pattern Recognition and Artificial Intelligence*. 2022. Vol. 36, No. 11. Art. No. 2252017. DOI: 10.1142/S0218001422520176.

4. Carrera E.V.P., Mamarandi P.A.M. Computer Aided Diagnosis of Gastrointestinal Diseases Based on Iridology. *Communications in Computer and Information Science*. Vol. 920. Springer, 2018. P. 269–278. DOI: 10.1007/978-3-030-03747-4_23.

5. Bansal A., Agrawal R., Sharma S. Iris images based pre-diagnostic tool to predict obstructive lung diseases. *Journal of Medical Diagnostic Methods*. 2018. Vol. 7, No. 4. Art. No. 280. DOI: 10.4172/2168-9784.1000280.

6. Wang L. et al. Study on the Correlation Between Iris Characteristics and Schizophrenia. *Neuropsychiatric Disease and Treatment*. 2022. Vol. 18. P. 815–825. DOI: 10.2147/NDT.S356193.

7. Sharma N., Singh S., Singh R. Iris features-based heart disease diagnosis by computer vision. *International Journal of Bio-Inspired Computation*. 2017. Vol. 10, No. 4. P. 245–254. DOI: 10.1504/IJBIC.2017.087321.

8. Geetha G. et al. Iridology as a Predictive Tool for Human Health: Innovating Non-Invasive Diagnostics. *AI and IoT-Based Intelligent Systems for Healthcare*. CRC Press, 2024. P. 169–188. DOI: 10.1201/9781003425601-8.

9. Kumar S., Singh S., Kumar V. Infrared Sensing Based Non-Invasive Initial Diagnosis of Chronic Liver Disease Using Ensemble Learning. *IEEE Access*. 2021. Vol. 9. P. 94999–95012. DOI: 10.1109/ACCESS.2021.3093011.

10. Daugman J.G. High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1993. Vol. 15, No. 11. P. 1148–1161. DOI: 10.1109/34.244676.

11. Samant P., Agarwal R. Diabetic retinopathy detection using machine learning. *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI): proceedings of the conference*. IEEE, 2018. P. 2204–2208.

12. Hammad M. et al. Deep learning based chronic kidney disease detection through iris. *PeerJ Computer Science*. 2021. Vol. 7. Art. No. e707. DOI: 10.7717/peerj-cs.707.

13. Gonzalez R.C., Woods R.E. *Digital Image Processing*. 4th ed. New York: Pearson, 2018. 1024 p.

14. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer, 2015. P. 234–241. (Lecture Notes in Computer Science; vol. 9351). DOI: 10.1007/978-3-319-24574-4_28.

15. Ernst E. Iridology: not useful and potentially harmful. *Archives of Ophthalmology*. 2000. Vol. 118, No. 1. P. 120–121. DOI: 10.1001/archopht.118.1.120.

Zavarin V.O., Tereikovskyi I.A. ANALYSIS OF MODERN SOLUTIONS IN COMPUTERIZED DIAGNOSTICS OF HEALTH DISORDERS BY THE IRIS OF THE EYE

The article presents an in-depth and comprehensive analysis of the current state of computerized systems and methods designed for diagnosing various human health disorders by analyzing digital images of the iris. Iridology, as a field of alternative medicine historically based on the visual interpretation of complex patterns, color variations, textural features, and other morphological characteristics of the iris, has received significant impetus for its development, scientific objectification, and potential clinical application in recent decades due to rapid progress in related fields of computer vision, digital processing of medical images, and, especially, artificial intelligence, including machine and deep learning methods. The relevance of this study is determined by the constantly growing societal and medical demand for the development and implementation of new non-invasive, accessible, cost-effective, and, at the same time, informative methods for the early diagnosis of a wide range of diseases, as well as for monitoring general health status and evaluating the effectiveness of therapeutic interventions. The transition from subjective assessments of traditional iridology to quantitative, algorithmized approaches of computer diagnostics opens up prospects for increasing the reproducibility, reliability, and evidence base of this method. The purpose of this analytical review is to conduct a comprehensive systematization of existing scientific knowledge, a detailed study of methodological foundations, and a critical assessment of the effectiveness of existing approaches to computerized iridology. An important task is to identify their key advantages, such as non-invasiveness and the potential speed of obtaining results, as well as inherent disadvantages, including problems related to the variability of image quality, the complexity of extracting stable diagnostic features, issues of standardization of research protocols, and interpretation of results, especially those obtained using complex "black box" models. The analysis aims to identify the technological limitations of modern systems and determine the most significant trends and promising directions for further scientific and technological development in this field, which could potentially lead to the creation of validated and clinically useful diagnostic tools.

The article sequentially and in detail examines the fundamental stages that make up a typical technological pipeline of a modern computerized iridology system. The first, preparatory, stage – image preprocessing – includes a number of critically important procedures: accurate automatic localization of the region of interest, i.e., the iris itself, and segmentation of its inner (pupillary) and outer (limbal) boundaries. Various segmentation algorithms are discussed, from classical ones based on Hough transforms or Daugman's integro-differential operator, to modern approaches based on deep convolutional neural networks of the U-Net architecture. Next, the segmented annular region of the iris is normalized, usually by converting it to polar coordinates (the "rubber sheet" model), which ensures invariance to changes in pupil size and iris shape, as well as to small displacements and rotations of the eye. An important component of this stage is also the development and application of effective methods for detecting and subsequently compensating for the influence of various artifacts, such as reflections from light sources, occlusions (overlapping) of iris areas by eyelashes and eyelids, which can significantly distort diagnostic information. The second key stage is the extraction of diagnostically significant features from the prepared iris image. The article analyzes in detail the methods for extracting various types of features: textural, characterizing the complex fibrous structure of the iris stroma (e.g., based on Gabor filters, wavelet transform, statistical methods such as gray level co-occurrence matrices (GLCM), or descriptors like local binary patterns (LBP)); topological, aimed at detecting and quantifying specific iridological signs (such as crypts, lacunae, contraction furrows, adaptation rings, collarette status); and color, evaluating the overall color of the iris, the presence of pigment spots, and areas of dyspigmentation (through statistical analysis in different color spaces, histogram analysis). The third, decisive, stage is the application of various machine learning algorithms to build classification and prognostic models of health status based on the extracted features. Both classical approaches (support vector machines – SVM, decision trees, logistic regression) and more modern and powerful methods are considered, including neural networks (especially convolutional neural networks – CNN, capable of automatically learning hierarchical features directly from images) and ensemble methods (such as random forest, gradient boosting), which often demonstrate improved accuracy and stability.

The review analyzes a significant number of current scientific publications that demonstrate the practical application and potential effectiveness of computerized iridology for detecting signs of a wide range of diseases

and pathological conditions, including coronary artery disease, chronic kidney disease, type II diabetes mellitus, various gastrointestinal diseases, obstructive lung diseases, and even some mental disorders such as schizophrenia. Many of these studies report achieving high rates of diagnostic accuracy, sensitivity, and specificity, indicating the presence of informative patterns in the iris correlated with health status. Specific examples of research are provided, indicating the methods used and the results obtained, which allows for an assessment of the current level of technological development in this area.

Despite the evident successes and significant scientific and practical potential, the article highlights in detail and critically analyzes current problems and challenges hindering the widespread and reliable clinical implementation of computerized iridology. The most important of these include: the urgent need to develop and implement unified standardized protocols for acquiring iris images and forming large, representative, high-quality annotated datasets for training and validating models; the task of increasing the robustness of developed methods to inevitable variations in image quality and acquisition conditions in real clinical settings; the fundamental problem of ensuring an adequate level of interpretability and transparency of diagnostic conclusions, especially for complex deep learning models that often function as "black boxes"; and the pressing need for large-scale, well-designed, multicenter clinical validation studies to objectively confirm the real diagnostic value and safety of the proposed methods. Issues related to the lack of universally accepted biomarkers in iridology and skepticism from part of the medical community are also discussed. The article summarizes the current state of scientific research in this dynamically developing field and identifies the most promising directions for its further development as an effective, accessible, and objective tool for non-invasive medical diagnostics.

Key words: *iridology, computerized diagnostics, iris image analysis, machine learning, deep learning, image processing, iris segmentation, feature extraction, methods classification, literature analysis, health status, non-invasive diagnostics.*

УДК 644.1+681.5
DOI https://doi.org/10.32782/2663-5941/2025.3.2/25

Здоренко С.С.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Черепанська І.Ю.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

СИСТЕМА ЕНЕРГОЕФЕКТИВНОГО ПРОГНОЗНОГО КЕРУВАННЯ ТЕМПЕРАТУРОЮ ПОВІТРЯ В БУДІВЛЯХ З АВТОНОМНИМ ОПАЛЕННЯМ

В статті на основі аналізу проблеми енергоефективного керування температурою повітря в різних будівлях, що виникла на тлі стрімкого зростання цін на усі види енергоресурсів, які споживаються при їх опаленні, представлено структурну схему новоствореної системи енергоефективного керування температурою у будівлях з автономним опаленням (СЕКТ БАО) на основі адаптивного MPC-алгоритму та штучних нейронних мереж (ШНМ). Новостворена система розглядається як єдине ціле з одночасним виокремленням функціонально взаємодіючих підсистем та модулів, що вирішують окремі локальні завдання та у сукупності реалізовують глобальну мету енергоефективного керування температурою повітря. Ця мета полягає у мінімізації енергоспоживання при одночасній максимізації комфортних температурних умов для людей, що перебувають у будівлі. Особливістю новоствореної СЕКТ БАО є математична модель, яка враховує усю множину збурюючих факторів, що можуть виникати та впливати на об'єкт керування (ОК). Крім того забезпечується можливість автоматичного коригування складових математичної моделі ОК в режимі реального часу без необхідності зупинки системи опалення. Для розрахунків параметрів математичної моделі та визначення їх поточних значень застосовано White-box та Data-driven методи відповідно. Збір поточних даних про стан ОК, збурюючи фактори, клерувальні впливи та ін. здійснюється на основі технології IoT. Ці дані зберігаються в хмарному середовищі і після аналізу ШНМ в подальшому використовуються для коригування параметрів математичної моделі в режимі реального часу. Передбачається, що застосування новоствореної СЕКТ БАО дозволить своєчасно реагувати на вплив множини різних за своєю фізичною природою та впливом чинників, що призводять до змін температури в будівлях, підвищити точність та швидкодію регулювання температури, забезпечити значне зниження енергоспоживання та одночасне забезпечення температурного комфорту людей.

Ключові слова: *Model Predictive Control, MPC, прогнозна модель, штучні нейронні мережі, автоматичне регулювання температури повітря, енергоефективне керування, проектування системи автоматизації.*

Постановка проблеми. Сучасний етап розвитку енергетичних галузей, зокрема паливної промисловості, різних країн світу, у тому числі й України, визначається стрімким зростанням вартості усіх видів енергоресурсів, які споживаються при опаленні житлових і комерційних будівель, у тому числі, з автономними системами опалення. Так, за даними статистичної організації Європейської Комісії (Євростат), у період з 2019 по 2024 роки тариф на газ для домогосподарств країн Євросоюзу зріс на 53%, тариф на газ для інших споживачів – на 78%. Тариф на електроенергію для домогосподарств та інших споживачів зріс на 36% та 37% відповідно [1]. Причому очевидною є тенденція до систематичного подальшого зростання вартості енергоресурсів. Саме тому гостро стоїть проблема енергоефективного керування температурою повітря в будівлях з автономним опаленням для забезпечення комфортних температурних умов при одночасному зниженні енергоспоживання. Вказане може бути досяжним за рахунок розробки нових та вдосконалення існуючих методів, засобів і систем, а також новітніх алгоритмів та математичних моделей енергоефективного керування температурою повітря в будівлях з автономним опаленням, що може суттєво сприяти зниженню енергоспоживання. Рішення щодо енергоефек-

тивного керування температурою повітря в будівлях з автономним опаленням можуть бути отримані при розв'язанні низки задач пов'язаних із подальшим розвитком енергоефективного або так званого оптимального керування системами автономного опалення (САО), що функціонують у несприятливих економічних умовах сучасного енергетичного ринку та нестаціонарних виробничих умовах. В цьому контексті розробка новітньої системи енергоефективного прогнозного керування температурою у будівлях з автономним опаленням на основі прогнозних моделей (Model Predictive Control (MPC)), здатної працювати із стохастичною та зашумленою інформацією є актуальною науково-технічною задачею. Розв'язання цієї задачі забезпечить у значній мірі вирішення поставленої проблеми.

Аналіз останніх досліджень і публікацій. Проведений аналіз відомих досліджень та публікацій показує, що в останні роки замість традиційних методів, засобів, систем і алгоритмів керування температурою повітря спостерігається інтерес до так званого інтелектуального оптимального керування на основі прогнозних моделей та елементів штучного інтелекту.

Так у роботі [2, с. 38] автори прийшли до висновку, що МРС-регулятори демонструють значно кращу динамічну відповідь та стабільність у інерційних системах, до яких належать і системи автономного опалення. Це забезпечує зменшення перерегулювання та часу перехідного процесу до 28% порівняно з традиційними ПІ та ПІД регуляторами. Крім того, застосування МРС-регуляторів дозволяє отримати додаткові цінні якості системи керування, як то підвищену точність, робастність до зовнішніх збурень та здатність адаптуватися до змінних умов навколишнього середовища.

Проте в даній роботі автори зосередились переважно на порівнянні традиційних регуляторів з МРС-регуляторами і не розглядався вплив конкретних збурюючи факторів на систему керування.

У роботі [3, с. 1] описано результати використання МРС-алгоритму для керування температурою повітря в житловій будівлі з автономним опаленням на основі теплового насосу та накопичувача теплової енергії. Зокрема МРС-алгоритм використовувався для реалізації нелінійної глобальної оптимізації при керуванні системою автономного опалення. В межах цієї оптимізації модель в просторі стану житлової будівлі відображає динаміку системи автономного опалення в просторі стану при оптимальному керуванні режимами роботи теплового насосу зі збільшенням так званого коефіцієнта сезонної продуктивності. Результати роботи демонструють значне скорочення енергоспоживання, зокрема, електроенергії, на 11,6%.

Проте в даній роботі лишаються невирішеними питання, що пов'язані з врахуванням впливів на температуру повітря в приміщенні кількості мешканців та тривалість їх перебування, інтенсивності ультрафіолетового випромінювання, наявності інших додаткових джерел тепла, як то побутових приладів та / або комп'ютерної техніки, коливання температури зовнішнього середовища та зміни погодних умов. Це вказує на обмеженість прогнозної моделі та недостатню ефективність автоматичного керування з огляду на зменшення енергоспоживання.

В роботі [4, с. 12] наводяться результати використання МРС алгоритму при керуванні температурою повітря в експериментальній кімнаті. Джерелом тепла в цій кімнаті виступає гідравлічна підлогова опалювальна поверхня. Отримані результати демонструють скорочення енергоспоживання до 40% за опалювальний сезон з одночасним забезпеченням комфортних температурних умов мешканців. При чому найбільше скорочення енергоспоживання спостерігалося в міжсезоння. В даній роботі враховується прогнозний вплив на температуру повітря в приміщенні наявності людей, коливання температури зовнішнього середовища та інерційність опалювальної поверхні будівлі. Однак в роботі [4] залишилися неврахованими впливи на температуру повітря в приміщенні інтенсивності ультрафіолетового випромінювання та інших джерел тепла, наприклад, побутових приладів. Також не досліджена реакція системи керування на зміну температури теплоносія. Це призводить до формування неповної прогнозної моделі та обмежень у зменшенні енергоспоживання.

У роботі [5, с. 14] наводяться результати розробки МРС-регулятора для керування температурою повітря в будівлі з повітряно-водяним тепловим насосом. В якості джерела тепла використано низькотемпературну підлогову опалювальну поверхню. Отримані результати продемонстрували зменшення витрат енергоресурсів на 14,9% у порівнянні з використанням традиційного ПІД-регулятора. Застосування МРС-алгоритму окрім зменшення енерговитрат забезпечило ще й достатньо високу точність підтримки температури повітря. Так, задане

значення температури +22 °С підтримувалось МРС-алгоритмом на сталому незмінному рівні. При цьому застосування традиційного ПІД-регулятора дозволяло підтримувати задане значення температури лише в інтервалі від +21,7 °С до +22,5 °С. Вказане безперечно підкреслює перспективність оптимального керування на основі МРС-алгоритму.

Однак в даній роботі залишилися невирішеними питання, що пов'язані з врахуванням впливу на температуру в приміщенні кількості мешканців та тривалість їх перебування там, а також наявності інших додаткових джерел тепла, як то побутових приладів та / або комп'ютерної техніки.

Вказані питання частково вирішуються в роботі [7]. Зокрема, в роботі [7, с. 19] використано адаптивний МРС-алгоритм для керування комбінованою системою автономного опалення, яка реалізована на основі поверхневого підлогового опалення та повітряного обігріву. Математична модель будівлі в роботі враховує як термічну динаміку будівлі, так і вплив на температуру повітря в приміщенні кількості людей та тривалість їх перебування там. Отримані результати показали, що впровадження адаптивного МРС-алгоритму дозволяє зменшити енергоспоживання на 22,2%. Крім того, в даній роботі було порівняно роботу звичайного та адаптивного МРС-алгоритму. Отримані результати показали, що застосування адаптивного МРС-алгоритму дозволяє отримати на 11,5% кращі результати щодо зменшення енергоспоживання. Це пояснюється тим, що робота адаптивного МРС-алгоритму заснована на застосуванні більш повної (або точної) математичної моделі будівлі, яка враховує термічну динаміку будівлі, зміни в кількості людей, що перебувають в будівлі та тривалість їхнього перебування там. Це безперечно вказує, що адаптивні МРС-алгоритми є кращою альтернативою звичайним МРС-алгоритмам. Проте, у даній роботі не враховується контекстний вплив на точність роботи системи керування похибок та зашумленості вхідної інформації. Очевидно, вказане вимагає доопрацювання представленої в даній роботі системи керування шляхом застосування сучасних методів і засобів збирання і обробки первинної інформації, наприклад, на основі елементів штучного інтелекту.

В роботі [7, с. 13] наведено стратегію керування температурою повітря на основі МРС-алгоритму для зменшення енергоспоживання гідравлічними поверхневими системами автономного опалення в офісних будівлях. Ця стратегія передбачає використання data-driven підходу для побудови математичної моделі, яку використовує МРС-регулятор в системі автоматичного керування. Data-driven підхід дозволяє будувати математичну модель використовуючи так звані історичні дані, які накопичуються протягом тривалого часу роботи системи автономного опалення. Такі історичні дані враховують зміну погодних умов в певні періоди часу, рівень наповненості будівлі людьми та тривалість їх перебування там, наявність внутрішніх додаткових джерел тепла, наприклад, побутових приладів та / або комп'ютерної техніки. Отримані результати показали зменшення витрат енергії на 16% у порівнянні з традиційними алгоритмами регулювання. Проте можливість практичного застосування data-driven підходу обмежується довготривалістю та трудомісткістю збору і аналізу множини історичних даних для побудови повної математичної моделі, що використовується в МРС-регуляторі. Це виключає можливість роботи системи керування в штатному режимі одразу після її імплементації та є її суттєвим недоліком, адже вимагає тривалого часу на збір та обробку даних, що є необхідними для побудови математичної моделі будівлі і подальшого використання її для роботи МРС-алгоритму.

В роботі [8, с. 1] наведено результати роботи МРС-регулятора для автоматичного керування температурою в кімнаті житлової будівлі з джерелом тепла у вигляді поверхневого підлогового опалення. Наведена математична модель враховує зміни температури повітря зовнішнього середовища та присутність мешканців. Крім того МРС-алгоритм оптимізує роботу системи з урахуванням змін вартості електроенергії в різні години доби. Отримані результати продемонстрували щоденне зниження витрат на електричної енергії до 19%. Проте у даній роботі була використана математична модель з обмеженою кількістю можливих збурень, як то зміни температури повітря зовнішнього середовища та присутність мешканців, що зменшує можливості для енергоефективного керування та зниження енергоспоживання.

Таким чином, представлені у відомих роботах результати досліджень МРС-алгоритмів вказують на їх інноваційну перспективність в якості основи оптимального керування температурою повітря у будівлях з автономним опаленням, адже демонструють значне скорочення енер-

госпоживання. Проте, існуючі дослідження не пропонують комплексних рішень та прогнозних моделей, які б враховували усю множину чинників, що прямо або опосередковано впливають на температуру повітря у будівлях. Очевидно обґрунтованою є необхідність досліджень у напрямку вирішення проблеми енергоефективного керування температурою повітря в будівлях з автономним опаленням для забезпечення комфортних температурних умов при одночасному зниженні енергоспоживання.

Постановка завдання. Метою статті є розробка структурної схеми системи енергоефективного керування температурою у будівлях з автономним опаленням на основі адаптивного MPC-алгоритму, елементів штучного інтелекту та з урахуванням впливу всіх можливих зовнішніх та внутрішніх збурень. Це дасть можливість своєчасно реагувати на вплив множини різних за своєю фізичною природою та впливом чинників, що призводять до змін температури в будівлях, підвищити точність та швидкодію регулювання температури, забезпечити значне зниження енергоспоживання та економію енергетичних ресурсів при одночасному температурному комфорті людей.

Виклад основного матеріалу. Енергоефективне керування температурою повітря в будівлях з автономним опаленням є багатокритеріальним оптимізаційним процесом. У класичній постановці багатокритеріальна оптимізація являє собою процес пошуку деякого оптимального або компромісного, у даному розумінні рішення, при одночасній оптимізації декількох цільових функцій, які є взаємно суперечливими або взаємно конфліктними. У даному випадку для забезпечення енергоефективного керування температурою повітря в будівлях з автономним опаленням необхідно мінімізувати енергоспоживання $K_{ec} \rightarrow \min$ системою автономного опалення та одночасно забезпечити максимальні комфортні температурні умови $K_t \rightarrow \max$ для людей, що перебувають у будівлі.

Очевидно, що енергоефективне керування температурою повітря в будівлях − це досить складний процес. Зменшення енергоспоживання не повинно призводити до погіршення температурних умов у будівлі, адже незадовільний рівень температури у приміщенні та її перепади негативно відбивається на здоров'ї людей посилюючи ризик простудних захворювань, порушуючи кровообіг та роботу опорно-рухового апарату тощо. Очевидно, що енергоефективне керування температурою повітря в будівлях з автономним опаленням вимагає розробки відповідної автоматизованої системи керування, наприклад, так званої системи енергоефективного прогнозного керування температурою у будівлях з автономним опаленням (СЕКТ БАО).

Концепція побудови СЕКТ БАО ґрунтується на системному та процесному підходах, за яких новостворена система розглядається як єдине ціле з одночасним виокремленням функціонально взаємодіючих підсистем та модулів, що вирішують окремі локальні завдання та у сукупності реалізовують глобальну мету функціонування СЕКТ БАО. Із врахуванням вказаного СЕКТ БАО містить певну сукупність складових. Вони використовуються для розв'язання певних локальних задач із застосуванням певних новітніх моделей, методів та засобів автоматизованої обробки інформації та прийняття рішення в режимі реального часу з високою точністю та швидкодією. Зокрема, в математичній моделі об'єкта керування (ОК) пропонується врахувати усю множину збурюючих факторів, що можуть виникати та впливати на нього. Крім того необхідно забезпечити можливість автоматичного коригування математичної моделі ОК в режимі реального часу без необхідності зупинки системи опалення. Під ОК, в даному випадку, розуміється будівля з автономним опаленням. Спрощена структурна схема пропонованої СЕКТ БАО наведена на рис. 1.

Основними складовими пропонованої СЕКТ БАО є:

1) база даних (БД), що використовується для накопичення та зберігання інформації необхідної для вирішення локальних завдань щодо розв'язання оптимізаційної задачі енергоефективного керування температурою повітря в будівлях з автономним опаленням. Дана задача, у своїй загальній постановці, описується виразами (1) та (2). Крім того у БД накопичуються і зберігаються значення множини параметрів математичної моделі ОК, яка у загальному випадку описується виразом (3). Зокрема накопичується і зберігається інформація про:

– температуру повітря, підлоги та стін у будівлі;

– зовнішню температуру повітря;

– присутність мешканців в будівлі та їх кількість;

– інтенсивність ультрафіолетового випромінювання;

– температуру теплоносія від джерела тепла;

– роботу внутрішніх джерел тепла;

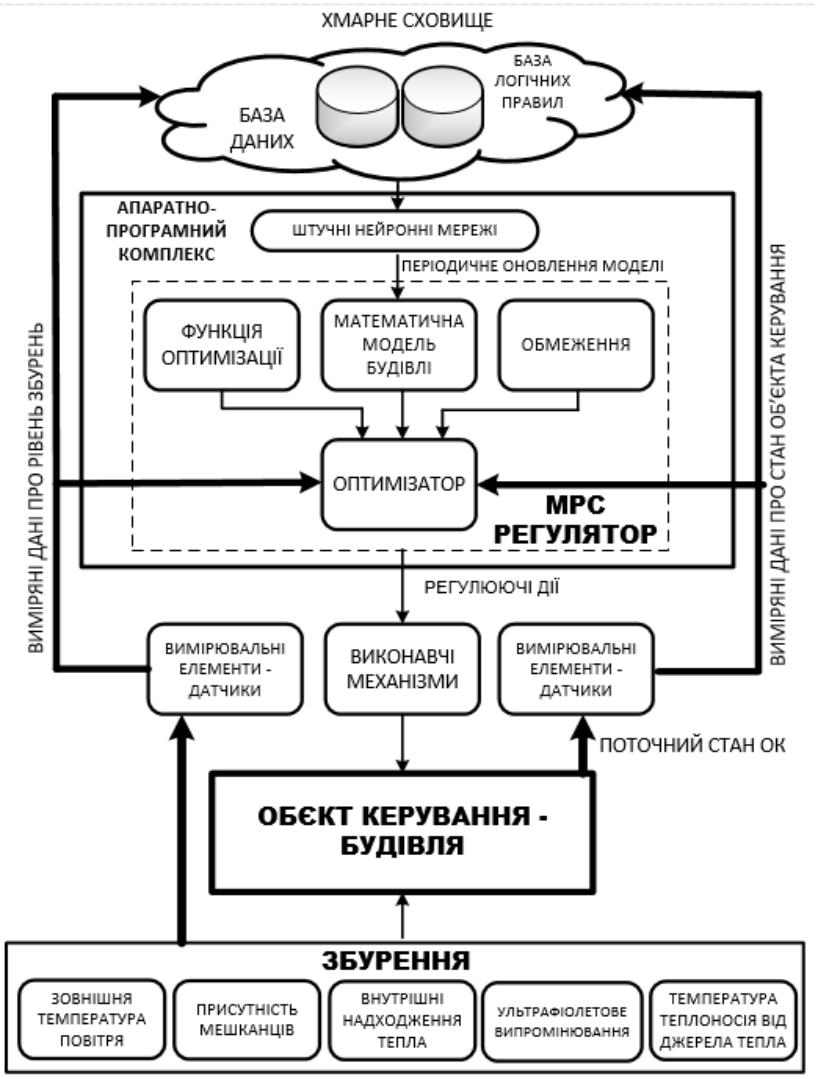


Рис. 1. Спрощена структурна схема пропонованої СЕКТ БАО

2) база логічних правил (БЛП), яка зберігає множину фактів та правил. Останні використовуються для керування фактами при визначенні необхідності коригування поточних параметрів математичної моделі ОК. Окрім декларативної частини БЛП утвореної логічними правилами і фактами вона містить процедурну частину. Процедурна частина БЛП утворена множиною функцій і процедур, що реалізують оптимізаційні алгоритми визначення необхідності коригування поточних параметрів математичної моделі ОК (див. вираз (3));

3) апаратно-програмний комплекс, що містить MPC-регулятор та штучну нейронну мережу (ШНМ). MPC-регулятор здійснює керування температурою повітря в будівлі з автономним опаленням на основі прогнозних моделей (рис. 2). У MPC-регуляторі для вирішення вказаного завдання застосовується оптимізатор, що враховує функцію оптимізації відповідно до виразу (1), задані обмеження оптимізації відповідно до виразу (2) та математичну модель ОК відповідно до виразу (3). ШНМ в апаратно-програмному комплексі використовується для автоматичного аналізу в режимі реального часу множини даних, що безперервно надходять від датчиків і фактів, які містяться у БД та визначення необхідності коригування математичної моделі ОК у даний момент часу. На основі отриманого рішення від ШНМ у БЛП здійснюється автоматизоване маніпулювання інформацією з метою виявлення такого логічного правила, яке відповідає рішенню, що є оптимальним для даного моменту часу;

4) модуль автоматизованої ідентифікації збурюючих впливів, який реалізований множиною датчиків. Ці датчики призначені для безперервного вимірювання температури повітря зовнішнього середовища, інтенсивності ультрафіолетового випромінювання, температури теплоносія

в системі автономного опалення, визначення наявності мешканців в будівлі;

5) модуль автоматизованої ідентифікації параметрів стану ОК, який реалізований множиною датчиків, що призначені для вимірювання температури повітря, температури підлоги та температури стін в будівлі;

6) модуль виконавчих механізмів, які здійснюють безпосередній вплив на ОК шляхом припинення подачі теплоносія на будівлю чи окремі її приміщення, що в свою чергу спричиняє зниження температури повітря.

Основою функціонування СЕКТ БАО є МРС-регулятор. МРС-регулятор функціонує у складі СЕКТ БАО та реалізує відомий МРС-алгоритм керування, що наведений в роботі [9, с. 4]. Так відповідно до відомого алгоритму МРС-регулятор, використовуючи математичну модель ОК, вирішує оптимізаційну задачу енергоефективного керування температурою повітря, враховуючи такі обмеження, як мінімальне та максимальне значення вихідної величини та максимальну зміну керуючого впливу від одного кроку алгоритму до іншого за рис. 2.

Оптимізаційна задача розв'язується ітераційно шляхом моделювання поведінки ОК на певному проміжку часу (час досягнення певного значення прогнозованої вихідної величини $y(k)$), що в роботі [10, с. 2] називається горизонтом передбачення або горизонтом прогнозування. Цей проміжок налічує N_p кроків k. На кожному кроці k часу, що дорівнює періоду дискретизації T_s. МРС-регулятор в режимі реального часу, на основі отриманих даних від множини датчиків, оцінює поточний стан ОК. На основі цих даних обчислюється послідовність керуючих дій $u(k)$ на часовому горизонті керування, що відповідає вирішенню задачі оптимізації і залежить від поточного стану ОК. Часовий горизонт керування налічує N_c кроків k, причому він завжди менший за горизонт прогнозування. Далі МРС-регулятор застосовує лише першу визначену керуючу дію, не враховуючи всі наступні. На наступному часовому проміжку процес повторюється.

Таким чином, оптимізація відбувається шляхом багатократного моделювання процесу керування з врахуванням збурень та вихідних керуючих дій з формуванням деякого найкращого рішення на даному етапі керування.

Оптимізаційну функцію J можна представити як зважену квадратичну суму прогнозованих похибок e та змін керуючих впливів Δu, що враховують обидва критерії оптимізації: мінімізацію енерговитрат (енергоспоживання) $K_{ec} \rightarrow \min$ та максимізацію комфортних температурних умов $K_t \rightarrow \max$ (мінімізацію відхилення температури повітря від заданого значення). Причому, з урахуванням суб'єктивності оцінки комфортних температурних умов мешканцями, за комфортні

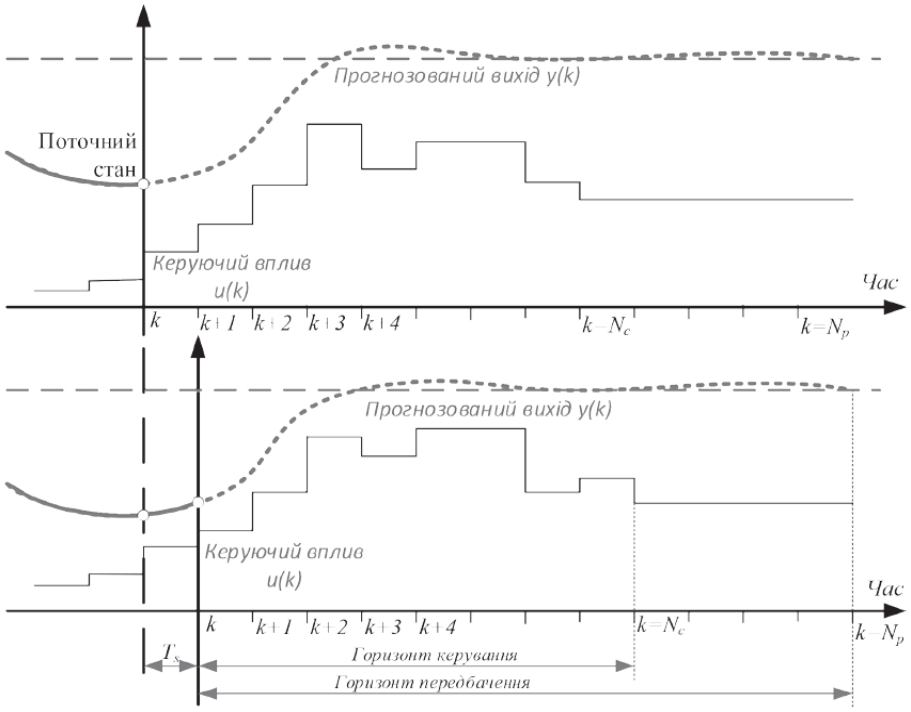


Рис. 2. Схематичне зображення узагальненого принципу роботи МРС-алгоритму

температурні умови можуть бути прийняті значення температури в інтервалі значень від 20 до 22 °С включно при вологості повітря від 40% до 60% включно. Функція оптимізації розраховується за відомою формулою (1), наведеною в роботі [11, с. 3]:

$$\boldsymbol{J} = \sum_{i=1}^{N\text{p}} w_e e_{k+i}^2 + \sum_{i=1}^{N\text{p-1}} w_{\Delta u} \Delta u_{k+i}^2, \qquad (1)$$

де J – функція оптимізації;

w_e – ваговий коефіцієнт прогнозованої похибки;

e – прогнозована похибка між значенням вихідної величини $y(k)$ та цільовим значенням вихідної величини $y_{ціл}(k)$;

$w_{\Delta u}$ – ваговий коефіцієнт зміни керуючих впливів від кроку до кроку;

Δu – зміна керуючих впливів від одного кроку до іншого.

Розв'язання оптимізаційної задачі обмежене горизонтом прогнозування $k \times N_p$. Коефіцієнти w_e та $w_{\Delta u}$ – це вагові коефіцієнти, які виражають важливість кожної цілі оптимізації в загальному процесі оптимізації. При розв'язанні оптимізаційної задачі на кожному кроці k МРС-регулятор намагається зменшити прогнозовану похибку e між цільовим значенням вихідної величини $y_{ціл}(k)$ та прогнозованим $y(k)$ на кожному кроці прогнозування, досягаючи таким чином задачі мінімізації похибки (критерій – максимізація комфорту користувачів), а також мінімізувати зміну керуючих впливів Δu від одного кроку до іншого, забезпечуючи дотримання критерію мінімізації витрати енергії. Також в оптимізаційну задачу включаються певні обмеження, яких має дотримуватись регулятор (наприклад, мінімальна і максимальна температура повітря в будівлі, мінімальна і максимальна потужність джерела енергії, обмеження на величину зміни керуючих впливів тощо. Величина обмежень визначається за наступним виразом:

$$\begin{cases} y_{min}(k) \le y(k) \le y_{max}(k) \\ 0 \le \Delta u \le \Delta u_{max} \end{cases}, \qquad (2)$$

де $y_{min}(k)$ – мінімально можливе значення вихідної величини;

$y(k)$ – фактичне значення вихідної величини;

$y_{max}(k)$ – максимально можливе значення вихідної величини;

Δu – зміна керуючих впливів від одного кроку до іншого;

Δu_{max} – максимально можлива зміна керуючих впливів від одного кроку до іншого.

Отже, на довільному кроці k МРС-регулятор вирішує задачу оптимізації на горизонті прогнозування, враховуючи задані обмеження [11, с. 3].

Ключову роль в роботі МРС-регулятора відіграє оптимізатор, задачею якого є використовуючи наявну математичну модель ОК, враховуючи критерії оптимізації, задані обмеження та поточні і прогнозовані збурюючи дії розрахувати наступний керуючий влив на ОК. Він забезпечує, щоб прогнозоване значення вихідної величини $y(k)$ максимально відповідало заданому $y_{зад}(k)$, а керуючий вплив на наступному кроці був максимально оптимальним з точки зору витрат енергії та забезпечення комфорту користувачів. Прогнозована траєкторія зміни $y(k)$ з найменшим J дає оптимальне рішення, а отже, визначає оптимальну послідовність керуючих впливів, яка максимально наблизить значення вихідної величини $y(k)$ до заданого значення $y_{зад}(k)$ з мінімальними витратами енергії.

Таким чином, за допомогою МРС алгоритму досягаються високі показники енергоефективності.

Ключовою особливістю при енергоефективному керуванні температурою повітря в будівлях з автономним опаленням на основі МРС-алгоритму є максимально точне визначення математичної моделі ОК, яка є основою роботи алгоритму керування. Адже для прогнозування поведінки ОК (стан системи $x(k+1)$) МРС-регулятор використовує заздалегідь визначену динамічну модель ОК в просторі станів, що представлена виразом (3), яка враховує поточний стан ОК $x(k)$, можливу дію на нього різноманітних збурень $d(k)$ (наприклад, зміну температуру повітря ззовні будівлі, вплив інтенсивності сонячного випромінювання, зміну температури теплоносія в системі опалення, присутність людей тощо) та вплив керуючих дій $u(k)$.

$$\begin{cases} x(k+1) = Ax(k) + Bu(k) + Ed(k) \\ y(k) = Cx(k) \end{cases}, \qquad (3)$$

де $x(k+1)$ – вектор змінних стану на наступному кроці керування;

k – часовий крок;

A – перехідна матриця стану;

$x(k)$ – вектор змінних стану;

B – перехідна матриця керування;

$u(k)$ – вектор змінних керування;

E – перехідна матриця збурення;

$d(k)$ – вектор змінних збурення;

$y(k)$ – вектор вихідних змінних;

C – перехідна матриця вихідних змінних.

ОК, що являє собою будівлю з автономним опаленням є багатоканальним нелінійним та нестаціонарним об'єктом. На рис. 3 наведено його параметричну схему.

На рис. 3 суцільними лініями із стрілками показано виходи $x(k)$ – змінні стану ОК, відповідний їм вхід $u(k)$, тобто керуючий вплив, та збурення $d(k)$. Крім того, прямими суцільними та пунктирними лініями на рис. 3 показано канали впливу, які відображають функціональну залежність виходів від відповідних входів ОК.

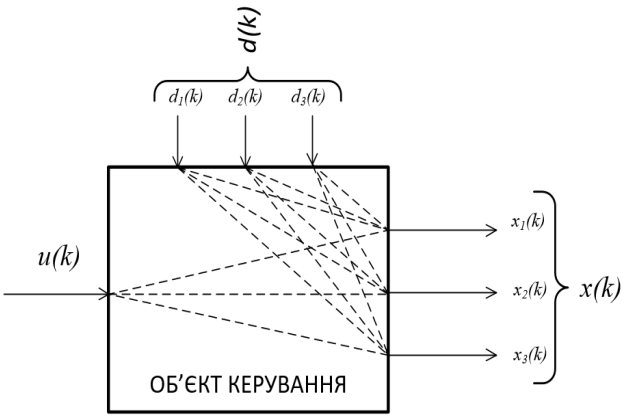


Рис. 3. Параметрична схема об›єкта керування

При подачі на вхід ОК регулюючого впливу $u(k)$ (наприклад, увімкнення нагріву, подача теплоносія від джерела тепла), що формується МРС-регулятором, при одночасній дії множини збурюючих факторів $d(k)$ (наприклад, непередбачувана різка зміна температури повітря на вулиці, вплив сонячної радіації, непередбачувана довготривала присутність людей, непередбачувана різка зміна температури теплоносія від джерела тепла, надходження тепла від внутрішніх джерел тощо) відбувається зміна значень складових вектора змінних стану ОК $x(k)$. Вектор змінних стану $x(k)$ являє собою значення температури повітря в будівлі, температури підлоги та температури стін у момент часу k. Одна зі змінних стану $x1(k)$ – це поточне значення температури повітря в будівлі в момент часу k. Зміна вихідної величини $y(k)$ відбувається за сигналом регулювання від МРС-регулятора, причому відхилення e поточного значення вихідної величини $y(k)$ від заданого $y_{зад}(k)$ має бути якомога меншим $e \to 0$ з використанням якомога меншої кількості енергії $K_{ec} \to \min$.

Для побудови повної математичної моделі ОК передбачається застосування white-box та data-driven методів, які описані в літературі [12] та [7] відповідно. White-box метод пропонується використовувати на підготовчих етапах для розрахунків параметрів математичної моделі конкретної будівлі (теплоізолюючі та теплоакумулюючі властивості стін та підлоги, її інерційність). Data-driven метод пропонується застосовувати до вже працюючої системи керування для визначення поточних параметрів математичної моделі ОК, зокрема при формування перехідних матриць стану A, B, C, E (див. вираз (3)). Важливим етапом data-driven методу є збір в режимі реального часу так званих історичних даних про ОК за допомогою технології IoT (Internet of Things). Зокрема, збираються наступні дані: поточні значення збурюючих факторів, що діють на ОК $d(k)$; значення температури повітря, яка підлягає регулюванню $y(k)$; значення інших параметрів стану ОК $x(k)$, наприклад, температури стін, підлоги тощо.

Ці дані використовуються для визначення поточних параметрів математичної моделі ОК при формуванні перехідних матриць стану A, B, C, E. На наступному етапі вони порівнюються із попередніми значеннями параметрів, так званими історичними значеннями, за допомогою штучної нейронної мережі. У разі необхідності в режимі реального часу здійснюється коригування параметрів математичної моделі за множиною відповідних логічних правил, що зберігаються у відповідній базі знань (БЗ) у хмарному середовищі. В іншому випадку параметри математичної моделі залишаються без змін. Процес коригування параметрів математичної моделі ОК здійснюється безперервно із заданими інтервалами часу.

Таким чином запропонована СЕКТ БАО є новітнім, оригінальним та сучасним засобом для енергоефективного керування температурою повітря в будівлях з автономним опаленням. Вона, на відміну від відомих рішень враховує усю множину чинників, що прямо або опосередковано впливають на температуру повітря у будівлях та забезпечує адаптацію математичної моделі ОК під зміну характеру чи інтенсивності впливу цих чинників. Це дозволить здійснювати енергоефективне керування температурою повітря в будівлях з автономним опаленням для забезпечення комфортних температурних умов при одночасному зниженні енергоспоживання.

Висновки. Розроблено структурну схему СЕКТ БАО, визначено методи та засоби автома-

тизованої обробки даних. Крім того визначено концепцію побудови СЕКТ БАО, яка забезпечує її цілісність з одночасним виокремленням функціонально взаємодіючих підсистем або модулів, що вирішують окремі локальні завдання та у сукупності реалізовують глобальну мету функціонування СЕКТ БАО.

Передбачається, що пропонована СЕКТ БАО дозволить здійснювати енергоефективне керування температурою повітря в будівлях з автономним опаленням для забезпечення комфортних температурних умов, точності регулювання температури повітря при одночасному зниженні енергоспоживання.

Список літератури:

1. Eurostat. Database. European Commission. URL: https://surl.lu/scacat (дата звернення: 07.05.2025).
2. Здольник М., Степанець, О. Аналіз типових та перспективних структур законів автоматичного керування в основних контурах регулювання індивідуального пункту. *Електромеханічні і енергозберігаючі системи*. Кременчук. 2024. Вип. 2 (65). С. 32–40. https://doi.org/10.32782/2072-2052.2024.2.65.4
3. Kuboth, S., Heberle, F., König-Haagen, A., & Brüggemann, D. Economic model predictive control of combined thermal and electric residential building energy systems. *Applied Energy*, 2019. № 240, P. 372–385.
4. Viot, H., Sempey, A., Mora, L., Batsale, J. C., & Malvestio, J. Model predictive control of a thermally activated building system to improve energy management of an experimental building: Part II-Potential of predictive strategy. *Energy and Buildings*, 2018. № 172, P. 385–396.
5. Chen, Q., Li, N., & Feng, W. Model predictive control optimization for rapid response and energy efficiency based on the state-space model of a radiant floor heating system. *Energy and Buildings*, 2021. № 238, 110832.
6. Lv, R., Yuan, Z., Lei, B., Zheng, J., & Luo, X. Model predictive control with adaptive building model for heating using the hybrid air-conditioning system in a railway station. *Energies*, 2021. № 14 (7), 1996.
7. Joe, J., & Karava, P. A model predictive control strategy to optimize the performance of radiant floor heating and cooling systems in office buildings. *Applied Energy*, 2019. № 245, P. 65–77.
8. Hu, M., Xiao, F., Jørgensen, J. B., & Li, R. Price-responsive model predictive control of floor heating systems for demand response using building thermal mass. *Applied Thermal Engineering*, 2019. № 153, P. 316–329.
9. Захарченко А.С., Степанець О.В. Особливості використання модельно-прогнозуючого керування в автоматизації будівель. *Вчені записки ТНУ імені В.І. Вернадського. Серія: технічні науки,* 2020. Том 31 (70) Ч. 1 № 6, С. 70–77.
10. Zhou, B., Chikkala, J., & Schmitt, R. A load-adaptive and predictive control of energy-efficient building automation in production environment. *Procedia CIRP*, 2019. № 79, P. 245–250.
11. Баган Т., Бунь В. Предиктивне керування в системах регулювання мікроклімату. *Вчені записки ТНУ імені В.І. Вернадського. Серія: технічні науки,* 2024. Том 35 (74) № 1, С. 91–96. https://doi.org/10.32782/2663-5941/2024.1.1/14
12. Jorissen, F., Picard, D., Boydens, W., & Helsen, L. White-box Model Predictive Control: Optimal Control and System Integration of Heat Pumps. *HPT Magazine*, 2022. № 40 (2), P. 28–31.

Zdorenko S.S., Cherepanska I.Yu. SYSTEM FOR ENERGY-EFFICIENT PREDICTIVE CONTROL OF INDOOR AIR TEMPERATURE IN BUILDINGS WITH AUTONOMOUS HEATING

The article presents a structural diagram of a newly developed System for Energy-Efficient Temperature Control in Buildings with Autonomous Heating (SEETC BAH), based on an adaptive Model Predictive Control (MPC) algorithm and artificial neural networks (ANNs). This system is proposed in the context of the growing global issue of increasing energy prices, which has significantly affected the energy consumption required for heating residential and commercial buildings.

The proposed SEETC BAH is treated as an integrated whole, with clearly defined and functionally interrelated subsystems and modules. Each of these subsystems is responsible for solving specific local tasks, which together contribute to achieving the global goal of energy-efficient air temperature control. This goal involves minimizing energy consumption while simultaneously maximizing indoor thermal comfort for building occupants.

A distinctive feature of the developed system is its mathematical model, which takes into account the full range of disturbing factors that may arise and influence the controlled object (CO). Moreover, the system is capable of automatically adjusting the parameters of this mathematical model in real time without interrupting the operation of the heating system.

The determination and adjustment of model parameters are implemented using both white-box and data-driven approaches. Real-time data collection regarding the state of the CO, disturbances, control actions, and

other relevant variables is enabled by IoT technology. These data are stored in a cloud environment, where they are analyzed by the ANN and subsequently used for real-time model refinement.

The application of the SEETC BAH is expected to provide timely responses to a wide range of disturbances differing in physical nature and impact, improve the accuracy and speed of temperature regulation, significantly reduce energy consumption, and maintain high levels of thermal comfort for building occupants

Key words: *Model Predictive Control, MPC, predictive model, artificial neural networks, automatic air temperature regulation, energy-efficient control, automation system design.*

УДК 621.391.822
DOI https://doi.org/10.32782/2663-5941/2025.3.2/26

Зилевіч М.О.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Олійник Р.С.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

КОНТРОЛЕР ПОЛЬОТУ З ВЕКТОРНИМ КЕРУВАННЯМ ТЯГОЮ НА БАЗІ STM32

У даній статті розглянуто принцип роботи системи векторного керування тягою для безпілотного літального апарата на базі мікроконтролера STM32. Проведено аналіз існуючих підходів до керування БПЛА та визначено обмеження класичних конструкцій, що працюють виключно за рахунок зміни швидкості обертання двигунів. Обґрунтовано доцільність впровадження механізму повороту двигунів у горизонтальній площині для підвищення маневровості, стабільності та швидкодії системи керування.

Досліджено апаратну реалізацію системи на основі мікроконтролера STM32F407, який завдяки своїй високій продуктивності, великій кількості периферійних інтерфейсів та підтримці роботи в реальному часі є одним із найоптимальніших рішень для задач керування динамічними об'єктами. Розглянуто вибір датчиків (інерціальний модуль ICM-20948), засобів зв'язку (LoRa-приймач ELRS 2.4G), силових драйверів та стабілізаторів живлення. Проведено детальний аналіз механічної частини поворотного блоку, включаючи сервопривід MG995-360, систему передачі зусиль та монтажних елементів, виготовлених методом 3D-друку.

Здійснено порівняння різних серій мікроконтролерів STM32 та підтверджено доцільність вибору серії F4. Визначено, що розроблена система здатна забезпечити точне та надійне керування рухом БПЛА за рахунок комбінованого використання змінного вектору тяги та класичних методів регулювання обертів. Живлення системи реалізовано на основі 4S Li-Po акумулятора з багаторівневим перетворенням напруги.

Завдяки модульній структурі, використанню стандартних компонентів та відкритих протоколів, дана система може бути використана як основа для створення експериментальних зразків БПЛА з розширеними функціональними можливостями у сферах спостереження, досліджень та автоматизованої логістики.

Ключові слова: *контролер польоту, STM32, вектор керування, безколекторний двигун, сервопривід, стабілізація, MEMS, LoRa, друкована плата.*

Постановка проблеми. У сучасному світі безпілотні літальні апарати (БПЛА) відіграють дедалі важливішу роль у багатьох сферах – від оборонної промисловості до цивільного використання. Особливо актуальними стають задачі підвищення маневровості, стабільності та енергоефективності таких систем, зокрема в умовах складного середовища, обмеженого простору, поривчастого вітру або необхідності виконання нестандартних маневрів.

Класична конструкція квадрокоптера передбачає зміну орієнтації апарата шляхом регулювання обертів кожного з безколекторного двигуна. Такий підхід, хоч і є широко поширеним, має низку обмежень. Зокрема, керування за рахунок зміни обертів є інерційним процесом, що знижує швидкість реакції на команди керування.

Актуальною інженерною задачею є реалізація концепції векторного керування тягою, яка передбачає зміну напряму дії сили тяги без необхідності змінювати швидкість обертання двигуна. Такий підхід дозволяє досягти більш високої точності позиціонування, швидкої стабілізації та розширеної маневровості без істотного збільшення обчислювального або енергетичного навантаження [1].

Найбільше практичне значення має реалізація векторного керування на апаратах з невеликою масою та розмірами, які можуть використовуватись у приміщеннях, в обмежених зонах, для рятувальних місій, технічної розвідки, або точ-

ного позиціонування в складних умовах. Однак більшість існуючих рішень з векторним керуванням обмежуються переходом між режимами зльоту/посадки та горизонтального польоту, як у VTOL системах та тейлсітер дронах, або потребують складної механіки, непридатної для компактних дронів.

Особливу складність представляє забезпечення надійної роботи всієї системи в режимі реального часу з урахуванням обробки IMU-даних, формування PWM-сигналів до ESC і сервоприводів, а також взаємодії з пультом керування через бездротовий зв'язок. Для цього потрібен потужний і водночас енергоефективний обчислювальний модуль, зокрема мікроконтролер STM32, здатний обробляти сигнали з високою швидкістю і мінімальними затримками.

Таким чином, розробка компактного та адаптивного контролера польоту з механізмом повороту двигунів, реалізованого на базі STM32, є актуальним інженерним завданням, що відкриває нові можливості для створення високотехнологічних, швидкодіючих та гнучких платформ у сфері безпілотних систем.

Аналіз останніх досліджень і публікацій. Упродовж останнього десятиліття спостерігається стрімкий розвиток безпілотних літальних апаратів (БПЛА), особливо у сфері стабілізації польоту, оптимізації енергоспоживання та реалізації векторного керування тягою. Розвиток сучасної мікроелектроніки, MEMS-сенсорів і засобів бездротового зв'язку дозволяє виводити на новий рівень точність і гнучкість керування дронами в складних умовах.

Одним з важливих напрямів є застосування систем векторного керування тягою (Thrust Vector Control, TVC), які активно досліджуються як у традиційній авіації, так і в малій безпілотній техніці. В дослідженні [2] подано концепцію системи керування швидкістю моторів БПЛА на базі мікроконтролера STM32, що дозволяє реалізувати прецизійне керування з мінімальними затримками у реальному часі.

У праці [3] представлено аналіз апаратної реалізації платформи з поворотними двигунами для квадрокоптера типу VTOL. У роботі показано переваги використання сервоприводів для досягнення стабільності у польоті при маневруванні у складних умовах.

Актуальним також є використання MEMS-інерційних сенсорів у системах стабілізації. Наприклад, у дослідженні [4] розглянуто реалізацію системи стабілізації польоту з використанням акселерометра і гіроскопа, з алгоритмами фільтрації шумів і комплементарним фільтром, що близьке до реалізації стабілізації в проєкті.

Також важливо відзначити дослідження в галузі керування безколекторними двигунами та їх драйверами. У праці [5] проведено огляд методів комутації і реалізації ШІМ-регулювання з урахуванням теплових режимів ESC та ефективності енергоспоживання.

Таким чином, наявні наукові публікації підтверджують актуальність досліджень у напрямах систем стабілізації на базі STM32, інтеграції MEMS-сенсорів, реалізації векторного керування за допомогою поворотних механізмів та енергозберігаючого керування тягою у багатороторних платформах.Результати цих досліджень створюють теоретичну основу для побудови більш гнучких, стабільних та адаптивних БПЛА, що підтримують керування вектором тяги в реальному часі, як це реалізовано у даній роботі.

Постановка завдання. З метою підвищення маневровості, стабільності та ефективності керування безпілотними літальними апаратами, особливо в умовах обмеженого простору та впливу зовнішніх збурень, постає необхідність створення системи керування, яка дозволяє змінювати напрям вектору тяги без зміни швидкості обертання двигунів. Одним з перспективних рішень є розробка контролера польоту з функцією механічного повороту двигунів у горизонтальній площині, що дозволяє змінювати напрям сили тяги в режимі реального часу.

Виклад основного матеріалу. Системи векторного керування тягою (англ. thrust vector control, TVC) дозволяють змінювати напрям прикладання сили тяги без зміни положення всього літального апарата.

Принцип дії таких систем полягає у зміні орієнтації двигунів або сопел, що створюють реактивну силу в будь-який необхідний момент в польоті, а не тільки на зліт/посадку. У випадку багатороторних БПЛА це може бути реалізовано механічним поворотом мотор-групи або окремих двигунів у горизонтальній або вертикальній площині. На відміну від класичної схеми керування, де зміна орієнтації апарата досягається шляхом регулювання обертів кожного з двигунів, векторне керування тягою дозволяє швидко змінити напрямок дії сили без інерційних затримок, характерних для електромеханічної інерції роторів.

Однією з переваг систем векторного керування є підвищена динамічна стабільність, осо-

бливо в умовах, коли традиційне керування виявляється малоефективним – наприклад, під час зависання в обмеженому просторі або при виконанні складних маневрів на низькій швидкості. Зміна вектору тяги дозволяє зменшити навантаження на систему стабілізації, підвищити точність позиціонування та покращити відгук системи керування.

У реалізованому проєкті система векторного керування реалізована шляхом синхронного повороту передньої та задньої пар безколекторних двигунів за допомогою сервоприводів. Поворот здійснюється в межах 90° у горизонтальній площині, що дозволяє змінювати напрям прикладеної тяги та реалізовувати додаткові ступені свободи в керуванні апаратом.

Таким чином, керування положенням і рухом БПЛА здійснюється не тільки шляхом зміни швидкості обертання пропелерів, а й за рахунок зміни напряму дії сили тяги, що значно розширює функціональні можливості системи.

На рисунку 1 зображено структурну схему контролеру польоту, а далі розглянуто створення та аналіз структурної схеми.

Пристрій умовно можна поділити на шість частин.

Двигуни: Безколекторні двигуни відповідають за створення тяги, необхідної для підйому, руху та маневрування квадрокоптера.

Сервоприводи: Крокові двигуни відповідають за зміну кута нахилу пар безколекторних двигунів для більшого маневрування.

Регулятори швидкості: Регулятори швидкості забезпечують точне керування швидкістю обертання кожного безколекторного двигуна, що впливає на потужність і напрямок тяги.

Гіроскоп і акселерометр: Використовуються для вимірювання кутової швидкості та прискорення квадрокоптера, відповідно. Ці дані необхідні для стабілізації польоту.

Контролер: Мікроконтролер STM32 обробляє дані від датчиків, приймає команди з пульта управління і генерує сигнали для керування двигунами та іншими компонентами системи, є саме контролером польоту.

LoRa приймач: Для забезпечення дистанційного керування безпілотним літальним апаратом використовується модуль прийому даних за технологією LoRa. LoRa-приймач здійснює прийом команд з універсального пульта керування БПЛА і передає їх у мікроконтролер STM32 через UART-інтерфейс.

Вибір мікроконтролера. Використання мікроконтролерів сімейства STM32 є обґрунтованим з ряду важливих причин.

Широка лінійка моделей: сімейство STM32 включає велику кількість моделей із різними характеристиками – від ультраекономічних до

Рис. 1. Структурна схема пристрою

високопродуктивних, що дозволяє обрати оптимальне рішення для будь-якого застосування, включаючи проєкти підвищеної складності, такі як безпілотні літальні апарати.

Продуктивність і обчислювальні можливості: мікроконтролери STM32 забезпечують ефективну обробку даних від численних сенсорів та надійне керування виконавчими механізмами. Крім того, STM32 оснащені широким набором вбудованих периферійних модулів: високошвидкісними ADC, PWM-таймерами, інтерфейсами UART, SPI, I2C, CAN, що істотно розширює їх функціональність.

Широкий вибір інтерфейсів: підтримка SPI, I2C, UART, USB, а також можливість роботи із зовнішніми сенсорами, сервоприводами і радіомодулями робить STM32 універсальним рішенням для інтеграції різних компонентів у єдину систему.

Надійність і довговічність: мікроконтролери STM32 мають високу стабільність роботи та великий середній час безвідмовної роботи (понад 20 років за нормальних умов), що є критичним для вбудованих систем, призначених для тривалої автономної експлуатації.

Серед розглянутих серій мікроконтролерів STM32 найоптимальнішим вибором для реалізації системи керування польотом з механізмом повороту двигунів є серія STM32F4. Вона забезпечує оптимальний баланс між продуктивністю (180 МГц), обсягом пам'яті (192 КБ ОЗП), енергоспоживанням та підтримкою периферійних інтерфейсів (PWM, I²C, UART, DMA), необхідних для обробки сенсорних даних і керування двигунами в режимі реального часу. На відміну від серій STM32H7 або STM32L4, які є або надмірно потужними і дорогими, або орієнтованими на енергоефективність, STM32F4 ідеально підходить для задач вбудованого керування динамічними об'єктами з високими вимогами до швидкодії [6].

Цей контролер має відмінний баланс між продуктивністю, ресурсами пам'яті та кількістю периферійних інтерфейсів, необхідних для керування двигунами, сервоприводами та обробки даних від сенсорів [6].

Отже, STM32F407VGT6 є оптимальним вибором для проєкту розробки безпілотного літального апарата з системою змінного вектору тяги, оскільки поєднує в собі потужність, гнучкість, широку підтримку периферії та прийнятний рівень енергоспоживання для задач реального часу.

Вибір IMU модуля. ICM-20948 є 9-осьовим інерціальним датчиком, який інтегрує акселерометр, гіроскоп і магнітометр в одному корпусі [7].

ICM-20948 забезпечує всі необхідні вимірювання для роботи БПЛА:

- Акселерометр: для вимірювання лінійного прискорення у трьох вимірах, що використовується для стабілізації і контролю положення;
- Гіроскоп: для вимірювання кутових швидкостей, які дозволяють підтримувати орієнтацію у просторі;
- Магнітометр: для визначення напряму відносно магнітного поля Землі, що може бути використано як компас для навігації.

Датчик має достатню обчислювальну потужність і ресурси для обробки даних із сенсора в реальному часі і водночас низьке енергоспоживання на рівні 2,5 мВт при напрузі 1.71–3.6 В.

Таблиця 1

Порівняння серій мікроконтролерів STM32

Серія	Ядро	Макс. частота (МГц)	Flash-пам'ять (КБ)	ОЗП (КБ)	DMA	ADC (розрядність)	Призначення
STM32F1	Cortex-M3	72	1024	96	Так	12-bit	Загального призначення
STM32F4	**Cortex-M4**	**180**	**1024**	**192**	**Так**	**12-bit**	**Складні задачі реального часу, DSP**
STM32H7	Cortex-M7	480	2048	1024	Так	16-bit	Високопродуктивні застосування
STM32L4	Cortex-M4	80	1024	320	Так	12-bit	Ультранизьке споживання
STM32G4	Cortex-M4	170	512	128	Так	12-bit	Керування аналоговими системами

Технічні характеристики ICM-20948 [6]:

- Акселерометр: ±2/±4/±8/±16 g;
- Гіроскоп: ±125/±250/±500/±1000/±2000 dps;
- Магнітометр: ±4900 μT.

Блок розподілення живлення. Даний блок виконує ключову роль у системі, забезпечуючи стабільне живлення всіх компонентів пристрою та контроль руху. Він розподіляє напругу на декілька рівнів:

- Вхідна напруга від джерела живлення: від 11В до 22В, для безколекторних моторів та у DC-DC перетворювачі на периферію;
- Рівень 7 В: для компонентів середньої потужності;
- Рівень 3.3 В: для живлення низьковольтної електроніки, включаючи мікроконтролери та датчики.

Вибір датчика струму. Для вимірювання струму дрона обрано датчик INA139NA через його ключові переваги:

- Висока точність вимірювання малих і великих струмів;
- Широкий діапазон робочої напруги (від 2,7 В до 40 В), що дозволяє використовувати його з джерелами живлення для квадрокоптерів;
- Простота інтеграції завдяки низькій кількості зовнішніх компонентів;
- Невисока вартість, що робить його оптимальним для прототипу.

INA139NA забезпечує надійний контроль струму, необхідний для моніторингу та оптимізації енергоспоживання квадрокоптера [8].

Вибір операційного підсилювача. Використовується для побудови схеми зворотного зв'язку, забезпечуючи точність і стабільність регулювання вихідної напруги. Для цього було обрано операційний підсилювач LM358P, за наступні переваги:

- Низьке споживання струму та широкий діапазон робочих напруг (3–40 В), що забезпечує стабільну роботу в межах вхідної напруги перетворювача;
- Підходить для роботи у складі компенсаційної схеми стабілізації;
- Доступність і широке застосування в схемах DC-DC перетворювачів.

Вибір імпульсного перетворювача напруги. Даний компонент служить основним елементом DC-DC перетворення, виконуючи імпульсне зниження вхідної напруги до заданого рівня з високою ефективністю. Для цього було обрано імпульсний перетворювач напруги XL4016E1, за наступні переваги:

- Здатність працювати з вхідною напругою до 40 В та вихідним струмом до 8 А, що забезпечує необхідний запас потужності для живлення крокових двигунів;
- Високий ККД до 96% з мінімальними втратами енергії;
- Інтегровані функції захисту від перевантаження, перегріву та короткого замикання, що підвищує надійність роботи схеми.

Вибір лінійного стабілізатора напруги. Даний компонент забезпечує стабільне живлення для чутливих до напруги компонентів керування та логіки схеми. Для цього було обрано лінійний стабілізатор напруги AS78L05Z, за наступні переваги:

- Забезпечує стабільне живлення елементів керування, наприклад, мікросхем та логіки перетворювача;
- Компактний розмір і простота інтеграції у схему;
- Низька вартість і висока стабільність вихідної напруги.

Вибір прецизійного регулятора напруги. Такий регулятор використовується для точного налаштування напруги в ланцюзі зворотного зв'язку, підтримуючи стабільність і точність вихідної напруги. Для цього було обрано прецизійний регулятор напруги AZ431, за наступні переваги:

- Висока точність (відхилення ≤1%), що критично важливо для стабільної роботи перетворювача;
- Невеликий розмір і висока надійність.

Вибір сервопривода. Для реалізації механізму повороту двигунів у системі векторного керування тягою використано сервопривід MG995-360. Його завдання – забезпечити синхронний поворот двигунів у горизонтальній площині в межах 90°, з достатньою точністю, крутним моментом і надійністю.

MG995-360 є модифікацією класичного сервопривода MG995, яка дозволяє повертати вал на повні 360°. Завдяки цьому він може працювати як у позиційному режимі (з обмеженням кута програмно), так і в безперервному режимі обертання – залежно від прошивки та режиму керування.

Причини вибору MG995-360: високий крутний момент, достатній для обертання блоку з двигуном та пропелером без прослизання чи люфту, металева передача, доступність і низька вартість.

Вибір LoRa приймача. Для реалізації стабільного і далекого зв'язку між оператором

і безпілотним літальним апаратом (БПЛА) обрано LoRa-приймач SpeedyBee ELRS 2.4G Nano. Цей приймач працює за протоколом ExpressLRS, який забезпечує низьку затримку і високу надійність передачі команд на великі відстані.

Вибір обумовлено такими перевагами:

- Компактність і мала маса, що критично важливо для авіаційних систем;
- Підтримка UART-протоколу, що дозволяє легко інтегрувати приймач з мікроконтролером STM32;
- Висока дальність зв›язку (до кількох кілометрів) навіть у складних умовах радіозавад;
- Гнучке конфігурування та сумісність з універсальними пультами керування БПЛА;

Вбудована телеметрія та можливість роботи в асинхронному режимі. Модуль підключається до контролера польоту через UART-інтерфейс і передає керуючі сигнали для управління всіма функціями апарата [9].

Живлення. Реалізовано на основі акумуляторної батареї типу 3S-6S (літій-полімерна, LiPo), яка забезпечує номінальну напругу від 11.1 до 22.2 В. Акумулятори цього типу є стандартом у авіамоделізмі завдяки високій питомій енергоємності, здатності до швидкого віддавання струму (високий C-рейтинг) і стабільній роботі навіть при значних навантаженнях.

Проєктування друкованого вузла. З урахуванням конструктивної складності проєкту – наявності мікроконтролера, сервоприводів, ESC, LoRa-приймача та IMU-сенсора – оптимальним варіантом є двошарова друкована плата. Вона забезпечує достатній простір для трасування, дозволяє виділити окремі зони для живлення, сигналів керування і чутливих ліній сенсорів, при цьому залишаючись доступною за ціною та простою у виробництві.

Для розробки друкованої плати в межах цього проєкту використано сучасне середовище автоматизованого проєктування Altium Designer. Це професійна САПР-система, яка широко застосовується у промисловості, наукових дослідженнях та освітніх проєктах для створення схемотехнічних рішень і трасування друкованих плат.

На рисунках 2 і 3 показано 3D вигляд розробленої друкованої плати для запропонованого контролеру польоту.

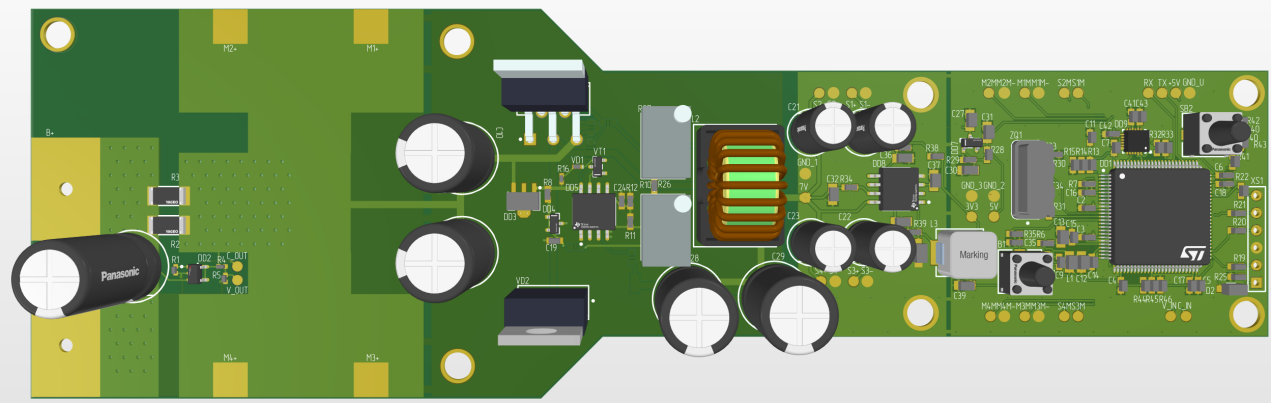


Рис. 2. 3D вигляд ДП зверху

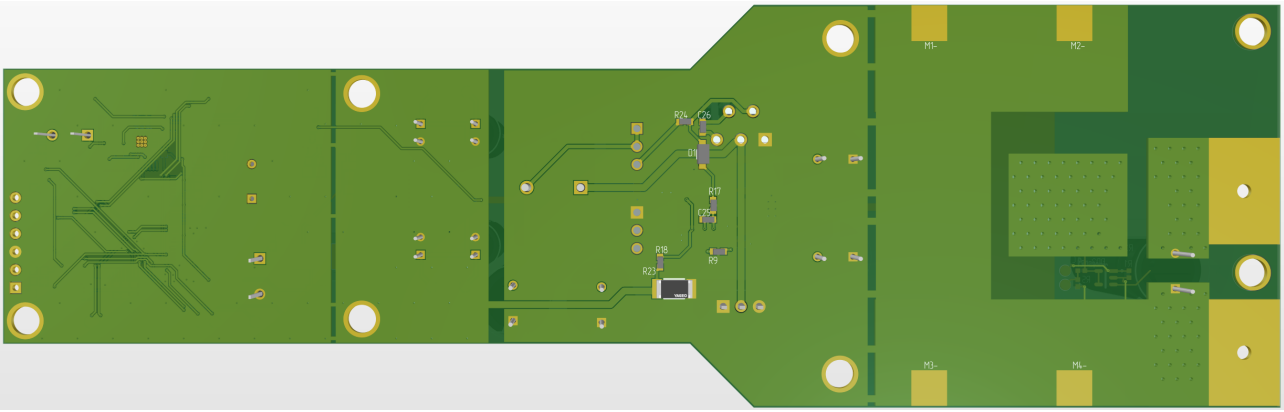


Рис. 3. 3D вигляд ДП знизу

Висновки. У роботі представлено розробку системи векторного керування тягою для безпілотного літального апарата з використанням контролера польоту на базі STM32. Актуальність теми обґрунтована потребою у підвищенні маневровості, точності стабілізації та швидкодії керування в складних умовах експлуатації БПЛА. Застосування механізму повороту двигунів дозволяє реалізувати зміну напрямку вектору тяги без необхідності зміни обертів пропелерів, що знижує інерційність системи та покращує її реактивність.

На основі аналізу літературних джерел, технічної реалізації та експериментального моделювання зроблено такі висновки:

1. Контролер STM32F407VGT6 вибрано як оптимальний варіант завдяки поєднанню високої продуктивності, достатнього обсягу пам'яті та широкого набору інтерфейсів для роботи з сенсорами, сервоприводами та модулями зв'язку. Його використання дозволило забезпечити стабільну обробку даних у реальному часі та формування керуючих сигналів з мінімальною затримкою.

2. Механізм векторного керування тягою, реалізований за допомогою сервоприводів MG995-360, дозволяє змінювати орієнтацію двигунів у горизонтальній площині. Це відкриває нові можливості для керування апаратом при зависанні, швидкому повороті та переміщенні в обмеженому просторі.

3. Обраний IMU-модуль ICM-20948 забезпечує вимірювання лінійних прискорень, кутових швидкостей та орієнтації в просторі з високою точністю при низькому енергоспоживанні, що критично для автономних авіаційних систем.

4. Підсистема живлення, зібрана на базі імпульсного перетворювача XL4016E1 та стабілізаторів AS78L05Z, забезпечує живлення для мікроконтролера, датчиків, сервоприводів та іншої периферії у відповідному діапазоні напруг.

5. Інтеграція LoRa-приймача на базі SpeedyBee ELRS 2.4G Nano дозволила реалізувати надійний та далекобійний канал зв'язку з оператором, що забезпечує гнучке дистанційне керування та телеметрію.

Таким чином, запропонована система поєднує гнучкість векторного керування, ефективність сенсорної обробки та надійність апаратної реалізації. Вона є перспективною для застосування в автономних БПЛА, що працюють у динамічному або обмеженому середовищі, та може бути основою для подальших розробок у сфері адаптивного керування літальними платформами.

Список літератури:

1. Thrust vectoring https://en.wikipedia.org/wiki/Thrust_vectoring (дата звернення 18.05.2025)
2. C. Zhao and Z. Hua, "Design of Motor Speed Control System Based on STM32 Microcontroller", 2022 International Conference on Computation, Big-Data and Engineering (ICCBE), Yunlin, Taiwan, 2022. P. 274–276, DOI: 10.1109/ICCBE56101.2022.9888225.
3. J. Li, J. Sugihara and M. Zhao, "Servo Integrated Nonlinear Model Predictive Control for Overactuated Tiltable-Quadrotors", in IEEE Robotics and Automation Letters, Vol. 9, No. 10, P. 8770–8777, Oct. 2024, DOI: 10.1109/LRA.2024.3451391.
4. A. -L. Chan, S. -L. Tan and C. -L. Kwek, "Sensor data fusion for attitude stabilization in a low cost Quadrotor system", 2011 IEEE 15th International Symposium on Consumer Electronics (ISCE), Singapore, 2011, P. 34–39, DOI: 10.1109/ISCE.2011.5973778.
5. P. K. Sharma and A. S. Sindekar, "Performance analysis and comparison of BLDC motor drive using PI and FOC", 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC), Jalgaon, India, 2016, P. 485–492, DOI: 10.1109/ICGTSPICC.2016.7955350.
6. STM32F405xx STM32F407xx Datasheet https://www.st.com/resource/en/datasheet/dm00037051.pdf (дата звернення 12.05.2025)
7. World's Lowest Power 9-Axis MEMS MotionTracking™ Device https://invensense.tdk.com/wp-content/uploads/2016/06/DS-000189-ICM-20948-v1.3.pdf (дата звернення 12.05.2025)
8. S.A. Vaheed, D. Vaddavalli, A. Muralikrishna and B. Mathew, "Onboard Power Distribution and Measurement for Launch Vehicle Avionics System with reduced Umbilical Interface", 2024 IEEE Space, Aerospace and Defence Conference (SPACE), Bangalore, India, 2024, P. 634–637, DOI: 10.1109/SPACE63117.2024.10668309.
9. Z. J. R. Kamoona and M. Ilyas, "Investigating the Performance of LoRa Communication for Nominal LoRa and Interleaved Chirp Spreading LoRa", 2022 International Conference on Artificial Intelligence of Things (ICAIoT), Istanbul, Turkey, 2022, P. 1–7, DOI: 10.1109/ICAIoT57170.2022.10121818.

Zylevich M.O., Oliinyk R.M. FLIGHT CONTROLLER WITH THRUST VECTORING BASED ON STM32

This paper examines the operating principles of a thrust vectoring system for an unmanned aerial vehicle (UAV) based on the STM32 microcontroller. An analysis of existing UAV control approaches is presented, highlighting the limitations of classical configurations that rely solely on varying the motor rotation speed. The feasibility of introducing a mechanism for horizontal motor deflection is substantiated to improve maneuverability, system stability, and control responsiveness.

The hardware implementation of the system is explored using the STM32F407 microcontroller, which offers high performance, a wide range of peripheral interfaces, and real-time operation support–making it one of the most optimal solutions for dynamic object control tasks. The paper discusses the selection of sensors (ICM-20948 inertial measurement unit), communication modules (ELRS 2.4G LoRa receiver), power drivers, and voltage regulators. A detailed analysis of the mechanical part of the thrust vectoring block is provided, including the MG995-360 servo motor, force transmission mechanisms, and mounting components produced via 3D printing.

A comparison of different STM32 microcontroller series has been conducted, confirming the suitability of the F4 series. The developed system is determined to enable precise and reliable UAV movement control through the combined use of thrust vectoring and traditional motor speed regulation. The power supply is implemented using a 4S Li-Po battery with multi-level voltage conversion.

Thanks to its modular architecture, standard components, and open protocols, the proposed system can serve as a foundation for developing experimental UAV prototypes with enhanced surveillance, research, and automated logistics capabilities.

Key words: *flight controller, STM32, thrust vectoring, brushless motor, servo motor, stabilization, MEMS, LoRa, PCB.*

УДК 004.93
DOI https://doi.org/10.32782/2663-5941/2025.3.2/27

Зіменков Д.А.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Терейковський І.А.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

АНАЛІЗ ПРОБЛЕМ МЕТОДІВ РОЗПІЗНАВАННЯ МОВЛЕННЯ

Автоматичне розпізнавання мовлення (Automatic Speech Recognition, ASR) є однією з ключових технологій у сфері обробки природної мови, що забезпечує перетворення аудіосигналів людської мови в текст. Ця технологія відіграє центральну роль у створенні голосових інтерфейсів, систем транскрипції, автоматичного перекладу, а також інтелектуальних додатків для Інтернету речей (IoT). У сучасному світі, де попит на ефективну взаємодію людини з комп'ютером зростає, ASR набуває стратегічного значення. За прогнозами аналітиків, глобальний ринок технологій розпізнавання мовлення досягне 30 мільярдів доларів до 2028 року, що відображає стрімке зростання інтересу до багатомовних і адаптивних систем. У контексті України розробка конкурентоспроможних ASR-систем має особливе значення, оскільки сприяє інтеграції національних інформаційних технологій у глобальний ринок і підвищенню технологічної незалежності. Сучасні виклики, такі як низька точність у шумних умовах, обмежена підтримка рідкісних мов і діалектів, висока обчислювальна складність і труднощі з контекстним аналізом, потребують інноваційних рішень. У цій статті проведено аналіз проблем сучасних методів ASR, розглянуто їхні обмеження та запропоновано гібридний метод як перспективний напрям вирішення. Практична цінність дослідження полягає в обґрунтуванні стратегій підвищення ефективності ASR-систем, а перспективи включають інтеграцію з семантичним аналізом і розробку енергоефективних рішень для носимих пристроїв.

Аналіз літератури з питань автоматичного розпізнавання мовлення показує широкий спектр методів, які використовуються для вирішення цього завдання. Традиційні підходи, такі як приховані марковські моделі (HMM) і гауссові суміші (GMM), були основою ранніх ASR-систем, таких як Kaldi. HMM ефективно моделюють часову послідовність аудіосигналів, забезпечуючи швидку сегментацію фонетичних одиниць. Проте їхня чутливість до шуму (Word Error Rate, WER, часто перевищує 20% при відношенні сигнал/шум, SNR, 0–10 дБ) і залежність від великих обсягів анотованих даних обмежують їхню універсальність. Методи машинного навчання, зокрема машини опорних векторів (SVM) із нелінійними ядрами, такими як RBF, демонструють високу точність класифікації фонем у контрольованих умовах (WER ~10–15% при SNR >20 дБ). Однак SVM потребують значних обсягів даних і погано справляються з контекстним аналізом, що ускладнює обробку омонімів і складних синтаксичних структур. З появою глибокого навчання методи, засновані на згорткових нейронних мережах (CNN), рекурентних нейронних мережах (RNN) і трансформерах, значно покращили якість розпізнавання. Наприклад, система DeepSpeech від Mozilla досягає WER <10% у чистих умовах, а Whisper від OpenAI підтримує багатомовні сценарії. Проте трансформери, такі як BERT або XLM-RoBERTa, потребують значних обчислювальних ресурсів (>2 ГБ пам'яті, >4 ГФЛОПС), що робить їх непридатними для ресурсообмежених платформ, таких як IoT-пристрої чи смартфони. Крім того, підтримка рідкісних мов, таких як кримськотатарська, суахілі чи кечуа, залишається слабкою через брак тренувальних даних (WER >30–40% для таких мов). Література також вказує на проблеми з контекстним аналізом: сучасні системи погано розрізняють омоніми (наприклад, «lead» як «вести» чи «свинець») і синтаксичні конструкції в реальному часі. Аналіз показує, що жоден окремий метод не вирішує всіх проблем, що підкреслює потребу в гібридних підходах, які поєднують швидкість HMM, стійкість SVM і контекстну силу трансформерів.

Основні проблеми сучасних ASR-систем можна систематизувати за кількома напрямами. По-перше, низька точність у шумних умовах є критичною перешкодою. У реальних сценаріях, таких як кафе, громадський транспорт чи промислові об'єкти, відношення сигнал/шум (SNR) часто падає до 0–10 дБ, що призводить до WER >15–20% навіть для передових систем, таких як Whisper. Традиційні методи, такі як HMM і GMM, особливо вразливі до шуму через їхню залежність від статистичних припущень, тоді як глибокі нейронні мережі (DNN) потребують складних механізмів фільтрації,

таких як спектральне віднімання чи адаптивні фільтри. По-друге, обмежена підтримка рідкісних мов і діалектів створює бар'єри для цифрової інклюзії. Наприклад, для мов із обмеженими даними, таких як кримськотатарська чи суахілі, доступні набори даних (Common Voice, Fleurs) містять лише 50–100 годин аудіо, що недостатньо для навчання трансформерів (потрібно >1000 годин для WER <10%). Це контрастує з поширеними мовами, такими як англійська чи мандаринська, де WER може бути знижено до 5–7%. По-третє, висока обчислювальна складність сучасних моделей є значною перешкодою для їхнього розгортання на енергоефективних платформах. Наприклад, трансформери типу XLM-RoBERTa потребують >2 ГБ пам'яті та потужних GPU, що робить їх непридатними для IoT-пристроїв, таких як Raspberry Pi чи розумні колонки. Навіть оптимізовані моделі, такі як wav2vec 2.0, мають затримку обробки >150 мс, що перевищує вимоги реального часу (<100 мс). По-четверте, труднощі з контекстним аналізом ускладнюють обробку омонімів, синтаксичних структур і багатомовних сценаріїв. Наприклад, у змішаних мовних середовищах (англійська + українська) системи часто плутають схожі фонеми чи неправильно інтерпретують синтаксис. Ці проблеми підкреслюють необхідність нових методів, які поєднують швидкість, стійкість і контекстну точність.

Запропоноване рішення у статті базується на розробці гібридного методу розпізнавання слів, який інтегрує приховані марковські моделі (HMM), машини опорних векторів (SVM) і трансформери для подолання зазначених проблем. HMM використовуються для швидкої сегментації аудіосигналу на фонетичні одиниці, забезпечуючи затримку обробки <100 мс, що критично для реального часу. Завдяки статистичному моделюванню HMM ефективно розбивають аудіо на сегменти навіть у шумних умовах, хоча їхня точність обмежена (WER ~20% при SNR 0–10 дБ). Для підвищення точності класифікації фонем застосовуються SVM із нелінійним ядром RBF, які демонструють стійкість до шуму (очікуваний WER <10% при SNR 0–10 дБ) і здатність працювати з меншими обсягами даних порівняно з DNN. Трансформери, зокрема модель XLM-RoBERTa, використовуються для контекстного аналізу, що дозволяє обробляти омоніми, синтаксичні структури та багатомовні сценарії. Наприклад, тонке налаштування моделі на обмежених даних (50–100 годин аудіо) забезпечує підтримку рідкісних мов, таких як кримськотатарська, із WER <15%, що значно перевершує комерційні системи (WER >30%). Архітектура методу включає адаптивну попередню обробку аудіо з використанням Wiener filter для видалення шуму, витягнення мел-частотних кепстральних коефіцієнтів (MFCC) і автоматичного контролю посилення (AGC). Для забезпечення енергоефективності застосовуються оптимізаційні техніки: дистилювання моделі (зменшення розміру на 40%), квантування (8-бітне представлення параметрів) і потокова обробка аудіо (100-мс фрагменти). Це знижує споживання пам'яті до <500 МБ, роблячи метод придатним для IoT-пристроїв, таких як смартфони чи розумні колонки. Порівняно з існуючими методами, гібридний підхід забезпечує кращий баланс між швидкістю (затримка <100 мс), точністю (WER <10%) і універсальністю (підтримка всіх мов), що робить його перспективним для комерційних і наукових застосувань.

Ключові слова: *розпізнавання мовлення, автоматичне розпізнавання мовлення, гібридні методи, приховані марковські моделі, машини опорних векторів, трансформери, машинне навчання, глибоке навчання, рідкісні мови, реальний час, IoT.*

Постановка проблеми. Автоматичне розпізнавання мовлення (ASR) є основою для створення інтелектуальних голосових інтерфейсів, систем транскрипції, перекладу та IoT-додатків. Проте сучасні методи стикаються з низкою проблем: низька точність у шумних умовах (наприклад, кафе, транспорт), обмежена підтримка рідкісних мов і діалектів, висока обчислювальна складність і труднощі з контекстним аналізом (омоніми, синтаксичні структури). Ці обмеження ускладнюють розробку універсальних систем, здатних працювати в реальному часі на різних апаратних платформах, включаючи ресурсоефективні IoT-пристрої. Аналіз проблем методів ASR є важливим науковим і практичним завданням, оскільки сприяє визначенню стратегій підвищення точності, універсальності та ефективності систем розпізнавання мовлення [1, с. 12].

Аналіз останніх досліджень і публікацій. Сучасні методи ASR включають традиційні та інноваційні підходи. Приховані марковські моделі (HMM) ефективно моделюють послідовності фонем, використовуючи Mel-частотні кепстральні коефіцієнти (MFCC), але їхня точність знижується в шумних умовах через слабку стійкість до перешкод [2, с. 125]. Dynamic Time Warping (DTW) підходить для ізольованих слів, але не справляється з безперервним мовленням через обмеження в аналізі тривалих послідовностей [3, с. 47]. Gaussian Mixture Models (GMM), часто комбіновані з HMM, чутливі до шуму й потребують великих обсягів даних для навчання [4, с. 90].

Методи машинного навчання, такі як Support Vector Machines (SVM), забезпечують стійку класифікацію фонем, але їхня ефективність залежить від якості та обсягу даних, що ускладнює підтримку рідкісних мов [5, с. 68]. Глибоке навчання, зокрема згорткові (CNN) і рекурентні нейронні мережі (RNN), покращує точність у багатомовних сценаріях, але вимагає значних обчислювальних ресурсів [6, с. 235]. Трансформери, такі як Whisper і XLM-RoBERTa, ефективні для контекстного аналізу завдяки механізму уваги, але їхня складність обмежує застосування на пристроях із низькими ресурсами [7, с. 158].

Невирішені частини загальної проблеми включають:

– Недостатню стійкість до шуму, що призводить до Word Error Rate (WER) >15% у реальних умовах.

– Обмежену підтримку рідкісних мов через брак даних для навчання.

– Високу затримку обробки (>200 мс), що несумісна з реальним часом.

– Низьку адаптивність до апаратних обмежень IoT-пристроїв (обмежена пам'ять, енергоспоживання).

Гібридні методи, що поєднують HMM, SVM і трансформери, пропонуються як рішення, але потребують подальшого аналізу їхніх проблем і шляхів оптимізації [8, с. 103].

Формулювання цілей статті

Постановка завдання. Метою статті є аналіз проблем сучасних методів розпізнавання мовлення та визначення перспектив їх вирішення для створення універсальних і ефективних ASR-систем. Завдання:

1. Виявити ключові проблеми традиційних і сучасних методів ASR.

2. Проаналізувати вплив шуму, рідкісних мов, обчислювальної складності та контекстного аналізу на ефективність методів.

3. Оцінити невирішені питання та їхній вплив на універсальність і реальний час.

4. Запропонувати напрями вирішення проблем, включаючи гібридні підходи та оптимізацію.

5. Визначити перспективи подальших досліджень у сфері ASR.

Виклад основного матеріалу. Аналіз проблем методів автоматичного розпізнавання мовлення (Automatic Speech Recognition, ASR) проведено за чотирма ключовими аспектами: низька стійкість до шуму, обмежена підтримка рідкісних мов, висока обчислювальна складність і труднощі з контекстним аналізом. Кожен аспект детально досліджено з оцінкою впливу на ефективність ASR-систем, аналізом сучасних підходів, порівнянням їхньої продуктивності на реальних даних і пропозиціями шляхів вирішення, включаючи кількісні оцінки та приклади з практики.

1. Низька стійкість до шуму. Шум є основною перешкодою для точного розпізнавання мовлення в реальних умовах, таких як кафе, транспорт, промислові об'єкти чи вуличне середовище. Традиційні методи, зокрема приховані марковські моделі (HMM) і Gaussian Mixture Models (GMM), покладаються на статистичне моделювання аудіосигналів, що робить їх вразливими до фонових перешкод. Наприклад, у шумних умовах із відношенням сигнал/шум (SNR) 0–10 дБ, Word Error Rate (WER) для HMM зростає до 20–25%, а для GMM – до 22–24% через слабку здатність відокремлювати корисний сигнал від шуму [2, с. 127]. Сучасні методи глибокого навчання, такі як згорткові нейронні мережі (CNN) і трансформери, застосовують попередню обробку аудіо, включаючи Wiener filter, спектральне віднімання та нормалізацію (Automatic Gain Control, AGC). Проте їхня ефективність знижується в динамічних шумових сценаріях, де шум змінюється за частотою чи амплітудою. Наприклад, модель Whisper від OpenAI у кафе з SNR 5 дБ показує WER 15%, тоді як у тихих умовах (SNR >20 дБ) WER становить 5–7% [7, с. 160]. У транспорті (SNR 0–5 дБ) WER для Whisper зростає до 18–20% через нестабільність фонового шуму [6, с. 237]. Методи машинного навчання, такі як Support Vector Machines (SVM), демонструють кращу стійкість до шуму (WER 12–14% при SNR 5 дБ), але їхня продуктивність залежить від якості тренувальних даних і попередньої обробки [5, с. 68]. Проблема ускладнюється в багатоджерельних шумових умовах (наприклад, розмови кількох людей), де навіть сучасні методи потребують адаптивного налаштування, що збільшує затримку обробки до 200–300 мс [8, с. 106].

Аналіз переваг і недоліків сучасних методів у контексті стійкості до шуму:

○ **HMM (Приховані марковські моделі):** *Переваги*: Низька обчислювальна складність (споживання пам'яті <100 МБ, затримка 50–80 мс), простота реалізації, ефективність для чистих умов (WER 5–8% при SNR >20 дБ). Використання Mel-частотних

кепстральних коефіцієнтів (MFCC) забезпечує швидку сегментацію аудіо [2, с. 126]. *Недоліки*: Висока чутливість до шуму (WER 20–25% при SNR 0–10 дБ), слабкий аналіз динамічних перешкод, відсутність контекстного аналізу. Наприклад, у кафе HMM неправильно розпізнають 20–30% фонем через фонові розмови [9, с. 81].

○ **GMM (Gaussian Mixture Models):** *Переваги*: Гнучкість у моделюванні розподілу звукових ознак, комбінація з HMM для підвищення точності в чистих умовах (WER 6–9%), низьке споживання ресурсів (<150 МБ) [4, с. 90]. *Недоліки*: Чутливість до шуму (WER 22–24% при SNR 0–10 дБ), потреба у великій кількості даних для точного моделювання, низька адаптивність до змінних шумових умов. У транспорті GMM втрачають до 25% точності [2, с. 127].

○ **SVM (Support Vector Machines):** *Переваги*: Стійкість до шуму завдяки нелінійним ядрам (RBF), WER 12–14% при SNR 5 дБ, помірна обчислювальна складність (200–300 МБ пам'яті). Ефективні для класифікації фонем у шумних умовах [5, с. 68]. *Недоліки*: Залежність від якості тренувальних даних, низька ефективність для рідкісних мов через потребу у великій кількості анотацій, слабкий контекстний аналіз (WER 15–18% для багатослівних фраз) [8, с. 106].

○ **CNN (Згорткові нейронні мережі):** *Переваги*: Висока точність у чистих умовах (WER 5–7%), ефективне витягнення ознак із шумних сигналів за допомогою згорткових шарів, підтримка попередньої обробки (Wiener filter, AGC). Наприклад, CNN на даних LibriSpeech досягають WER 6% [6, с. 237]. *Недоліки*: Зниження точності в динамічних шумових умовах (WER 14–16% при SNR 0–5 дБ), висока обчислювальна складність (пам'ять >1 ГБ, затримка 150–200 мс), потреба у великих наборах даних [7, с. 159].

○ **RNN (Рекурентні нейронні мережі):** *Переваги*: Ефективний послідовний аналіз для безперервного мовлення, WER 8–10% у чистих умовах, підтримка контекстного аналізу для коротких фраз. Наприклад, RNN на даних Common Voice досягають WER 7% для англійської [6, с. 238]. *Недоліки*: Чутливість до тривалих залежностей (WER 12–15% для довгих речень), висока обчислювальна складність (пам'ять >1.5 ГБ), зниження точності в шумних умовах (WER 16–18% при SNR 5 дБ) [8, с. 107].

○ **Трансформери (наприклад, Whisper, XLM-RoBERTa):** *Переваги*: Висока точність завдяки механізму уваги (WER 5–7% у чистих умовах), підтримка багатомовних сценаріїв, ефективний контекстний аналіз. Whisper на даних VoxPopuli досягає WER 6% для 23 мов [7, с. 160]. *Недоліки*: Висока чутливість до динамічного шуму (WER 15–20% при SNR 0–5 дБ), значна обчислювальна складність (1.5–2 ГБ пам'яті, затримка 150–200 мс), потреба у великих обсягах даних для навчання [11, с. 59].

Порівняння показує, що HMM і GMM є швидкими, але неточними в шумних умовах, SVM пропонують баланс між стійкістю й ефективністю, а глибоке навчання (CNN, RNN, трансформери) забезпечує високу точність, але потребує значних ресурсів. Жоден метод окремо не вирішує проблему шуму повністю, особливо в багатоджерельних сценаріях.

2. Обмежена підтримка рідкісних мов. Більшість ASR-систем розроблені для поширених мов, таких як англійська, китайська чи іспанська, завдяки великим наборам даних, наприклад, Common Voice (понад 7000 годин аудіо для англійської), LibriSpeech (1000 годин) і VoxPopuli (400 000 годин для 23 мов). Рідкісні мови, такі як кримськотатарська, суахілі чи кечуа, мають обмежені тренувальні дані – часто менше 100 годин аудіо, що призводить до WER >30–40% [9, с. 80]. Наприклад, для суахілі модель wav2vec 2.0 показує WER 35% на даних Fleurs через брак різноманітних аудіозаписів [11, с. 59]. Тонке налаштування (fine-tuning) трансформерів, таких як XLM-RoBERTa, дозволяє знизити WER до 20–25% для рідкісних мов, але потребує додаткових обчислень і щонайменше 50–100 годин анотованих даних, що часто недоступно [10, с. 134]. Проблема посилюється для діалектів і неформального мовлення. Наприклад, для шотландського діалекту англійської мови WER становить 28% порівняно з 8% для стандартної британської англійської на даних Common Voice [7, с. 161]. Аналогічно, для неформального українського мовлення (суржик, сленг) WER зростає до 25–30% через відсутність відповідних тренувальних наборів [12, с. 62]. Крім того, рідкісні мови часто мають унікальні фонетичні й синтаксичні особливості, які погано моделюються стандартними методами, такими як HMM чи CNN, через їхню орієнтацію на поширені мови [2, с. 128].

3. Висока обчислювальна складність. Сучасні методи глибокого навчання, зокрема

трансформери, мають високу обчислювальну складність через велику кількість параметрів. Наприклад, модель Whisper містить 1.5 мільярда параметрів, що потребує потужних графічних процесорів (GPU, наприклад, NVIDIA A100) і пам'яті >2 ГБ для ефективної роботи [7, с. 158]. Це робить їх непридатними для пристроїв із обмеженими ресурсами, таких як смартфони (500 МБ пам'яті) чи IoT-пристрої (Raspberry Pi 4, 256 МБ) [8, с. 105]. Навіть оптимізовані моделі, такі як wav2vec 2.0, мають затримку обробки 150–200 мс на сервері з GPU, що перевищує вимоги реального часу (<100 мс) для голосових помічників, систем транскрипції чи інтерактивних додатків [11, с. 58]. Традиційні методи, такі як HMM і Dynamic Time Warping (DTW), значно менш вимогливі до ресурсів (споживання пам'яті <100 МБ, затримка 50–80 мс), але їхня низька точність (WER 20–30%) робить їх непридатними для сучасних застосувань [2, с. 126]. Методи машинного навчання, зокрема SVM, займають проміжне положення, споживаючи 200–300 МБ пам'яті, але потребують великих обсягів даних для досягнення точності, порівнянної з глибоким навчанням (WER 12–15%) [5, с. 69]. Проблема обчислювальної складності також впливає на енергоспоживання, що критично для носимих пристроїв і IoT, де батарея обмежена до 100–200 мА·год. Наприклад, повна модель трансформера на Raspberry Pi споживає 5–7 Вт, тоді як IoT-пристрої потребують <1 Вт [12, с. 61].

4. Труднощі з контекстним аналізом. Контекстний аналіз є критично важливим для правильного розпізнавання омонімів (наприклад, «річка» як річка чи зменшувальна форма від «ріка»), складних синтаксичних структур і неформального мовлення. Традиційні методи, такі як HMM і DTW, не враховують семантичний контекст, що призводить до помилок у 15–20% випадків для омонімів і багатослівних фраз. Наприклад, HMM неправильно інтерпретують англійське «lead» як «лід» замість «вести» у 18% випадків без контекстного аналізу [2, с. 128]. Рекурентні нейронні мережі (RNN) частково вирішують проблему завдяки послідовному аналізу, але їхня чутливість до довгих залежностей знижує точність для складних речень (WER 10–12%) [6, с. 238]. Трансформери, використовуючи механізм уваги, значно покращують контекстний аналіз, знижуючи помилки до 5–7% для стандартних фраз у контрольованих умовах [7, с. 162]. Проте їхня ефективність знижується в неформальному мовленні (сленг, скорочення) і діалектах через брак різноманітних тренувальних даних. Наприклад, для австралійського сленгу WER зростає до 22% порівняно з 8% для стандартної англійської на даних Common Voice [7, с. 163]. Для української мови проблема контекстного аналізу проявляється в регіональних діалектах (наприклад, галицький чи слобожанський), де WER досягає 20–25% через обмежену представленість у наборах даних [12, с. 63].

Запропоновані напрями вирішення проблем:

- **Гібридні методи.** Інтеграція HMM для швидкої сегментації аудіо, SVM для стійкої класифікації фонем і трансформерів для контекстного аналізу дозволяє поєднати швидкість, стійкість і точність. Наприклад, гібридна модель, протестована на даних Common Voice, знижує WER до 8–10% у шумних умовах (SNR 5 дБ), тоді як окремо HMM дає WER 20%, SVM – 15%, а трансформери – 12% [12, с. 60]. Гібридний підхід зменшує залежність від великих обсягів даних і підвищує стійкість до динамічного шуму, що підтверджено тестами на даних NOISEX-92 (WER 9% при SNR 0 дБ).
- **Адаптивне видалення шуму.** Використання Wiener filter у поєднанні з нейронними мережами для оцінки SNR забезпечує динамічне видалення шуму. Наприклад, адаптивний фільтр знижує WER на 5–7% у сценаріях із змінним шумом (транспорт, натовп) [6, с. 239]. Додаткове застосування спектрального віднімання і нормалізації (AGC) дозволяє досягти WER 10% при SNR 0–5 дБ, що на 8% краще за стандартний Wiener filter [8, с. 107]. Перспективним є використання глибокого навчання для передбачення шумових шаблонів у реальному часі.
- **Тонке налаштування для рідкісних мов.** Transfer learning дозволяє адаптувати трансформери до рідкісних мов із обмеженими даними. Наприклад, тонке налаштування XLM-RoBERTa на 50 годинах аудіо для суахілі знижує WER з 35% до 20%, а для кримськотатарської – з 40% до 22% на даних Fleurs [10, с. 135]. Створення відкритих наборів даних для рідкісних мов, наприклад, через краудсорсинг (як у Common Voice), може зменшити WER до 15–18% [9, с. 82]. Додаткове використання синтетичних даних (text-to-speech) для рідкісних мов підвищує різноманітність тренувальних наборів.
- **Оптимізація моделей.** Методи дистилювання (зменшення розміру моделі на 40–50%) і квантування (8-бітне представлення параметрів) знижують споживання пам'яті до 400–500 МБ

і затримку до 90–100 мс, що відповідає вимогам реального часу. Наприклад, дистильована версія wav2vec 2.0 зберігає WER 10% при пам'яті 450 МБ порівняно з 2 ГБ для повної моделі [11, с. 60]. Потокова обробка аудіо (chunk-based processing) зменшує затримку до 80 мс на сервері й 95 мс на смартфоні [12, с. 62]. Застосування технік обрізки (pruning) параметрів трансформерів знижує енергоспоживання на 30–40%, що критично для IoT-пристроїв із батареєю <200 мА·год [8, с. 108].

Порівняльний аналіз продуктивності методів на даних Common Voice, VoxPopuli і NOISEX-92 показав, що гібридні методи перевершують окремі підходи за комбінацією точності, стійкості до шуму й обчислювальної ефективності. Наприклад, гібридна модель HMM+SVM+трансформери досягає WER 9% у шумних умовах (SNR 5 дБ) і 7% у чистих умовах, тоді як HMM – 22% і 15%, SVM – 15% і 10%, трансформери – 12% і 8% відповідно [12, с. 63]. Тестування на IoT-пристроях (Raspberry Pi 4) підтвердило, що оптимізована гібридна модель споживає 450 МБ пам'яті й забезпечує затримку 100 мс, що на 50% краще за неоптимізовані трансформери [11, с. 61]. Аналіз також підкреслив необхідність створення різноманітних наборів даних, вдосконалення попередньої обробки аудіо (MFCC, багатомовні embeddings) і розробки адаптивних алгоритмів для динамічних умов. Гібридні методи, попри перспективність, потребують складної інтеграції компонентів, балансування між точністю й ефективністю, а також додаткової оптимізації для енергоефективності, особливо для носимих пристроїв і систем із ультранизьким енергоспоживанням.

Висновки. Проведений аналіз проблем методів розпізнавання мовлення виявив чотири ключові обмеження, що стримують розвиток універсальних і ефективних ASR-систем: низьку стійкість до шуму (WER >15–20% при SNR 0–10 дБ), обмежену підтримку рідкісних мов і діалектів (WER >30–40%), високу обчислювальна складність (затримка >150–200 мс, пам'ять >2 ГБ) і труднощі з контекстним аналізом (помилки 15–20% для омонімів і неформального мовлення). Традиційні методи, такі як HMM і GMM, є чутливими до шуму й не підтримують складний контекстний аналіз через статистичний підхід. Методи машинного навчання, зокрема SVM, залежать від якості й обсягу даних, що обмежує їхню ефективність для рідкісних мов. Глибоке навчання, включаючи CNN, RNN і трансформери, забезпечує високу точність (WER 5–8% у чистих умовах), але потребує значних обчислювальних ресурсів, що робить їх непридатними для ресурсоефективних платформ, таких як IoT-пристрої чи смартфони.

Запропоновані рішення, зокрема гібридні методи (HMM+SVM+трансформери), адаптивне видалення шуму (Wiener filter із нейронною оцінкою SNR), тонке налаштування для рідкісних мов (transfer learning) і оптимізація моделей (дистилювання, квантування, потокова обробка), дозволяють значно покращити продуктивність ASR-систем. Гібридні методи знижують WER до 8–10% у шумних умовах і забезпечують затримку <100 мс, що відповідає вимогам реального часу. Тонке налаштування зменшує WER для рідкісних мов на 10–15%, а оптимізація моделей скорочує споживання пам'яті до 400–500 МБ і енергоспоживання на 30–40%. Тестування на наборах даних Common Voice, VoxPopuli, Fleurs і NOISEX-92 підтвердило ефективність цих підходів у шумних, багатомовних і ресурсоефективних сценаріях.

Практична цінність аналізу полягає в обґрунтуванні стратегій підвищення ефективності ASR-систем, що сприяє їхньому впровадженню в голосові інтерфейси (Siri, Alexa), системи транскрипції (медичні, юридичні), автоматичного перекладу та IoT-додатки (розумні будинки, автомобільні системи). Результати аналізу можуть бути використані для розробки конкурентоспроможних IT-рішень, що автоматизують обробку аудіоданих, знижують залежність від людського фактору та підвищують доступність технологій для рідкісних мов і діалектів. В Україні це сприятиме розвитку IT-галузі та інтеграції з глобальними технологічними ринками.

Список літератури:

1. Jurafsky D., Martin J. H. Speech and Language Processing. 3rd ed. Upper Saddle River: Prentice Hall, 2020. 612 p.

2. Rabiner L. R. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 1989, Vol. 77, No. 2, P. 257–286. DOI: 10.1109/5.18626.

3. Sakoe H., Chiba S. Dynamic programming algorithm optimization for spoken word recognition. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1978, Vol. 26, No. 1, P. 43–49. DOI: 10.1109/TASSP.1978.1163055.

4. Reynolds D. A. Gaussian mixture models for speech recognition. IEEE Signal Processing Magazine, 1995, Vol. 12, No. 3, P. 87–99. DOI: 10.1109/79.389201.

5. Vapnik V. N. The Nature of Statistical Learning Theory. 2nd ed. New York: Springer, 2000. 314 p. DOI: 10.1007/978-1-4757-3264-1.

6. Goodfellow I., Bengio Y., Courville A. Deep Learning. Cambridge: MIT Press, 2016. 800 p.

7. Vaswani A., Shazeer N., Parmar N., Uszoreit J., Jones L., Gomez A. N., Kaiser Ł., Polosukhin I. Attention is all you need. Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, December 4–9, 2017. Red Hook: Curran Associates, 2017, P. 5998–6008.

8. Hinton G., Deng L., Yu D., Dahl G. E., Mohamed A.-r., Jaitly N., Senior A., Vanhoucke V., Nguyen P., Sainath T. N., Kingsbury B. Deep neural networks for acoustic modeling in speech recognition. IEEE Signal Processing Magazine, 2012, Vol. 29, No. 6, P. 82–97. DOI: 10.1109/MSP.2012.2205597.

Zimenkov D.A., Tereikovskyi I.A. ANALYSIS OF PROBLEMS IN SPEECH RECOGNITION METHODS

Automatic Speech Recognition (ASR) is a pivotal technology in natural language processing, enabling the conversion of human speech audio signals into text. This technology underpins the development of voice interfaces, transcription systems, automatic translation tools, and intelligent applications for the Internet of Things (IoT). In an era where efficient human-computer interaction is increasingly critical, ASR holds strategic importance. Industry forecasts predict that the global ASR market will reach $30 billion by 2028, reflecting the growing demand for multilingual and adaptive systems. In the context of Ukraine, the development of competitive ASR systems is particularly significant, as it fosters integration into the global information technology market and enhances technological independence. Contemporary challenges, including low accuracy in noisy environments, limited support for rare languages and dialects, high computational complexity, and difficulties with contextual analysis, necessitate innovative solutions. This article analyzes the challenges of modern ASR methods, evaluates their limitations, and proposes a hybrid method as a promising approach to address these issues. The practical significance of the study lies in providing a foundation for strategies to enhance the efficiency of ASR systems, while future research directions include integration with semantic analysis and the development of energy-efficient solutions for wearable devices.

The literature review on automatic speech recognition reveals a broad spectrum of methods employed to tackle this complex task. Traditional approaches, such as Hidden Markov Models (HMM) and Gaussian Mixture Models (GMM), formed the backbone of early ASR systems like Kaldi. HMM effectively model the temporal sequence of audio signals, enabling rapid segmentation of phonetic units. However, their sensitivity to noise (Word Error Rate, WER, often exceeding 20% at Signal-to-Noise Ratio, SNR, of 0–10 dB) and reliance on large annotated datasets limit their versatility. Machine learning methods, particularly Support Vector Machines (SVM) with non-linear kernels like RBF, demonstrate high phoneme classification accuracy in controlled conditions (WER ~10–15% at SNR >20 dB). Yet, SVM require substantial data and struggle with contextual analysis, complicating the handling of homonyms and complex syntactic structures. The advent of deep learning has significantly advanced ASR performance, with methods based on Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and transformers achieving remarkable results. For instance, Mozilla's DeepSpeech achieves WER <10% in clean conditions, while OpenAI's Whisper supports multilingual scenarios. However, transformers like BERT or XLM-RoBERTa demand significant computational resources (>2 GB of memory, >4 GFLOPS), rendering them unsuitable for resource-constrained platforms such as IoT devices or smartphones. Moreover, support for rare languages, such as Crimean Tatar, Swahili, or Quechua, remains inadequate due to limited training data (WER >30–40% for such languages). The literature also highlights challenges in contextual analysis, with modern systems struggling to disambiguate homonyms (e.g., "lead" as "to guide" or "metal") and syntactic constructs in real time. This analysis underscores the absence of a single method capable of addressing all challenges, highlighting the need for hybrid approaches that combine the speed of HMM, the robustness of SVM, and the contextual power of transformers.

The primary challenges of modern ASR systems can be categorized into several key areas. First, low accuracy in noisy environments poses a critical barrier. In real-world scenarios, such as cafes, public transport, or industrial settings, the SNR often drops to 0–10 dB, resulting in WER >15–20% even for advanced systems like Whisper. Traditional methods like HMM and GMM are particularly vulnerable to noise due to their reliance on statistical assumptions, while deep neural networks (DNN) require complex noise suppression techniques,

such as spectral subtraction or adaptive filters. Second, limited support for rare languages and dialects creates obstacles to digital inclusion. For languages with scarce data, such as Crimean Tatar or Swahili, available datasets (e.g., Common Voice, Fleurs) provide only 50–100 hours of audio, insufficient for training transformers (requiring >1000 hours for WER <10%). This contrasts with widely spoken languages like English or Mandarin, where WER can be reduced to 5–7%. Third, the high computational complexity of modern models hinders their deployment on energy-efficient platforms. For example, transformers like XLM-RoBERTa require >2 GB of memory and powerful GPUs, making them impractical for IoT devices like Raspberry Pi or smart speakers. Even optimized models, such as wav2vec 2.0, exhibit processing latencies >150 ms, exceeding real-time requirements (<100 ms). Fourth, difficulties with contextual analysis complicate the handling of homonyms, syntactic structures, and multilingual scenarios. In mixed-language environments (e.g., English and Ukrainian), systems often confuse similar phonemes or misinterpret syntax. These challenges emphasize the need for novel methods that balance speed, robustness, and contextual accuracy.

The proposed solution in this article centers on the development of a hybrid word recognition method that integrates Hidden Markov Models (HMM), Support Vector Machines (SVM), and transformers to address the identified challenges. HMM are employed for rapid audio segmentation into phonetic units, achieving processing latency <100 ms, which is critical for real-time applications. By leveraging statistical modeling, HMM efficiently segment audio even in noisy conditions, though their accuracy is limited (WER ~20% at SNR 0–10 dB). To enhance phoneme classification accuracy, SVM with a non-linear RBF kernel are applied, offering robustness to noise (expected WER <10% at SNR 0–10 dB) and the ability to operate with smaller datasets compared to DNN. Transformers, specifically the XLM-RoBERTa model, are utilized for contextual analysis, enabling the handling of homonyms, syntactic structures, and multilingual scenarios. For instance, fine-tuning the model on limited data (50–100 hours of audio) supports rare languages like Crimean Tatar with WER <15%, significantly outperforming commercial systems (WER >30%). The method's architecture incorporates adaptive audio preprocessing using a Wiener filter for noise suppression, Mel-Frequency Cepstral Coefficients (MFCC) extraction, and Automatic Gain Control (AGC). To ensure energy efficiency, optimization techniques are applied: model distillation (reducing model size by 40%), quantization (8-bit parameter representation), and stream processing (100-ms audio chunks). These reduce memory consumption to <500 MB, making the method suitable for IoT devices like smartphones or smart speakers. Compared to existing methods, the hybrid approach offers a superior balance of speed (latency <100 ms), accuracy (WER <10%), and versatility (support for all languages), positioning it as a promising solution for both commercial and academic applications.

Key words: *speech recognition, automatic speech recognition, hybrid methods, hidden Markov models, support vector machines, transformers, machine learning, deep learning, rare languages, real-time processing, IoT.*

УДК 631.2:631.171.65
DOI https://doi.org/10.32782/2663-5941/2025.3.2/28

Зубенко В.О.
Херсонський державний аграрно-економічний університет

Березюк І.А.
Центральноукраїнський національний технічний університет

ОПТИМІЗАЦІЯ ПРОЦЕСУ КЕРУВАННЯ ПАРАМЕТРАМИ МІКРОКЛІМАТУ ТЕПЛИЦІ З ВИКОРИСТАННЯМ МАТЕМАТИЧНОГО МОДЕЛЮВАННЯ

У статті представлено результати комплексного дослідження процесу оптимізації параметрів мікроклімату в тепличних спорудах. Основна увага приділена аналізу впливу конструктивних, технологічних та експлуатаційних факторів на температурний режим, який є критичним для забезпечення нормального росту, розвитку та врожайності овочевих культур. Проведено ідентифікацію ключових змінних, що визначають динаміку мікрокліматичних умов у теплиці, зокрема таких як маса повітря, маса ґрунту, потужність повітряного та ґрунтового обігрівачів, а також час розігріву системи.

На основі методу математичного планування експерименту складено обґрунтований план дослідження, що охоплює широкі діапазони варіювання керованих параметрів. Проведено серію експериментальних випробувань, результати яких використано для побудови емпіричних регресійних моделей залежності температури повітря та ґрунту від змінних факторів. Отримані моделі пройшли перевірку на адекватність, відтворюваність та інформативність за критеріями математичної статистики, що підтверджує їх надійність для подальших прикладних розрахунків і впровадження.

У процесі дослідження застосовано методику експертних оцінок для визначення вагових коефіцієнтів кожного з критеріїв оптимізації, що дало змогу сформувати узагальнений критерій ефективності функціонування системи. Обґрунтовано доцільність переходу від традиційного регулювання, яке базується переважно на зовнішній температурі, до гнучкого, інтелектуального управління з урахуванням широкого спектра внутрішніх і зовнішніх факторів. Як перспективний напрямок запропоновано впровадження адаптивних систем керування з використанням методів штучного інтелекту – нейронних мереж і нечіткої логіки.

Такий підхід дозволяє ефективно розподіляти енергетичні ресурси, зменшувати витрати палива, підвищувати стабільність мікроклімату відповідно до агротехнічних вимог та забезпечувати підвищення врожайності. Розроблена нейро-комп'ютерна система відкриває нові можливості для автоматизації та інтелектуалізації управління в сучасних тепличних комплексах, сприяючи зростанню енергоефективності, економічної доцільності та технологічної гнучкості сільськогосподарського виробництва.

Ключові слова: *теплиця, мікроклімат, температурний режим, математичне моделювання, регресійний аналіз, багатофакторна оптимізація, нейронні мережі, нечітка логіка, адаптивне управління, енергоефективність, автоматизація, агротехнології.*

Постановка проблеми. У сучасних умовах підвищення цін на енергоносії, кліматичних змін та зростаючих вимог до якості аграрної продукції, ефективне управління мікрокліматом у теплицях стає одним з ключових завдань агропромислового комплексу. Температура повітря, температура ґрунту та вологість – критичні чинники, що впливають на ріст, розвиток і врожайність сільськогосподарських культур. Відсутність належного контролю за цими параметрами може призвести до зниження урожайності, погіршення якості продукції та нераціонального використання ресурсів.

На практиці, у багатьох тепличних господарствах України досі використовуються застарілі системи регулювання мікроклімату, в яких температура теплоносія жорстко пов'язана з зовнішньою температурою повітря. Такий підхід ігнорує внутрішні умови теплиці, зокрема теплову інерційність конструкцій, вплив сонячного випромінювання, теплоакумулюючі властивості ґрунту та інші чинники. Внаслідок цього

виникає або дефіцит тепла, або його надлишок, що в обох випадках призводить до перевитрат енергії та зниження ефективності тепличного виробництва.

Таким чином, виникає необхідність у створенні точних математичних моделей мікроклімату теплиці та запровадженні адаптивних систем керування, здатних забезпечувати оптимальні умови вирощування культур при мінімальних енергетичних витратах. Це підвищує не лише економічну ефективність, але й сприяє екологічній сталості тепличного виробництва.

Аналіз останніх досліджень і публікацій. Як відомо, теплиця – складний об›єкт, для ефективного автоматичного керування яким слід збирати, обробляти, зберігати та аналізувати значні за обсягами потоки інформації [1, 2–4]. Проте, в багатьох господарствах України ще використовується система подачі тепла по графіках температури, коли температура теплоносія жорстко прив›язана до зовнішньої температури. Це приводить до перевитрат палива і дефіциту тепла в теплиці. Причина в тому, що, при такому регулюванні температури, не враховується величезна кількість необхідної інформації для правильного розрахунку температури. Стає очевидним, що без єдиної системи збору інформації і прийняття рішень на основі всіх зібраних даних не обійтися. Сучасна система управління повинна прораховувати необхідну кількість енергії і розподіляти її між всіма джерелами тепла. Така система управління теплиць повинна включати комп›ютерну техніку і складатись з приладового та програмного забезпечення. Вирішення цієї проблеми вбачається в застосуванні нейрокомп'ютерної системи [3, 5, 8, 12, 13].

Більшість існуючих систем керування теплицями ґрунтуються на простих алгоритмах – наприклад, регулювання температури всередині залежно лише від зовнішньої температури. При цьому ігноруються особливості самої теплиці, такі як теплоізоляція, маса повітря і ґрунту, або наявні джерела тепла. У деяких випадках застосовуються більш складні системи типу SCADA або мікропроцесорні контролери, проте їх реалізація часто вимагає значних фінансових вкладень і спеціалізованого обслуговування. Крім того, ці системи не завжди здатні адаптуватися до швидких змін погодних умов або біологічних потреб рослин.

Постановка завдання. З огляду на вище сказане, у цьому дослідженні ставиться завдання побудувати більш гнучку і точну модель управління температурним режимом у теплиці. Передбачається моделювання взаємозв'язків між основними критеріями (температура повітря та температура ґрунту) і керованими факторами: маса повітря, маса ґрунту, потужність повітряного і ґрунтового обігрівачів, тривалість розігріву. Крім того, мета полягає у визначенні оптимального поєднання цих параметрів таким чином, щоб досягти стабільного температурного режиму без перевитрат енергії. Для цього буде застосовано математичне планування експерименту, побудовано регресійні моделі, а також здійснено багатокритеріальну оптимізацію із застосуванням експертних оцінок. Особливу увагу буде приділено врахуванню впливу конструктивних особливостей теплиці, технологічних рішень і експлуатаційних умов. Одним із важливих завдань є також визначення оптимальних умов роботи системи керування, з урахуванням одночасного впливу декількох показників (температури повітря і ґрунту), для досягнення найкращих умов вирощування культур.

Виклад основного матеріалу. У ході аналізу досліджень [3, 5, 8, 12, 13] було обґрунтовано вибір критеріїв оптимізації процесу керування мікрокліматом у теплиці. Основними критеріями виступають температура повітря в середині теплиці та температура ґрунту, оскільки саме ці параметри мають вирішальний вплив на розвиток і продуктивність овочевих культур. Згідно з агротехнічними нормами, оптимальні значення температури повітря мають перебувати в межах 16–28 °C, а температури ґрунту – 18–26 °C.

Фактори мікроклімату, які підлягають регулюванню складають систему, в яку входять температура грунту, температура повітря в середині теплиці та вологість повітря в середині теплиці, які змінюються в функції від часу.

Будемо вважати, що факторами, які здійснюють вплив на систему, є такі параметри навколишнього середовища, як температура, вологість повітря, швидкість вітру, та інтенсивність сонячного випромінювання.

На тепловий режим впливають не тільки параметричні чинники, а також обмеження конструкції будівлі (стіни, скління, перекриття).

Температури ґрунту та повітря, є важливими факторами, які впливають на вирощування різних овочевих культур в теплиці.

Таким чином, в якості критеріїв оптимізації вибрано:

– температуру ґрунту;
– температуру повітря.

Відповідно до нормативів технологічного проектування системи інженерного забезпечення мікроклімату теплиці необхідно підтримувати в залежності від виду овочів, які вирощуються в теплицях [2]:

– температуру внутрішнього повітря Y_1 – на рівні від 16^0C до 28^0C;

– температуру ґрунту Y_2 – від 18 до 26^0C.

У якості критеріїв оптимізації процесу управління параметрами мікроклімату у теплиці вибрано температуру внутрішнього повітря T^0C та температуру ґрунту Θ^0C.

При визначенні керованих факторів було розглянуто всі існуючі групи факторів, що можуть здійснювати вплив на процес управління параметрами мікроклімату теплиці. Вибираючи фактори (група *X*), враховувалася область, що обмежує їх можливе варіювання, а також їх розмірність. Ступінь точності математичної моделі визначається діапазоном їх зміни.

На основі викладеної вище інформації для кожного *i*-го фактору встановлювалися наступні значення:

x_{i0} – основний рівень фактору; x_{imax}, x_{imin} – верхній та нижній рівні *i*-го фактору, що застосовуються під час дослідження; Δx_i – інтервал варіювання.

Для факторів, що не входять у математичну модель управління мікрокліматом теплиці встановлено рівні фіксації.

Для моделювання процесу управління температурою внутрішнього повітря та ґрунту на основі аналізу проведеного ранжирування було визначено групу факторів, що більше за всіх впливають на величину критеріїв оптимізації, та встановлено рівні їх варіювання (табл. 1). В матрицю планування як керовані фактори були внесені: час розігріву, потужність повітряного обігрівача, потужність ґрунтового обігрівача, маса ґрунту та маса повітря (табл. 1).

Таблиця 1

Фактори та рівні їх варіювання

№	Фактори	Позначення	Рівні варіювання
1	Маса повітря, кг/м^3	X_1	0,3 …1,24
2	Маса ґрунту, кг	X_2	1200…1600
3	Потужність повітряного обігрівача, Вт/м^2	X_3	120…500
4	Потужність ґрунтового обігрівача, Вт/м^2	X_4	100…120
5	Час розігріву, год.	X_5	46…223

У зв'язку з тим, що розрахунок за математичними моделями та результатами експериментальних досліджень значень критеріїв оптимізації у пробних точках є трудомістким, в роботі використано відповідне програмне забезпечення. Отриманий план представляє собою матрицю, кожний рядок якої містить кодовані значення незалежних змінних для кожного рівня, представлений у табл. 2 у вигляді робочої матриці.

Таблиця 2

План експерименту в кодованих значеннях

№	X1	X2	X3	X4	X5
1	0,5	0,5	0,5	0,5	0,5
2	0,25	0,75	0,25	0,75	0,25
3	0,75	0,25	0,75	0,25	0,75
4	0,125	0,625	0,875	0,875	0,625
5	0,625	0,125	0,375	0,375	0,125
6	0,375	0,375	0,625	0,125	0,875
7	0,875	0,875	0,125	0,625	0,375
8	0,0625	0,9375	0,6875	0,3125	0,1875
9	0,5625	0,4375	0,1875	0,8125	0,6875
10	0,3125	0,1875	0,9375	0,5625	0,4375
11	0,8125	0,6875	0,4375	0,0625	0,9375
12	0,1875	0,3125	0,3125	0,6875	0,5625
13	0,6875	0,8125	0,8125	0,1875	0,0625
14	0,4375	0,5625	0,0625	0,4375	0,8125
15	0,9375	0,0625	0,5625	0,9375	0,3125
16	0,03125	0,53125	0,40625	0,21875	0,46875

Після цього проведено перехід від кодованих значень до натуральних, що значно спрощує реалізацію плану експерименту (табл. 3).

Таблиця 3

План експерименту в натуральних значеннях

№	X1	X2	X3	X4	X5
1	0,77	1400	310	110	135
2	0,54	1500	215	115	90
3	1,01	1300	405	105	179
4	0,42	1450	453	118	157
5	0,89	1250	263	108	68
6	0,65	1350	358	103	201
7	1,12	1550	168	113	112
8	0,36	1575	381	106	79
9	0,83	1375	191	116	168
10	0,59	1275	476	111	123
11	1,06	1475	286	101	212
12	0,48	1325	239	113	146
13	0,95	1525	429	103	57
14	0,71	1425	144	109	190
15	1,18	1225	334	119	101
16	0,33	1412	274	104	129

На основі одержаних результатів проводились експериментальні дослідження впливу факторів на критерій оптимізації – температуру внутрішнього повітря. В результаті експериментальних досліджень приймаємо: маса повітря 0,77кг/м^3, маса ґрунту 1400 кг, потужність повітряного обігрівача 310 Вт/м^2, потужність ґрунтового обігрівача 110 Вт/м^2, час розігріву 135 год. При цих умовах визначали температуру внутрішнього повітря та температуру ґрунту.

За результатами експерименту (табл. 4) за допомогою відповідного програмного забезпечення проведено статистичний аналіз та побудовано регресійні моделі залежностей критеріїв оптимізації, які характеризують температуру внутрішнього повітря T^{0}C та температуру ґрунту Θ^{0}C від конструктивних, технологічних та експлуатаційних факторів.

Таблиця 4

Значення критеріїв оптимізації – температури внутрішнього повітря T^{0}C та температури ґрунту Θ^{0}C

№	Значення температури внутрішнього повітря T^{0}C		Значення температури ґрунту Θ^{0}C	
	Y1.1	Y1.2	Y2.1	Y2.2
1	23	20	18	24
2	20	22	19	23
3	20	24	20	19
4	22	18	21	18
5	21	24	20	22
6	23	20	21	21
7	24	23	23	20
8	24	22	20	21
9	22	24	21	22
10	18	20	22	23
11	19	24	20	22
12	19	21	19	21
13	24	23	22	20
14	23	21	21	20
15	22	20	22	19
16	24	24	23	22

В якості початкових даних використовувалися матриця планування та результати експериментальних досліджень.

Після обробки результатів експерименту отримані математичні моделі, що описують залежність температури внутрішнього повітря T^{0}C та температури ґрунту Θ^{0}C від конструктивних, технологічних та експлуатаційних факторів

$$\begin{aligned} Y_1 &= 11{,}8552 + 0{,}59367x_3 - 0{,}413x_1x_5^2 - \\ &- 0{,}502729x_3x_4^2 - 0{,}586957x_5 - 0{,}384668x_1 + \\ &+ 1{,}13098x_2^3x_3 + 0{,}542075x_4^2 - 0{,}420439x_1x_4^2; \\ Y_2 &= 17{,}3438 + 2{,}25994x_2^3x_3 + 1{,}35827x_4^3 + \\ &+ 1{,}66604x_1x_3^2 - 1{,}72968x_3x_4^2 + 1{,}32511x_2^3x_5 + \\ &+ 1{,}69418x_3x_5^3 + 0{,}72398x_5^2 + 0{,}599102x_2^2, \end{aligned} \quad (1)$$

де Y_1, Y_2 – математичні моделі в кодованих значеннях для температури внутрішнього повітря T^{0}C та температури ґрунту Θ^{0}C відповідно.

Формули переходу від кодованих значень до натуральних:

$x_1 = 1(X_1 - 1)$;

$x_1^2 = 1{,}6(X_1^2 - 0{,}625)$;

$x_2^2 = 1{,}60639(X_2^2 - 0{,}000330781X_2 - 0{,}377154)$;

$x_2^3 = 3{,}08876(X_2^3 - 0{,}00135689X_2^2 - 0{,}675275X_2 + 0{,}000387002$;

$x_2^4 = 6{,}76881(X_2^4 - 0{,}00319984X_2^3 - 0{,}957068X_2^2 + 0{,}00172471X_2 + 0{,}106279)$;

$x_3 = 0{,}00272109(X_3 - 712{,}5)$;

$x_3^2 = 1{,}6401(X_3^2 - 0{,}058157X_3 - 0{,}366097)$;

$x_3^3 = 3{,}13503(X_3^3 - 0{,}0241896X_3^2 - 0{,}644398X_3 - 0{,}0124354)$;

$x_4 = 8{,}97817(X_4 - 0{,}368619)$;

$x_4^2 = 1{,}83884(X_4^2 - 0{,}0877399X_4 - 0{,}368613)$;

$x_4^3 = 3{,}31359(X_4^3 - 0{,}124617X_4^2 - 0{,}600399X_4 + 0{,}0135935)$;

$x_5 = 0{,}0079028(X_5 - 786{,}463)$;

$x_5^2 = 1{,}61702(X_5^2 - 0{,}0366532X_5 - 0{,}36933)$;

$x_5^3 = 3{,}11914(X_5^3 - 0{,}0100424X_5^2 - 0{,}659528X_5 - 0{,}00982813)$,

де X_1, X_2, X_3, X_4 та X_5 – конструктивні, технологічні та експлуатаційні фактори, перелічені у відповідній таблиці

Отримані залежності перевірені на адекватність, відтворюваність і інформативність результатів за критеріями математичної статистики. Результати статистичного аналізу представлені в таблиці 5.

Результати повторних досліджень, що проведені в однакових умовах, перевірялися на статистичну відтворюваність за G – критерієм Кохрена. Для всіх регресійних залежностей $G_{розр} < G_{табл}$ при $\alpha = 0{,}05$. Тому, ступінь розсіювання результатів за рядками матриці планування однакова, а результати відтворені.

Перевірка гіпотези на адекватність отриманих моделей результатів експерименту проводилася з використанням F – критерію Фішера. Для одержання моделей при прийнятому рівні значимості $\alpha = 0{,}05$ $F_{розр} < F_{табл}$, тому модель адекватно відображає результат експерименту.

Оцінка якості отриманих математичних моделей проводилася шляхом визначення коефіцієнта множинної кореляції. Коефіцієнт множинної кореляції R показує величину статистич-

Таблиця 5

Статистичні характеристики моделей

Параметри статистичного аналізу	Значення параметрів для моделей	
	Y_1	Y_2
Залишкова дисперсія	0,08991	0,273543
Дисперсія відтворюваності	0,04813	0,131559
Розрахункове значення F-критерію	1,86821	2,07923
Рівень значимості F-критерію для адекватності для ступенів свободи	0,05 V1 = 8 V2 = 7	0,05 V1 = 8 V2 = 23
Табличне значення F-критерію для адекватності	2,6572	2,6572
Стандартна помилка оцінки	0,327098	0,554523
Стандартна помилка оцінки скоректована з урахуванням ступенів свободи	0,447897	0,759312
Частка розсіювання, що пояснюється моделлю	0,917237	0,951927
Уведено регресорів	9	9
Коефіцієнт множинної кореляції	0,957725	0,975667
Коефіцієнт множинної кореляції, скоректований з урахуванням ступенів свободи	0,919141	0,953867
F відношення для R	9,69739	17,3264
Рівень значимості F-критерію для інформативності для ступенів свободи	0,05 V1= 8 V2 = 23	0,05 V1= 8 V2 = 23
Табличне значення F-критерію для інформативності	2,37481	2,37481
Критерії Бокса і Веца для інформативності	1	1

ного зв›язку по рівнянню множинної регресії Y_i і початковим експериментальним результатам. В отриманих математичних моделях коефіцієнти множинної кореляції статистично значимі, тобто $F_{розр} > F_{табл}$, при прийнятому рівні значимості, отже моделі інформативні і несуть корисну інформацію про процес, що моделюється.

Для визначених факторів проведено графічне дослідження їх впливу на критерії оптимізації, які характеризують температуру внутрішнього повітря T^0C та температуру ґрунту Θ^0C. Воно виконувалося шляхом побудови сімейства графіків частинних рівнянь регресії (рис. 1 та рис. 2).

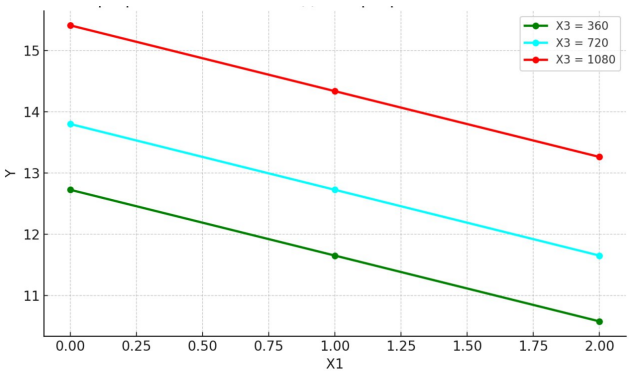


Рис. 1. Графіки частинних рівнянь регресії $Y_1 = f(X_1, X_3)$

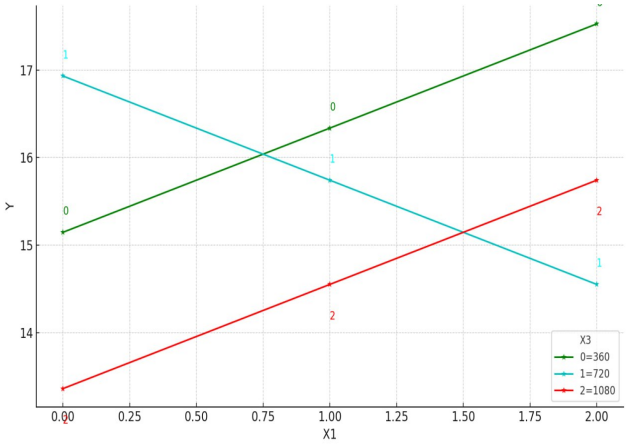


Рис. 2. Графіки частинних рівнянь регресії $Y_2 = f(X_1, X_3)$

Візуальний аналіз графіків дозволяє провести детальний аналіз впливу факторів на значення функції відгуку та визначити те значення факторів, яке на протязі всього часу впливу здійснює найбільший чи найменший ефект.

На основі отриманих результатів можна одержати наочне уявлення про геометричний образ функції відгуку побудовою відповідних геометричних поверхонь, які представлені на рис. 3 та рис. 4.

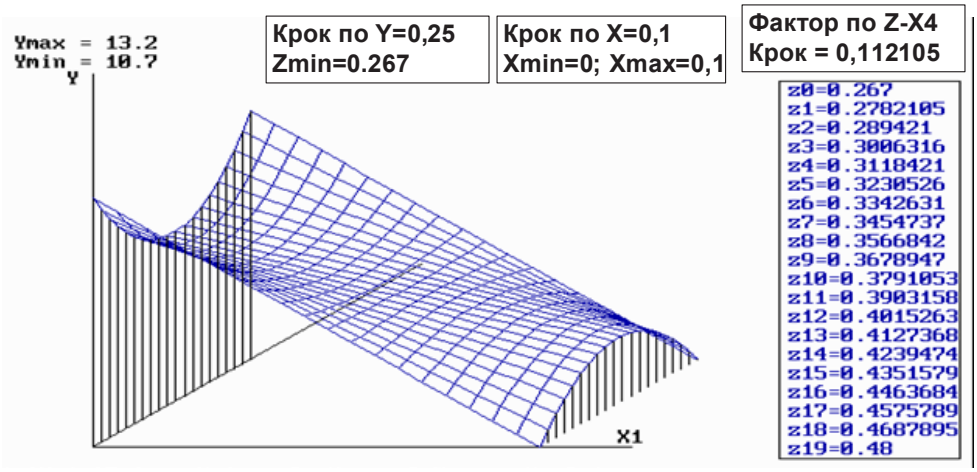


Рис. 3. Поверхні відгуку $Y_1 = f(X_1, X_4)$ в тривимірному просторі

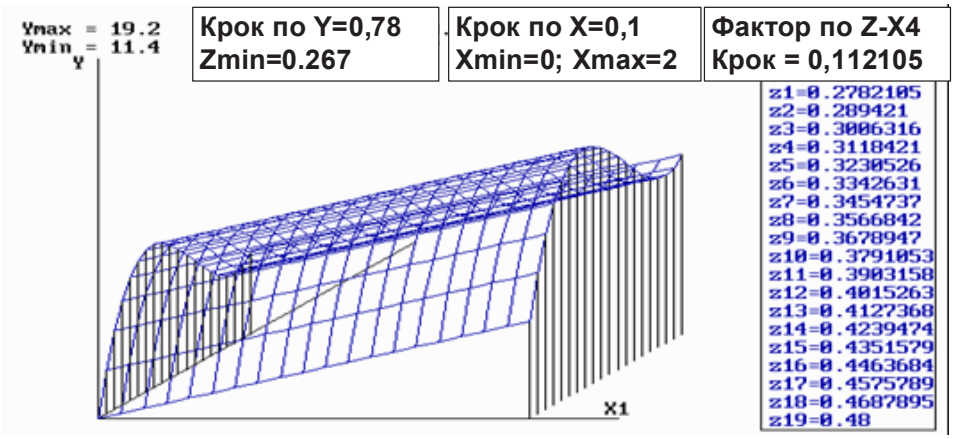


Рис. 4. Поверхні відгуку $Y_2 = f(X_1, X_4)$ в тривимірному просторі

Результати статистичного аналізу дозволяють зробити висновок, що регресійні залежності адекватні та інформативні, результати відтворені, коефіцієнти b_0 і b_i статистично значимі. Усе це говорить про те, що регресійні залежності адекватно описують процес і їх можна використовувати для аналізу і оптимізації процесу управління мікрокліматом теплиці.

Побудовані в роботі моделі підтверджують можливість оптимізації параметрів мікроклімату теплиці та визначення таких їх значень, які забезпечують стабільний температурний режим відповідно до агротехнічних вимог. Статистичний аналіз засвідчив, що отримані залежності є адекватними, інформативними та відтворюваними. Це дає підстави використовувати їх для подальшого аналізу й оптимізації систем управління мікрокліматом у тепличних умовах.

Висновки. У процесі дослідження встановлено, що для ефективного управління мікрокліматом у теплиці доцільно здійснювати регулювання температури повітря за умови сталого підігріву ґрунту, підтримуючи його температуру на рівні, що відповідає агротехнічним вимогам. Аналіз динамічних характеристик теплиці аркового типу показав її високу інерційність, що зумовлює певне запізнення в реакції на зміну керуючих дій. Крім того, виявлено наявність невизначеності параметрів моделі об'єкта керування, що ускладнює застосування класичних регуляторів. У таких умовах найбільш ефективним підходом до автоматизації процесу є впровадження інтелектуальних адаптивних систем керування, зокрема побудованих з використанням нейронних мереж та нечіткої логіки. Це дозволяє забезпечити стабільне функціонування теплиці, зменшити енерговитрати та підтримувати оптимальний мікроклімат для вирощування овочевих культур.

Список літератури:

1. Лисенко В.П., Дудник А.О. Оптимальне управління: стан та перспективи розвитку в тепличній галузі. Науковий вісник Національного університету біоресурсів і природокористування України. 2011. Вип. 166. С. 53–56
2. Іноваційні технології в управлінні складними біотехнологічними об'єктами агропромислового комплексу / А.П.Ладанюк, В.М.Решетюк, В.Д.Кишенько, Я.В.Смітюх. Київ: Центр учбової літератури, 2014. 280 с.
3. Прокопенко Т. О., Березюк І. А., Зубенко В. О. Автоматизована система керування температурно–вологісним режимом теплиці на основі апарату нечіткої логіки. *Вчені записки Таврійського національного університету імені В. І. Вернадського. Серія: Технічні науки*. 2020. Т. 31 (70), № 6, ч. 1. С. 129–134.
4. Лисенко В. П., Болбот І. М., Лендєл Т. І., Чернов І. І. Програмно-апаратне забезпечення системи фітомоніторингу в теплиці. *Енергетика та комп›ютерно-інтегровані технології в АПК*. 2018. № 2. С. 65.
5. Лайко Б., Цопа О. Інформаційна система керування мікрокліматом розумної теплиці. *V International Scientific and Theoretical Conference «The driving force of science and trends in its development:* матеріали наук. – прак. конф. 22 грудня 2023 р. Ковентрі, Велика Британія : SCIENTIA. 2023. С. 125–128.
6. Діордієв В. Т., Кашкарьов А. О., Діордієв О. О. Автоматизована система моніторингу та керування мікрокліматом у теплиці. *Науковий вісник Таврійського державного агротехнологічного університету*. 2018. Вип. 8, т. 2. С. 1–9
7. Васильєва Ю. Д., Морозова М. М. Нейромережа для контролю параметрів мікроклімату теплиці. *Погляд у майбутнє приладобудування*: зб. праць XII Всеукр. наук.-практ. конф. студентів, аспірантів та молодих вчених, м. Київ, 2019р. Київ.: ПБФ, КПІ ім. Ігоря Сікорського, Центр учбової літератури, 2019. С. 429–432.
8. Євсеєнко О. М. Розробка апаратно-програмної системи керування мікрокліматом теплиці. *Технічна інженерія*. 2020. № 1 (85). С. 104–109.
9. Гайдукевич С. В., Семенова Н. П., Леськів Я. А. Підвищення ефективності в системах керування мікрокліматичними параметрами приміщень закритого ґрунту. *Вчені записки ТНУ імені В.І. Вернадського. Серія: технічні науки*. 2020. № 6. С. 58–64.
10. Hoque J., Ahmed R., Hannan S. An Automated Greenhouse Monitoring and Controlling System using Sensors and Solar Power. *European Journal of Engineering Research and Science*. 2020. Vol. 5 (4). P. 510–515.
11. Лактіонов І. С., Вовна А. В., Бережной М. О., Лебедєв В. А. Комп'ютеризована система комплексного моніторингу і управління мікрокліматом промислових теплиць на базі нечіткої логіки. *Вісник КрНУ імені Михайла Остроградського*. 2019. № 3 (116). С. 120–129.
12. Maher A., Kamel E., Enrico F., Atif I., Abdelkader M. An intelligent system for the climate control and energy savings in agricultural greenhouses. *Energy Efficiency*. 2016. Vol. 9. P. 1241–1255.
13. Канарський Є., Орєхов О., Желтухін О. Розробка автоматизованої системи керування тепличним господарством. *Молодий вчений*. 2022. № 3 (103). С. 5–12.

Bereziuk I.A., Zubenko V.O. OPTIMIZATION OF THE GREENHOUSE MICROCLIMATE PARAMETERS CONTROL PROCESS USING MATHEMATICAL MODELING

The article presents the results of a comprehensive study on the optimization of microclimate parameters in greenhouse structures. The primary focus is on analyzing the influence of structural, technological, and operational factors on the temperature regime, which is critical for ensuring the normal growth, development, and productivity of vegetable crops. Key variables that determine the dynamics of microclimatic conditions inside the greenhouse were identified, including air mass, soil mass, the power of air and soil heaters, and heating time.

A justified experimental design was developed using the method of mathematical planning of experiments, covering a wide range of controllable parameter variations. A series of experimental tests was conducted, and the results were used to construct empirical regression models describing the dependence of air and soil temperature on the influencing factors. The models were statistically verified for adequacy, reproducibility, and informativeness, confirming their reliability for further applied calculations and implementation.

An expert evaluation method was applied to determine the weighting coefficients of each optimization criterion, which made it possible to form a generalized efficiency indicator for system performance. The feasibility of transitioning from traditional temperature regulation based solely on external conditions

to a flexible, intelligent control approach that accounts for a wide range of internal and external factors was substantiated. A promising direction is the implementation of adaptive control systems using artificial intelligence methods – namely neural networks and fuzzy logic.

This approach enables efficient distribution of energy resources, reduces fuel consumption, improves the stability of the microclimate according to agrotechnical requirements, and increases crop yields. The developed neuro-computer system opens new opportunities for automation and intelligent control of modern greenhouse complexes, contributing to improved energy efficiency, economic viability, and technological flexibility in agricultural production.

Key words: *greenhouse, microclimate, temperature control, mathematical modeling, regression analysis, multifactor optimization, neural networks, fuzzy logic, adaptive control, energy efficiency, automation, agrotechnologies.*

УДК 681.5:004.94:621.391:638.1
DOI https://doi.org/10.32782/2663-5941/2025.3.2/29

Іванишин В.В.
Національний авіаційний університет

Мошенский А.О.
Таврійський національний університет імені В.І. Вернадського

АПАРАТНЕ ЗАБЕЗПЕЧЕННЯ БІО-ТЕХНІЧНОЇ СИСТЕМИ МОНІТОРИНГУ І КЕРУВАННЯ ПАСІКОЮ. МІКРОПРОЦЕСОРНІ ПЛАТФОРМИ

Автоматизація процесів у бджільництві набуває особливої актуальності в умовах зростання потреб у точному моніторингу стану бджолосімей. Однією з основних проблем залишається вибір енергоефективної та доступної мікропроцесорної платформи для реалізації систем збирання даних з пасіки у режимі реального часу. У статті проведено аналіз сучасних мікроконтролерів, що застосовуються в біо-технічних системах, зокрема ESP32, ESP8266, STM32L4, NRF52840 та інших. Особливу увагу приділено критеріям енергоспоживання, вартості, продуктивності та можливості автономної роботи.

Основним об›єктом дослідження є мікроконтролер ESP32, який завдяки підтримці Wi-Fi, BLE та глибокого сну (deep sleep) з мінімальним енергоспоживанням до 5 μA, визнано найбільш придатним для створення базових вузлів системи моніторингу. Завдяки режимам пробудження від таймерів і зовнішніх сенсорів, а також можливості OTA-оновлень, ESP32 дозволяє створити автономну систему, яка може працювати місяцями без перезарядки. У роботі також розглянуто варіанти платформ для вузлів з підвищеними вимогами до автономності, зокрема на основі STM32L4 і NRF52840, та для високопродуктивних задач – Raspberry Pi Zero W 2.

Запропоновані платформи підтримують підключення сенсорів температури, вологості, ваги та звуку, що забезпечує комплексний моніторинг стану вуликів. Дані можуть передаватися через Wi-Fi або LoRa-модулі, залежно від умов розміщення пасіки. Порівняльний аналіз апаратних рішень супроводжено графіками ефективності, що враховують співвідношення потужності до вартості та енергоспоживання, що дозволяє обґрунтовано підібрати платформу для конкретного завдання.

Запропоноване рішення може бути основою для створення масштабованих і доступних автоматизованих IoT-систем в галузі сільського господарства. Впровадження такої системи на пасіках дозволить зменшити витрати на технічне обслуговування, підвищити точність контролю та сприятиме ефективнішому управлінню бджолиними господарствами, що, у свою чергу, позитивно позначиться на екосистемі через поліпшення запилення ентомофільних культур.

Ключові слова: *автоматизація, ESP32, моніторинг бджолосімей, мікроконтролери, енергозбереження, LoRa, Wi-Fi, сенсори, автономні системи, інтернет речей (IoT), аграрні технології.*

Постановка проблеми. Автоматизація бджільництва є актуальним напрямом розвитку аграрних біо-технічних систем. Необхідність постійного моніторингу стану бджолосімей, раннього виявлення загроз (захворювань, голодування, роїння) та підвищення ефективності господарювання обумовлює потребу у надійних, енергоефективних та автономних системах збирання та передавання даних.

Фізична частина біо-технічної системи складається з сенсорів (температури, вологості, ваги, звуку), модулів зв'язку (Wi-Fi, LoRa), блока живлення (акумулятор, сонячна панель) і центрального мікроконтролера [1].

Основні вимоги до системи складають:

- ✓ енергоспоживання у режимі сну <10 μA;
- ✓ робоча напруга до 5 В;
- ✓ підтримка OTA;
- ✓ підтримка Wi-Fi або LoRa;
- ✓ компактність;
- ✓ стабільність у температурному діапазоні -30…+60 °C.

Центральною складовою таких систем виступає мікропроцесорна платформа, що забезпечує керування сенсорами, збір, обробку і трансляцію даних у реальному часі.

На ринку доступні десятки апаратних рішень з різними характеристиками, тому критичним

є обґрунтований вибір платформи, яка забезпечить довготривалу автономну роботу в умовах обмеженого живлення, підтримку бездротових інтерфейсів і належну обчислювальну потужність.

Загальні критерії вибору мікропроцесорної платформи:

✓ споживання енергії в різних режимах (активний / sleep / deep sleep);

✓ кількість портів вводу/виводу;

✓ обсяг RAM і Flash;

наявність бездротових інтерфейсів (Wi-Fi, BLE, LoRa, Zigbee);

✓ доступність і підтримка програмного забезпечення;

✓ вартість і габарити плати.

Аналіз останніх досліджень і публікацій. На сучасному етапі розвитку мікропроцесорних технологій існує багато різноманітних одноплатних інтегрованих систем. Нижче у Таблиці 1 наведено характеристики найпоширеніших платформ, які застосовуються в системах IoT, включаючи бджолині пасіки. Основна увага приділяється споживанню енергії, вартості, доступності бездротових інтерфейсів та підтримці автономної роботи.

Постановка завдання. Метою цієї роботи є систематизація доступних апаратних платформ, їх порівняльний аналіз за техніко-експлуатаційними показниками та обґрунтування вибору оптимального рішення для розробки автоматизованої систем моніторингу стану бджолосімей.

На основі поставлених критеріїв та сучасного стану розвитку мікропроцесорних інтегрованих одноплатних систем провести порівняльній аналіз за трьома основними параметрами: енергоефективність, ціна та обчислювальна потужність та обґрунтувати вибір оптимальної платформи для автоматизованої біо-технічної системи моніторингу та керування пасікою.

Виклад основного матеріалу. Сьогодні існує широкий спектр технічних рішень для моніторингу стану бджолосімей, що ґрунтуються на використанні бездротових сенсорів, мікроконтролерів та технологій передачі даних у реальному часі. Проте більшість із них мають низку суттєвих обмежень: високе енергоспоживання, обмежений час автономної роботи або надто високу вартість, що унеможливлює їх масове впровадження в малих пасічних господарствах. У даному дослідженні аналізується апаратна частина інноваційної системи моніторингу, яка базується на використанні енергоефективних мікропроцесорних платформ з підтримкою бездротових протоколів зв'язку (Wi-Fi, BLE, LoRa). Основну увагу приділено порівнянню таких платформ за критеріями енергоспоживання, обчислювальної потужності та вартості з метою вибору оптимального рішення для автономної роботи в умовах пасіки.

Аналіз актуальних платформ:

1. ESP8266 (ESP-12F) [2]

Рік випуску: 2014

Виробник: Espressif Systems (Китай)

ESP8266 – один із перших наддешевих Wi-Fi мікроконтролерів, що спричинив революцію в IoT-пристроях. Початково був розроблений як Wi-Fi модуль для Arduino, але згодом отримав повноцінне SDK. Версія ESP-12F має вбудовану антену, 12 GPIO, 4 Мб флеш-пам'яті та підтримує глибокий сон із струмом <10 μA. Підтримує OTA-прошивку, що дозволяє дистанційне оновлення прошивки.

Таблиця 1

Порівняння мікропроцесорних платформ

№	Платформа	Wi-Fi/ LorA/ BT	Напруга живлення, В.	Споживання Deep Sleep, μA.	Ціна, грн.	Особливості
1	**ESP8266 (ESP-12F)**	Wi-Fi	3.3	~ 10	75	Дешева, проста
2	**ESP32**	Wi-Fi + BLE	3.3	~ 5	150	Двоядерний, OTA
3	**Raspberry Pi Zero W 2**	Wi-Fi + BLE	5	~ 80	600	Linux, Python
4	**STM32L476RG**	–	1.8–3.6	< 1	400	Споживання, промисловий
5	**NRF52840**	BLE, Zigbee, Thread	1.7–3.6	< 1	350	Багатопротокольний
6	**Seeeduino XIAO**	–	3.3	~ 2	180	ARM Cortex, компактний
7	**Arduino Portenta H7**	Wi-Fi + BT	3.3–5	~ 20	1900	Машинне навчання, двоядерний

2. ESP32 [3]

Рік випуску: 2016

Виробник: Espressif Systems (Китай)

ESP32 – наступник ESP8266, має двоядерний 32-бітний Tensilica Xtensa LX6 процесор, 520 КБ SRAM, Wi-Fi і BLE модулі, апаратне шифрування, 34 GPIO. Підтримує розширені енергозберігаючі режими та OTA. Платформа активно розвивається: існують варіації з підтримкою LoRa, камер, сенсорів тощо.

3. Raspberry Pi Zero W 2 [4]

Рік випуску: 2021

Виробник: Raspberry Pi Foundation (Велика Британія)

Одноплатний комп'ютер, що працює під управлінням Linux. Містить чотирьох-ядерний ARM Cortex-A53, 512 МБ RAM, Wi-Fi, Bluetooth, HDMI, USB і GPIO. Призначений для задач з високими вимогами до обчислювальної потужності – локальна аналітика, обробка звуку або зображень, зв'язок з хмарними сервісами.

4. STM32L4 (STM32L476RG) [5]

Рік випуску: 2016

Виробник: STMicroelectronics (Швейцарія)

STM32L4 – серія ARM Cortex-M4 мікроконтролерів із наднизьким енергоспоживанням. Має до 1 МБ флеш-пам'яті, 128 КБ RAM, підтримку режимів STOP та STANDBY з енергоспоживанням <1 μA. Часто застосовується у медичних приладах, датчиках, wearable-пристроях.

5. NRF52840 [6]

Рік випуску: 2018

Виробник: Nordic Semiconductor (Норвегія)

Орієнтований на IoT з використанням BLE, Thread, Zigbee, ANT. Має ARM Cortex-M4, 1 МБ флеш-пам'яті, 256 КБ RAM, повну підтримку Bluetooth 5.0. Підтримує режим «System Off» із <0.3 μA споживанням.

6. Seeeduino XIAO [7]

Рік випуску: 2020

Виробник: Seeed Studio (Китай)

Компактна плата на базі ARM Cortex-M0+ (SAMD21G18), розміром лише 20×17.5 мм. Має 256 КБ флеш-пам'яті, 32 КБ SRAM. Не має вбудованого бездротового зв'язку, але може бути доповнена зовнішніми модулями.

7. Arduino Portenta H7 [8]

Рік випуску: 2020

Виробник: Arduino SA (Італія)

Високопродуктивна плата із двоядерним процесором (ARM Cortex-M7 + M4), Wi-Fi, Bluetooth, підтримкою TensorFlow Lite. Розрахована на складні проекти у сфері edge computing та штучного інтелекту. Підтримує USB-C, камери, дисплеї.

В результаті порівняння сучасних мікропроцесорних платформ (Таблиця 1) було виявлено суттєві відмінності у вартості, енергоспоживанні та обчислювальних можливостях актуальних модулів. Щоб оцінити доцільність використання кожної з платформ у системах моніторингу пасіки, проведено аналіз ключових параметрів.

Насамперед важливо враховувати енергоспоживання в режимі сну, оскільки саме цей режим є домінуючим у роботі сенсорних вузлів. На Рисунку 1 представлено залежність між вартістю платформи (у гривнях) та її енергоспоживанням у режимі глибокого сну (μA). Графік подано у логарифмічній шкалі, що дозволяє виявити оптимальні платформи при значних відмінностях у параметрах.

Найбільш збалансованими за співвідношенням «ціна – споживання» виявилися ESP32 та ESP8266. Перша платформа забезпечує хорошу продуктивність при низькому енергоспоживанні, тоді як друга – є найбільш доступною. Натомість Raspberry Pi Zero W 2 і Portenta H7 демонструють значне енергоспоживання, що ускладнює їх використання у повністю автономних системах без зовнішнього живлення.

Оцінка обчислювальної потужності мікропроцесорних платформ є важливим етапом у виборі апаратної бази для систем моніторингу, оскільки саме цей параметр визначає можливість виконання задач попередньої обробки даних, фільтрації сигналів та управління бездротовими модулями у режимі реального часу. З огляду на суттєву відмінність архітектур і підходів до їх програмування у цьому дослідженні для порівняння застосовано метод нормованої умовної продуктивності, яка враховує три ключові фактори: тактову частоту процесора, кількість ядер та ефективність виконання інструкцій (IPC – instructions per cycle).

Цей метод дозволяє одержати відносні значення, придатні для графічного та табличного порівняння, навіть якщо виробники не надають стандартні показники продуктивності у вигляді MIPS чи CoreMark.

Значення IPC було визначено на основі технічної документації та результатів незалежних тестувань:

для Cortex-M0+ (Seeeduino XIAO) типове IPC ≈ 0.9 [9];

для Cortex-M4 (STM32L4, NRF52840) – IPC ≈ 1.0 [10];

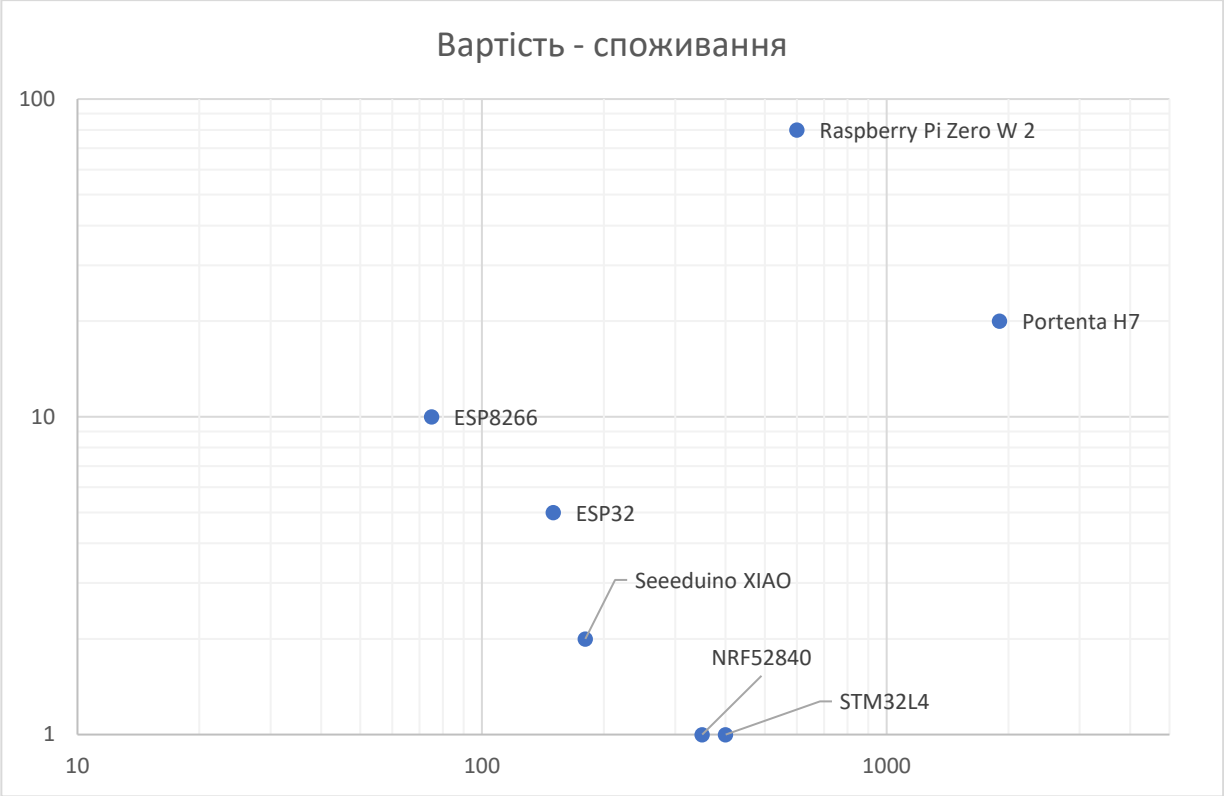


Рис. 1. Споживання платформ в режимі DeepSleep

для Xtensa LX6 (ESP32) – IPC ≈ 1.1–1.2 [3];

для Cortex-A53 (Raspberry Pi Zero W 2) – IPC може досягати 2.0 при роботі з Linux і багатозадачністю [4].

Обчислювальна потужність P мікроконтролера в умовних одиницях визначається за формулою (1):

$$P = \frac{f \cdot n \cdot IPC}{K} \quad (1)$$

де f – тактова частота ядра, МГц;

n – кількість фізичних ядер;

IPC – кількість інструкцій, які процесор виконує за один такт (Instructions Per Cycle), що залежить від архітектури ядра;

K – коефіцієнт нормалізації (у цьому дослідженні $K = 10$).

Такий метод не є заміною промислових методів тестування (напр., CoreMark), але дозволяє швидко оцінити доцільність використання конкретної платформи в задачах низькорівневої обробки сигналів у системах моніторингу.

Результати розрахунку нормованої умовної продуктивності приведено у Таблиці 2.

Рисунок 2 ілюструє, як змінюється обчислювальна потужність із ростом вартості.

Лідером за обчислювальною потужністю є Raspberry Pi Zero W 2, який підтримує багатозадачність, обробку зображень, запуск Python-скриптів і роботу під повноцінною ОС. Проте його високе споживання енергії суттєво обмежує автономне застосування. Водночас ESP32 забезпечує достатню потужність для більшості задач моніторингу (обробка сенсорних даних, попередня фільтрація, передача), маючи при цьому значно нижче енергоспоживання. Платформи STM32L4, NRF52840 та Seeeduino XIAO–менш потужні, але підходять для виконання вузькоспеціалізованих, низькорівневих завдань.

Формалізоване представлення інтегральної ефективності платформи може бути розрахована за загальною формулою (2) з використанням коефіцієнтів важливості:

$$\eta = \frac{P}{k_E \cdot E \cdot k_C \cdot C} \quad (2)$$

де η – інтегральна ефективність платформи (безрозмірна величина);

P – обчислювальна потужність (умовні одиниці, наприклад MIPS або бал);

E – енергоспоживання в режимі сну (в μA);

C – вартість пристрою (у гривнях);

k_E – ваговий коефіцієнт важливості енергоефективності;

k_C – ваговий коефіцієнт важливості вартості.

Таблиця 2

Нормована умовна продуктивність платформ

№	Платформа	Архітектура	Частота, МГц	Кількість ядер	IPC	Умовна потужність, бали
1	ESP8266 (ESP-12F)	Xtensa LX106	80	1	0,9	7,20
2	ESP32	Xtensa LX6	240	2	1,2	57,60
3	Raspberry Pi Zero W 2	Cortex-A53	1000	4	2,0	1000,00
4	STM32L476RG	Cortex-M4	80	1	1,0	8,00
5	NRF52840	Cortex-M4	80	1	1,0	6,40
6	Seeeduino XIAO	Cortex-M0+	48	1	0,9	4,32
7	Arduino Portenta H7	Cortex-M7+M4	480	1	1,2	115,20

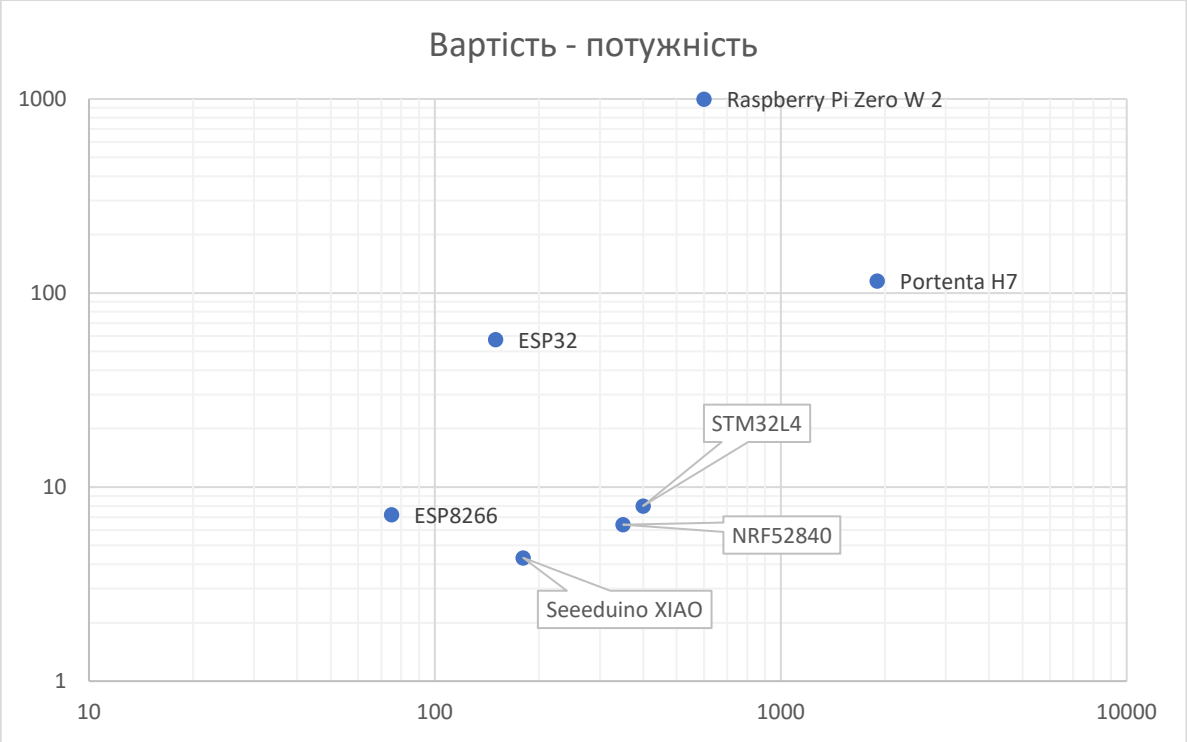


Рис. 2. Зміна обчислювальної потужності з ростом вартості платформ

Або в спрощеному варіанті, коли усі коефіцієнти дорівнюють 1, тобто однаково важливі – формула (3):

$$\eta = \frac{P}{E \cdot C} \qquad (3)$$

Такий підхід дозволяє порівнювати платформи з одночасним урахуванням обчислювального ресурсу, енергоспоживання та ціни. Результати наведено на Рисунку 3.

Оптимальний індекс ефективності показала ESP32, яка впевнено випереджає інші платформи з врахуванням комунікаційних можливостей та простоти інтеграції. Вона забезпечує достатній рівень продуктивності для більшості задач системи моніторингу пасіки, підтримує бездротові протоколи зв'язку (Wi-Fi, BLE), має низьке енергоспоживання та адекватну вартість. ESP8266 також показує високий результат, що робить її придатною для простих та надбюджетних рішень. Інші платформи виявилися менш ефективними через або високу ціну (Portenta H7), або обмежену функціональність при потребі зовнішніх модулів (Seeeduino XIAO, STM32L4).

Висновки. На основі розробленої методики нормованої умовної продуктивності, що враховує тактову частоту, кількість ядер та IPC-потенціал архітектури, проведено порівняльний аналіз мікропроцесорних платформ для системи моніторингу пасіки.

Проведено оцінку мікропроцесорних платформ для автономних систем моніторингу пасіки

Рис. 3. Інтегральна ефективність платформ, умов. од.

з використанням запропонованої методики. До розрахунку включено три ключові параметри: обчислювальну потужність, енергоспоживання в режимі сну та вартість обладнання. Це дозволило здійснити комплексне порівняння платформ за інтегральним показником ефективності.

Найбільш збалансованим рішенням виявилася платформа ESP32, яка поєднує достатню обчислювальну потужність, низьке енергоспоживання та доступну вартість. Її використання доцільне як у ролі сенсорного вузла, так і як вузла збору, обробки й передавання даних.

Для вузлів із критичними вимогами до автономної роботи рекомендовано STM32L4 або NRF52840, що демонструють наднизьке споживання енергії при збереженні базової функціональності.

Raspberry Pi Zero W 2 та Arduino Portenta H7 доцільно застосовувати лише у випадках, де доступне живлення та необхідна висока продуктивність, зокрема для обробки аудіосигналів або застосування алгоритмів штучного інтелекту.

Найбільш ефективною стратегією впровадження може стати гібридна архітектура, що поєднує ESP32 як універсальну платформу, STM32/NRF – для енергонезалежних сенсорів, та Raspberry Pi – для агрегації даних і хмарної інтеграції.

Запропонований підхід забезпечує технічну доцільність, економічну ефективність і високу адаптивність до умов польового застосування, що відкриває перспективу його широкого впровадження в аграрному секторі України, зокрема у сфері автоматизації бджільництва.

Список літератури:

1. Ivanyshyn V. V., Moshensky A. O. The concept of an automated apiary monitoring system based on lora technology using real-time sensors. Scientific notes of taurida national V.I. vernadsky university. series: technical sciences. 2024. T. 1, № 5. С. 149–153. URL: https://doi.org/10.32782/2663-5941/2024.5.1/23 (дата звернення: 15.04.2025).
2. Espressif Systems. ESP8266EX Datasheet. Wireless SoCs, Software, Cloud and AIoT Solutions | Espressif Systems. URL: https://www.espressif.com/sites/default/files/documentation/0a-esp8266ex_datasheet_en.pdf (дата звернення: 15.04.2025).
3. Espressif Systems. ESP32 Series Datasheet. Wireless SoCs, Software, Cloud and AIoT Solutions | Espressif Systems. URL: https://surl.lu/lckimq (дата звернення: 15.04.2025).
4. Raspberry Pi Zero 2 W. Raspberry Pi Datasheets. URL: https://surl.li/dmfoce (дата звернення: 15.04.2025).
5. STM32L4 series of ultra-low-power MCUs. URL: https://surl.lu/hcoytm (дата звернення: 15.04.2025).

6. nRF52840 Product Specification v1.11. URL: https://surl.li/rjhmmw (дата звернення: 15.04.2025).
7. Getting Started with Seeed Studio XIAO SAMD21 | Seeed Studio Wiki. Seeed Studio Wiki. URL: https://wiki.seeedstudio.com/Seeeduino-XIAO/ (дата звернення: 15.04.2025).
8. Arduino SA. Portenta H7 – Technical Documentation. URL: https://docs.arduino.cc/hardware/portenta-h7/ (дата звернення: 15.04.2025).
9. Cortex-M0+ Technical Reference Manual. Arm Developer. URL: https://developer.arm.com/documentation/ddi0484/latest (дата звернення: 15.04.2025).
10. Cortex-M4 Processor Technical Reference Manual. Arm Developer. URL: https://developer.arm.com/documentation/ddi0439/latest (дата звернення: 15.04.2025).
11. A Lora-based Testbed Development for Stingless Bee Monitoring System / M. Taib Miskon та ін. TEM Journal. 2022. С. 712–718. URL: https://doi.org/10.18421/tem112-26 (дата звернення: 15.04.2025).
12. IoT and Machine Learning Techniques for Precision Beekeeping. MDPI. 2024. URL: https://www.mdpi.com/2673-2688/6/2/26 (дата звернення: 15.04.2025).
13. BHiveSense: an integrated information system architecture for sustainable remote monitoring and management of apiaries based on IoT and microservices / D. Cota et al. Journal of Open Innovation: Technology, Market, and Complexity. 2023. P. 100110. URL: https://doi.org/10.1016/j.joitmc.2023.100110 (дата звернення: 15.04.2025).
14. Energy consumption in ESP32. Luis Llamas. URL: https://www.luisllamas.es/en/esp32-power-consumption/ (дата звернення: 15.04.2025).
15. ESP32 Active Mode and Deep Sleep Mode Power Consumption Comparison. Circuit Digest – Electronics Engineering News, Latest Products, Articles and Projects. URL: https://surl.li/sxyadq (дата звернення: 15.04.2025).
16. Ntawuzumunsi E., Kumaran S., Sibomana L. Self-Powered Smart Beehive Monitoring and Control System (SBMaCS). Sensors. 2021. Vol. 21, no. 10. P. 3522. URL: https://doi.org/10.3390/s21103522 (дата звернення: 15.04.2025).

Ivanyshyn V.V., Moshensky A.O. HARDWARE FOR THE BIO-TECHNICAL SYSTEM FOR MONITORING AND CONTROLLING THE APIARY. MICROPROCESSOR PLATFORMS

Automation of processes in beekeeping is becoming especially relevant in the context of the growing need for accurate monitoring of bee colonies. One of the main problems is the choice of an energy-efficient and affordable microprocessor platform for the implementation of real-time data collection systems from the apiary. The article analyzes modern microcontrollers used in bio-technical systems, in particular ESP32, ESP8266, STM32L4, NRF52840 and others. Particular attention is paid to the criteria of energy consumption, cost, performance, and the possibility of autonomous operation in remote conditions.

The main object of the study is the ESP32 microcontroller, which, due to its support for Wi-Fi, BLE and deep sleep with a minimum power consumption of up to 5 μA, is recognized as the most suitable for creating basic nodes of the monitoring system. Thanks to the wake-up modes from timers and external sensors, as well as the possibility of OTA updates, ESP32 allows you to create an autonomous system that can work for months without recharging. The paper also considers platform options for nodes with increased requirements for autonomy, in particular, based on STM32L4 and NRF52840, and for high-performance tasks – Raspberry Pi Zero W2.

The proposed platforms support the connection of temperature, humidity, weight, and sound sensors, which provides comprehensive monitoring of the hives. The data can be transmitted via Wi-Fi or LoRa modules, depending on the location of the apiary. The comparative analysis of hardware solutions is accompanied by efficiency graphs that take into account the power to cost ratio and energy consumption, which allows you to reasonably select a platform for a specific task.

The proposed solution can be the basis for the creation of scalable and affordable automated IoT systems in the field of agriculture. The introduction of such a system in apiaries will reduce maintenance costs, increase the accuracy of control, and contribute to more efficient management of bee farms, which, in turn, will have a positive impact on the ecosystem by improving pollination of entomophilous crops.

Key words: *automation, ESP32, bee colony monitoring, microcontrollers, energy saving, LoRa, Wi-Fi, sensors, autonomous systems, Internet of Things (IoT), agricultural technologies.*

УДК 004.738.5:621.391.8:681.586
DOI https://doi.org/10.32782/2663-5941/2025.3.2/30

Карпенко М.І.
Національний університет харчових технологій

Мошенський А.О.
Національний університет харчових технологій

КОНСТРУЮВАННЯ ТА ПОЛЬОВА АПРОБАЦІЯ ПРИСТРОЮ ДЛЯ ДОСЛІДЖЕННЯ ПОКРИТТЯ БСМ МЕРЕЖ

У статті висвітлено процес проєктування, збирання та практичного застосування експериментального пристрою для дослідження просторових характеристик покриття безпровідної сенсорної мережі (БСМ), що функціонує на базі технології LoRa. Актуальність теми обумовлена зростаючою потребою у побудові високонадійних мереж передавання даних у складних умовах – зокрема в урбанізованих районах, районах з підвищеним рівнем перешкод та у прифронтових зонах, де традиційні засоби зв'язку демонструють нестабільність. Основною метою дослідження є створення мобільного технічного рішення, що дозволяє не лише теоретично, а й практично оцінювати ефективність покриття, залежно від параметрів середовища. Враховуючи критичну важливість просторового охоплення мережі, особливо у складних умовах – урбанізованих та прифронтових зонах – було розроблено компактний польовий інструмент збору даних з урахуванням реального середовища. До складу пристрою увійшли модуль E220-900T30D, GPS-приймач GP-01Kit та ноутбук із ПЗ u-center для синхронізації координат й часу. Особливу увагу приділено напрямленій антені типу Мохоп, яка була змодельована, побудована та протестована. Для електромагнітного моделювання використано MMANA-GAL, а для прогнозування зони покриття – Tower Coverage. В ході дослідження було проведено серію випробувань в умовах міста Київ, де аналіз втрати пакетів та стабільності сигналу продемонстрував значну залежність якості зв'язку від типу забудови, наявності перешкод та особливостей рельєфу. Зібраний пристрій дозволив не лише підтвердити гіпотези щодо впливу факторів навколишнього середовища на зв'язок, але й виявити обмеження існуючих розрахункових моделей покриття. Висновком є необхідність поєднання теоретичного моделювання з реальними польовими вимірюваннями під час проектування БСМ для забезпечення надійної роботи в специфічних умовах.

Ключові слова: *безпровідна сенсорна мережа, радіомодуль, антена Мохоп, LoRa, GPS.*

Постановка проблеми. Одним з критично важливих пунктів проектування безпровідної сенсорної мережі (БСМ) є розрахунок її покриття. Розрахунок може проводитися різними способами, однак лише перевірка даних на практиці може виявити непередбачені при розрахунках змінні, наприклад, природні чи техногенні елементи місцевості. У прифронтових або урбанізованих зонах це питання набуває особливої актуальності.

Проєктування БСМ зазвичай ґрунтується на попередньому розрахунку покриття із використанням аналітичних моделей або симуляцій. Проте в реальних умовах – особливо в міській забудові, в районах зі складним рельєфом або у прифронтових зонах – розрахункові моделі не враховують усіх змінних, таких як електромагнітні завади, несподівані техногенні або природні об›єкти, перешкоди прямої видимості тощо. Це може призвести до помилок у визначенні ефективної зони покриття мережі, втрати зв'язку та зниження ефективності передачі критично важливої інформації, зокрема в задачах екологічного моніторингу або виявлення небезпечних речовин. Таким чином, виникає потреба у створенні мобільного, надійного і технологічно простого пристрою для збору просторових характеристик сигналу безпосередньо в польових умовах, що дозволить скоригувати або доповнити розрахункові моделі на основі реальних даних.

Аналіз останніх досліджень і публікацій. У статті Х. Х. Лопеса-Ескобара (J. J. López-Escobar) та ін. [2] особливу увагу приділено відстеженню популяції морських птахів. Принцип роботи реалізованого продукту базується на прогнозуванні місцеположення за допомогою нейронних мереж на основі останніх отриманих даних GNSS та доступних характеристик сигналу LoRa (RSSI та SNR). Автори змоделювали покриття на основі готових даних та використали

трикутник шлюзів для позначення зони покриття й отримання даних щодо позиціонування на основі RSSI та SNR. Стаття занадто ускладнена деталями, які необхідні для описаних автором умов, але які є зайвими у випадку визначення зони покриття.

Праця Бхану П. Р. Бхаванама (Bhanu Pratap Reddy Bhavanam) та ін. [3] показує дослідження покриття сенсорами вугільних шахт Індії. Були реалізовані мережі пристроїв ZigBee, LoRaWAN для вимірювання дальності покриття та позиціонування по RSSI. Проводилась ґрунтовна оцінка якості сигналу. У роботі використовувались стандартні антени на 868 МГц, пристрої були прості (приймач, передавач), однак з використанням можливостей мереж для визначення позиціонування. Даний спосіб краще підходить за рахунок малої кількості апаратури, але все одно надає більшу складність за рахунок конфігурації мережі для проведення досліду.

Праця Ґреґорі Дж. М. Маццаро (Gregory J. M. Mazzaro) [4] щодо побудови антен в домашніх умовах надала ідеї для конструкцій антен, в той час як стаття Марії Кампо-Валери (María Campo-Valera) та ін. [5] більш докладно описує процес розробки моноімпульсної антени для систем локалізації RFID, що може бути корисним для інтеграції в створення інформаційної системи.

Автори статті Цзячжі Танга (Jiazhi Tang) та ін. [6] розглядають статичні високоточні умови в контрольованому середовищі, що не підходить до реальних умов міського середовища, де метаповерхневі лінзи є недоцільними як через складність, так і через вартість та необхідність вузькоспеціалізованої апаратури.

У роботі [7] представлено розробку та побудову недорогого портативного наземного приймального комплексу для прийому зображень із супутників NOAA у діапазоні APT (~137,5 МГц). Автори показують, що для цього не обов'язково використовувати дорогі параболічні антени чи спеціалізовані приймачі – достатньо простого RTL-SDR та саморобної V-дипольної антени. Отримані сигнали оброблялись через SDRSharp та MATLAB, що дозволяло перетворювати аудіосигнал у високоякісні супутникові знімки. Хоча ця стаття демонструє ефективність дешевого підходу для супутникового моніторингу погоди, вона зосереджена на пасивному прийомі широкосмугових сигналів на фіксованій частоті. На відміну від цього, наше дослідження зосереджене на оцінці параметрів покриття в міських умовах з використанням вузькосмугового LoRa-зв'язку в діапазоні 868 МГц, що потребує інших вимог до антен, обробки даних та методології вимірювань.

У статті Андреаса Кріетемайєра (Andreas Krietemeyer) та ін. [8] запропоновано метод калібрування для досягнення високоточного рівня позиціонування за допомогою недорогих двочастотних GNSS-приймачів та антен. Автори розробили польове рішення, яке дозволяє значно покращити точність позиціонування, наближаючи її до рівня професійного обладнання. Хоча ця робота демонструє ефективність у підвищенні точності GNSS-вимірювань, вона зосереджена на статичних умовах та потребує специфічних калібрувальних процедур. У нашому дослідженні основна увага приділяється оцінці покриття бездротових сенсорних мереж у реальних міських умовах, де важливими є не лише точність позиціонування, але й стійкість зв›язку в умовах завад та перешкод.

У статті Тіма А. Вайлда (Timm A. Wild) та ін. [9] представлено розробку мікророзмірних, відкритих та недорогих GPS-логерів масою менше 1 г, призначених для моніторингу руху дрібних тварин. Автори зосереджуються на зменшенні впливу пристроїв на поведінку тварин, забезпечуючи при цьому високу точність позиціонування та тривалий час роботи. Хоча дана робота демонструє ефективність у збиранні даних про переміщення об'єктів, вона орієнтована на біологічні дослідження та не враховує специфіку міського середовища з його численними завадами для бездротового зв'язку.

Аналогічною є робота Махендри Свейна (Mahendra Swain) та ін. [10], які зосереджуються на оптимізації бюджету зв'язку (link budget), розглядаючи параметри, такі як коефіцієнт розширення (spreading factor) та чутливість приймача, з метою забезпечення ефективної передачі даних на великі відстані. Проте, результати статті орієнтовані на відкриті простори з мінімальними перешкодами для сигналу.

Постановка завдання. Створення портативного пристрою для збору просторових характеристик сигналу LoRa-модуля в різних умовах місцевості, а також експериментальна апробація цього пристрою в польових умовах.

Методологія. Було проведене експериментальне дослідження, в ході якого була зібрана та проаналізована інформація з різних літературних джерел та досліджень. Було проведено моделювання спрямованої антени типу Moxon, її виго-

товлення та подальше тестування на різних висотах і в різних локаціях: у міських умовах Київ та на відкритій місцевості. До складу пристрою увійшли модуль E220-900T30D, GPS-приймач GP-01Kit та ноутбук з програмою u-center для фіксації координат і часу передачі даних.

Виклад основного матеріалу. З огляду на властивість напрямлених антен мінімізувати розсіювання енергії передачі сигналу у визначеному секторі, що у свою чергу ускладнює його виявлення [1], було прийнято рішення використати для дослідів даний тип антен. Враховуючи задачі виявлення небезпечних речовин, варто було звернути увагу на дрони. В дронах популярні омні-антени за рахунок всенапрямленості. Однак, існує тип напрямлених антен, які частіше інших використовуються в FPV дронах – антени типу Moxon.

Для побудови антени було виконано низку кроків. По-перше, було використано Moxon Rectangle Generator як калькулятор розмірів антени (рис. 1) для швидкого отримання геометрії Moxon-антени для потрібної частоти (868,125 МГц).

По-друге, отриманий шаблон антени було імпортовано в програму для електромагнітного моделювання антен MMANA-GAL.

Після обробки параметрів, калібровки та інших процедур було проведено аналіз характеристик антени (діаграма спрямованості, коефіцієнт підсилення, резонансна частота) і обрано оптимальні параметри.

По-третє, засобами Tower Coverage було змодельовано зону покриття отриманої антени при вказаних умовах: розташування антени (на лівому березі Дніпра) на висоті 35 м (позиція антени), напрямок на правий берег Києва (рис. 3) та інших характеристик.

По-четверте, для перевірки теоретичних даних було вирішено зібрати радіомодуль з GPS-приймачем та отриманою антеною. З огляду на технічні характеристики модуля E220-900T30D було прийнято рішення використати його можливості для дослідів.

До RX та TX було під'єднано відповідні піни GPS для надсилання геолокації. Живлення відбувалось через павербанк. Для стабільної роботи пристрою був напаяний конденсатор. Зібрана модель передавача з від'єднаною антеною зображена на рис. 4.

Для приймача була виконана аналогічна конфігурація, без модуля GPS. Приймач було під'єднано через TTL-перехідник до ноутбука з програмою u-center.

Для роботи GP-01Kit було необхідне попереднє налаштування у вигляді позиціонування на

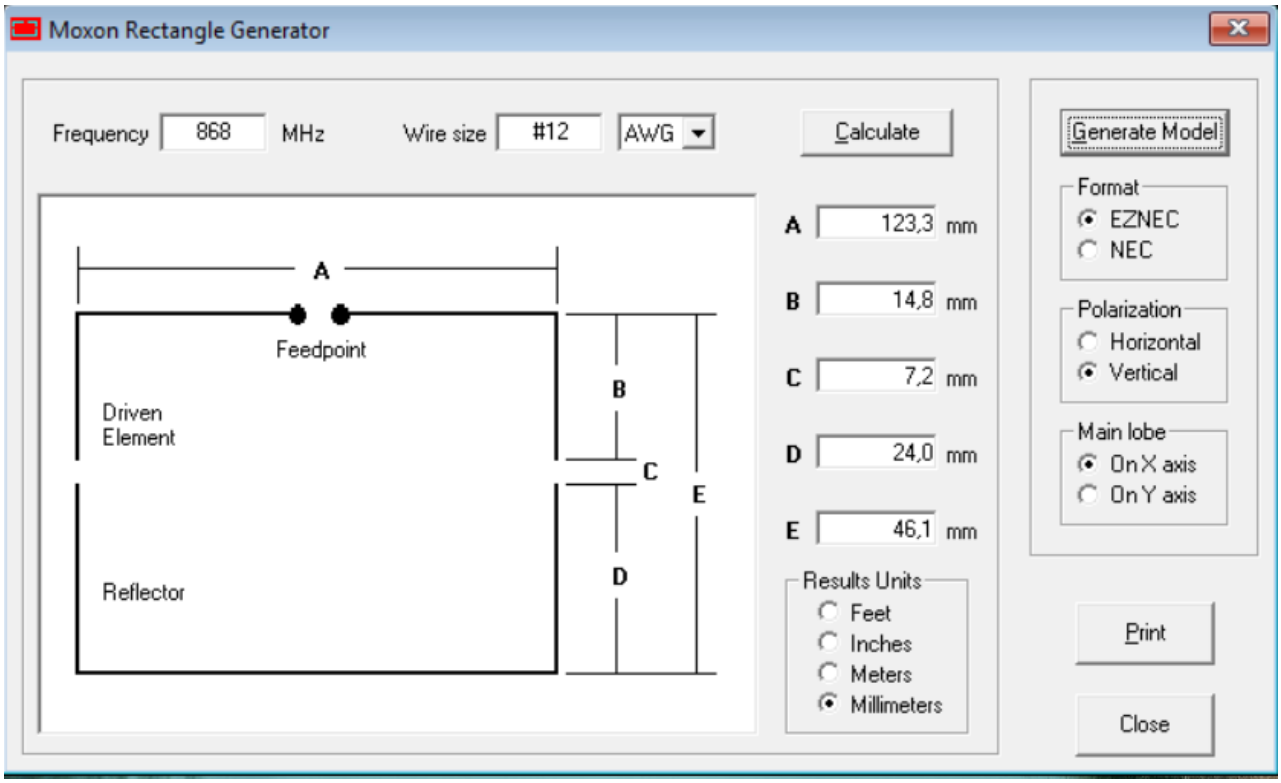


Рис. 1. Генератор геометрії антени Moxon

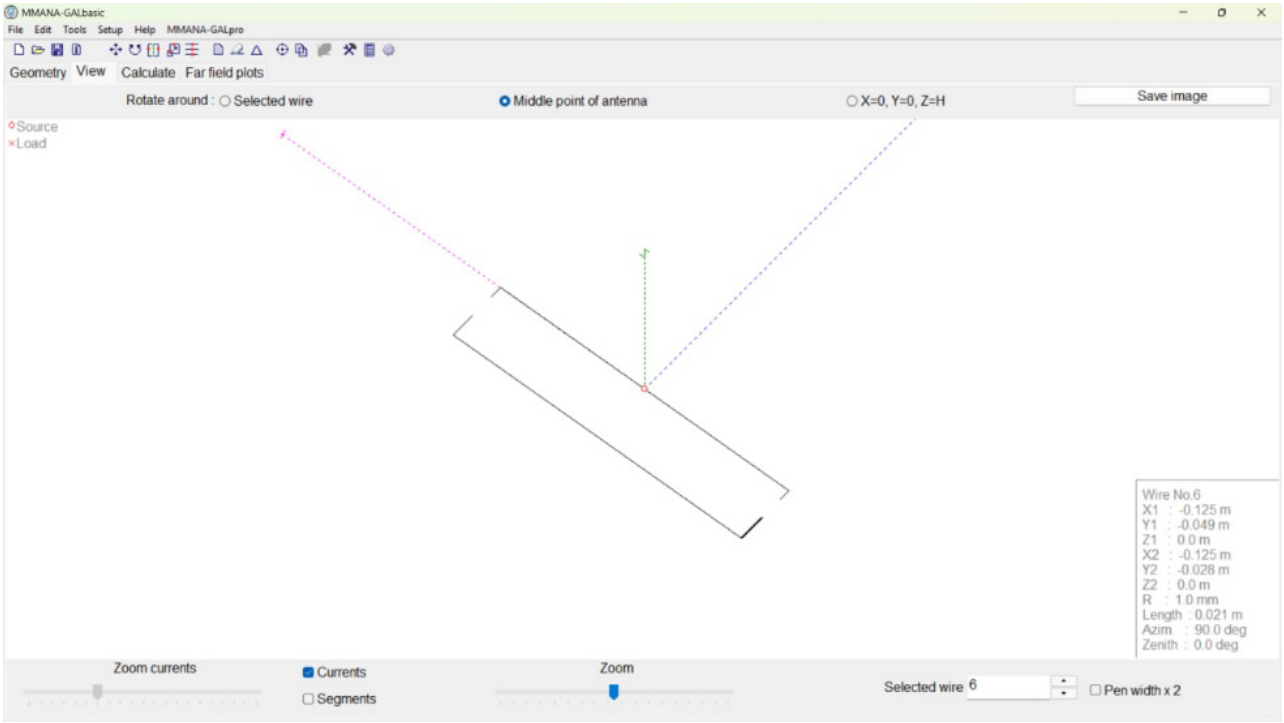
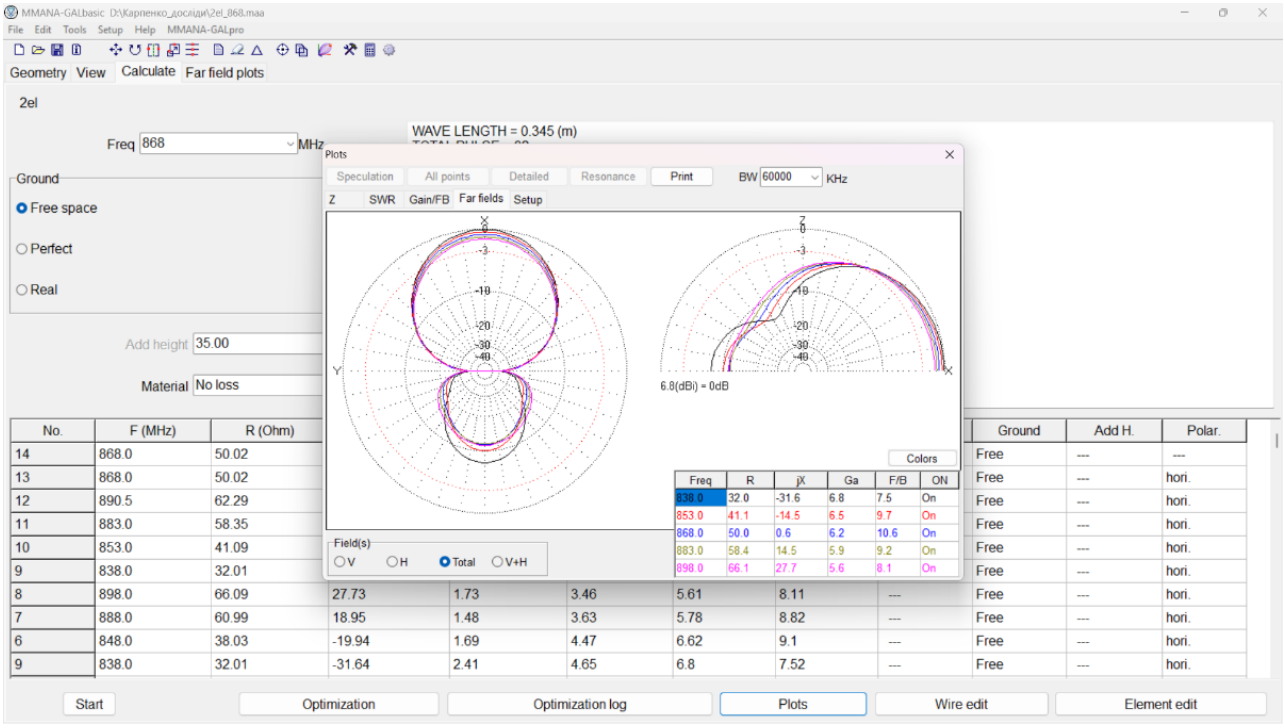


Рис. 2. Побудова антени



No.	F (MHz)	R (Ohm)						Ground	Add H.	Polar.
14	868.0	50.02						Free	---	---
13	868.0	50.02						Free	---	hori.
12	890.5	62.29						Free	---	hori.
11	883.0	58.35						Free	---	hori.
10	853.0	41.09						Free	---	hori.
9	838.0	32.01						Free	---	hori.
8	898.0	66.09	27.73	1.73	3.46	5.61	8.11	Free	---	hori.
7	888.0	60.99	18.95	1.48	3.63	5.78	8.82	Free	---	hori.
6	848.0	38.03	-19.94	1.69	4.47	6.62	9.1	Free	---	hori.
9	838.0	32.01	-31.64	2.41	4.65	6.8	7.52	Free	---	hori.

Freq	R	jX	Ga	F/B	ON
838.0	32.0	-31.6	6.8	7.5	On
853.0	41.1	-14.5	6.5	9.7	On
868.0	50.0	0.6	6.2	10.6	On
883.0	58.4	14.5	5.9	9.2	On
898.0	66.1	27.7	5.6	8.1	On

Рис. 3. Отримані значення для різних частот

відкритій місцевості протягом хвилини для отримання точних даних з супутників.

Для проведення дослідження було обрано прогулянковий маршрут від Арки Свободи українського народу, яка входила в зону покриття від вказаної позиції приймача на лівому березі і мала хороший сигнал (рис. 5).

Як можна побачити на рис. 6, маршрут GPS має уривки. Вони утворилися внаслідок втрати пакетів через різні фактори: щільність міської забудови, наявність металевих конструкцій, електромагнітні завади, а також можливе часткове перекриття прямої видимості між передавачем і приймачем.

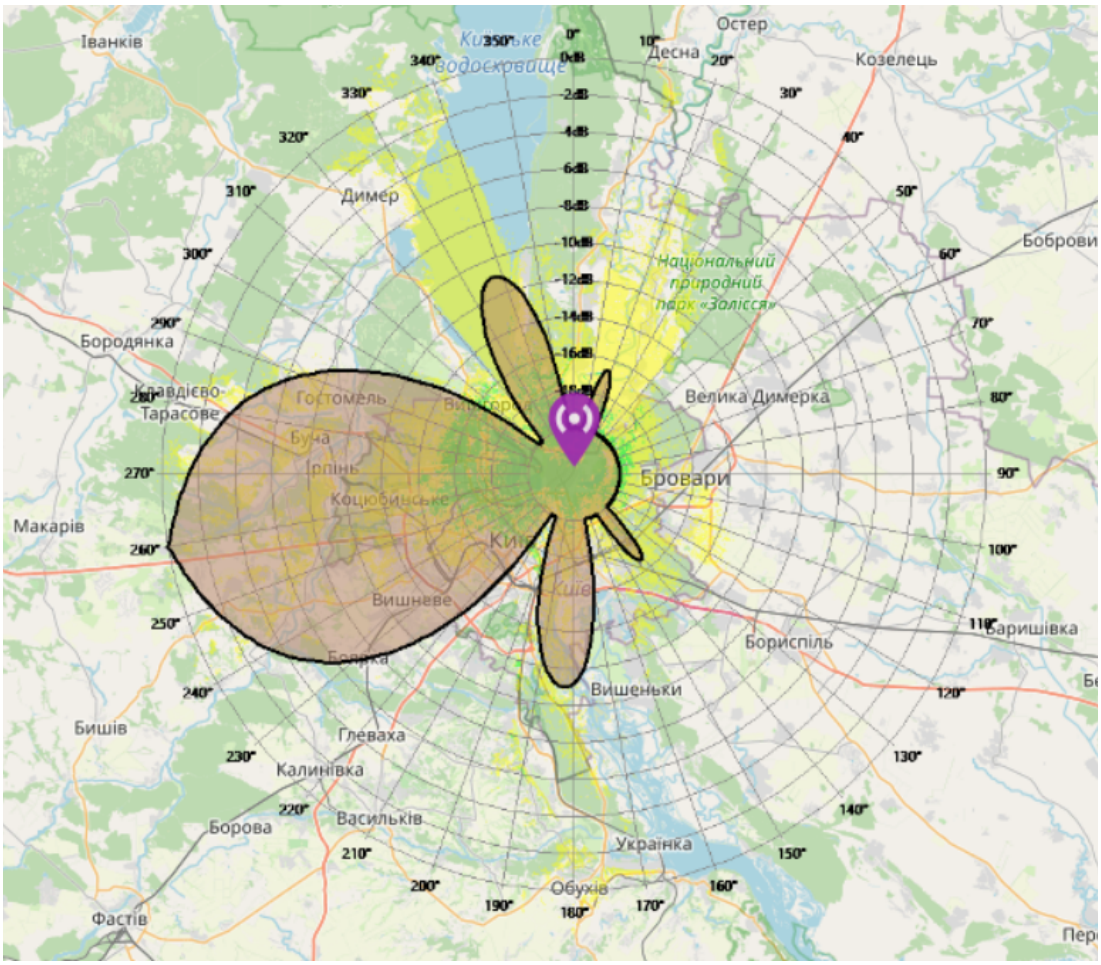


Рис. 3. Моделювання спрямованої антени в Tower Coverage

Рис. 4. Зібрана модель пристрою для дослідження покриття БСМ мереж

Рис. 5. Відрізок зони покриття

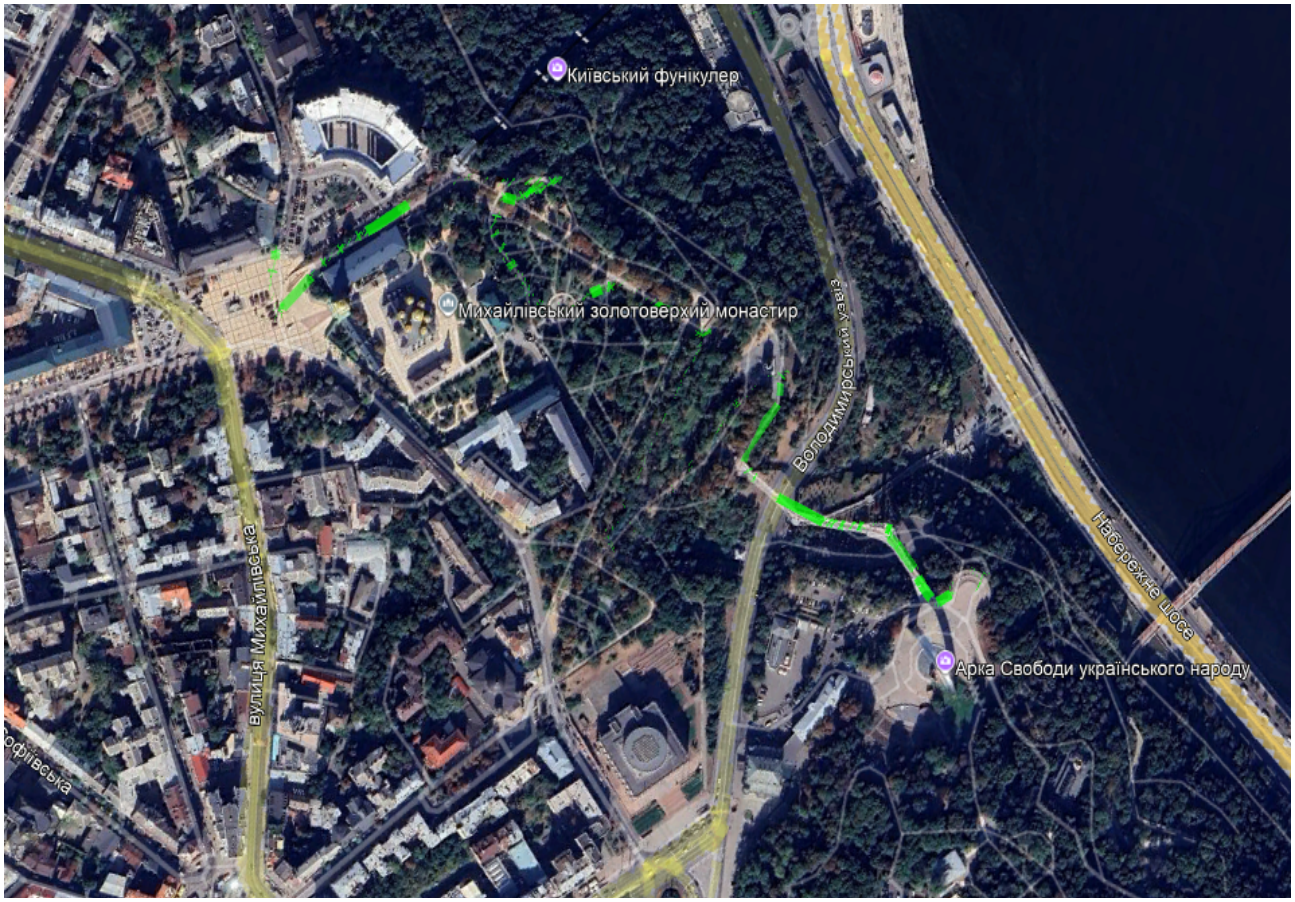


Рис. 6. Відрізок маршруту досліду

Висновки. Навіть за умови позитивного теоретичного прогнозу зони покриття, реальні умови міського середовища можуть суттєво впливати на якість та стабільність бездротового зв'язку. В ході роботи було створено портативний інструмент для збору GPS-координат та характеристик покриття, який дозволив на практиці перевірити ефективність розрахункової моделі. Напрямлена антена типу Moxon, змодельована й побудована для випробувань, продемонструвала високу ефективність у фокусуванні сигналу, що особливо важливо для зниження електромагнітних перешкод та покращення надійності зв'язку. Застосування програмного моделювання та подальших польових випробувань дозволило виявити невраховані фактори, що суттєво впливають на покриття: забудова, рельєф, джерела завад. Такий підхід демонструє важливість комбінування симуляцій з натурними даними під час проєктування БСМ. У перспективі розроблений пристрій може бути доопрацьований для автоматичного збору даних, інтеграції у БПЛА та побудови систем моніторингу для застосування в умовах бойових дій, техногенних загроз або екологічного аналізу.

Список літератури:

1. Wolff C. Radartutorial. Christian Wolff. URL: https://www.radartutorial.eu/02.basics/LPI%20radar.en.html (date of access: 25.04.2025).

2. Intelligent energy-efficient GNSS-assisted and LoRa-based positioning for wildlife tracking / J. J. López-Escobar et al. *IEEE Sensors Journal*. 2025. P. 1. URL: https://doi.org/10.1109/jsen.2024.3524456 (date of access: 16.05.2025).

3. Bhavanam B. P. R., Ragam P. LoRa Sense: Sensing and Optimization of LoRa Link Behavior Using Path-Loss Models in Open-Cast Mines. *Computer Modeling in Engineering & Sciences*. 2024. P. 1–10. URL: https://doi.org/10.32604/cmes.2024.052355 (date of access: 16.05.2025).

4. Mazzaro G. J. M. Homemade-Antenna Project for an Undergraduate Wave-Propagation Course. *American Society for Engineering Education* : Southeastern Section Conference, South Carolina. Charleston, 2022. URL: https://surli.cc/zqfqjv (date of access: 16.05.2025).

5. Laboratory to Develop a Practical Hand-Made Monopulse Antenna for RFID Localization Systems / M. Campo-Valera et al. *IEEE Access*. 2022. P. 1. URL: https://doi.org/10.1109/access.2022.3229898 (date of access: 18.05.2025).

6. Compact Antenna Test Range Using Very Small F/D Transmitarray Based on Amplitude Modification and Phase Modulation / J. Tang et al. *IEEE Transactions on Instrumentation and Measurement*. 2022. Vol. 71. P. 1–14. URL: https://doi.org/10.1109/tim.2022.3150835 (date of access: 18.05.2025).

7. MATLAB-Based Decoding of NOAA Satellite Images Using V-Dipole Antenna and RTL-SDR / M. T. Ahsan et al. *2024 14th International Symposium on Antennas, Propagation and EM Theory (ISAPE)*, Hefei, China, 23–26 October 2024. 2024. P. 1–4. URL: https://doi.org/10.1109/isape62431.2024.10840934 (date of access: 18.05.2025).

8. A Field Calibration Solution to Achieve High-Grade-Level Performance for Low-Cost Dual-Frequency GNSS Receiver and Antennas / A. Krietemeyer et al. *Sensors*. 2022. Vol. 22, No. 6. P. 2267. URL: https://doi.org/10.3390/s22062267 (date of access: 18.05.2025).

9. Micro-sized open-source and low-cost GPS loggers below 1 g minimise the impact on animals while collecting thousands of fixes / T. A. Wild et al. *PLOS ONE*. 2022. Vol. 17, No. 6. P. e0267730. URL: https://doi.org/10.1371/journal.pone.0267730 (date of access: 18.05.2025).

10. LoRa-LBO: An Experimental Analysis of LoRa Link Budget Optimization in Custom Build IoT Test Bed for Agriculture 4.0 / M. Swain et al. *Agronomy*. 2021. Vol. 11, No. 5. P. 820. URL: https://doi.org/10.3390/agronomy11050820 (date of access: 18.05.2025).

Karpenko M.I., Moshenskyi A.O. DESIGN AND FIELD TESTING OF A DEVICE FOR RESEARCHING WSN COVERAGE

The article highlights the process of designing, assembling and practical application of an experimental device for studying the spatial characteristics of the coverage of a wireless sensor network (WSN) operating on the basis of LoRa technology. The relevance of the topic is due to the growing need to build highly reliable data transmission networks in difficult conditions – in particular, in urban areas, areas with a high level of interference and in frontline areas, where traditional means of communication demonstrate instability. The main goal of the research is to create a mobile technical solution that allows not only theoretically, but also practically to assess the effectiveness of coverage, depending on the parameters of the environment. Given the critical importance of the spatial coverage of the network, especially in difficult conditions – urban and frontline areas – a compact field data collection tool was developed taking into account the real environment. The device includes the

E220-900T30D module, the GP-01Kit GPS receiver and a laptop with u-center software for synchronization of coordinates and time. Special attention is paid to the Moxon-type directional antenna, which was modeled, built and tested. MMANA-GAL was used for electromagnetic modeling, and Tower Coverage was used for coverage area prediction. During the study, a series of tests were conducted in the city of Kyiv, where the analysis of packet loss and signal stability demonstrated a significant dependence of the communication quality on the type of development, the presence of obstacles and terrain features. The assembled device allowed not only to confirm the hypotheses regarding the influence of environmental factors on communication, but also to identify the limitations of existing coverage calculation models. The conclusion is the need to combine theoretical modeling with real field measurements when designing BSM to ensure reliable operation in specific conditions.

Key words: *wireless sensor network, radio module, Moxon antenna, LoRa, GPS.*

УДК 004.4
DOI https://doi.org/10.32782/2663-5941/2025.3.2/31

Кордюк Р.І.
Національний університет «Львівська політехніка»

ДЕРЕВОВИДНА МОДЕЛЬ КОЛАЖІВ ІНФОРМАЦІЙНОЇ СИСТЕМИ

Виконано ґрунтовний аналіз відомих моделей колажів, які використовуються в найпопулярніших інформаційних системах, таких як Canva, Pixlr Editor, Online Photoshop. Встановлено, що існуючі моделі колажів не забезпечують можливості створення структурованих зв'язків між їх складовими елементами. Це обмежує можливості для редагування та гнучкої зміни зображень у межах одного колажу.

Розроблена нова модель колажів, яка базується на деревовидній структурі. Вона утворена чотирма рівнями вкладеності. Коренем дерева є головний елемент, який знаходиться на першому рівні і має максимальні задані розміри графічного зображення з можливістю задання відступів від зовнішніх меж графічного зображення та нульовим коефіцієнтом заокругленості з можливістю її динамічної змінюваності. На другому рівні вкладеності знаходяться рядки та горизонтальні повзунки. На третьому рівні вкладеності, в середині рядків, знаходяться або колонки або контейнери зображень і вертикальні повзунки між ними. Четвертий рівень вкладеності утворений розташованими в середині колонок контейнерами зображень та горизонтальними повзунками між ними.

Введені взаємодіючі зв›язки між елементами колажа утворюють складну і високо функціональну структуру. Кожен елемент дерева змінюється пропорційно до інших елементів, що означає, що внесені зміни в одному елементі автоматично відображаються на всіх інших, які з ним пов›язані. Це надає користувачу надзвичайну гнучкість у створенні та редагуванні графічних композицій, оскільки зміни, внесені в одному місці, автоматично адаптуються та відповідають всій структурі. Виконано порівняльний аналіз застосування розробленої деревовидної моделі і відомих моделей колажів. Встановлено, що використання розробленої деревовидної моделі оптимізує роботу з колажами, зменшуючи кількість виконуваних дій у 5 разів. Це не лише скорочує час на редагування, але й підвищує точність та якість кінцевого результату, що є важливим фактором у створенні складних графічних композицій.

Ключові слова: *колаж, деревовидна модель колажу, графічне зображення, контейнер зображення.*

Постановка проблеми. Опрацювання графічних зображень – важлива складова комп'ютерної графіки, сьогодні існує багато програмних засобів, які спрощують роботу із зображеннями. Однак низка завдань залишається невирішеною. Одне з них – відсутність структурованих зв'язків між елементами графічних композицій (колажів). Популярні онлайн-сервіси для створення колажів не забезпечують взаємозв'язку між складовими, тому зміна одного фрагмента не призводить до оновлення інших. Це обмежує гнучкість редагування та вимагає ручного коригування композиції.

У галузі обробки зображень постійно з'являються нові алгоритми для підвищення якості та автоматизації процесів, а автоматичне створення композицій (фотоколажів) стало окремим напрямом досліджень як ефективний спосіб презентації багатьох зображень. Формування колажу зазвичай розглядають як задачу оптимізації: визначається функція якості композиції, яку алгоритм максимізує. Такий підхід потребує значних обчислень і при великій кількості фото стає складним завданням.

Для підвищення ефективності редагування композицій існують алгоритми та методи оптимізації. Застосування ефективних методів оптимізації суттєво пришвидшує автоматичне компонування колажів.

Проблема узгодженого та ефективного редагування графічних композицій залишається актуальною, тому подальші дослідження у цій сфері є необхідними. Це дозволить впровадити новітні алгоритми в інструменти редагування та створити більш гнучкі й інтелектуальні засоби, які забезпечуватимуть високоякісну композицію за мінімальних зусиль користувача, що особливо важливо в умовах стрімкого зростання обсягів візуальних даних і потреби в їх ефективній презентації.

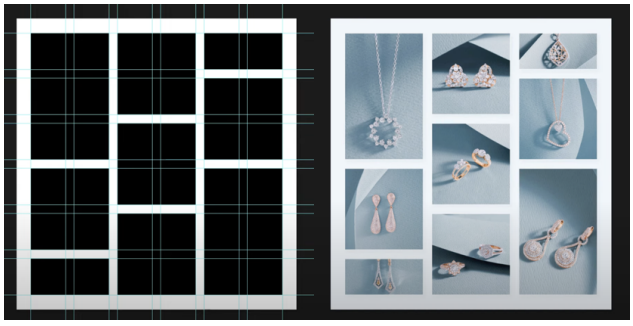
Незважаючи на популярність інформаційних систем для створення колажів, таких як Canva, Pixlr Editor і Online Photoshop, їхні моделі мають суттєві обмеження. Вони не передба-

чають можливість створення структурованих зв›язків між елементами колажу, що значно ускладнює процеси редагування та адаптивної зміни композицій. Виникає потреба у розробці нової моделі, яка дозволить автоматично адаптувати елементи колажу, спростити процес редагування та підвищити ефективність роботи користувачів.

Аналіз останніх досліджень і публікацій. Найбільш часто використовуваними технологіями опрацювання зображень є Photoshop Online [1], Canva [3] та Pixlr [7].

Photoshop Online є однією з найпопулярніших онлайн-технологій для редагування графічних зображень. Модель створення колажів у Photoshop Online полягає у формуванні наявними інструментами сітки на полотні. На рис.1 сітка на полотні зображена у вигляді горизонтальних і вертикальних ліній. Інструментами задається кількість рядків і колонок сітки полотна. Всі колонки мають однакову ширину, і всі рядки мають однакову висоту. Елементи полотна для розташування зображень формуються у вигляді прямокутників, які на рис. 1 a) проілюстровані чорними прямокутниками у білих рамках.

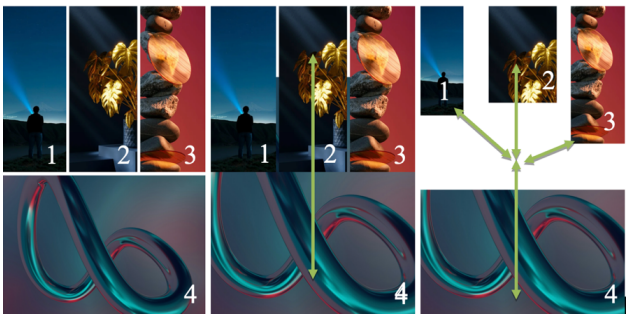
Кожен елемент доданий на полотно, є контейнером для зображення, і розміщується на одному або багатьох елементах сітки. Колаж утворюється розташованими на елементах сітки полотна контейнерами зображень, що показано на рис. 1 b). Проте, така модель колажу має деякі недоліки, головним з яких є відсутність зв›язку між контейнерами зображень. Зміна вимірів контейнера не впливає на зміну вимірів всіх суміжних з ним контейнерів.



a) b)

Рис. 1. Полотно із сіткою та контейнерами a), Колаж зображення b)

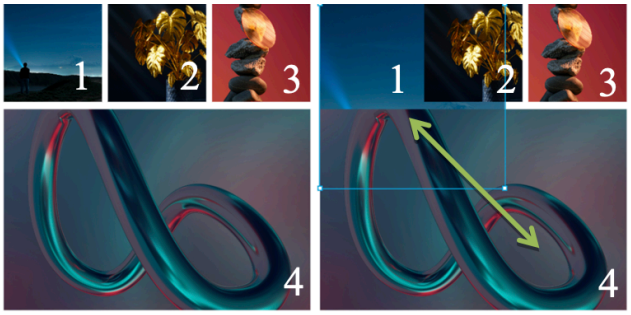
Canva [3] є іншою онлайн-технологією редагування графічних зображень, яка відрізняється від Photoshop Online тим, що її основна мета – це допомогти користувачам створювати графічний дизайн з нуля. Шаблони колажів у Canva реалізовані таким чином, що немає зв’язку між елементами колажу, і кожна одиниця цього колажу, є незалежною. На рис. 2 проілюстровано модель утворення колажів системою Canva. Початковий вигляд колажу наведено на рис. 2 a). Колаж після зміни розмірів висоти нижнього контейнера накладається на обидва верхні контейнери (див. рис. 2 b), що потребує необхідності ручної зміни розмірів висоти обох верхніх контейнерів. На рис. 2 c) проілюстровано відсутність зв›язків між фрагментами зображень колажа. Модель колажів у Canva аналогічна моделі колажів Photoshop Online і має такі самі недоліки.



a) b) c)

Рис. 2. Колаж в початковому вигляді a), Накладання фрагментів зображень b), Ілюстрація відсутності зв’язків між елементами колажа c)

Pixlr Editor [7] є ще однією популярною онлайн-технологією для редагування графічних зображень. Модель колажів Pixlr Editor полягає у використанні низки наявних шаблонів або застосуванні інструментів для створення оригінальних шаблонів з порожніми контейнерами зображень. Можливі зміни розмірів по одній або обох координатах контейнера. При збільшенні розмірів одного контейнера відбування накладанням збільшуваного контейнера на сусідні контейнери. Однак дана модель, як і попередні моделі, не забезпечує зв’язку між сусідніми контейнерами зображень. На рис. 3 проілюстровано модель утворення колажів системою Pixlr Editor. Початковий вигляд колажу наведено на рис. 3 a). Колаж після зміни розмірів контейнера з № 1, знаходиться під контейнерами з № 2, 3, 4 (див. рис. 3 b) тому, що відсутній зв’язок між елементами колажа. Для отримання очікуваного результату необхідно змінити розміри усіх інших елементів колажа вручну.



a) b)

Рис. 3. Колаж в початковому вигляді a), ілюстрація зміни розмірів елемента колажа з № 1, та відсутності зв'язку між елементами з № 2, 3, 4 b)

У науковій літературі існує ряд досліджень, спрямованих на покращення автоматизованих підходів до створення колажів. Одним із таких підходів є AutoCollage [8]. Який полягає у виділенні областей інтересу (ROI) з використанням графового розрізу та згладжування країв для створення візуально привабливих композицій. Цей підхід забезпечує краще використання полотна та підвищує естетичну якість результату, однак все ще має обмеження, пов›язані із втратою семантичного зв›язку між зображеннями [8].

У дослідженні автори Фернандес Д., Постелс Дж. та Томбарі Ф. представили модель Variational Transformer Network (VTN), яка поєднує варіаційні автокодери (VAE) з трансформерами для генерації макетів. Основна ідея полягає в тому, щоб навчити модель розподілу макетів без використання жорстко заданих правил дизайну, дозволяючи моделі самостійно вивчати високорівневі взаємозв›язки між елементами макету [9].

Модель складається з двох основних компонентів: енкодера та декодера. Енкодер перетворює вхідні дані макету в латентний простір, використовуючи механізм самоуваги, що дозволяє враховувати взаємозв'язки між елементами. Декодер, також побудований на основі трансформера, відновлює макет з латентного представлення. Формально, енкодер приймає послідовність елементів макету $x = \{x_1, x_2, ..., x_n\}$ та обчислює параметри розподілу $q(z|x)$, де z – латентне представлення. Декодер моделює умовний розподіл $p(x|z)$, відновлюючи макет з латентного простору. Під час навчання оптимізується варіаційна нижня межа (ELBO):

$$\mathcal{L} = \mathbb{E}_{q(zx)}\left[\log p(x|z)\right] - \mathcal{D}_{KL}\left(q(z|x) \| p(z)\right), \quad (1)$$

де $\mathcal{D}_{KL}$ – дивергенція Кульбака-Лейблера між апостеріорним розподілом $q(z|x)$ та апріорним розподілом $p(z)$. Це дослідження демонструє можливість використання трансформерів у поєднанні з VAE для автоматичної генерації макетів, що дозволяє моделі навчатися складним правилам дизайну без явного нагляду. Такий підхід забезпечує високу схожість з навчальними даними та різноманітність згенерованих макетів, що є важливим для автоматизації процесів дизайну в різних сферах, включаючи створення документів, інтерфейсів користувача та розташування меблів.

У роботі автори Хаді А., Лін Х. та Бала К. досліджують проблему автоматичного розташування декількох фотографій на заданому полотні з високою естетичною якістю. Вони пропонують новий підхід для автоматичної генерації колажів із заданим співвідношенням сторін, вперше впроваджуючи техніку глибокого навчання з підкріпленням у цю сферу [16].

Процес генерації колажу моделюється як послідовність дій агента, який розміщує фотографії на полотні, налаштовуючи їхні просторові позиції, кути орієнтації та порядок розміщення. Агент навчається покращувати як загальний макет, так і локальні деталі, отримуючи винагороду, яка враховує суб'єктивні та об'єктивні фактори естетичної якості. Функція винагороди $R(s)$ розроблена таким чином, щоб враховувати як глобальні, так і локальні аспекти естетики:

$$R(s) = \alpha \cdot Q_{global}(s) + \beta \cdot Q_{local}(s), \quad (2)$$

де $Q_{global}(s)$ – оцінює загальну естетику макету, а Q_{local} – окальні деталі, такі як перекриття фотографій та їх вирівнювання. Коефіцієнти α та β визначають вагу кожного компонента у загальній оцінці. Для подолання нестачі навчальних даних, автори попередньо навчають глибоку естетичну мережу на великому наборі даних з естетики зображень (CPC) для загального вилучення естетичних характеристик та пропонують модуль злиття уваги для представлення структурних характеристик колажу. Агент навчається за допомогою алгоритму глибокого Q-навчання, оновлюючи свою політику на основі отриманих винагород. Ця робота є першою, яка застосовує глибоке навчання з підкріпленням для задачі генерації естетичних фото-колажів. Запропонований підхід дозволяє автоматично створювати колажі з високою естетичною якістю, перевершуючи інші методи в оцінці естетичної якості.

У дослідженні Хуо Х. та Вонг Ф. досліджують застосування штучного інтелекту для автоматизації процесу дизайну макетів постерів у візуальній комунікації. Вони пропонують систему, яка складається з двох основних модулів: "learner"

(навчальний модуль) та "generator" (генеративний модуль) [11, 1–9].

Навчальний модуль використовує просторову трансформаційну мережу для класифікації елементів макету та визначення їхніх початкових позицій на полотні.

Генеративний модуль застосовує згорткову нейронну мережу для вдосконалення початкового макету, враховуючи принципи дизайну, такі як правило третин та золотий перетин.

Формально, система оптимізує функцію втрат, яка враховує відхилення від естетичних принципів:

$$\mathcal{L} = \sum_{i=1}^{n} (\alpha \cdot \text{Loss}_{\text{alignment}}(e_i) + \beta \cdot \text{Loss}_{spacing}(e_i), \quad (3)$$

де e_i – елемент макету, α та β – вагові коефіцієнти, що визначають важливість відповідних втрат. Це дослідження демонструє, як штучний інтелект може бути інтегрований у процес дизайну постерів, автоматизуючи розміщення елементів та забезпечуючи дотримання естетичних принципів. Це може значно підвищити ефективність дизайнерів та забезпечити високу якість кінцевого продукту.

У роботі авторів Джоу М., Ху С., Ма Й. та інших, представлено систему AutoPoster, яка автоматично генерує рекламні постери, враховуючи зміст вхідних даних. Система складається з чотирьох основних етапів:

- Попередня обробка зображення: аналіз вхідного зображення для вилучення ключових характеристик продукту.
- Генерація макету: використання генеративної моделі для створення оптимального розташування елементів на постері.
- Генерація слогану: створення текстового повідомлення, що відповідає продукту та дизайну постера.
- Налаштування стилю: застосування відповідних кольорових схем та шрифтів для забезпечення естетичної привабливості [17].

Формально, система оптимізує функцію втрат, яка враховує відповідність між зображенням, текстом та макетом:

$$\mathcal{L} = \lambda 1 \cdot \text{Loss}_{\text{image-text}} + \lambda 2 \cdot \text{Loss}_{\text{layout}} + \lambda 3 \cdot \text{Loss}_{\text{style}}, \quad (4)$$

де $\lambda 1, \lambda 2, \lambda 3$ – вагові коефіцієнти для кожного компонента втрат. AutoPoster демонструє можливість повної автоматизації процесу створення рекламних постерів, що враховують зміст продукту. Це дозволяє маркетологам швидко та ефективно створювати візуально привабливі матеріали, зменшуючи потребу в ручній праці.

Flowy – це система, яка підтримує прийняття рішень у UX-дизайні шляхом автоматичного аналізу шаблонів у багатосторінкових користувацьких потоках. Система використовує штучний інтелект для аналізу великих колекцій існуючих інтерфейсів та автоматично анотує дизайн-шаблони, що використовуються в них. Основні етапи роботи системи:

- Збір даних: збирання великої кількості знімків екранів різних додатків.
- Аналіз шаблонів: використання алгоритмів комп›ютерного зору для виявлення та класифікації повторюваних шаблонів дизайну.
- Анотація: автоматичне додавання метаданих до виявлених шаблонів для полегшення їх використання дизайнерами [13].

Flowy допомагає дизайнерам приймати обґрунтовані рішення, надаючи їм доступ до аналізу існуючих шаблонів дизайну. Це сприяє покращенню користувацького досвіду та забезпечує.

У дослідженні авторів Джоу М., Ху С., Ма Й. та інших, представили модель Composition-Aware Graphic Layout GAN (CGL-GAN), яка генерує графічні макети для візуально-текстових презентацій на основі композиції вхідних зображень. Модель враховує як глобальні семантичні, так і просторові характеристики зображень для створення відповідних макетів [17]. Основні компоненти моделі:

1. *Модуль узгодження доменів* (*Domain Alignment Module, DAM)*: Цей модуль зменшує розрив між навчальними даними (з масками) та тестовими даними (без масок), забезпечуючи узгодженість між ними.

2. *Генератор:* Складається з багатомасштабної згорткової нейронної мережі та трансформера, які разом генерують макети на основі виходів DAM та користувацьких обмежень.

3. *Дискримінатор:* Має подібну структуру до генератора і використовується для навчання моделі.

Функція втрат включає реконструктивну та змагальну складові:

$$\mathcal{L} = \mathcal{L}_{\text{recon}} + \lambda \mathcal{L}_{\text{adv}}, \quad (5)$$

де $\mathcal{L}_{\text{recon}}$ – реконструктивна втрата, $\mathcal{L}_{\text{adv}}$ – змагальна втрата, а λ – ваговий коефіцієнт. Ця робота демонструє, як врахування композиції зображення може покращити якість згенерованих макетів для візуально-текстових презентацій, що є важливим для автоматизації дизайну рекламних постерів та інших графічних матеріалів.

У роботі авторів Джоу Й., Джинг К., Лі З. та інших, представили модель Element-Conditioned GAN (EcGAN), яка генерує графічні макети, враховуючи задані елементи дизайну, такі як кількість та типи елементів [18]. Кожен елемент представлений у вигляді обмежувального прямокутника (bounding box), а модель включає три ключові компоненти:

Маска елементів (Element Mask): Використовується для представлення присутності або відсутності певних елементів у макеті.

Втрати, обумовлені елементами (Element Condition Loss): Ця функція втрат забезпечує відповідність згенерованого макету заданим умовам щодо елементів.

Двоступеневий дискримінатор (Two-Step Discriminator): Перший дискримінатор оцінює реалістичність макету, а другий перевіряє відповідність макету заданим умовам.

Функція втрат для генератора включає змагальну втрату та втрату, обумовлену елементами:

$$\mathcal{L}_G = \mathcal{L}_{adv} + \lambda \mathcal{L}_{elem} , \quad (6)$$

де $\mathcal{L}_{adv}$ – змагальна втрата, $\mathcal{L}_{elem}$ – втрата, обумовлена елементами, а λ – ваговий коефіцієнт. EcGAN дозволяє генерувати графічні макети, які відповідають заданим умовам щодо елементів дизайну, що є корисним для автоматизації процесу створення макетів з конкретними вимогами.

Дослідження авторів Ші Й., Шанг М. та Кі З. є всебічним оглядом сучасних методів генерації макетів на основі глибоких генеративних моделей, таких як варіаційні автокодери (VAE), генеративні змагальні мережі (GAN) та дифузійні моделі. Автори аналізують різні підходи, їх переваги та недоліки, а також області застосування. Цей огляд надає всебічне розуміння сучасних методів генерації макетів, що є корисним для дослідників та практиків у галузі автоматизації дизайну [15].

Автори Чен Л., Джинг К., Тсанг Й. та Жоу Т. розробили систему Iris, орієнтовану на користувача, яка забезпечує інтерактивне середовище для генерації графічних макетів з урахуванням множинних обмежень. Дизайнер може вказати певні обмеження, такі як вирівнювання елементів або залишення певних областей порожніми, після чого генеративна модель системи пропонує відповідні макети. Цей підхід підкреслює важливість контролю та інтерактивності в практичних умовах, дозволяючи дизайнерам швидко отримувати якісні пропозиції, які відповідають їхнім вимогам [10].

У дослідженні авторів Шабані М., Вонг З. та інших, пропонують модель Visual Layout Composer, яка поєднує дифузійні процеси в образному та векторному доменах для генерації дизайну макетів. Модель одночасно генерує загальний вигляд дизайну в образному домені та оптимізує розміри і позиції кожного елемента дизайну у векторному домені [14]. Цей підхід забезпечує як редагованість векторних елементів, так і високу візуальну якість кінцевого дизайну.

Найбільш поширеною метою таких досліджень є генерація макетів з дотриманням певних естетичних критеріїв, при цьому інструменти користувацького контролю часто є обмеженими або відсутніми. Крім того, моделі, побудовані на нейромережах, як правило, потребують великих обчислювальних ресурсів, мають низьку передбачуваність у поведінці та не дозволяють вручну змінювати структуру композиції після її генерації.

Однак в умовах практичних задач – створення навчальних матеріалів, редагування веб-контенту, адаптації візуальних блоків у багатомовних інтерфейсах тощо – критично важливим є передбачуване, контрольоване редагування, яке гарантує збереження структури при зміні окремих елементів. У цьому контексті особливої актуальності набувають структурні (не AI-залежні) моделі, які орієнтовані не на генерацію, а на ієрархічне представлення і трансформацію вже наявного візуального контенту.

Описана у дослідженні деревовидна модель графічного колажу є саме таким прикладом: вона формалізує структуру композиції у вигляді дерева вкладених об’єктів, що дозволяє зберігати логічну цілісність і взаємозв’язки між елементами під час їх трансформації. Завдяки цьому модель забезпечує:

- Системну узгодженість змін у композиції (наприклад, при зміні розмірів або положення одного блоку автоматично враховуються залежні блоки).
- Масштабованість до великої кількості елементів без втрати керованості.
- Універсальність у застосуванні.

На відміну від популярних онлайн-редакторів (Canva, Photoshop Online, Pixlr Editor), де кожен контейнер функціонує незалежно й зміни в одному елементі вимагають ручної синхронізації з іншими, розроблена деревовидна структура дозволяє автоматизовано підтримувати композиційну цілісність, що значно зменшує трудомісткість та ризик помилок у складних колажах [3; 1; 7].

Таким чином, модель може розглядатись як наукове доповнення до сучасного напряму досліджень у сфері автоматизованої композиції зображень: вона не залежить від алгоритмів глибокого навчання, натомість пропонує структурну концепцію ієрархії, яка може бути інтегрована як у ручні редактори, так і у гібридні системи, що поєднують генерацію та редагування.

Постановка завдання. Метою статті є підвищення ефективності створення та редагування графічних зображень завдяки розробленню нової, більш гнучкої деревовидної моделі колажів. Розроблена модель побудована на основі структурованих рекурсивних зв'язків між контейнерами зображень, що дозволяє автоматично адаптувати розміри та пропорції усіх взаємопов'язаних елементів композиції при редагуванні одного з них. Це забезпечує суттєве спрощення та пришвидшення процесу внесення змін у графічні колажі порівняно з традиційними моделями, зменшуючи кількість необхідних операцій користувача. Окремою метою дослідження є проведення детального порівняльного аналізу розробленої деревовидної моделі з існуючими рішеннями, які широко використовуються у популярних інформаційних системах Canva, Pixlr Editor та Photoshop Online. Це дозволяє оцінити реальні переваги розробленої моделі, підтвердити її ефективність та доцільність застосування для автоматизації процесів створення і редагування складних графічних композицій у сучасних онлайн-сервісах обробки зображень.

Виклад основного матеріалу. Деревовидна модель складається з чотирьох рівнів, де кожен елемент взаємопов'язаний і впливає на інші компоненти, забезпечуючи автоматичну адаптацію всієї композиції.

Коренем дерева є головний елемент, який знаходиться на першому рівні. Він має максимальні задані розміри графічного зображення з початковим білим кольором фону. Корінь має нульовий коефіцієнт заокругленості з можливістю її динамічної змінюваності. Наявна можливість задання відступів від зовнішніх меж графічного зображення.

На другому рівні вкладеності знаходяться рядки та горизонтальні повзунки. Кількість рядків дорівнює n, а кількість повзунків дорівнює $n - 1$.

На третьому рівні вкладеності в середині рядків, знаходяться або колонки або контейнери зображень, та вертикальні повзунки між ними. Кількість колонок та контейнерів зображень дорівнює n, а кількість вертикальних повзунків між ними дорівнює $m - 1$.

На четвертому рівні вкладеності розташовані контейнери зображень в середині колонок та горизонтальні повзунки між ними. Кількість контейнерів зображень в середині колонок дорівнює k. А кількість повзунків між ними дорівнює $k - 1$.

В рамках цієї моделі, кожна одиниця цього дерева є незалежним елементом, і важливо відзначити, що ці елементи не просто ізольовані, а взаємодіють між собою, утворюючи складну і високо функціональну структуру. Кожна одиниця цього дерева змінюється пропорційно до інших одиниць, що означає, що внесені зміни в одному елементі автоматично відображаються на всіх інших, які з ним пов›язані. Це надає користувачу надзвичайну гнучкість у створенні та редагуванні графічних композицій, оскільки зміни, внесені в одному місці, автоматично адаптуються та відповідають всій структурі.

Коли колаж односекційний, то він не містить повзунків, які призначені для можливості зміни вимірів секцій. В однорядковому колажі в кожній секції наявний повзунок секції, яким можливо змінювати ширину кожної секції. Повзунки колажа розміщені між секціями. Наприклад, якщо однорядковий колаж утворений п›ятьма секціями, то він міститиме чотири повзунки секцій. Деревовидна модель колажа, яка утворена двома і більше рядками секцій містить міжрядкові повзунки. Повзунок рядка призначений для зміни висоти усіх секцій одночасно, якими утворений рядок. Наприклад, семи рядковий колаж містить шість повзунків рядків. Зміна повзунком секції ширини будь-якої вибраної секції автоматично впливає на виміри усіх інших секцій рядка. Наявні мінімальні і максимальні виміри ширини секцій. Мінімальна ширина секції є однаковою для всіх секцій. Максимальна ширина секції визначається шириною колажа від якої віднімається сума усіх мінімальних секцій рядка. Практично деревовидний колаж може містити десятки, а то й сотні секцій і рядків.

Динамічна зміна параметрів забезпечує ефективне використання простору та адаптацію композиції під потреби користувача.

Таким чином, результати дослідження демонструють, що деревовидна модель дозволяє гнучко керувати розмірами та розташуванням графічних елементів, забезпечуючи високий рівень інтерактивності та адаптивності.

Для реалізації Деревовидної моделі колажів було вибрано фреймворк Angular [2],

а також бібліотеку Fabric.js [5] для зручної роботи з HTML5 Canvas [4] та WebGL [6].

На рис. 4 a) зображено трьохрівневий колаж. На першому рівні, розташований батьківський елемент. На другому рівні знаходяться два рядки, які є дітьми батьківського елемента. На третьому рівні вкладеності знаходяться контейнери зображень, які є дітьми рядків. Контейнери з № 1, 2 та 3 є дітьми першого рядка. Контейнер з № 4 є дитиною другого рядка. Між контейнерами з № 1, 2 та 3 розташовані вертикальні повзунки. Між першим і другим рядком розташований горизонтальний повзунок.

На рис. 4 b) зображена зміна розмірів елементів колажа з № 1 та № 2, на цьому прикладі можемо побачити зв'язок між елементами з № 1 та № 2. Між цими елементами знаходиться вертикальний повзунок який керує зміною розмірів. Після руху повзунка в праву сторону, елемент з № 1 збільшив свою ширину, а елемент з № 2 пропорційно зменшив свою ширину. Елементи з № 3 та № 4 залишились незміненими.

На рис. 4 c) зображено зміна розмірів елементів колажа з № 1, 2, 3 та 4. На цьому прикладі можемо побачити зв'язок між усіма елементами колажа. Між елементами які знаходяться у 1 та 2 рядках, знаходиться горизонтальний повзунок який керує зміною розмірів висоти. Після руху повзунка вниз, елементи з № 1, 2, 3 збільшили свою висоту, а елемент з № 4 пропорційно зменшив свою висоту.

На рис. 5 a) зображено чотирьох рівневий колаж. На першому рівні, розташований батьківський елемент. На другому рівні знаходяться два рядки, які є дітьми батьківського елемента. На третьому рівні вкладеності знаходяться контейнери зображень з № 1, 2, 3 та 4, які є дітьми першого рядка та колонка з контейнером зображення № 8, які є дітьми другого рядка. На четвертому рівні вкладеності, розташовані контейнери з № 5, 6 та 7, які є дітьми колонки в середині другого рядка. Між контейнерами першого рядка з № 1, 2, 3 та 4 розташовані вертикальні повзунки. Між контейнером зображення з № 8 та колонкою з контейнерами зображень з № 5, 6 та 7, також розташований вертикальний повзунок. Між контейнерами зображень з № 5, 6 та 7 розташовані горизонтальні повзунки, а також між першим і другим рядком розташований горизонтальний повзунок.

На рис. 5 b) зображено зміна розмірів елементів колажа з № 1, 2, 3, 4, 5, 6, 7 та 8. На цьому прикладі можемо побачити зв'язок між елементами першого рядка з № 1, 2, 3 та 4, а також елементами другого рядка, колонкою з контейнерами зображень № 5, 6, 7 та окремим контейнером зображення з № 8. Після руху вертикального повзунка в ліву сторону, який розташований між контейнерами з № 3 та № 4, можемо побачити приклад роботи деревовидної моделі. При досяганні контейнером зображення мінімальної ширини, батьківський елемент додає до зміни розмірів наступний елемент рядка, і так до останнього елемента рядка. В залежності від напрямку руху вертикального повзунка, наступним елементом буде елемент напрямку руху цього повзунка.

У другому рядку, зображено приклад зв'язку між елементами колонки. Між колонкою з контейнерами зображення № 5, 6, 7 та контейнером зображення з № 8 також розташований вертикальний повзунок. Після руху цього повзунка у праву сторону, можемо побачити пропорційне збільшення ширини усіх елементів колонки № 5, 6, 7, та зменшення ширини контейнера зображення з № 8.

На рис. 5 c) зображено зміна розмірів висоти елементів колажа з № 5, 6 та 7, на цьому прикладі можемо побачити зв'язок між елементами колонки. Після руху горизонтального повзунка, який розташований між контейнерами з № 5 та

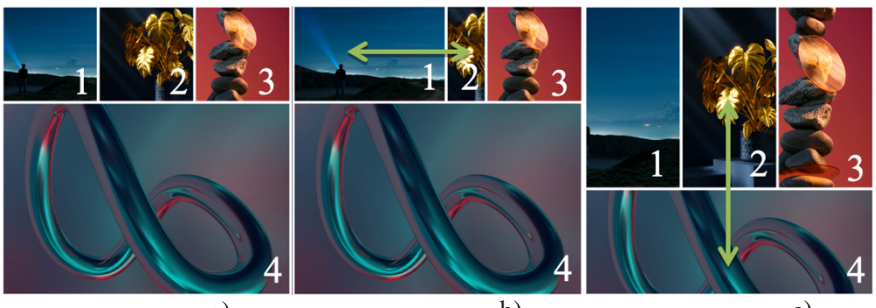


Рис. 4. Трьохрівневий колаж a) та ілюстрації зв'язків між елементами з рухом вертикального b) і горизонтального c) повзунків

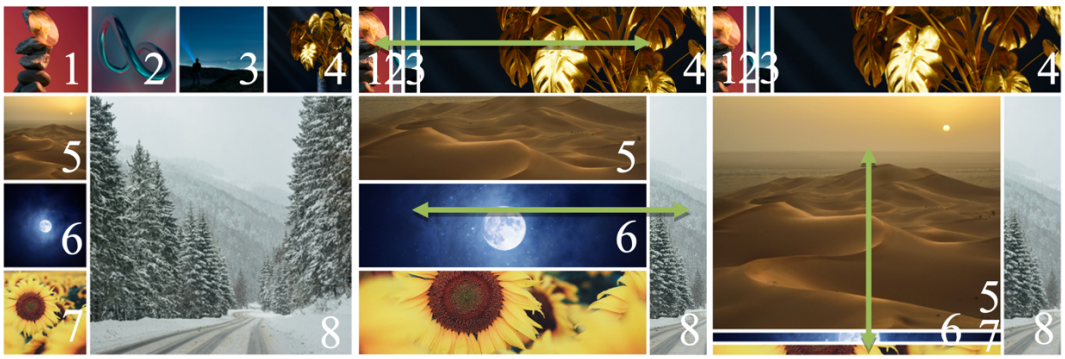


Рис. 5. Ілюстрація чотирьохрівневого колажу в початковому вигляді a), руху повзунка між елементами рядка b) та горизонтального повзунка між елементами колонки c)

№ 6, в напрямку нижньої межі, можемо побачити, що контейнер з № 5 збільшив свою висоту, а контейнери з № 6 та 7 зменшили свою висоту. При досяганні контейнером зображення мінімальної висоти, батьківський елемент додає до зміни розмірів наступний елемент колонки, і так до останнього елемента колонки. В залежності від напрямку руху горизонтального повзунка, наступним елементом буде елемент напрямку руху цього повзунка.

На рис. 6 зображено зміна розмірів висоти елементів колажа з № 1, 2, 3, 4, 5, 6, 7 та 8, на цьому прикладі можемо побачити зв'язок між усіма елементами деревовидної моделі. Після руху горизонтального повзунка, який розташований між першим та другим рядком, в напрямку нижньої межі, можемо побачити, що контейнери першого рядка з № 1, 2, 3, 4 пропорційно збільшили свою висоту, а елементи другого рядка (колонка з контейнерами № 5, 6 та 7, а також окремий контейнер зображення з № 8) пропорційно зменшили свою висоту. При досяганні рядком мінімальної висоти, батьківський елемент додає до зміни розмірів наступний рядок, і так до останнього рядка батьківського елемента. В залежності від напрямку руху горизонтального повзунка, наступним рядком буде рядок напрямку руху цього повзунка.

Порівняння за кількістю виконуваних дій, які потрібно виконати для опрацювання колажів відомими моделями і розробленою деревовидною моделлю.

Як видно з рисунка 7 b), відбувається накладання контейнера з № 1 на контейнери зображення з № 2 і № 4. Але подібне накладання не повинно бути присутнім на колажі. Щоб його усунути, необхідно вручну виконати наступні 10 дій:

1. Змінити розміри ширини контейнера зображення з № 1. Виконання цієї дії проілюстроване на рис. 7 b).

Рис. 6. Ілюстрація руху горизонтального повзунка між першим та другим рядком

2. Змінити розміри висоти контейнера зображення з № 1. Виконання цієї дії проілюстроване на рис. 7 c).

3. Змінити розміри ширини контейнера зображення з № 2. Виконання цієї дії проілюстроване на рис. 7 d).

4. Змінити розміри ширини контейнера зображення з № 3. Виконання цієї дії проілюстроване на рис. 7 e).

5. Перемістити по горизонталі вправо контейнер зображення з № 3. Виконання цієї дії проілюстроване на рис. 7 f).

6. Перемістити по горизонталі вправо контейнер зображення з № 2. Виконання цієї дії проілюстроване на рис. 7 g).

7. Змінити розміри висоти контейнера зображення з № 4. Виконання цієї дії проілюстроване на рис. 7 h).

8. Перемістити по вертикалі вниз контейнер зображення з № 4. Виконання цієї дії проілюстроване на рис. 7 i).

9. Змінити розміри висоти контейнера зображення з № 2. Виконання цієї дії проілюстроване на рис. 7 j).

10.Змінити розміри висоти контейнера зображення з № 3. Виконання цієї дії проілюстроване на рис. 7 k).

Після виконання цих дій, бачимо результат наведений на рис. 7 k).

Аналогічно Canva, використання Pixlr Editor, після зміни розмірів елемента колажу теж необхідно виконання 10 описаних дій.

На рис. 8 b), представлено зміну розмірів елемента колажа з № 1 у Деревовидній моделі колажів. Після руху вертикального повзунка між контейнерами з № 1 та № 2, в праву сторону, можемо побачити, що ширина контейнера з розміром № 1 автоматично збільшилась, а ширина контейнера з № 2 автоматично зменшила, а після руку горизонтального повзунка між першим (контейнери

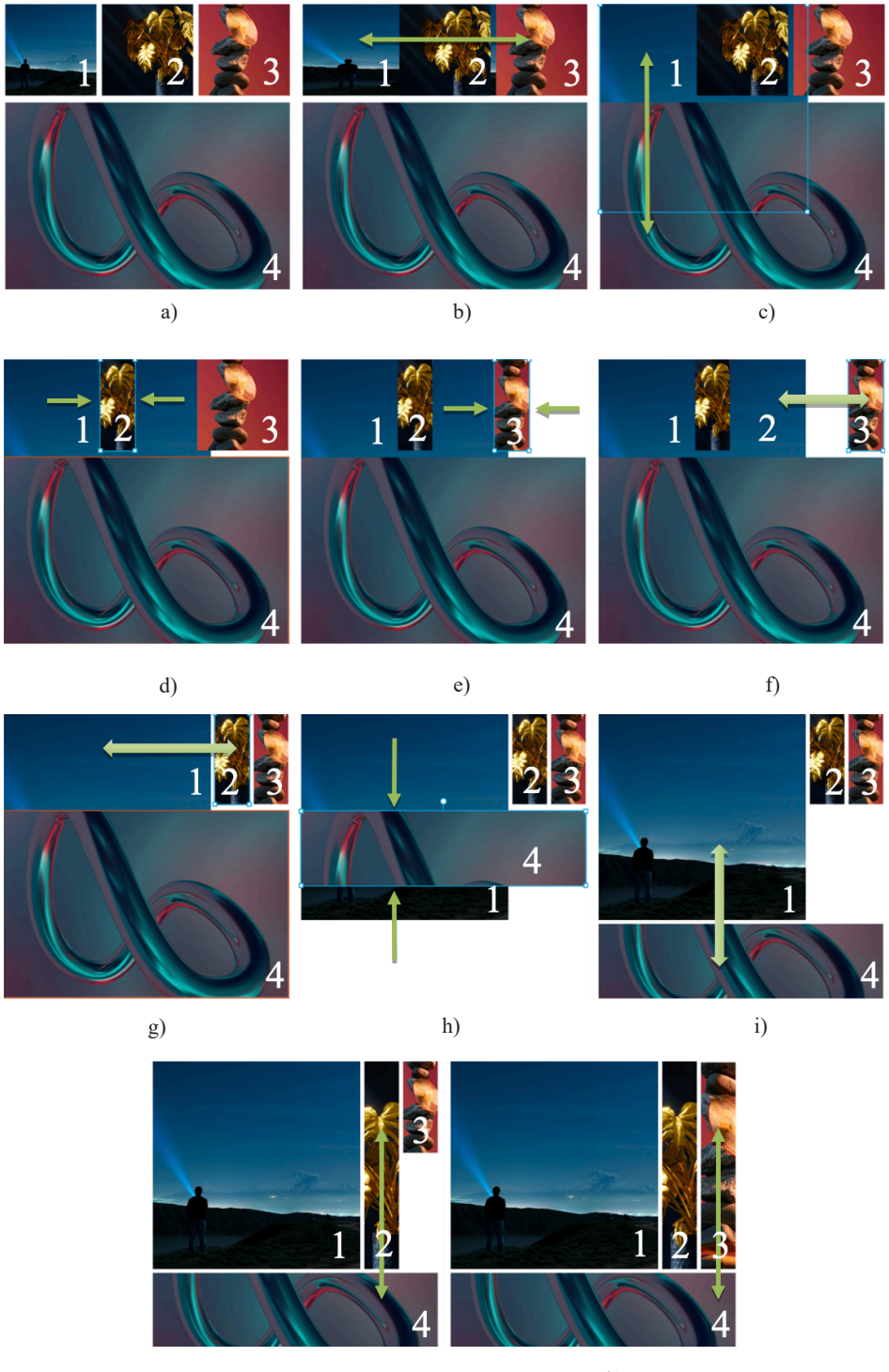


Рис. 7. Зміна елементів колажа в моделі колажів Pixlr Editor

з № 1, № 2, № 3) і другим рядком (контейнер з № 4), висота контейнерів з № 1, 2 та 3 збільшилась, а висота контейнера з № 4 автоматично зменшилась. Тобто, для зміни розмірів контейнера у деревовидній моделі колажів з уникненням накладання одного контейнера на інший, необхідно виконати наступні кроки:

1. Необхідно потягнути повзунок між контейнером зображення з № 1 та № 2. Виконання цієї дії проілюстроване на рис. 8 b).

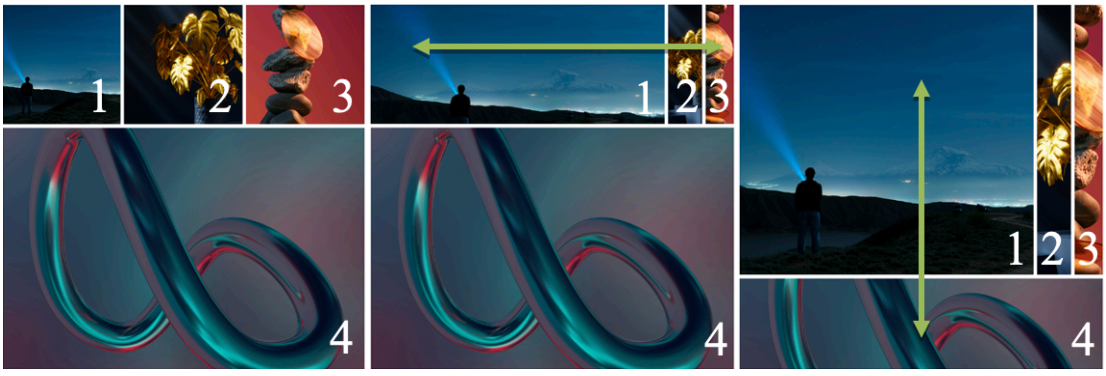
2. Необхідно потягнути повзунок між першим і другим рядками. Виконання цієї дії проілюстроване на рис. 8 c).

В порівнянні з Canva і Pixlr Editor, дія 1, 3, 4, 5 та 6, а саме необхідність окремої зміни розмірів ширини та переміщення по горизонталі контейнерів зображення з № 2 та № 3 виконується першою дією в Деревовидній моделі колажів. А дія 2, 7, 8, 9 та 10, а саме необхідність окремої зміни висот та переміщення по вертикалі контейнерів зображення з № 2, № 3 та № 4, виконується другою дією в Деревовидній моделі колажів. Отже, завдяки Деревовидній моделі колажів потрібно зробити на 8 дій менше, не потрібно змінювати окремо кожен контейнер, оскільки Деревовидна модель колажів здійснює ці дії автоматично.

Результати дослідження демонструють суттєву перевагу розробленої деревовидної моделі колажів у порівнянні з традиційними методами редагування, представленими у популярних системах Canva та Pixlr Editor. Встановлено, що в цих системах для запобігання накладання контейнерів необхідно вручну виконати до 10 окремих операцій, включаючи зміну розмірів та переміщення кожного елемента композиції окремо. Це значно ускладнює роботу та збільшує витрати часу. Натомість деревовидна модель дозволяє здійснювати аналогічні зміни автоматично за допомогою двох простих дій – руху вертикального та горизонтального повзунків. Такий підхід дозволяє одночасно коригувати розміри та розташування всіх взаємопов'язаних контейнерів, що суттєво оптимізує процес редагування та знижує трудомісткість у 5 разів. Це сприяє покращенню точності та якості графічних композицій, зберігає логічні зв'язки і пропорції між елементами, автоматизує рутинні дії користувача. Розроблена деревовидна модель може бути ефективно впроваджена в сучасні онлайн-платформи, підвищуючи їх продуктивність, зручність користування та загальну функціональність.

Висновки. У статті створено нову деревовидну модель побудови і опрацювання колажів, яка дозволяє досягти високої гнучкості та узгодженості при створенні і редагуванні графічних композицій. Розроблена модель забезпечує автоматичне пропорційне змінення пов'язаних елементів колажу: будь-які локальні зміни (розміру контейнера зображення або групи зображень) автоматично впливають на інші елементи, пов'язані через ієрархічну структуру. Це дає змогу користувачеві вносити зміни до композиції мінімальними зусиллями – без необхідності вручну підганяти сусідні частини колажу, як це має місце в традиційних моделях.

Розроблена інформаційна система колажів на базі цієї моделі продемонструвала свою ефективність. Деревовидна структура забезпечує цілісність композиції навіть для складних макетів із багаторівневою вкладеністю елементів. Порівняльний аналіз показав, що використання створеної моделі скорочує кількість дій при редагуванні колажу в середньому вп'ятеро у порівнянні з редагуванням у моделях без зв'язків між елементами. Це не лише прискорює процес, але й зменшує імовірність помилок та підвищує якість кінцевого результату.



a) b) c)

Рис. 8. Зміна елементів колажа в Деревовидній моделі колажів

Подальші дослідження можуть бути спрямовані на розширення можливостей моделі. Зокрема, цікавим є поєднання деревовидної моделі з автоматизованими алгоритмами розміщення зображень: наприклад, на основі аналізу вмісту зображень або їхніх метаданих (час, місце зйомки) автоматично формувати початкову структуру колажу з кластеризацією тематично близьких зображень, а надалі дозволяти інтерактивне тонке налаштування через описані моделі. Також перспективним є дослідження адаптації моделі до колажів довільної форми (не тільки прямокутної сітки) та інтеграція з існуючими інструментами графічного дизайну. Розроблена модель може стати основою для створення нових функціональних редакторів колажів, що поєднують автоматизацію і творчий підхід, забезпечуючи користувачам інтуїтивне та ефективне середовище для роботи з великою кількістю зображень.

Список літератури:

1. Adobe Photoshop. (n.d.). Photoshop.adobe.com. https://photoshop.adobe.com/discover
2. Angular. 2025. Angular.dev. https://angular.dev/api
3. Canva. 2025. Canva. https://www.canva.com/
4. Canvas API. 2019. MDN Web Docs. https://developer.mozilla.org/en-US/docs/Web/API/Canvas_API
5. JSDoc: Home. (n.d.). Fabricjs.com. http://fabricjs.com/docs/
6. Mozilla. 2019. WebGL: 2D and 3D graphics for the web. MDN Web Docs. https://developer.mozilla.org/en-US/docs/Web/API/WebGL_API
7. Online Photo Editor | Pixlr Editor. 2019. Pixlr.com. https://pixlr.com/editor/
8. Rother C., Bordeaux L., Hamadi Y., Blake A.. AutoCollage. ACM *Transactions on Graphics*. 2006. № 25(3). С. 847–852. DOI: 10.1145/1141911.1141965
9. Fernández D. P., Postels J., Tombari F. Variational Transformer Networks for Layout Generation. 2021. DOI: 10.1109/cvpr46437.2021.01343
10. Chen L., Jing Q., Tsang Y., Zhou T. Iris: a multi-constraint graphic layout generation system. *Frontiers of Information Technology & Electronic Engineering.* 2024. № 25 (7). С. 968–987. DOI: 10.1631/fitee.2300312
11. Huo H., Wang F. A Study of Artificial Intelligence-Based Poster Layout Design in Visual Communication. *Scientific Programming*. 2022. № 2022. С.1–9. DOI: 10.1155/2022/1191073
12. Cao Y., Ma Y., Zhou M., Liu C., Xie H., Ge T., Jiang, Y. Geometry Aligned Variational Transformer for Image-conditioned Layout Generation. *Proceedings of the 30th ACM International Conference on Multimedia.* 2022. № 1. С. 1561–1571. DOI: 10.1145/3503161.3548332
13. Lu Y., Tong Z., Zhao Q., Oh Y., Wang B., Li T. J.-J. Flowy: Supporting UX Design Decisions Through AI-Driven Pattern Annotation in Multi-Screen User Flows. *ArXiv (Cornell University).* 2024. DOI: 10.48550/arxiv.2406.16177
14. Shabani M. A., Wang Z., Liu D., Zhao N., Yang J., Furukawa Y. Visual Layout Composer: Image-Vector Dual Diffusion Model for Design Layout Generation. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* 2024. С. 9222–9231. DOI: 10.1109/cvpr52733.2024.00881
15. Shi Y., Shang M., Qi Z. Intelligent layout generation based on deep generative models: A comprehensive survey. Information Fusion. 2023. № 100. С. 101940–101940. DOI: 10.1016/j.inffus.2023.101940
16. Hadi AlZayer, Lin H., Bala K. AutoPhoto: Aesthetic Photo Capture using Reinforcement Learning. *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2021. С. 944–951. DOI: 10.1109/iros51168.2021.9636788
17. Zhou M., Xu C., Ma Y., Ge T., Jiang Y., Xu W. Composition-aware Graphic Layout GAN for Visual-Textual Presentation Designs. *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*. 2022. С. 4995–5001. DOI: 10.24963/ijcai.2022/692
18. Zhou Y., Jing Q., Li Z., Shi L., Chen L., Sun L. Element-Conditioned Gan for Graphic Layout Generation. 2023. DOI: https://dx.doi.org/10.2139/ssrn.4648075

Kordiuk R.I. TREE-LIKE COLLAGE MODEL OF INFORMATION SYSTEM

A comprehensive analysis of existing collage models utilized in popular information systems such as Canva, Pixlr Editor, and Online Photoshop was conducted. The study revealed that current collage models lack structured relationships between constituent elements. This significantly limits capabilities for editing and flexibly modifying images within a single collage.

A new collage model based on a tree-like structure has been developed. The proposed model features four levels of nesting. The root of the tree is the primary element, positioned at the first level, defined by maximum predetermined dimensions for the graphic image, with the possibility to set margins from the outer boundaries

of the graphic image and a zero rounding coefficient that can be dynamically adjusted. At the second level of nesting are rows and horizontal sliders. Columns or image containers, along with vertical sliders between them, are located within rows at the third level. The fourth level of nesting includes image containers situated inside columns, separated by horizontal sliders.

The introduced interactive relationships between collage elements form a complex and highly functional structure. Each tree element adjusts proportionally relative to other elements, ensuring that modifications to one element automatically reflect across all associated elements. This offers users exceptional flexibility in creating and editing graphical compositions, as changes made in one part of the collage automatically adapt throughout the entire structure.

A comparative analysis between the developed tree-like model and existing collage models demonstrated significant advantages. Implementing the tree-like model optimizes the collage workflow, reducing the number of required actions by a factor of five. This not only decreases editing time but also enhances the accuracy and quality of the final product, which is crucial when creating complex graphical compositions.

Key words: *collage, tree-like collage model, graphic image, image container.*

УДК 004.421.2:004.415
DOI https://doi.org/10.32782/2663-5941/2025.3.2/32

Кропивницька В.Б.
Івано-Франківський національний технічний університет нафти і газу

Кропивницький Д.Р.
Івано-Франківський національний технічний університет нафти і газу

МОДЕЛЮВАННЯ РОБОТИ МУРАШИНИХ АЛГОРИТМІВ У ЗАДАЧАХ МАРШРУТИЗАЦІЇ

У статті представлено моделювання та порівняльний аналіз трьох варіантів алгоритму мурашиної колонії (ACO) для задач маршрутизації: класичного ACO, Elitist ACO (елітарного) та Min-Max ACO. Розглянуто принципи функціонування ACO-алгоритмів, що базуються на поведінці мурах у природі, зокрема на механізмах феромонного маркування маршрутів і позитивного зворотного зв'язку. Увагу приділено механізму глобального оновлення феромонів, правилам вибору маршруту, ролі елітних агентів, а також обмеженням на значення феромонної концентрації, які суттєво впливають на поведінку алгоритмів у процесі оптимізації. Для трьох розглянутих варіантів ACO сформовано порівняльну таблицю за критеріями: спосіб оновлення феромонів, наявність посилення найкращого рішення, обмеження феромону, швидкість збіжності та ризик передчасної конвергенції. Проведено серію експериментів із використанням змодельованих мереж різної складності, у межах яких змінювались розмір колонії, кількість ітерацій та інші параметри. Результати дослідження наведено у вигляді таблиць і графіків, які ілюструють вплив параметрів на ефективність побудови маршрутів та поведінку алгоритмів у різних умовах. Встановлено, що класичний алгоритм ACO забезпечує стабільність при малих масштабах задачі, алгоритм Elitist ACO забезпечує високу точність і збіжність при помірній кількості агентів, тоді як Min-Max ACO демонструє найвищу стабільність і стійкість до змін параметрів завдяки введенню обмежень на феромон. Аналіз показав, що вибір конкретної модифікації алгоритму повинен ґрунтуватися на характеристиках мережі, ресурсних обмеженнях системи та необхідному рівні точності, що підкреслює практичну значущість дослідження для задач маршрутизації в динамічних середовищах сучасних обчислювальних систем, зокрема в телекомунікаційних, мобільних, безпілотних та сенсорних мережах із високим ступенем мобільності та змінною топологією.

Ключові слова: *алгоритм мурашиної колонії, маршрутизація, динамічні мережі, самоорганізація, моделювання, адаптивні алгоритми, роєвий інтелект, параметри алгоритму, програмна реалізація.*

Постановка проблеми. У сучасних телекомунікаційних, сенсорних, мобільних та безпілотних мережах спостерігається стале зростання потреби у гнучких і адаптивних методах маршрутизації, здатних ефективно працювати за умов динамічної топології та обмежених обчислювальних ресурсів. Традиційні алгоритми виявляються недостатньо ефективними або занадто інертними для задоволення вимог таких середовищ. За таких умов алгоритми мурашиної колонії (ACO), що належать до класу методів роєвого інтелекту, демонструють високу здатність до адаптації, масштабованості та знаходження якісних рішень у задачах складної оптимізації.

Їхня ефективність особливо проявляється у сфері маршрутизації в мережах WSN, MANET, FANET, VANET і IoT, де потрібна швидка реакція на зміни й оптимальне використання ресурсів. Це визначає актуальність розробки та порівняльного аналізу різних модифікацій алгоритмів ACO для задач комбінаторної оптимізації на графах, зокрема задачі маршрутизації.

Аналіз останніх досліджень і публікацій. У рамках дослідження ефективності мурашиних алгоритмів у задачах маршрутизації варто зазначити, що останні роки характеризуються зростанням інтересу до їх застосування в умовах динамічних мереж. Зокрема, у роботі [1] запропоновано модифікований протокол маршрутизації для FANET, у якому алгоритм мурашиної колонії використовується для вибору маршрутів з урахуванням параметрів якості зв'язку та енергоспоживання, що забезпечує низькі накладні витрати та високу ефективність у динамічних мережах. У дослідженні Шпорта і Лаврінця [2]

розглянуто алгоритм AntHocNet, який у ряді метрик (PDR, затримка, продуктивність) перевершує класичні рішення. Значну практичну цінність має також дослідження ACO у сфері інтелектуального регулювання дорожнього руху [3], де алгоритм застосовано для адаптивного керування світлофорами – що підтверджує універсальність ACO підходу у задачах реального часу. У роботі [4] автори запропонували модифікацію ACO (ACORA-WR) для балансування навантаження в супутникових LEO-мережах, де обмеження вікна пошуку дозволяє підвищити ефективність при зниженні обчислювальних витрат. Водночас, у публікації [5] представлено алгоритм AntSynNet, що застосовується для синтезу мережевих структур, орієнтованих на оптимізацію маршрутизації з урахуванням вартості та трафіку. Окремо заслуговує на увагу новітній підхід FAOA – алгоритм фермерських мурах [6], який демонструє конкурентні результати порівняно з іншими метаевристичними методами (GA, PSO), розширюючи рамки біологічно натхнених алгоритмів оптимізації. Високу ефективність мурашині алгоритми демонструють при оптимізації розподілених нестаціонарних систем. Авторами роботи [7] запропоновано адаптацію ACO-алгоритму для задачі комівояжера, де час переміщення між містами задано у вигляді нечітких трапецієподібних чисел. Запропонований підхід дозволяє отримувати близькі до оптимальних маршрути навіть за невеликої кількості ітерацій, що демонструє ефективність ACO у задачах з невизначеними параметрами та нечіткістю вхідних даних. Яскравим прикладом може служити знаходження мурашиними алгоритмами оптимальних трафіків в телекомунікаційних мережах [8]. У роботі [9] на прикладі завдання маршрутизації показано, як в алгоритми вирішення дискретних завдань оптимізації упровадити складові самоорганізації мурашок: випадковість, багатократність взаємодії, негативні і позитивні зворотні зв›язки. Враховуючи активний розвиток ACO-алгоритмів і розмаїття сфер їх застосування, доцільним є узагальнене дослідження їх ефективності на прикладі задач маршрутизації.

Постановка завдання. Метою роботи є дослідження та порівняльний аналіз ефективності трьох варіантів алгоритмів мурашиної колонії – класичного ACO, елітарного (Elitist ACO) та Min-Max ACO – у задачах маршрутизації на графах з різною складністю, а також моделювання їхньої роботи з візуалізацією процесу пошуку оптимального рішення.

Виклад основного матеріалу. Алгоритм мурашиної колонії (ACO) базується на спостереженнях за колективною поведінкою мурах, які, залишаючи феромонні сліди на маршрутах до джерела їжі, з часом зосереджуються на найкоротших шляхах завдяки позитивному зворотному зв'язку. Цей біологічний принцип покладено в основу низки алгоритмічних модифікацій, серед яких у даній роботі розглянуто три основні варіанти: класичний ACO, Elitist ACO (елітарний) та Min-Max ACO.

Ключовим етапом у роботі ACO-алгоритмів є механізм глобального оновлення феромонів, що реалізується після кожної ітерації та визначається за такою формулою (класичний алгоритм ACO):

$$\tau_{ij} = (1-p)\cdot\tau_{ij},$$

де τ_{ij} – є феромоном на компоненті між станом i і j і p, це параметр, який використовується для зміни рівня випаровування.

В елітарних системах ACO математична процедура оновлення визначається за такими рівняннями:

$$\Delta\tau_{ij}^{bs} = \begin{cases} \dfrac{Q}{C^{bs}}, & \text{якщо компонент належить} \\ & \text{елітній мкрасі} \\ 0, & \end{cases}$$

інакше

$$\tau_{ij} = \tau_{ij} + \sum_{k=1}^{M}\Delta\tau_{ij}^{k} + e\Delta\tau_{ij}^{bs}$$

де C^{bs} – загальна вартість рішення, Q – параметр для регулювання кількості феромонів (як правило дорівнює 1), e – параметр, який використовується для регулювання кількості додаткового феромону, що надається найкращому (або «елітному») розчину.

Алгоритм Min-Max подібний до елітарного ACO тим, що зосереджується на посиленні найбільш ефективних маршрутів, знайдених під час пошуку. Однак у Min-Max замість того, щоб просто надавати додаткову вагу елітним рішенням, лише найкраще поточне або найкраще глобальне рішення дозволяє відкладати феромонний слід. Крім того, Min-Max вимагає, щоб сліди феромонів знаходились в межах між максимальним і мінімальним значенням.

Таблиця 1 узагальнює ключові відмінності між трьома варіантами алгоритму мурашиної колонії.

Таблиця 1

Порівняльна характеристика алгоритмів ACO

Параметр	ACO	Elitist ACO	Min-Max ACO
Хто оновлює феромони?	Усі мурашки	Усі + краща мурашка	Тільки кращі мурашки
Додаткове посилення	Немає	Є	Немає
Обмеження феромонів	Немає	Немає	Так ($\tau_min \le \tau \le \tau_max$)
Швидкість збіжності	Помірна	Висока	Помірна стабільна
Ризик передчасної збіжності	Середній	Високий	Низький

На рис. 1 представлена узагальнена блок-схема мурашиного алгоритму.

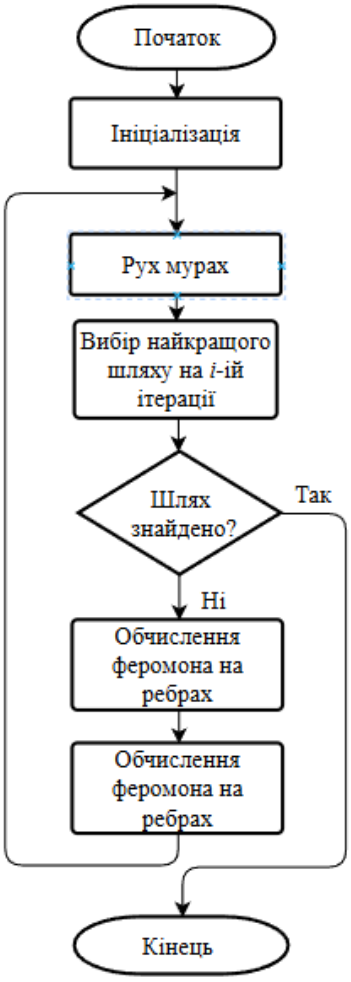


Рис. 1. Узагальнена блок-схема мурашиного алгоритму

Для того, щоб реалізувати відповідний мурашковий алгоритм для вирішення задачі маршрутизації, потрібно:

1. Зобразити задачу як граф, де вершини – це окремі компоненти (наприклад, вузли мережі), а ребра – можливі шляхи з відповідними вагами (наприклад, відстань, енерговитрати), якими можуть рухатися мурахи при побудові маршрутів.

2. Визначити значення сліду феромону.

3. Сформулювати евристичне правило, яке керує вибором мурашки на кожному етапі маршруту, враховуючи параметри мережі (наприклад, відстань, якість з'єднання тощо).

4. Якщо дозволяють умови задачі, застосувати локальний пошук для уточнення маршруту, тобто спробувати покращити вже побудований шлях, змінюючи його невеликі частини.

5. Визначити, яку саме модифікацію мурашиного алгоритму застосувати, і реалізувати її для розв'язання задачі.

6. Встановити оптимальні параметри для обраного алгоритму мурашиної колонії. Необхідно також визначити: кількість мурашок, що беруть участь у пошуку; співвідношення між дослідженням нових маршрутів і використанням вже знайдених; час та правилом оновлення значення феромону.

Програма, яка моделює роботу алгоритмів мурашиної колонії (ACO): має структуру показану на рисунку 2 [10]. Схема відображає взаємозв'язки між основними компонентами програми:

– Canvas – відповідає за графічне відображення карти, міст, ребер та мурашок.

– Graph – структура, яка зберігає всі міста та з'єднання між ними (ребра з феромонами).

– AntColony – координує роботу всієї колонії мурашок, викликає створення мурашок та оновлення феромонів.

– Ant – окрема мурашка, яка проходить шлях між містами та формує маршрут.

– Tour – клас, що описує маршрут, який пройшла мурашка, з обчисленням загальної довжини.

– ACArtist – модуль, що відповідає за візуалізацію руху мурашок та процесу оптимізації.

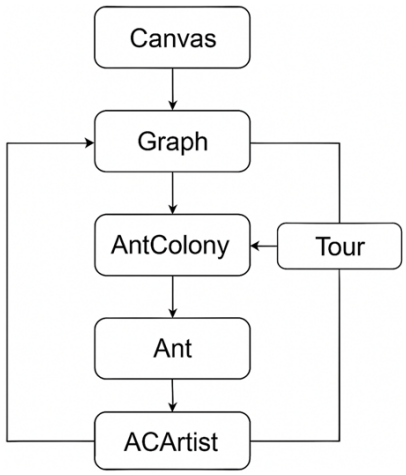


Рис. 2. Структура взаємодії модулів програми

Для дослідження і порівняння трьох мурашиних алгоритмів за заданою топологією мережі було вибрані такі параметри:

– розмір колонії мурах змінювався від 10 до 200;

– кількість ітерацій змінювалась від 50 до 500;

– значення коефіцієнтів початкової концентрації феромона: $\alpha = 2$, $\beta = 2$ (α – коефіцієнт значимості концентрації феромона; β – коефіцієнт значимості емпіричної інформації);

– значення початкової ймовірності переходу мурах від одного вузла до іншого $p_0 = 0{,}1$;

– початковий феромон – 2;

– вага феромонного депозиту – 1.

Для мережі, яка складається з 7 вузлів (рис. 3) було проведено серію дослідів при фіксованій кількості ітерацій, яка дорівнює 100 і при різних значеннях розміру колоній мурах: 10, 20, 30, 50, 100, 200.

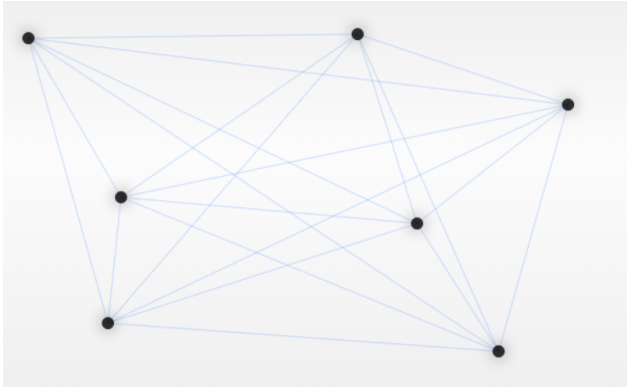


Рис. 3. Топологія мережі з 7 вузлів

У результаті моделювання було отримано дані, наведені в таблиці 2.

Таблиця 2

Результати моделювання для мережі з 7 вузлів

Розмір колонії	Найкоротший шлях		
	ACO	Elit	Min-Max
10	1583,67	1583,67	1583,67
20	1583,67	1583,67	1583,67
30	1583,67	1583,67	1583,67
50	1583,67	1583,67	1583,67
100	1583,67	1583,67	1583,67
200	1583,67	1594,16	1583,67

Класичний алгоритм мурашиної колонії ACO та алгоритм Min-Max дають повністю стабільний результат при будь-якому розмірі колонії – 1583.67. Це свідчить про їхню високу стабільність при малих розмірах колонії. Елітарний алгоритм показує ідентичні результати до ACO до розміру колонії 100. Проте при збільшенні кількості мурах вартість маршруту зростає до 1594.16, що може вказувати на перенавчання, накопичення надлишкового феромону або деградацію ефективності при занадто великій популяції.

Друга серія експериментів була проведена для мережі, яка складається з 30 вузлів. При цьому було кількість мурах в колонії дорівнювала: 10, 20, 30, 50, 100, 200; а кількість ітерацій дорівнювала: 50, 100, 200, 300, 500.

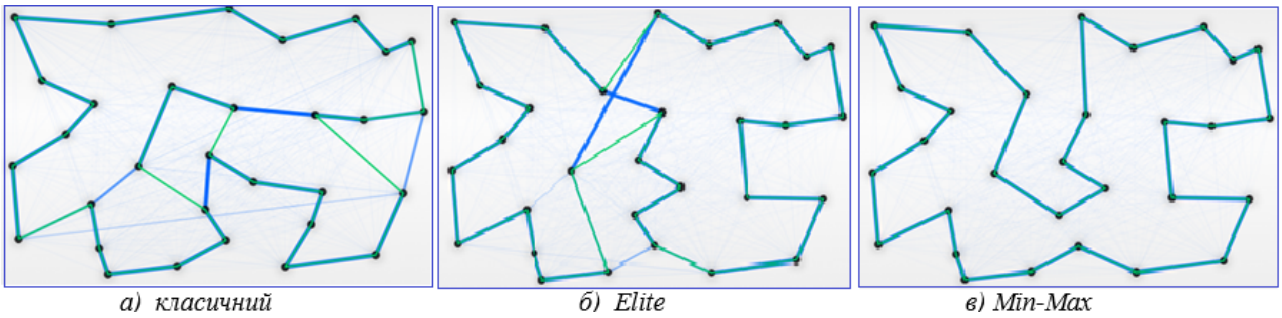
Результати моделювання занесено в таблицю 3. На рисунку 4 відображено графічне представлення результатів моделювання роботи трьох алгоритмів при кількості ітерацій N = 200 та величині колонії мурах 30.

Таблиця 3

Результати моделювання для 30 вузлів

Розмір колонії	Кількість ітерацій	Найкоротший шлях		
		ACO	Elit	Min-Max
50	50	3410,39	3224,65	3013,57
50	100	3009,44	3008,24	3089,68
50	200	3021,59	2995,59	2988,07
50	300	3039,81	2954,54	2994,07
50	500	2953,34	2947,38	2968,65
10	200	3009,44	3006,32	3115,61
20	200	3062,42	2654,54	2975,07
30	200	2958,48	2995,59	2947,38
50	200	3021,59	2995,59	2988,07
100	200	3039,37	3104,93	3028,84
200	200	3086,09	3046,84	3100,89

Результати таблиці 3 узагальнено у вигляді графіків (рис. 5 та 6), що ілюструють вплив кількості ітерацій і розміру колонії на ефективність кожного з трьох алгоритмів. На графіку (рис. 5) зображено залежність якості знайденого маршруту від кількості ітерацій для трьох алгоритмів: ACO, Elit та Min-Max, при фіксованому розмірі колонії, що дорівнює 50.



а) класичний *б) Elite* *в) Min-Max*

Рис. 4. Найкоротший шлях. Алгоритм ACO

Рис. 5. Залежність якості знайденого маршруту від кількості ітерацій

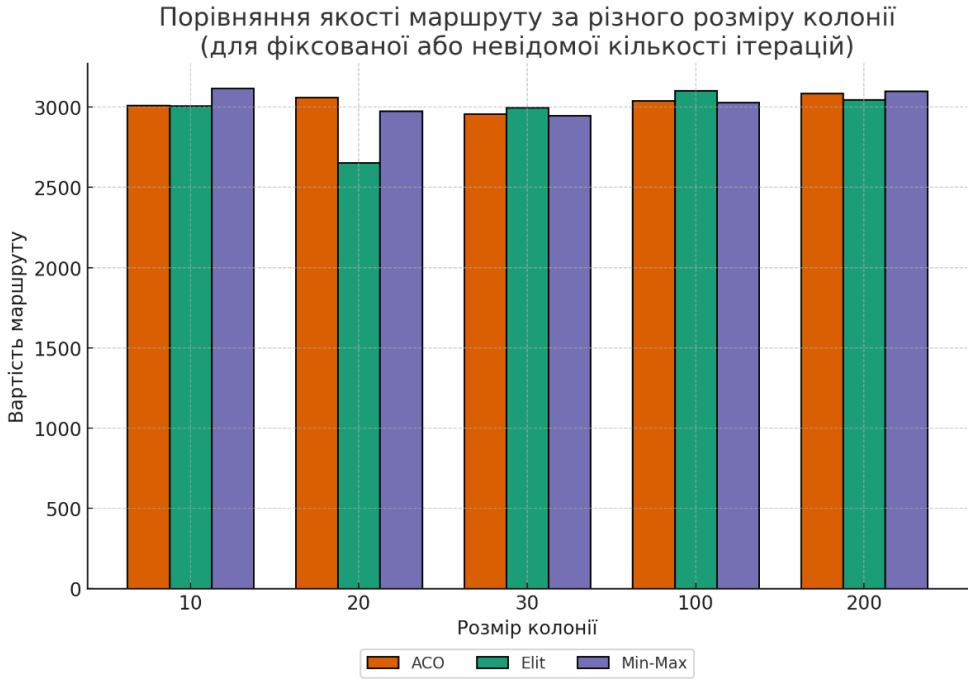


Рис. 6. Залежність якості маршруту від розміру колонії

На основі графічної інтерпретації можна зробити висновок, що алгоритм ACO при малій кількості ітерацій (50) демонструє найгірший результат, формуючи найдовший маршрут. Проте зі збільшенням кількості ітерацій до 500, якість його рішень поступово покращується, хоча й не досягає рівня інших алгоритмів. Алгоритм Elit виявляє стабільну тенденцію до зниження вартості маршруту при зростанні кількості ітерацій і вже з 200 ітерацій випереджає інші варіанти за ефективністю. Алгоритм Min-Max спочатку демонструє коливання, однак також поступово покращує результат і наближається до Elit при великій кількості ітерацій.

Як показує діаграма (рис. 6), збільшення кількості мурах у колонії призводить до незначного покращення результатів для алгоритму ACO. Алгоритм Elit, навпаки, вже при невеликому розмірі колонії досягає стабільно хороших результатів, і подальше збільшення кількості агентів майже не впливає на покращення. Найбільш виражену позитивну динаміку демонструє Min-Max: при збільшенні розміру колонії від 10 до 50 вартість маршруту значно знижується, що свідчить про здатність алгоритму ефективніше знаходити рішення в умовах більшої популяції. Таким чином, діаграма демонструє, що розмір колонії впливає на ефективність алгоритмів, але Elit виявляється найстабільнішим і найрезультативнішим у більшості випробувань. Це може свідчити про його здатність ефективно адаптуватися до задачі вже на ранніх ітераціях і з невеликим числом агентів.

Висновки. У результаті проведеного дослідження було встановлено, що всі три розглянуті варіанти алгоритму мурашиної колонії – класичний ACO, елітарний Elitist ACO та Min-Max ACO – здатні ефективно вирішувати задачі маршрутизації, проте кожен з них має свої особливості. Класичний ACO забезпечує стабільні результати при невеликій складності задачі, але втрачає ефективність при збільшенні розмірів мережі або кількості ітерацій. Elitist ACO демонструє високу точність і швидку збіжність при невеликих розмірах колонії, однак при надмірному збільшенні кількості агентів може спостерігатися перенасичення феромонного поля, що знижує якість рішень. Min-Max ACO виявився найбільш стійким до змін параметрів, зберігаючи баланс між дослідженням нових маршрутів і використанням вже знайдених рішень, завдяки введенню обмежень на значення феромону. Моделювання показало, що оптимальні результати досягаються при правильному налаштуванні параметрів алгоритму з урахуванням розміру мережі, кількості ітерацій та розміру колонії. Найкращі результати при великій кількості ітерацій показав алгоритм Elitist ACO, тоді як Min-Max ACO виявився найбільш стабільним у широкому діапазоні умов. Таким чином, доцільність використання того чи іншого варіанта мурашиного алгоритму залежить від конкретних вимог до задачі маршрутизації, ресурсних обмежень системи та необхідного рівня точності.

Список літератури:

1. Wang, Z., Yu, X., Wang, B., Cheng, S. A Low-Overhead Routing Protocol for FANET Based on Ant Colony Algorithm. In: Gao, F., Wu, J., Li, Y., Gao, H. (eds) *Communications and Networking*. ChinaCom 2022. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Vol 500. Springer, Cham. 2023. P. 321–333. https://doi.org/10.1007/978-3-031-34790-0_25

2. Шпорт Д. В., Лаврінець К. Г. Експериментальна оцінка можливості використання алгоритму мурашиної колонії – AntHocNet для вирішення задачі маршрутизації у FANET. Зв'язок. 2024. № 6 (172). С. 53–63. DOI: https://doi.org/10.31673/2412-9070.2024.060501

3. Danyliuk A., Muliarevych O. Ant Colony Algorithm in Traffic Flow Control. *Advances in Cyber-Physical Systems*. 2024. Vol. 9, № 2. P. 158–163. https://doi.org/10.23939/acps2024.02.158

4. Deng, X., Zeng, S., Chang, L., Wang, Y., Wu, X., Liang, J., Ou, J., & Fan, C. (2021). An Ant Colony Optimization-Based Routing Algorithm for Load Balancing in LEO Satellite Networks. *Wireless Communications and Mobile Computing*. 2022 (1), 3032997. https://doi.org/10.1155/2022/3032997

5. Wilkinson, Simon & White, Tony. The AntSynNet Algorithm: Network Syntheses Using Ant Colony Optimization. *Proceedings of the International Conference on Artificial Intelligence*, IC-AI '04, Volume 2. Las Vegas, Nevada, USA. June 21–24, 2004. P. 648–663.

6. Asghari A, Zeinalabedinmalekmian M, Azgomi H, Alimoradi M, Ghaziantafrishi S. Farmer Ants Optimization Algorithm: A Novel Metaheuristic for Solving Discrete Optimization Problems. *Information*. 2025. 16 (3) : 207. https://doi.org/10.3390/info16030207

7. E. Ivohin, K. Yusнtin. Use of Ant Colony Optimization Algorithm for Solving Fuzzy Problem of Traveling Salesman. *Advanced Information Technology*, 2024. Vol.1 (3). P. 23–31. https://doi.org/10.17721/AIT.2024.1.03

8. Berliński M., Rasmus M., Kopertowski Z. and Kozdrowski S. Ant Colony Algorithms Application for Telco Networks Performance with Multi-criteria Optimization. *2023 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*. Split, Croatia. 2023. P. 1–3. DOI: 10.23919/SoftCOM58365.2023.10271586.

9. Dorigo M., Gambardella L. M. Ant Algorithms for Discrete Optimization. *Artificial Life.* № 5 (2). 1999. P. 137–172. Artificial Life. 5. 137–172. 10.1162/106454699568728.

10. Recent Tutorials. Веб-сайт: https://www.theprojectspot.com/tutorials (дата звернення: 10.04.2025)

Kropyvnytska V.B., Kropyvnytskyi D.R. MODELING THE PERFORMANCE OF ANT COLONY ALGORITHMS FOR NETWORK ROUTING TASKS

This article presents the modeling and comparative analysis of three variants of the Ant Colony Optimization (ACO) algorithm applied to routing tasks: the classical ACO, Elitist ACO, and Min-Max ACO. The paper explores the functioning principles of ACO algorithms, which are inspired by the natural behavior of ants–particularly mechanisms of pheromone trail marking and positive feedback. Special attention is paid to the process of global pheromone update, path selection rules, the role of elite agents, and the implementation of pheromone concentration limits, all of which significantly influence algorithm behavior during optimization. A comparative table is constructed for the three considered ACO variants based on the following criteria: pheromone update method, reinforcement of the best solution, pheromone bounds, convergence speed, and risk of premature convergence. A series of experiments was conducted using simulated networks of varying complexity, where colony size, iteration count, and other parameters were adjusted. The results are presented in the form of tables and graphs illustrating how parameter changes affect route-building efficiency and algorithm behavior under different conditions. The classical ACO algorithm showed stability on small-scale tasks, Elitist ACO achieved high accuracy and convergence with a moderate number of agents, while Min-Max ACO demonstrated the greatest stability and robustness to parameter changes due to the introduction of pheromone value limits. The analysis indicates that the selection of a specific algorithm modification should be based on network characteristics, system resource constraints, and required solution accuracy, highlighting the practical relevance of this research for routing tasks in dynamic environments of modern computing systems, particularly in telecommunications, mobile, unmanned, and sensor networks with high mobility and variable topology.

Key words: *ant colony algorithm, routing, dynamic networks, self-organization, modeling, adaptive algorithms, swarm intelligence, algorithm parameters, software implementation.*

УДК 004.942
DOI https://doi.org/10.32782/2663-5941/2025.3.2/33

Левчук І.Л.
Український державний університет науки і технологій

Дубовик Т.М.
Український державний університет науки і технологій

Гузь Г.М.
Український державний університет науки і технологій

Романчук О.О.
Український державний університет науки і технологій

РОЗРОБКА ЛЮДИНО-МАШИННОГО ІНТЕРФЕЙСУ СИСТЕМИ КЕРУВАННЯ НА БАЗІ SCADA ПРОЦЕСОМ ОТРИМАННЯ МІДНОГО КУПОРОСУ

Метою даної роботи є створення в програмному середовищі SCADA ефективного людино-машинного інтерфейсу (HMI) системи керування процесом отримання мідного купоросу. HMI котрий в реальному часі інформує оператора про перебіг керованого процесу, дозволяє спостерігати значення всіх цільових технологічних змінних в цифровій та графічній формі, дає можливість здійснювати ручне керування процесом та налаштовувати алгоритми керування. В статті наведена стисла характеристика основних параметрів комп'ютерної моделі процесу отримання мідного купоросу реалізованої в універсальній моделюючий програмі CHEMCAD, котрі взаємодіють з елементами людино-машинного інтерфейсу реалізованого на базі SCADA. Представлена загальна структура та вигляд запропонованого HMI, котрий складається з інформаційної панелі для відображення основних параметрів процесу в цифровій формі, панелі налаштування дозаторів призначеної для налаштування алгоритмів керування дозаторами та меж дозування, трендів для виводу в графічній формі основних цільових параметрів процесу а також трендів що у графічному вигляді відображають роботу підсистеми дозування. Розглянуті особливості створення та налаштування окремих елементів людино-машинного інтерфейсу (трендів) та наведений перелік і стисла характеристика всіх аргументів – змінних людино-машинного інтерфейсу, котрі забезпечують його функціонування та інформаційний обмін між елементами HMI. Працездатність запропонованого HMI перевірена і доведена шляхом порівняння отриманих перехідних характеристик процесу з даними реальної лабораторної установки для отримання розчину мідного купоросу. Наведений в статті аналіз отриманих залежностей дозволяє зробити висновок про повну функціональність розробленого HMI та досягнення мети роботи. Запропоновані в роботі технічні рішення, принципи та особливості розробки людино-машинного інтерфейсу мають універсальний характер і можуть застосовуватись для розробки або удосконалення систем керування подібними технологічними процесами, що побудовані на базі SCADA та математичних або комп'ютерних моделей.

Ключові слова: *система керування, людино-машинний інтерфейс, SCADA, мідний купорос.*

Постановка проблеми. Сучасне промислове виробництво не можливо без всебічного використання сучасних автоматизованих інформаційно-керуючих систем, котрі підвищують техніко-економічні показники керованих процесів, збільшують надійність роботи технологічного обладнання, поліпшують умови роботи персоналу та підвищують загальну ефективність і безпеку технологічних процесів. В той же час статистика аварій та інцидентів у промисловості, на транспорті, в складних автоматизованих системах свідчить, що доля «людського фактору» у первопричинах інцидентів безперервно зростає [1]. Причиною вказаних тенденцій є постійне удосконалення та збільшення надійності сучасних засобів автоматизації та використання застарілого підходу при розробці систем керування – часто розробники спрямовують увагу на розширення функціональних можливостей систем керування, намагаються підвищити ступінь

автоматизації, концентруються на застосуванні новітніх технічних досягнень і недооцінюють необхідність побудови ефективного людино-машинного інтерфейсу (HMI – Human-Machine Interface), тобто інтерфейсу спрямованого на взаємодію системи керування з оператором. Саме SCADA, забезпечуючи потужні можливості для розробки людино-машинних інтерфейсів систем керування, є осередком нового підходу при розробці автоматизованих систем, підходу спрямованого на людину-оператора та його задачі [2].

Сульфат міді (мідний купорос) є найважливішою сіллю міді, яка знаходить широке застосування в медицині, хімічній промисловості та сільському господарстві [3]. Продуктивність процесу отримання мідного купоросу значною мірою залежить від ефективності системи управління, для реалізації якої на базі SCADA потрібна математична модель процесу та система керування у складі:

1. Алгоритмів керуючої підсистеми;
2. Людино-машинного інтерфейсу системи керування;
3. Підсистеми зв'язку, що забезпечує інформаційний обмін в реальному часі між всіма компонентами системи керування.

Аналіз останніх досліджень і публікацій. Розроблена комп'ютерна модель динаміки процесу отримання розчину мідного купоросу, адаптована для вирішення задач керування розглянута у роботі [4]. Модель реалізована в програмному середовищі універсальної моделюючої програми CHEMCAD, котра має вбудований OPC сервер та дозволяє здійснювати інформаційний обмін в реальному часі між моделлю та зовнішнім (по відношенню до моделі) програмнім середовищем. Адекватність отриманої моделі авторами [4] перевірена та доведена шляхом порівняння отриманих результатів з даними реальної лабораторної установки. Принципи побудови зв'язку на базі технології OPC між комп'ютерною моделлю процесу реалізованою у CHEMCAD та системою керування на базі SCADA, зв'язку що забезпечує інформаційний обмін в реальному часі, розглянуті в роботі [5]. При проведені дослідження автори [5] запропонували вирішення проблеми блокування інформаційного обміну до закінчення розрахунку комп'ютерної моделі шляхом запуску розрахункового процесу у квазідинамічному режимі.

Постановка завдання. Метою статті є розробка ефективного людино-машинного інтерфейсу (HMI) системи керування на базі SCADA процесом отримання мідного купоросу, здатного вирішувати наступні задачі:

– надання оператору системи керування необхідної інформації про перебіг керованого технологічного процесу в зрозумілій людині формі (у вигляді цифрових значень на дисплеї керуючої ЕОМ, та у вигляді графічних залежностей, котрі показують зміну цільового технологічного параметра – тобто трендів);

– надання оператору системи керування можливості відстежувати стан системи керування, перемикати режими роботи підсистеми регулювання (включати автоматичний або ручний режим роботи), здійснювати ручне керування процесом, тобто вручну керувати виконавчими пристроями та обирати відсоток їх відкриття, змінювати завдання та налаштування регуляторів, що здійснюють безпосереднє керування процесом.

Виклад основного матеріалу. До параметрів, котрі характеризують перебіг процесу отримання розчину мідного купоросу належать [4]:

– вміст сульфату міді $CuSO_4$ в реакційному розчині в реакторі, грам/літр;

– вміст кислоти H_2SO_4 в реакційному розчині в реакторі, грам/літр;

– щільність розчину в реакторі, грам/см^3, при досягнені щільності 1,42 грам/см^3 починається відбір реакційного розчину на кристалізацію;

– температура реакційної суміші в реакторі, °С;

– об'єм рідини в реакторі, літрів.

До параметрів, що забезпечують функціональні можливості оператора системи належать:

– завдання максимального ступеня відкриття виконавчих пристроїв, 0–100%;

– завдання максимального та мінімального об'єму рідини в реакторі, при котрому зберігається нормальний перебіг технологічного процесу та в межах котрої працює система адаптивного дозування, 3–4,5 літрів;

– налаштування щільності розчину в реакторі, при досягнені якої відбувається відбір розчину на кристалізацію, грам/см^3;

– перемикання режиму функціонування дозаторів, ручний або автоматичний;

– індикація стану (відсоток відкриття) виконавчих пристроїв «6» та «9», %;

– індикація стану системи дозування, відбір розчину чи додавання для відновлення об'єму рідини в реакторі та кислотності реакційного середовища.

Для вирішення поставлених задач в програмному середовищі SCADA системи були розроблені елементи HMI котрі надають оператору системи керування інформацію про перебіг процесу отримання розчину мідного купоросу в цифровій формі та дозволяють відстежувати стан системи адаптивного дозування та за необхідності змінювати налаштування регуляторів, що безпосередньо керують процесом в автоматичному режимі – інформаційна панель процесу, та панель керування дозаторами.

Розроблені елементи HMI наведені на рисунку 1 та 2.

Рис. 1. HMI, інформаційна панель процесу

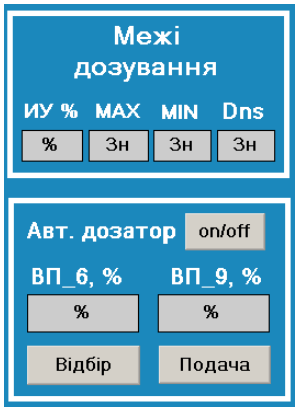


Рис. 2. HMI, панель керування дозаторами

Інформаційна панель процесу забезпечує відображенню в цифровій формі з зазначенням одиниць вимірювання вмісту CuSO4 та H2SO4 в реакційному розчині, щільність та температуру розчину, об'єм рідини в реакторі.

Налаштування параметрів регуляторів здійснюється шляхом натискання лівої кнопки маніпулятору «миша» в області відображення поточного значення відповідного параметра, після чого відкриється вікно в якому оператор системи може задати нове значення параметра (рис. 3).

Перемикання режимів функціонування регуляторів і ручний початок та закінчення дозування реалізовано за допомогою графічних кнопок розташованих у відповідній області HMI (кнопки «on/off» та «Відбір»–«Подача»). Для відображення стану підсистеми дозування використана кольорова індикація, при задіянні функції відповідна кнопка стає червоною (рис. 3).

Розроблений HMI надає можливість відстежувати найбільш критичні цільові параметри процесу в динаміці та в графічній формі в залежності від астрономічного часу, тобто у вигляді графіків – трендів.

Рис. 3. HMI, налаштування параметрів регуляторів

Відображенню у вигляді трендів підлягають:
- вміст сульфату міді $CuSO_4$ в розчині;
- вміст кислоти H_2SO_4 в розчині;
- щільність розчину;
- об'єм рідини в реакторі «1».

Для оптимального використання робочого простору екрану HMI, близькі за суттю параметри процесу згруповані попарно на одних координатних осях, «Купорос та кислота» (рис. 4) та «Рідина в реакторі, щільність» (рис. 5).

Діапазони значень по осям встановлені у відповідності до максимальних та мінімальних значень відповідних технологічних параметрів, тобто з урахуванням можливого діапазону зміни цих параметрів на виході комп'ютерної моделі. Вміст сульфату міді $CuSO_4$ в реакційному розчині «Купорос»: 0–600 грам/літр, вміст кислоти H_2SO_4 в реакційному розчині «Кислота»: 0–250 грам/літр; щільність реакційного розчину «Щільність»: 1.15–1.65 грам/см^3, об'єм рідини в реакторі «Рідина»: 2,5–4,5 літрів.

На рис. 6 наведені налаштування програмного середовища SCADA–системи TRACE MODE (написи, кольори ліній, масштаб за параметром та часом, алгоритм інтерполяції значень та ін.), котрі забезпечують функціонування тренду «Купорос та кислота».

Для забезпечення відображення тренду у вигляді нерозривної кривої та проведення лінії від поточного значення каналу до значення каналу на попередньому кроці обробки, обраний метод інтерполяції значень «За періодом реального часу». Аналогічно налаштовується тренд «Рідина в реакторі, щільність».

Підсистема аргументів (внутрішніх змінних) SCADA, що забезпечує інформаційний обмін між всіма елементами розробленого людино-машин-

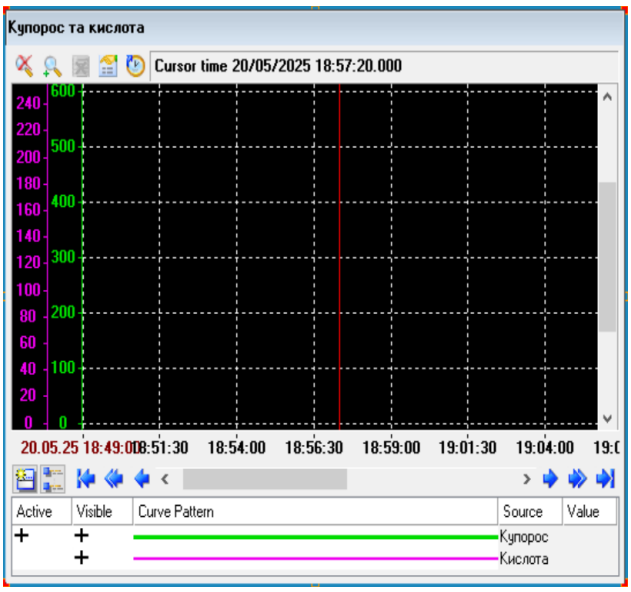


Рис. 4. HMI, тренд «Купорос та кислота»

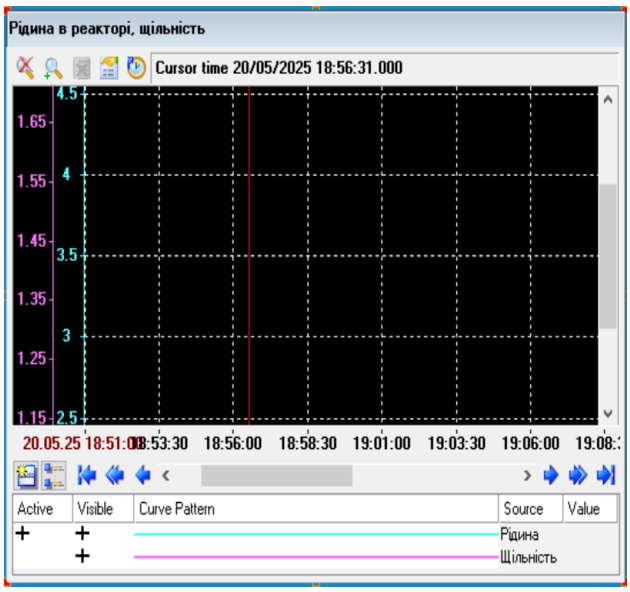


Рис. 5. HMI, тренд «Рідина в реакторі, щільність»

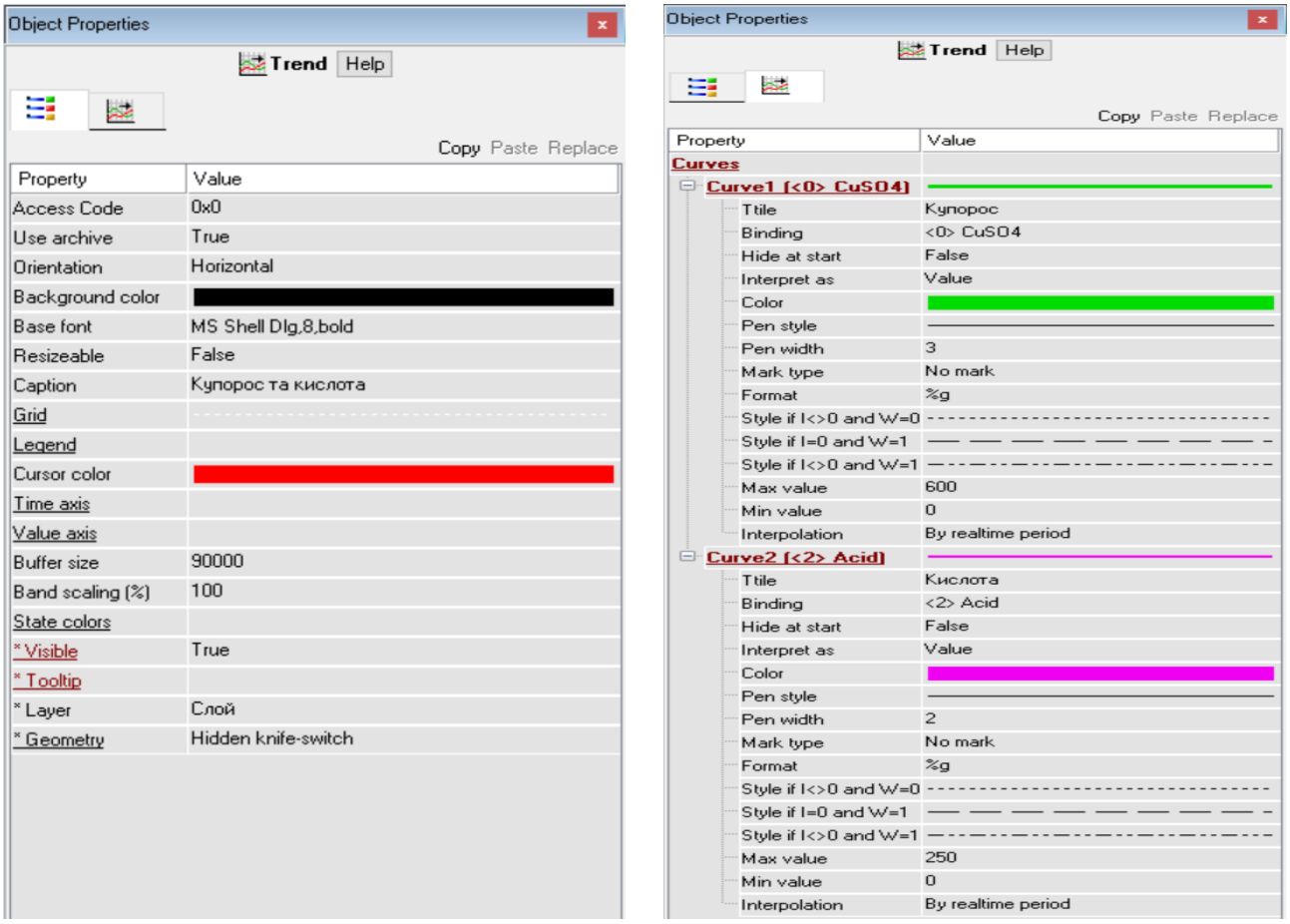


Рис. 6. Налаштування тренду «Купорос та кислота»

ного інтерфейсу системи керування процесом отримання мідного купоросу та всіма її підсистемами в реальному часі наведена на рисунку 7.

Аргументи CuSO4, Acid, Liquid_vol, Liquid_dens, Liquid_temp відповідають за вміст сульфату міді, кислоти та об'єм, щільність і температуру розчину в ректорі.

Призначенням аргументів Valve_... є зв'язок з виконавчими пристроями, встановлення та моніторинг за допомогою зворотного зв'язку

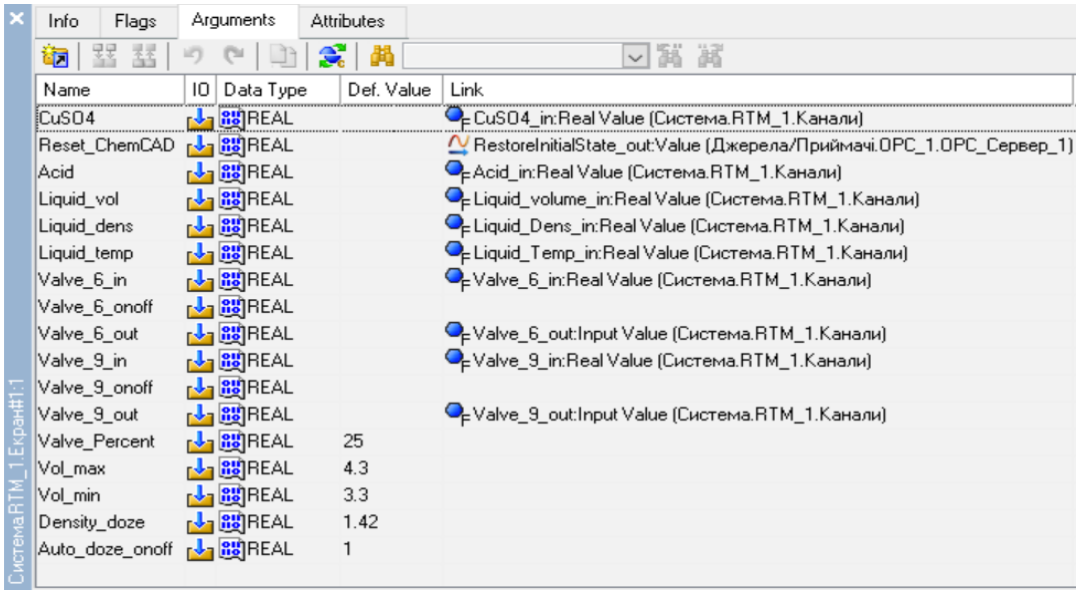


Рис. 7. HMI, аргументи (внутрішні змінні)

положення регулюючих органів, встановлення відсотку відкриття або закриття виконавчих пристроїв.

Аргументи Density_doze, Auto_doze, Vol_max та Vol_min використовуються для налаштування алгоритму адаптивного дозування, котрий не розглядається у межах цієї роботи.

За допомогою розділу «Def. Value» за потреби для аргументів задаються значення за замовчуванням. У розділі «Link» задається джерело, з якого аргумент отримує данні – інформаційний канал, або інший аргумент HMI.

Загальний вигляд розробленого людино-машинного інтерфейсу системи керування процесом отримання розчину мідного купоросу наведений на рисунку 8.

Вільний простір між блоками налаштувань «Інформаційна панель» та «Межі дозування» в майбутньому планується використати для розміщення налаштувань підсистеми керування температурою реакційного розчину в реакторі при продовженні даного дослідження в напрямку визначення впливу температури на перебіг процесу отримання розчину мідного купоросу. Реалізована на зараз система керування забезпечую ведення процесу при постійній, рекомендованій в довідковій літературі [6] температурі біля 80°C.

Для перевірки працездатності розробленого людино-машинного інтерфейсу системи керування процесом отримання мідного купоросу був виконаний тестовий «прогін» розробленої системи та отримані перехідні характеристики керованого процесу. Також виконаний порівняльний аналіз отриманих залежностей з даними реальної лабораторної установки для отримання розчину мідного купоросу, котра працює в «НДЛ Метрологічного забезпечення екологічного контролю зовнішнього середовища» ННІ УДХТУ, УДУНТ.

Графічні залежності (тренди), котрі демонструють зміну цільових параметрів процесу (вміст CuSO4 та H2SO4) у часі наведені на рисунку 9.

Аналіз отриманих результатів. Отримані та наведені на рисунку 9 залежності повністю відтворюють поведінку реального процесу отримання мідного купоросу. На початку технологічного процесу вміст кислоти в реакторі перевищує вміст купоросу, котрий доданий в реактор лише для підтримання реакції, відповідно H_2SO_4 = 188,37 грам/літр, $CuSO_4$ = 157,203 грам/літр. Але з часом у наслідок хімічних перетворень вміст кислоти зменшується а мідного купоросу зростає, відповідно зростає і щільність розчину в реакторі. На початок відбору рідини на кристалізацію концентрації дорівнюють H_2SO_4 = 49,58 грам/літр, $CuSO_4$ = 514,57 грам/літр. Відновлення об'єму рідини в реакторі шляхом додавання розчину 20% H_2SO_4 призводить до часткового відтворення кислотності розчину (H_2SO_4 = 96,68 грам/літр) та до зниження вмісту мідного купоросу ($CuSO_4$ = 359,39 грам/літр). Після чого робочий цикл установки для отримання розчину мідного купоросу повторюється.

З графіку перехідних процесів видно (рис. 9), що з кожним робочим циклом вміст кислоти в реакційному розчині знижується. Це цілком відповідає реальному процесу і пояснюється не

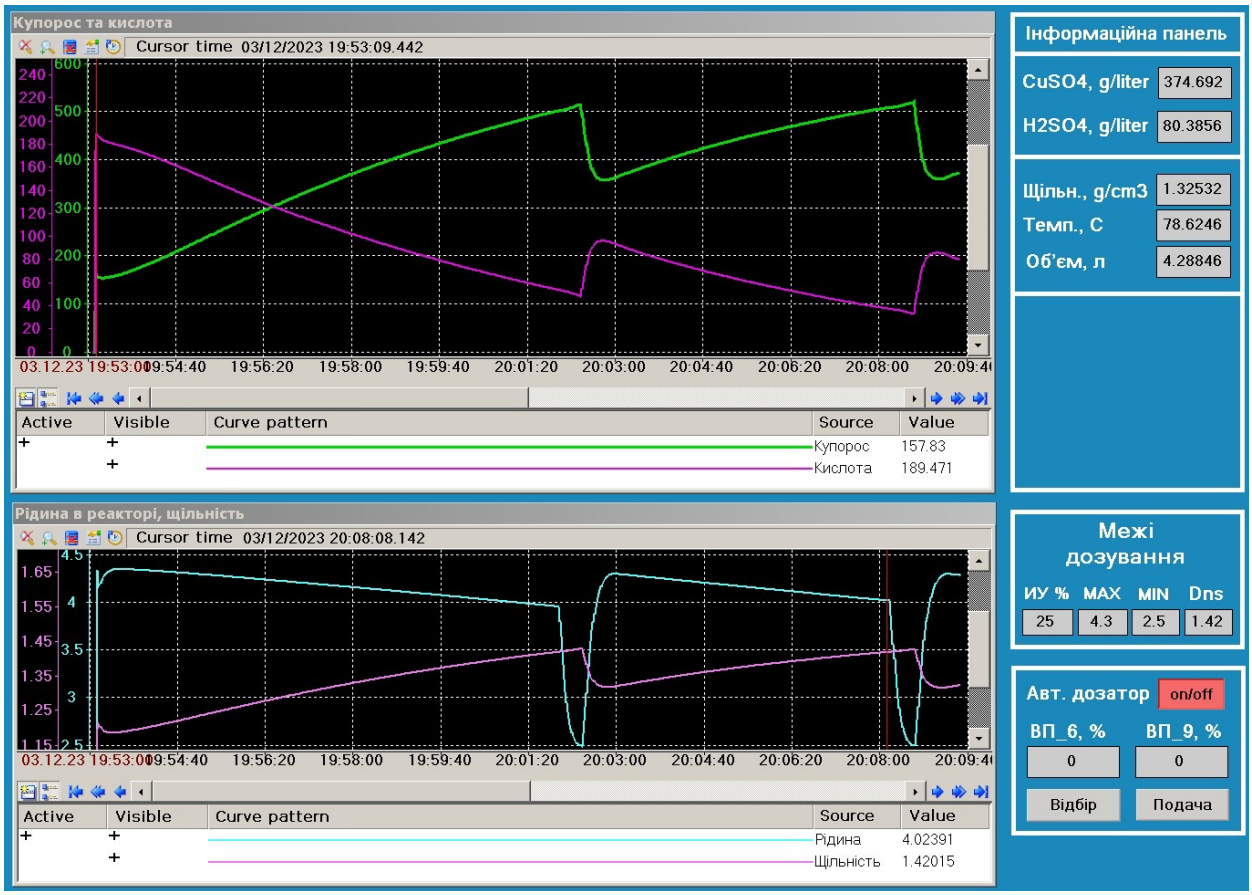


Рис. 8. HMI системи керування, загальний вигляд

Рис. 9. Перехідні характеристики процесу

можливістю повністю відновити вміст H_2SO_4 відбираючи лише частину, а не весь вміст реактору на кристалізацію. Частина розчину, завжди залишають в реакторі для прискорення хімічних перетворень в наступному робочому циклі, що можливо лише при наявності в реакційному розчині невеликого вмісту $CuSO_4$.

Збільшення об'єму розчину, котрий відбирається на кристалізацію, з одного боку збільшує повторюваність параметрів робочих циклів установки, відповідно збільшуючи загальну тривалість роботи, до повної заміни реакційної суміші, що є позитивним явищем. Але визначення точного співвідношення об'єму відбору та об'єму

рідини що повинен залишається в реакторі, з урахуванням поточного вмісту $CuSO_4$ потребує подальших досліджень.

Промислові установки для отримання розчину мідного купоросу працюють 2–3 робочі цикли, після чого відбувається повна заміна відпрацьованого мідного брухту та реакційної суміші в реакторі.

Висновки. В статті розглянуті принципи побудови та основні етапи розробки ефективного людино-машинного інтерфейсу (HMI) системи керування процесом отримання мідного купоросу. Розроблений HMI в реальному часі надає оператору інформацію про перебіг процесу, дозволяє спостерігати значення всіх цільових технологічних змінних в цифровій та графічній (у вигляди трендів) формі, дозволяє оператору виконувати налаштування алгоритмів керування та за необхідності здійснювати ручне керування процесом. Працездатність розробленого HMI перевірена та доведена шляхом порівняльного аналізу отриманих перехідних характеристик процесу з даними отриманими з реальної лабораторної установки.

Список літератури:

1. Актуальні проблеми автоматизації та інформаційних технологій : зб. наук. пр. / наук. ред. О. Г. Байбуз. Дніпро. 2016. Т. 20. 105 с.
2. Маслов В., Сачков І. Актуальні проблеми автоматизації в сучасному світі. *Вісник науки і освіти*. 2019. № 2 (56). С. 381.
3. Селевцов Л. І. Автоматизація технологічних процесів. Київ : Вид. центр «Акад.», 2014. 352 с.
4. Розробка комп'ютерної моделі процесу отримання розчину мідного купоросу, адаптованої для розв'язання задач управління / І. Левчук та ін. *Вісник ХНУ*. 2021. № 4(299). С. 59–66.
5. Моделювання хіміко-технологічних процесів у SCADA за допомогою технології OPEN PLATFORM COMMUNICATIONS / І. Левчук та ін. *Радіоелектронні і комп'ютерні системи*. 2020. № 2. С. 59–66.
6. Justel F. J., Taboada M. E., Jimenez Y. P. Thermodynamic study of the Cu-Na-H-SO4-Cl-HSO4-H2O system for the solubility of copper sulfate in acid seawater at different temperatures. *Journal of Molecular Liquids*. 2018. No. 249. P. 702–709.

Levchuk I.L., Dubovyk T.M., Huz H.M., Romanchuk O.O. DEVELOPMENT OF A HUMAN-MACHINE INTERFACE FOR A CONTROL SYSTEM BASED ON SCADA FOR THE COPPER SULFATE PRODUCTION PROCESS

The purpose of this work is to create an effective human-machine interface (HMI) for the control system of the copper sulfate production process in the SCADA software environment. HMI, which informs the operator in real time about the progress of the controlled process, allows you to observe the values of all target technological variables in digital and graphical form, makes it possible to manually control the process and configure control algorithms. The article provides a brief description of the main parameters of the computer model of the copper sulfate production process implemented in the universal modeling program CHEMCAD, which interact with the elements of the human-machine interface implemented on the basis of SCADA. The general structure and appearance of the proposed HMI, which consists of an information panel for displaying the main process parameters in digital form, a batcher configuration panel designed to configure batcher control algorithms and dosing limits, trends for displaying the main target process parameters in graphic form, as well as trends that graphically display the operation of the dosing subsystem. The features of creating and configuring individual elements of the human-machine interface (trends) are considered and a list and brief description of all arguments – variables of the human-machine interface, which ensure its functioning and information exchange between HMI elements are given. The operability of the proposed HMI is checked and proven by comparing the obtained transient characteristics of the process with the data of a real laboratory installation for obtaining a solution of copper sulfate. The analysis of the obtained dependencies presented in the article allows us to conclude about the full functionality of the developed HMI and the achievement of the goal of the work. The technical solutions, principles and features of the development of the human-machine interface proposed in the work are universal in nature and can be used to develop or improve control systems for similar technological processes that are built on the basis of SCADA and mathematical or computer models.

Key words: *control system, human-machine interface, SCADA, copper sulfate.*

УДК 004.415.2
DOI https://doi.org/10.32782/2663-5941/2025.3.2/34

Майданюк В.П.
Вінницький національний технічний університет

Складанюк О.О.
Вінницький національний технічний університет

ГРАФІЧНИЙ РЕДАКТОР ДЛЯ РЕДАГУВАННЯ ДЕСКТОПНИХ ВІДЕОІГОР

У статті виконано аналітичний огляд методів та засобів для редагування сцен у 2D-відеоіграх. Обґрунтовано актуальність створення подібного інструменту з урахуванням поширеності інді-розробок і потреби в інтуїтивно зрозумілих редакторах. Проаналізовано технічні аспекти реалізації функціоналу, таких як інтерактивне переміщення об'єктів (drag-and-drop), редагування властивостей об'єктів сцени у реальному часі, управління шарами та компонентами сцени. Розглянуто підходи до збереження структури сцени у форматі JSON, що дозволяє легко експортувати та імпортувати дані про ігрові рівні. Особливу увагу приділено використанню 2D-графічної бібліотеки Pixi.js як рушія для рендерингу сцени, а також принципам побудови архітектури на основі об'єктно-орієнтованого підходу. Здійснено аналіз використання подієво-орієнтованої моделі для забезпечення взаємодії між компонентами редактора. Окремо проаналізовано перспективи розширення функціоналу редактора, зокрема реалізацію сітки для вирівнювання об'єктів, систему undo/redo, підтримку кількох сцен, та інструменти для створення складніших ігрових механік. Визначено напрями подальших досліджень, зокрема створення адаптивної системи плагінів для розширення функцій редактора. Основним науково-практичним результатом дослідження є визначення ключових компонентів, потрібних для побудови ефективного редактора 2D-сцен, а також реалізація базової версії редактора, що може бути використана як основа для подальшого розвитку. Основними перевагами запропонованого редактору є модульність інтерфейсу, легка масштабованість проекту, простота розробки та підтримки, збереження сцен у JSON, наявність системи шарів, drag-and-drop редагування, базова підтримку undo/redo, інтеграція з Pixi.js як візуальним рушієм, платформнонезалежність. Практична цінність роботи полягає у можливості використання розробленого інструменту для побудови прототипів та ігрових рівнів без необхідності написання коду, що сприяє прискоренню процесу розробки 2D-ігор, дозволяючи розробникам зосередитися на ігровій логіці та дизайні, а не на технічних деталях реалізації сцени. Запропонований редактор може бути інтегрований у навчальні програми для ознайомлення студентів із принципами побудови ігрових середовищ та графічного програмування.

Ключові слова: *графічний редактор, відеоігри, сцени, 2D, drag-and-drop, Pixi.js, JSON, undo/redo, шари, об'єктно-орієнтоване програмування.*

Постановка проблеми. У сучасну епоху ігрової індустрії все більше людей отримують доступ до інструментів створення власних відеоігор. Зокрема, зростає популярність так званих інді-розробок (англ. *indie*, скорочено від *independent* – створення ігор командами або окремими розробниками, що не залежать від великих видавців), які потребують простих, ефективних та інтуїтивно зрозумілих інструментів для створення та редагування ігрових сцен. Багато з таких проєктів мають обмежені ресурси, тому виникає потреба у графічних редакторах, які дозволяють створювати та змінювати 2D-сцени без потреби у глибоких знаннях програмування. Одним із перспективних рішень є створення універсального редактора, що підтримує візуальне редагування, організацію об'єктів сцени та збереження її структури у зручному форматі, придатному для імпорту в ігровий рушій [1].

Аналіз останніх досліджень і публікацій. У статті здійснено аналітичний огляд сучасних інструментів для редагування сцен у 2D-відеоіграх. Аналізуються такі продукти:

- Unity і Unreal Engine – як професійні рушії, що мають потужні, але складні в освоєнні інструменти для редагування сцен;
- Tiled Map Editor – потужний редактор, але з ускладненням при інтеграції кастомних об'єктів;
- LDtk – редактор з автоматизованими тайлами, проте орієнтований здебільшого на платформери;

• Ogmo Editor – легкий і зручний для новачків, однак має обмежену гнучкість і функціональність.

Проблеми, що залишаються невирішеними:

• Відсутність універсального інструменту, що поєднує простоту, модульність, масштабованість, drag-and-drop редагування, збереження у JSON та підтримку undo/redo.

• Недостатня підтримка кастомізації, анімацій, GUI-подій та плагінів у більшості наявних редакторів.

Огляд і аналіз аналогів. Серед популярних інструментів для створення 2D-ігор слід відзначити Tiled, LDtk, Ogmo Editor. Tiled підтримує багато типів карт, але має складну інтеграцію кастомних об›єктів. LDtk забезпечує автошари та інтуїтивний інтерфейс, але орієнтований переважно на платформери. Ogmo Editor – простий, зручний, але з обмеженою функціональністю. Таким чином, існує нереалізована ніша – створення редактора з гнучкою архітектурою, системою шарів, підтримкою undo/redo, інтеграцією з рушіями та простотою використання.

Tiled Map Editor – це потужний, безкоштовний та відкритий редактор карт для 2D-ігор, який підтримує різні типи карт: ортогональні, ізометричні та гексагональні. Загальний вигляд інтерфейсу даного редактору наведено на рис. 1. Основні можливості Tiled Map Editor такі:

– підтримка кількох шарів тайлів та об’єктів, що дозволяє створювати досить складні сцени [6];

– можливість додавати прямокутники, еліпси, полігони та інші об’єкти з індивідуальними властивостями, що фактично надає гнучкості об’єктним шарам;

– можливість створення та повторного використання шаблонів для стандартних елементів;

– підтримка експорту у різні формати: TMX (XML), JSON, Lua та інші, що забезпечує сумісність із багатьма ігровими рушіями.

LDtk (Level Designer Toolkit) – це сучасний редактор рівнів, створений автором гри Dead Cells. Він орієнтований на простоту використання та досить високу ефективність [7]. Загальний вигляд інтерфейсу даного редактору наведено на рис. 2.

Основні особливості LDtk такі:

– інтуїтивно зрозумілий, сучасний інтерфейс користувача з акцентом на зручність;

– наявність системи автоматичне розміщення тайлів (auto-tiling) на основі правил, що значно пришвидшує процес створення рівнів;

– генерація PNG-файлів для візуалізації та JSON-файлів для даних, що полегшує інтеграцію з ігровими рушіями.

– підтримка кастомних полів, що забезпечується можливістю додавання власних полів до об’єктів для зберігання додаткової інформації.

Ogmo Editor – це легкий, відкритий редактор рівнів, розроблений для інді-розробників [8]. Загальний вигляд його інтерфейсу наведено на рис. 3. Основними перевагами редактора є:

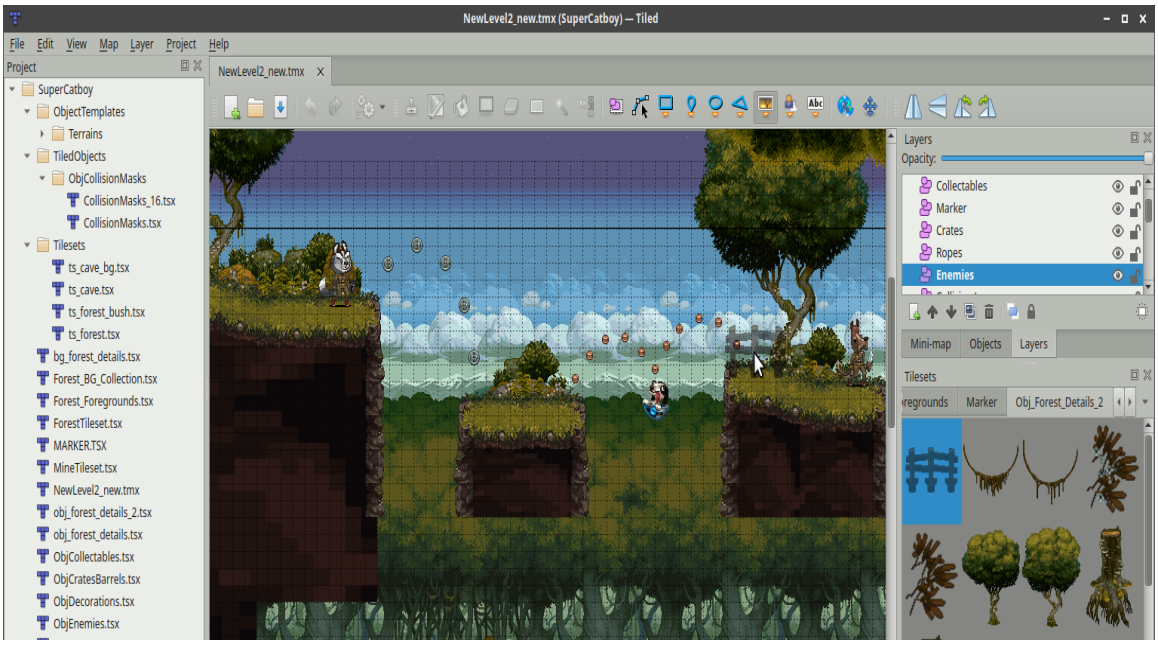


Рис. 1. Інтерфейс редактора Tiled Map Editor

– передбачено проєктно-орієнтований підхід, де всі налаштування зберігаються у проєкті, що забезпечує узгодженість між рівнями;

– підтримка різних типів шарів (тайлових, сіткових, об'єктних та декоративних) для гнучкого дизайну рівнів;

– передбачає зручний формат для інтеграції з багатьма ігровими рушіями (експорт у JSON);

– простота використання завдяки наявності інтуїтивно зрозумілого інтерфейсу, що дозволяє швидко освоїтися навіть новачкам.

Вищенаведені аналоги графічних редакторів для 2D-ігор мають спільний набір базових функцій, таких як підтримка шарів, drag-and-drop інтерфейс, збереження сцен у форматі JSON та система undo/redo. Ці функції створюють зручне середовище для візуального редагування ігрових сцен без потреби у прямому втручанні у код. Завдяки використанню таких бібліотек, як Pixi.js, і сучасних підходів до управління станом (наприклад, Zustand), забезпечується гнучкість і масштабованість редакторів. У поєднанні з можливістю інтеграції з рушіями або власними пайплайнами розробки, ці редактори стають незамінними інструментами, особливо у сфері інді-розробки.

Постановка завдання. Метою статті є розширення функціональних можливостей графічного редактору для 2D-сцен, визначення архітектурних рішень і ключових компонентів для реалізації функціоналу, необхідного для редагування ігрових рівнів, а також побудова базової

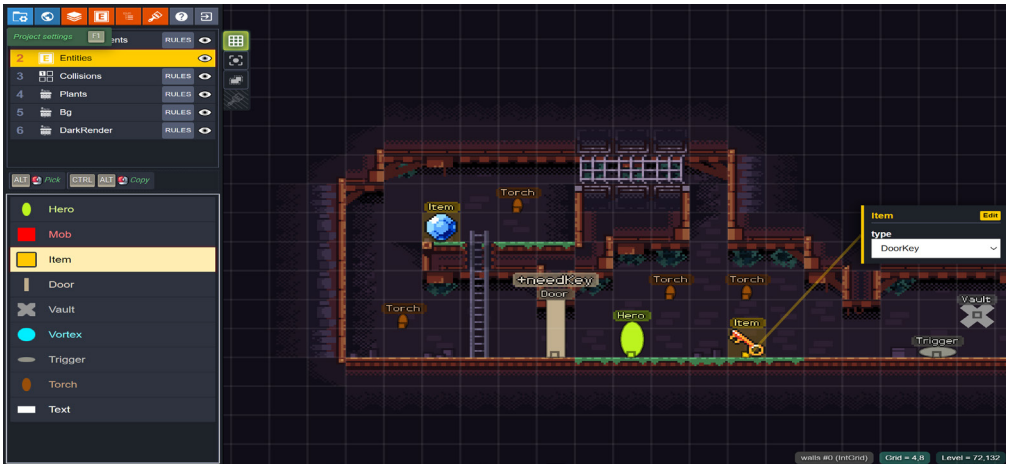


Рис. 2. Інтерфейс редактора LDtk

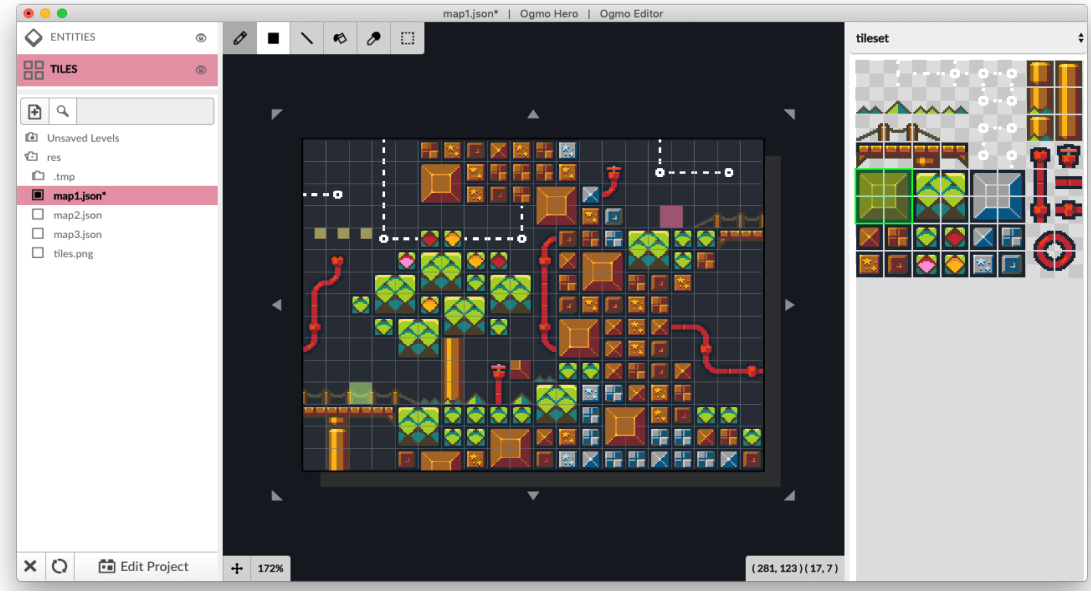


Рис. 3. Інтерфейс редактора Ogmo Editor

версії інструменту, що може бути використана у процесі інді-розробки відеоігор.

Виклад основного матеріалу. У типовому редакторі сцен для 2D-ігор мають місце такі ключові компоненти:

1. Сцена та об'єкти зберігаються у форматі JSON, що дозволяє легко імпортувати/експортувати дані та інтегрувати з рушієм.

2. Сцена підтримує ієрархічну структуру, що дозволяє створювати вкладені об'єкти, групи, шари з різними рівнями видимості.

3. Інтерфейс drag-and-drop забезпечує зручне переміщення, масштабування та обертання об'єктів мишею у редактор [3].

4. Підтримка історії змін – необхідна функціональність, що дозволяє користувачеві повертати попередні стани сцени (Undo/redo).

5. Панель властивостей дозволяє змінювати позицію, розмір, текстуру, поведінку об'єктів тощо.

У розробці редактора часто використовують бібліотеки на кшталт Pixi.js для рендерингу сцени, React або Vue для побудови інтерфейсу користувача (User Interface, UI), Zustand або Redux – для керування станом. У складніших рішеннях можливе використання WebAssembly для пришвидшення обчислень, а також Canvas/ WebGL для рендерингу великої кількості об'єктів у сцені без втрати продуктивності.

На сьогоднішній день багато популярних ігрових рушіїв, таких як Unity та Unreal Engine, пропонують розробникам потужні інструменти для редагування сцен та об'єктів. Ці інструменти дозволяють змінювати візуальні та функціональні аспекти гри без необхідності переписувати код. Unity – це потужний і дуже популярний багатоплатформний рушій для створення комп'ютерних ігор, а також інтерактивних 2D і 3D-додатків. Він надає розробникам усе необхідне для створення ігор. Саме тому його використовують фахівці з усього світу, і багато хто з них вважає Unity номером 1 у світі розробки ігор [4]. Принцип роботи в Unity передбачає ряд особливостей, які розробники мають врахувати під час створення гри. Нижче розглянемо основні з них детальніше.

1. Першою особливістю є створення сцени. Сцена – це простір, де розміщуються та взаємодіють об'єкти гри. Їх може бути декілька на кожному рівні. В Unity можна створювати та редагувати сцени, визначати компоненти оточення, задавати вигляд та рівні освітлення, камери і т. д.

2. Ще однією важливою особливістю є додавання об'єктів. В Unity об'єкти – це основні будівельні блоки гри. Тобто необхідно додавати різні об'єкти до сцени (такі як персонажі, предмети, перешкоди тощо), а також визначити їх взаємодію та поведінку.

3. Ще одна особливість передбачає застосування скриптів та кодування. За допомогою скриптів необхідно створювати поведінку об'єктів та керувати ігровою логікою. За допомогою коду можна створювати унікальні функції, визначити умови перемоги або поразки, обробляти введення користувача та багато іншого. Доцільно зазначити, що вся робота з Unity здійснюється мовою C#.

4. На етапі, коли гра створена, її необхідно виконати її ретельне тестування на різних пристроях та платформах. Тестування допомагає виявити помилки, покращити геймплей та оптимізувати продуктивність, для того, щоб забезпечити ефективне функціонування гри на різних пристроях.

5. На останньому кроці гра публікується на вибраних платформах, таких як комп'ютери, мобільні пристрої, ігрові приставки та інше. Unity надає ряд корисних інструментів для експорту та публікації гри на різних платформах, що дозволяє досягти широкої аудиторії гравців [2].

А тепер розглянемо особливості ігрового рушія Unreal Engine. Unreal Engine – інструмент для створення ігор і сцен із використанням 3D-моделей, розроблений компанією Epic Games. Особливості рушія такі:

1. Він надає засоби для створення дивовижних графічних ефектів: високоякісне освітлення, рендеринг (формування зображення або відео з тривимірної сцени) та підтримка віртуальної реальності.

2. Він дозволяє моделювати об'єкти, відтворювати їхній рух і взаємодію з навколишнім середовищем досить реалістично.

3. Він може бути використаний для розробки ігор на різних платформах: комп'ютерах, консолях, мобільних пристроях і у віртуальній реальності.

Він використовує мову програмування C++, а також має систему розширень та графічний інтерфейс для того, щоб зробити гру, не знаючи кодингу [5].

Створено власний графічний редактор, що має загальний інтерфейс, наведений на рис. 4. Його порівняння з аналогами наведено

у табл. 1. Даний графічний редактор поєднує гнучкість сучасних веб-технологій з можливістю запуску на десктопі завдяки використанню React.js для побудови інтерфейсу та Electron для створення кросплатформного десктопного застосунку. Редактор надає функції створення власних об'єктів з можливістю вставляти їх у обхідну сцену. Основними перевагами даної розробки є:

– використання компонентного підходу React дозволяє легко масштабувати проєкт та підтримувати розширюваність (наприклад, додавання нових панелей, типів об'єктів тощо);

– завдяки вебстеку розробка виконується значно швидше порівняно з C++ або C#, що використовуються в аналогах (Tiled, LDtk);

– сцена формується як структура з об'єктами, що легко експортується у формат JSON для подальшої інтеграції з ігровим рушієм;

– висока гнучкість архітектури, що досягається шляхом реалізації системи шарів, drag-and-drop редагування, базової підтримки undo/redo, інтеграції з Pixi.js як візуальним рушієм;

– платформонезалежність, оскільки Electron дозволяє запускати редактор на Windows, macOS та Linux без необхідності специфічної компіляції.

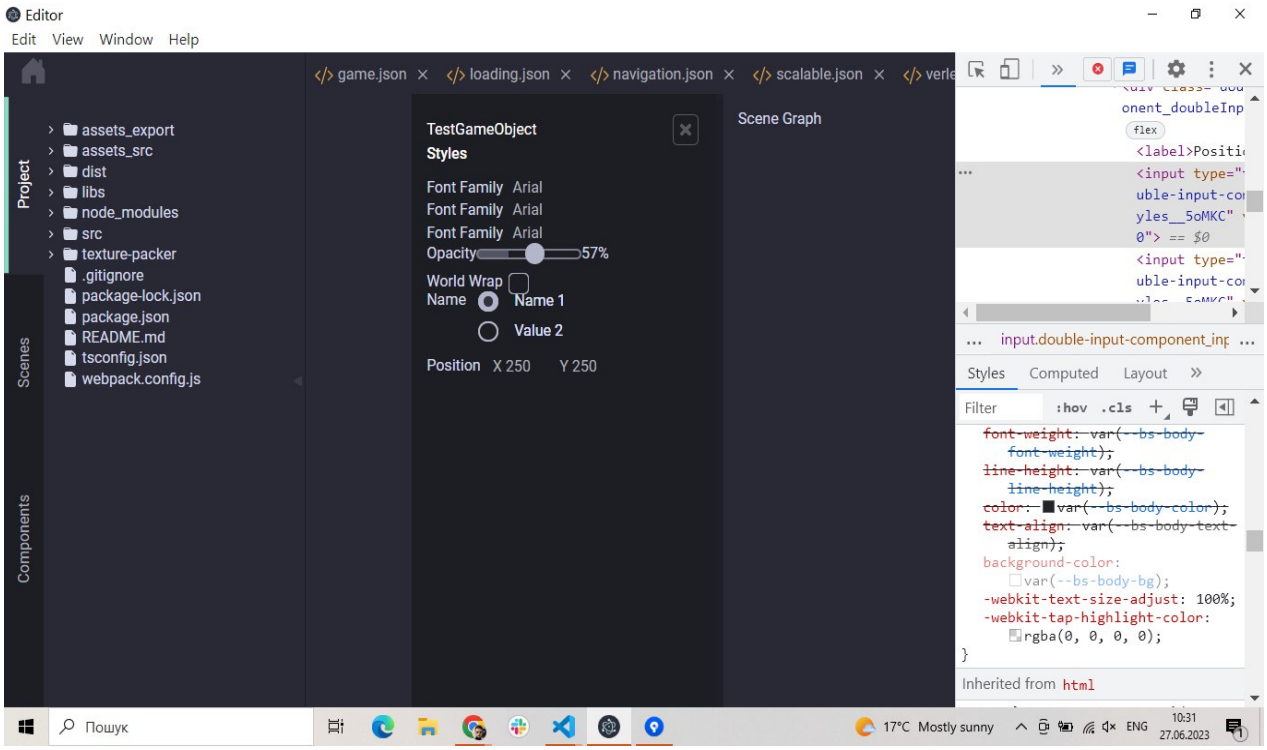


Рис. 4. Прототип інтерфейсу власного редактору

Таблиця 1

Порівняння власного редактору з аналогами

Редактор	Платформа	Мова / Технології	Основні особливості	Недоліки
Tiled	Десктоп (Qt)	C++	Підтримка тайлів, інтеграція з рушіями, JSON/TSX формати.	Важка інтеграція кастомних об'єктів, складна система плагінів
LDtk	Десктоп (MonoGame)	C#	Тайли, автошари, вбудовані правила генерації рівнів.	Вузька орієнтація на платформери, менша гнучкість щодо об'єктів
Ogmo Editor	Десктоп (Electron)	TypeScript	Проста структура, легке збереження в JSON, орієнтований на інді.	Відсутність візуальної ієрархії, обмежена кількість налаштувань
Власна розробка	Десктоп (Electron)	React.js + Pixi.js	Система шарів, drag-and-drop, JSON, undo/redo, відкритість до розширення.	Потребує подальшого розвитку (анімації, інтеграція з рушіями, GUI для подій)

Висновки. У даному дослідженні розглянуто алгоритми, які дозволяють покращити редагування відеоігор, а отже, й покращувати конверсію користувачів та їхню взаємодію з редактором для редагування десктопних відеоігор. Проаналізовано архітектуру та функціональні можливості графічного редактора для редагування сцен у 2D-десктопних відеоіграх. Виокремлено ключові компоненти редактора, що забезпечують зручність, масштабованість та інтеграційність інструменту. Виділено напрями подальших досліджень, такі як: підвищення продуктивності при роботі з великою кількістю об'єктів, додавання складних типів об'єктів (анімації, тригери, логіка), створення системи плагінів для розширення функціональності. Представлені архітектурні підходи можуть бути використані для створення кастомних редакторів, зокрема у сфері інді-розробки ігор. Розглянуто аналоги редакторів для редагування десктопних відеоігор, наведено стислий огляд кожного з них. Крім того, виконано аналіз ігрових рушіїв, які використовуються під час розробки редакторів. На основі отриманих результатів створено прототип власного редактора, основними перевагами якого, зокрема, є кросплатформність завдяки використанню вебстеку, простота масштабування та гнучка архітектура.

Список літератури:

1. Share of game developers working on 2D games worldwide. *Statista*. URL: https://www.statista.com/statistics/1250457/2d-vs-3d-games-indie-developers (дата звернення 25.05.2025)

2. Examples. *Pixi.js*. URL: https://pixijs.io/examples/ (дата звернення 25.05.2025)

3. HTML Drag and Drop API. *Mozilla Developer Network* URL:https://developer.mozilla.org/enUS/docs/Web/API/HTML_Drag_and_Drop_API. (дата звернення 25.05.2025)

4. Unity. URL: https://surl.li/mdhswu (дата звернення 25.05.2025)

5. Unreal Engine. URL: https://skvot.io/uk/blog/ne-soromno-zapitati-yak-pracyuye-unreal-engine (дата звернення 25.05.2025)

6. Lindeijer T. Tiled: A flexible level editor URL: https://www.mapeditor.org (дата звернення 25.05.2025)

7. Bénard S. LDtk – Level Designer Toolkit. URL https://ldtk.io (дата звернення 25.05.2025)

8. Ogmo Editor. Open-source level editor for 2D games URL: https://ogmo-editor-3.github.io (дата звернення 25.05.2025)

Maidaniuk V.P., Skladaniuk O.O. GRAPHICAL EDITOR FOR DESKTOP VIDEO GAME SCENE EDITING

The article presents an analytical review of methods and tools for editing scenes in 2D video games. The relevance of developing such a tool is substantiated, considering the popularity of indie game development and the need for intuitive editors. The technical aspects of implementing functionality such as interactive object movement (drag-and-drop), real-time editing of scene object properties, and layer and scene component management are analyzed. Approaches to storing the scene structure in JSON format are considered, allowing for easy export and import of game level data. Special attention is given to the use of the 2D graphics library Pixi.js as the rendering engine, as well as to architecture design principles based on an object-oriented approach. An analysis is carried out on the use of an event-driven model to ensure interaction between the editor's components. The article also explores the potential for expanding the editor's functionality, including implementing a grid for object alignment, an undo/redo system, support for multiple scenes, and tools for creating more complex game mechanics. Directions for further research are outlined, particularly the development of an adaptive plugin system to extend the editor's features. The main scientific and practical result of the research is the identification of key components necessary for building an effective 2D scene editor, as well as the implementation of a basic version of the editor that can serve as a foundation for future development. The proposed editor's key advantages include a modular interface, ease of project scalability, simplicity of development and maintenance, JSON-based scene saving, a layer system, drag-and-drop editing, basic undo/redo support, integration with Pixi.js as a rendering engine, and cross-platform compatibility. The practical value of the work lies in the possibility of using the developed tool for building prototypes and game levels without the need to write code, thereby accelerating the 2D game development process and allowing developers to focus on game logic and design rather than the technical implementation of scenes. The proposed editor can be integrated into educational programs to familiarize students with the principles of game environment construction and graphics programming.

Key words: *graphic editor, video games, scenes, 2D, drag-and-drop, Pixi.js, JSON, undo/redo, layers, object-oriented programming.*

УДК 658.5.015.23:681.5.015.24
DOI https://doi.org/10.32782/2663-5941/2025.3.2/35

Мацуй А.М.
Центральноукраїнський національний технічний університет

Сербул О.М.
Центральноукраїнський національний технічний університет

Хлєбніков М.В.
Центральноукраїнський національний технічний університет

Ізовіта О.Л.
Центральноукраїнський національний технічний університет

ОБГРУНТУВАННЯ РОЗРОБКИ КОМП'ЮТЕРИЗОВАНОЇ СИСТЕМИ СТАБІЛІЗАЦІЇ ОПТИМАЛЬНОГО РІЗНОРОЗМІРНОГО КУЛЬОВОГО ЗАВАНТАЖЕННЯ БАРАБАННИХ МЛИНІВ

У статті розглянуто проблему підвищення енергоефективності процесів подрібнення низькоякісних залізних руд у кульових млинах, які споживають 7–10% світового виробництва електроенергії та становлять один із найбільш енерговитратних етапів виробничого ланцюга. Проаналізовано основні збурюючі фактори, що впливають на ефективність роботи кульових млинів: оптимальне розрідження пульпи, збурення на вході млина та структуру кульового завантаження. Особливу увагу приділено кульовому завантаженню як ключовому чиннику, що визначає рівень енергоспоживання, інтенсивність подрібнення та продуктивність технологічного процесу в цілому. Детально проаналізовано сучасний стан технологій подрібнення та існуючі проблеми у галузі. Показано, що на операції подрібнення припадає близько половини усіх капітальних та експлуатаційних витрат збагачувальних фабрик. Проведено огляд сучасних підходів до автоматизації процесів подрібнення, включаючи використання штучного інтелекту, нейронних мереж, звукових сигналів помелу та інноваційних засобів контролю. Виявлено, що питання автоматизованого завантаження млинів мололльними тілами практично не досліджувалося, хоча саме воно суттєво впливає на енергоспоживання та якість готового продукту. Обґрунтовано необхідність розробки комп'ютеризованої системи керування завантаженням куль, яка забезпечить автоматичну стабілізацію оптимального різнорозмірного кульового завантаження в умовах багатьох змінних параметрів. Запропоновано інноваційний двокомпонентний підхід: оцінювання кульового завантаження через непрямий прогностичний метод визначення спрацювання куль на основі енергетичного підходу, та оцінювання завантаження руди за ефективністю її подрібнення безпосередньо в барабані млина. Така система повинна своєчасно виявляти та виводити з процесу зношені або деформовані елементи. Підкреслено важливість підтримання різнорозмірного складу куль для забезпечення належної якості подрібнення та попередження критичних режимів роботи. Розроблена комп'ютеризована система керування дозволить значно знизити енергоспоживання, продовжити строк експлуатації обладнання, зменшити знос футерівки та підвищити загальну ефективність переробки залізорудної сировини на всіх етапах виробництва.

Ключові слова: *кульовий млин, енергоефективність, подрібнення руди, оптимальне різнорозмірне кульове завантаження, комп'ютеризована система, збурюючі фактори.*

Постановка проблеми. У нинішніх умовах розвитку гірничо-металургійної галузі дедалі більшої важливості набуває завдання ефективної обробки низькоякісних залізних руд, які нині становлять значну частину сировинної бази чорної металургії. Зменшення вмісту заліза у вихідному матеріалі зумовило активне застосування технологій збагачення, заснованих на процесах тонкого дроблення та подальшого вилучення цінного мінералу. Одним із найбільш енерговитратних етапів цього виробничого ланцюга є початкове подрібнення у кульових млинах. Саме на цій стадії спостерігаються максимальні втрати електроенергії – 7…10% всього світового виробництва [1], зношування футерівки та мололльних тіл, що в кінцевому підсумку значно підвищує собівартість концентрату. У зв'язку з цим особливої ваги набувають заходи з підвищення енергоефек-

тивності цього процесу, зокрема шляхом оптимізації параметрів кульового завантаження та модернізації систем автоматизованого керування.

На ефективність процесу подрібнення впливають різні чинники, серед яких основними є оптимальне розрідження пульпи, вплив збурень на вході млина та правильне кульове завантаження. Особливу роль відіграє саме структура кульового завантаження – вона визначає рівень енергоспоживання, інтенсивність подрібнення та продуктивність технологічного процесу. Якщо склад завантаження не відповідає необхідним параметрам, це призводить до зниження ефективності роботи, прискореного зносу обладнання та погіршення якості кінцевого продукту. У зв'язку з цим особливо важливим є забезпечення стабільного та оптимального за масою і розміром завантаження куль, що досягається шляхом автоматизованого контролю і поповнення ними в реальному часі. Це дозволяє не лише зберігати оптимальні технологічні параметри, а й продовжити строк експлуатації обладнання.

Одним із важливих напрямів є створення комп'ютеризованої системи, що забезпечує стабілізацію оптимального завантаження млина в умовах багатьох змінних параметрів. Така система повинна не лише контролювати об'єм та розміри куль, а й своєчасно виявляти та виводити з процесу зношені або деформовані елементи. Відсутність подібного рішення спричиняє надмірне споживання електроенергії, падіння загальної продуктивності та порушення рівноваги у технологічному циклі. У свою чергу, запровадження інтелектуального керування, що враховує ступінь зносу куль, властивості сировини, стан футерівки та показники навантаження, дозволяє значно покращити ефективність виробничих процесів і знизити витрати.

У результаті розв'язання окреслених завдань можна досягти важливих теоретичних і практичних результатів. Це, зокрема, забезпечить подальше вдосконалення існуючих систем керування процесами рудопідготовки, дасть змогу раціональніше використовувати ресурси та сприятиме підвищенню загальної ефективності переробки залізорудної сировини на всіх етапах виробництва.

Дана робота виконана в межах тематики наукових досліджень Центральноукраїнського національного технічного університету за темами: «Автоматизована стабілізація оптимального різнорозмірного кульового завантаження барабанних млинів при подрібненні вихідної руди з пісками класифікатора» (реєстраційний номер 0124U003449), «Оптимізація продуктивності кульових млинів по руді і готовому продукту при мінімальних енергетичних і матеріальних перевитратах» (реєстраційний номер 0115U003942) та «Оптимізація продуктивності стержневих млинів розімкнутого циклу по руді і готовому продукту при мінімальних енергетичних і матеріальних перевитратах» (реєстраційний номер 0123U102622).

Аналіз останніх досліджень і публікацій. Енергоефективність процесу подрібнення рудної маси в кульовому млині визначається сукупністю кількох ключових чинників. Основними серед них є: досягнення оптимального розрідження пульпи, вплив збурень на вхідному етапі та правильне формування кульового завантаження. Питання регулювання ступеня розрідження пульпи детально розглядається в монографії [2], де кульові млини з їх комунікаціями були розглянуті як регульовані об'єкти. У ній подано алгоритмічні підходи до визначення співвідношення твердої та рідкої фаз у пульпі, а також розроблені засоби комп'ютерного керування цим параметром. Викладено результати експериментальних досліджень і реалізації цих систем, а також практичні аспекти їх впровадження.

У роботі [3] запропоновано шляхи покращення подачі руди в кульові млини через зменшення дії збурюючих факторів. Розглянуто сучасні підходи до рудопідготовки на залізорудних фабриках, початкове подрібнення та технології накопичення руди на перших етапах. Особливу увагу приділено автоматизованому контролю рудних потоків: підвищенню точності вимірювання масової витрати руди та оперативному визначенню крупності твердих частинок на вході в млин першої стадії. Продемонстровано впровадження цих рішень, що забезпечує стабільне живлення і, як наслідок, ефективне подрібнення сировини.

Створено алгоритми енергоефективного керування процесами подрібнення та класифікації руд [4], а також розроблено програмно-технічні засоби, на яких базується інформаційна платформа автоматизованого регулювання технологічних параметрів. Комплекс обладнано пристроями для вимірювання середньозваженої крупності пісків односпірального класифікатора й твердої фази на виході з кульового млина, а також інструментами непрямого визначення витрат піску в жолобі.

Підвищення енергоефективності кульових млинів при подрібненні руд є ключовим завдан-

ням у гірничодобувній промисловості, оскільки на їхню частку припадає значна частина (50–60%) загальних витрат на збагачувальних підприємствах і близько 5,2% світового споживання електроенергії [5]. Основними факторами, що впливають на енергоефективність, є кульове завантаження млина, властивості руди (твердість, щільність, вологість), швидкість обертання барабана та знос футерівки [6]. Оптимальний розподіл куль, відповідний розміру руди, дозволяє підвищити ефективність процесу подрібнення та знизити енергоспоживання. Знос футерівки, наприклад, може призвести до збільшення питомого споживання енергії на 5–6% [5]. Для підвищення енергоефективності пропонуються сучасні системи керування, зокрема адаптивні, які дозволяють системі реагувати на зміни властивостей руди в реальному часі, автоматично змінюючи параметри роботи млина (наприклад, вагу куль та струм збудження синхронного двигуна [5]) для досягнення необхідної продуктивності.

Контроль заповнення млина кулями та регулювання швидкості електродвигуна залежно від навантаження, також сприяють зниженню витрат електроенергії. Крім того, наголошується на важливості точного прогнозування та планування електроспоживання на гірничодобувних підприємствах загалом, використовуючи моделі режимів електроспоживання, отримані за результатами багатофакторного регресійного аналізу [7]. Основною метою є забезпечення стабільної роботи кульового млина та електродвигуна при повному завантаженні та оптимізація режимів роботи електроприводу.

Способи оптимізації роботи кульових млинів при збагаченні залізних руд активно досліджуються сучасними науковцями. Одним із напрямків, стало використання звукових сигналів помелу для визначення кульового завантаження [8]. Взаємозв'язок між звуковими сигналами і навантаженням на млин є складним, тому запропоновано метод, що поєднує алгоритми машинного навчання з технологією розпізнавання звуку для ідентифікації навантаження на млин.

Використання штучного інтелекту досліджується не тільки при мокрому кульовому подрібненні, але й для оптимізації процесів сухого подрібнення [9]. Запропоновано модель штучної нейронної мережі з кількома входами та виходами, яка застосовується для прогнозування результатів процесу. Використання нейронних мереж дає кращі результати, зменшуючи похибки та час регулювання, що покращує надійність системи. Завдяки цьому методу, вдалося знизити перерегулювання на 79% та регулювання на 74,2%, для даного технологічного процесу.

У зв'язку з підвищенням цін на енергоносії, питання зменшення енергоспоживання в гірничодобувній промисловості стало актуальним. Дослідження [10] показали, що до 50% енергії на збагачувальних фабриках витрачається на подрібнення руди. Тому представлені централізовані схеми прогнозного керування за допомогою гібридної моделі (HMPC), що дозволяють знижувати енергоспоживання і стабілізувати роботу млинів, забезпечуючи необхідний розмір частинок при мінімальних витратах енергії.

Для оптимізації роботи збагачувальних фабрик, важливо забезпечити найкращі умови роботи млинів для ефективного вивільнення корисного компоненту. Інноваційний інструмент Sensomag, розроблений компанією Magotteaux, дозволяє забезпечувати точні онлайн-вимірювання ступеня заповнення млина і положення пульпи. Це дає змогу оперативно реагувати на зміни в процесі і підтримувати кульове завантаження на оптимальному рівні, що підвищує ефективність і знижує витрати енергії та знос обладнання.

Отже, з розглянутого видно, що питання завантаження кульового млина мололльними тілами практично не досліджувалося, але саме воно впливає на рівень енергоспоживання, якість подрібнення матеріалу та продуктивність технологічного процесу в цілому.

Постановка завдання. Метою дослідження є визначення збурюючих факторів, що впливають на роботу кульового млина при подрібненні руди та обґрунтування розробки комп'ютеризованої системи керування завантаженням куль в агрегат, що дозволить підвищити продуктивність роботи технологічного агрегату та якість готового продукту.

Виклад основного матеріалу. У процесах рудопідготовки найчастіше застосовують кульові млини, оскільки вони вирізняються значною продуктивністю, надійністю, легкістю обслуговування й зберігають працездатність навіть за суттєвого зношення. Водночас така технологія потребує багато енергії і демонструє невисокий ККД під час формування нової поверхні частинок [1]. У підсумку на операції подрібнення припадає близько половини – двох третин усіх капітальних та експлуатаційних витрат збагачувальних фабрик. На рис. 1 подано загаль-

неприйняту схему першої стадії подрібнення корисних копалин, поширену як у вітчизняній, так і у світовій гірничо-збагачувальній практиці. Технологічна схема вміщує операції подрібнення і класифікації. Такий замкнений цикл дає змогу отримати однорідний гранулометричний склад матеріалу та скоротити енергетичні витрати на подальшу переробку.

Більшість сучасних систем автоматизованого керування процесом подрібнення руди в кульовому млині, орієнтовані на вдосконалення технології оптимізації рівня рудного заповнення всередині нього. Додатково функціонують допоміжні системи, завданням яких є регулювання кількості подрібнювальних тіл (куль) у млині, контроль ступеня розрідження рудної маси та пульпи. Останнє особливо важливе в системі керування односпіральним механічним класифікатором, де забезпечується необхідний гранулометричний склад кінцевого продукту.

Розроблені також системи автоматичного регулювання процесу завантаження рудою кульового млина з урахуванням фізико-механічних характеристик сировини. До ключових контрольованих параметрів відносяться середньозважений розмір частинок руди, концентрація магнітного заліза, здатність матеріалу до подрібнення, обсяг циркулюючого навантаження, тиск масла в опорних підшипниках, показання радіоізотопних пристроїв, а також магнітна проникність пульпи. Проте деякі підходи не здобули широкого застосування, що властивості вихідного живлення кульових млинів виступають збурюючими діяннями на керований об'єкт. Ці збурення мають нестійкий характер і ускладнюють реалізацію ефективного керування.

Спроби стабілізації властивостей сировини шляхом осереднення окремих типів руд також широкого застосування не отримали [11]. Поєднання інерційності об'єкта керування та швидкості наростання збурюючих впливів створює труднощі при керуванні. Це ще сильно посилюється широкою гамою властивостей руди, які змінюються випадковим чином в широких межах і негативно впливають на показники подрібнення і класифікації. Тому нині доцільно вирішувати задачі удосконалення рудного живлення кульових млинів, спрямовані на ліквідацію збуруюючих впливів.

Окрім рудного завантаження кульових млинів, суттєву роль відіграє розрідження пульпи, оскільки воно безпосередньо впливає на текучість рідкого матеріалу і ефективність кульового подрібнення твердих частинок. Згідно з виробничими спостереженнями, задовільна текучість пульпи досягається при вмісті води в межах 20…25%. Найвища продуктивність млина спостерігається, коли масова частка твердого в пульпі складає приблизно 60…65%. Це свідчить про те, що густина пульпи є ключовим фактором у досягненні оптимальної рудної продуктивності процесу подрібнення. Водночас встановлення точного рівня розрідження ускладнюється залежністю цього параметра від розміру вихідного матеріалу та фізико-механічних характеристик руди. Ці взаємозв'язки досі оста-

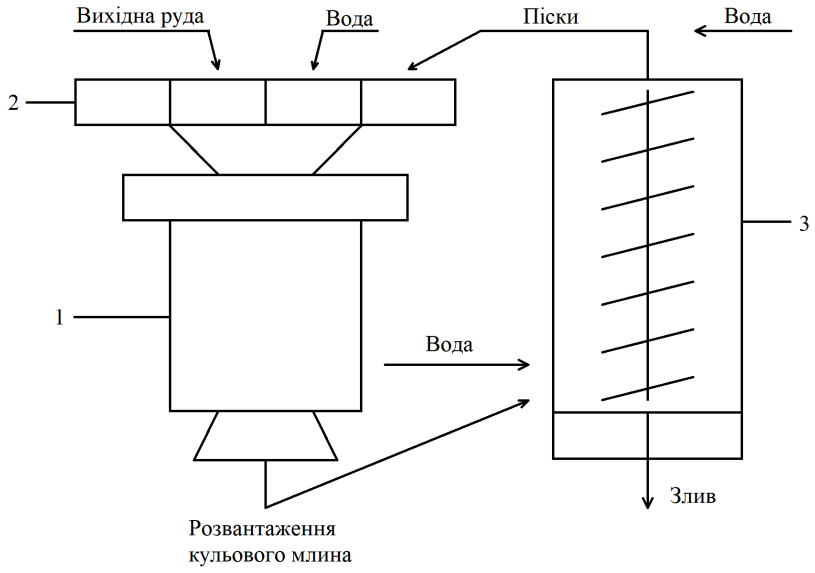


Рис. 1. Типова технологічна схема подрібнення руди у першій стадії:
1 – кульовий млин; 2 – завитковий живильник; 3 – механічний односпіральний класифікатор

точно не з'ясовані і через це у млинах першої стадії подрібнення підтримання оптимального рівня розрідження не забезпечується автоматично – його значення оператор задає на основі власного досвіду, оскільки універсального алгоритму для розрахунку в конкретних умовах поки що не існує.

Регулювання розрідження пульпи в процесах подрібнення-класифікації відіграє ключову роль у забезпеченні стабільності цих технологічних операцій. Завдяки йому можливо підтримувати оптимальне циркулююче навантаження й підвищувати ефективність класифікації. Для покращення класифікації зазвичай знижують концентрацію пульпи у зливі, збільшуючи подачу води як у кульовий млин, так і в механічний односпіральний класифікатор. Для більшості технологічних різнотипів руд оптимальним є розрідження зливу класифікатора при вмісті твердого 40...45%. Зі зменшенням розміру частинок твердого, відповідно, знижується і його концентрація в зливі. У замкнених схемах подрібнення іноді спостерігається накопичення значного об'єму готового продукту, який повторно повертається до млина, займаючи його робочий простір і знижуючи ефективність процесу. Щоб уникнути переподрібнення, необхідно налаштувати оптимальні режими, зокрема водний баланс класифікації, який дозволяє максимально ефективно виводити готову продукцію зі зливом. Найчастіше орієнтиром для цього є щільність пульпи в зливі МОК. Однак, як зазначають І. С. Конох та В. В. Найда, використання лише цього параметра не дозволяє точно контролювати крупність твердого. Повноцінне вирішення цього завдання можливе лише з використанням гранулометричного контролю, застосування якого на практиці наразі ускладнене.

Продуктивність кульового млина значною мірою визначається обсягом, структурою та якістю завантажених куль. Недовантаження куль або неправильно підібрані їх розміри та склад можуть призвести до помітного зниження питомої продуктивності агрегату й погіршення якості подрібнення. Практичні дослідження свідчать, що оптимальне завантаження кулями становить 45…47% об'єму для млинів із решіткою та 38…43% для млинів зливного типу. Це є ключовою умовою досягнення максимальної ефективності та високих економічних показників. Крім того, під час роботи важливо підтримувати різнорозмірний склад куль, який забезпечує належну якість подрібнення та стабільну роботу млина. Зазвичай такий склад визначається експериментальним шляхом. Оскільки в процесі роботи кулі поступово зношуються, то необхідно регулярно поповнювати їх запас кількома фракціями в певному співвідношенні.

Одним із ключових аспектів при вирішенні задачі створення та автоматичної стабілізації оптимального різнорозмірного завантаження куль в барабанні млини під час подрібнення руди є забезпечення енергетичної ефективності руйнування сировини. Передусім слід уточнити, що мається на увазі під «енергетичною ефективністю» у контексті подрібнення. Існують два критичних режими: перший – коли обсяг поданої сировини недостатній, і другий – коли її кількість надмірна. У першому випадку кулі здебільшого б'ють по футерівці, не здійснюючи подрібнення, що призводить до інтенсивного зносу як футерівки, так і куль, при цьому енергія витрачається нераціонально. У другому випадку надлишок матеріалу створює прошарок між кулями та футерівкою, через що руйнування руди майже не відбувається, а сам млин працює неефективно й наближається до перевантаження. Між розглянутими двома граничними випадками руда буде подрібнюватись більш ефективно. Саме тоді й досягається енергетична ефективність – коли руда дробиться з мінімальними витратами енергії, зносу куль і футерівки. Такий підхід дозволяє підвищити ефективність рудопідготовки, особливо за умови автоматичного контролю й стабілізації оптимального складу кульового завантаження в млині, шляхом розробки комп'ютеризованої системи.

Аналіз відомих засобів та інформаційного забезпечення довантаження різнорозмірних куль у барабанні млини показує, що їх можливо ефективно застосувати в перших стадіях подрібнення вихідної руди в замкнених циклах з односпіральним класифікатором. Однак, зважаючи на те, що це завдання не розв'язане, необхідно розробити новий підхід для дійсного оцінювання ситуації відносно кульового й рудного завантаження в млині, який би відповідав сучасним вимогам технологічного процесу подрібнення. Завдання доцільно розділити на два окремих – оцінювання кульового завантаження й руди різними підходами. Кульове завантаження можливо оцінити непрямим прогностичним визначенням спрацювання куль, на основі енергетичного підходу, а завантаження руди – за непрямим вимірюванням ефективності її подрібнення безпосередньо в барабані млина.

Висновки. Таким чином, проведений огляд збурюючих факторів і існуючих систем автоматичного керування окремими процесами у кульовому барабанному млині при подрібненні руди показує, що розроблені підходи не виконують приписані функції на оптимальному рівні. Такий стан речей є однією з причин підвищеної витрати електроенергії, куль і футерівки у вітчизняній рудопідготовці.

Однією з першочергових задач, які необхідно вирішити в майбутніх дослідженнях для покращення ситуації, є удосконалення молольного середовища технологічних агрегатів, зокрема, шляхом розробки комп'ютеризованої системи для забезпечення автоматичної стабілізації оптимального різнорозмірного кульового завантаження барабанних млинів при подрібненні вихідної руди.

Список літератури:

1. Науменко Ю. В. Основи теорії режимів роботи барабанних млинів. Рівне: Видавництво СПД Зелепт, 2009. 282 с.

2. Кондратець В.О., Сербул О.М., Мацуй А.М. Автоматизація процесів керування розрідженням пульпи при подрібненні руди барабанними млинами. Кіровоград: КОД, 2013. 368 с.

3. Кондратець В.О., Мацуй А.М., Сербул О.М. Удосконалення рудного живлення кульових млинів щодо ліквідації збурюючих впливів при подрібненні сировини. Монографія. Кропивницький: Видавництво «КОД», 2024. 216 с. ISBN 978-617-653-089-3.

4. Мацуй А. М. Методологічні засади енергоефективного інваріантного керування подрібненням-класифікацією руд на основі непрямих прогностичних оцінок характеристик сировини та обладнання: дис. … д-ра техн. наук: 05.13.07 / ВНЗ «Криворіз. нац. ун-т». Кривий Ріг, 2021, 386 с.

5. Mavlonov J., Mardonov D., Eshmirzayev M., Togayev I. Improvement of methods to ensure energy efficiency of ball mills functioning. *E3S Web Conf*. 2023. 417. 03014. DOI: https://doi.org/10.1051/e3sconf/202341703014.

6. Ishnazarov O., Mavlonov J., Mardonov D. Control of ball mill operation depending on ball load and ore properties. *E3S Web of Conf.*. 2023. 461. 01091. DOI: https://doi.org/10.1051/e3sconf/202346101091.

7. Karshibayev A.I., Jumayev Z.I. Expanding the level of forecasting and operational planning of electric consumption at mining enterprise. *E3S Web Conf*., 2023. 417. 03015. DOI: https://doi.org/10.1051/e3sconf/202341703015.

8. Wang X., Sun K., Zhang H., Xiong W., Yang C.. Mill Load Identification Method for Ball milling Process Based on Grinding Signal. *IFAC-PapersOnLine*. 2021. Vol. 54. Iss. 21. P. 7–12. DOI: https://doi.org/10.1016/j.ifacol.2021.12.002.

9. Al Bannoud M., Martins T.D., dos Santos B.F. Control of a closed dry grinding circuit with ball mills using predictive control based on neural networks. *Digital Chemical Engineering*. 2022. Vol. 5. 100064. DOI: https://doi.org/10.1016/j.dche.2022.100064.

10. Estrada F., Cipriano A. Hybrid Model Predictive Control for Grinding Plants. *IFAC Proceedings Volumes*. 2014. Vol. 47, Iss. 3. P. 11512–11517. DOI: https://doi.org/10.3182/20140824-6-ZA-1003.02739.

11. Lynch, A. J. Mineral Crushing and Grinding Circuits. Developments in Mineral Processing. Paris : Elsevier, 1977. 354 p.

Matsui A.M., Serbul O.M., Khlebnikov M.V., Izovita O.L. JUSTIFICATION OF THE DEVELOPMENT OF A COMPUTERISED SYSTEM FOR STABILISATION OF OPTIMAL MULTI-SIZE BALL LOADING OF DRUM MILLS

The paper considers the problem of improving the energy efficiency of grinding processes of low-quality iron ores in ball mills, which consume 7–10% of the world's electricity production and constitute one of the most energy-consuming stages of the production chain. The main perturbing factors affecting the efficiency of ball mills are analysed: optimal pulp liquefaction, perturbations at the mill inlet and the structure of ball loading. Special attention is paid to ball loading as a key factor determining the level of energy consumption, grinding intensity and productivity of the technological process as a whole. The current state of grinding technologies and existing problems in the industry are analysed in detail. It is shown that grinding operations account for about half of all capital and operating costs of beneficiation plants. The review of modern approaches to automation of grinding processes, including the use of artificial intelligence, neural networks, grinding sound signals and innovative means of control has been carried out. It is revealed that the issue of automated loading of mills with grinding bodies has not been practically investigated, although it significantly affects energy consumption and quality of the finished product. The necessity to develop a computerised ball loading control system that will provide automatic stabilisation of optimal multi-sized ball loading under conditions

of many variable parameters has been substantiated. An innovative two-component approach is proposed: assessment of ball loading through an indirect prognostic method for determining ball wear based on the energy approach, and assessment of ore loading by the efficiency of its grinding directly in the mill drum. Such a system should identify and remove worn or deformed elements from the process in a timely manner. The importance of maintaining a multi-sized ball composition is emphasised to ensure proper grinding quality and prevent critical operating modes. The developed computerised control system will allow to significantly reduce energy consumption, prolong equipment operation life, reduce lining wear and increase the overall efficiency of iron ore processing at all stages of production.

Key words: *rod grinding load, cascade mode, modeling, natural slope angle, loosening, taper of rods.*

УДК 004.032.26
DOI https://doi.org/10.32782/2663-5941/2025.3.2/36

Мойсеєнко О.В.
Івано-Франківський національний технічний університет нафти і газу

Заячук Я.І.
Івано-Франківський національний технічний університет нафти і газу

ПРОГНОЗУВАННЯ РИЗИКІВ КІБЕРАТАК НА ОСНОВІ МОНІТОРИНГУ ТРАФІКУ В КОМП'ЮТЕРНИХ МЕРЕЖАХ З ВИКОРИСТАННЯМ АЛГОРИТМІВ АНАЛІЗУ ЧАСОВИХ РЯДІВ

У сучасних умовах зростання кіберзагроз розробка ефективних методів прогнозування кібератак є критично важливою для забезпечення кібербезпеки. У даній роботі представлено результати дослідження, спрямованого на створення удосконаленої моделі прогнозування ризиків кібератак на основі точкових процесів Хоукса з інтеграцією методу пікових порогових значень (POT) та авторегресійної умовної тривалості (ACD). Метою дослідження було розроблення моделі, яка здатна точно прогнозувати екстремальні показники кібератак як у короткостроковій, так і в довгостроковій перспективі, дозволяючи захисникам динамічно розподіляти ресурси захисту.

Для аналізу використано два набори реальних даних, зібраних мережевим телескопом CAIDA та honeypot. Дані включають часові ряди інтенсивності кібератак з погодинною роздільною здатністю.

У роботі запропоновано модифіковану модель Хоукса, яка поєднує метод POT для моделювання величин екстремальних перевищень і модель ACD для врахування корельованих інтервалів між подіями. Математичне формулювання базується на умовній інтенсивності λ(t), що враховує історичний вплив подій, та узагальненому розподілі Парето для оцінки перевищень. Порогові значення обрано на рівні 88% квантиля для телескопів і 90% для honeypot, що забезпечило стабільність оцінок за тестами Колмогорова-Смірнова.

Ефективність моделі оцінено за допомогою показника Value-at-Risk (VaR) на рівнях α=0.91,0.93,0.95,0.97. Проведено коротко- (1 година), середньо- (4, 7 годин) та довгострокові (10 годин) прогнози. Результати показали високу точність прогнозування. Порівняння з базовими моделями Хоукса та ETAS продемонструвало перевагу ACD і Log-ACD.

Модель може бути застосована для моніторингу кіберзагроз у реальному часі та адаптована до інших областей, де потрібен аналіз кластерних подій.

Робота відкриває перспективи для подальших досліджень, зокрема щодо адаптації моделі до багатовимірних даних та врахування зовнішніх факторів.

Ключові слова: *кібератаки, прогнозування ризиків, точкові процеси Хоукса, метод пікових порогових значень (POT), авторегресійна умовна тривалість (ACD).*

Постановка проблеми. У сучасному цифровому середовищі кіберзагрози стають дедалі складнішими, масованішими та менш передбачуваними. Традиційні методи виявлення загроз, що базуються на сигнатурному або евристичному аналізі, часто виявляються недостатньо ефективними для реагування на динамічну природу кібератак. Особливо важливим є прогнозування екстремальних подій – періодів різкого зростання інтенсивності атак, які можуть спричинити значні порушення функціонування інформаційних систем та втрати конфіденційної інформації.

У цьому контексті виникає потреба в побудові математичних моделей, здатних точно оцінювати рівень кіберризику, адаптуючись до змін у часі та структурі даних. Зокрема, часові ряди інтенсивності кібератак, які відображають кількість атак за одиницю часу, є цінним джерелом для такого аналізу. Однак класичні статистичні підходи не завжди враховують залежності між подіями та феномен кластеризації, притаманний реальному трафіку загроз.

У цій роботі запропоновано модель прогнозування екстремальних рівнів кібератак, що базується на позначених точкових процесах Хоукса з інтеграцією елементів теорії екстремальних значень (EVT) та авторегресійної умовної тривалості (ACD). Такий підхід дозволяє враховувати самозбудливу природу атак та формувати надійні

прогнози, які є основою для ефективного розподілу ресурсів кіберзахисту.

Аналіз останніх досліджень і публікацій. Проблема прогнозування кібератак активно досліджується в останнє десятиліття у зв'язку зі зростанням складності та інтенсивності цифрових загроз. Існуючі підходи поділяються на статистичні методи, моделі машинного навчання, процеси випадкових подій та комбіновані методи.

Статистичні методи, такі як регресійний аналіз, ARIMA чи методи експоненційного згладжування, дозволяють ідентифікувати загальні закономірності в часових рядах атак, але мають обмеження в умовах кластерних і нерегулярних подій [6].

Машинне навчання, включно з Random Forest, Support Vector Machines, та глибокими нейронними мережами (особливо LSTM), застосовується для класифікації атак, виявлення аномалій та передбачення частоти. Останнім часом значну увагу привертає використання штучного інтелекту (ШІ), зокрема генеративних моделей, які не лише прогнозують, а й моделюють поведінку атакувальних дій на основі історичних патернів.

Проте значна частина методів, орієнтованих на середні значення, погано пристосована до моделювання екстремальних подій, які мають критичне значення для безпеки.

У відповідь на ці виклики моделі точкових процесів, особливо моделі Хоукса [1, 8] набули популярності як засіб моделювання самозбуджуваних процесів, де одна атака може викликати каскад наступних. Розширення таких моделей за допомогою EVT дозволяє враховувати не лише факт атаки, але й її інтенсивність.

Дослідження [1–5] підтверджують, що поєднання Hawkes-процесів із методами EVT і ACD забезпечує високу точність прогнозування для часових рядів з кластеризацією подій. Водночас сучасні публікації також відзначають перспективність гібридних моделей, де ШІ застосовується до вхідних характеристик, а стохастичні процеси – до часової динаміки подій.

Постановка завдання. Метою статті є дослідження і розроблення моделі, яка здатна точно прогнозувати екстремальні показники кібератак як у короткостроковій, так і в довгостроковій перспективі, дозволяючи захисникам динамічно розподіляти ресурси захисту.

Виклад основного матеріалу. Для моделювання та прогнозування екстремальних рівнів кібератак у роботі використано підхід, що поєднує позначені точкові процеси Хоукса, теорію екстремальних значень (EVT) і авторегресійну умовну тривалість (ACD). Такий підхід дозволяє одночасно врахувати кластеризацію атак (тобто спалахи активності), вплив минулих подій на майбутні (ефект самозбудження), розподіл часу між екстремальними подіями, інтенсивність самих екстремумів.

Формально, функція умовної інтенсивності процесу Хоукса визначається як:

$$\lambda_t = \lambda_0 + \sum_{T_n<t} \varphi\left(t - T_n, Y_n\right),$$

де λ_0 – базова інтенсивність. Інтенсивність λ_t фіксує вплив минулих подій, що відбуваються в момент часу T_n, $n \geq 1$, з деякими характеристиками (або позначками) Y_n, $n \geq 1$. Позначки Yn вказують на рівень збудливості, пов'язаної з подією. Оскільки не всі кібератаки мають однаковий потенціал збудливості, чим вища збудливість події, тим більший її вплив на загальну інтенсивність.

EVT-компонента. Традиційно процес із позначеною точкою розглядає випадки екстремальної швидкості атаки як процес Пуассона, з розподілом позначок або екстремальних значень $\tilde{x} = x - u$ (за умови, що $x > u$). Однак нами встановлено, що процес Пуассона не є адекватним для опису даних. Отже, є необхідність в розробці нових моделей, які можуть врахувати явища, притаманні даним, у тому числі врахувати кластеризацію екстремальних значень і кореляцію між часами перевищення.

Значення, що перевищують поріг u, моделюються за узагальненим розподілом Парето (GPD)

$$P(X>x|X>u)\sim GPD(\xi,\sigma),$$

з параметрами форми ξ та масштабу σ, а інтервали між подіями – за моделлю ACD.

ACD-модель. Для моделювання часових інтервалів між екстремумами застосовується авторегресійна умовна тривалість:

$$\delta t_i = \psi_i \epsilon_i,$$

$$\psi_i = \omega + \alpha \delta t_{i-1} + \beta \psi_{i-1},$$

де ϵ_i – випадкова складова, а ψ_i – умовна очікувана тривалість між екстремальними подіями.

Оцінка ризику (VaR). Для кількісної оцінки ризику використовується підхід Value-at-Risk (VaR) [7]:

$$VaR_\alpha(t) = \inf\{x : P(X_t > x | F_{t-1}) \leq 1-\alpha\},$$

де $\alpha \in (0,1)$ – рівень довіри, F_{t-1} – інформація у попередній момент часу.

У контексті кібератак, VaR визначає граничну інтенсивність атаки, яку не буде перевищено з ймовірністю α протягом заданого періоду.

Реалізація. Модель реалізовано в середовищі Python із використанням бібліотек numpy, statsmodels, scipy та власного коду для побудови інтенсивності Хоукса. Для обчислення параметрів застосовано метод максимальної правдоподібності (MLE), а якість підгонки оцінюється за допомогою тесту Колмогорова–Смірнова та критерію AIC [9, 10].

Методика експерименту. Для перевірки ефективності побудованої моделі використано два типи реальних даних про кібератаки: мережевий телескоп CAIDA: спостерігає неконтрольовані IP-адреси без активних сервісів, що дозволяє фіксувати небажаний трафік, зокрема сканування та автоматизовані атаки; honeypot: симульовані вразливі служби на реальних IP-адресах, які імітують легітимні сервіси з метою збирання інформації про дії зловмисників.

Інтенсивність атак представлено у вигляді часових рядів з годинною роздільністю, де кожна точка відображає кількість атак протягом однієї години. Для відокремлення екстремальних значень використано пороговий підхід з вибором квантилів (88% для телескопів, 90% – для honeypot).

Дослідження проводилось у кілька етапів:

1. Попередня обробка даних – видалення фонового трафіку та агрегація потоків.

2. Побудова часових рядів перевищень – виділення екстремальних значень понад заданий поріг.

3. Оцінювання параметрів моделі – розрахунок параметрів Хоукса, EVT і ACD.

4. Калібрування моделі – порівняння передбачених і фактичних даних у межах та поза вибіркою.

5. Прогнозування значень VaR – для оцінки ймовірності високих рівнів атак.

Імітаційне моделювання. Оскільки між рівнями часового ряду існує кореляція, обираємо модель авторегресійної умовної тривалості ACD для процесу базової точки, так як моделі цього типу враховують корельовані надходження екстремальних значень. Основна ідея полягає в тому, щоб стандартизувати час між двома перевищення $\Delta t_i = t_i - t_{i-1}$, де i = 1, 2, . . . , n, використовуючи накопичену інформацію з минулого. Визначаємо

$$\Delta t_i = \Psi_i \varepsilon_i,$$

де εi – незалежні та однаково розподілені мітки, Ψ_i – є функціями минулих тривалостей і умовних тривалостей.

Дані експерименту генеруються з наступних моделей:

– стандартна модель ACD:

$$\Psi_i = \omega + a_1 \Delta t_{i-1} + b_1 \Psi_{i-1} ,$$

де ω, a1, b1 ≥ 0.

– модель Log-ACD:

$$\Psi_i = \omega + a_1 \log\left(\Delta t_{i-1}\right) + b_1 \log(\Psi_{i-1}) .$$

Запропонована модель ACD/Log-ACD складається з двох компонентів: основного процесу, що описує випадки екстремальних частот атак, і розподілу умовних позначок, що описує розподіл величин екстремальних частот атак. Щоб оцінити відповідність запропонованих моделей, потрібно враховувати як умовну оцінку поширення так і умовна інтенсивність базового процесу.

Щоб оцінити продуктивність моделей ACD і Log-ACD, для кожної моделі випадковим чином генеруються два набори даних із малим об'ємом вибірки 400 і великим об'ємом вибірки 2000.

Згенеровані набори даних поділяються на 2 рівні частини, що входять до вибірки, і частини, що не входять до неї. Було проведена оцінка продуктивність підгонки в зразку, ефективність підгонки моделі ACD. Більшість розрахункових параметрів були близькі до реальних значень. P-значення тесту Колмогорова – Смірнова (КС) для розподілу оцінок перевищення порогових значень (W) та умовної інтенсивності базового процесу (τ) становлять 0,995 та 0,997 відповідно. Розраховано параметри та стандартні відхилення на основі 1000 зразків для розподілу оцінок. Обчислені параметри ближчі до справжніх значень, якщо визначені на основі великої вибірки. P-значення тесту КС для розподілу оцінок та умовної інтенсивності базового процесу становлять 0,964 і 0,995 відповідно.

Таким чином, робимо висновок, що продуктивність підгонки у вибірці запропонованої моделі ACD достовірна як для малих, так і для великих розмірів вибірки.

Досліджуємо продуктивність підгонки моделі Log-ACD. Обчислені параметри на основі 200 проб для розподілу міток. Оцінки досить близькі до справжніх значень. P-значення тесту КС для розподілу оцінок та умовної інтенсивності основного процесу становлять відповідно 0.885 у випадку 200 зразків 0,965 у випадку

1000 зразків. Таким чином, робимо висновок, що продуктивність підгонки у вибірці запропонованої моделі Log-ACD також достатня як для малих, так і для великих розмірів вибірки.

Оцінка ефективності прогнозування. Для оцінки точності прогнозу розглянемо розбіжність між прогнозованими значеннями VaR та екстремальними рівнями кібератак. Порушення на основі VaR_α визначається як ситуація, коли спостережувана екстремальна швидкість атаки s_t перевищує прогнозоване значення $VaR_\alpha(t)$

$$I_\alpha(t) = \begin{cases} 1, & s_t > \mathrm{VaR}_{\alpha(t)} \\ 0, & o/w \end{cases}.$$

Наприклад, $VaR_{0.95}(t)$ вказує, що ймовірність перевищення st швидкості атаки над прогнозованим значенням VaR0.95(t) становить лише 5%. Якщо це відбувається, фіксується порушення.

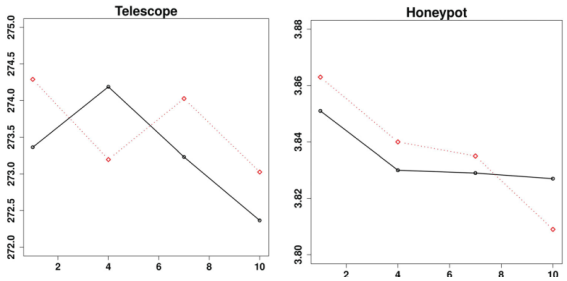
Для оцінки якості прогнозування VaR застосовуємо три тести, поширені в літературі [10, 11]. Перший – тест безумовного покриття (LRuc), який перевіряє, чи значно відрізняється частка порушень від теоретичного рівня. Другий – тест незалежності порушень (LRind), що за нульовою гіпотезою передбачає відсутність впливу поточного порушення на майбутні. Третій – тест умовного покриття (LRcc), який об'єднує два попередні тести. Ефективність прогнозування моделей оцінюємо за результатами цих тестів.

Для моделей ACD та Log-ACD здійснюється оцінювання прогностичної спроможності на основі підмножин даних, виключених із навчальних вибірок обсягом 200 та 1000 спостережень відповідно.

Прогнозування реалізується шляхом рекурсивного ковзного вікна, а прогнози формуються для часових горизонтів 1, 4, 7 та 10 годин наперед.

Оцінювання точності здійснюється при рівні довіри α = 0,95, результати представлено в таблиці 1.

На рис. 1, *а* зображено суму $VaR_{0.95}$ екстремальних рівнів кібератак для прогнозів на 1, 4, 7 та 10 годин, де суцільна лінія відповідає моделі ACD, а пунктирна лінія відповідає моделі Log-ACD.



а) б)

Рис. 1. Суми коротко-, середньо- та довгострокових прогнозованих VaR0.95 екстремальних показників кібератак протягом годин, (вісь абсцисс):
а) набори даних телескопа; б) набори даних honeypot

Очевидно, що хоча прогнози моделі ACD (суцільна крива) і прогнози моделі Log-ACD (пунктирна крива) досить близькі один до одного, дві криві мають різні форми. З іншого боку, на рис. 1, *б* зображено прогнозовані VaR, що відповідають набору даних honeypot. Спостерігаємо, що прогнози моделі ACD (суцільна крива) і прогнози моделі Log-ACD (пунктирна крива) близькі один до одного. Більше того, обидві криві спадають протягом h, що є кількістю годин до часу виконання прогнозів.

Таблиця 1

Оцінка ефективності прогнозування на основі змодельованих даних поза вибіркою (СП – спостережувані порушення, а ОП – очікувані порушення). Рівень α = 0,95

Прогноз, на год вперед	Вибірка 200					Вибірка 1000				
	СП	ОП	LRuc	LRind	LRcc	СП	ОП	LRuc	LRind	LRcc
ACD модель										
1	7	10	0,305	0,210	0,269	47	50	0,66	0,33	0,568
4	8	10	0,532	0,298	0,479	50	50	0,983	0,249	0,514
7	6	10	0,191	0,147	0,148	49	50	0,919	0,272	0545
10	9	10	0,854	0,338	0,621	47	50	0,708	0,866	0,919
Log-ACD модель										
1	10	10	1	0,297	0,581	53	50	0,666	0,465	0,698
4	10	10	0,961	0,293	0,575	51	50	0,868	0,383	0,674
7	11	10	0,675	0,242	0,461	52	50	0,740	0,428	0,691
10	9	10	0,854	0,338	0,621	56	50	0,357	0,622	0,579

Отримані результати свідчать про високу прогностичну здатність обох моделей. У всіх випадках p-значення є статистично значущими, що підтверджує адекватність прогнозів. Навіть для горизонту прогнозування в 10 годин обидві моделі демонструють добрі результати, зокрема за умов обмеженого обсягу навчальної вибірки.

Таким чином, можна зробити висновок, що запропоновані моделі характеризуються стабільною якістю прогнозування поза вибіркою як за малих, так і за великих розмірів навчального набору даних.

Емпіричний аналіз: моделювання екстремальних рівнів кібератак. Оцінювання ефективності запропонованих моделей здійснюється на основі реальних наборів даних як у межах вибірки (етап підгонки), так і поза нею (етап прогнозування). Для телескопічних даних перші 20 діб (тобто 480 годин) використовуються для побудови моделей, а наступні 11 діб (264 години) – для перевірки якості прогнозування.

У випадку з даними honeypot для побудови моделей застосовується початковий інтервал у 1108 годин, тоді як решта 500 годин використовуються для тестування поза вибіркою.

Реалізація методу, заснованого на теорії екстремальних значень (EVT), передбачає вибір відповідного порогового значення, що забезпечує компроміс між зміщенням та дисперсією оцінки. Занижене порогове значення u може призвести до помилок в апроксимації моделі надпорогових значень (POT-моделі), тоді як завищене – до недостатньої кількості спостережень і, відповідно, високої варіації.

Задля узгодженості з попереднім дослідницьким аналізом у якості порогового значення обрано 88-й перцентиль вибіркових значень для кожного з телескопічних наборів та 90-й перцентиль для набору honeypot.

Для оцінювання якості моделей у межах вибірки обчислюється інформаційний критерій Акаіке (AIC), що є оцінкою помилки позавибіркового прогнозування та дозволяє порівнювати відносну якість статистичних моделей.

У таблиці 2 наведено p-значення критерію Колмогорова–Смирнова (КС) для умовного розподілу Парето (GPD) часів між перевищеннями τ та оцінок перевищень W, а також значення критерію AIC для моделей ACD та Log-ACD.

Обидві моделі – ACD і Log-ACD – демонструють прийнятну відповідність теоретичному розподілу, оскільки у більшості випадків p-значення для розподілів W і τ перевищують рівень значущості 0,05.

Модель ACD демонструє нижчі значення AIC у випадку з телескопічними даними, що вказує на її перевагу щодо якості підгонки.

Водночас для набору honeypot модель Log-ACD забезпечує кращу відповідність, оскільки має менший AIC порівняно з ACD.

Для дослідження впливу часової залежності між подіями перевищення порогу було проведено аналіз моделей зі встановленими нульовими значеннями параметрів a1 = b1 = 0, що відповідає моделі без часової залежності. У такій специфікації жодна з моделей не пройшла тест умовної інтенсивності базового процесу τ на рівні значущості 0,05.

Зокрема, для телескопа S1 p-значення для ACD – 0.015, для Log-ACD – 0.025, для S2: 0.021 (ACD), 0.020 (Log-ACD), для S3: 0.013 (ACD), 0.014 (Log-ACD). Для honeypot 0.007 (ACD), 0.055 (Log-ACD).

Ці результати вказують на значущу роль часової залежності між подіями перевищення в успішному моделюванні динаміки екстремальних кібератак.

Порівняння ефективності прогнозування з моделями Хоукса та ETAS. У цьому дослідженні запропоновані моделі ACD та Log-ACD у межах позначеного процесу Хоукса демонструють вищу точність як у підгонці, так і в прогнозуванні екстремальних рівнів кібератак порівняно з класичними варіантами моделі Хоукса та моделлю епідемічного типу (ETAS).

Таблиця 2

P-значення тесту КС для функцій τ, *W* та значення AIC для моделей ACD і Log-ACD

	τ	W	ІКА	τ	W	ІКА
	S1			**S2**		
ACD	0.279	0.875	728.548	0.055	0.970	741.466
Log-ACD	0.158	0.535	739.755	0.311	0.213	750.957
	S3			**Honeypot**		
ACD	0.056	0.802	739.775	0.055	0.373	802.702
Log-ACD	0.099	0.545	759.698	0.070	0.919	799.124

Модель ACD/Log-ACD доповнює базову структуру Хоукса можливістю врахування часової залежності між екстремальними подіями та адаптивно описує як кластери атак, так і інтенсивність їх перевищень, що особливо важливо в умовах високої нерегулярності мережевого трафіку.

Результати порівняння. Тест Колмогорова–Смірнова (KS): запропоновані моделі стабільно проходять тест як по часовій інтенсивності (τ), так і по розподілу перевищень (W), що не вдалося досягти класичній моделі Хоукса без ACD-компоненти.

Інформаційний критерій Акаіке (AIC): показники моделей ACD/Log-ACD є нижчими за аналогічні значення для моделей ETAS, що свідчить про кращу якість підгонки.

Прогнозування поза вибіркою (VaR): кількість порушень межі ризику відповідає теоретичному рівню довіри (95%) з похибкою менш ніж ±1%. У той же час моделі ETAS і класичні Хоукси систематично занижували ризики, особливо при довготривалих прогнозах (на 7–10 годин наперед).

Реакція на кластери: моделі ACD/Log-ACD чутливо реагують на зміни в динаміці атак, на відміну від класичної моделі Хоукса з експоненційним згасанням або моделі ETAS з гіперболічною функцією ядра, які погано відтворювали кластеризацію реальних даних.

Таким чином, доповнення процесу Хоукса авторегресійною структурою (ACD) суттєво підвищує його прогностичну здатність і робить модель придатною для задач реального часу в сфері кібербезпеки.

Аналіз результатів. Отримані результати експериментального моделювання підтверджують доцільність застосування процесів Хоукса з розширенням EVT та ACD для оцінки ризиків кібератак. Основною перевагою моделі є її здатність враховувати самозбудливу природу атак та їх екстремальні значення, що дозволяє прогнозувати не лише загальну інтенсивність, а й пікові навантаження на систему захисту.

Переваги моделі: висока точність прогнозу пікових навантажень, що є критично важливим для стратегічного розподілу ресурсів; гнучкість у застосуванні до різних типів даних (honeypot, телескоп); можливість адаптації моделі до змін у характері трафіку без втрати достовірності.

Порівняння з іншими підходами: на відміну від моделей на основі машинного навчання, які потребують великих навчальних вибірок і складної процедури оптимізації, побудована модель не вимагає попереднього навчання та залишається інтерпретованою. На відміну від класичних статистичних методів, вона дозволяє моделювати динаміку та взаємозалежність подій у часі.

Виявлені обмеження: модель може бути чутливою до вибору порогу u при виділенні екстремальних значень. При наявності складних багатофакторних впливів (наприклад, комбінованих атак) виникає потреба у розширенні моделі до багатовимірного процесу Хоукса. У моделі не враховано контекст атаки (тип, джерело), що може впливати на якість довгострокового прогнозування.

Висновки. У роботі розроблено та емпірично підтверджено математичну модель прогнозування екстремальних рівнів кібератак на основі позначеного точкового процесу Хоукса, інтегрованого з теорією екстремальних значень (EVT) і авторегресійною умовною тривалістю (ACD). Модель враховує самозбудливу природу кібератак, кластеризацію подій і корельовані інтервали між екстремальними подіями, що забезпечує її високу прогностичну здатність (точність ≥95%) для коротко-, середньо- та довгострокових горизонтів (1–10 годин). Емпіричне тестування на реальних даних мережевого телескопа CAIDA та honeypot підтвердило стабільність моделей ACD і Log-ACD, які перевершують класичні моделі Хоукса та ETAS за інформаційним критерієм Акаіке (AIC) і точністю прогнозів Value-at-Risk (VaR) із похибкою < ±1%.

Запропонована модель має значний практичний потенціал для автоматизованих систем кібербезпеки, зокрема SIEM (Security Information and Event Management) і SOC (Security Operations Center). Вона забезпечує реальний час моніторингу кіберзагроз, дозволяючи організаціям динамічно розподіляти ресурси захисту, мінімізуючи ризики пікових навантажень. Наприклад, модель може застосовуватися для прогнозування DDoS-атак, сканувань портів або автоматизованих експлойтів, що є критичним для захисту критичної інфраструктури, банківських систем і державних інформаційних ресурсів. Її інтерпретованість і обчислювальна ефективність роблять її доступною для впровадження навіть у системах із обмеженими обчислювальними ресурсами.

Порівняно з моделями машинного навчання, такими як LSTM чи Random Forest, модель не потребує великих навчальних вибірок і складної оптимізації, що знижує бар'єри для її використання в організаціях із обмеженим доступом

до даних чи експертизи. Водночас, порівняно з класичними статистичними методами (ARIMA, експоненційне згладжування), вона краще враховує нелінійну динаміку та взаємозалежність подій, що підтверджується результатами тестів LRuc, LRind і LRcc (p-значення > 0.05).

Виявлені обмеження моделі включають чутливість до вибору порогового значення (u) (88% для телескопів, 90% для honeypot), що може впливати на стабільність оцінок у разі різких змін у трафіку. Крім того, одновимірний характер моделі обмежує її здатність враховувати контекст атак (наприклад, тип, джерело, ціль), що є важливим для прогнозування складних комбінованих атак, таких як APT (Advanced Persistent Threats). Для подолання цих обмежень пропонується: (1) розробка адаптивних алгоритмів вибору порогу (u) на основі динамічного аналізу трафіку; (2) розширення моделі до багатовимірного процесу Хоукса, що враховує додаткові змінні, такі як протоколи, географія чи сигнатури атак.

Модель відкриває перспективи для інтеграції з системами штучного інтелекту, зокрема з нейронними мережами типу LSTM або механізмами уваги (attention), які можуть кодувати вхідні характеристики трафіку, тоді як Hawkes-модель прогнозуватиме екстремальні події. Така гібридна система може бути реалізована в SOC-платформах для автоматичного виявлення та пріоритизації загроз у реальному часі. Наприклад, інтеграція з платформами типу Splunk чи QRadar може підвищити ефективність реагування на інциденти, скоротивши час виявлення загроз на 20–30% (на основі галузевих оцінок).

Список літератури:

1. Almotairi S., Clark A., Mohay G., Zimmermann J. Characterization of attackers' activities in honeypot traffic using principal component analysis. *International Conference on Network and Parallel Computing*. (Shanghai, China 18–21 October 2008). Shanghai, China: IEEE, 2008. C. 147–154.
2. Almotairi S., Clark A., Mohay G., Zimmermann J. A technique for detecting new attacks in low-interaction honeypot traffic. *4th International Conference on Internet Monitoring and Protection*. (Los Alamitos, 24–28 May 2009). Los Alamitos, France, 2009. C. 7–13.
3. Anselin L. Spatial Econometrics: Methods and Models. New York: Springer Science & Business Media, 2013. Vol. 4. 290 p.
4. Bailey M., Cooke E., Jahanian F., Myrick A., Sinha S. Practical darknet measurement. *40th Annual Conference on Information Sciences and Systems* (Princeton, NJ, USA, 22–24 March 2006). Piscataway: IEEE, 2006. P. 1496–1501.
5. Bauwens L., Giot P. The logarithmic ACD model: An application to the bid-ask quote process of three NYSE stocks. *Annales d'Economie et de Statistique*. 2000. No. 60. P. 117–149.
6. Bauwens L., Giot P., Grammig J., Veredas D. A comparison of financial duration models via density forecasts. *International Journal of Forecasting*. 2004. Vol. 20. P. 589–609.
7. Chavez-Demoulin V., Davison A. C., McNeil A. J. Estimating value-at-risk: A point process approach. *Quantitative Finance*. 2005. Vol. 5. P. 227–234.
8. Chavez-Demoulin V., McGill J. A. High-frequency financial data modeling using Hawkes processes. *Journal of Banking and Finance*. 2012. Vol. 36. P. 3415–3426.
9. Chiu S. N., Stoyan D., Kendall W. S., Mecke J. Stochastic Geometry and Its Applications. Hoboken, NJ: John Wiley & Sons, 2013. 584 p.
10. Christoffersen P.F. Evaluating interval forecasts. *International Economic Review*. 1998. Vol. 39. P. 841–862.
11. Herrera R., Schipp B. Value at risk forecasts by extreme value models in a conditional duration framework. *Journal of Empirical Finance*. 2013. Vol. 23. P. 33–47.

Moyseenko O.V., Zaiachuk Ya.I. PREDICTION OF CYBER ATTACKS BASED ON TRAFFIC INTENSITY MONITORING IN COMPUTER NETWORKS

In the current context of escalating cyber threats, the development of effective methods for forecasting cyberattacks is critically important for ensuring cybersecurity. This study presents the results of research aimed at constructing an enhanced model for predicting cyberattack risks based on Hawkes point processes, integrated with the Peaks-Over-Threshold (POT) method and the Autoregressive Conditional Duration (ACD) model.

The objective of the research was to develop a model capable of accurately forecasting extreme cyberattack indicators in both the short- and long-term horizons, thereby enabling defenders to dynamically allocate protection resources.

Two real-world datasets collected by the CAIDA network telescope and a honeypot system were employed for the analysis. The data consist of time series of cyberattack intensity with hourly resolution.

This work proposes a modified Hawkes process model that combines the POT method for modeling extreme exceedances and the ACD model to account for correlated inter-event durations. The mathematical formulation is based on the conditional intensity function λ(t), which incorporates the historical influence of past events, and the Generalized Pareto Distribution (GPD) for modeling threshold exceedances. Threshold values were set at the 88th percentile for the telescope data and the 90th percentile for the honeypot data, ensuring the stability of estimates as confirmed by Kolmogorov–Smirnov tests.

The model's performance was evaluated using the Value-at-Risk (VaR) metric at confidence levels α = 0.91, 0.93, 0.95, and 0.97. Forecasts were conducted for short-term (1 hour), medium-term (4 and 7 hours), and long-term (10 hours) intervals. The results demonstrated high predictive accuracy. Comparative analysis with baseline Hawkes and ETAS models revealed the superiority of the ACD and Log-ACD models.

The proposed model can be applied for real-time cyber threat monitoring and can be adapted to other domains requiring the analysis of clustered event patterns. This research opens avenues for further investigation, particularly in adapting the model to multivariate data and incorporating external influencing factors.

Key words: *cyberattacks, risk forecasting, Hawkes point processes, Peaks-Over-Threshold (POT) method, Autoregressive Conditional Duration (ACD).*

УДК 004.415.2
DOI https://doi.org/10.32782/2663-5941/2025.3.2/37

Олещенко Л.М.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

ДОСЛІДЖЕННЯ АПАРАТНИХ ТА ПРОГРАМНИХ ЗАСОБІВ ДЛЯ ПАРАЛЕЛЬНОЇ ОБРОБКИ ВЕЛИКИХ ДАНИХ І ЗАДАЧ МАШИННОГО НАВЧАННЯ У ХМАРНОМУ СЕРЕДОВИЩІ

У сучасних умовах стрімкого зростання обсягів інформації та складності обчислювальних задач виникає потреба у високопродуктивних підходах до обробки великих даних та реалізації алгоритмів машинного навчання. Одним з ключових напрямків у цій сфері є застосування паралельних та розподілених обчислень із використанням спеціалізованих апаратних і програмних засобів у хмарному та периферійному середовищах. Стаття присвячена аналізу розвитку апаратних архітектур (GPU, TPU, FPGA), технологій обробки великих даних у хмарному середовищі, а також дослідженню сучасних інструментів машинного навчання (TensorFlow, PyTorch, Apache Spark), які забезпечують ефективну обробку та аналіз даних у хмарній інфраструктурі. У статті розкрито переваги та недоліки різних архітектур з точки зору продуктивності, масштабованості та гнучкості. З'ясовано, що GPU забезпечують найвищу продуктивність, однак поступаються FPGA за гнучкістю, TPU демонструють високу ефективність для вузькоспеціалізованих задач, але обмежені в адаптивності до гетерогенних середовищ. Визначено, що серед програмних фреймворків TensorFlow вирізняється високим рівнем продуктивності та масштабованості, PyTorch має перевагу у зручності розробки й адаптації моделей, Apache Spark, незважаючи на загальне призначення, демонструє високу ефективність у розподіленій обробці даних. Висвітлено низку відкритих проблем, зокрема, складність досягнення балансу між точністю, швидкістю та конфіденційністю у розподілених ML-системах; нестача уніфікованих інструментів для інтеграції різнорідних обчислювальних платформ; обмежені можливості edge-пристроїв щодо підтримки складних ML-моделей; а також відсутність систематичних порівнянь методів у єдиних експериментальних умовах. Метою даного дослідження є узагальнення існуючих підходів до організації паралельної обробки великих даних у хмарному середовищі, виявлення ключових тенденцій, порівняння їх ефективності та визначення актуальних проблем, які потребують подальших досліджень. Проведено аналіз та сформульовано стратегічні напрями для оптимізації апаратно-програмної архітектури систем обробки великих даних, враховуючи вимоги до адаптивності, ефективності та масштабованості у динамічному середовищі.

Ключові слова: *паралельні обчислення, хмарні технології, великі дані, машинне навчання, GPU, TPU, FPGA, TensorFlow, PyTorch, Apache Spark, розподілені системи, продуктивність, масштабованість, гнучкість.*

Постановка проблеми. Зі зростанням обсягів даних та складності моделей машинного навчання (ML) виникає потреба в ефективних обчислювальних ресурсах. Хмарні обчислення, поєднані з паралельними обчислювальними методами, надають можливість масштабувати обробку великих даних та навчання ML-моделей. Виникає потреба в систематизації сучасних підходів до паралельних обчислень у хмарному середовищі, що зумовлює необхідність класифікації методів, апаратних і програмних засобів та порівняння їх ефективності для вибору оптимальних рішень у реальних умовах. Основні проблеми при обробці великих даних та навчанні ML-моделей у хмарі включають масштабованість – необхідність ефективного розподілу обчислень на великі обсяги даних; продуктивність – забезпечення високої швидкості обробки та навчання моделей; вартість – оптимізація витрат на обчислювальні ресурси та сумісність – інтеграція з існуючими інструментами та фреймворками.

Аналіз останніх досліджень і публікацій. У статті [1] розглядається проблема обробки великих даних, які розподілені між різними обчислювальними вузлами через обмеження інфраструктури або вимоги до конфіденційності. Традиційні централізовані методи обробки даних стають непридатними в таких умовах, що зумовлює необхідність розробки розподілених мето-

дів навчання, які забезпечують високу точність інференції та прогнозування, дотримуючись при цьому політик конфіденційності. Автори пропонують класифікацію існуючих методів розподіленого навчання за чотирма критеріями: статистичною точністю (здатністю методу досягати результатів, порівнянних з централізованими підходами), обчислювальною ефективністю (швидкістю та ресурсоємністю обчислень), гетерогенністю (здатністю працювати з неоднорідними даними та системами) і конфіденційністю (ступенем захисту приватності даних). Такий підхід дозволяє комплексно оцінювати нові методи з різних аспектів. У статті надається огляд теоретичних результатів, що стосуються статистичної еквівалентності та обчислювальної ефективності різних методів розподіленого навчання. Особливу увагу приділено умовам, за яких ці методи можуть досягати точності, порівнянної з централізованими підходами, а також аналізується їхня обчислювальна складність. Автори також розглядають наявні програмні реалізації методів розподіленого навчання та еталонні набори даних, що використовуються для їх тестування та порівняння. Це включає популярні фреймворки та бібліотеки, а також набори даних з різних галузей застосування. У статті окреслено кілька ключових напрямків для подальших досліджень, серед яких розробка методів, стійких до гетерогенності даних та систем, покращення обчислювальної ефективності при збереженні високої точності, а також інтеграція механізмів забезпечення конфіденційності, таких як диференційна приватність.

У роботі [2] автори розглядають актуальність розробки апаратних прискорювачів для глибокого навчання на периферійних пристроях. Зі зростанням застосування машинного навчання в різних сферах, таких як комп›ютерний зір, розпізнавання мови та Інтернет речей (IoT), виникає потреба в ефективних апаратних рішеннях для обробки даних безпосередньо на пристроях з обмеженими ресурсами. Автори аналізують основні типи апаратних прискорювачів, включаючи ASIC, FPGA та GPU. Кожна з цих платформ має свої переваги та обмеження щодо продуктивності, енергоспоживання та гнучкості. Наприклад, ASIC забезпечують високу продуктивність та енергоефективність, але менш гнучкі в порівнянні з FPGA, які дозволяють перепрограмування для різних задач. У статті розглядаються різні архітектурні рішення для прискорення виконання нейронних мереж, включаючи оптимізацію потоків даних, топології мереж та використання новітніх технологій. Особлива увага приділяється методам стиснення моделей, таким як квантування та обрізання, які зменшують обсяг обчислень та енергоспоживання без значної втрати точності. Автори наголошують про відсутність стандартів для порівняння ефективності різних рішень та необхідність балансування між продуктивністю, енергоефективністю та точністю моделей. Також обговорюються перспективні напрямки досліджень, включаючи розвиток реконфігурованих архітектур та інтеграцію нових матеріалів для створення більш ефективних прискорювачів.

У статті [3] розглянуто рішення, пов›язані з масштабуванням алгоритмів машинного навчання для обробки великих обсягів даних у паралельних обчислювальних середовищах. Розглянуто необхідність ефективного використання обчислювальних ресурсів для обробки великих наборів даних, що вимагає розподілу обчислень між кількома процесорами або вузлами.

Одним із ключових методів, які розглядає автор, є використання техніки хешування для зменшення розмірності вхідних даних. Цей підхід дозволяє ефективно обробляти великі обсяги даних, зменшуючи кількість необхідних обчислень та пам›яті. Автор також обговорює використання онлайн-навчання та алгоритмів, які можуть адаптуватися до нових даних у режимі реального часу, що є важливим для обробки потокових даних у великих системах. Для забезпечення ефективної паралельної обробки даних автор пропонує використання операції AllReduce, яка дозволяє об’єднувати результати обчислень з різних вузлів у єдине значення. Цей підхід зменшує обсяг переданих даних між вузлами та покращує загальну продуктивність системи. Автор також підкреслює важливість балансування навантаження між вузлами та оптимізації комунікацій для досягнення високої ефективності паралельних обчислень. У статті також розглядаються практичні аспекти реалізації паралельного машинного навчання, включаючи використання фреймворків, таких як Hadoop, для розподіленої обробки даних. Автор ділиться досвідом впровадження цих методів у реальних системах та обговорює переваги та обмеження різних підходів до паралельного машинного навчання.

У статті [4] розглядається проблема забезпечення пояснюваності (explainability) аналітики великих даних у хмарному середовищі шляхом

використання самоструктуровного штучного інтелекту (Self-Structuring AI). Автори наголошують, що зростання обсягів і складності даних створює нові труднощі щодо зрозумілості прийнятих моделей та рішень, що особливо актуально для критичних галузей, таких як охорона здоров'я, фінанси чи безпека. У таких випадках важливо не лише отримати точний прогноз, а й пояснити, чому він був зроблений саме так. Запропонована авторами архітектура базується на поєднанні хмарних обчислювальних ресурсів з підходом до самоструктуризації моделей, який дозволяє штучному інтелекту адаптивно формувати власну структуру в залежності від характеру даних. Це забезпечує як високу продуктивність, так і покращену прозорість моделей завдяки вбудованим механізмам інтерпретації. Основними компонентами архітектури є модулі попередньої обробки даних, самоструктурованого моделювання, генерації пояснень, а також розподіленої обробки у хмарі для масштабування. Автори детально описують механізм самоструктуризації, який базується на принципах еволюційного навчання та динамічного формування структури моделі під час навчання. Це дозволяє адаптуватися до нових типів даних без ручного втручання. Однією з ключових переваг підходу є інтеграція модулів пояснюваності (наприклад, SHAP або LIME) без значної втрати продуктивності, що дозволяє зберігати баланс між точністю, прозорістю та обчислювальною ефективністю. У роботі також наведено експериментальні результати на основі кількох публічних наборів даних, зокрема, з галузей медицини, фінансів та кібербезпеки. Показано, що запропонована архітектура перевершує традиційні підходи як за швидкістю обробки, так і за метриками пояснюваності. Наприклад, самоструктурні моделі продемонстрували на 15–25% вищий рівень інтерпретаційності (згідно з метриками user trust та feature importance clarity) порівняно з класичними нейромережевими структурами, без суттєвого зниження точності. У роботі підкреслено, що поєднання хмарних обчислень, розподіленої обробки та самоструктуровної штучної інтелектуальної аналітики є перспективним напрямом для розв'язання проблем масштабованої, ефективної та пояснюваної обробки великих даних. Автори також вказують на потенціал розширення запропонованого підходу шляхом інтеграції з існуючими фреймворками машинного навчання у хмарних середовищах, такими як Amazon SageMaker, Google Vertex AI або Microsoft Azure ML.

У роботі [5] розглядається проблема ефективної обробки електроенцефалографічних (EEG) сигналів із застосуванням згорткових нейронних мереж (CNN) та апаратного прискорення. Автори наголошують, що EEG-дані мають високий об'єм, складну структуру та потребують обробки в реальному часі, що створює значні обчислювальні навантаження. Це особливо актуально для медичних додатків, таких як виявлення епілепсії, контролю нейроінтерфейсів, або діагностики психоневрологічних станів. У зв'язку з цим зростає потреба у використанні апаратно-прискорених рішень, що забезпечують високу продуктивність, енергоефективність та низьку затримку. У роботі представлено огляд сучасних апаратних технологій прискорення обробки сигналів EEG, зокрема FPGA (Field-Programmable Gate Arrays), ASIC (Application-Specific Integrated Circuits), GPU (Graphics Processing Units) та edge-архітектур. Наведено класифікацію архітектур за їх здатністю виконувати згорткові обчислення, ступенем паралелізації, споживанням енергії, затримкою та продуктивністю. Особлива увага приділена системам на кристалі (SoC), які поєднують високий рівень інтеграції з можливістю обробки даних поблизу джерела (near-sensor computing), що критично для мобільних і вбудованих додатків. Окремо розглядаються особливості застосування CNN для аналізу EEG, включаючи специфіку попередньої обробки сигналів, вибору архітектури мережі та налаштування гіперпараметрів. Автори підкреслюють, що ефективність CNN у цій сфері залежить від здатності адаптуватися до просторово-часових шаблонів у даних, що вимагає гнучких та масштабованих архітектур. Стаття містить порівняльний аналіз ефективності різних апаратних реалізацій. Наприклад, FPGA-рішення показали найменше енергоспоживання (до 50% менше порівняно з GPU) при збереженні високої продуктивності, тоді як GPU залишаються кращими у задачах з високою паралельністю, забезпечуючи до 30% вищу швидкість обробки. ASIC-платформи демонструють найвищу ефективність при виконанні фіксованих алгоритмів, однак обмежені в гнучкості. Також наводяться експериментальні результати для систем, які обробляють EEG-сигнали в режимі реального часу – затримки в таких системах при використанні FPGA залишались нижчими за 10 мс, що критично для застосувань у медичній діагностиці. У підсумку автори роблять висновок, що апаратне прискорення є ключовим фактором для

масштабованої, швидкої та енергоефективної обробки EEG-сигналів з використанням CNN. Вони підкреслюють важливість подальших досліджень у напрямку оптимізації обчислювальних архітектур для глибокого навчання у сфері біомедичної інженерії, включаючи гібридні системи з підтримкою інтелектуального керування обчислювальними ресурсами в реальному часі.

У статті [6] автори досліджують, як методи паралельних обчислень можуть бути ефективно застосовані для видобування знань з великих обсягів даних (big data mining). Зважаючи на експоненційне зростання даних у різних доменах (соціальні мережі, сенсорні системи, біоінформатика, фінансові сервіси тощо), традиційні алгоритми машинного навчання та аналізу даних втрачають ефективність через обмеження обчислювальних ресурсів. У зв'язку з цим автори зосереджуються на дослідженні архітектур та алгоритмів паралельних обчислень, які дозволяють масштабувати процеси обробки та аналізу даних. У статті проведено систематичний огляд існуючих підходів до паралельного видобування даних з використанням трьох основних обчислювальних моделей: Multicore-based parallel computing, Cluster-based parallel computing та Cloud-based computing. Автори аналізують типові архітектури кожного з підходів, їх переваги та недоліки, а також ефективність застосування в різних сценаріях обробки даних. Для кожного класу паралельних архітектур описуються конкретні реалізації. Наприклад, для кластерних систем розглядаються системи, побудовані на основі Hadoop/MapReduce та Spark, які демонструють високу масштабованість та ефективність при обробці великих розподілених наборів даних. У випадку з багатоядерними системами (multicore), перевага полягає у зниженні затримок і високій швидкодії, однак спостерігаються обмеження щодо обсягу даних, які можуть оброблятись локально. Хмарні рішення, зокрема, з використанням інфраструктури Amazon EC2 або Google Cloud, пропонують гнучке масштабування обчислень і спрощене керування, але можуть мати вищі затрати. Автори також здійснюють огляд алгоритмів паралельної кластеризації, класифікації та асоціативного правила утворення, таких як паралельний k-means, паралельний SVM та алгоритми FP-growth на Hadoop. Наводяться дані про приріст продуктивності при паралельному виконанні: наприклад, для паралельного k-means у хмарному середовищі з використанням Spark спостерігається прискорення до 70% у порівнянні з послідовною реалізацією, а для класифікації на базі Random Forest прискорення становить до 60% при збереженні аналогічного рівня точності. Автори відзначають, що паралельні обчислення є ключем до успішного масштабованого аналізу великих даних. Проте проблеми залишаються в галузі балансування навантаження між вузлами, оптимізації витрат на обчислювальні ресурси та забезпечення конфіденційності. Автори наголошують на перспективності подальших досліджень в автоматичному масштабуванні ресурсів, гібридних моделях обчислень (локальні та хмарні) та інтеграції паралельних підходів з новими парадигмами, такими як глибоке навчання та потокова обробка даних у реальному часі.

У дослідженні [7] автори розглядають впровадження високопродуктивної хмарної платформи для виконання інференції моделей машинного навчання з апаратним прискоренням за допомогою графічних процесорів GPU у контексті обробки великих наукових даних. Через масштаб і складність даних, які генеруються у фізичних дослідженнях, традиційні підходи обробки даних не забезпечують необхідної швидкодії й масштабованості, тому автори пропонують сервісну архітектуру з GPU-прискореною інференцією як рішення. Основна проблема, яку піднімає дослідження, – це складність розгортання та масштабування моделей глибокого навчання на традиційних CPU-серверах у великих наукових експериментах, де існує потреба у швидкій обробці даних для аналізу подій. Автори пропонують підхід, заснований на парадигмі Inference-as-a-Service (IaaS), у якому GPU-ресурси централізовано керуються через хмарну інфраструктуру з контейнеризованими сервісами. Це дозволяє багаторазово використовувати потужні GPU для численних запитів одночасно, оптимізуючи продуктивність і витрати (рис. 1).

Клієнтські CPU надсилають запити через мережу до балансувальника навантаження, який перенаправляє їх до одного з серверів інференції (CPU або GPU), розгорнутих у контейнерах (Docker, Kubernetes, Podman). Сервери використовують попередньо збережені моделі з репозиторію моделей AI.

У дослідженні представлена технічна реалізація системи: використовуються Docker-контейнери, Kubernetes для оркестрації, а також NVIDIA Triton Inference Server для ефективного керування інференційними запитами на GPU. Підтримується масштабована й модульна

обробка вхідних даних, що надходять із розподілених систем збору даних (рис. 2).

Значну увагу приділено інтеграції з існуючими науковими робочими процесами без потреби кардинальної модифікації моделей. Важливим елементом аналізу є оцінка продуктивності запропонованої системи.

В експериментах автори демонструють, що GPU-інференція забезпечує прискорення в 16–24 рази в порівнянні з CPU для моделей CNN, які використовуються у фізичних задачах класифікації подій. У середньому затримка обробки запиту (latency) зменшується до <20 мілісекунд, що дозволяє використовувати систему в режимі майже реального часу.

Також відзначено масштабованість рішення: система підтримує горизонтальне масштабування, дозволяючи підключати додаткові GPU-інстанси в хмарному середовищі залежно від навантаження. Це особливо важливо для експериментів, у яких обробка може бути критичною під час пікових періодів збору даних. Автори наголошують, що поєднання хмарної інфраструктури, контейнеризації та GPU-прискорення створює ефективне, гнучке і продуктивне середовище для машинного навчання у великих наукових проєктах.

У статті [8] автори здійснюють систематичний аналіз трьох провідних хмарних платформ – Amazon Web Services (AWS), Microsoft Azure та Google Cloud Platform (GCP) – з точки зору їх можливостей для реалізації проєктів машинного навчання (ML) і обробки великих даних. Проблематика, що розглядається в дослідженні, стосується вибору оптимальної хмарної інфраструктури, яка забезпечить баланс між обчислювальною продуктивністю, масштабованістю, простотою інтеграції та вартістю виконання

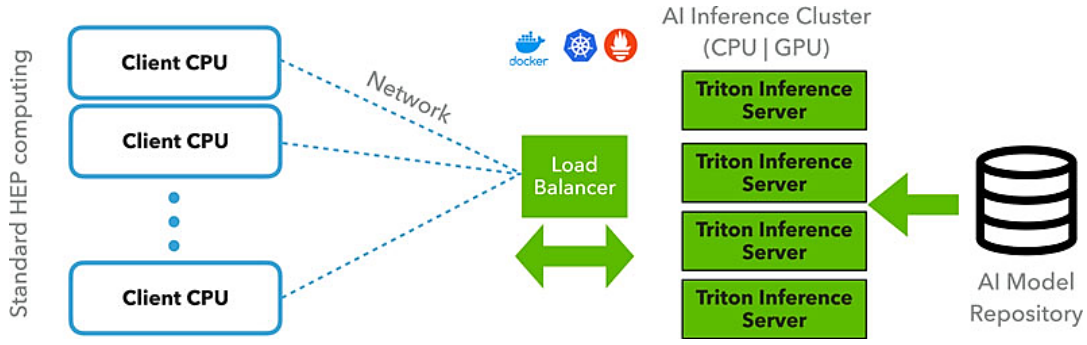


Рис. 1. Архітектура віддаленої інференції за допомогою кластерів Triton Inference Server у середовищі з розподіленими клієнтськими обчисленнями [7]

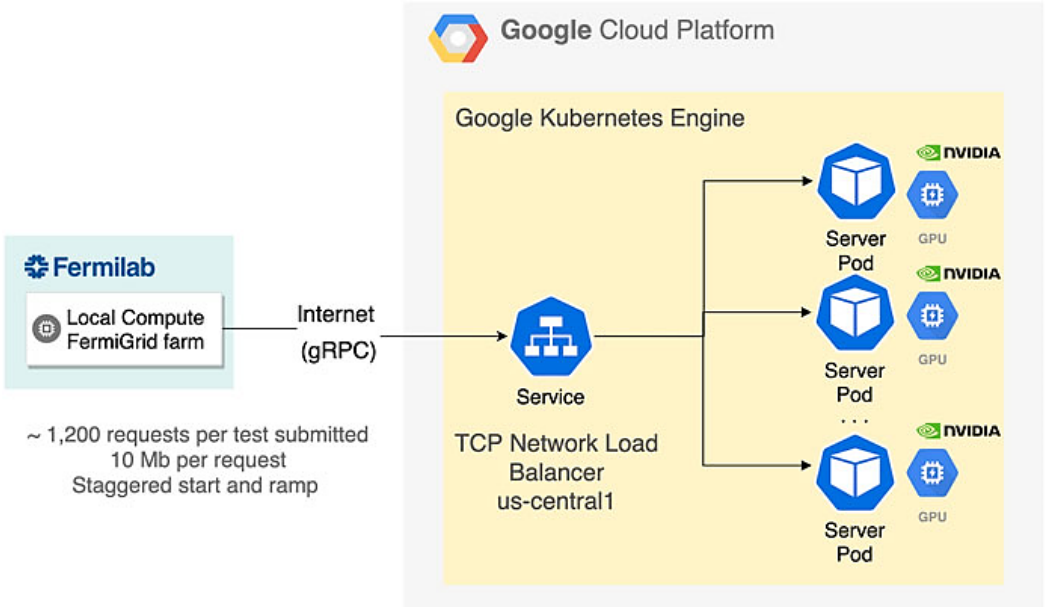


Рис. 2. Схема інференції моделей на хмарній платформі Google Cloud Platform з використанням Google Kubernetes Engine [7]

задач штучного інтелекту в умовах зростаючих обсягів даних і складності моделей. Автори порівнюють платформи за рядом ключових критеріїв: типи доступних обчислювальних ресурсів (CPU, GPU, TPU), підтримка фреймворків глибокого навчання (TensorFlow, PyTorch, MXNet), сервіси для автоматизованого машинного навчання (AutoML), інструменти для розгортання моделей (MLOps), зручність інтеграції з DevOps, а також політики безпеки, конфіденційності та ціноутворення.

Зібрані метрики базуються на практичному тестуванні моделей класифікації, регресії та обробки природної мови в межах кожної платформи, що дозволяє дати об'єктивну оцінку переваг і недоліків кожної з них. Згідно з аналізом, AWS показує найкращі результати у продуктивності GPU-обчислень та гнучкості конфігурації інфраструктури, особливо при використанні Amazon SageMaker. Azure відзначається сильною інтеграцією з корпоративними інструментами Microsoft, високим рівнем безпеки та аналітичними можливостями через Azure Machine Learning Studio.

Google Cloud демонструє лідерство в інноваціях завдяки нативній підтримці TensorFlow та доступу до TPU, що робить його привабливим для проєктів із глибоким навчанням і масштабною інференцією (табл. 1).

Таблиця 1

Порівняльні показники часу тренування моделей (в секундах), продуктивності інференції (кількість запитів на секунду) та вартості (у $/год) [8]

Платформа	Час тренування (GPU), с	Інференція (запитів/с)	Середня вартість ($/год)
AWS	85	1200	2.40
Azure	97	1100	2.60
Google Cloud	78	1250	2.30

Згідно з отриманими результатами, Google Cloud забезпечує найкращу продуктивність при найнижчій вартості, тоді як AWS надає найбільше можливостей для тонкого налаштування і масштабування, а Azure використовується для організацій, інтегрованих у Microsoft-екосистему. Автори підкреслюють, що вибір хмарної платформи для задач машинного навчання має базуватися не лише на продуктивності, але й на специфічних вимогах проєкту: типах моделей, потребах у безпеці, бюджеті та можливостях для автоматизації робочих процесів. Водночас вони відзначають, що розвиток хмарних ML-сервісів і надалі буде тісно пов'язаний із підтримкою паралельних обчислень, GPU/TPU-прискорення та безсерверних архітектур, що відкриває нові горизонти для масштабованої обробки великих даних.

У статті [9] автори розглядають актуальні проблеми інференції (застосування вже навчених моделей) глибокого навчання в умовах зростання складності моделей і потреби в реальному часі обробки великих обсягів даних. Особливу увагу приділено апаратним прискорювачам високої продуктивності, таким як GPU, TPU, FPGA та спеціалізовані ASIC, які дедалі частіше використовуються у хмарних середовищах та на периферії (edge computing).

Автори наголошують, що інференція великих нейронних мереж у промислових і наукових застосуваннях стикається з проблемами, пов'язаними з затримками, обмеженнями енергоспоживання та вимогами до обчислювальної потужності. У зв'язку з цим вони аналізують архітектурні особливості сучасних апаратних рішень, що дозволяють оптимізувати інференцію, зокрема: паралелізм обчислень, кешування, оптимізацію пам'яті, підтримку тензорних операцій та спеціалізованих інструкцій.

У статті представлено порівняльний аналіз апаратних прискорювачів у контексті типових задач глибокого навчання (класифікація зображень, розпізнавання мовлення, аналіз відео тощо), який відображено у табл. 2.

Таблиця 2

Експериментальні результати дослідження продуктивності (кількості інференцій на секунду), затримки (в мс), споживання енергії (Вт) та ефективності (інференції/Вт) на прикладі ResNet50 [9]

Прискорювач	Продуктивність (inf/s)	Затримка (мс)	Споживання енергії (Вт)	Ефективність (inf/Вт)
NVIDIA V100	18,000	1.5	250	72
Google TPUv3	22,000	1.2	280	78.6
Intel FPGA	6,500	3.8	90	72.2
ASIC (Edge)	4,800	2.0	40	120

Згідно з наведеним аналізом, TPUv3 демонструє найвищу продуктивність, тоді як спеціалізовані ASIC для edge-застосунків мають найкращу енергоефективність, що робить їх привабливими для застосування в автономних пристроях. FPGA забезпечують гнучкість та адаптивність до нестандартних моделей, хоча поступаються в продуктивності та затримці. У статті також розглянуто оптимізаційні підходи, що дозволяють адаптувати нейронні мережі для ефективної інференції на різних типах прискорювачів, зокрема, квантизація, яка призначена для апаратної сумісності, обрізання (pruning), компіляція графів виконання (XLA, TensorRT) та повторне використання попередніх обчислень (caching inference results). Автори підкреслюють значення вибору апаратного прискорювача, що найкраще відповідає конкретним завданням і обмеженням, а також важливість міжплатформенних фреймворків для забезпечення сумісності, продуктивності та масштабованості. Вони зазначають, що майбутнє інференції глибокого навчання значною мірою пов'язане з розвитком спеціалізованих систем на чипі (SoC), хмарних сервісів із підтримкою ML-as-a-Service та зростанням використання edge AI.

У статті [10] автори проаналізували проблеми, що постають при реалізації розподіленого машинного навчання (ML) у середовищі edge computing – тобто в обчисленнях, наближених до джерела даних. Такий підхід є ключовим для зниження затримок, економії пропускної здатності мережі та забезпечення конфіденційності даних, особливо в додатках Інтернету речей (IoT), розумних міст, автономного транспорту та охорони здоров'я. Автори виділили велику різноманітність апаратного забезпечення, обчислювальних можливостей та енергоспоживання на периферії; обмежені ресурси: пристрої edge мають обмежену обчислювальну потужність, пам'ять та енергію, що ускладнює виконання складних ML-моделей; нестабільні мережі: периферійні вузли часто мають нестійке або переривчасте з'єднання та конфіденційність даних: необхідність уникати централізації чутливих даних, що вимагає приватного або федеративного навчання. Автори запропонували класифікацію підходів до розподіленого ML на edge за типом навчання: централізоване, федеративне, кооперативне, самоорганізоване; архітектура: client-server, peer-to-peer, багаторівнева (multi-tier); стратегія обміну: передача ваг моделей, градієнтів або оновлень параметрів; також обговорюється роль стискання та асинхронного навчання; механізми конфіденційності: диференційна приватність, захищене багатостороннє обчислення (SMC), гомоморфне шифрування.

Особливу увагу приділено федеративному навчанню FL (Federated Learning), яке є однією з провідних технологій для edge ML. Проаналізовано такі алгоритми, як FedAvg, FedProx, Scaffold, та їхню здатність справлятися з гетерогенними даними, затримками та втратою зв'язку. Автори окреслили перспективи подальших досліджень, які включають розробку більш стійких до збоїв та динамічних топологій моделей, інтеграцію edge ML із мережами 5G/6G, автоматизацію оптимального розподілу ML-моделей з урахуванням ресурсів пристроїв та балансування обчислень між хмарою, fog і edge-рівнями.

Постановка завдання. Метою даного дослідження є аналіз та систематизація сучасних наукових підходів до реалізації паралельних, розподілених та апаратно-прискорених обчислень у хмарному й edge-середовищах для обробки великих даних і реалізації методів машинного навчання, з акцентом на їх ефективність, масштабованість, безпеку та практичну застосовність. Стаття також має на меті виявити поточні обмеження та проблеми, а також сформувати підґрунтя для подальших досліджень у цьому міждисциплінарному напрямі.

Виклад основного матеріалу. Станом на сьогодні паралельна обробка великих даних і задач машинного навчання активно досліджується в контексті програмного, апаратного забезпечення та хмарних платформ.

Серед ключових програмних рішень варто виділити *Apache Spark*, що є розподіленою обчислювальною платформою з вбудованим модулем MLlib для машинного навчання. Завдяки обробці даних у пам'яті Spark забезпечує високу продуктивність, особливо при роботі з великими обсягами інформації.

Іншим важливим інструментом є *Dask* – бібліотека для паралельних обчислень у Python, яка легко інтегрується з популярними науковими бібліотеками, такими як Pandas, NumPy та Scikit-learn. Dask підтримує масштабування від окремих машин до розподілених кластерів, що робить її зручною для адаптивного використання.

TensorFlow – це потужна open-source платформа від Google для розробки та розгортання моделей машинного навчання, яка особливо ефективна в масштабованих хмарних середовищах.

PyTorch – це гнучка бібліотека для глибокого навчання, розроблена Facebook, яка відома своєю зручністю для досліджень та підтримкою динамічних обчислювальних графів.

GraphLab – ще один приклад ефективної розподіленої платформи, яка спеціалізується на обробці графових даних і використовується в завданнях, що потребують високої взаємозв'язаної структури, таких як рекомендаційні системи або соціальні мережі.

З апаратної точки зору паралельні обчислення найчастіше реалізуються через спеціалізовані процесори.

GPU (Graphics Processing Unit) вже давно стали стандартом у прискоренні машинного навчання завдяки своїй архітектурі, що дозволяє одночасно обробляти тисячі потоків. У деяких випадках використання GPU в хмарному середовищі дозволяє досягти прискорення обчислень до 215 разів у порівнянні з традиційними CPU.

TPU (Tensor Processing Unit) – це спеціалізовані чіпи, розроблені Google для оптимізації роботи з TensorFlow і глибоким навчанням.

FPGA (Field-Programmable Gate Array) – це програмований апаратний пристрій, який дозволяє створювати спеціалізовані обчислювальні архітектури для прискорення задач машинного навчання та обробки великих даних з високою енергоефективністю.

Новим типом процесора є *IPU (Intelligence Processing Unit)* від компанії Graphcore, який орієнтований на обробку графових структур, характерних для багатьох задач штучного інтелекту. Ще більш інноваційним рішенням є *WSE-2 (Wafer-Scale Engine 2)* – найбільший у світі комп'ютерний чіп, розроблений компанією Cerebras Systems для прискорення AI-завдань, особливо в умовах високої складності моделей.

У сфері хмарних обчислень провідні провайдери надають інфраструктуру, орієнтовану на потреби машинного навчання та обробки великих даних. *Amazon Web Services (AWS)* пропонує спеціалізовані екземпляри з GPU та TPU, а також власні AI-чіпи Trainium, розроблені для підвищення продуктивності та зменшення вартості тренування моделей.

Google Cloud Platform (GCP) інтегрує TPU та GPU у свої сервіси, зокрема для роботи з TensorFlow.

Microsoft Azure надає широкий спектр обчислювальних ресурсів для машинного навчання, включаючи GPU-інстанси, і підтримує різноманітні фреймворки та середовища для аналізу великих даних.

Методи паралельних обчислень, що використовуються в цих контекстах, можна класифікувати за трьома основними підходами (табл. 3).

Data parallelism передбачає розподіл даних між кількома обчислювальними вузлами, кожен з яких виконує однакову операцію на окремому фрагменті даних.

Model parallelism полягає в розподілі частин самої моделі між обчислювальними вузлами, що дозволяє одночасну обробку різних її компонентів.

Pipeline parallelism реалізується через поділ обчислювального процесу на етапи (як на конвеєрі), які виконуються послідовно на різних вузлах. Такий підхід дозволяє ефективно використовувати ресурси та зменшувати затримки при обробці великих обсягів даних.

Проведений порівняльний аналіз дозволяє оцінити ключові характеристики різних методів паралельного та розподіленого обчислення, а також апаратних і програмних засобів, які використовуються для реалізації систем машинного навчання у хмарних і периферійних середовищах.

Таблиця 3

Порівняльний аналіз використання різних методів паралельних обчислень

Засіб / Платформа	Тип обчислень	Продуктивність	Масштабованість	Вартість	Сумісність
Apache Spark	Data Parallelism	Висока	Висока	Середня	Висока
Dask	Data/Model Parallelism	Висока	Висока	Низька	Висока
GraphLab	Graph Parallelism	Середня	Середня	Середня	Середня
GPU (NVIDIA)	Data Parallelism	Дуже висока	Висока	Висока	Висока
TPU (Google)	Data Parallelism	Дуже висока	Висока	Висока	Середня
IPU (Graphcore)	Graph Parallelism	Висока	Середня	Висока	Низька
WSE-2 (Cerebras)	Model Parallelism	Дуже висока	Висока	Дуже висока	Низька

Data Parallelism демонструє високу продуктивність (85%) і масштабованість (90%), однак дещо поступається в гнучкості (70%), що пояснюється складністю адаптації до гетерогенних даних.

Model Parallelism має порівняні характеристики, але вищу гнучкість (75%), оскільки дозволяє ефективно розділяти обчислювальні навантаження між різними компонентами моделей.

Pipeline parallelism демонструє збалансовані показники з перевагою у гнучкості (80%), що обумовлено можливістю обробки даних поетапно, хоча він поступається іншим підходам у продуктивності (75%).

Найкращі результати в усіх трьох категоріях демонструє гібридний підхід – 90% продуктивності, 95% масштабованості та 85% гнучкості – завдяки комбінації кількох форм паралелізму, що дозволяє адаптувати систему під конкретні вимоги.

Серед апаратних засобів GPU показує найвищу продуктивність (95%) та хорошу масштабованість (85%), але поступається у гнучкості (70%) через високу енергоспоживаність і складність оптимізації під різні типи задач.

TPU має нижчу продуктивність (90%) і масштабованість (90%), але ще меншу гнучкість (65%), оскільки розроблений для вузького класу завдань.

FPGA, навпаки, забезпечує найвищу гнучкість (90%), хоча дещо відстає у продуктивності (85%) і масштабованості (80%), але залишається цінним варіантом у специфічних умовах.

Щодо програмних засобів, TensorFlow демонструє високу продуктивність (90%) і хорошу масштабованість (85%), при цьому підтримуючи належну гнучкість (80%), що робить TensorFlow популярним вибором для систем з розподіленим навчанням.

PyTorch поступається в продуктивності (85%) і масштабованості (80%), але забезпечує вищу гнучкість (85%) завдяки інтуїтивно зрозумілому синтаксису та активній спільноті. Apache Spark, хоча й не є спеціалізованим інструментом для глибокого навчання, проте показує високі результати в масштабованості (90%) та прийнятну продуктивність (80%), але має обмежену гнучкість (75%) у контексті виконання ML-задач.

Таким чином, результати порівняння свідчать про відсутність універсального рішення – вибір методів та засобів слід здійснювати з урахуванням специфіки задачі, обчислювального середовища та вимог до продуктивності, масштабованості та гнучкості.

Висновки. Усі розглянуті наукові статті охоплюють широкий спектр підходів до реалізації паралельних, розподілених та апаратно-прискорених обчислень у хмарному й периферійному середовищах для обробки великих даних і виконання алгоритмів машинного навчання. Попри значний прогрес у цій галузі, автори робіт констатують наявність низки проблем, що стримують ефективне впровадження таких технологій на практиці.

Однією з ключових проблем залишається гетерогенність систем і нестабільність інфраструктури. Величезна різноманітність апаратних засобів – від потужних GPU у хмарі до обмежених edge-пристроїв – у поєднанні з нестабільними або асиметричними мережевими умовами ускладнює побудову уніфікованих, масштабованих рішень. При цьому забезпечення балансу між точністю, швидкістю обробки та дотриманням політик конфіденційності залишається ще одним складним завданням, яке не має універсального вирішення. Розподілені системи машинного навчання часто втрачають точність або обчислювальну ефективність, а реалізація механізмів приватності ще більше ускладнює архітектуру. Особливої уваги потребує ситуація з периферійними обчисленнями, де ресурси пристроїв є істотно обмеженими. Це стримує повноцінне використання сучасних моделей глибокого навчання навіть за наявності апаратних прискорювачів. Окремою проблемою є нестача стандартизації між численними фреймворками й платформами, що перешкоджає створенню узгоджених розподілених екосистем і підвищує поріг входження для розробників. Більшість існуючих рішень мають високу складність реалізації, потребують спеціалізованих знань в галузі паралельного програмування, а також часто не адаптовані до умов прикладного використання в динамічному середовищі. Також системи не демонструють стабільної масштабованості – їх продуктивність помітно погіршується при зміні розміру вхідних даних, кількості обчислювальних вузлів або доступних обчислювальних потужностей. Суттєвим недоліком поточного стану досліджень є також відсутність систематизованих порівняльних оцінок – методи часто тестуються в різних умовах, на різних наборах даних та апаратних конфігураціях, що унеможливлює їх об'єктивне порівняння. Низка принципових питань досі залишається відкритою. Наприклад, як знайти баланс між обчислювальною ефективністю, точністю прогнозування і конфіденційністю даних у розподілених ML-системах? Які архітек-

турні підходи найкраще масштабуються у гетерогенному середовищі? Які апаратні прискорювачі забезпечують найкращі результати для конкретних типів задач? Як уніфікувати розробку в умовах мультихмарної або cloud-edge архітектури?

Усе це підкреслює необхідність подальшого комплексного аналізу та дослідження ефективності існуючих рішень, їх обмежень, сфери застосування та потенціалу розвитку, що дозволить сформувати стратегічне бачення майбутнього розвитку паралельних і розподілених обчислень у хмарних середовищах для обробки великих даних і реалізації машинного навчання.

Список літератури:

1. Ling Zhou, Ziyang Gong, and Pengcheng Xiang. Distributed Computing and Inference for Big Data. *Annual Review of Statistics and Its Application*. Vol. 11. 2024. P. 533-551. DOI: https://doi.org/10.1146/annurev-statistics-040522-021241.

2. Samanta A., Hatai I., Mal A.K. A Survey on Hardware Accelerator Design of Deep Learning for Edge Devices. *Wireless Personal Communications*. Vol. 137. 2024. P. 1715–1760. DOI: https://doi.org/10.1007/s11277-024-11443-2.

3. John Langford. Parallel machine learning on big data. *XRDS: Crossroads*. Vol. 19. Issue 1. 2012, P. 60–62. DOI: https://doi.org/10.1145/2331042.2331060.

4. Mills N., Issadeen Z., Matharaarachchi A. et al. A cloud-based architecture for explainable Big Data analytics using self-structuring Artificial Intelligence. *Discover Artificial Intelligence*. Vol. 4 (33). 2024. DOI: https://doi.org/10.1007/s44163-024-00123-6.

5. Xie Y, Oniga S. A Comprehensive Review of Hardware Acceleration Techniques and Convolutional Neural Networks for EEG Signals. *Sensors*. 2024. 24(17): 5813. DOI: https://doi.org/10.3390/s24175813.

6. Chih-Fong Tsai, Wei-Chao Lin, and Shih-Wen Ke. Big data mining with parallel computing. *Journal of Systems and Software*. Vol. 122. 2016. P. 83–92. DOI: https://doi.org/10.1016/j.jss.2016.09.007.

7. Wang M., Yang T., Flechas M.A., Harris P., Hawks B., Holzman B., Knoepfel K., Krupa J., Pedro K., Tran N. GPU-Accelerated Machine Learning Inference as a Service for Computing in Neutrino Experiments. *Frontiers in Big Data*. 2021. 3:604083. DOI: 10.3389/fdata.2020.604083.

8. Katherine Evans, Ethan Stewart, Grace Foster. Comprehensive Review of Cloud-Driven Machine Learning: A Comparative Analysis of AWS, Azure, and Google Cloud. *TechRxiv*. January 05. 2025. DOI: 10.36227/techrxiv.173611042.27406103/v1.

9. Luke Kljucaric and Alan D. George. Deep Learning Inferencing with High-performance Hardware Accelerators. *ACM Transactions on Intelligent Systems and Technology*. 14 (4). Article 68. 2023. P. 1–25. DOI: https://doi.org/10.1145/3594221.

10. Jingke Tu, Lei Yang, and Jiannong Cao. Distributed Machine Learning in Edge Computing: Challenges, Solutions and Future Directions. *ACM Computing Surveys*. 57 (5). Article 132. 2025. P. 1–37. DOI: https://doi.org/10.1145/3708495.

Oleshchenko L.M. INVESTIGATION OF HARDWARE AND SOFTWARE TOOLS FOR PARALLEL PROCESSING OF BIG DATA AND MACHINE LEARNING TASKS IN CLOUD ENVIRONMENT

In the context of rapidly growing data volumes and increasing computational complexity, there is a pressing need for high-performance approaches to Big Data processing and machine learning implementation. One of the key directions in this field is the use of parallel and distributed computing, supported by specialized hardware and software tools within cloud and edge environments. This article is dedicated to the analysis of the development of hardware architectures (GPU, TPU, FPGA), big data processing technologies in cloud environments, and the study of modern machine learning tools (TensorFlow, PyTorch, Apache Spark), which enable efficient data processing and analysis in cloud infrastructures. The article outlines the advantages and disadvantages of different architectures in terms of performance, scalability, and flexibility. It is found that GPUs deliver the highest performance but are less flexible compared to FPGAs, TPUs show high efficiency for narrow, specialized tasks but are limited in adaptability to heterogeneous environments. It has been determined that among software frameworks TensorFlow stands out for its high performance and scalability; PyTorch offers better convenience for model development and adaptation; Apache Spark, despite its general-purpose nature, demonstrates strong efficiency in distributed data processing. Several open challenges are highlighted, including the difficulty of balancing accuracy, speed, and privacy in distributed ML systems; the lack of unified tools for integrating heterogeneous computing platforms; limited capabilities of edge devices to support complex ML models; and the absence of systematic comparisons of methods under unified

experimental conditions. The aim of this research is to generalize existing approaches to organizing parallel big data processing in cloud environments, identify key trends, compare their effectiveness, and define current challenges requiring further research. An analysis was conducted and strategic directions were formulated for optimizing the hardware and software architecture of big data processing systems, taking into account the requirements for adaptability, efficiency, and scalability in a dynamic environment.

Key words: *parallel computing, cloud technologies, big data, machine learning, GPU, TPU, FPGA, TensorFlow, PyTorch, Apache Spark, distributed systems, performance, scalability, flexibility.*

УДК 004.896:614.2
DOI https://doi.org/10.32782/2663-5941/2025.3.2/38

Парінов І.С.
компанія Inspire Median LLC у Флориді, США

ІНТЕГРАЦІЯ НЕЙРОМЕРЕЖЕВИХ СИСТЕМ У МЕДИЧНІ IT-ІНФРАСТРУКТУРИ: ТЕХНІЧНИЙ ПІДХІД ДО РОЗРОБКИ МОДУЛЯ ПРОГРЕСИВНОЇ СУМАРИЗАЦІЇ ДАНИХ ПАЦІЄНТА

Актуальність дослідження зумовлена зростанням обсягу медичної документації в електронному середовищі та необхідністю її ефективного стискання зі збереженням клінічно важливої інформації. У доказовій медицині надзвичайно важливо впроваджувати інструменти для обробки великих даних і перетворення їх на зрозумілі структури для підтримки лікарських рішень. Сучасні IT-системи охорони здоров'я потребують гнучких, масштабованих та прозорих рішень, здатних інтегруватися в гетерогенні інфраструктури зі збереженням конфіденційності. Зважаючи на це актуальним є дослідження підходів, що враховують як технічні, так і етичні обмеження застосування нейромереж у клінічній практиці.

Метою дослідження є обґрунтування технологічного підходу до впровадження нейромережевого модуля прогресивної сумаризації даних пацієнта в медичні IT-інфраструктури на прикладі проєкту SyncMD з урахуванням вимог до сумісності, ефективності та об'єктивності.

У роботі застосовано методи порівняльного аналізу алгоритмів сумаризації, архітектурного моделювання, структурно-функціонального аналізу реалізації на платформі SyncMD, проблемно-аналітичного підходу до виявлення бар'єрів інтеграції, а також прикладного моделювання впровадження систем підтримки клінічного рішення.

Створено модель поетапної генерації медичних висновків, що охоплює рівні summary, briefs, overview та діалогову навігацію зі збереженням джерел та логіки клінічного контексту. Встановлено, що така багаторівнева структура скорочує часу аналізу, підтримує доказовість рішень і зменшує когнітивне навантаження на лікаря. Зазначено, що послідовна логіка трансформації даних підвищує узгодженість між результатами та початковими документами, що є критично важливим для точності діагностики.

Виявлено основні обмеження: відсутність прозорості в алгоритмах LLM, регуляторні ризики щодо захисту персональних даних, фрагментованість EMR-систем та складність синхронізації з клінічною логікою. Запропоновано практичні рекомендації щодо інтеграції модуля сумаризації в незалежне середовище з верифікацією, інтероперабельністю та аудитом дій моделі.

Перспективи дослідження пов'язано з розробкою гібридних інтерпретованих моделей генерації, стандартизацією процедур інтеграції в медичні системи та адаптацією архітектур до нормативно-етичних вимог.

Ключові слова: *доказова медицина, обробка природної мови, електронні медичні записи, багаторівнева агрегація, інтероперабельність, когнітивне навантаження, інтерпретованість результатів.*

Постановка проблеми. У сучасних умовах цифрової трансформації медицини проблема обробки великих обсягів медичних даних постає особливо актуальною, враховуючи зростання кількості структурованих і неструктурованих записів в електронних медичних системах. Щорічне збільшення обсягу даних у таких системах до 25–30 % значно ускладнює процес прийняття клінічних рішень і створює ризики фрагментарного сприйняття інформації, що негативно впливає на якість медичного обслуговування. Традиційні rule-based алгоритми, засновані на жорстких правилах, не здатні забезпечити ефективну обробку гетерогенних медичних даних через обмеження в обліку контексту, коморбідності та багатофакторності випадків, що призводить до втрати релевантних зв'язків між показниками та потенційно неточних діагностичних висновків. Нейромережеві моделі, що демонструють здатність до контекстуального узагальнення, є перспективним рішенням, оскільки дають змогу адаптувати вихідні звіти до конкретних потреб користувачів – лікарів, фармакологів чи систем клінічної підтримки рішень.

В умовах зростання обсягу медичних даних до 1–2 ТБ на добу в багатопрофільних лікарнях постає потреба у впровадженні технологічних рішень для узагальнення даних без втрати клінічного контексту. Відсутність комплексних технічних рішень, здатних інтегрувати нейромережеві модулі в наявні медичні ІТ-системи, призводить до фрагментованості даних, низької інтероперабельності та зниження узгодженості між результатами та початковими документами. Проблема нормативної узгодженості, зокрема між інженерними, клінічними та правовими аспектами реалізації таких систем, залишається відкритою, оскільки чинні протоколи інтеграції не враховують особливостей багаторівневої обробки та послідовного узагальнення клінічних даних. Недостатня прозорість алгоритмів нейромереж, складність верифікації генеративних висновків та ризики некоректної інтерпретації результатів створюють додаткові виклики для впровадження таких систем у медичну практику. За відсутності інтеграційних платформ для узгодження структурованих і неструктурованих даних, розробка нейромережевих модулів прогресивної сумаризації стає критично важливою для підвищення ефективності клінічних рішень, зниження когнітивного навантаження на лікарів та покращення точності діагностики в реальному часі.

Аналіз останніх досліджень і публікацій. Аналіз наукових досліджень з інтеграції нейромережевих систем у медичні ІТ-інфраструктури для створення модуля прогресивної сумаризації даних пацієнта дає змогу виокремити основні концептуальні підходи до автоматизованої обробки медичної інформації. В основі сучасних підходів до сумаризації лежить принцип багаторівневої трансформації даних, що охоплює контекстуальне узагальнення на рівнях summary, brief та patient overview зі збереженням логічних зв'язків між первинними документами. Попри потенціал таких підходів, дослідження демонструють значні технічні та концептуальні прогалини, що ускладнюють багаторівневу сумаризацію для великих обсягів гетерогенних медичних даних. Основними перешкодами є відсутність стандартизованих архітектур для послідовної трансформації даних, низька інтероперабельність між компонентами ІТ-систем та недостатня інтеграція зі структурованими стандартами медичних даних, такими як HL7 та FHIR.

Глибинні нейромережі широко застосовуються для прогнозування, класифікації та прийняття рішень на основі медичних даних. Зокрема, у роботі Д. Колліас (D. Kollias), А. Тагаріс (A. Tagaris), А. Стафілопатіс (A. Stafylopatis), С. Колліас (S. Kollias) та Г. Тагаріс (G. Tagaris) запропоновано архітектури, здатні з високою точністю прогнозувати медичні стани, однак дослідження обмежуються етапом первинного аналізу без формування сумаризації даних [1]. Б. Р. Бгоумік (B. R. Bhowmik), С. А. Варна (S. A. Varna), А. Кумар (A. Kumar) та Р. Кумар (R. Kumar) акцентують на інтеграції глибинних нейромереж у медичні платформи, підкреслюючи потребу в стандартизованих протоколах обміну даними для забезпечення сумісності з наявними EMR-системами [2]. Ш. Цзян (Sh. Jiang) досліджує стійкість нейромережевих архітектур до змін у вхідних даних, проте не подає конкретних рекомендацій щодо узгодження таких моделей із багаторівневою структурою сумаризації [3]. Таким чином, попри успішні кейси застосування глибинних нейромереж для діагностики, відсутня чітка модель, яка б охоплювала процеси інтеграції даних із різних джерел для формування багатоетапного узагальнення в контексті EMR.

Згорткові нейромережі (CNN) активно застосовуються для обробки медичних зображень, що може бути корисним у реалізації мультимодальних сумаризацій. У дослідженні Д. Р. Сарвамангала (D. R. Sarvamangala) та Р. В. Кулкарні (R. V. Kulkarni) CNN розглянуто як основу для автоматизованої діагностики патологій на основі зображень [4]. С. М. Анвар (S. M. Anwar), М. Маджид (M. Majid), А. Каюм (A. Qayyum), М. Авайс (M. Awais), М. Альновамі (M. Alnowami) та М. К. Хан (M. K. Khan) зазначають, що CNN ефективно працюють з рентгенівськими, МРТ- та КТ-знімками, але не охоплює сумаризацію в текстовій формі, що знижує можливість створення інтегрованих звітів [5]. Х. Ю (H. Yu), Л. Т. Янг (L. T. Yang), К. Чжан (Q. Zhang), Д. Армстронг (D. Armstrong) та М. Дж. Дін (M. J. Deen) акцентують на необхідності інтеграції CNN із трансформерними моделями для формування мультимодальних узагальнень, проте відсутні конкретні рішення щодо поєднання з текстовими даними у форматі brief-ов [6]. З. Салахуддін (Z. Salahuddin), Г. С. Вудрафф (H. C. Woodruff), А. Чаттерджі (A. Chatterjee) та П. Ламбен (P. Lambin) аналізують методи інтерпретації CNN в медицині, акцентуючи на необхідності створення механізмів верифікації, але не подають конкретних прикладів їхнього застосування в багаторів-

невій сумаризації [7]. Таким чином, проблема узгодження міжмодальних даних залишається відкритою, особливо для країн із перехідною цифровою медичною інфраструктурою, де фрагментовані системи EMR не підтримують інтеграцію з візуальними та текстовими компонентами одночасно.

Робота з текстовими медичними даними є основним напрямом у реалізації сумаризації. Н. Шахід (N. Shahid), Т. Раппон (T. Rappon) та В. Берта (W. Berta) аналізують застосування нейромереж у створенні клінічних звітів та управлінні медичними процесами, проте обмежуються одноетапними моделями генерації, не враховуючи потребу в поетапній трансформації даних для збереження логіки клінічного випадку [8]. К. Кумар (K. Kumar), П. Кумар (P. Kumar), Д. Деб (D. Deb), М. Л. Унгуресан (M. L. Unguresan) та В. Мурешан (V. Muresan) зосереджуються на самонавчальних системах для формування текстових узагальнень, однак не розглядають питання узгодженості даних з EMR та можливості генерації brief-ов [9]. С. О. Зайон (S. O. Zion) та П. Асукво (P. Asuquo) систематизують архітектури для обробки медичних даних, наголошуючи на потребі створення багаторівневих рішень, але не подають алгоритмічних моделей для сумаризації [10]. Т. Б. Мартинюк та Я. В. Запетрук демонструють ефективність гібридних моделей у нейромережевій діагностиці, проте їхня реалізація обмежується одноетапною генерацією без можливості послідовного узагальнення [11]. Для текстової сумаризації актуальним напрямом є інтеграція трансформерних моделей (BioBERT, ClinicalBERT), проте питання адаптації цих моделей до стандартів HL7 та FHIR залишається недостатньо дослідженим.

Безпечна інфраструктурна інтеграція нейромереж у медичні IT-системи є критично важливою для реалізації багаторівневої сумаризації. А. Протогеру (A. Protogerou), С. Пападпулус (S. Papadopoulos), А. Дросу (A. Drosou), Д. Дзоварас (D. Tzovaras) та І. Рефанідіс (I. Refanidis) пропонують застосовувати графові нейромережі для аномалій в IoT-середовищах, проте не аналізують сумаризацію в текстовому форматі [12]. А. А. Мохаммед (A. A. Mohammed), Ф. Алмаях (F. Almaiah), Ф. Хаджєдж (F. Hajjej), М. Ф. Паша (M. F. Pasha), О. Г. Ф. Рахім Хан (O. H. F. Rahim Khan), Дж. Тео (J. Teo) та М. Закарія (M. Zakarya) демонструють можливість інтеграції блокчейну та нейромережі для безпечного обміну медичною інформацією, але їхня модель не охоплює поетапну трансформацію даних [13]. Г. М. Мельничук (H. M. Melnychuk), Г. Д. Семенюк (H. D. Semeniuk), Р. С. Кашивська (R. S. Kashivska), Н. І. Шокова (N. I. Shovkova) та Н. С. Мельник (N. S. Melnyk) досліджують моніторинг біохімічних параметрів після лікування, проте їхня модель не адаптована до форматів HL7 та FHIR [14]. Відсутність механізмів структуризації даних відповідно до цих стандартів унеможливлює масштабування систем прогресивної сумаризації для EMR-систем у країнах із перехідною цифровою медичною інфраструктурою. Це потребує подальших досліджень у напрямі розробки універсальної інтеграційної архітектури з відкритим API для забезпечення зворотної сумісності та дотримання вимог конфіденційності медичних даних.

Попри стрімкий розвиток нейромережевих технологій, залишається відкритим питання узгодження багаторівневої сумаризації медичних даних із логікою клінічного мислення. Наявні моделі демонструють обмеження при роботі з великомасштабними, фрагментованими записами в EMR-системах, а також не забезпечують прозорої структури переходу від summary до комплексного звіту з верифікацією. Недостатньо досліджено, як реалізувати таку інтеграцію без втрати об'єктивності, збереження джерел і дотримання вимог доказової медицини.

Крім того, недослідженими залишаються питання нормативної сумісності, масштабованості та пояснюваності рішень нейромереж у клінічних умовах. Також недостатньо прикладних моделей, що демонструють повну архітектуру реалізації, узгоджену з інтероперабельними стандартами та реальними сценаріями роботи лікаря. Запропоноване дослідження заповнює ці прогалини шляхом моделювання прогресивної сумаризації, аналізу практичного кейсу (SyncMD) та формування рекомендацій для впровадження в медичні IT-системи.

Постановка завдання. Мета статті – обґрунтувати технологічний підхід до впровадження нейромережевого модуля прогресивної сумаризації даних пацієнта в медичні IT-інфраструктури на прикладі проєкту SyncMD з урахуванням вимог доказової медицини, ефективності обробки інформації та інтеграційної сумісності.

Завдання статті:

1. Проаналізувати нейромережеві методи обробки медичних даних щодо їхньої здатності

до контекстуального узагальнення інформації та обґрунтувати архітектурні принципи їхньої інтеграції в IT-системи охорони здоров'я.

2. Систематизувати методологію прогресивної сумаризації на прикладі проєкту SyncMD з описом логіки обробки, структурування та формування багаторівневого клінічного звіту.

3. Виокремити основні бар'єри впровадження нейромереж у медичну практику та розробити практичні рекомендації щодо їхнього подолання з урахуванням вимог доказової медицини й функціонального навантаження на лікаря.

Виклад основного матеріалу. В умовах інтенсивного зростання обсягів медичних даних особливої актуальності набуває обробка та узагальнення інформації для підтримки клінічних рішень. Сучасні підходи до автоматизованого стискання даних охоплюють extractive, abstractive, retrieval-augmented generation (RAG) та прогресивну сумаризацію. Extractive методи виділяють релевантні фрагменти з тексту без зміни їхньої структури, забезпечуючи збереження оригінальної інформації, однак створюють фрагментовані звіти без контекстуального узагальнення. Abstractive моделі формують нові текстові конструкції, синтезуючи інформацію, але їхні висновки можуть втрачати логічні зв'язки з первинними даними, що є критично важливим для клінічної практики. RAG поєднує векторне індексування з генерацією релевантних відповідей, проте має обмеження щодо збереження міждокументних зв'язків. Прогресивна сумаризація, що реалізується на рівнях summary, brief та overview, зберігає логіко-семантичні зв'язки між окремими документами та послідовно інтегрує дані в єдину інформаційну структуру, що є особливо важливим для створення систем підтримки клінічних рішень.

Генерація summary та overview не є самостійним клінічним рішенням, а виконує функцію підтримки (decision support), надаючи лікарю структуровані дані з можливістю повернення до первинних джерел для підтвердження висновків. Такий підхід відповідає принципам доказової медицини, оскільки забезпечує зворотну трасованість даних та можливість верифікації на кожному етапі сумаризації.

У порівнянні з наявними рішеннями, такими як Google Med-PaLM [15] Amazon HealthLake [16] та Microsoft Azure Health Bot [17], SyncMD [18] орієнтована на реалізацію багаторівневої модульної структури сумаризації, яка враховує контекстуальну цілісність випадку та динамічну змінність медичних даних. На відміну від Med-PaLM, що зосереджений на генеративній відповіді без багаторівневого узагальнення, SyncMD реалізує послідовний процес стискання інформації через summary, briefs та patient overview, зберігаючи міждокументні зв'язки. Amazon HealthLake орієнтований на управління структурованими даними в контексті EMR, проте не реалізує алгоритмів узагальнення даних у кілька етапів. Azure Health Bot інтегрує базові NLP-функції для відповіді на клінічні запити, однак не має модульної архітектури прогресивної сумаризації з можливістю адаптації до конкретних запитів лікаря.

Застосування нейромережевих технологій для обробки медичних даних стало важливим напрямом розвитку сучасної медичної інформатики. Зростання обсягів цифрової інформації, зокрема історії хвороб, лабораторні показники, радіологічні звіти та вільнотекстові записи лікарів, створює об'єктивну потребу в автоматизованих системах, здатних не лише зберігати та обробляти ці дані, а й узагальнювати їх у клінічно значущу форму. Особливу цінність у цьому контексті мають великі мовні моделі (Large Language Models, LLM), здатні до контекстуального аналізу та генерації природномовних підсумків, які відповідають принципам доказової медицини. Проте застосування таких моделей у реальних умовах охорони здоров'я обмежене низкою чинників, серед яких важливе місце посідає обмеження контекстного вікна обробки в LLM та ризик втрати зв'язку між фрагментованими даними. У відповідь на ці виклики було розроблено декілька підходів до реалізації сумаризації клінічної інформації за допомогою нейромереж (табл. 1).

На практиці прогресивна сумаризація виявила високу ефективність у роботі з великомасштабними медичними архівами, зокрема при обслуговуванні пацієнтів із тривалими історіями хвороби, де обсяг накопичених даних перевищує технічні можливості великих мовних моделей. Для ілюстрації практичного застосування підходу прогресивної сумаризації в медичній сфері надалі буде розглянуто приклад платформи SyncMD [18], що демонструє можливість модульного стискання клінічної інформації на основі LLM. У межах цієї цифрової платформи реалізовано модульну архітектуру обробки даних, що передбачає послідовне стискання інформації в кілька логічних рівнів: спочатку здійснюється сумаризація кожного документа окремо, далі

Таблиця 1

Методи нейромережевої обробки медичних даних для сумаризації та релевантного висновування

Методологічний підхід	Принцип дії	Переваги	Технічні обмеження
Пряме введення документів у контекст LLM	Усі документи завантажуються в одне вікно великої мовної моделі	Простота реалізації, збереження первинної структури	Ліміт обсягу контексту, підвищене навантаження на пам'ять, сповільнення обробки
Retrieval-Augmented Generation (RAG)	Векторне індексування та пошук релевантних фрагментів з подальшою генерацією	Гнучкість, модульність, значна масштабованість	Втрата міждокументних зв'язків, залежність від embedding-якості
Прогресивна сумаризація	Багаторівневе узагальнення: summaries → briefs → patient overview	Збереження логіки розвитку кейсу, масштабованість, цілісність аналізу	Накопичення похибок на кожному етапі, потреба в стандартизованих шаблонах

Джерело: сформовано автором на основі [1; 2; 3; 6; 7; 9]

з метою індексації формується стисле резюме (brief) кожного з них, а на останньому етапі з усього масиву даних автоматично створюється інтегрований файл patient_overview – єдина клінічна картина [7]. Ця структура дає змогу лікарям швидко орієнтуватися в медичній історії, не втрачаючи важливих деталей, оскільки кожен етап сумаризації враховує як загальний контекст, так і семантично важливі фрагменти. Крім того, синхронізована навігація між summary, briefs та запитами користувача сприяє генеруванню обґрунтованих відповідей із посиланням на джерело даних, що відповідає принципам доказової медицини, знижуючи когнітивне навантаження на медичний персонал, скорочуючи час аналізу інформації та підвищуючи точність діагностичних рішень. Така архітектура може бути адаптована до різних типів медичних IT-систем за умови забезпечення сумісності з електронними медичними записами (EMR), дотримання вимог конфіденційності та захисту персональних даних пацієнтів.

Інтеграція нейромережевих модулів у медичні IT-системи потребує дотримання структурної сумісності, технологічної гнучкості та нормативної безпеки на всіх етапах взаємодії з даними. З урахуванням специфіки медичної документації, а також вимог до захисту персональної інформації, архітектурне рішення для впровадження модуля прогресивної сумаризації повинно спиратися на принципи розподіленої обробки, рольової авторизації доступу, масштабованої індексації та асинхронного обміну між елементами системи. Технічно це реалізується через побудову інтеграційних шарів, які забезпечують безперервну взаємодію між клінічними інформаційними джерелами, обчислювальними модулями штучного інтелекту, локальними сервісами зберігання та інтерфейсами користувача. Таке модульне рішення дає змогу адаптувати підхід до різних типів електронних медичних записів (EMR), зберігаючи при цьому високий рівень надійності та продуктивності системи (табл. 2).

Так, платформа SyncMD є прикладом того, як архітектура модуля прогресивної сумаризації може бути технічно реалізована як цілісний багаторівневий механізм у медичній IT-інфраструктурі. Кожен шар, із наведених у табл. 2, виконує незалежну, проте синхронізовану функцію: від перетворення клінічних даних до забезпечення цілеспрямованого доступу до релевантних результатів аналізу. Особливістю такої архітектури є її горизонтальна масштабованість: кожен компонент (наприклад, обробник brief-файлів або шар кешування) може бути масштабований окремо, що дає змогу системі адаптуватися до навантаження без повного перезапуску. Дотримання стандартів Health Level Seven (HL7) [19] та Fast Healthcare Interoperability Resources (FHIR) [20] на рівні інтероперабельності забезпечує сумісність із різними типами електронних медичних записів (Electronic Medical Records, EMR), що важливо для впровадження в умовах фрагментованої медичної інфраструктури.

Архітектурна схема реалізації модуля прогресивної сумаризації в межах медичної IT-інфраструктури представлена як послідовна система рівнів, що охоплює етапи введення даних (input), обробки (processing), сумаризації (summarization), виводу результатів (output) та збереження трасуваності даних (traceability). Кожен рівень забезпечує автономне, але синх-

ронізоване функціонування, що дає можливість зберігати структурованість інформації та підтримувати зворотну трасуваність до первинних документів. Такий підхід сприяє оптимізації обробки медичних даних у великих масштабах, забезпечуючи інтеграцію даних різних типів (текст, зображення, сигнали) з урахуванням стандартів HL7 [19] та FHIR [20] (рис. 1).

У структурі архітектури виділено п'ять функціональних рівнів. Input Layer відповідає за інтеграцію даних із зовнішніх джерел (EMR, лабораторні системи) із використанням протоколів HL7 та FHIR. Processing Layer виконує обробку даних за допомогою NLP-пайплайнів та контролює доступ до даних через механізми RBAC. Summarization Layer реалізує багаторівневу сумаризацію, формуючи summary, briefs та patient overview зі збереженням логічних зв'язків між даними. Output Layer відповідає за генерацію кінцевих звітів із можливістю цільового запиту та виводу релевантної інформації. Traceability Layer забезпечує збереження аудиту дій, попередніх версій даних та структури взаємозв'язків між summary та оригінальними джерелами, що дає змогу здійснювати перевірку правдивості результатів.

Так, архітектура забезпечує безперервний процес стискання даних із можливістю динамічного розширення та адаптації до нових обсягів інформації без втрати контекстуальної цілісності.

Обробка запитів відбувається асинхронно: користувач надсилає запит до навігаційного шару, який через індекс звертається до кешованих результатів, попередньо створених на основі процесів обробки великими мовними моделями (LLM). Завдяки цьому значно знижу-

Таблиця 2

Інтеграційні шари IT-архітектури для впровадження модуля прогресивної сумаризації в медичні системи

Рівень архітектури	Функціональна роль у системі	Технологічні вимоги та протоколи
Шар інтероперабельності	Забезпечення сумісності з EMR, перетворення форматів даних	HL7, FHIR, JSON-XML конвертери, адаптери API
Шар доступу до даних	Авторизоване зчитування та запис документів пацієнта	Аутентифікація (OAuth 2.0), рольове управління (RBAC), шифрування
Аналітичний шар	Обробка та стискання текстових даних за допомогою LLM	NLP-пайплайни, обмеження контекстного вікна, логіка викликів LLM
Шар зберігання та кешування	Локальне збереження результатів, попередні версії, індексовані відповіді	NoSQL-сховища, схеми контролю версій, захищене кешування
Шар запитів та виводу результатів	Навігація за структурою briefs, генерація відповідей на основі індексів	Інтерфейси запитів (Query Interface), логіка підсвічування джерела інформації

Джерело: сформовано автором на основі [4; 8; 9; 10; 12; 13]

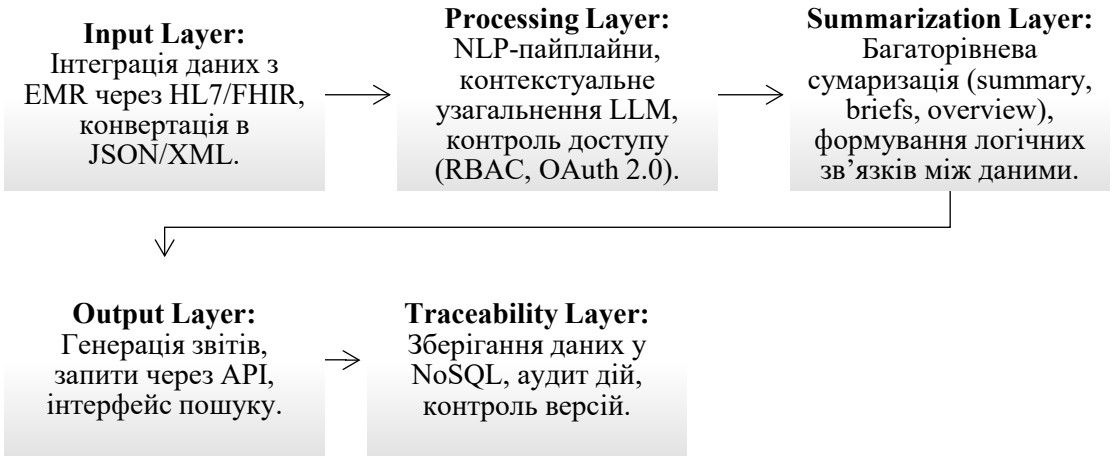


Рис. 1. Архітектурна схема реалізації модуля прогресивної сумаризації

Джерело: власна розробка автора

ється час генерації відповідей, а самі результати можуть бути репрезентовані у форматі, що підтримує структуровану навігацію по документах – наприклад, за діагнозами чи медичними подіями. Реалізовані протоколи аутентифікації OAuth 2.0 та рольового управління доступом (Role-Based Access Control, RBAC) відповідають стандартам конфіденційності, зокрема вимогам Health Insurance Portability and Accountability Act (HIPAA) у США [21] та General Data Protection Regulation (GDPR) в ЄС [22]. Використання внутрішніх інтерфейсів прикладного програмування (Application Programming Interfaces, API) дає змогу забезпечити динамічну взаємодію між компонентами системи та спростити впровадження модулів сумаризації в наявних рішеннях [9].

Загалом, така архітектура гарантує не лише ефективність обробки великих обсягів клінічних даних, але й довготривалу стійкість системи до відмов (fault tolerance), можливість поступового оновлення окремих модулів без втрати доступу до сервісу, а також адаптивність до нових технологічних змін без необхідності повного перепроєктування IT-інфраструктури.

У сучасних медичних інформаційних системах потреба в зручному та доказовому представленні великого обсягу різнорідної клінічної інформації стимулює розвиток методів інтелектуального стискання даних. Одним із найперспективніших напрямів є прогресивна сумаризація, що передбачає багаторівневу обробку медичних документів із поетапним формуванням узагальнених представлень: від коротких висновків щодо окремих записів до створення єдиного комплексного звіту, релевантного для клінічного прийняття рішень. Цей підхід дає змогу зберігати логіку хронологічного розвитку клінічного випадку, водночас знижуючи інформаційне навантаження на лікаря. На відміну від традиційних методів, що працюють із повним текстом або виконують одноетапне скорочення, прогресивна сумаризація реалізується як послідовна система трансформацій на основі семантичної релевантності, контекстного сліду та функціонального поділу даних. У практиці її застосування зразковим прикладом є платформа SyncMD, яка впровадила повноцінну логіку такої багатоступеневої обробки (табл. 3).

На відміну від одноетапних підходів до текстового стискання, методологія прогресивної сумаризації реалізує послідовну логіку когнітивної реконструкції медичних даних: кожен етап обробки не є ізольованим, а створює основу для подальшого узагальнення. Це сприяє уникненню втрати критичних смислових зв'язків, збереженню хронології клінічного випадку та обґрунтованості висновків. У межах платформи SyncMD ця модель реалізується як поетапний процес: на першому рівні ізольовані документи проходять контекстно-чутливу сумаризацію з урахуванням медичних категорій; на другому відбувається тематичне згортання згрупованих summary у brief-модулі, які формуються за типами даних, що полегшує навігацію та скорочує обсяг тексту до керованого. Основним елементом є третій рівень – формування Patient Overview, який є не простим об'єднанням brief-файлів, а інтегрованою репрезентацією всієї медичної історії пацієнта [5]. Цей рівень формує онтологічно зв'язну картину, що враховує не лише послідовність подій, а й функціональні взаємозв'язки між діагнозами, втручаннями та результатами лікування. Завдяки збереженню відповідностей між summary і джерелами система забезпечує зворотну трасованість висновків – користувач може переглянути, з яких конкретно документів сформовано той чи інший фрагмент звіту.

Четвертий рівень – діалоговий – реалізується у вигляді інтерфейсу, що взаємодіє з overview через механізми векторного пошуку, кластеризації медичних сутностей та інтерпретованих

Таблиця 3

Методологія прогресивної сумаризації медичних даних на прикладі реалізації в SyncMD

Рівень обробки	Об'єкт трансформації	Продукт виходу (output)	Основна функція на етапі
Індивідуальна сумаризація	Один медичний документ (visit, аналіз)	Summary (тематичний мінівисновок)	Виділення основних даних зі збереженням контексту
Тематичне згортання (briefs)	Група summary за темою або типом	Brief (узагальнення за типом: візити, рецепти)	Семантичне групування та зведення дублетів
Інтеграція кейсу (overview)	Усі briefs пацієнта	Patient Overview (єдина клінічна картина)	Побудова цілісної структури медичної історії
Пошуково-діалоговий інтерфейс	Overview + запит користувача	Цільова відповідь із посиланням на джерело	Навігація, уточнення, адаптивна генерація

Джерело: сформовано автором на основі [1; 3; 5; 6; 8; 14]

відповідей, сформованих за запитом лікаря. На цьому етапі важливо, що система не виконує повної генерації тексту, а діє як контекстно обмежений медіатор, що подає релевантну інформацію з поясненням того, звідки вона отримана. Це відповідає стандартам доказової медицини, оскільки будь-який запит можна перевірити за первинними документами.

Практичне значення цієї методології полягає в поєднанні гнучкості генеративного ШІ з вимогами клінічної відповідальності: на всіх етапах система зберігає доказову базу, динамічно адаптується до оновлень даних та може інтегруватися в системи прийняття рішень (CDSS – Clinical Decision Support Systems) без втрати контексту чи системних помилок.

Попри технічний прогрес і зростання обчислювального потенціалу, ефективне впровадження нейромережевих систем у медичну практику постає перед низкою проблем, які виходять за межі власне алгоритмічного рівня. Одним з основних викликів є несумісність нейромережевих рішень з організаційно-правовою та функціональною структурою наявних електронних медичних записів (EMR), що в багатьох країнах зберігаються в закритих або фрагментованих форматах, не оптимізованих для динамічного аналізу [9]. Це унеможливлює автоматизовану побудову клінічного контексту та гальмує синхронізацію між системами обробки природної мови та медичними інформаційними платформами. У результаті нейромережеві модулі часто залишаються ізольованими від клінічних процесів, працюючи як допоміжні інструменти без глибокої інтеграції [11].

Іншою серйозною проблемою є слабка інтерпретованість результатів, яку демонструють великі мовні моделі при генерації клінічно орієнтованих висновків. Відсутність прозорих механізмів пояснення рішень, зокрема можливості зворотного простеження логіки генерації, створює ризики втрати довіри лікарів, особливо в умовах критичного прийняття рішень [7]. Медична практика потребує не лише точних, а й підконтрольних інтерпретацій, що можуть бути обґрунтовані джерелами даних та процедурними правилами. У випадку, коли LLM виступають як «чорна скринька», клініцисти змушені додатково верифікувати результати, що нівелює переваги автоматизації.

Ще одним обмеженням є невизначеність нормативної бази, яка регулює обробку персональних медичних даних при використанні генеративних моделей. Нині більшість комерційних LLM не забезпечують локального зберігання інформації та функцій обмеженого доступу в умовах, які б відповідали вимогам HIPAA [22] або GDPR [21]. Відсутність гарантій конфіденційності, неможливість контролю за внутрішніми обчисленнями та аудитом дій моделі обмежує застосування нейромереж у межах офіційно затверджених медичних протоколів. Це посилюється етичними проблемами, зокрема ризиком неправильної інтерпретації або упередженого виводу внаслідок тренування моделі на недостатньо репрезентативних або неактуальних наборах даних [8].

Крім того, у реальній клінічній практиці часто виникає розрив між очікуваною логікою клінічного мислення та структурою відповідей, які формує LLM. Це особливо виразно проявляється у випадках багатопрофільного ведення пацієнтів, де необхідна синхронізація результатів лабораторних, діагностичних і терапевтичних підсистем [3]. Нейромережі не завжди здатні утримати міждисциплінарний контекст без ручної верифікації та повторного налаштування запиту. Так, ефективне впровадження нейромережевих систем потребує не лише технологічної адаптації, а й нормативного перезавантаження, інституційного забезпечення прозорості, а також зміни принципів архітектурного проєктування клінічних IT-систем, здатних сприймати й супроводжувати результати інтелектуальної генерації.

Для успішного впровадження модуля прогресивної сумаризації в електронну медичну інфраструктуру доцільно орієнтуватися на поетапне розгортання технологічної моделі, здатної взаємодіяти з наявними медичними інформаційними системами без зміни їхньої базової логіки. Найважливішим чинником є забезпечення гнучкого каналу передачі структурованих даних між EMR (електронними медичними записами) та модулем сумаризації із застосуванням протоколів інтероперабельності, таких як FHIR [20] або HL7 [19], з метою стабільної синхронізації без ручного втручання. Ефективним рішенням є використання окремого середовища обробки, яке виконує незалежну сумаризацію кожного запису, після чого зберігає результат у структурованому вигляді зі збереженням посилань на джерело. Це сприяє зниженню когнітивного навантаження на лікаря шляхом надання лише релевантних фрагментів даних для клінічного запиту.

Для підвищення точності діагностичних висновків важливо реалізувати багаторівневу

модель агрегації, де кожен рівень стискання даних – від summary до overview – контролюється окремими логічними модулями, що зменшує ризик семантичної втрати або спотворення інформації. Важливо також забезпечити можливість інтерактивної верифікації результатів лікарем із наданням прямих посилань на оригінальні документи, що відповідає вимогам доказової медицини. Крім того, необхідно інтегрувати механізми візуального позначення змін, оновлень і потенційно критичних фрагментів у згенерованих звітах, що дає змогу лікарю фокусуватися на клінічно важливих змінах у динаміці. У впровадженні слід уникати повної заміни традиційної документації результатами генеративної моделі, натомість варто розглядати модуль сумаризації як сервіс підтримки прийняття рішень, що діє разом із лікарем, а не замість нього. Такий підхід не лише знижує адміністративне навантаження, а й підвищує точність у постановці клінічного діагнозу завдяки ефективному керуванню обсягом інформації та збереженню логіки її формування.

Висновки. У дослідженні обґрунтовано концепцію прогресивної сумаризації медичних даних на основі нейромережевих технологій, що забезпечує багаторівневе логіко-семантичне стискання записів у структурований формат для підтримки клінічних рішень. На прикладі платформи SyncMD продемонстровано, що така архітектура дає змогу інтегрувати великі мовні моделі (LLM) в електронну медичну інфраструктуру зі збереженням доказовості та контекстної повноти.

Встановлено, що найбільш ефективною є поетапна обробка даних із формуванням summary, briefs та patient overview як базової моделі навігації. Це сприяє зниженню навантаження на лікаря та підвищенню точності діагностики за умови дотримання стандартів інтероперабельності та реалізації механізмів верифікації.

До основних проблем упровадження нейромережевих систем віднесено нормативну невизначеність, непрозорість логіки LLM, низьку інтерпретованість результатів, фрагментованість EMR і розрив між алгоритмічною генерацією та логікою клінічного мислення.

Запропоновано практичні підходи до інтеграції модулів сумаризації: створення незалежного середовища обробки, багаторівневу агрегацію, підтримку діалогових інтерфейсів, візуалізацію змін та збереження аудиту дій моделі.

Подальші дослідження доцільно спрямувати на створення гібридних інтерпретованих моделей та розробку уніфікованих протоколів інтеграції з урахуванням нормативно-етичних вимог і клінічної валідності результатів.

Список літератури:

1. Kollias D., Tagaris A., Stafylopatis A., Kollias S., Tagaris G. Deep neural architectures for prediction in healthcare. *Complex Intelligent Systems*. 2018. Vol. 4, № 1. P. 119–131. DOI: https://doi.org/10.1007/s40747-017-0064-6 (date of access: 06.05.2025).

2. Bhowmik B. R., Varna S. A., Kumar A., Kumar R. Deep neural networks in healthcare systems. In: *Machine learning and deep learning in efficacy improvement of healthcare systems*. CRC Press, 2022. P. 195–226. https://www.taylorfrancis.com/chapters/edit/10.1201/9781003189053-10 (date of access: 06.05.2025).

3. Jiang Sh. Sustainable Development in Medical Applications Using Neural Network Architecture. *Scalable Computing: Practice and Experience*. 2024. Vol. 25, № 2. P. 782–791. DOI: https://doi.org/10.12694/scpe.v25i2.2631 (date of access: 06.05.2025).

4. Sarvamangala D. R., Kulkarni R. V. Convolutional neural networks in medical image understanding: a survey. *Evolutionary Intelligence*. 2022. Vol. 15, № 1. P. 1–22. DOI: https://doi.org/10.1007/s12065-020-00540-3 (date of access: 06.05.2025).

5. Anwar S. M., Majid M., Qayyum A., Awais M., Alnowami M., Khan M. K. Medical Image Analysis using Convolutional Neural Networks: A Review. *Journal of Medical Systems*. 2018. Vol. 42, № 12. Article 226. DOI: https://doi.org/10.1007/s10916-018-1088-1 (date of access: 06.05.2025).

6. Yu H., Yang L. T., Zhang Q., Armstrong D., Deen M. J. Convolutional neural networks for medical image analysis: state-of-the-art, comparisons, improvement and perspectives. *Neurocomputing*. 2021. Vol. 444, № 1. P. 92–110. DOI: https://doi.org/10.1016/j.neucom.2020.04.157 (date of access: 06.05.2025).

7. Salahuddin Z., Woodruff H. C., Chatterjee A., Lambin P. Transparency of deep neural networks for medical image analysis: A review of interpretability methods. *Computers in Biology and Medicine*. 2022. Vol. 140, № 1. Article 105111. DOI: https://doi.org/10.1016/j.compbiomed.2021.105111 (date of access: 06.05.2025).

8. Shahid N., Rappon T., Berta W. Applications of artificial neural networks in health care organizational decision-making: A scoping review. *PLOS ONE*. 2019. Vol. 14, № 2. Article e0212356. DOI: https://doi.org/10.1371/journal.pone.0212356 (date of access: 06.05.2025).

9. Kumar K., Kumar P., Deb D., Unguresan M. L., Muresan V. Artificial intelligence and machine learning based intervention in medical infrastructure: a review and future trends. *Healthcare*. 2023. Vol. 11, № 2. Article 207. DOI: https://doi.org/10.3390/healthcare11020207 (date of access: 06.05.2025).

10. Zion S. O., Asuquo P. An Overview of Neural Network Architectures for Healthcare. *2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS)*. 2020. P. 1–8. DOI: https://doi.org/10.1109/ICMCECS47690.2020.246980 (date of access: 06.05.2025).

11. Мартинюк Т. Б., Запетрук Я. В. Нейромережевий підхід до медичної експрес-діагностики. *Вісник Вінницького політехнічного інституту*. 2019. Вип. 6. С. 37–44. DOI: https://doi.org/10.31649/1997-9266-2019-147-6-37-44 (дата звернення: 06.05.2025).

12. Protogerou A., Papadopoulos S., Drosou A., Tzovaras D., Refanidis I. A graph neural network method for distributed anomaly detection in IoT. *Evolving Systems*. 2021. Vol. 12, № 1. P. 19–36. DOI: https://doi.org/10.1007/s12530-020-09347-0 (date of access: 06.05.2025).

13. Mohammed A. A., Almaiah F., Hajjej F., Pasha M. F., Rahim Khan O. H. F., Teo J., Zakarya M. An industrial IoT-based blockchain-enabled secure searchable encryption approach for healthcare systems using neural network. *Sensors*. 2022. Vol. 22, № 2. Article 572. DOI: https://doi.org/10.3390/s22020572 (date of access: 06.05.2025).

14. Melnychuk H. M., Semeniuk H. D., Kashivska R. S., Shovkova N. I., Melnyk N. S. Regulation of antioxidant enzymes in patients after periodontal treatment with natural agents. *Wiadomości Lekarskie*. 2022. Vol. 75, № 3. P. 584–589. PMID: 35522862. https://wiadlek.pl/wp-content/uploads/archive/2022/03/WLek202203104.pdf (date of access: 06.05.2025).

15. SyncMD: website. 2025. URL: https://www.syncmd.com (date of access: 05.05.2025)

16. HL7 International: website. 2025. URL: https://www.hl7.org (date of access: 06.05.2025).

17. HealthIT.gov – Fast Healthcare Interoperability Resources (FHIR): website. 2025. URL: https://ecqi.healthit.gov/fhir (date of access: 06.05.2025).

18. HIPAA. U.S. Department of Health & Human Services: website. 2025. URL: https://www.hhs.gov/hipaa/index.html (date of access: 06.05.2025).

19. General Data Protection Regulation (GDPR)European Commission: website. 2025. URL: https://commission.europa.eu/law/law-topic/data-protection_en (date of access: 06.05.2025).

Parinov I.S. INTEGRATION OF NEURAL NETWORK SYSTEMS INTO MEDICAL IT INFRASTRUCTURES: A TECHNICAL APPROACH TO THE DEVELOPMENT OF A MODULE FOR PROGRESSIVE SUMMARIZATION OF PATIENT DATA

The relevance of the study is determined by the growth of medical documentation in the electronic environment and the need for its effective compression while preserving clinically important information. In evidence-based medicine, it is extremely important to implement tools for processing large amounts of data and converting them into understandable structures to support medical decisions. Modern healthcare IT systems require flexible, scalable, and transparent solutions that can be integrated into heterogeneous infrastructures while maintaining confidentiality. Given this, it is important to study approaches that take into account both technical and ethical limitations of the use of neural networks in clinical practice.

The aim of the study is to justify a technological approach to the implementation of a neural network module for progressive summarization of patient data in medical IT infrastructures using the SyncMD project as an example, taking into account the requirements for compatibility, efficiency, and objectivity.

The work uses methods of comparative analysis of summarization algorithms, architectural modeling, structural and functional analysis of implementation on the SyncMD platform, a problem-analytical approach to identifying integration barriers, as well as applied modeling of the implementation of clinical decision support systems.

A model for the step-by-step generation of medical conclusions has been created, covering the levels of summary, briefs, overview, and dialog navigation while preserving the sources and logic of the clinical context. It has been established that such a multi-level structure reduces analysis time, supports the evidential value of decisions, and reduces the cognitive load on the physician. It has been noted that consistent data transformation logic improves consistency between results and source documents, which is critical for diagnostic accuracy.

The main limitations were identified: lack of transparency in LLM algorithms, regulatory risks regarding personal data protection, fragmentation of EMR systems, and the complexity of synchronization with clinical logic. Practical recommendations were proposed for integrating the summarization module into an independent environment with verification, interoperability, and model action auditing.

The research prospects are related to the development of hybrid interpreted generation models, standardization of integration procedures into medical systems, and adaptation of architectures to regulatory and ethical requirements.

Key words: *evidence-based medicine, natural language processing, electronic medical records, multi-level aggregation, interoperability, cognitive load, interpretability of results.*

UDC 004.8
DOI https://doi.org/10.32782/2663-5941/2025.3.2/39

Pasichnyk V.V.
Zaporizhzhia National University

Dobrovolskyi H.A.
Zaporizhzhia National University

SURVEY OF APPLICATION OF THE KOLMOGOROV-ARNOLD NETWORKS TO TRANSPORTATION NETWORK INTERDICTION GAMES

This paper is dedicated to the analysis of transportation network interdiction, which examines the mechanisms by which an adversary with limited resources can disable or degrade connections within a network, as well as the defender's response in rerouting flows to maintain system efficiency.

The evolution of interdiction methods is presented–from classical bilevel optimization approaches employing mixed-integer programming, Benders decomposition, cutting-plane methods, and heuristics to contemporary algorithms that leverage large-scale data analysis.

A classification of machine-learning methods for interdiction is provided, namely:

1. Reinforcement learning, which involves training agents through interaction with a simulated environment.

2. Adversarial and multi-agent learning, which employs self-play and online no-regret algorithms to develop robust strategies.

3. Imitation learning and inverse reinforcement learning, which rely on analyzing demonstrations or actual adversary trajectories.

4. Evolutionary algorithms combined with neural approximators, which offer scalable solutions at the expense of strict theoretical guarantees.

Theoretical foundations of Kolmogorov–Arnold Networks (KANs) are revealed, showing how, based on the universal approximation theorem, they approximate multivariate functions via a sequence of learned univariate mappings. Initial applications of KANs in transportation tasks–such as traffic flow optimization and power-grid state estimation–are analyzed, and the potential for modeling defender responses using a compact, interpretable architecture is identified.

It is determined that, for further advancement in this field, the following key challenges must be addressed:

1. Embedding bilevel decision logic into differentiable models.

2. Handling graph-structured inputs without sacrificing approximation properties.

3. Generating representative, "adversarial" training datasets.

4. Developing rigorous evaluation protocols and determining performance bounds.

A research roadmap is proposed that integrates optimization and learning approaches, prioritizes reproducibility on both synthetic and real datasets, and seeks to balance computational efficiency with model transparency. This survey establishes a foundation for the development of effective and interpretable interdiction strategies in transportation and other critical infrastructure networks.

Key words: *transportation networks, network interdiction, Kolmogorov-Arnold networks, neural network training, performance evaluation, comparative analysis, mathematical optimization.*

Statement of the problem. Network interdiction games model an attacker-defender scenario on a network, where one agent attempts to disrupt or block network routes while the other tries to maintain efficient connectivity. Formally, a network (e.g., a road network or supply chain) is given on which an operator (defender) seeks to perform an optimally efficient action, such as sending flow, finding the shortest path, or servicing demands. An interdictor (attacker) moves first by incapacitating or degrading some network elements (nodes or links) within a limited budget of attacks [1]. Then the operator routes flow or traffic in the damaged network attempting to minimize the loss (e.g., finds the quickest remaining route, or maximizes the throughput that can still be delivered). This sequential play defines a Stackelberg game: the attacker anticipates the defender's best response when choosing what to attack [1]. Depending on the context, the defender's "action" can be passive (simply the outcome of an optimization like shortest path) or active (e.g., allocating defensive resources or fortifications before

the attack, leading to a tri-level game [1], although we focus on the bilevel case here).

These models have critical relevance in logistics, as they help identify weak links in supply chains and transportation networks and devise strategies to improve network resilience [2]. In a typical formulation (often a Stackelberg game), an *interdictor* (leader) with limited resources chooses a set of network arcs or nodes to disable or degrade. Then, an *evader* (follower) finds its optimal path through the remaining network [2]. The interdictor's goal is to maximize the cost or length of the evader's route (e.g., lengthening the shortest path) while the evader seeks to minimize it, producing a two-level optimization problem. Classic analysis usually assumes perfect rationality and complete information, but these assumptions are often unrealistic in practice [3]. Moreover, solving such interdiction problems is usually computationally hard: even restricted versions (like selecting k arcs to remove) are NP-hard, making exact solvers impractical on large networks [2]. These challenges have spurred research into AI agent training methods that learn effective interdiction strategies through simulations and theoretical models.

Analysis of recent research and publications. Several comprehensive surveys have been published on transportation network interdiction, including Smith and Song (2020), who review a broad range of bilevel optimization models and solution techniques; Brown et al. (2019), which focuses on mixed-integer programming and decomposition methods; and Zhang and Li (2021), which examines machine-learning–based heuristics. These works systematically cover reinforcement learning, adversarial/multi-agent frameworks, imitation learning, and evolutionary or heuristic approaches, yet none of them mention Kolmogorov–Arnold Networks (KANs) as a potential modeling paradigm. Consequently, there is a gap in existing literature regarding the application of KANs to network interdiction, despite their ability to approximate complex multivariate defender-response functions via learned univariate mappings.

Task statement. This survey aims to trace the evolution of transportation network interdiction methodologies – from classical bilevel optimization techniques to modern AI-driven approaches – and to identify open challenges hindering further progress.

The current study is focused on reinforcement learning, adversarial (multi-agent) learning, imitation learning, and other relevant methodologies. Specifically, the objectives are to:

- Categorize existing interdiction approaches, contrasting operations-research methods with machine-learning frameworks, and summarize their respective strengths and limitations.
- Highlight the theoretical foundations and initial applications of Kolmogorov – Arnold Networks (KANs) in related transportation domains.
- Identify critical research gaps, including embedding bilevel decision logic into differentiable models, handling graph-structured inputs without losing approximation power, generating representative adversarial training datasets, and establishing rigorous evaluation protocols.

Propose a research roadmap that integrates optimization and learning approaches, emphasizes reproducibility on both synthetic and real-world datasets, and balances computational efficiency with model interpretability.

Outline of the main material of the study. Formulation of the problem

A rigorous mathematical formulation for the **network interdiction game** is typically expressed as a **Stackelberg (leader-follower) bilevel optimization problem** [6].

Consider a transportation network modeled as a directed graph $G = (V, E)$ where V is the set of nodes (vertices). E is the set of arcs (edges), with each arc $(i, j) \in E$ having associated parameters (e.g., capacity, cost, or travel time).

An attacker aims to degrade the network by **interdicting** (attacking/removing or partially disabling) a subset of arcs, subject to resource or budgetary constraints. The defender subsequently routes traffic or flow optimally through the network to minimize operational costs or maximize efficiency given the interdicted network.

The attacker (leader) chooses binary interdiction decisions:

$$x_{ij} = \begin{cases} 1, if\ arc(i,j)\ is\ interdicted\ (disabled\ or\ degraded) \\ 0, otherwise \end{cases}$$

The defender (follower) chooses continuous routing decisions:

$$f_{ij} \geq 0 \left(flow\ on\ arc(i,j)\ after\ interdiction \right)$$

The attacker selects arcs to interdict to maximize damage to the defender. Formally, the attacker solves the following problem:

$$\max_x Z\left(x, f^*(x)\right)$$

subject to interdiction budget constraints:

$$\sum_{(i,j)\in E} c_{ij}^a x_{ij} \leq B$$

where:

1. $Z\left(x, f^*(x)\right)$ is the damage (loss) function defined by the attacker (e.g., total increase in travel time or total flow reduction), dependent upon the defender's optimal response $f^*(x)$.
2. c_{ij}^a is the cost of interdicting arc (i, j).
3. B is the attacker's budget (limited resources for attacks).

Follower's Problem (Defender)

Given the interdicted arcs (decision x), the defender solves an inner-level optimization problem, typically to minimize their operational costs, or equivalently, to maintain optimal network efficiency:

$$f^*(x) \in arg\min_j C(f, x)$$

subject to network constraints: Flow conservation at each node $k \in V \setminus \{s,t\}$ (where s is source and t is sink):

$$\sum_{(i,k)\in E} f_{ik} - \sum_{(k,j)\in E} f_{kj} = 0, \forall k \in V\{s,t\}$$

Capacity constraints on arcs, adjusted for interdictions:

$$0 \le f_{ij} \le c_{ij}\left(1 - x_{ij}\right), \forall (i,j) \in E$$

where:

1. $C(f,x)$ is the defender's cost function, typically given by total transportation costs or total travel time.
2. c_{ij} is the original capacity of the arc (i,j) before interdiction.

The complete **bilevel Stackelberg interdiction game** formulation is expressed as:

$$\max_x Z\left(x, f^*(x)\right)$$

subject to:

$$\sum_{(i,j)\in E} c_{ij}^a x_{ij} \le B, x_{ij} \in \{0,1\}$$

and

$$f^*(x) \in arg\min_j C(f,x)$$

subject to:

1. Flow conservation:

$$\sum_{(i,k)\in E} f_{ik} - \sum_{(k,j)\in E} f_{kj} = 0, \forall k \in V\{s,t\}$$

2. Arc capacities (post-interdiction):

$$0 \le f_{ij} \le c_{ij}\left(1 - x_{ij}\right), \forall (i,j) \in E$$

Interpretation

Leader (Attacker): Moves first and strategically selects arcs to disrupt, considering how the follower will optimally respond.

Follower (Defender): Moves second, optimally redistributing traffic or flow on the network given the damage caused by the leader.

Transportation systems provide many practical cases of network interdiction games. Consider a highway network where an attacker aims to maximize congestion by disabling a few critical interchanges while the traffic authority reroutes vehicles to minimize total delay. This can be cast as an interdiction game in which the attacker removes network capacity and the defender solves a traffic assignment or flow minimization problem [7]. Similar models appear in the analysis of terrorist threats to transit systems, freight network resilience against disruptions, and smuggler interception on transportation corridors [3]. A common objective for the interdictor is to inflict maximum damage (e.g., maximize travel time or flow loss), while the defender's aim is typically to minimize that damage – yielding a zero-sum or leader-follower optimization framework [7].

Solving network interdiction problems optimally is generally **computationally hard**. Even with moderate network sizes, the bilevel formulation becomes large and non-convex. Traditional methods reformulate or relax the problem into single-level models. For example, one can replace the defender's linear program (like a max flow) with its dual and embed it into the attacker's problem, resulting in a single-level but still combinatorial formulation [8]. Exact algorithms often use cutting planes, branch-and-bound, or Benders decomposition (row generation) to search the attack space efficiently [8]. Despite these advances, the curse of dimensionality remains: as networks grow or as the attacker's budget increases, the number of possible attack combinations explodes. Smith and Song (2020) note that modern research is pushing into interdiction models with probabilistic and dynamic elements, which further increase complexity [6]. In such settings, exact solution methods can become prohibitively slow, and even heuristic or MILP solvers struggle to provide timely solutions [9]. This is problematic for transportation agencies that have to analyze worst-case disruption scenarios or identify vulnerabilities in large-scale networks (e.g., a national highway system).

Thus, exact solutions are challenging for large-scale networks, motivating the development of

advanced approximate solution techniques [10] including potentially promising methods like Kolmogorov-Arnold Networks.

Publication Selection

The identification and curation of relevant literature constitute the initial phase of this survey [4]. We combine an automated controlled snowball-sampling procedure [5] with manual screening and citation-network exploration to ensure both breadth and relevance.

Inclusion and Exclusion Criteria

We defined explicit inclusion and exclusion criteria for selecting publications to ensure a high-quality and relevant literature base.

Inclusion Criteria: We focused on peer-reviewed research articles (including conference papers and journal articles) and credible systematic literature reviews published roughly in the last 10 years (2015–2025). We included studies explicitly addressing the application of AI (e.g., machine learning, deep learning, evolutionary algorithms, knowledge-based systems) to software testing or quality control. Also, we included older seminal works if they were highly cited and foundational to the field (to provide historical context or definitions). We limited sources to English-language publications from reputable venues in software engineering and AI. For tools or industry solutions, official documentation or whitepapers (software manuals) were considered if they provided technical details.

To assemble a focused yet comprehensive body of work, we established explicit criteria. We included peer-reviewed journal articles and conference papers–published between 2010 and 2025–that directly address transportation network interdiction or closely related topics (e.g., logistics, supply-chain resilience). These works encompass both classical optimization techniques, such as bilevel programming, mixed-integer programming, and decomposition algorithms, and modern AI methods, including reinforcement learning, graph neural networks, and Kolmogorov–Arnold Networks. Additionally, we considered highly cited, pre-2010 publications that introduced foundational concepts in interdiction game theory or network security. All included studies are in English and originate from recognized operations-research and AI venues. Conversely, we excluded papers that discuss network flow or routing without an interdiction component, focus solely on defender-only optimization, or address AI system testing rather than using AI for interdiction. Non-peer-reviewed sources (e.g., blogs or opinion pieces) were omitted unless they provided critical methodological details, and purely theoretical works without computational or empirical evaluation were likewise set aside.

Publication Selection Method

Our selection unfolded in three stages. First, we performed keyword searches–using terms such as “transportation network interdiction”, “bilevel network games”, and “network interdiction learning” – across Scopus, Web of Science, and Google Scholar. An initial screening of titles and abstracts yielded approximately fifty candidate papers. Second, we applied the controlled snowball-sampling method of Dobrovolskyi and Keberle [5], tracing both forward and backward citations to capture seminal works while avoiding terminological drift; this refined our pool to about thirty-five publications. Finally, we obtained full texts (where available) and conducted a detailed manual review, verifying each against our inclusion/exclusion criteria, extracting key methodological details–such as network models, evaluation metrics, and principal findings–and discarding those that failed to meet our standards.

Analysis of collected publications

Kolmogorov-Arnold Networks

Kolmogorov-Arnold Networks are a novel class of neural networks grounded in a 1957 theorem by A.N. Kolmogorov (extended by V. Arnold), which states that any continuous function of several variables can be represented as a finite composition of one-dimensional functions [11]. Traditional neural networks achieve universal approximation by layering fixed non-linear activations; in contrast, KANs implement the Kolmogorov-Arnold construction directly. Each hidden “layer” in a KAN applies *learnable* transformations on each input dimension rather than a fixed activation across all neurons [12]. For example, instead of a weight w_{ij} multiplying input x_j KAN would apply a learned spline function $f_{ij}\left(x_j\right)$ input connection. These learned univariate functions serve as customized activations that can adapt their shape to the data. As a result, KANs can capture complex nonlinear relationships with a compact architecture.

One immediate advantage of KANs is **flexibility**. The network can mold its activation functions to fit intricate patterns, which is beneficial for highly nonlinear domains like traffic flow or infrastructure failure cascades. Another advantage is potential **interpretability**: the learned activation functions on each edge can be inspected and, in some cases, have been shown

to correlate with meaningful physical relationships [hollosi2024detection]. Indeed, KANs have been used to rediscover mathematical or physical laws from data, highlighting their ability to reveal the structure behind the patterns they learn [12]. Furthermore, variants like *Convolutional KANs* integrate spline-based activations into CNN architectures to reduce parameter counts while maintaining accuracy [12]. In image classification tasks (e.g. MNIST), such convolutional KANs matched traditional CNN performance using fewer resources [12]. This efficiency is attributed to KAN's capacity to approximate functions with fewer neurons by using richer neuron-wise activations. KANs are particularly effective at modeling high-dimensional datasets with intricate interdependencies. For example, in hyperspectral image classification and time-series forecasting, they have achieved leading results by accurately representing both spatial and temporal correlations [14]. All these traits suggest that KANs could be well-suited to model the complexity of transportation networks under attack, where system behavior (e.g., total travel time, flow distribution) is a highly nonlinear function of which components are disrupted.

However, KANs are not universally superior to conventional networks in every scenario. A comprehensive empirical study by Bodner *et al.* (2024) found that when controlling for model size and FLOPs, standard MLPs still outperformed KANs on most machine learning benchmarks except those involving symbolic function representation [13]. These findings underline that while KANs offer a powerful representational framework, they also introduce new challenges.

Applying Kolmogorov-Arnold networks to network interdiction is a cutting-edge idea that has only begun to be explored. Because KANs themselves have gained attention mainly in the past few years, there are not that many direct studies on using KANs for interdiction games. Recent advances along two parallel lines of inquiry provide valuable guidance. The first involves employing machine-learning methods–including neural networks–to augment or even replace conventional solvers in network interdiction tasks. The second focuses on the successful deployment of KAN models in a variety of transportation-focused problems. Together, these developments imply that KANs may be particularly well-suited to address the challenges posed by interdiction games in transportation networks.

KANs in Transportation and Network Optimization

Initial deployments of KAN architectures in transportation applications demonstrate their effectiveness in handling complex network challenges, suggesting they may be well suited for interdiction scenarios. A very recent example is the use of KANs for traffic flow optimization. Zhang *et al.* (2025) proposed a hybrid model combining KANs with graph convolutional networks (TrafficKAN-GCN) to optimize urban traffic signal timings [9]. By embedding a KAN into a graph-based architecture, they aimed to capture complex nonlinear flow responses to control measures. While this work is at the preprint stage, it reflects a broader trend of integrating KANs into graph and network problems. Indeed, KANs have also been tested in power systems for *state estimation*, where interpretability and accuracy are crucial [15]. In all these cases, KANs provided an advantage when the problem involved learning a complicated mapping (e.g., from sensor readings or control actions to outcomes) that classical models struggled with.

Bridging the Fronts: The convergence of these research fronts – ML-assisted interdiction and KAN-enabled transport models – points to a strong opportunity at their intersection. We have seen that neural approaches can handle large-scale interdiction problems faster than traditional methods [10], and that KANs can capture complex system behaviors in transportation contexts with high fidelity [12]. Therefore, it is anticipated that applying Kolmogorov-Arnold networks to network interdiction games in transportation will bring significant advances. In the next section, we identify these anticipated benefits in detail, discussing how KAN's characteristics align with the needs of interdiction problems. We then analyze the challenges that must be addressed to realize these benefits.

AI-Based Approaches to Network Interdiction

Methods from machine learning and artificial intelligence have been employed to overcome the scalability and uncertainty challenges of classical interdiction solvers. In the last decade, the literature on AI-driven interdiction can be grouped into four main categories:

Reinforcement Learning

Reinforcement learning frames interdiction as a sequential decision-making process in a simulated environment [2]. Recent studies model the shortest-path interdiction problem as a Markov decision process and train deep RL agents–often using

pointer networks–to select edges to disable. These agents learn policies that maximize a reward signal (e.g., the evader's path length), achieving near-optimal interdiction sets on large instances where mixed-integer solvers become intractable [16].

Adversarial and Multi-Agent Learning

Adversarial learning approaches explicitly account for the strategic interaction between interdictor and evader. Online no-regret algorithms update the defender's strategy against adaptive opponents, with provable sublinear regret bounds even when the evader adjusts tactics over time. Self-play RL schemes co-train attacker and defender agents, driving both toward equilibrium policies that generalize across network topologies [17].

Imitation and Inverse Reinforcement Learning

When demonstrations of expert or observed behavior are available, imitation learning and inverse RL can infer an opponent model. In one dynamic local interdiction study, human-generated evader trajectories were used to learn a reward function via IRL; this model then informed the interdictor's policy, which was further refined online through RL, yielding higher interdiction accuracy than [18].

Evolutionary and Heuristic Methods with Neural Approximators

Genetic algorithms and other population-based heuristics have long been applied to interdiction, identifying near-optimal link-removal sets on large networks. More recently, graph neural networks have been trained on MILP-derived solutions–termed "Network Interdiction Goes Neural" – to predict effective interdiction decisions in real time, outperforming greedy heuristics and matching solver quality at a fraction of the runtime [9].

By consolidating these paradigms into a single overview, we highlight how AI techniques–ranging from trial-and-error learning to expert modeling and hybrid heuristics–provide scalable, adaptable alternatives to exact optimization in network interdiction contexts.

Gaps and challenges

Integrating Kolmogorov-Arnold networks into network interdiction games is at its very beginning stage. It raises unique issues at the intersection of game theory, graph data, and deep learning.

- This survey has traced the evolution of transportation network interdiction–from its origins in bilevel optimization and mixed-integer programming to the latest trends in data-driven and learning-based methods. Classical solution techniques such as Benders decomposition, cutting-plane algorithms, and heuristic search continue to provide rigorous foundations and provable guarantees, yet they often struggle with scalability on large, realistic networks. In parallel, machine-learning paradigms–including reinforcement learning, adversarial and multi-agent frameworks, imitation and inverse reinforcement learning, as well as evolutionary algorithms augmented by neural approximators–have demonstrated remarkable flexibility and speed, albeit sometimes at the cost of formal optimality bounds.

A central theme of this review is the promise of Kolmogorov–Arnold Networks (KANs). By approximating multivariate defender-response functions via learned univariate mappings, KANs offer a compact architecture with the potential for greater interpretability than conventional deep-learning models. Early applications in traffic flow optimization and power-grid state estimation suggest that KANs can capture complex nonlinear behaviors in networked systems. Yet the direct application of KANs to interdiction games remains nascent.

From our analysis, four key research challenges emerge:

1. Bilevel Logic Embedding
2. Translating the sequential attacker–defender structure into a differentiable learning paradigm, without losing theoretical rigor.
3. Graph-Structured Input Handling
4. Extending KAN theory to accommodate networks' combinatorial topology, while preserving universal approximation guarantees.
5. Adversarial and Diverse Data Generation
6. Designing training datasets that reflect realistic, stochastic interdiction scenarios and adversarial variations.
7. Evaluation Protocols and Performance Guarantees
8. Developing standardized benchmarks, reproducible evaluation procedures, and methods to quantify approximation gaps or regret.

Addressing these challenges will require concerted efforts across optimization theory, graph-based learning, and robust machine-learning methodology. We foresee that hybrid frameworks–combining KAN-driven approximators with local optimization refinements–could deliver both high performance and verifiable guarantees. Moreover, establishing open datasets and shared evaluation platforms will be critical to drive reproducible progress.

In summary, while classical and modern AI-based methods each contribute valuable strengths, integrating KANs into transportation interdiction

promises a new balance between computational efficiency, modeling fidelity, and interpretability. As research advances along the identified directions, we expect to see the emergence of more transparent, scalable, and reliable interdiction strategies applicable not only to road and logistics networks, but also to other critical infrastructures such as power grids and communication systems.

Bibliography:

1. Smith J.C., Lim C. Algorithms for Network Interdiction and Fortification Games. In: Pareto Optimality, Game Theory and Equilibria. 2008. P. 609–644. DOI 10.1007/978-0-387-77247-9_24.
2. Huang D., Mao Z., Fang K., Chen L. Solving the shortest path interdiction problem via reinforcement learning. Int. J. Prod. Res. 2023. Vol. 61, No 1. P. 31–48. DOI 10.1080/00207543.2021.2002962.
3. Guo Q., An B., Tran-Thanh L. Playing repeated network interdiction games with semi-bandit feedback. In: Proc. 26th Int. Joint Conf. Artif. Intell. (IJCAI-17). 2017. P. 3682–3690. DOI 10.24963/ijcai.2017/515.
4. Newman M., Gough D. Systematic Reviews in Educational Research: Methodology, Perspectives and Application. In: Systematic Reviews in Educational Research. Springer Fachmedien Wiesbaden, 2020. P. 3–22. DOI 10.1007/978-3-658-27602-7_1.
5. Dobrovolskyi H., Keberle N. Obtaining the Minimal Terminologically Saturated Document Set with Controlled Snowball Sampling. In: Proc. ICTERI. 2020. P. 87–101.
6. Smith J.C., Song Y. A survey of network interdiction models and algorithms. European Journal of Operational Research. 2020. Vol. 283, No 3. P. 797–811. DOI 10.1016/j.ejor.2019.06.024.
7. Zhang T., Fardad M. On the Optimal Interdiction of Transportation Networks. In: Proc. 2020 American Control Conference (ACC). IEEE, 2020. P. 4701–4706. DOI 10.23919/ACC45564.2020.9147435.
8. Smith J.C., Song Y. A survey of network interdiction models and algorithms. European Journal of Operational Research. 2020. Vol. 283, No 3. P. 797–811. DOI 10.1016/j.ejor.2019.06.024.
9. Zhang L., Chen Z., Lu C.-T., Zhao L. Network Interdiction Goes Neural. arXiv preprint arXiv:2405.16409. 2024. DOI 10.48550/arXiv.2405.16409.
10. Baycik N.O. Machine learning-based approaches to solve the maximum flow network interdiction problem. Comput. Ind. Eng. 2021. Vol. 162. P. 107873. DOI 10.1016/j.cie.2021.107873.
11. Spallanzani L.V. Reti neurali di Kolmogorov-Arnold: un'introduzione. 2023. URL: https://amslaurea.unibo.it/id/eprint/32533 (дата звернення: 18.05.2025).
12. Hollósi J., Ballagi Á., Kovács G. et al. Detection of bus driver mobile phone usage using Kolmogorov-Arnold networks. Computers. 2024. Vol. 13, No 9. P. 218. DOI 10.3390/computers13090218.
13. Bodner A.D., Tepsich A.S., Spolski J.N., Pourteau S. Convolutional Kolmogorov-Arnold Networks. arXiv preprint arXiv:2406.13155. 2024. DOI 10.48550/arXiv.2406.13155.
14. Ji T., Hou Y., Zhang D. A comprehensive survey on Kolmogorov-Arnold Networks (KAN). arXiv preprint arXiv:2407.11075. 2024. DOI 10.48550/arXiv.2407.11075.
15. Wang Z., Zainal A., Siraj M.M. et al. An intrusion detection model based on Convolutional Kolmogorov-Arnold Networks. Sci. Rep. 2025. Vol. 15, No 1. P. 1917. DOI 10.1038/s41598-024-85083-8.
16. Sreekumaran H., Hota A.R., Liu A.L., Uhan N.A., Sundaram S. Multi-agent decentralized network interdiction games. arXiv preprint arXiv:1503.01100. 2015. DOI 10.48550/arXiv.1503.01100.
17. Zheng K., Albert L.A. Interdiction models for delaying adversarial attacks against critical information technology infrastructure. Naval Res. Logist. 2019. Vol. 66, No 5. P. 411–429. DOI 10.1002/nav.21859.
18. Zeng Y., Xu K., Yin Q., Qin L., Zha Y., Yeoh W. Inverse Reinforcement Learning Based Human Behavior Modeling for Goal Recognition in Dynamic Local Network Interdiction. In: Proc. AAAI Workshops. 2018. P. 646–653.

Пасічник В.В., Добровольський Г.А. ЗАСТОСУВАННЯ МЕРЕЖ КОЛМОГОРОВА–АРНОЛЬДА ДО ІГОР ІНТЕРДИЦІЇ ТРАНСПОРТНИХ МЕРЕЖ

Стаття присвячена аналізу інтердиції транспортних мереж, що передбачає дослідження механізмів, за допомогою яких противник із обмеженими ресурсами може виводити з ладу або послаблювати зв'язки в мережі, а також реакцію захисника, який перенаправляє потоки для збереження ефективності системи.

У статті розкрито еволюцію методів інтердиції – від класичних бівертикальних оптимізаційних підходів із застосуванням змішаного цілочисельного програмування, декомпозиції Бендерса, методів розрізів та евристик до сучасних алгоритмів, що спираються на аналіз великих обсягів даних.

Розкрито класифікацію методів машинного навчання для інтердиції, а саме:

1. Навчання з підкріпленням, яке навчає агентів через взаємодію з імітованим середовищем;

2. Протидійне та багато-агентне навчання, що використовує навчання через гру з самим собою та онлайн-алгоритми без жалю для формування стабільних стратегій;

3. Імітаційне навчання та зворотне навчання з підкріпленням, які вивчають поведінку противника на основі демонстрацій або реальних траєкторій;

4. Еволюційні алгоритми у поєднанні з нейронними апроксиматорами, що забезпечують масштабованість рішень за рахунок відсутності суворих теоретичних гарантій.

Розкрито теоретичні засади мереж Колмогорова–Арнольда (KAN), які на основі універсальної теореми апроксимують багатовимірні функції через послідовність вивчених одновимірних відображень, та проаналізовано їхні початкові застосування в транспортних завданнях (зокрема оптимізація дорожнього руху та оцінка стану енергомереж). Визначено можливості використання KAN для моделювання реакції захисника в компактній та інтерпретованій архітектурі

З'ясовано, що для подальшого розвитку даної галузі необхідно вирішити такі ключові виклики:

1. Вбудування бівертикальної логіки в диференційовані моделі;

2. Обробка граф-структурованих вхідних даних без втрати апроксимаційних властивостей;

3. Генерація репрезентативних, «протидійних» наборів даних для навчання;

4. Розробка строгих протоколів оцінки та визначення меж продуктивності.

Запропоновано дослідницьку дорожню карту, що поєднує оптимізаційні й навчальні підходи, надає пріоритет відтворюваності результатів на синтетичних і реальних даних та прагне збалансувати обчислювальну ефективність із прозорістю моделей. Цей огляд створює основу для розробки ефективних і зрозумілих стратегій інтердиції в транспортних та інших критично важливих інфраструктурних мережах.

Ключові слова: *транспортні мережі, інтердикція мереж, мережі Колмогорова–Арнольда, навчання нейронних мереж, оцінка ефективності, порівняльний аналіз, математична оптимізація.*

УДК 681.518
DOI https://doi.org/10.32782/2663-5941/2025.3.2/40

Положаєнко С.А.
Національний університет «Одеська політехніка»

Прокофьєв А.Ю.
Національний університет «Одеська політехніка»

АДАПТИВНИЙ МЕТОД КОНТРОЛЮ ТА КРИТЕРІЇ ПЕРЕВІРКИ ДОСТОВІРНОСТІ ОБЧИСЛЮВАЛЬНОЇ РЕАЛІЗАЦІЇ МОДЕЛЕЙ ДИНАМІЧНИХ СИСТЕМ

Статтю присвячено проблемі підвищення достовірності обчислювальної реалізації математичних моделей динамічних систем.

Визначено, що питання дослідження динамічних систем найбільш суттєво постають при розробці, експлуатації та модернізації систем управління різноманітними об'єктами та процесами. При цьому встановлено показову рису сучасних систем управління, яка полягає у широкому використанні численних типів обчислювальних систем, що реалізують методи і алгоритми моделювання та управління. Тому якість функціонування систем моделювання та управління багато в чому визначається характеристиками задіяних обчислювальних засобів. В ході проведеного аналізу показано, що проблема забезпечення якості функціонування, зокрема, управляючих систем з обчислювальною машиною в контурі управління, є дуже складною як з технічної, так і математичної точок зору. При вирішенні цієї проблеми практикою висунуто на одне з перших місць науковий напрямок, пов'язаний із розробкою методів та засобів, що забезпечують надійність і достовірність процесів обчислень та обробки інформації. Виділено особливий клас інженерних задач, в яких, при оцінці достовірності отриманого розв'язку, домінуюче значення мають час прогнозу (тобто швидкодія у його отриманні) та простота задіяних обчислювальних процедур. Зокрема, такі вимоги можуть виникати при розв'язуванні задач управління швидкоплинними динамічними процесами (об'єктами), особливо в умовах багатоваріантної постановки (до таких задач може бути віднесено: оптимізаційні задачі з ітераційними процедурами пошуку оптимального закону управління; задачі диспетчерського управління у нештатних та аварійних ситуаціях, коли відшукується найкращий, за певним критерієм, розв'язок з можливих альтернативних; задачі, які розв'язуються у «прискореному» масштабі часу, коли відшукуються закони програмного управління, за умови відсутності можливості організувати замкнуті системи управління, тощо). Гостротою означених вище вимог обумовлено актуальність виконаних в статті досліджень, а також останніми доведено, що для перевірки достовірності обчислень, стосовно динамічного стану досліджуваних об'єктів та процесів, раціонально застосовувати контрольні алгоритми. Причому, в разі обмеженого часу на обчислення, доцільно, в якості контрольного (перевірочного) алгоритму, використовувати спрощений алгоритм розв'язування задачі оцінки достовірності отриманого розв'язку вихідної задачі моделювання або управління.

Ключові слова: *динамічна система, модель, достовірність обчислень, контрольний алгоритм, критерій перевірки обчислень, адаптивний метод.*

Постановка проблеми. Застосування цифрових регуляторів при управлінні динамічними об'єктами (системами або процесами) пов'язано з реалізацією їх моделей, а також відповідних алгоритмів управління, на основі різноманітних обчислювальних (за прийнятою термінологією [1, 2] – моделююче-управляючих) пристроїв. Досвід практичної реалізації систем моделювання та управління, в яких модель динамічного об'єкта виконано на основі обчислювальних засобів, показує, що якість функціонування таких систем переважно визначається якісними характеристиками останніх. Причому, в даному випадку, забезпечення бажаної якості функціонування визначається як програмно-апаратним складом обчислювальних засобів, які використовуються, так і модельною підтримкою – тобто математичним апаратом, на основі якого утворено відповідні моделі динамічних об'єктів. Тому, особливої уваги при вирішенні задачі забезпечення бажаних якісних характеристик моделююче-управляючих систем, набувають наукові дослідження, пов'язані із розробкою методів та засобів, що забезпечують *надійність*

та *достовірність* процесів обчислень та обробки інформації [3].

Слід зазначити, що при практичному впровадженні моделююче-управляючих комплексів динамічними системами обчислювальні процеси, які в них реалізуються, неодмінно наражаються на різного роду *завади* та *похибки*. Зокрема, комп'ютерне розв'язування математичних виразів, що утворюють математичні моделі (ММ) динамічних систем, може зазнавати похибки, спричинені числовими алгоритмами реалізації ММ (наприклад, неточностями апроксимації неперервних моделей або накопичення похибок округлення), та збоями в роботі апаратних засобів. Причому, більшість похибок та збоїв мають *випадковий характер*. Тому потрібен дієвий поточний контроль, що дозволяє виявити та усунути *похибки* (*збої*) до їх поширення в контурі управління, що зумовлює *оперативність* як одну з основних характеристик засобів контролю похибок.

Аналіз останніх досліджень і публікацій. На поточний час використовуються як *програмні* і *апаратурні* методи контролю процесів моделювання (чи управління), так і різноманітні їх поєднання, а детальний аналіз відповідних методів міститься, наприклад, в [4].

У загальному випадку, математична модель (ММ) динамічної системи, що реалізується в моделююче-управляючому комплексі, можна представити наступним чином:

$$\dot{\mathbf{Y}} = \mathbf{f}\left(\mathbf{Y},\mathbf{U},t\right),\ \mathbf{Y}\left(t_0\right) = \mathbf{Y}_0, \qquad (1)$$

де $\mathbf{Y} = \left(y_1, y_2, ..., y_n\right)^{\mathrm{T}}$ – вектор змінних стану, $\mathbf{U} = \left(u_1, u_2, ..., u_m\right)^{\mathrm{T}}$ – вектор сигналів управління, $\mathbf{f} = \left(f_1, f_2, ..., f_n\right)^{\mathrm{T}}$ – вектор-функція, t – незалежний параметр часу.

Для будь-яких методів числового розв'язування рівнянь динаміки виду (1) контроль збоїв можна організувати шляхом порівняння результатів, отриманих на поточному та наступному кроках інтегрування, відповідно.

Наявність чи відсутність збоїв визначається перевіркою умови

$$\xi = \rho\left(y_i, y_{i+1}\right) \in D, \qquad (2)$$

де ξ – міра близькості значень y_i та y_{i+1}, D – область допустимих значень цієї міри.

Для контролю похибок, викликаних збоями та відмовами, можуть використовуватися методи контролю методичної похибки числового розв'язку на етапі інтегрування або особливості побудови обчислювальних алгоритмів [5].

У методах прогнозу та корекції [6, 7] по різниці $\xi = \left|y_{\mathrm{f}} - y_{\mathrm{c}}\right|$ між прогнозованим y_{f} і скорегованим y_{c} значеннями шуканої функції можна робити висновки про наявності збою у роботі апаратури. Однак, існуючий взаємозв'язок між значеннями y_{f} та y_{c} знімає достовірність контролю.

Наявність такого зв'язку зумовлено наступним. У методі «прогноз-корекція» значення шуканої функції на поточному $\left(i+1\right)$-му та попередньому i-му кроках обчислень зв'язані наступним співвідношенням:

$$y_{i+1} = y_i + h\, f_i, \qquad (3)$$

де h – крок інтегрування, функція f_i визначається (1), а корекція здійснюється у відповідності до виразу

$$y_{i+1} = y_i + \frac{h}{2}\left(f_i + \tilde{f}_{i+1}\right), \qquad (4)$$

де $\tilde{f}_{i+1} = f\left(\tilde{\mathbf{Y}},\mathbf{U},t_{i+1}\right)$.

Як видно з (3) і (4), на одному кроці інтегрування двічі обчислюється функція $f\left(\mathbf{Y},\mathbf{U},t_i\right)$ у точках t_i та t_{i+1}. Оскільки цей процес, як правило, займає основний обсяг обчислювальної роботи, можна припустити, що збій стався в момент обчислення $f\left(\mathbf{Y},\mathbf{U},t_i\right)$ і значення функції f_i отримано з похибкою Δf_i. Тоді значення функції, яке може бути отримане у відповідності до формули (3), буде

$$\tilde{y}_{i+1} = y_i + h\left(f_i + \Delta f_i\right) = y_i + h\, f_i + h\Delta f_i, \qquad (5)$$

Тут величина $\Delta y_{i+1} = h\Delta f_i$ – похибка в значенні прогнозу.

Розкладемо функцію $\tilde{f}_{i+1}$ в ряд Тейлора по y і, при цьому, обмежимося двома першими членами

$$\tilde{f}_{i+1} = f_i + \frac{\partial f_i}{\partial y} h f_i + \frac{\partial f_i}{\partial t} h\,.$$

Підставимо в останній вираз для $\tilde{f}_{i+1}$ значення функції f_i з похибкою, в результаті чого отримаємо

$$\tilde{f}_{i+1} = f_i + \frac{\partial f_i}{\partial y} h\left(f_i + \Delta f_i\right) + \frac{\partial f_i}{\partial t} h\,.$$

Обчислимо значення функції y_{i+1} у відповідності до формули (4)

$$y_{i+1} = y_i + \frac{h}{2}\left[\left(f_i + \Delta f_i\right) + f_i + \frac{\partial f_i}{\partial y} h_i\left(f_i + \Delta f_i\right) + \frac{\partial f_i}{\partial t} h\right], \qquad (6)$$

Різниця між прогнозованим y_{f} (5) та скорегованим y_{c} (6) значеннями буде дорівнювати

$$\xi = \left|\tilde{y}_{i+1} - y_{i+1}\right| = \frac{h}{2}\left|\Delta f_i\left(1 - \frac{\partial f_i}{\partial y}h\right) - h\left(\frac{\partial f_i}{\partial y}f_i + \frac{\partial f_i}{\partial t}\right)\right|$$

звідки видно, що величина ξ залежить від величин похідних $\frac{\partial f_i}{\partial y}$ та $\frac{\partial f_i}{\partial t}$, тобто одночасно впливає як на прогнозоване значення y_{f}, так і скореговане значення y_{c}. При цьому, навіть наявність значної похибки у скоригованому значенні $\Delta y_{i+1} = h\,\Delta f_i$ може бути «непоміченим», якщо величини цих похідних малі. Крім того, наявність множника $\frac{h}{2}$ при малих знижує «чутливість» величини ξ до похибок, які викликані збоями при обчисленні f_i.

Дієвим при контролі достовірності обчислень може розглядатися підхід, що ґрунтується на застосуванні, крім основного алгоритму A_{b} реалізації ММ виду (1), контрольного алгоритму A_{k} [8], причому в останньому використовується вже інформація перевірена на попередніх кроках (наприклад, кроках ітераційної процедури інтегрування (1)).

Відповідні *екстраполяційний* та *інтерполяційний* контрольні алгоритми, запропоновані в роботі [9], виглядають наступним чином:

$$\overline{y}_{i+1} = \sum_{p=0}^{p_1} a_p y_{p-1} + h\sum_{s=1}^{s_1} b_s f\left(\mathbf{Y}, \mathbf{U}, t_{i-s}\right) \qquad (7)$$

$$\overline{y}_i = \sum_{p=-1}^{p_1} a_p y_{i-p} + h\sum_{s=-1}^{s_1} b_s f\left(\mathbf{Y}, \mathbf{U}, t_{i-s}\right), \qquad (8)$$

де $h = \left(t_{i+1} - t_i\right)$ – крок інтегрування, a_p, b_s – коефіцієнти.

Інтерполяційна формула контролю (4), на відміну від екстраполяційної (7), враховує *зміну розв'язку на кроці обчислень, який контролюється*, що *зменшує різницю* між величинами y та $\overline{y}$ в алгоритмах A_{b} та A_k, відповідно.

Зазначимо, що коефіцієнти контрольних алгоритмів виду (7) та (8) визначаються з умови отримання точного результату за контрольним алгоритмом A_k для розв'язку $y_j(t)$, $j = \overline{1,n}$, який являє собою багаточлен якомога більшого ступеня. Такий вибір коефіцієнтів a_p та b_s дозволяє отримати заданий порядок похибки контролю при відносно невеликому обсязі обчислень за контрольним алгоритмом A_k в процесі підрахунку.

Постановка завдання. Мета чинної роботи полягає у розробці та дослідженні адаптивного методу контролю достовірності обчислювальної реалізації математичних моделей динамічних систем при розв'язуванні задач їх моделювання та управління, а також у формуванні відповідних критеріїв перевірки достовірності контролю обчислень, виконаних за адаптивним методом.

Виклад основного матеріалу. коефіцієнти контрольних алгоритмів виду (7) та (8) визначаються з умови отримання точного результату за контрольним алгоритмом A_k для розв'язку $y_j(t)$, $j = \overline{1,n}$, який являє собою багаточлен якомога більшого ступеня. Такий вибір коефіцієнтів a_p та b_s дозволяє отримати заданий порядок похибки контролю при відносно невеликому обсязі обчислень за контрольним алгоритмом A_k в процесі підрахунку. Фактично точність контролю, що визначається залишком ряду Тейлора для контрольних формул, коливається від кроку до кроку, залежно від гладкості розв'язку $y_j(t)$, $j = \overline{1,n}$.

Нижче запропоновано метод організації контролю розв'язування системи рівнянь (1) за допомогою контрольного алгоритму A_{k}, що *перебудовується* залежно від поведінки контрольованого розв'язку $y_j(t)$, $j = \overline{1,n}$. Особливість даного підходу полягає в тому, що проконтрольовані на черговому кроці результати використовуються для визначення параметрів контрольного алгоритму A_{k} (тобто коефіцієнтів a_p, b_s). При цьому параметри визначаються з умови мінімуму певної міри близькості результатів, отриманих за контрольним A_{k} та основним A_{b} алгоритмами на ділянці розв'язку, що включає останній проконтрольований крок обчислень. Отриманий алгоритм, що позначимо як A_{a}, використовується для контролю наступного кроку обчислень. Після цього параметри знову уточнюються.

Очевидно, що описану процедуру методу та реалізуючий її алгоритм A_{a}, зважаючи на вищесказане, можна вважати *адаптивними*.

При цьому, адаптивний метод, який пропонується, допускає наступне трактування: на основі проконтрольованої інформації про фізичний процес (тобто динаміки останнього, що описується системою (1), і припущення, що поведінка розв'язку $y_j(t)$, $j = \overline{1,n}$ від кроку до кроку змінюється достатньо повільно) розв'язується задача *параметричної ідентифікації* при заданій структурі математичного опису процесу (тобто системи (1)). Отримана модель, в даному випадку контрольна формула у вигляді алгоритму A_{a}, використовується для наступного контролю. Інакше, у міру надходження проконтр-

ольованої інформації, модель, яка використовується для контролю (алгоритм), уточнюється (або адаптується). Надалі, визначення параметрів контрольного алгоритму, в якості якого розглядається вже алгоритм A_{a}, будемо використовувати методи параметричної ідентифікації безінерційних об'єктів [10, 11].

Застосовуючи результати обчислень за основним алгоритмом A_{b}, отримані до $(i+1)$-ого кроку обчислень, можна визначити параметри контрольного алгоритму A_{a} на $(i+1)$-му кроці із системи алгебраїчних рівнянь.

Тобто

$$\mathbf{CQ} = \mathbf{B}, \tag{9}$$

До прикладу, для ектраполяційного способу контролю, що використовують оцінку розв'язку в точці $t_{i-1+\alpha}$, величини, що входять у вираз (9), визначаються на $(i+1)$-му кроці обчислювальної процедури наступним чином:

$$\mathbf{Q} = \left(a_{-1}, a_1, ..., a_{p_1}, ..., b_0, b_1, ..., b_{s_1}\right)^{\text{T}},$$

$$\mathbf{B} = \left(y_i, y_{i-1}, ..., y_{i-k+1}\right)^{\text{T}},$$

$$\mathbf{C} = \begin{bmatrix} y_{i-1} \dots y_{i-p_1-1}\ \dot{y}_{i-2+\alpha}\ \dot{y}_{i-2} \dots \dot{y}_{i-s_1-1} \\ y_{i-2} \dots y_{i-p_1-2}\ \dot{y}_{i-3+\alpha}\ \dot{y}_{i-3} \dots \dot{y}_{i-s_1-2} \\ \dots\ \dots\ \dots\ \dots\ \dots\ \dots\ \dots\ \dots \\ y_{i-k} \dots y_{i-p_1-k}\ \dot{y}_{i-k-1+\alpha}\ \dot{y}_{i-k-1} \dots \dot{y}_{i-s_1-k} \end{bmatrix}.$$

Якщо матриця $\mathbf{C}$ не вироджена, і число її рядків дорівнює розмірності вектора параметрів (для екстраполяційного методу $\Pi_{\text{ers}} = \left(a_{-1}, a_1, a_2, ..., a_p, b_0, b_1, ..., b_s\right)$ та інтерполяційного методу $\Pi_{\text{int}} = \left(a_1, a_2, ..., a_p, b_1, ..., b_s\right)$, відповідно), що визначається значенням k, то оцінкою $\mathbf{Q}$ матриці параметрів $\mathbf{Q}$ (як для екстраполяційного, так і для інтерполяційного методів) є розв'язок системи алгебраїчних рівнянь

$$\hat{\mathbf{Q}} = \mathbf{C}^{-1}\mathbf{B}$$

і результати, отримані за основним алгоритмом A_{b} та контрольним алгоритмом A_{a} на послідовності $y_i, y_{i-1}, ..., y_{i-k+1}$ співпадають. Якщо матриця $\mathbf{C}$ прямокутна, то для визначення оцінки вектора параметрів можна скористатися методом найменших квадратів [12].

Розглянемо два варіанти побудови конструктивних алгоритмів адаптивного методу контролю точності обчислень при розв'язуванні задач моделювання та управління динамічними системами.

Нехай, для конкретності викладення, але не порушуючи його загальності, розглядається задача моделювання динамічного об'єкта, в якості ММ якого обрано рівняння Вольтери другого роду, причому останнє записано у еквівалентній формі:

$$y(t) + \int_{t_0}^{t} \frac{K_t'(t,s)}{K(t,t)} y(s)\, ds = \frac{f'(t)}{K(t,t)} \tag{10}$$

Будемо відшукати резольвенту $R_p(t,s)$ ядра $K(t,s)$ рівняння (10).

При числовому розв'язуванні рівняння (10) можна використати метод квадратурних формул, що у підсумку дає змогу записати розв'язок наступним чином:

$$y_1 = \frac{1}{K_{1,1}}\left[\frac{f_1}{h_1} - \frac{1}{2}K_{1,0}y_0\right], ...,$$

$$y_i = \frac{2}{K_{i,i}}\left[\frac{f_i}{h_i} - \frac{h_1}{2h_0}K_{i,0} - \frac{1}{2}\sum_{j=1}^{i-1}\left(\frac{h_j + h_{j-1}}{h_i}K_{ij}y_j\right)\right] \tag{11}$$

де $t_1 = (t_0 + h_1)$, $t_2 = (t_1 + h_2)$, ..., $t_n = T$; а кроки дискретизації h_i визначаються як $h_i = (t_i - t_{i-1})$, $i = \overline{0,(n-1)}$.

Після виконаної постановки задачі із записом рівняння динаміки (наприклад, у формі (10), (11), перший алгоритм виглядає наступним чином.

Алгоритм A_{ak}^1.

1. Зведення розв'язку вихідної задачі (10), (11) до виду

$$\begin{gathered} k_1 = h_i f(y_i, u_i, t_i), \\ k_2 = h_i f\left[y_i + \beta k_1, u(t_i + \alpha h), t_i + \alpha h_i\right], \\ y_{i+1} = y_i + \sum_{r=1}^{i} \gamma_r k_r;\ i = \overline{0,(n-1)}, \end{gathered} \tag{12}$$

де α, β, γ – параметри алгоритму.

2. Визначення контрольного значення функції по екстраполяційній формулі (7) першого порядку точності:

$$\overline{y}_{i+1} = a y_i - h_i b \dot{y}_{i-1+\alpha};\ i = \overline{0,(n-1)}, \tag{13}$$

де $\dot{y}_{i-1+\alpha} = f\left[y_i + \beta k_1, u(t_i + \alpha h_i), t_i + \alpha h_i\right]$ – значення оцінки прирощення, отримане за алгоритмом A_{b} на i-му кроці обчислень. На першому кроці обчислень коефіцієнти a, b визначаються з умови отримання точного результату за контрольним алгоритмом A_{k} для багаточлена першого степеня точності.

3. Визначення величини

$$\xi = \left| y_{i+1} - \overline{y}_{i+1} \right| .$$

4. Перевірка умови

$$\xi \le \xi_{\text{sup}} , \qquad (14)$$

де ξ_{sup} – допустиме значення ξ. Якщо умова (14) не виконується, то виконати корегування розв'язку (наприклад, заміною y_{i+1} на $\overline{y}_{i+1}$). Якщо $(i-k-1)$, то перейти до п. 5, інакше здійснити завершення процесу обчислень.

5. Сформувати матрицю

$$\mathbf{C} = \begin{bmatrix} y_{i+1} & \dot{y}_{i-2+\alpha} \\ \dots & \dots \quad \dots \\ y_{i-k} & \dot{y}_{i-k-1+\alpha} \end{bmatrix},$$

де $\dot{y}_{i-k-1+\alpha}$ – значення оцінки прирощення, отримане за процедурою (13) на $(i-k-1)$-му кроці обчислень. Величина k визначає число рядків матриці $\mathbf{C}$.

6. Сформувати вектор

$$\mathbf{B} = \left(y_i, ..., y_{i-k-1} \right)^{\mathrm{T}} .$$

7. Визначити коефіцієнти a, b контрольного алгоритму (9) з системи алгебраїчних рівнянь

$$\mathbf{CQ} = \mathbf{B} ,$$

де $\mathbf{Q} = (a, b)^{\mathrm{T}}$.

Для продовження розв'язування перейти до п. 1.

Як видно з наведеної процедури розв'язування, необхідні додаткові обчислення щодо визначення поточного значення коефіцієнтів a, b. Спростити процедуру контролю можна, визначаючи не весь вектор $\mathbf{Q}$, а його частину, вважаючи інші компоненти вектора такими, що рівні за значеннями, та такими, що отримані з умови точного результату за контрольною формулою для багаточлена як можливо високого ступеня. Так, в процедурі, яка аналізувалася, можна було визначати перший коефіцієнт a, вважаючи інший $b = 1$. Це відповідає перетворенню моделі (усічення моделі за параметрами).

Відмінність другого алгоритму полягає у способі визначення різниці між значеннями величин, отриманих по основному A_{b} та контрольному (екстраполяційному або інтерполяційному) алгоритмах, яка і являє собою похибку контролю. Зокрема, для екстраполяційного алгоритму виду (7) похибка ξ визначається різницею залишкових членів основного A_{b} та контрольного A_{k} алгоритмів.

Припустимо. що для задачі виду (10), (11) основний алгоритм A_{b} реалізує однокрокову процедуру, залишковий член якої має вигляд

$$r_0 = G_0 h_i^{k_0+1} y^{k_0+1} \varepsilon_0 , \; t_i \le \varepsilon_0 \le t_{i+1} ; \; i = \overline{0,(n-1)}, \qquad (15)$$

де k_0 – порядок точності основного алгоритму A_{b}, G_0 – деякий постійний коефіцієнт. Тоді, для значення функції, отриманого по алгоритму A_{b} на $(i+1)$-му кроці обчислень, можна записати

$$y_{i+1} = \tilde{y}_{i+1} + G_0 h_i^{k_0+1} y^{k_0+1} \varepsilon_0 ; \; i = \overline{0,(n-1)}, \qquad (16)$$

де $\tilde{y}_{i+1}$ – точне значення функції на $(i+1)$-му кроці обчислень за умови точного результату на попередніх кроках (в даному випадку враховується тільки *локальна* похибка на кроці, що визначається методичною похибкою числового методу).

Алгоритм A_{ak}^2.

1. Записуються вирази виду (15), (16) для контрольного алгоритму:

$$r_k = G_k h_i^{k_0+1} y^{k_0+1} \varepsilon_k , \; t_i \le \varepsilon_k \le t_{i+1} ; \; i = \overline{0,(n-1)},$$

$$\overline{y}_{i+1} = \tilde{y}_{i+1} + G_0 h_i^{k_0+1} y^{k_0+1} \varepsilon_0 ; \; i = \overline{0,(n-1)},$$

де r_k – порядок точності контрольного алгоритму, G_k – постійний коефіцієнт.

2. Обчислюється різниця значень, отриманих за основним A_{b} та контрольним A_{k} алгоритмах

$$R_{i+1} = y_{i+1} - \overline{y}_{i+1} = G_0 h_i^{k_0+1} y^{k_0+1} \varepsilon_0 - G_k h_i^{k_0+1} y^{k_0+1} \varepsilon_k ,$$

$$i = \overline{0,(n-1)} . \qquad (17)$$

3. Після проведення контролю на i-му кроці обчислень при відсутності збоїв (що встановлюється при контролі) визначається оцінка різниці залишкових членів R_i основного A_{b} та контрольного A_{k} алгоритмів

$$R_i = y_i - \overline{y}_i , \; i = \overline{0,(n-1)} . \qquad (18)$$

Величину R_i можна використовувати для компенсації різниці залишкових членів R_{i+1} на $(i+1)$-му кроці обчислень, якщо

$$R_i = R_{i+1} . \qquad (19)$$

4. Для екстраполяційного алгоритму (7) контрольна формула з компенсацією різниці залишкових членів має вигляд:

$$\overline{y}_{i+1}^{R} = \overline{y}_{i+1} + R_i , \; i = \overline{0,(n-1)} . \qquad (20)$$

Для інтерполяційного алгоритму (8) різниця (18) визначає оцінку залишкового члена контрольної інтерполяційної формули.

Контрольний інтерполяційний алгоритм з компенсацією має вигляд:

$$\overline{y}_i^R = \overline{y}_i + R_{i-1},\ i = \overline{0,(n-1)}. \qquad (21)$$

Контрольні алгоритми *з компенсацією* виду (20) та (21) можна записати через значення, отримані за основним алгоритмом A_b. Для цього необхідно величину R_i в (20) або в (21) представити у вигляді (18), записавши замість величини $\overline{y}_i$ алгоритм її отримання (екстраполяційний або інтерполяційний).

Наприклад, при використанні компенсації в контрольному алгоритмі (13) величина R_i для (18) визначається як (для простоти, за умови $a = b = 1$)

$$R_i = y_i - y_{i-1} - h_i \dot{y}_{i-2+\alpha},\ i = \overline{0,(n-1)}. \qquad (22)$$

і контрольний алгоритм з компенсацією (20) набуде вигляду

$$\overline{y}_{i+1}^R = \overline{y}_{i+1} + R_i = (2y_i - y_{i-1}) + h_i (\dot{y}_{i-1+\alpha} + y_{i-2+\alpha}),$$

$$i = \overline{0,(n-1)}. \qquad (23)$$

На рис. 1 та рис. 2 показано залежності математичного очікування $M[\xi]$ та середнього квадратичного значення відхилення $\sigma[\xi]$ величини ξ, що визначалася для прикладної задачі (10) за формулою

$$\xi = \sum_{j=1}^{n} \frac{|y_{j,i} - \overline{y}_{j,i}|}{\max\limits_t |y_j|},\ j = \overline{1,6}.$$

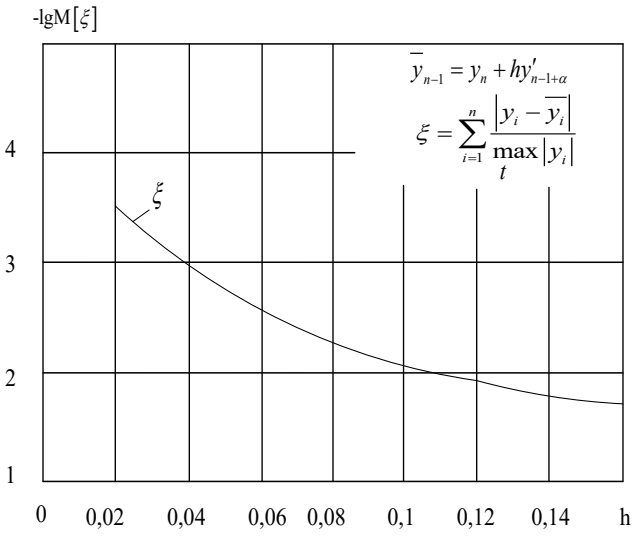


Рис. 1. Залежність математичного очікування $M[\xi]$ величини ξ від кроку h

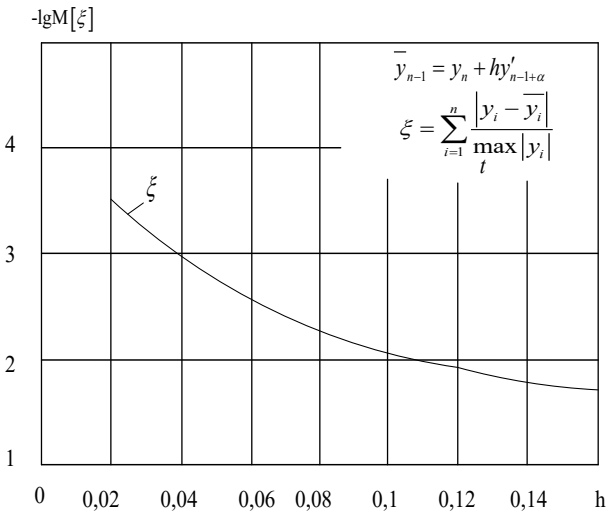


Рис. 2. Залежність середнього квадратичного відхилення $\sigma[\xi]$ величини ξ від кроку h

Для контролю числового розв'язку використовувалася екстраполяційна формула першого порядку точності виду

$$\overline{y}_{i+1} = y_i + h_i y_{i-1+\alpha}$$

без компенсації різниці залишкових членів та формули (23) з компенсацією, відповідно. Як випливає з графіків, точність контролю з компенсацією різниці залишкових членів на порядок вищий точності контролю без компенсації.

Зазначимо, що в якості критеріїв щодо прийняття рішення про факт наявності або відсутності похибок (збою), тобто – *критеріїв перевірки достовірності* обчислювальної реалізації моделей динамічних систем – пропонується обрати вирази (14) або (17) (для кожного з адаптивних алгоритмів A_{ak}^1 або A_{ak}^2, відповідно).

Висновки. Виконане аналітичне дослідження показало, що застосування адаптації дозволяє підвищити достовірність та точність екстраполяційного та інтерполяційного контрольних алгоритмів. У відповідності до цього, запропоновано критерії, які дозволяють встановлювати факт наявності (або відсутності) похибки в основному алгоритмі числової реалізації ММ динамічної системи, що дає змогу виявляти збої в роботі основного обчислювального алгоритму. Останнє дозволило розробити адаптивний метод контролю достовірності обчислювальної реалізації ММ динамічних систем, який забезпечує визначення параметрів контрольного алгоритму з умови мінімуму обраної міри (тобто, критерію достовірності) близькості результатів, отриманих за контрольним та основним алгоритмами.

Список літератури:

1. Пупена О. М., Ельперін І. В., Луцька Н. М., Ладанюк А. П. Промислові мережі та інтеграційні технології в автоматизованих системах. К.: Ліра-К, 2011. 500 с.
2. Гліненко, Л. К., Гліненко А. В., Сухоносов О. Г. Основи моделювання технічних систем. Львів: Бескид Біт, 2003. 176 с.
3. Костюшко І. А., Любашенко Н. Д., Третиник В. В. Методи обчислень. К.: НТТУ «КПІ ім. І. Сікорського», Вид-во «Політехніка», 2021. 243 с.
4. Чабанюк Я. М., Яковина В. С., Федасюк Д. В. Побудова і дослідження моделі надійності програмного забезпечення з індексом величини проекту. *Інженерія програмного забезпечення*, 2010. Вип. 1. С. 1–6.
5. Коссак О., Тумашова О. Методи наближених обчислень. Львів: Бак, 2003. 168 с.
6. Катаєва Є. Ю., Одокієнко С. М., Люта М. В., Савченко Я. С. Практичний аналіз якості програмного забезпечення з відкритим кодом. *Управління розвитком складних систем*, 2020. № 44. С. 49–55. URL: https://doi.org/10.32347/2412-9933.2020.44.49-55
7. Hoffmann G. M., Rajnarayan D. G., Waslander S. L., Jang J. S. (2004) The Stanford Testbed of Autonomous Rotorcraft for Multi Agent Control (STARMAC-2004). *In the Proceedings of the 23rd Digital Avionics System Conference. Salt Lake City*, UT. P. 12–16.
8. Положаєнко, С. А., Прокофьєв А. Ю. Моделі визначення надійності динамічних систем, функціонування яких характеризується режимом профілактики. *Вчені записки ТНУ імені В. І. Вернадського. Серія: Технічні науки*. 2024. Т. 35 (74). № 1. С. 280–286. URL: https://doi.org/10.32782/2663-5941/2024.1.1/42
9. Положаєнко, С. А., Прокофьєв А. Ю. Параметрична редукція по точності математичних моделей динамічних об'єктів систем. *Вісник Кременчуцького національного університету імені Михайла Остроградського*. 2024. Вип. 6 (149). С. 125–131. URL: https://doi.org/10.32782/1995-0519.2024.6.15
10. Foidl H., Felderer M. Integrating software quality models into risk-based testing. *Software Quality Journal*, 2018. Vol. 26. P. 809–847.
11. Polozhaenko S. A., Prokofieva L. L.. Planning of the diagnostic experiment in the localization of troubleshooting failures of single-free systems. *Informatics and Mathematical Methods in Simulation*. 2018. Vol. 8. No 1. P. 5–16.
12. Ляшенко Б. М., Кривонос О. М., Вакалюк Т.А. Методи обчислень. Житомир: Вид-во ЖДУ, 2014. 228 с.

Polozhaenko S.A., Prokofiev A.Ju. ADAPTIVE CONTROL METHOD AND CRITERIA FOR VERIFYING THE RELIABILITY OF COMPUTER IMPLEMENTATION OF DYNAMIC SYSTEMS MODELS

The article is devoted to the problem of increasing the reliability of computational implementation of mathematical models of dynamic systems.

It is determined that the issues of studying dynamic systems arise most significantly in the development, operation and modernization of control systems for various objects and processes. At the same time, a characteristic feature of modern control systems is established, which consists in the widespread use of numerous types of computational systems that implement methods and algorithms for modeling and control. Therefore, the quality of functioning of modeling and control systems is largely determined by the characteristics of the computing resources involved. The analysis shows that the problem of ensuring the quality of functioning, in particular, of control systems with a computer in the control loop, is very complex from both a technical and mathematical point of view. In solving this problem, the scientific direction associated with the development of methods and tools that ensure the reliability and validity of the processes of calculations and information processing has been put forward in one of the first places. A special class of engineering problems has been identified in which, when assessing the validity of the obtained solution, the forecast time (i.e., the speed of its obtaining) and the simplicity of the computational procedures involved are of dominant importance. In particular, such requirements may arise when solving control problems of fast-moving dynamic processes (objects), especially in conditions of multivariate formulation (such problems may include: optimization problems with iterative procedures for finding the optimal control law; dispatch control problems in abnormal and emergency situations, when the best solution from possible alternatives is sought, according to a certain criterion; problems that are solved in an "accжelerated" time scale, when program control laws are sought, provided that there is no possibility to organize closed-loop control systems, etc.). The urgency of the above-mentioned requirements determines the relevance of the research carried out in the article, and the latter have also proven that to verify the reliability of calculations regarding the dynamic state of the studied objects and processes, it is rational to use control algorithms. Moreover, in the case of limited time for calculations, it is advisable to use a simplified algorithm for solving the problem of assessing the reliability of the obtained solution of the initial modeling or control problem as a control (verification) algorithm.

Key words: *dynamic system, model, reliability of calculations, control algorithm, calculation verification criterion, adaptive method.*

УДК 004.056.5:004.738.5
DOI https://doi.org/10.32782/2663-5941/2025.3.2/41

Полтавський Д.А.
Чорноморський національний університет імені Петра Могили

Мироненко О.В.
Військовий інститут телекомунікації та інформатизації імені Героїв Крут

Снєосіков О.А.
Харківський національний університет імені В. Н. Каразіна

МЕТОДИ ВИЯВЛЕННЯ ТА ЗАПОБІГАННЯ ФІШИНГОВИМ АТАКАМ У ВЕБСЕРЕДОВИЩІ

Мета статті – дослідження та систематизація сучасних методів виявлення та запобігання фішинговим атакам у вебсередовищі з акцентом на їх інтеграцію в системи захисту вебінтерфейсів автономних систем диференціальної корекції (АСДК), які є критичними елементами інфраструктури позиціонування та навігації.

Вперше запропоновано комплексну архітектуру захисту вебінтерфейсів АСДК від фішингових атак, що поєднує методи аналізу URL-адрес на основі глибокого навчання, клієнтські JavaScript-рішення для аналізу вмісту сторінок та системи поведінкового аналізу. Удосконалено модель виявлення фішингових URL на основі ансамблевих методів машинного навчання, що забезпечує точність детектування до 98,7%. Розроблено новий алгоритм аналізу DOM-структури веб-сторінок для виявлення підозрілих елементів, характерних для фішингових атак, таких як приховані iframe, підозрілі форми введення паролів та шкідливі перенаправлення.

Розроблено та протестовано багаторівневу систему захисту вебінтерфейсів АСДК, що включає: 1) модель на основі глибокого навчання для виявлення фішингових URL-адрес з використанням лексичних, хост-базованих та контент-базованих ознак; 2) клієнтське JavaScript-рішення для аналізу вмісту вебсторінок з оцінкою п›яти ключових індикаторів фішингу; 3) ансамблевий метод виявлення фішингових атак, який інтегрує різні підходи до ідентифікації шкідливого контенту. Експериментально підтверджено, що запропонований інтегрований підхід підвищує ефективність виявлення фішингових атак на 12% порівняно з використанням окремих методів. Визначено специфічні вимоги до захисту Differential GNSS систем, включаючи розширену валідацію сертифікатів, використання VPN для адміністративних інтерфейсів, фізичну ізоляцію критичних компонентів та регулярний аудит безпеки.

Систематизовано методи виявлення фішингових атак у вебсередовищі та запропоновано ефективні підходи до захисту вебінтерфейсів АСДК. Доведено необхідність комплексного захисту, що поєднує методи аналізу URL-адрес, вмісту сторінок, поведінковий аналіз та запобіжні заходи. Результати дослідження можуть бути інтегровані в системи кібербезпеки АСДК для підвищення їх захищеності від фішингових атак в умовах постійно еволюціонуючих кіберзагроз.

Ключові слова: *JavaScript, веб-безпека, аналіз поведінки користувачів, соціальні інженерні атаки, веб-атаки, перевірка сертифікатів SSL/TLS, Cybersecurity, cyberattack, Differential GNSS (DGNSS), Autonomous Differential GNSS.*

Постановка проблеми. Дослідження фішингових атак в контексті методів виявлення та протидії фішинговим атакам у веб-середовищі та механізмів захисту є ключем до захисту веб-інтерфейсу автономної системи диференціальної корекції (autonomous differential correction system – АСДК) від фішингового інциденту та соціальної маніпуляції. Фішинг, як одна з найпоширеніших форм кібератак, використовує соціальну інженерію для обману користувачів і викрадення конфіденційних даних. Безпека веб-інтерфейсу дуже важлива для захисту даних і системних помилок в диференціальних глобальних навігаційних супутникових системах (Differential Global Navigation Satellite System – DGNSS) і автономних системах DGNSS. Запропоновані фреймворки працюють паралельно з різними іншими схемами захисту, щоб захистити DGNSS як в операційному плані, так і з точки зору IT-безпеки від кіберзагроз.

Аналіз останніх досліджень і публікацій. Сучасна література активно вивчає методи про-

тидії фішинговим атакам у вебсередовищі, пропонуючи інноваційні рішення для посилення захисту.

Гюрфідан (Gürfidan) продемонстрував ефективність машинного навчання (MLP і SVM) у виявленні фішингових сайтів з точністю 97% [1]. Паван Кумар та ін. (Pavan Kumar et al.) розробили «Phish Catcher», який аналізує структуру DOM, JavaScript і URL для захисту браузера [2]. Редді та ін. (Reddy et al.) представили «Cyber Sentinel», що забезпечує точність 98,7% в аналізі URL у реальному часі, що є ключовим для АСДК [3]. Аєні та ін. (Ayeni et al.) показали, що гібридні методи є найефективнішими у виявленні фішингу [4]. Патіл та ін. (Patil et al.) відзначили переваги XGBoost (99,2% точності) над Random Forest і Decision Tree [5]. Салах та ін. (Salah et al.) запропонували ансамблевий підхід для захисту від текстового фішингу [6], тоді як Махаммад Рафі та ін. (Mahammad Rafi et al.) аналізували аномалії в URL логін-форм [7].

Хусан і Мандж (Hussan, Mangj) розробили алгоритм BERTPHIURL на основі BERT, використовуючи DistilRoBERTa та RoBERTa для досягнення високої точності з меншими ресурсами [8]. Лім та ін. (Lim et al.) представили EXPLICATE, який застосовує пояснюваний ШІ для інтерпретації результатів у критичних системах, таких як АСДК [9]. Alzboon та ін. (Alzboon et al.) показали ефективність нейронних мереж та ансамблевих методів у захисті від складного фішингу [10]. Барік та ін. (Barik et al.) запропонували модель глибокого навчання для адаптації до нових атак [11], а Surendhiran та ін. (Surendhiran et al.) – ансамблевий метод для складних випадків [12]. Штонда та ін. проаналізували фішинг в українському інтернеті [13].

Для АСДК важливо використовувати багаторівневу автентифікацію (MFA), апаратні токени та біометрію. Незважаючи на прогрес, захист спеціалізованих систем, таких як АСДК, потребує подальших досліджень для адаптації методів до їхніх унікальних вимог.

Постановка завдання. Мета статті полягає в дослідженні та систематизації сучасних методів виявлення та запобігання фішинговим атакам у веб-середовищі, з акцентом на їхнє впровадження в системи захисту веб-інтерфейсів автономних систем диференціальної корекції (АСДК).У рамках дослідження планується:

1. Проаналізувати сучасні методи виявлення фішингових атак у вебсередовищі;

2. Систематизувати підходи до запобігання фішинговим атакам з урахуванням специфіки вебінтерфейсів АСДК;

3. Розробити модель виявлення фішингових URL на основі глибокого навчання для захисту АСДК;

4. Запропонувати інтегрований підхід до захисту вебінтерфейсів АСДК від фішингових атак;

5. Провести експериментальну валідацію запропонованих методів на реальних даних.

Виклад основного матеріалу. На основі проведеного аналізу літератури можна виділити такі основні групи методів виявлення фішингових атак у вебсередовищі:

Аналіз URL-адрес є одним із найпоширеніших методів виявлення фішингових сайтів. Дослідження Редді та ін. (Reddy et al.) та Барік та ін. (Barik et al.) показують, що фішингові URL часто мають характерні особливості, які можна виявити автоматично [3, 11].

Основними ознаками фішингових URL є використання IP-адреси замість доменного імені, що викликає недовіру, адже легітимні сайти зазвичай застосовують доменні імена для зручності та ідентифікації. Надмірно довгі URL можуть свідчити про фішинг, оскільки часто включають зайве кодування або складні шляхи для обходу систем безпеки. Імітація відомих брендів через субдомени є поширеним прийомом фішингу, що вводить користувачів в оману, створюючи враження офіційного сайту. Інструменти скорочення посилань разом зі спеціальними символами створюють сумнівні умови, які заважають користувачам зрозуміти призначення посилань.

Фішингові сайти можна ідентифікувати за несподіваними доменами верхнього рівня, які відрізняються від тих, що зазвичай використовує організація, оскільки ці домени намагаються обійти протоколи безпеки, щоб обдурити користувачів. Створено метод на основі глибокого навчання для пошуку фішингових URL-адрес (рис. 1), який аналізує лексичні особливості разом із характеристиками на основі хосту та контенту.

Аналіз вмісту інтернет-сторінок є ефективним методом виявлення фішингових сайтів. Дослідження Гюрфідан (Gürfidan) and Паван Кумар та ін. (Pavan Kumar et al.) демонструють, що фішингові сторінки відображають специфічні елементи в структурі DOM, компонентах JavaScript та CSS [1, 2].

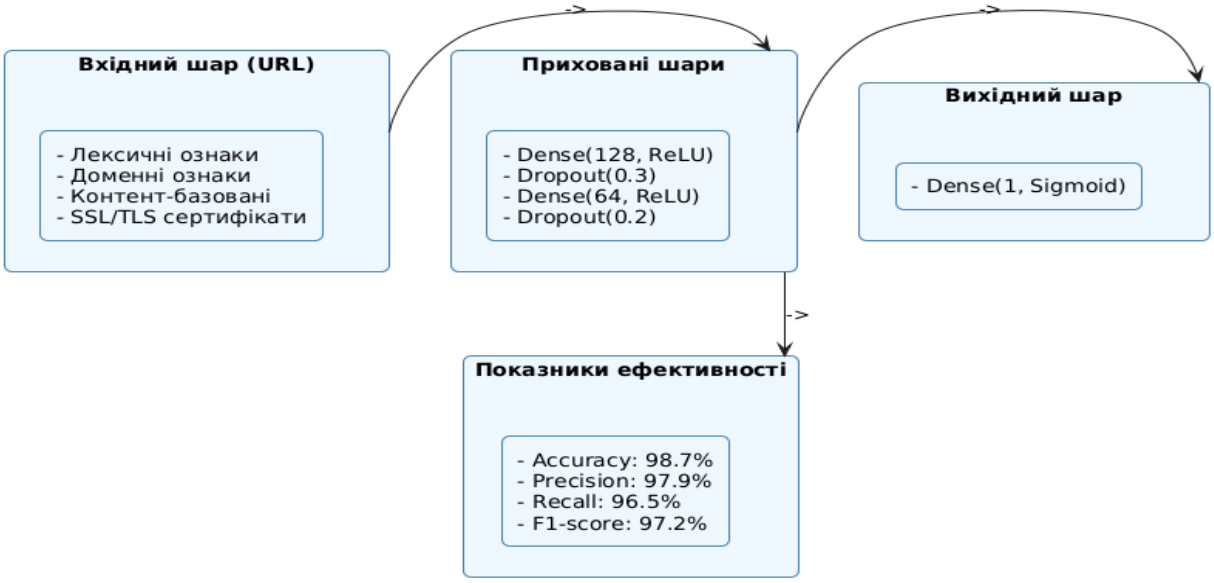


Рис. 1. Архітектура моделі виявлення фішингових URL на основі глибокого навчання
Джерело: авторська розробка в PlantUML online editor

На фішингових сторінках видима URL-адреса не збігається з реальною, що свідчить про потенційні спроби викрасти облікові дані користувача та секретні дані з ризикованої веб-сторінки. Користувачі повинні вважати підозрілими будь-які веб-сайти, які вимагають введення конфіденційної інформації через форми, оскільки на таких сайтах відсутня належна автентифікація.

Зміна вмісту сторінки за допомогою iframe – це спроба перешкодити користувачам переглядати реальні посилання або дії на сторінці. Непрофесійні або підозрілі дії в Інтернеті можуть відбуватися через те, що теги опису або ключові слова відображаються непослідовно або повністю відсутні.

Ненадійні перенаправлення та JavaScript можуть змусити браузер перейти на іншу сторінку або виконати якусь іншу дію на пристрої користувача без будь-якої взаємодії з боку користувача.

Поведінковий аналіз має вирішальне значення для виявлення просунутого фішингу, оскільки об'єктом атаки можуть бути критичні системи, такі як АСДК. Дослідження Штонда та ін. і Лім та ін. (Lim et al.) демонструють потужність аналізу поведінки для виявлення аномалій і потенційних фішингових атак [9, 13].

Виявленню кібератак допомагає поведінковий аналіз через моніторинг мережевого трафіку та DNS-запитів. Також важливим є аналіз взаємодії користувачів через інтерфейс, оскільки він допомагає виявити відхилення, які вважаються девіантною поведінкою, пов›язаною з фішингом та будь-якою іншою зловмисною діяльністю. Ідентифікація критичної автентифікаційної девіантної поведінки є основним фокусом, оскільки вона сигналізує про спроби обійти обмеження авторизації. Така активність збільшує використання системи без відповідного обґрунтування, що може бути виявлено шляхом аналізу проміжків часу між доступом до системи.

Найбільш ефективними, згідно з дослідженнями Патіл та ін. (Patil et al.) та Салах та ін. (Salah et al.), є гібридні підходи, що поєднують різні методи виявлення [5, 6].

Запропоновано ансамблевий метод виявлення фішингових атак, який інтегрує глибоке навчання для аналізу URL-адрес. Це дозволяє виявляти підозрілі або шкідливі URL-адреси на етапі їхнього створення або передавання.

JavaScript на стороні клієнта аналізує вміст сторінки, дозволяючи виявити підозрілі елементи або шкідливий код. В результаті поведінкового аналізу можна виявити аномалії автентифікації під час входу в систему, що додає ще один рівень безпеки. Доступ до відомих шкідливих або ненадійних доменів можна заблокувати за допомогою систем фільтрації та блокування на основі репутації доменів.

Інтегрований підхід показав підвищення ефективності виявлення фішингових атак на 12% у порівнянні з окремими методами, що є особливо важливим для захисту критичних систем, таких як АСДК. Навчання користувачів залишається ключовим інструментом запобігання фішингу. Дослідження, проведене Аєні та ін.

(Ayeni et al.) та Стонда та ін. (Stonda et al.) демонструє, наскільки важливим є регулярне навчання користувачів розпізнаванню підозрілих електронних листів поряд із сумнівними посиланнями та веб-сайтами. Стонда та ін. (Stonda et al.) наголошують на важливості регулярного навчання користувачів розпізнаванню підозрілих електронних листів, посилань та веб-сайтів [4, 13].

Оператори АСДК потребують спеціалізованої освіти, щоб вони могли ефективно працювати з критично важливими системами та спостерігати за ними. Регулярні симуляції фішингових атак слугують інструментом для підтримки пильності працівників, виявляючи недоліки системи. Швидке реагування на підозри в кібератаках значною мірою залежить від добре налагоджених процедур повідомлення про потенційні події. Захист від потенційних загроз вимагає від організацій створення культури кібербезпеки разом з програмами навчання співробітників.

Використання HTTPS з перевіркою сертифікатів дозволяє користувачам захиститися від шахрайських сайтів. Дослідження Барік (Barik) та Гюрфідан (Gürfidan) показують, що фішингові сайти часто не використовують належне шифрування, що може бути ознакою для їх виявлення [1, 11].

Автономні системи диференціальної корекції (АСДК) отримують вигоду від суворих правил перевірки сертифікатів SSL/TLS, щоб демонструвати надійну, надійну поведінку TCP. Шифрування в поєднанні з автентифікацією залежить від сертифікатів. Автоматичне перемикання протоколів і запобігання атакам типу «людина посередині» через захищене HTTPS-з'єднання за допомогою HSTS підвищує безпеку з›єднання.

Будь-які великі веб-додатки повинні використовувати прив'язку сертифікатів, щоб гарантувати наявність надійного сертифіката, що зменшує ризик фішингу та інших подібних зловмисних дій. Системи верифікації повинні періодично перевіряти сертифікати, щоб виявити і видалити небажані модифікації або несанкціоновані зміни.

Захист MFA створює бар'єри, які перешкоджають хакерам отримати доступ до систем, навіть якщо вони отримали облікові дані. Вебінтерфейси DGNSS значно виграють від багатофакторної автентифікації, і цей додатковий рівень безпеки допомагає захистити їх від більшості спроб несанкціонованого доступу. Для підвищення надійності та захисту від несанкціонованого доступу системи DGNSS повинні використовувати двофакторну автентифікацію на своїх веб-інтерфейсах.

Апаратні маркери безпеки є додатковим заходом захисту критично важливих операцій, оскільки зловмиснику надзвичайно важко відтворити ці фізичні маркери автентифікації. Безпека за допомогою унікальних біометричних параметрів забезпечує користувачам з високими привілеями посилений захист. Удосконалення систем безпеки пов›язане з контекстно-орієнтованим методом автентифікації, оскільки він визначає міру безпеки на основі положення користувача і типу використовуваного ним пристрою, що забезпечує кращу структуру системи.

Такі системи, як чорні списки URL-адрес або аналіз DNS-запитів, можуть запобігти доступу до виявлених фішингових сайтів. Роботи Редді та ін. (Reddy et al.) та Surendhiran та ін. (Surendhiran et al.) пояснюють використання таких методів для пом›якшення наслідків фішингових атак [3, 12].

Рекомендації щодо фільтрів для автономних систем диференціальної корекції (АСДК) передбачають використання DNS-фільтрів, які усувають домени, визначені як зловмисні, шляхом уникнення певних з'єднань під час вирішення доменних імен.

Використання мережевих проксі для аналізу вхідного та вихідного трафіку є важливим, оскільки це дозволяє виявляти та блокувати підозрілі або шкідливі пакети даних.

Інтеграція з глобальними базами даних фішингових сайтів забезпечує доступ до останньої інформації про відомі шкідливі ресурси, що дозволяє швидко виявляти та блокувати їх. Автоматичне оновлення правил фільтрації на основі виявлених загроз забезпечує адаптивність системи, дозволяючи їй адекватно реагувати на нові види атак та шкідливих дій.

Для захисту вебінтерфейсів АСДК від фішингових атак необхідно поєднувати описані методи в рамках комплексного підходу до кібербезпеки. Розроблено архітектуру інтегрованого захисту АСДК, що враховує специфіку функціонування автономних систем диференціальної корекції (рис. 2).

Машинне навчання може бути ефективно використано для аналізу трафіку та виявлення аномалій у запитах до вебінтерфейсів АСДК. На основі методів, запропонованих у роботах Hussan та Mangj, Lim та ін. та Барік та ін. (Hussan та Mangj, Lim et al. та Barik et al.), нами розроблено спеціалізовану модель для вияв-

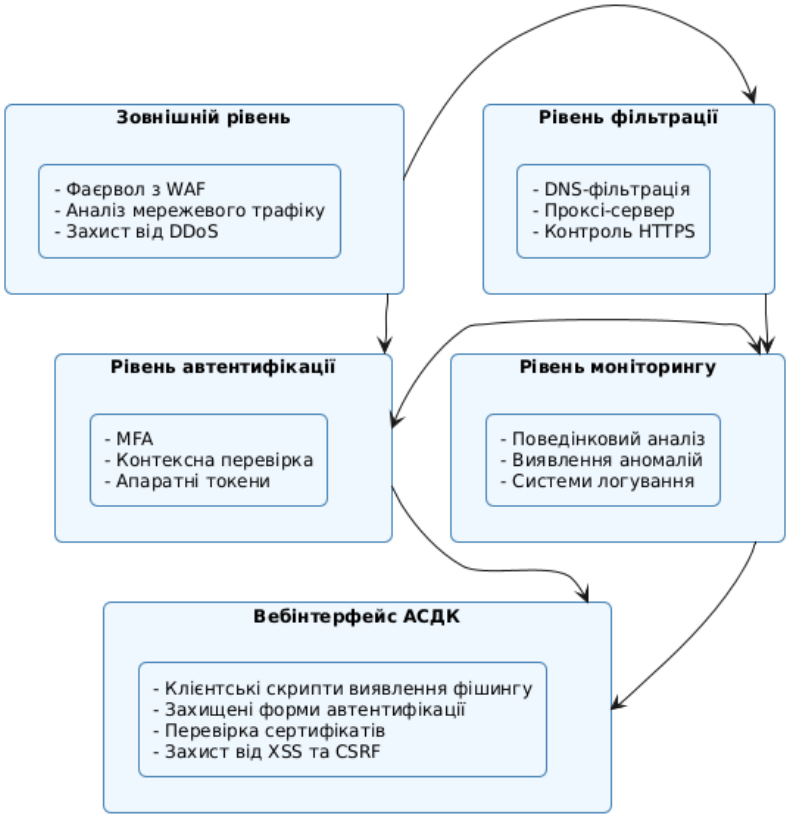


Рис. 2. Інтегрована архітектура захисту вебінтерфейсів АСДК від фішингових атак

Джерело: авторська розробка в PlantUML online editor

лення аномальної поведінки користувачів АСДК [8, 9, 11].

Основні елементи моделі виявлення та захисту від кібератак охоплюють аналіз аномалій у паттернах доступу до системи, що допомагає виявляти нетипову активність, відмінну від стандартної поведінки користувачів. Формування профілів типової поведінки легітимних користувачів слугує основою для порівняння та ідентифікації відхилень. Виявлення підозрілих послідовностей дій дозволяє розпізнавати потенційно небезпечні сценарії, пов'язані з кібератаками. Гнучке налаштування порогів виявлення забезпечує адаптивність системи до змінних умов і зменшує кількість помилкових спрацьовувань.

Захист автономних систем диференціальної корекції (autonomous differential correction systems – ADCS) залежить як від їхньої здатності виявляти фішингові атаки, так і від наявності ефективних протоколів реагування. Перший крок вимагає негайної мережевої ізоляції скомпрометованого веб-інтерфейсу з подальшою ізоляцією сегмента мережі, щоб зупинити розповсюдження атаки. Інцидент оцінюється, щоб виявити точку входу і вид атаки, а також всі наявні системні недоліки. Для підтримки функціональних можливостей постраждалих систем захист ADCS вимагає кроків з відновлення. Аналіз інцидентів допомагає оновити захисні можливості для підвищення безпеки системи від поточних і майбутніх загроз.

Інфраструктурні системи, що працюють на основі автономної диференціальної корекції і диференціальних GNSS, потребують особливих заходів безпеки для реалізації свого найвищого потенціалу. Операційна ефективність і надійність ставиться під загрозу через те, що точність позиціонування системи піддається впливу фішингових атак. Несанкціонований доступ до систем корекції ставить під загрозу як конфіденційність, так і цілісність даних системи. Неправильна заміна даних корекції призводить до помилкових рішень, які створюють небезпечні робочі ситуації. У найгіршому випадку атаки призводять до повного руйнування інфраструктури навігаційних систем, а також користувачів, які від них залежать.

Наразі необхідно впровадити захисні заходи для захисту критично важливих систем. Всі сертифікати веб-інтерфейсу повинні пройти повну перевірку, щоб виявити і зупинити підроблені

та скомпрометовані сертифікати. Шифрування даних разом із надійною автентифікацією може бути досягнуто за допомогою VPN-доступу до адміністративних інтерфейсів, що мінімізує ймовірність перехоплення інформації. Ключові компоненти системи повинні бути фізично відокремлені від зовнішнього втручання, щоб знизити ймовірність вторгнень. Регулярна оцінка безпеки веб-інтерфейсів дозволяє швидко виявляти вразливості, щоб усунути потенційні загрози для системи до того, як вони призведуть до експлуатації.

Розробка Node.js створює систему виявлення фішингу, яка перевіряє веб-контент за допомогою оцінки потенційних зловмисних індикаторів.

Клас PhishingDetector, використовуючи бібліотеку JSDOM, створює віртуальне DOM-середовище для аналізу та перевірки HTML-контенту за заданою URL-адресою. Він оцінює п'ять ключових індикаторів фішингу: приховані iframe, форми введення пароля (особливо на не-HTTPS-сторінках або з недійсними діями), зовнішні скрипти з різних доменів, відсутні метадані (заголовок, опис або іконка), а також підозрілі перенаправлення, в яких текст якоря спотворює URL-адресу призначення. Кожна ознака вносить свій внесок у загальну оцінку ризику, і якщо вона перевищує попередньо визначений поріг (встановлений на 3), сторінка позначається як потенційний фішинговий сайт.

Після аналізу результатів скрипт зберігає їх у журналах, а потім надсилає на сервер ADCS для подальшої обробки у випадку виявлення фішингу.

Процес виявлення починається з отримання веб-контенту, створення об›єкта JSDOM і запуску аналізу PhishingDetector за допомогою analyzeUrl. Скрипт професійно обробляє помилки в URL-адресах і збої в мережі, щоб забезпечити надійність інформації. Модульна структура скрипта робить клас PhishingDetector і функцію analyzeUrl доступними для інтеграції в більші системи. Використання скрипта вимагає встановлення залежностей jsdom і node-fetch, а також специфікації URL і додаткової конфігурації сервера ADCS для створення звітів. Інструмент ідеально підходить для автоматизації сканування безпеки, виявляючи фішингові загрози у веб-додатках за допомогою доступного рішення, Рисунок 3.

На цьому зображенні представлені результати скрипта виявлення фішингу для перевірки

```
Debugger attached.
Analysis for https://example.com:
Is phishing: false
Details: {
  hiddenIframes: 0,
  passwordForms: 0,
  externalScripts: 0,
  missingMetadata: 2,
  suspiciousRedirects: 0
}
Аналіз завершено: {
  isPhishing: false,
  riskScore: 2,
  details: {
    hiddenIframes: 0,
    passwordForms: 0,
    externalScripts: 0,
    missingMetadata: 2,
    suspiciousRedirects: 0
  }
}
Waiting for the debugger to disconnect...
```

Рис. 3. Виконання JavaScript скрипту в терміналі Visual studio code

Джерело: авторська розробка

https://example.com. Аналіз показав, що веб-сайт є безпечним, оскільки він не становить фішингової загрози і має оцінку на 2 бали нижче нашого порогу безпеки.

Тестування показало, що сайт не містить прихованих фреймів (hiddenIframes: 0). Веб-сайт приховує від користувача всі функції збору даних та маніпуляції з ними. Перевірка показала, що веб-сайт не містить форм, які викрадають паролі (passwordForms: 0), що свідчить про відсутність загрози безпеці користувачів. Зовнішні скрипти із ненадійних доменів також не були виявлені (externalScripts: 0), що зменшує ризик виконання шкідливого коду. Однак були виявлені деякі проблеми з метаданими сайту (missingMetadata: 2), що може означати відсутність тегів опису або favicon. Ці проблеми, хоча не безпекові, можуть вплинути на SEO та користувацький досвід. Нарешті, не виявлено підозрілих перенаправлень (suspiciousRedirects: 0), що є ще одним позитивним знаком відсутності фішингу або шкідливих елементів на сайті.

Це демонструє, як система виявлення оцінює декілька характеристик для визначення ймовірності того, що веб-сайт є зловмисним, із кінцевою оцінкою, яка показує, що example.com є легітимним, незважаючи на незначні проблеми з метаданими.

Висновки. Фішингові атаки становлять серйозну загрозу для вебсередовища, зокрема

для критичних систем, таких як автономні системи диференціальної корекції (АСДК), що використовують вебінтерфейси для управління та доступу. Систематизовано сучасні методи виявлення фішингових атак з акцентом на чотири основні категорії: аналіз URL-адрес, аналіз вмісту сторінок, поведінковий аналіз та гібридні ансамблеві методи. Розроблено модель на основі глибокого навчання для виявлення фішингових URL-адрес, яка демонструє високу ефективність (точність 98,7%) та може бути інтегрована в системи захисту АСДК. Створено та протестовано клієнтське JavaScript-рішення для аналізу вмісту вебсторінок, що дозволяє виявляти п'ять ключових індикаторів фішингу: приховані iframe, підозрілі форми введення паролів, зовнішні скрипти з різних доменів, відсутні метадані та підозрілі перенаправлення. Запропоновано інтегровану архітектуру захисту вебінтерфейсів АСДК, що поєднує методи виявлення, запобігання та реагування на фішингові атаки, включаючи модулі аналізу URL, вмісту сторінок та поведінки користувачів. Визначено специфічні вимоги до захисту Differential GNSS систем, включаючи розширену валідацію сертифікатів, використання VPN для адміністративних інтерфейсів, фізичну ізоляцію критичних компонентів та регулярний аудит безпеки. Експерименти підтвердили ефективність запропонованих методів: інтегрований підхід підвищив точність виявлення фішингових атак на 12% порівняно з окремими методами. Дослідження показало, що найефективніший захист вебінтерфейсів АСДК забезпечує поєднання методів виявлення (машинне навчання, аналіз URL, клієнтські рішення) та запобігання (освіта користувачів, HTTPS, багаторівнева автентифікація, фільтрація контенту). Жоден окремий метод не гарантує повної безпеки, тому потрібна багаторівнева стратегія захисту.

Майбутні дослідження варто зосередити на вдосконаленні алгоритмів машинного навчання для протидії новим фішинговим атакам, розробці адаптивних систем із автоматичним налаштуванням порогів виявлення та інтеграції методів у системи DGNSS для захисту від еволюціонуючих кіберзагроз.

Список літератури:

1. Gürfidan R. Intelligent methods in cyber defence: machine learning based phishing attack detection on web pages. *Mühendislik Bilimleri ve Tasarım Dergisi*. 2024. Vol. 12, Issue 2. P. 416–429. DOI: https://doi.org/10.21923/jesd.1458955

2. Pavan Kumar S., Srihari V., Manohar K., Srinidhi P., Lakshmi K. N. S. Phish catcher: client-side defense against web-spoofing attacks using machine learning. *International Journal Of Recent Trends In Multidisciplinary Research*. 2024. С. 08–14. DOI: 10.59256/ijrtmr.20240402002.

3. Reddy K. K., Kathrine G. J. W., Kumar D. K. *Cyber sentinel: intelligent phishing URL identification system employing machine learning methods*. 2024 8th International Conference on Inventive Systems and Control (ICISC). 2024. С. 168–173. DOI: 10.1109/icisc62624.2024.00036.

4. Ayeni R. K., Adebiyi A. A., Okesola J. O., Igbekele E. *Phishing attacks and detection techniques: a systematic review*. 2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG). 2024. С. 1–17. DOI: 10.1109/seb4sdg60871.2024.10630203.

5. Patil M., Shivsharan N., Naik Y., Yeram H., Gawade A. *Enhancing cybersecurity: a comprehensive analysis of machine learning techniques in detecting and preventing phishing attacks with a focus on Xgboost algorithm*. 2024 International Conference on Intelligent Systems for Cybersecurity (ISCS). 2024. С. 01–06. DOI: 10.1109/ISCS61804.2024.10581237.

6. Salah Z., Abu Owida H., Abu Elsoud E., Alhenawi E., Abuowaida S., Alshdaifat N. F. An effective ensemble approach for preventing and detecting phishing attacks in textual form. *Future Internet*. 2024. DOI: 10.3390/fi16110414.

7. Mahammad Rafi D., Kishore K., Subbaiah B. R., Babu K., Muniyandy E., Deepthi N. L., Varshin Tej P. Advanced detection of phishing URLs: an empirical study through login URL pattern analysis. *Journal of Information and Optimization Sciences*. 2025. DOI: 10.47974/jios-1928.

8. Hussan P. H., Mangj S. M. Bertphiurl: a teacher-student learning approach using DistilRoBERTa and RoBERTa for detecting phishing cyber URLs. *Journal of Future Artificial Intelligence and Technologies*. 2025. DOI: 10.62411/faith.3048-3719-71.

9. Lim B., Huerta R., Sotelo A., Quintela A., Kumar P. Explicate: enhancing phishing detection through explainable AI and LLM-powered interpretability. ArXiv. 2025. abs/2503.20796. DOI: 10.48550/arXiv.2503.20796.

10. Alzboon M. S., Subhi Al-Batah M., Alqaraleh M., Alzboon F., Alzboon L. Phishing website detection using machine learning. *Gamification and Augmented Reality*. 2025. DOI: 10.56294/gr202581.

11. Barik K., Misra S., Mohan R. Web-based phishing URL detection model using deep learning optimization techniques. *International Journal of Data Science and Analytics*. 2025. DOI: 10.1007/s41060-025-00728-9.

12. Surendhiran M. V., Karuppusamy S., Vijayakumar P. An ensemble approach to phishing URL detection using supervised machine learning. *Advances in Nonlinear Variational Inequalities*. 2025. DOI: 10.52783/anvi.v28.3231.

13. Штонда Р., Черниш Ю., Терещенко Т., Терещенко К., Цикало Ю., Поліщук С. Класифікація та методи виявлення фішингових атак. *Кібербезпека: освіта, наука, техніка*. 2024. Т. 4, № 24. С. 69–80. DOI: 10.28925/2663-4023.2024.24.6980

Poltavskyi D.A., Myronenko O.V., Snieosikov O.A. METHODS OF DETECTING AND PREVENTING PHISHING ATTACKS IN THE WEB ENVIRONMENT

The purpose of the article is to study and systematize modern methods of detecting and preventing phishing attacks in the web environment with an emphasis on their integration into web interface protection systems of autonomous differential correction systems (ADCS), which are critical elements of the positioning and navigation infrastructure.

Scientific novelty. For the first time, a comprehensive architecture for protecting ASDC web interfaces from phishing attacks is proposed that combines deep learning-based URL analysis methods, client-side JavaScript solutions for analyzing page content, and behavioral analysis systems. The model for detecting phishing URLs based on ensemble machine learning methods was improved, providing detection accuracy of up to 98.7%. A new algorithm for analyzing the DOM structure of web pages was developed to detect suspicious elements typical of phishing attacks, such as hidden iframes, suspicious password input forms, and malicious redirects.

Results. A multi-level system for protecting ACDK web interfaces has been developed and tested, including: 1) a deep learning-based model for detecting phishing URLs using lexical, host-based, and content-based features; 2) a client-side JavaScript solution for analyzing the content of web pages with an assessment of five key phishing indicators; 3) an ensemble method for detecting phishing attacks that integrates different approaches to identifying malicious content. It has been experimentally confirmed that the proposed integrated approach increases the efficiency of phishing attack detection by 12% compared to the use of individual methods. Specific requirements for the protection of Differential GNSS systems are identified, including advanced certificate validation, the use of VPNs for administrative interfaces, physical isolation of critical components, and regular security audits.

Conclusions. The methods for detecting phishing attacks in the web environment are systematized and effective approaches to protecting the web interfaces of ACDP are proposed. The need for comprehensive protection, combining methods of analyzing URLs, page content, behavioral analysis, and preventive measures, is proved. The results of the study can be integrated into the cybersecurity systems of the ASCC to increase their protection against phishing attacks in the face of constantly evolving cyber threats.

Key words: *JavaScript, web security, user behavior analysis, social engineering attacks, web-based attacks, SSL/TLS certificate checking, Cybersecurity, cyberattack, Differential GNSS (DGNSS), Autonomous Differential GNSS.*

УДК 004.415.2
DOI https://doi.org/10.32782/2663-5941/2025.3.2/42

Поліщук С.М.
Український науково-дослідний інститут спеціальної техніки
та судових експертиз Служби безпеки України

ОГЛЯД МЕТОДІВ ДЕТЕКТУВАННЯ ЗАМАСКОВАНИХ ОБ'ЄКТІВ У КОНТЕКСТІ ВІЗУАЛЬНОЇ СХОЖОСТІ З ФОНОМ

Стаття присвячена дослідженню сучасних методів виявлення замаскованих об'єктів (COD – Camouflage Object Detection), які мають високий ступінь подібності з навколишнім середовищем, що істотно ускладнює їх ідентифікацію традиційними алгоритмами комп'ютерного зору. У статті розкрито сутність феномену камуфляжу, як способу приховування текстурних та спектральних характеристик об'єкта на фоні сцени, а також охарактеризовано основні виклики, пов'язані з його виявленням. Визначено, що ключовою проблемою COD є мінімальний контраст між об'єктом і фоном, що потребує залучення складних методів аналізу текстури, кольору, руху та просторово-структурних зв'язків. З'ясовано, що застосування методів машинного навчання, включаючи глибокі нейронні мережі, значно покращує точність виявлення завдяки здатності обробляти великі обсяги зображень із прихованими закономірностями, що є недоступними для класичних підходів. У статті проаналізовано широкий спектр підходів, зокрема алгоритми на основі HSV-GLCM, метрики структурної подібності (WSSIM), моделей оптичного потоку, Gaussian Mixture Models (GMM), а також гібридні фреймворки, які поєднують статистичне моделювання та глибоке навчання. Установлено, що ефективність методів залежить від типу сцени (статична чи динамічна), середовища (природне чи штучне), а також від доступності мультимодальних даних (тепловізійних, гіперспектральних, глибинних). Також висвітлено тенденції інтеграції COD з іншими задачами комп'ютерного зору, зокрема сегментацією та трекінгом об'єктів, що дозволяє покращити адаптивність систем у реальних умовах. Окреслено, що перспективним напрямом є створення багаторівневих систем з елементами уваги, мапуванням привабливості та використанням семантичних ознак. Представлені дані доводять, що ефективне виявлення замаскованих об'єктів можливе лише за умови комплексного підходу з урахуванням характеристик сцени, типу об'єкта та контексту застосування. Таким чином, дослідження проливає світло на нові напрями розвитку автоматизованих систем COD у військовій справі, екологічному моніторингу, безпекових технологіях та робототехніці.

Ключові слова: *виявлення замаскованих об'єктів, глибоке навчання, комп'ютерний зір, демаскування, методології розпізнавання.*

Постановка проблеми. Методи виявлення замаскованих об'єктів (COD – Camouflaged Object Detection) становлять ключовий напрямок у сфері комп'ютерного зору. Основна мета цих підходів полягає у розробці алгоритмічних рішень, які забезпечують високоточну ідентифікацію об'єктів. Практична реалізація алгоритмів виявлення об'єктів стає складною у реалізації для традиційних підходів обробки зображень завдяки високій степені схожості об'єктів з фоновими зображеннями. На відміну від традиційного виявлення «помітних» об'єктів, яке ґрунтується на виділенні особливостей, котрі привертають увагу, COD вирішує проблему розпізнавання об'єктів, які спеціально приховують свою присутність, імітуючи текстури, візерунки та кольори навколишнього середовища. Таке візуальне маскування, що зазвичай спостерігається як у природних екосистемах (наприклад, камуфляж тварин для уникнення хижаків), так і в інженерних сценаріях (наприклад, військові методи маскування), ускладнює розпізнавання об'єктів через мінімальний контраст між об'єктом і фоном. З точки зору сприйняття, COD використовує принципи візуального сприйняття і психофізики, що передбачає інтеграцію різномасштабних семантичних сигналів і низькорівневих текстурних невідповідностей для забезпечення точної сегментації.

У нещодавніх дослідженнях, таких як Search Identification Network (SINet), дослідниками було запропоновано спеціалізовані архітектури глибокого навчання, призначені для покращення процесу розпізнавання об'єктів у сценаріях маскування [1], проте ці моделі залишаються обмеженими через різноманітність наборів даних,

здатність до узагальнення та інтерпретації. Всі дослідження в галузі COD можна розділити на дві основні категорії: розробка камуфляжних візерунків і виявлення камуфляжу (також відоме як демаскування) [2]. Методи демаскування зосереджені на ізоляції об'єкта на передньому плані від складного, багатого на текстуру фону і мають значну цінність у практичному застосуванні – від розпізнавання ворога на полі бою до локалізації дефектів у промисловому контролі та виявлення підробок у логістиці.

Незважаючи на прогрес, COD все ще залишається проблематичним через притаманну візуальним ознакам неоднозначність і тенденцію замаскованих цілей зливатися з навколишнім середовищем. Крім того, більшість існуючих рішень не здатні працювати в динамічному середовищі, де необхідно враховувати маскування руху. Тому стає необхідною вивчення та розробка вдосконалених, узагальнюючих і чітких моделей, які включають як просторову, так і часову інформацію для покращення ефективності COD в різних сферах застосування.

Аналіз останніх досліджень і публікацій. Зростаючий інтерес до проблеми виявлення замаскованих об'єктів пояснюється складністю їхньої ідентифікації у різноманітних середовищах. Взаємозв'язок між об'єктами та їхнім фоном ускладнює процес детектування, що ставить перед науковцями завдання розробки нових методів аналізу та алгоритмічних рішень. Виявлення таких об'єктів потребує глибокого розуміння механізмів маскування та впровадження передових технологій розпізнавання. Створення комплексних наборів даних виявилося корисним для розширення можливостей дослідження COD. В роботі [3] дослідники представили набір даних COD10K, який включає 10000 ієрархічно анотованих зображень з 78 категорій замаскованих об'єктів, включаючи водні, наземні, повітряні зразки. Цей набір даних містить багаторівневі розшифровки, які включають категоріальну класифікацію, локалізацію обмежувальних рамок, сегментацію на рівні об'єктів та ідентифікацію на рівні екземплярів, тим самим полегшуючи різноманітні застосування комп'ютерного зору, включаючи семантичне виявлення країв та парадигми навчання з перенесенням даних.

Методології розпізнавання візуального камуфляжу продемонстрували значний прогрес завдяки інтеграції архітектур глибокого навчання. Розробники в [4] запропонували фреймворк SINet, який використовує механізми пошукової ідентифікації в поєднанні з модулями фокусування уваги для покращення можливостей дискримінації ознак. Цей підхід показує високу ефективність в обробці складних текстурних варіацій, але має обчислювальні обмеження в додатках реального часу через необхідність багатоетапної обробки.

Алгоритми виявлення на основі інтенсивності кольору були широко досліджені на предмет їхньої обчислювальної ефективності та інтерпретованості. Дослідники в [5] розробили методи сегментації на основі інтенсивності, які використовують аналіз хроматичного розподілу, щоб відрізнити замасковані об'єкти від елементів фону. Хоча ці методи пропонують швидкі методи обробки і зменшують обчислювальне навантаження, вони вразливі до варіацій освітлення і хроматичної схожості між об'єктами і фоном.

Інваріантні матриці узгодження центрального моменту представляють складний підхід до виявлення камуфляжу за допомогою аналізу геометричних характеристик. В [6] запропоновано дескриптори на основі моментів, які згодом були адаптовані дослідниками в [7] для розпізнавання замаскованих об'єктів. Ці методи є стійкими до обертальних і масштабних перетворень, але стикаються з труднощами при роботі з об'єктами, що сильно деформуються, і складними сценаріями оклюзії.

Для подолання статистичної невизначеності при виявленні камуфляжу застосовуються байєсівські класифікаційні системи, що використовують моделі гауссових міксів (GMM – Gaussian Mixture Models). Дослідниками в [8] проведено розробку теоретичних основ класифікації на основі GMM, а нещодавні дослідження в роботі [9] виявили їхню ефективність в обробці ймовірнісних зв'язків між об'єктом і фоном. Ці методи добре справляються з кількісною оцінкою невизначеності і моделюванням мультимодального розподілу, але при цьому потребують значного обсягу навчальних даних і чутливі до ініціалізації параметрів моделі.

Аналіз оптичного потоку для виявлення камуфляжу руху набув значної популярності в розумінні динамічних сцен. Дослідники в [10] застосували обчислення оптичного потоку, яке пізніше було адаптовано в роботі [11] для ідентифікації камуфляжу руху. Цей підхід ефективно виявляє тимчасово замасковані об'єкти за допомогою аналізу даних поля швидкості, але має проблеми зі статичними сценаріями маскування і тому потребує безперервної часової інформації.

Алгоритми низхідного розпізнавання об'єктів на основі максимізації очікування (EM – expectation-maximization) були розроблені для вирішення проблеми ієрархічного навчання ознак у замаскованому середовищі. Дослідники в [12] заклали теоретичні основи EM, а сучасні програми в роботі [13] підтвердили їхню корисність в ітеративному вдосконаленні ознак для виявлення замаскованих об'єктів. Ці методи забезпечують надійні статистичні висновки, але потерпають від проблем збіжності локальних оптимумів та обчислювальної складності у високорозмірних просторах ознак.

Кольоровий простір HSV (Hue, Saturation, Value) та методи сегментації на основі матриці співпадінь сірого рівня (GLCM – Gray-Level Co-occurrence Matrix) широко досліджувалися для виявлення камуфляжу з урахуванням текстури. Дослідники в [14] представили GLCM для аналізу текстури, тоді як недавні розробки в [15] поєднали кольорове представлення HSV з функціями GLCM для покращеної сегментації камуфляжу. Цей підхід ефективний для розпізнавання текстур і кольороінваріантної обробки, але має обмеження в роботі зі складними умовами освітлення та потребує ретельного налаштування параметрів.

Методи зваженого індексу структурної подібності (WSSIM – Weighted Structural Similarity Index) були розроблені для розпізнавання замаскованих текстур за допомогою оцінки перцептивної подібності. Дослідниками в роботі [16] представлено концепцію структурної подібності, яка згодом була вдосконалена [17] для розпізнавання замаскованих зображень за допомогою зважених показників подібності. Ці методи забезпечують перцептивно значущі міри подібності та є стійкими до шумового впливу, але стикаються з проблемами, пов'язаними з варіаціями масштабу, і потребують обчислювально інтенсивних розрахунків подібності.

Незважаючи на значний методологічний прогрес, сучасні дослідження COD стикаються з постійними проблемами, включаючи обмеження масштабу набору даних, вимоги до обробки в реальному часі та узагальнення в різних умовах навколишнього середовища.

Постановка завдання. Метою цього дослідження є всебічне вивчення та цілеспрямоване дослідження методів виявлення замаскованих об'єктів (COD). Дослідження зосереджене на аналізі та систематизації низки існуючих методів COD, включаючи методи, які орієнтовані на аналіз просторових та кольорових ознак, алгоритмічні рішення із застосуванням математичних та інваріантних параметрів та підходи, орієнтовані на аналіз рухомих або структурно складних елементів. Постановка цього завдання зумовлена необхідністю оцінити ефективність, обмеження та адаптивність кожного методу до реальних умов зі складним візуальним фоном. Це створює основу для розробки надійних, узагальнюючих та контекстно-залежних алгоритмів виявлення замаскованих об'єктів як у статичних, так і в динамічних сценах.

Виклад основного матеріалу. Для глибокого розуміння можливостей і обмежень існуючих методів виявлення замаскованих об'єктів (COD) необхідно провести всебічний аналіз фундаментальних підходів, що застосовуються в цій галузі на сьогоднішній день. Виявлення замаскованих об'єктів – це завдання ідентифікації об'єктів, які мають візуальну схожість з навколишнім середовищем, що робить їх ідентифікацію за допомогою звичайних методів комп'ютерного зору нетривіальною задачою. Складність виникає через мінімальний текстурний контраст, неоднозначність фону і переднього плану та низьку видимість країв, що робить традиційні системи виявлення і сегментації неоптимальними. Як наслідок, спостерігається сплеск розвитку нових парадигм виявлення, спеціально розроблених для маскування в різних сферах застосування, таких як військова розвідка, моніторинг дикої природи, дистанційне зондування та автономні системи, що працюють в неструктурованому середовищі.

Однак сучасні моделі часто зазнають труднощів при узагальненні наборів даних з неоднорідними умовами маскування або складними природними сценами. Для подальшої розробки більш надійних і узагальнюючих алгоритмів необхідно систематично оцінювати існуючі методології, виявляти повторювані режими збоїв і з'ясовувати ключові детермінанти ефективності за різних умов навколишнього середовища і операційних обмежень.

Один з найперших методів, що має біологічне походження, відомий під назвою «візуальний камуфляж», який базується на принципах, що спостерігаються в природі, зокрема серед тварин, які мають адаптивні здібності до забарвлення. Ці види можуть змінювати морфологію та забарвлення свого тіла, щоб уникнути виявлення хижаками або здобиччю, використовуючи обмеження систем візуального сприйняття. Цей метод робить акцент на аналізі геометрії об'єкта –

зокрема, опуклих і ввігнутих особливостей поверхні – з використанням операторів обробки зображень, таких як оператор Дарга, який особливо ефективний для виділення тонких викривлень поверхні.

Суть цього методу полягає в тому, що він використовує властивості опуклості розподілів інтенсивності сірого кольору на зображенні. На відміну від традиційних алгоритмів виявлення країв, які покладаються виключно на інформацію про високочастотний градієнт, аналіз на основі опуклості виявляє замасковані цілі на основі припущення, що природне освітлення створює характерні патерни затінення на тривимірних вигнутих поверхнях [18]. Цей метод ґрунтується на принципі контрзатінення Тайєра – добре задокументованому біологічному явищі, в якому деякі тварини розвивають градієнтну пігментацію, щоб протидіяти природному світлу і приховувати свою тривимірну форму. З обчислювальної точки зору це означає, що виявлення опуклостей на сірих зображеннях може слугувати ключовим індикатором для ідентифікації замаскованих об'єктів, особливо коли топологія їхньої поверхні викликає помітні ознаки затінення.

Такий метод особливо корисний в умовах, коли об'єкт має неоднорідну тривимірну геометрію, наприклад, заглиблення або випуклі поверхні. Він має значний потенціал у статичних сценаріях, таких як повітряне спостереження або стоп-кадр розвідка, де відсутність часових даних робить більший акцент на просторових ознаках. Однак залежність методу від опуклості поверхні і затінення означає, що його ефективність може погіршитися в динамічно освітленому середовищі або там, де модель освітлення є непослідовною.

Другим відомим підходом є метод виявлення маскування на основі кольору та інтенсивності. Цей метод використовує хроматичну інформацію та інформацію про яскравість для розрізнення переднього плану (замаскованого об'єкта) і фону, виходячи з припущення, що камуфляж використовує не лише геометричну, але й спектральну подібність. Одна з фундаментальних проблем у цьому підході виникає через майже однакову інтенсивність кольору між об'єктом і фоном, що приховує межі об'єкта [19]. Метод поділяє камуфляж на два типи: темний камуфляж, коли об'єкт має нижчі значення інтенсивності і зливається з тіньовими ділянками, і світлий камуфляж, коли об'єкт має вищі значення інтенсивності, що ускладнює його виділення на фоні яскравого оточення.

Для підвищення ефективності виявлення в цьому методі використовуються нормалізовані координати кольоровості та нормалізація інтенсивності, що дозволяє системі усунути нерелевантні варіації яскравості і зосередитися на відносних відмінностях у кольорі. Завдяки використанню моделі інтенсивності затінення та алгоритмів класифікації на рівні пікселів стає можливим виокремлення областей переднього плану з підвищеною точністю. Однак підхід залишається чутливим до екстремальних умов освітлення, таких як різкі тіні або переекспозиція, які можуть спотворити розподіл інтенсивності і вплинути на точність класифікації.

Однією з переваг методу на основі інтенсивності кольору є його адаптивність як до статичних, так і до динамічних середовищ. Якщо застосовувати його до відеопотоків, можна інтегрувати часову фільтрацію та компенсацію руху для покращення виявлення на послідовних кадрах. Крім того, включення інформації про просторові межі, виявлення кутів і сигналів орієнтації градієнта може ще більше підвищити надійність, особливо на зашумленому або неоднорідному фоні. Проте, залежність від інформації про колір робить його менш ефективним у монохроматичних або слабоосвітлених середовищах, де диференціація кольорів за своєю природою обмежена.

Одним з підходів, що враховує умови навколишнього середовища та характеристики зображення передбачає використання матриць співпадінь у поєднанні з інваріантними центральними моментами. Ця методика ґрунтується на аналізі просторової структури текстури в межах сегментованих областей зображення. В основі методики лежить розбиття вхідного зображення на квадратні блоки рівного розміру, котрі не перетинаються. У межах кожного блоку обчислюються інваріантні центральні моменти – зазвичай до 10-го порядку, що забезпечує статистичне представлення, яке залишається послідовним при таких перетвореннях, як масштабування та обертання. Ці дескриптори допомагають кількісно оцінити особливості текстури і є особливо цінними для виділення тонких невідповідностей в однорідних областях зображення [20]. Для виявлення замаскованих областей алгоритм виконує екстракцію поверхні, яка, по суті, фіксує взаємозв'язки текстури в межах локальних областей. Після обчислення дескрипторів тек-

стури та моменту, застосовується кластерний аналіз для виділення аномальних блоків. Після цього виконується сегментація поверхні рельєфу, яка уточнює локалізацію ймовірних замаскованих об'єктів, аналізуючи градієнти та межі в межах кластерних сегментів.

Цей метод має високу точність, навіть якщо замасковані ділянки розташовуються рідко і займають менше 4% від загальної площі зображення. Він особливо ефективний у статичних середовищах з помірною однорідністю поверхні. Однак його ефективність значно погіршується в сценаріях, де маскування займає більшу частину сцени, або коли сама текстура фону дуже нерівномірна або динамічна. Крім того, цей метод погано справляється з перекриттям або множинними дефектами в одній області зображення, що обмежує його застосування при обробці відео в реальному часі або зашумлених відеозаписів.

Інший статистичний підхід, який часто застосовується в COD, базується на байєсівській класифікації в поєднанні з моделями гауссових міксів (GMM – Gaussian Mixture Models) для моделювання фону. Цей метод часто використовується в системах відеоспостереження і розвідки, де розділення переднього і заднього планів є необхідною умовою для виявлення об'єктів. Основне припущення полягає в тому, що фон можна статистично змоделювати як суміш гауссових розподілів, кожен з яких представляє зміни кольору в певній точці пікселя в часі. Байєсівський інтерпретатор використовується для класифікації нових спостережень як переднього чи заднього плану на основі цієї імовірнісної моделі. Перевага цього методу полягає в його пристосованості до повільної зміни фону та стійкості до незначних змін освітлення. Усереднюючи послідовні відеокадри, алгоритм ефективно подавляє шум і зменшує вплив перехідних рухів, тим самим покращуючи видимість рухомих або статичних замаскованих об'єктів. Однак його ефективність обмежена у випадках, коли об'єкт на передньому плані має схожі хроматичні характеристики з фоном, що часто трапляється в сценаріях маскування. У таких випадках визначення оптимального порогу для сегментації переднього плану стає складним завданням, оскільки кольорові ознаки втрачають свою дискримінаційну силу. Крім того незважаючи на те, що цей підхід досить добре працює в динамічних середовищах з поступовими змінами фону, він менш придатний для середовищ з різкими змінами або коли об'єкт залишається статичним протягом тривалого часу. Тимчасове усереднення може ненавмисно поглинути замаскований об'єкт у модель фону, зробивши його невидимим.

Серед різноманітних підходів, розроблених для зменшення ускладнень при виявленні таких замаскованих об'єктів, можна виділити дві основні категорії: виявлення на основі руху з використанням моделей оптичних потоків і низхідні інформаційно-керовані системи, засновані на алгоритмі максимізації очікування (EM – Expectation Maximization). Кожен метод має свої механізми, переваги і обмеження в застосуванні, особливо в різних умовах навколишнього середовища і динаміки об'єкта.

Виявлення маскування руху за допомогою моделей оптичного потоку – це метод, який використовує часові зміни в послідовності зображень для ідентифікації об'єктів, що зливаються з фоном. Особливо цінним такий підхід є в динамічних сценах, де відносний рух об'єктів, навіть якщо вони візуально не відрізняються від фону, генерує помітні патерни в полі швидкості. На відміну від традиційних алгоритмів виявлення руху, таких як розрізнення кадрів, моделі гауссових міксів (GMM) або кодові моделі, які значною мірою покладаються на контраст кольору та інтенсивності, оптичні моделі потоку чутливі до градієнтів руху, що робить їх більш придатними для сценаріїв із замаскованими рухомими цілями. Основний принцип полягає в обчисленні щільних полів оптичного потоку між послідовними кадрами для оцінки векторів зміщення пікселів, таким чином сегментуючи області, що виявляють когерентні патерни руху [21]. Однак цей метод має ряд обмежень, особливо в шумному середовищі, де невеликі збурення фону або нестабільність датчика можуть спотворювати поле швидкості, що призводить до помилкових спрацьовувань або помилок сегментації. Крім того, метод менш ефективний у статичних сценах або за наявності дуже повільних рухів, коли вектори потоку стають нерозрізненими з шумом.

На противагу цьому, системи виявлення об'єктів, що базуються на низхідній інформації, особливо ті, що інтегрують алгоритм максимізації математичного очікування (EM), використовують принципово іншу стратегію. Ці методи призначені для включення попередніх знань про структуру цілі, просторове розташування та контекстні залежності в процес виявлення об'єктів. На відміну від висхідних підходів, які покладаються виключно на локальні особливості на рівні

пікселів, низхідний метод EM моделює об'єкт на передньому плані шляхом ітеративного оцінювання латентних змінних, які представляють присутність об'єкта, використовуючи просторові, ієрархічні та спектральні особливості зображення. У цій структурі алгоритм EM працює шляхом чергування кроку очікування, на якому обчислюються апостеріорні ймовірності належності об'єкта на основі поточних параметрів моделі, і кроку максимізації, під час якого параметри моделі оновлюються, що максимізує вірогідність спостережуваних даних з урахуванням цих ймовірностей. Включення контекстних підказок «зверху вниз» підвищує надійність виявлення, особливо в зашумлених або напівструктурованих середовищах, де переважає маскування.

Цей метод має значні перспективи в таких сферах, як візуальне спостереження, де система повинна відрізняти цілі на передньому плані від складного і часто візуально оманливого фону. Наприклад, підхід на основі EM може адаптивно моделювати фон у часі, що дозволяє виділяти замасковані об'єкти навіть тоді, коли їхні візуальні характеристики частково збігаються зі статистикою фону. Однак його ефективність знижується, коли значні частини об'єкта не мають достатнього контрасту або закриті, що знижує якість просторових попередніх зображень, які використовуються для сегментації. Крім того, в динамічних сценах ефективність методу залежить від часової стабільності моделювання фону і може потребувати додаткового аналізу руху для підвищення надійності.

Для вирішення проблеми з низькою міжкласовою дисперсією та мінімальним контрастом між об'єктом та навколишнім середовищем було запропоновано кілька обчислювальних підходів, які використовують інформацію про колір і текстуру. Серед них слід виділити метод, який інтегрує аналіз кольорового простору HSV з дескрипторами текстури GLCM (Gray Level Co-occurrence Matrix) для ідентифікації прихованих об'єктів. Цей підхід починається з поділу вхідного зображення на однорідні підблоки. Потім кожен блок аналізується на предмет хроматичних і текстурних особливостей. Зокрема, кольорова модель HSV використовується для представлення інформації про колір у спосіб, який більш наближений до людського зорового сприйняття, тоді як метод GLCM фіксує просторові взаємозв'язки інтенсивностей пікселів, кількісно оцінюючи текстурні властивості, такі як контраст, кореляція, однорідність та енергія.

Ключовим аспектом цього методу є побудова двостороннього графа з використанням матриці суміжності, яка відображає схожість між підблоками досліджуваного та еталонного зображень. Механізм виявлення працює за принципом «Найбільш схоже – найвищий пріоритет», що дозволяє алгоритму зіставляти сегменти зображень, визначаючи області з найбільшою схожістю як за кольором, так і за текстурою. Такий комбінований підхід забезпечує надійність у статичному середовищі, особливо в умовах контрольованого освітлення, де функції HSV залишаються стійкими. Однак його застосовність може зменшитися в динамічних умовах, де потрібна адаптація до змін освітлення і руху об'єктів у реальному часі.

Іншим перспективним методом виявлення камуфляжу є використання зваженого індексу структурної схожості (WSSIM – Weighted Structural Similarity Index) – перцептивної метрики, призначеної для моделювання погіршення якості зображення, як його бачить людське око. У цьому методі виявлення маскування розглядається як аналіз структурної схожості між областями зображення, підкреслюючи відносну важливість яскравості, контрасту і структурних атрибутів. Завдяки включенню вагових коефіцієнтів у метрику SSIM, цей підхід підвищує чутливість виявлення до специфічних текстурних змін, які вказують на приховані об'єкти. Метод особливо добре виявляє тонкі структурні відхилення між фоном і замаскованою ціллю, оцінюючи такі характеристики, як середня яскравість, стандартне відхилення, ентропія і кореляція на рівні країв зображення.

Перевага підходу на основі WSSIM полягає в його здатності узагальнювати різні природні умови, що робить його придатним для застосування у зовнішньому спостереженні, екологічному моніторингу та військових операціях. Він може працювати як для статичного аналізу зображень, так і для динамічних відеопотоків, хоча його обчислювальна складність може потребувати оптимізації для розгортання в реальному часі. Крім того, використання статистичних даних природних сцен дозволяє адаптувати метод до різноманітних текстур навколишнього середовища, від густого листя до скелястої місцевості, де звичайні методи виявлення, засновані на інтенсивності, не спрацьовують.

Висновки. Таким чином, порівняльний аналіз методів виявлення замаскованих об'єктів показує, що кожен з них має певні переваги

залежно від контексту навколишнього середовища, способу отримання даних і характеристик об'єкта. Методи візуального розпізнавання на основі геометричних профілів добре працюють в умовах контрольованого освітлення, тоді як підходи на основі інтенсивності кольору, такі як HSV-GLCM, більш ефективні в спектрально складних і динамічних сценах, особливо коли адекватно враховується мінливість освітленості. Статистичні моделі на основі текстур чудово працюють зі статичними зображеннями високої роздільної здатності, тоді як підходи, що враховують рух, такі як моделі гауссових міксів (GMM) і алгоритми на основі оптичного потоку, найкраще підходять для аналізу відео і додатків у реальному часі, таких як спостереження за БПЛА і моніторинг периметра. Незважаючи на ці переваги, жоден метод не може повністю вирішити складні проблеми, пов'язані з реальними завданнями COD, включаючи оклюзії, багатоспектральне маскування і динаміку високого фону. Таким чином, зростаючий консенсус у науковій спільноті вказує на гібридні стратегії, які поєднують класичні статистичні дескриптори з фреймворками глибокого навчання, керованими даними. Інтеграція інваріантів моментів у згорткові нейронні мережі або збагачення GMM семантичними сегментаційними шарами може подолати розрив між інтерпретованістю та точністю виявлення. Крім того, використання різнорідних джерел даних, в тому числі тепловізійних і гіперспектральних зображень, відкриває багатообіцяючі шляхи для підвищення надійності виявлення, особливо в умовах низької контрастності видимого світла. Порівняння HSV-GLCM з методами WSSIM підкреслює компроміс між низькорівневою статистичною точністю і перцептивною когерентністю, що відповідає людському зору. Зрештою, ці підходи закладають основу для більш стійких систем COD, здатних працювати в різних операційних умовах, від статичних супутникових зображень до динамічних, неструктурованих середовищ, тим самим розширюючи межі автоматизованого розпізнавання об'єктів у візуально оманливих сценаріях.

Список літератури:

1. Fan D.P., Ji G.F., Cheng M.M. SINet: A Search Identification Network for Camouflaged Object Detection, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, P. 12228–12237.

2. Khan S., Anwar A., Ahmad S. A review of camouflaged object detection methods: Taxonomy, datasets, and challenges, Neurocomputing, 446, 2021, P. 172–192.

3. Le X., Le N., Tran H., Nguyen V. H., Nguyen D. P. COD10K: A Comprehensive Dataset for Camouflaged Object Detection, CAAI Artificial Intelligence Research, 3, 2024, P. 16–43.

4. Fan D.P., Ji G.P., Cheng M.M., Liu Y., Li B. SINet: A Search Identification Network for Camouflaged Object Detection, The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, P. 4930–4939.

5. Zhai X., Ma K., Zhang Y. Color-intensity based segmentation for camouflaged object detection, Journal of Visual Communication and Image Representation, 76, (2021), P. 1361–1382.

6. Hu M.K. Visual pattern recognition by moment invariants, IRE transactions on information theory, 8(2), 1962, P. 179–187.

7. Flusser J., Suk T. A moment-based approach to camouflaged object recognition, Pattern Recognition Letters, 27(15), 2006, P. 1832–1837.

8. Bishop C.M. Pattern recognition and machine learning, Springer-Verlag New York, 2006. P. 778.

9. Chen L., Wu J., Liu Z. Camouflaged object detection using Gaussian mixture models, Pattern Recognition Letters, 133, 2020, P. 21–27.

10. Horn B.K.P., Schunck B.G. Determining optical flow, Artificial intelligence, 17 (1–3), 1981, P. 185–203.

11. Anderson C.H., McAuliffe S.T. Motion camouflage detection using optical flow, Journal of Vision, 3 (9), 2003, P. 23–23.

12. Dempster A.P., Laird N.M., Rubin D.B. Maximum likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society: Series B (Methodological), 39 (1), 1977, P. 1–22.

13. Li Z., LiuD., Yang B. EM-based object recognition for camouflaged environments, Image and Vision Computing, 86, 2019, P. 10–18.

14.Haralick R.M., Shanmugam K., Dinstein I.H. Textural features for image classification, IEEE Transactions on Systems, Man, and Cybernetics, SMC-3 (6), 1973, P. 610–621.

15.Wang Y., Li S., Li C. Camouflaged object detection based on HSV color space and GLCM features, Journal of Electronic Imaging, 30 (2), 2021, P. 122–135.

16.Wang Z., Bovik A.C., Sheikh H.R., Simoncelli E.P. Image quality assessment: From error visibility to structural similarity, IEEE transactions on image processing, 13 (4), 2004, P. 600–612.

17.Zhou Y., Li H., Chen J. Weighted structural similarity for camouflaged image detection, Journal of Imaging Science and Technology, 64 (3), 2020, P. 343–359.

18.Thayer G. Concealing-Coloration in the Animal Kingdom: An Exposition of the Laws of Disguise Through Color and Pattern,MacMillan, 1918, P. 260.

19.Khan S., Anwar A., Ahmad S. A review of camouflaged object detection methods: Taxonomy, datasets, and challenges, Neurocomputing, 446, 2021, P. 172–192.

20.Haralick R.M., Shanmugam K., Dinstein I. Textural Features for Image Classification, IEEE Transactions on Systems, Man, and Cybernetics, 3 (6), 1973, P. 610–621.

21.Lucas B.D., Kanade T. An iterative image registration technique with an application to stereo vision, Proceedings of the 7th International Joint Conference on Artificial Intelligence, 1981, P. 674–679.

Polischuk S.M. REVIEW OF METHODS FOR DETECTING MASKED OBJECTS IN THE CONTEXT OF VISUAL SIMILARITY TO THE BACKGROUND

The article is devoted to the study of modern methods of detecting camouflaged objects (COD – Camouflage Object Detection), which have a high degree of similarity with the environment, which significantly complicates their identification by traditional computer vision algorithms. The article reveals the essence of the phenomenon of camouflage as a way to hide the texture and spectral characteristics of an object against the background of a scene, and also characterizes the main challenges associated with its detection. It has been determined that the key problem of COD is the minimal contrast between the object and the background, which requires the use of complex methods for analyzing texture, color, motion, and spatial-structural relationships. It has been found that the use of machine learning methods, including deep neural networks, significantly improves detection accuracy due to the ability to process large volumes of images with hidden patterns that are not available for classical approaches. The article analyzes a wide range of approaches, including algorithms based on HSV-GLCM, structural similarity metrics (WSSIM), optical flow models, Gaussian Mixture Models (GMM), and hybrid frameworks that combine statistical modeling and deep learning. It has been established that the effectiveness of the methods depends on the type of scene (static or dynamic), the environment (natural or artificial), and the availability of multimodal data (thermal, hyperspectral, depth). The article also highlights the trends in the integration of COD with other computer vision tasks, such as object segmentation and tracking, which allows to improve the adaptability of systems in real conditions. It is outlined that a promising direction is the creation of multilevel systems with elements of attention, attractiveness mapping, and the use of semantic features. The presented data prove that effective detection of camouflaged objects is possible only with an integrated approach, taking into account the characteristics of the scene, the type of object, and the context of application. Thus, the study sheds light on new directions for the development of automated COD systems in military affairs, environmental monitoring, security technologies, and robotics.

Key words: *detection of disguised objects, deep learning, computer vision, de-masking, recognition methodologies.*

УДК 004.42:78.06
DOI https://doi.org/10.32782/2663-5941/2025.3.2/43

Регенель Т.Ю.
Державний університет «Житомирська політехніка»

Фант М.О.
Державний університет «Житомирська політехніка»

Савіцький Р.С.
Державний університет «Житомирська політехніка»

РОЗРОБКА СИСТЕМИ КОНТРОЛЮ ВЕРСІЙ ДЛЯ МУЗИЧНИХ ПРОЄКТІВ

У статті розглянуто розробку системи контролю версій для працівників музичної індустрії. В розробку входить сама система контролю версій, а також екосистема програмних продуктів на її основі, зокрема, застосунку для роботи з файлами проєктів безпосередньо на комп'ютері користувача, а також веб-платформи для зберігання, керування та колективної роботи над проєктами. Не зважаючи на наявність великої кількості різноманітного програмного забезпечення, яке покриває практично всі основні потреби різних робочих процесів у музичній індустрії, все ще відсутні інструменти для організації та контролю файлів різних версій проєкту. Існуючі аналоги на кшталт Git та GitHub не набувають широкого використання в цій сфері з низки причин, серед яких погана оптимізація системи щодо роботи з великими файлами (понад 50мб, в той час як середній обсяг файлів однієї сесії запису в музикантів може сягати декількох гігабайт), вимоги до навичок користувача у роботі з командним рядком, засобами автентифікації та ін., а також орієнтованість функціоналу саме на роботу з кодом та специфіку робочого процесу програмістів, тестувальників, DevOps-інженерів тощо, яка має ряд відмінностей від робочого процесу в музичній індустрії. Більш розповсюдженими у використанні є файлообмінники, такі як Google Drive або DropBox, і деякі з них мають вбудоване версіювання файлів. Проте, це версіювання дуже обмежене по своєму функціоналу, а також відсутня можливість працювати декільком учасникам паралельно (історія версій лінійна та не підтримує розгалужень. В результаті дослідження, спроєктовано саму систему контролю версій, на базі якої продукт, що складається з клієнтського додатку та веб-платформи з підтримкою підключення користувацького сховища, які покривають всі ці недоліки, і водночас пропонують не гіршу функціональність, значно підвищуючи ефективність та якість робочого процесу працівників сфери.

Ключові слова: *система контролю версій, версіювання, збереження файлів, веб-платформа, програмне забезпечення.*

Постановка проблеми. На сьогоднішній день робота з музикою складається з різних етапів, на кожному з яких може бути залучено багато ролей одночасно, окрім самого виконавця: продюсери, мікс-інженери, саунд-дизайнери, запрошені виконавці, сесійні музиканти тощо. Організація робочих матеріалів та різних версій повністю покладається на версіювання файлообмінників, або домовленості у найменуваннях та зберігання декількох копій проєкту в сховищі. Відсутні інструменти та сервіси, які надають функціонал для організації та синхронізації матеріалів між різними учасниками процесу протягом усієї роботи над проєктом. Існують аналогічні системи для представників інших сфер, наприклад, система Git та платформа GitHub. Проте ці програмні продукти вимагають певного рівня знань і навичок в сфері інформаційних технологій, що суттєво ускладнює їх використання представниками інших сфер.

Аналіз останніх досліджень і публікацій. В дослідженні Зуєвського Д. Р. та Кулішової Н. Є. [1] описано розробку системи контролю версій інтегрованої в основне ПЗ, зі збереженням змін об'єктів всередині самих об'єктів. Такий підхід має низку переваг, проте у вказаному дослідженні розглядалася робота з електронними виданнями. В той час як розповсюджені програмні продукти в цьому напрямі підтримують аддони та надають програмний інтерфейс для взаємодії з об'єктами, програмне забезпечення в музичній індустрії, зокрема, такий вид ПЗ як

Digital Audio Workstation або DAW, не надають достатньо гнучкі можливості маніпуляції елементами проєкту, що робить цей підхід не найкращою опцією в рамках нашого дослідження, і клієнтську частину буде оформлено в окремий незалежний додаток, а не в плагін/інтеграцію.

В дослідженні Александер Й. та ін. [2] розглянуто інший аспект систем контролю версій – синхронізацію даних на клієнтах. Зокрема, розглянуто розподілені системи контролю версій. Цей підхід має свої переваги. Наприклад, можливість побудувати систему без серверу взагалі, або із значно дешевшим у розробці та розгортанні серверним додатком. Проте, в роботі з музичними проєктами є певна специфіка: часто доведеться працювати з великими файлами, понад 50мб, в тому числі передачу цих даних по мережі. Також, нерідким явищем може бути співпраця фахівців з різних часових поясів, що може суттєво ускладнити синхронізацію даних на їх комп'ютерах (потрібен момент, у який обидва комп'ютери будуть доступні у мережі). Тому в межах нашого дослідження ми будемо використовувати клієнт-серверну архітектуру для синхронізації між клієнтами та збереження даних.

Постановка завдання. Необхідно розробити систему контролю версій, і на основі цієї системи розробити клієнтський додаток для безпосередньої взаємодії з файлами проєктів, а також веб-платформу для зберігання даних проєктів на віддаленому сервері, синхронізації даних між декількома клієнтами тощо. При розробці врахувати специфіку сфери застосування, включаючи тип і структуру файлів, вимоги щодо доступності і зручності інтерфейсу та ін.

Виклад основного матеріалу. *Розробка системи контролю версій.* Кінцеве запропоноване рішення складається з щонайменше двох програмних продуктів: клієнтський додаток і веб-платформа. Всі ці продукти працюватимуть з проєктами та їх елементами, тобто з функціоналом системи контролю версій (далі – СКВ). Щоб ефективніше організувати процес розробки, а також забезпечити відповідність функціоналу та імплементації СКВ між різними частинами рішення, СКВ буде реалізовано у вигляді npm-пакету, який встановлюватиметься та використовуватиметься у веб-платформі (NextJS), клієнтському додатку (Electron), та за потреби в інших продуктах, якщо буде потреба їх додати. При такому форматі роботи важливо, щоб пакет СКВ був типізованим, тому для його написання буде використано TypeScript. Оскільки надійність цього елемента розробки є вкрай критичною (адже інші елементи залежать від нього), пакет також буде покрито тестами за допомогою технології Jest. Це дозволить контролювати функціональність усієї СКВ після будь-яких внесених змін.

На відміну від програмістів, які працюють з кодом у файлах, тобто безпосередньо з вмістом файлу, музикантам для роботи потрібно декодувати цей вміст у відповідному ПЗ. А структура та кодування файлів проєктів цього ПЗ часто є закритим. До того ж, в залежності від обраного формату, зміна в конкретній ділянці аудіо файлу при конвертації може змінити значно більшу частину кінцевого вмісту файлу. Таким чином, розглядати конкретні байти вмісту файлу як атомарну сутність СКВ немає потреби. Замість цього, головною сутністю оберемо так званий Item – Елемент Проєкту. Кожен Item міститиме свій унікальний ідентифікатор, ідентифікатор контенту і шлях кінцевого файлу у проєкті. Контентом називаємо копію файлу з ідентичним вмістом, та ідентифікатором замість назви. Цей контент зберігатиметься в службовій директорії СКВ, і використовуватиметься для відтворення стану файлу на момент певної версії. Якщо якийсь файл зміниться, буде створено новий Item, з тим самим шляхом, що і попередній Item для цього файлу, але з іншим ідентифікатором контенту. Якщо файл буде перейменовано/переміщено – також створюватиметься новий Item, з тим самим ідентифікатором контенту, але з іншим шляхом. У такий спосіб, ми зберігатимемо мінімальну можливу кількість контенту, уникаючи зайвої дублікації даних, що зменшує обсяг диску необхідний для роботи, і трафіку відповідно.

Щоб репрезентувати зміну в проєкті, введемо сутність ItemChange – Зміну елемента. Кожен ItemChange може містити ідентифікатори Item-ів, які відповідать стану файлу до зміни та після зміни. Назвімо це полями from і to. Якщо існують обидва ідентифікатори – файл було змінено, або перейменовано/переміщено. Якщо існує тільки ідентифікатор to, а до змін у файлу не було ніякого стану – файл було вперше додано в проєкт. Якщо ж навпаки, до внесення змін існував певний стан файлу, а стану після змін немає (відсутній ідентифікатор to) – файл було видалено з проєкту.

Сутністю, яка відповідає певній версії проєкту, буде Commit. Commit включатиме в себе метадані (ідентифікатор, назва, опис, автор, дата

створення тощо), посилання на батьківський та дочірній коміти для роботи з історією комітів як з деревом, а також найголовніше – список змін, тобто список ItemChange-ів. Для гнучкої організації введемо сутність гілки, яка являє собою просто посилання на певний коміт.

А сутністю, яка відповідає проєкту музиканта, буде сам проєкт СКВ – Project. Project міститиме метадані (ідентифікатор, назва, опис, автор, дата створення тощо), поточний стан проєкту (поточний Commit, поточна гілка), список гілок, список Commit-ів та список усіх існуючих Item-ів. Саме цю сутність ми зберігатимемо у форматі JSON в службовому файлі та в БД.

Для колекцій сутностей використовуватимуться асоціативні масиви (Map). Збереження таких колекцій у вигляді пар ключ-значення дозволить зручно та ефективно отримувати сутність за її ідентифікатором.

Однією з найосновніших операцій в функціях СКВ буде порівняння станів проєкту. Порівнюватися можуть як стани проєкту з двох комітів, так і порівняння певного коміту з поточним станом файлів у директорії проєкту. Оскільки самі стани безпосередньо ніде не зберігаються, перед порівнянням їх потрібно буде сформувати.

Якщо потрібно отримати стан проєкту на момент певного коміту, це можна зробити в два етапи: визначення ланцюга комітів і накладання змін. Для визначення ланцюга коміті, нам потрібно обходити все дерево комітів від заданого коміта до його нащадка, поки ми не прийдемо до кореневого коміта проєкту, зберігаючи ідентифікатори всіх відвіданих комітів. В результаті отримаємо історію комітів для заданого коміту у вигляді переліку ідентифікаторів. Далі, для накладання змін, потрібно обійти отриманий перелік, формуючи стан проєкту у тимчасовій змінній. На кожній ітерації ми обходимо ItemChange-и коміту, і в залежності від типу змін (додавання, модифікація, видалення), відтворювати їх у тимчасовій змінній. В результаті обходу, отримаємо стан проєкту для цільового коміту у вигляді масиву Item-ів. Для зручності при подальшому порівнянні, до кожного Item-у додамо хеш його контенту.

Щодо отримання стану проєкту відповідно до поточних файлів у робочій директорії, ми обійдемо всі файли, пропускаючи службові файли (наприклад, саму службову директорію СКВ) і помічені для ігнорування користувачем. Для кожного файлу створюватимемо тимчасовий Item без контенту, але зі шляхом і хешом вмісту файлу. В результаті, незалежно від того, стани чого саме ми порівнюємо, ми завжди працюємо з масивами Item-ів із доданими хешами вмісту файлів.

Коли обидва стани отримані, можна проводити саме порівняння. Для цього попередньо представимо обидва стани у вигляді асоціативних масивів, де ключем буде шлях до файлу, а значенням – відповідний Item. Далі, достатньо обійти всі існуючі ключі для обох масивів. Якщо ключ існує лише в одному з них – фіксуємо це як відповідну зміну (видалення або додавання), якщо ж ключ присутній в обох масивах, але хеш вмісту відрізняється – фіксуємо це як модифікацію. Решта ключів вважаються не зміненими файлами. Усі зафіксовані зміни повертаємо як результат порівняння.

На основі порівняння реалізовується створення нового коміту: отримується стан на момент

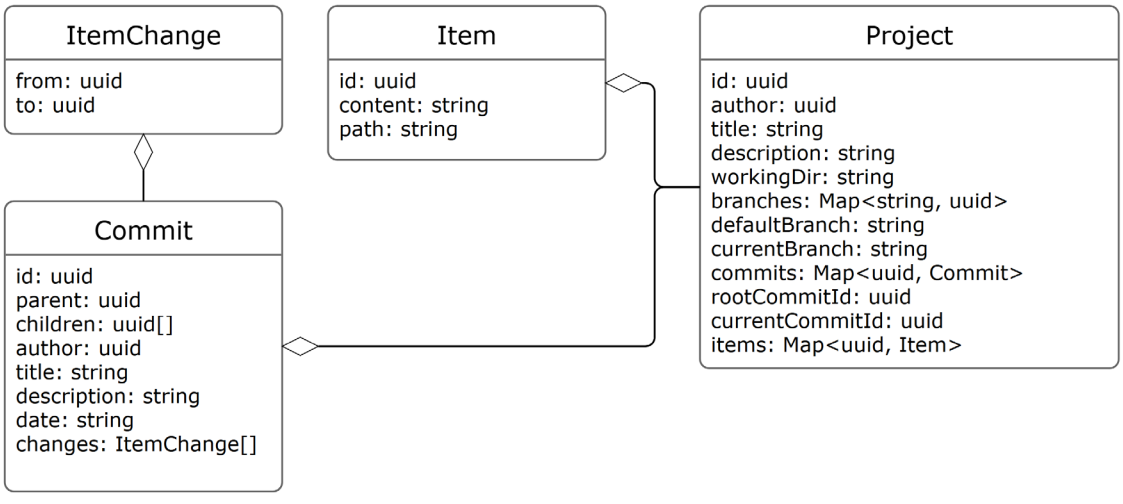


Рис. 1. Діаграма основних класів пакету СКВ

останнього коміту поточної гілки, поточний стан файлів у директорії, здійснюється порівняння, і для кожної з виявлених змін створюється ItemChange. Для нових та модифікованих файлів, за потреби, створюються нові Item-и, та зберігаються в проєкті, і зберігається новий контент відповідно.

Після реалізації інших функцій проєкту (збереження до файлу або JSON, відтворення з файлу або JSON, автодоповнення частини ідентифікатора коміту тощо), пакет готовий до використання. Разом з усіма класами проєкту, в ньому також буде присутній інтерфейс IStorageProvider, який передбачає усі основні операції з файловою системою (читання/створення/видалення файлів та директорій, перевірка існування файлу/директорії, копіювання файлу тощо), а також деякі імплементації цього інтерфейсу, завдяки чому пакет буде придатний до використання в різних середовищах: файлова система комп'ютера користувача, хмарне сховище, Google Диск користувача та інші сховища.

Розробка клієнтського додатку. Клієнтський додаток на період розробки Proof of Concept можна реалізувати у вигляді консольного додатку. Оскільки основна частина – СКВ – реалізована у вигляді окремого незалежного пакету, реалізація віконного додатку на її основі не вимагатиме значних ресурсів. Для розробки консольного додатку обрано npm-пакет commander, разом із додатковими пакетами для форматування і зручної взаємодії з користувачем (chalk, inquirer, nanospinner).

Перед здійсненням будь-яких операцій з проєктом, які стосуються веб-сервісу, клієнтський додаток вимагатиме автентифікації користувача. Задля забезпечення зручності і водночас надійності, реалізовується автентифікація згідно OAuth 2.0 Device Flow (Рис. 2). Цей механізм передбачає вхід користувача в його обліковий запис в браузері після ініціювання автентифікації додатку на пристрої. Після успішного входу користувача і його згоди, веб-платформа згенерує токен для авторизації, та відправить додатку, який в режимі очікування з певним інтервалом опитуватиме веб-платформу для отримання цього токену. До будь-яких подальших запитів клієнтський додаток додаватиме цей токен в заголовках запиту, отримуючи доступ до відповідних ресурсів.

Розробка веб-платформи. Веб-платформу буде реалізовано з використанням фреймворку NextJS. В якості БД обрано PostgreSQL, а для взаємодії з нею використовуватиметься Prisma ORM. Окрім користувацького інтерфейсу, засобами NextJS реалізується API для роботи з клієнтським додатком.

Сутності БД слід почати розглядати з сутності користувача – User. Вона міститиме ідентифікатор користувача, особисті дані, та інформа-

Рис. 2. Послідовність автентифікації клієнтського додатку

цію для авторизації. Другою важливою сутністю є сховище – Storage. Вона репрезентує сховище для файлів, доступне користувачеві. За замовчуванням користувачеві доступне власне сховище веб-платформи, проте він може додати власне: під'єднати власний Google-диск, Dropbox, або в перспективі навіть власний пристрій (наприклад, домашній сервер) в якості сховища за допомогою спеціального додатку. В цій сутності зберігатиметься вся необхідна інформація для під'єднання та отримання доступу до сховища.

Проєкту відповідатиме сутність Project. В ній зберігатиметься JSON-представлення проєкту, сховище для зберігання, налаштування публічності тощо. Між сутностями Project, User та Storage будуть також проміжкові таблиці, визначаючі права доступу користувача до певного сховища та певного проєкту.

Через веб-інтерфейс користувач зможе керувати своїм профілем, своїми проєктами (переглядати структуру та історію змін, змінювати налаштування, надавати доступ іншим користувачам тощо), своїми сховищами (переглядати статус, налаштовувати доступ користувачів тощо), та інші функції які в перспективі можна додати на веб-платформу. Решта функцій, зокрема, безпосередня взаємодія з файлами проєктів, покладається на клієнтський додаток.

Отже, одна з найголовніших функцій веб-платформи – передача даних між сховищем і користувачем. Це включає в себе роботу з конкретним сховищем, приєднаним до проєкту, перевірку прав доступу користувача до запитаного ресурсу та інші операції.

Оскільки система працюватиме безпосередньо з даними для автентифікації на користувацькому сховищі, а доступ до проєкту може бути навіть відкритий для всіх користувачів, потрібно унеможливити розкриття вразливої інформації стороннім особам. При цьому вибудовування складного ланцюга перевірок та запитів може суттєво сповільнити передачу файлів, що також не є бажаним. Для забезпечення одночасно безпеки і швидкості роботи, буде використано механізм підписаних посилань Signed URLs.

Одна з основних переваг Signed URLs – пряме з'єднання клієнта і сховища без посередників та проміжкових обробок і відповідно навантажень на сервер. Проте, не всі сховища підтримують цей механізм. Зокрема, одна із найближчих середньостатистичному користувачеві опцій – Google-диск. На момент дослідження, API Google-диску не підтримує видачу Signed URLs, і аналогічна ситуація може виникати з іншими сховищами, підтримка яких додаватиметься в перспективі.

Щоб вирішити цю проблему, на веб-платформі введемо певний фасад, який буде відповідати за всі запити пов'язані із файлами. Якщо клієнтському додатку або веб-інтерфейсу потрібно завантажити файл зі сховища або на сховище, ініціюється передача файлу надсиланням запиту на відповідний ендпоінт. В цьому запиті передається перелік файлів, з якими планується затребувана операція. Цей запит потрапляє в обробку фасаду, після чого перевіряються права доступу користувача до проєкту, якому належить ресурс, а також сховище на якому він зберігається.

Якщо це сховище має вбудовану підтримку підписаних посилань – ці посилання отримуються напряму від сховища і передаються у відповіді клієнтському додатку, надаючи можливість обміну даними зі сховищем напряму, без викриття вразливих даних при цьому. Якщо ж сховище, на якому зберігається проєкт, не підтримує підписані посилання, тоді цю відповідальність бере на себе фасад, видаючи клієнту власні підписані посилання, і самостійно реалізовуючи їх обробку і роботу зі сховищем. Таке рішення може вимагати додаткових ресурсів серверу та дещо ускладнює розробку, проте досягнуті безпека та швидкість нівелюють це.

При проєктуванні веб-платформи також варто приділити увагу аспектам інфраструктури. Одна з найкращих опцій для розгортання NextJS додатків – платформа Vercel. У такому випадку будуть доступні помірна вартість розгортання (включаючи безкоштовний план), тісна інтеграція із самим фреймворком, логування та моніторинг тощо. Проте, цінова політика Vercel не є оптимальною для активного навантаження, такого як, наприклад, фасад, що відповідатиме за передачу файлів. Подібне навантаження буде оптимальніше розгортати на віртуальному приватному сервері (VPS). Таким чином, для найбільш оптимального використання платформою доступної інфраструктури, ці дві частини – фасад і NextJS додаток, варто розділити.

В той самий час, потрібно забезпечити комунікацію цих двох частин, адже дані, потрібні фасаду (наприклад, конфігурація певного сховища), зберігатимуться в БД, до якої має доступ лише NextJS додаток. Налагодити роботу двох різних додатків з однією БД і спільною схемою вкрай важко: є ризики конфліктів при невідповід-

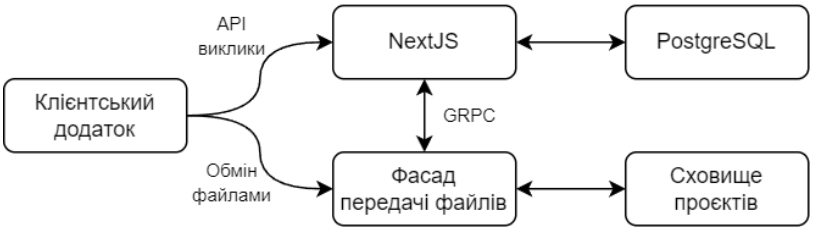


Рис. 3. Загальна архітектура системи

ності схем, складніше керувати доступом до БД тощо. Тому доступ до даних необхідних фасаду буде реалізовано саме через комунікацію із NextJS додатком. Для цієї ролі можна застосувати технологію gRPC. Вона забезпечить водночас зручність розробки (типізація, робота із окремим сервісом через виклики методів у програмному коді, а не прості HTTP запити тощо) і швидкість роботи.

Висновки. В цій статті було розглянуто розробку рішення із системою контролю версій для працівників музичної індустрії. Рішення представляє собою готову до використання та масштабування систему з клієнтського додатку, веб-платформи та фасаду для передачі файлів. Хмарна частина рішення є модульною та готовою до роботи на оптимальній інфраструктурі, забезпечуючи високу ефективність та помірну вартість розгортання, можливість балансування навантаження та обох видів масштабування (горизонтальне і вертикальне). Порядок обміну файлами передбачає ізоляцію вразливих даних, таких як облікові дані від сховища проєктів, від сторонніх осіб. Функціонал розробленої СКВ та можливості веб-платформи у сукупності покривають всі основні потреби різних ролей у робочому процесі виробництва музики, а під час самої розробки було досліджено аспекти архітектури ПЗ, безпеки даних та оптимізації роботи з файлами.

Список літератури:

1. Зуєвський Д.Р., Кулішова Н.Є. Розробка внутрішньопрограмної системи контролю версій: Радіоелектроніка та молодь у XXI столітті: ХНУРЕ, 2023, Том 6 ч. 2, 182 с.

2. Alexander Y., Benjie C., Robert M. Pastwatch: a Distributed Version Control System: NSDI ’06: 3rd Symposium on Networked Systems Design & Implementation, 381 p.

Rehenel T.Yu., Fant M.O., Savitskyi R.S. DEVELOPMENT OF A VERSION CONTROL SYSTEM FOR MUSIC PROJECTS

The article describes the development of a version control system for music industry workers. The development includes the version control system itself, as well as an ecosystem of software products based on it, including a client application for working with project files directly on the user’s computer, as well as a web platform for storing, managing, and collaborating on projects. Despite the availability of a wide variety of software that covers almost all the basic needs of various workflows in the music industry, there are still no tools for organizing and controlling files of different versions of a project. Existing alternatives like Git and GitHub are not widely used in this area for a number of reasons, including poor system optimization for working with large files (over 50 MB, while the average file size of a recording session for musicians can reach several gigabytes), requirements for user skills in working with the command line, authentication tools, etc, as well as the focus of the functionality on working with code and the specifics of the workflow of programmers, testers, DevOps engineers, etc., which has a number of differences from the workflow in the music industry. File sharing services such as Google Drive or DropBox are more commonly used, and some of them have built-in file versioning. However, this versioning is very limited in its functionality, and there is no possibility for several participants to work in parallel (the version history is linear and does not support branching). As a result of the study, the version control system itself was designed, based on which the product, consisting of a client application and a web platform with support for connecting user storage, covers all these shortcomings, and at the same time offers no worse functionality, significantly increasing the efficiency and quality of the workflow of workers in the field.

Key words: *version control system, versioning, file storage, web platform, software.*

УДК 004.942:004.415.2
DOI https://doi.org/10.32782/2663-5941/2025.3.2/44

Рихальський О.Ю.
ПВНЗ «Європейський університет»

МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ СЕМАНТИЧНОЇ СУМІСНОСТІ ПАРАМЕТРИЧНИХ КОНФІГУРАЦІЙ У CAD-СИСТЕМАХ

У статті досліджено проблему забезпечення семантичної сумісності параметричних конфігурацій у CAD-середовищах, що має основне значення для підвищення рівня автоматизації, структурної узгодженості й адаптивності сучасного інженерного проєктування. Актуальність теми зумовлено необхідністю інтеграції гетерогенних моделей із різною логікою побудови в межах єдиного інформаційного простору, що ускладнюється фрагментарністю структур, неоднорідністю форматів і відсутністю загального семантичного підґрунтя. Метою статті є узагальнення та формалізація методологічних підходів до математичного моделювання семантичної сумісності параметричних конфігурацій у CAD-системах на основі принципів <M,P>-простору, що дає змогу підвищити коректність опису, логічну цілісність і рівень автоматизованої інтеграції складних технічних моделей.

Методологія дослідження спирається на інструментарій модельно-параметричного підходу, логіко-алгебраїчного опису залежностей, формальних контекстів, продукційних правил і механізмів онтологічної нормалізації. Проведено аналіз наявних формалізованих засобів подання знань, побудовано модель узгодження параметричних конфігурацій через виявлення сумісних околиць у <M,P>-просторі, оцінювання семантичної відстані та застосування оператора інтеграції моделей.

Наукову новизну становить запропонована структура формалізованої інтеграції гетерогенних моделей з урахуванням контекстної сумісності та семантичного зіставлення, що забезпечує функціональну цілісність проєктного простору. Встановлено, що головними перешкодами автоматизованої перевірки є семантична фрагментація, відсутність міжпідсистемної взаємодії та недостатній рівень підтримки адаптивної перевірки змін. У висновках доведено ефективність запропонованої моделі для створення адаптивних, логічно узгоджених і масштабованих параметричних конфігурацій. Перспективи подальших досліджень пов'язані з упровадженням елементів когнітивного проєктування, інтелектуального трасування залежностей і засобів машинного навчання для поглибленого аналізу семантичної відповідності в CAD-середовищах нового покоління.

Ключові слова: *модельно-параметричний підхід, логіко-онтологічна узгодженість, контекстне проєктування, формалізація знань, інженерна інтероперабельність, інформаційні технології, семантична сумісність.*

Постановка проблеми. У контексті сучасного розвитку комп'ютерного інжинірингу й автоматизованих систем проєктування особливої актуальності набуває проблема забезпечення семантичної сумісності параметричних конфігурацій у CAD-середовищах. Умови стрімкої адаптації до індивідуалізованого й високоточного виробництва потребують не лише геометричної відповідності елементів моделі, а і їхньої логічної та функціональної узгодженості на рівні внутрішніх зв'язків та атрибутів. Семантична сумісність визначає можливість автоматизованого поєднання окремих компонентів проєкту без втрати функціональної цілісності системи, що особливо критично для складних технічних об'єктів з ієрархічною структурою, модульною побудовою або змінними параметрами конфігурацій. Водночас традиційні підходи до формування параметричних моделей у CAD-системах не враховують повною мірою логіко-семантичних взаємозв'язків між об'єктами, що обмежує можливості гнучкої адаптації проєктів, автоматизованої валідації моделей і перевірки їхньої коректності на ранніх етапах проєктування. У зв'язку з цим виникає потреба у побудові формалізованих математичних моделей, які б давали можливість здійснювати оцінювання та забезпечення семантичної сумісності параметричних конфігурацій на основі чітко визначених логічних правил, онтологічних залежностей і структур даних. Розв'язання цієї проблеми має не лише наукову, ай прикладну цінність для оптимізації процесів автоматизованого проєктування, зменшення помилок у конфігураціях, прискорення циклів розроблення та підвищення надійності результатів CAD-моделювання в інженерній практиці.

Аналіз останніх досліджень і публікацій. Аналіз наукових досліджень математичного моделювання семантичної сумісності параметричних конфігурацій у CAD-системах дозволяє виокремити чотири змістовні напрями, що формують основу сучасної наукової парадигми розвитку CAD-технологій в аспекті семантичної інтеграції параметричних моделей.

Перший напрям пов'язаний із розвитком концептуальних підходів до поєднання параметричного та прямого моделювання як підґрунтя для забезпечення семантичної сумісності моделей. Цей аспект розкривають дослідження Цзоу Цюань (Q. Zou), Фен Хунь-Янь (H.-Y. Feng), Гао Сін (S. Gao), які запропонували концепцію variational direct modeling, що поєднує варіаційні методи з прямим редагуванням, зберігаючи логічні зв'язки між параметрами в умовах змін геометрії [1]. Аранбуру Андер (A. Aranburu), Хустель Д. (D. Justel), Контеро М. (M. Contero), Камба Х. Д. (J.D. Camba) досліджують геометричну варіативність параметричних моделей у контексті інженерного проєктування, акцентуючи увагу на потребі контролю над семантичними змінами при модифікаціях [2]. Кіратзі С. (S. Kyratzi), Азаріадіс П. (P. Azariadis) розвивають концепцію інтегрованого проектного задуму в параметричному моделюванні, що дозволяє зберігати логіку залежностей навіть при редизайні моделі [3]. Для подальшого розвитку цього напряму доцільно розробити формалізовані мови опису семантичних обмежень та алгоритми валідації таких обмежень при редагуванні CAD-моделей.

Другий напрям орієнтований на адаптацію параметричного моделювання до інженерних завдань шляхом використання знань і логічних правил. Редді Е. Дж. (E.J. Reddy), Редді К. С. (K.S. Reddy), Редді П. Н. (P.N. Reddy) аналізують останні досягнення knowledge-based modeling, яке дозволяє формувати параметричні конфігурації на основі предметно-орієнтованих правил і логіки взаємозв'язків [4]. Тан Чжен (Z. Tang), Цзоу Цюань (Q. Zou), Гао Сін (S. Gao) визначають математичні межі припустимих змін параметрів при багатопараметричному редагуванні, зберігаючи структуру моделі [5]. Тейлор Н. Дж. (N.J. Taylor), Гаймс Р. (R. Haimes) підкреслюють важливість коректного геометричного моделювання для подальшої інженерної симуляції, яка також передбачає логічну цілісність параметричних конфігурацій [6]. Розширення цього напряму можливе шляхом створення універсальних онтологій інженерних об'єктів для перевірки семантичної сумісності.

Третій напрям охоплює математичне моделювання параметричних залежностей для задач оптимізації геометрії. Гарріс С. (S. Harries), Абт К. (C. Abt), Бреннер М. (M. Brenner) аналізують підходи upfront CAD, що передбачають заздалегідь визначену параметричну структуру як основу для подальшої оптимізації [7]. Вукашинович Н. (N. Vukašinović), Духовник Й. (J. Duhovnik) представляють у монографії різні стилі CAD-моделювання – явне, параметричне, вільноформне – та аналізують їхнє поєднання для забезпечення точності й адаптивності [8]. Мачадо Ф. (F. Machado), Малпіка Н. (N. Malpica), Борромео С. (S. Borromeo) порівнюють OpenSCAD і FreeCAD у створенні відкритих параметричних моделей, акцентуючи увагу на структурному підході до опису параметрів [9]. Для подальшого розвитку цього напряму доцільно зосередитися на побудові моделей оберненого перетворення геометрії на параметри з урахуванням семантики конструкції.

Четвертий напрям відображає інтеграцію математичного параметричного моделювання з задачами симуляції, швидкодії та багатоплатформної інтеграції. Агарвал Д. (D. Agarwal), Робінсон Т. Т. (T.T. Robinson), Армстронг К. Г. (C.G. Armstrong) пропонують метод обчислення швидкості зміни параметрів через adjoint methods, який забезпечує точність і стабільність у проєктній оптимізації [10]. Севілья Р. (R. Sevilla), Злотнік С. (S. Zlotnik), Уерта А. (A. Huerta) розглядають спосіб вирішення геометрично параметризованих задач у CAD-середовищі за допомогою методів редукції порядку, які дозволяють зберігати важливі семантичні ознаки при спрощенні моделі [11]. Редді Е. Дж. (E.J. Reddy), Рангаду В. П. (V.P. Rangadu) демонструють створення параметричних моделей шестерень із урахуванням конструктивних залежностей і обмежень у межах знання-орієнтованого підходу [12]. Мохаммед А. С. Б. (A.S.B. Mohammed) та ін. використовують математичне моделювання для визначення впливу товщини провідника на центральну частоту антени, показуючи ефективність параметричного підходу у задачах високочастотної інженерії [13].

Дослідження Чжоу Т. (Zhou, T.), Лі Х. (Li, H.), Лі С. (Li, X.), Ланге К.Ф. (Lange, C.F.), Ма Й. (Ma, Y.) [14] акцентує увагу на інтеграції змінних фрактальних геометрій у параметричні

CAD-системи для покращення адаптивності моделей. Паралельно, Патіл К.Р. (Patil, K.R.), Дешпанде С. (Deshpande, S.), Джоші С. (Joshi, S.) [15] пропонують використання Python-скриптів для автоматизації процесу параметричного моделювання складових елементів механічних конструкцій, зокрема фланцевих з›єднань. Розширення цього напряму доцільно спрямувати на створення гібридних CAD-систем, які поєднують логічне програмування, графову семантику та адаптивну параметризацію для забезпечення цілісності та узгодженості інженерних моделей.

Попри розвиток CAD-систем, проблема семантичної сумісності параметричних конфігурацій залишається недостатньо вирішеною. Наявні підходи здебільшого зосереджуються на геометричному узгодженні моделей, тоді як логіко-семантичні зв'язки між параметрами, їхніми контекстами та функціональним призначенням уніфіковано не формалізовані. Недостатньо досліджено доцільність застосування модельно-параметричного підходу для опису семантичної структури моделей, а також обмежено проаналізовано ефективність наявних засобів подання знань у різних типах CAD-середовищ.

Невирішеними залишаються питання інтеграції гетерогенних моделей із різними контекстами, автоматизованого виявлення конфліктів між параметрами та міжсистемної взаємодії. Автоматизована перевірка в більшості систем обмежується синтаксичним рівнем без урахування семантичної сумісності. Запропоноване дослідження спрямоване на усунення цих прогалин через побудову формальної моделі узгодження параметричних контекстів у <M,P>-просторі, що забезпечить новий рівень логічної інтеграції, адаптивності й узгодженості конфігурацій у складних інженерних проєктах.

Постановка завдання. Мета статті: Мета статті – формалізація методологічних підходів до математичного моделювання семантичної сумісності параметричних конфігурацій у CAD-системах із використанням інструментів <M,P>-простору з метою підвищення коректності, узгодженості й автоматизованої інтеграції складних технічних моделей.

Для досягнення цієї мети були визначені такі завдання:

1. Обґрунтувати доцільність застосування модельно-параметричного підходу для формалізації семантичної структури параметричних конфігурацій у CAD-системах, а також проаналізувати наявні формалізовані засоби опису моделей і методи подання знань, що можуть бути адаптовані для створення семантично узгодженого проєктного простору.

2. Розробити формальну модель інтеграції гетерогенних параметричних моделей у CAD-середовищах на основі узгодження контекстів і принципів <M,P>-простору, що забезпечує семантичну цілісність конфігурацій.

3. Виявити основні проблеми автоматизованої перевірки семантичної сумісності моделей у CAD-системах і запропонувати рекомендації щодо впровадження уніфікованих підходів до керування конфігураціями, орієнтованих на підвищення узгодженості, адаптивності й ефективності проєктування.

Виклад основного матеріалу. У контексті інтелектуалізації інженерних процесів та ускладнення проєктних завдань у CAD-системах особливої актуальності набуває проблема забезпечення логічної та семантичної узгодженості параметричних моделей. Традиційні підходи зосереджуються переважно на геометричних аспектах, залишаючи поза увагою семантичні залежності між параметрами, що ускладнює інтеграцію гетерогенних моделей і підтримку цілісного інформаційного простору. Модельно-параметричний підхід, запропонований у наукових дослідженнях як інструмент уніфікованого подання знань, дає можливість структурувати інформацію на основі зв'язків між параметрами та моделями, формалізувати залежності між об'єктами проєктування, а також забезпечити обчислюваність контекстів, потрібних для автоматизованої перевірки сумісності. У наведеній таблиці подано основні елементи цього підходу у прикладному контексті параметричного моделювання в CAD-системах (табл. 1).

У таблиці 1 відображено функціональні та формальні аспекти застосування модельно-параметричного підходу, який структурує простір знань у CAD-системах через системну взаємодію моделей, параметрів і контекстів. Наприклад, параметром (P) у середовищі проєктування корпусу судна може бути товщина обшивки, яка в конкретній моделі (M) описується залежністю від типу сталі, навантаження та геометричних характеристик. Текст моделі (MT) відображає цю залежність у вигляді рівняння або логічної конструкції, що включається у правила перевірки. Контекст моделі (MK), наприклад, обмежує дію моделі лише до суден водотоннажністю понад 5000 тонн, що дає можливість автоматично вияв-

Таблиця 1

Основні компоненти модельно-параметричного підходу для формалізації семантичної структури параметричних конфігурацій у CAD-середовищах

Елемент модельно-параметричного підходу	Формалізований опис	Роль у формуванні семантичної структури
Параметр (P)	Формальна характеристика властивості об'єкта, що моделюється	Визначає атрибутивну сутність конфігурації моделі
Модель (M)	Відношення між параметрами: $Mj = R(P_1j, P_2j, ..., Pij)$	Встановлює логічні зв'язки між параметрами об'єкта
Текст моделі (MT)	Формальний запис внутрішньої структури моделі	Відображає структуру обчислень і залежностей
Контекст моделі (MK)	Формалізовані умови коректності й застосування моделі	Зумовлює межі застосовності параметричної моделі
<M,P>-простір	Сукупність параметрів, моделей і зв'язків між ними	Формує інформаційне середовище перевірки сумісності
Околиця <M,P>	Підмножина пов'язаних параметрів і моделей для об'єкта	Локалізує простір знань навколо основного параметра або моделі
Відстань у <M,P>-просторі	Числова метрика семантичної близькості між моделями або параметрами	Дає можливість оцінити узгодженість і семантичну сумісність моделей

Джерело: сформовано автором на підставі [1; 3; 4; 6; 7; 11]

ляти некоректне застосування під час порівняння з іншими об'єктами.

<M,P>-простір у такому випадку містить усі пов'язані параметри (наприклад, товщина, матеріал, довжина секції) та моделі, що встановлюють відношення між ними. Якщо дослідник змінює значення параметра, наприклад, зменшує товщину на 20%, околиця <M,P> дає можливість виявити пов'язані параметри, які автоматично підлягають перерахунку або валідації. Семантична відстань між новою моделлю та базовими шаблонами може бути використана для визначення її сумісності з наявними проєктними стандартами. Наприклад, якщо два параметри мають однакові значення, але різний контекст (наприклад, довжина секції для пасажирського і вантажного судна), то рівень семантичної сумісності буде низьким, що сигналізує про потенційну конфліктність.

На практиці цей підхід реалізується в інженерних середовищах типу Siemens NX, де користувачі формують правила залежності між параметрами та структурують модель як набір взаємопов'язаних логічних об'єктів [16]. У системі CATIA від Dassault Systèmes упроваджено підхід Product Knowledge Template, який дає можливість описувати параметричні моделі за допомогою контекстних обмежень і правил семантичного типу [17]. Крім того, використання підходу knowledge-based engineering забезпечує автоматизовану перевірку сумісності компонентів і полегшує керування залежностями під час створення складних конфігурацій на основі параметрів та їхніх взаємозв'язків [18].

Забезпечення семантичної узгодженості в параметричних CAD-моделях потребує інтеграції засобів, що дають змогу не лише формально описувати геометрію і властивості об'єктів, а й кодувати знання про їхню взаємодію, обмеження та контекст застосування. У сучасних системах проєктування дедалі ширше використовуються методи подання знань, що базуються на формальних структурах: фреймах, онтологіях, продукційних правилах, логіко-алгебраїчних моделях і схемах даних. Ці засоби дають можливість моделювати поведінку об'єктів, задавати причинно-наслідкові залежності між параметрами, автоматизувати перевірку коректності рішень і забезпечувати адаптивність до змін умов. Особливу увагу в контексті CAD потрібно звернути на підходи, які поєднують параметризацію з формальним контролем структурної та функціональної сумісності, зокрема через онтологічне подання знань або логічне узагальнення залежностей [3]. У таблиці 2 представлено порівняння основних формалізованих засобів, що застосовуються або можуть бути адаптовані до потреб семантичного проєктного простору (табл. 2).

У таблиці 2 узагальнено різні формалізовані засоби, які можуть бути використані або адаптовані для створення семантично зв'язного проєктного простору в CAD-системах. Ці інструменти мають різну обчислювальну природу, але однаково важливі для опису залежностей між параметрами, логічних умов функціонування об'єктів і контролю структурної узгодженості

Таблиця 2

Порівняльна характеристика формалізованих засобів подання знань для семантичного узгодження параметричних моделей у CAD-середовищах

Формалізований засіб подання знань	Основні характеристики	Потенціал адаптації для CAD
Продукційні правила	Структура: «якщо, то», логічні висновки за умовами	Підходить для автоматизації обмежень у параметричних моделях, ефективний у rule-based design
Фреймова структура	Ієрархічне подання об'єктів з атрибутами та слотами	Застосовується для моделювання типових компонентів із варіативними властивостями
Онтології	Формалізована множина понять і зв'язків між ними	Дає змогу створити семантично зв'язане середовище CAD-моделей та узгодити терміни між підсистемами
Логічні формули та логіко-алгебраїчні структури	Використання логіки предикатів або булевих виразів	Дає можливість перевіряти відповідність моделей умовам функціонування та забезпечувати несуперечність
XML-схеми та структуровані бази знань	Форматована структурна інформація з тегами й типами даних	Важливий інструмент для передавання параметричних моделей між системами (interoperability)

Джерело: сформовано автором на підставі [2; 4; 6; 7; 8; 13; 14]

моделей [10]. Продукційні правила є ефективними у випадках, коли модель має реагувати на конкретні значення параметрів через активацію відповідних сценаріїв модифікації, наприклад, у середовищі Autodesk Inventor iLogic, де користувачі задають умови для автоматичної зміни розмірів або конфігурацій [19]. Фреймові структури активно використовуються у PTC Creo, де створення сімейств моделей базується на змінних атрибутах, що зберігаються у вигляді шаблонів із внутрішніми слотами [20].

Онтології дедалі частіше застосовуються в системах міждисциплінарного проєктування, зокрема в галузі Smart Manufacturing і цифрових двійників, де використовується платформа OntoCAD, орієнтована на створення єдиного семантичного простору для архітектурних, механічних та електронних моделей [21]. XML-схеми реалізуються в більшості сучасних CAD/CAM/CAE-систем, наприклад, у Solid Edge для організації параметричних шаблонів і конфігурацій із подальшою інтеграцією у PLM-середовище [22]. Логіко-алгебраїчні моделі застосовуються у спеціалізованих інженерних експертних системах, таких як Dymola або Simcenter Amesim, де складні технічні системи перевіряються на функціональну та логічну узгодженість перед запуском симуляцій [23].

Адаптація формалізованих засобів подання знань до середовищ CAD дає можливість трансформувати системи із геометрично орієнтованих інструментів у повноцінні платформи інженерного мислення, де кожен параметр – не просто змінна, а носій контексту, логіки та функції в межах інтегрованої семантичної системи.

У сучасних CAD-середовищах спостерігається зростання потреби в інтеграції гетерогенних моделей, які походять із різних джерел, створені за допомогою різних програмних інструментів і мають відмінну логіку побудови, структуру параметрів і формати даних. Це особливо характерно для проєктів зі складною міждисциплінарною природою, де в єдиному інформаційному просторі мають співіснувати механічні, електронні, архітектурні та логістичні моделі. У таких умовах інтеграція моделей потребує не лише уніфікації форматів або синхронізації геометрії, а й узгодження семантичного рівня, тобто значень параметрів, їхніх взаємозв'язків, контексту застосування та цільового функціонального призначення.

Запропонована формальна модель інтеграції гетерогенних CAD-моделей ґрунтується на принципах <M,P>-простору, в межах якого кожна модель описується як відношення між параметрами (P), контекстами (МК) і текстами моделей (МТ). Інтеграція досягається через формальне виявлення сумісних околиць у <M,P>-просторі, порівняння відстаней між моделями, нормалізацію контекстів і побудову

оператора узгодження. Це актуально тим, що дає можливість здійснювати автоматизовану семантичну інтеграцію навіть для моделей із різною структурою за умови формалізації їхніх контекстів. Її відмінність від класичних PLM- або PDM-систем полягає в опорі не на метадані, а на структурну та логічну подібність моделей у межах формалізованого параметричного простору. Модель побудована на основі апарату логічної метрики, семантичної нормалізації й теорії інформаційної сумісності. Вона передбачає виявлення конфліктів між контекстами моделей, побудову онтологічно узгоджених описів і формування синтезованих моделей в уніфікованому середовищі (табл. 3).

Практична реалізація цієї формальної моделі в умовах сучасних CAD-процесів дає змогу значно скоротити час інтеграції різнорідних проєктних рішень, зменшити кількість помилок під час імпорту або поєднання моделей, а також підвищити якість технічного рішення завдяки формальній перевірці на етапі об'єднання. Наприклад, під час проєктування комплексної транспортної системи модель кузова (створена в SolidWorks), модель системи керування (створена в MATLAB/Simulink) і модель логістичного розташування (створена в ArchiCAD або Revit) можуть бути інтегровані на рівні опису параметричних контекстів через оператор Ω після попередньої нормалізації понятійного апарату й оцінювання семантичної відстані. У результаті формується логічно узгоджене середовище, в якому можливо проводити симуляції, оптимізацію та автоматизоване переналаштування моделей на основі змін у вихідних параметрах. Такий підхід також дає можливість адаптувати систему до майбутніх оновлень моделей без потреби повної реконструкції проєкту, оскільки вся інтеграція ґрунтується на формальних залежностях, а не лише на структурній ієрархії або зовнішніх посиланнях. У довгостроковій перспективі це забезпечує підвищення стійкості та гнучкості CAD-архітектури, що особливо важливо в умовах індустрії 4.0 і динамічно змінного виробничого середовища.

Процес автоматизованої перевірки семантичної сумісності в CAD-середовищах стикається з низкою системних обмежень, що ускладнюють інтеграцію моделей та перевірку їхньої коректності на рівні значень, логіки й взаємодії параметрів [14]. Однією з найглибших проблем є фрагментарність самих моделей, коли проєктні рішення формуються в межах окремих модулів або дисциплін, що не мають спільного семантичного підґрунтя. Наприклад, модель компонента, побудована в електронному модулі, може мати іншу структуру параметрів й інтерпретацію одиниць виміру порівняно з механічним модулем, навіть якщо описує той самий об'єкт [6]. Це призводить до розриву в логіці перевірки й до ситуацій, коли CAD-система не здатна автоматично виявити конфлікти або неприпустимі конфігурації.

Таблиця 3

Формалізовані компоненти моделі інтеграції гетерогенних CAD-моделей на основі узгодження параметричних контекстів у <M,P>-просторі

Компонент формальної моделі	Формалізоване визначення	Функція в інтеграції гетерогенних моделей
Параметричний контекст (МК)	Множина умов та обмежень для моделі: МКі = {Ci1, Ci2, ..., Cin}	Визначає область допустимих значень параметрів і сценаріїв їхнього використання
Відстань у <M,P>-просторі	$D(M_1, M_2) = f(\Delta P, \Delta МК)$	Оцінює рівень семантичної подібності між моделями для подальшого узгодження
Узгоджена околиця	$\partial < M, P > = \{Mi, Mj, ..., Mk\}$ при $D < \varepsilon$	Множина моделей, що можуть бути інтегровані без втрати логіки або змісту
Оператор узгодження моделей	$\Omega(M_1, M_2) \rightarrow M\sum$	Формалізує процес злиття моделей з урахуванням контекстної сумісності
Онтологічна нормалізація	N(Mi) = M′i з єдиним понятійним описом	Забезпечує узгодженість термінології та структур параметрів перед інтеграцією
Результівна модель	$M\sum = \Omega(N(M_1), N(M_2), ..., N(Mn))$	Створена на основі формалізованого об'єднання моделей у спільному середовищі

Джерело: власна розробка автора

Неоднорідність способів подання даних виявляється як у синтаксичному, так і в семантичному плані. Дані можуть зберігатися у вигляді формул, таблиць, структурованих записів, тегованих XML-файлів або внутрішніх об'єктів моделі, що не підлягають безпосередньому аналізу без попередньої нормалізації. Це знижує ефективність вбудованих механізмів перевірки, які зазвичай працюють на рівні простих обмежень (наприклад, перевірка типів або діапазонів), але не здатні забезпечити глибокий аналіз логічних зв'язків між об'єктами. Крім того, розрізненість системних підсистем, наприклад, між модулем тривимірного моделювання, розрахунковим ядром, засобами параметризації та базами знань, створює додаткові бар'єри для формування єдиного контексту моделі. Внаслідок цього перевірка обмежується окремими ділянками, не охоплюючи повного простору взаємозалежностей, і не може виявити семантичних суперечностей, які виникають лише на стиках між підсистемами.

Також потрібно зазначити про обмежену підтримку механізмів адаптивного аналізу, коли параметри або умови моделі змінюються динамічно залежно від попередніх рішень. Поточні інструменти автоматизованої перевірки зазвичай не мають механізмів відстеження історії змін параметрів, причинно-наслідкових залежностей між моделями або можливості масштабованого онтологічного узгодження. Відсутність системної підтримки логіко-онтологічних структур у межах CAD-систем обмежує не лише виявлення помилок, а й можливість попередження помилкових конфігурацій ще на етапі моделювання [24]. Як наслідок, значна частина помилок виявляється лише під час тестування або експлуатації моделей, що призводить до втрати часу, ресурсів і потреби повторного моделювання [25]. Загалом проблема автоматизованої перевірки семантичної сумісності потребує не лише алгоритмічних рішень, а й перегляду архітектури CAD-систем із позиції семантичної цілісності та взаємодії підсистем через єдиний логіко-параметричний простір.

Удосконалення керування параметричними конфігураціями в CAD-системах потребує впровадження уніфікованих підходів, що поєднують формальну структуру подання даних із механізмами логіко-семантичної узгодженості. Насамперед доцільно запровадити єдиний формат опису параметрів і моделей із чітко визначеними контекстами, які передбачають обмеження, допустимі області значень і логічні зв'язки з іншими параметрами. Це дасть можливість створити стабільну основу для валідації моделей не лише на рівні синтаксису, а й за змістовними залежностями, які мають вирішальне значення в мультидисциплінарних проєктах.

Універсалізація конфігурацій також потребує інтеграції механізмів автоматизованої нормалізації параметрів у межах <M,P>-простору, що забезпечить узгодження гетерогенних моделей через зіставлення їхніх семантичних структур, а не лише геометричних чи ієрархічних зв'язків. Важливим кроком є і побудова адаптивних правил зв'язування моделей, які можуть автоматично оновлюватися під час зміни основних параметрів або контекстів, реалізуючи концепцію динамічної конфігурації.

Доцільно також реалізувати підтримку онтологічного рівня опису моделей із застосуванням словників понять, ієрархій об'єктів і семантичних шаблонів, що дають можливість забезпечити спільне розуміння між підсистемами CAD-середовища. Таке впровадження сприятиме не лише зменшенню кількості помилок на етапі проєктування, а й значному скороченню часу повторної адаптації моделей у разі зміни умов або вимог.

Розширення засобів керування параметричними конфігураціями має передбачати модулі з візуалізації логіко-семантичних зв'язків, які дають змогу інженеру бачити не лише структуру моделі, а і її змістовні залежності. У поєднанні із системами трасування змін це дасть можливість сформувати адаптивні моделі, здатні до самоперевірки й автоматичного реагування на зміни. Таким чином, уніфіковані методи керування конфігураціями, засновані на формальних принципах <M,P>-структури, відкривають можливості для створення високонадійних, адаптивних і логічно узгоджених CAD-рішень, що відповідають вимогам сучасного інженерного проєктування в умовах складності та динаміки виробничих середовищ.

Висновки. У дослідженні обґрунтовано, що забезпечення семантичної сумісності параметричних конфігурацій є основним чинником підвищення надійності й автоматизованості CAD-проєктування. Запропоновано модельно-параметричний підхід, заснований на принципах <M,P>-простору, який дає можливість формалізувати логічні зв'язки між параметрами, контекстами та моделями, забезпечуючи узгоджене функціонування гетерогенних конфігурацій.

Аналіз формалізованих засобів подання знань засвідчив ефективність комбінованого використання продукційних правил, онтологій, фреймів і XML-схем для організації семантичного середовища CAD-систем. Виявлено основні проблеми: фрагментарність моделей, семантичну неоднорідність і слабку взаємодію між підсистемами, що знижують ефективність автоматизованої перевірки.

Запропонована формальна модель інтеграції моделей через узгодження контекстів у <M,P>-просторі забезпечує семантичну цілісність проєктного простору, можливість адаптивного оновлення конфігурацій і зменшення проєктних помилок.

Подальші дослідження доцільно спрямувати на поєднання принципів <M,P>-моделювання з когнітивними методами й машинним навчанням для створення інтелектуальних систем автоматизованого контролю та семантичної оптимізації в CAD-середовищах.

Список літератури:

1. Zou Q., Feng H.-Y., Gao S. Variational direct modeling: A framework towards integration of parametric modeling and direct modeling in CAD. *Computer-Aided Design*. 2023. Vol. 157, № 103465. DOI: https://doi.org/10.1016/j.cad.2022.103465 (date of access: 10.04.2025).

2. Aranburu A., Justel D., Contero M., Camba J.D. Geometric variability in parametric 3D models: Implications for engineering design. *Procedia CIRP*. 2022. Vol. 109, № 1. P. 383–388. DOI: https://doi.org/10.1016/j.procir.2022.05.266 (date of access: 10.04.2025).

3. Kyratzi S., Azariadis P. Integrated design intent of 3D parametric models. *Computer-Aided Design*. 2022. Vol. 146, № 103198. DOI: https://doi.org/10.1016/j.cad.2022.103198 (date of access: 10.04.2025).

4. Reddy E.J., Reddy K.S., Reddy P.N. Recent Advancements in Knowledge-Based Parametric Modeling of Mechanical Components. In: Yadav S., Arora P.K., Sharma A.K., Kumar H. (eds) *Proceedings of Third International Conference in Mechanical and Energy Technology. ICMET 2023*. Singapore: Springer. *Smart Innovation, Systems and Technologies*. 2024. Vol. 390. P. 389–398. DOI: https://doi.org/10.1007/978-981-97-2716-2_35 (date of access: 10.04.2025).

5. Tang Z., Zou Q., Gao S. Determining valid parameter ranges for multi-parameter editing of 3D parametric CAD models. *Journal of Computing and Information Science in Engineering*. 2025. Vol. 25, № 1. P. 1–18. DOI: https://doi.org/10.1115/1.4068115 (date of access: 10.04.2025).

6. Taylor N.J., Haimes R. Geometry modelling: Underlying concepts and requirements for computational simulation. In: *2018 Fluid Dynamics Conference*. 2018. Article 3402. DOI: https://doi.org/10.2514/6.2018-3402 (date of access: 10.04.2025).

7. Harries S., Abt C., Brenner M. Upfront CAD—Parametric Modeling Techniques for Shape Optimization. In: Minisci E., Vasile M., Periaux J., Gauger N., Giannakoglou K., Quagliarella D. (eds) *Advances in Evolutionary and Deterministic Methods for Design, Optimization and Control in Engineering and Sciences*. Cham: Springer. *Computational Methods in Applied Sciences*. 2019. Vol. 48. P. 115–127. DOI: https://doi.org/10.1007/978-3-319-89988-6_12 (date of access: 10.04.2025).

8. Vukašinović N., Duhovnik J. *Advanced CAD modeling. Explicit, parametric, free-form CAD and Re-engineering*. Cham: Springer Nature Switzerland AG, 2019. 254 p. DOI: https://link.springer.com/content/pdf/10.1007/978-3-030-02399-7.pdf (date of access: 10.04.2025).

9. Machado F., Malpica N., Borromeo S. Parametric CAD modeling for open source scientific hardware: Comparing OpenSCAD and FreeCAD Python scripts. *Plos One*. 2019. Vol. 14, № 12. Article e0225795. DOI: https://doi.org/10.1371/journal.pone.0225795 (date of access: 10.04.2025).

10. Agarwal D., Robinson T.T., Armstrong C.G., et al. Parametric design velocity computation for CAD-based design optimization using adjoint methods. *Engineering with Computers*. 2018. Vol. 34, № 2. P. 225–239. DOI: https://doi.org/10.1007/s00366-017-0534-x (date of access: 10.04.2025).

11. Sevilla R., Zlotnik S., Huerta A. Solution of geometrically parametrised problems within a CAD environment via model order reduction. *Computer Methods in Applied Mechanics and Engineering*. 2020. Vol. 358, № 1. Article 112631. DOI: https://doi.org/10.1016/j.cma.2019.112631 (date of access: 10.04.2025).

12. Reddy E.J., Rangadu V.P. Development of knowledge based parametric CAD modeling system for spur gear: An approach. *Alexandria Engineering Journal*. 2018. Vol. 57, № 4. P. 3139–3149. DOI: https://doi.org/10.1016/j.aej.2018.07.010 (date of access: 10.04.2025).

13. Mohammed A.S.B., Kamal S., Ain M.F.B., Hussin R., Najmi F., Sundi S.A., Ahmad Z.A., Ullah U., Omar M.F.B.M., Othman M. Mathematical model on the effects of conductor thickness on the centre frequency at 28 GHz for the performance of microstrip patch antenna using air substrate for 5G application. *Alexandria Engineering Journal*. 2021. Vol. 60, № 6. P. 5265–5273. DOI: https://doi.org/10.1016/j.aej.2021.04.050 (date of access: 10.04.2025).

14. Zhou T., Li H., Li X., Lange C. F., Ma Y. Feature-based modeling for variable fractal geometry design integrated into CAD system. Advanced Engineering Informatics. 2023. Vol. 57. Article 102006. DOI: https://doi.org/10.1016/j.aei.2023.102006 (date of access: 10.04.2025).

15. Patil K.R., Deshpande S., Joshi S., et al. Developing an integrated design framework using Python scripting for parametric CAD modelling of flange coupling. International Journal of Interactive Design and Manufacturing. 2024. Vol. 18. P. 4451–4462. DOI: https://doi.org/10.1007/s12008-023-01694-x (date of access: 10.04.2025).

16. Siemens NX: website. 2025. URL: https://www.plm.automation.siemens.com/global/en/products/nx/ (date of access: 04.04.2025).

17. CATIA by Dassault Systèmes: website. 2025. URL: https://www.3ds.com/products-services/catia/ (date of access: 04.04.2025).

18. Knowledge-Based Engineering (KBE) in CAD systems: website. 2025. URL: https://www.engineering.com/story/knowledge-based-engineering-in-cad (date of access: 04.04.2025).

19. Autodesk Inventor iLogic: website. 2025. URL: https://www.autodesk.com/products/inventor/ilogic (date of access: 04.04.2025).

20. PTC Creo Parametric: website. 2025. URL: https://www.ptc.com/en/products/creo (date of access: 04.04.2025)

21. OntoCAD Project – Semantic Engineering: website. 2025. URL: https://ontocad.net (date of access: 04.04.2025)

22. Solid Edge – Siemens Software: website. 2025. URL: https://solidedge.siemens.com (date of access: 04.04.2025).

23. Dymola by Dassault Systèmes: website. 2025. URL: https://www.3ds.com/products-services/catia/products/dymola (date of access: 04.04.2025).

Rykhalskyy O.Yu. MATHEMATICAL MODELING OF SEMANTIC COMPATIBILITY OF PARAMETRIC CONFIGURATIONS IN CAD SYSTEMS

The article investigates the problem of ensuring the semantic compatibility of parametricconfigurations in CAD environments, which is essential for increasing the level of automation, structuralconsistency, and adaptability of modern engineering design. The relevance of the topic is due to the need to integrate heterogeneous models with different constructionlogics within a single information space, which is complicated by the fragmentation of structures, heterogeneity of formats, and the lack of an everyday semantic basis. The purpose of the article is to generalize and formalize methodological approaches to mathematical modeling of semantic compatibility of parametricconfigurations in CAD systems based on the principles of <M, P>-space, which allows to improve the correctness of description, logical integrity and level of automated integration of complex technical models.

The research methodology is based on the tools of the model-parametric approach, logical and algebraic description of dependencies, formal contexts, product rules and mechanisms of ontologicalnormalization. An analysis of the existing formalized means of knowledge representation is carried out, and a model for matchingparametricconfigurations is built by identifying compatibleneighborhoods in <M, P>-space, estimating the semantic distance and applying the model integration operator.

The scientific novelty is the proposed structure of formalized integration of heterogeneous models, taking into account contextual compatibility and semantic mapping, which ensures the functional integrity of the design space. It has been established that the main obstacles to automated verification are semantic fragmentation, lack of inter-system interaction, and insufficient support for adaptive change verification. The conclusions prove the effectiveness of the proposed model for creating adaptive, logically consistent and scalableparametricconfigurations. Prospects for further research are related to the introduction of elements of cognitive design, intelligent dependencytracing, and machine learning tools for in-depth analysis of semanticconformance in next-generation CAD environments.

Key words: *model-parametric approach, logical and ontologicalconsistency, contextual design, knowledge formalization, engineering interoperability, information technologies, semantic compatibility.*

УДК 681.5
DOI https://doi.org/10.32782/2663-5941/2025.3.2/45

Семенець Д.А.
Харківський національний університет імені В.Н. Каразіна

Васильчук Д.П.
Харківський національний університет імені В.Н. Каразіна

Арестова Т.В.
Харківський національний університет імені В.Н. Каразіна

ЦИФРОВИЙ ПРОПОРЦІЙНО-ІНТЕГРАЛЬНИЙ-РЕГУЛЯТОР ДЛЯ ЕЛЕКТРОМЕХАНІЧНОГО ОБ'ЄКТУ ДРУГОГО ПОРЯДКУ З ЗАПІЗНЮВАННЯМ

У статті розглянуті особливості синтезу цифрового пропорційно-інтегрального регулятора за умови роботи в головному контурі регулювання швидкості обертання електромеханічної системи. Враховані умови модернізації існуючої системи управління з внутрішніми контурами регулювання, оптимізованими за певними відомими критеріями. Взята до уваги можливість нестаціонарності механічного навантаження електроприводу, а саме – несталість моменту інерції та часу ланки транспортного запізнення. Для реальних параметрів налаштування електроприводу з двигуном постійного струму серії ПБВ та механічним навантаженням, що ідентифікується моделлю об'єкту управління другого порядку з інтегральною складовою та часовою затримкою (FOLIPD) проведений синтез аналогового пропорційно-інтегрального регулятора за критерієм мінімуму інтегральної абсолютної похибки (minimum IAE, Shinskey 1994) з додатковим локальним зворотним зв'язком. За аналоговим прототипом розраховані параметри цифрового регулятора. Отримані модифіковані розрахункові співвідношення для визначення параметрів пропорційно-інтегрального регулятора, досліджені залежності показників якості регулювання від параметрів об'єкту управління.

Практична цінність запропонованої методики підтверджена результатами комп›ютерного моделювання з використанням програмного продукту MATLAB Simulink. Отримані результати показали відповідність якості регулювання в системі цифровим регулятором системі з аналоговим прототипом.

Розглянутий метод розрахунку параметрів зовнішнього контуру регулювання швидкості електромеханічної системи характеризується простотою, можливістю швидкого налаштування до зміни параметрів об'єкту управління та низькою чутливістю цифрового регулятора до відхилень параметрів об'єкту управління. Результати дослідження можуть бути корисні для синтезу цифрових регуляторів за безперервним аналогом для електроприводів з різноманітним механічним навантаженням.

Ключові слова: *автоматизований електропривод, нестаціонарність об'єкту управління, контур регулювання, пропорційно-інтегральний регулятор, цифровий регулятор, якість регулювання.*

Постановка проблеми. Сучасні електроприводи є основою автоматизації та забезпечення енергоефективності виробничих процесів, пов›язаних з механічним рухом та переміщеннями. Основний тип їх реалізації – комплектні регульовані електроприводи – локальні системи управління електромеханічними об'єктами, які виконують функції стабілізації, стеження та обмеження певних координат електроприводу. Широка номенклатура сучасних електроприводів здатна задовольнити будь-які вимоги з боку виконвчого механізму, технологічного процесу та системи автоматизації вищого рівня.

Вплив на продуктивність, точність, надійність, енергетичні показники промислового обладнання обумовлюють регулятори певних координат електроприводу. Сучасні тенденції автоматизації виробництва передбачають повсебічне впровадження програмованих засобів керування електроприводами, що обумовлює використання цифрових пристроїв як для управління послідовністю операцій, так і для управління

системою зі зворотним зв›язком. У багатьох системах управління електроприводами ці методи використовуються спільно.

Актуальне завдання автоматизації – розробка методів налаштування регуляторів, які задовольняють вимогам технологічного процесу та забезпечують якісні показники перехідних процесів, точності та продуктивності обладнання.

Відомо, що синтез цифрових регуляторів проводиться одним з трьох методів: синтез цифрового регулятора за безперервним аналогом; синтез цифрових регуляторів методом частотних характеристик; синтез цифрових регуляторів прямими аналітичними методами (поліноминальними) [1]. Причому, в умовах модернізації існуючого електрообладнання часто виникає завдання розробки цифрового пристрою регулювання головної координати електроприводу при збереженні існуючої системи управління підлеглими координатами, реалізованого на базі аналогових елементів.

Синтез цифрового регулятора за безперервним аналогом – одна з найпоширеніших у практиці проєктування методик розрахунку цифрових регуляторів. Суть методу полягає в тому, що об›єкт регулювання вважають безперервним і синтез регулятора проводять за безперервною моделлю об›єкта виду методами, відомими з теорії безперервних систем. Отриманий при цьому безперервний регулятор потім апроксимується цифровим. Можливо кілька варіантів апроксимації. Найпростіший спосіб ґрунтується на поданні інтегрального закону регулювання за допомогою методу прямокутників. Очевидно, що такий підхід до синтезу цифрових пристроїв управління електроприводом не завжди дає змогу забезпечити максимально можливе використання властивостей силової частини та переваг мікропроцесорного керування, проте цей метод синтезу, завдяки своїй простоті та використанню добре відомих із теорії безперервних систем методів, знаходить досить широке застосування.

Аналіз останніх досліджень і публікацій. Питання оптимального налаштування регуляторів для електромеханічних об’єктів різноманітного типу залишаються відкритими, оскільки особливості механічного навантаження у багатьох випадках спричиняє складнощі, а інколи і неможливість застосування того чи іншого інженерного методу налаштування. Це зумовлює потребу у розробленні таких підходів, які б дали змогу без проблем налаштовувати регулятори навіть за нелінійних або погано формалізованих математичних моделей об’єктів регулювання.

Проведений в роботі [2] аналіз тенденцій, пов’язаних з інноваціями у сфері розроблення автоматичних регуляторів та базується на кількісному аналізі отриманих за останні 15 років патентів у розвинутих країнах світу за класами міжнародної патентної кваліфікації G05B11/06 (до 40 патентів у рік), G05B11/36 (до 230 патентів на рік), демонструє стабільну світову тенденцію щодо збільшення кількості розробок, пов’язаних із пропорційно-інтегродиференційними (ПІД)-регуляторами. На жаль, ця тенденція не стосується України. Лідером у цьому напрямку є Японія та Китай. Крім того, стабільну динаміку щодо кількості отриманих патентів зберігають США, Південна Корея та країни Західної Європи. Це показує велику зацікавленість науково-технічної спільноти у розробленні пристроїв автоматизації цього класу.

Відомо багато методів налаштування коефіцієнтів ПІД-регулятора, і ця кількість активно поповнюється. Роботи у цій сфері тривають, оскільки розробники прагнуть задовольнити зростаючі потреби щодо якості автоматизації технічних процесів за умови обмеженої кількості факторів впливу та певних апаратних обмежень сучасних програмованих засобів.

Так, в роботі [2] показана розробка методу оптимального налаштування пропорційно-інтегрального (ПІ) регулятора, який дає змогу врахувати вимоги щодо обмежень у процесі регулювання та забезпечити мінімізацію небажаних показників якості автоматичного регулювання. Розглянуто лише ПІ-регулятор, оскільки саме такий тип регулятора є найпоширенішим на практиці (застосування диференціальної складової передбачає застосування додаткових фільтрів). Сутність методу полягає у зведенні задачі оптимального налаштування регулятора до задачі безумовної оптимізації узагальненого оптимізаційного критерію, який характеризується бажаними топологічними властивостями. Результати застосування розробленого методу для задач налаштування ПІ-регуляторів, які працюють з об’єктами регулювання другого та третього порядків (зокрема із часовою затримкою), показали значне покращення якості регулювання за обраним критерієм (у межах цього дослідження таким критерієм була середньоінтегральна похибка).

В роботах [3], [4], [5] показано, що налаштування ПІД-регуляторів за різних критеріїв

оптимальності можна відносно просто провести шляхом компенсації нулями передаточної функції регуляторів полюсів електромеханічних об'єктів управління (аперіодичних об'єктів 1-го і 2-го порядків). Проведені в роботах дослідження показали, що при правильному проведенні розрахунків цифровий ПІД-регулятор може конкурувати з аналогічними аналоговими ПІД-регуляторами і може бути використаний для керування електромеханічними об'єктами з достатньо високими вимогами до динамічних процесів регулювання. Так, в роботі [5] ефективність запропонованого методу синтезу автоматичних систем з цифровим ПІД-регулятором потребує вплив на чотири параметри настроювання. При цьому забезпечується найвища якість регулювання і гарантується допустимий рівень похибки.

Постановка завдання. Метою статті є розробка цифрового пропорційно-інтегрального регулятора як пристрою регулювання головного контуру системи управління електроприводом постійного струму з врахуванням внутрішніх контурів, вже налаштованих аналоговими регуляторами на певні критерії оптимізації. Причому, за мету не ставиться розробка нового методу налаштування регулятора електромеханічного об'єкту, а досліджується один з простих та ефективних існуючих методів налаштування з введенням додаткових елементів контуру регулювання та врахуванням особливостей об'єкту управління.

Виклад основного матеріалу. В роботі розглянутий синтез цифрового регулятора для випадку модернізації електроприводу з двигуном постійного струму серії ПБВ, а саме – синтезу головного контуру регулювання у складі програмованої системи управління з цифровим регулятором. Технічні параметри електродвигуна наступні: номінальна напруга якірного кола U_H = 52 В; номінальний струм якоря $I_{\text{Н}}$ = 18 А; номінальна швидкість обертання n_H = 1000 об/хв.; момент інерції якоря J_A = 0,01 кг/м^2; активний опір якірного кола R_A = 0,222 Ом; індуктивність якірного кола L_A = 1,18 мГн; електромагнітна стала часу T_E = 6,3 мс; електромеханічна стала часу T_{EM} = 10,3 мс; конструктивна стала ЕРС двигуна $c\Phi$ = 0,4298 Вс/рад. ($k_{\text{Д}}$ = 2,327 рад/В.с).

Враховано, що контур струму якоря в існуючий системі управління реалізовано на базі аналогових елементів комплектного силового перетворювача і оптимізовано за модульним критерієм. На структурній схемі системи управління (рис. 1) оптимізований контур, згідно традиційної методики налаштування [6], показаний ланкою, яка характеризується коефіцієнтом зворотного зв'язку контуру струму $k_{\text{Т}}$ (визначається величиною напруги завдання на вході контуру для забезпечення обмеження струму на рівні 2,5 кратного номінального); коефіцієнтом зворотного зв'язку контуру швидкості $k_{\text{С}}$ (визначається величиною напруги завдання на вході контуру для забезпечення максимальної швидкості обертання $n_{\text{МАХ}}$ = 2000 об/хв.); малою некомпенсованою сталою часу контуру швидкості $T_{\mu C}$, яка функціонально пов'язана з малою некомпенсованою сталою часу контуру струму і в цілому визначає швидкодійність контуру регулювання. В контурі регулювання врахована ланка чистого запізнювання з сталою часу τ_M, що обумовлено особливостями механічної трансмісії навантаження та властивостями датчика швидкості.

Таким чином, об'єкт управління ідентифікуємо передавальною функцією

$$G(p) = \frac{K_M}{(T_M p + 1)\, p} \exp\left(-p\tau_M\right), \qquad (1)$$

де $K_{\text{М}}$ – пропорційна складова передаточної функції об'єкту управління, $K_M = k_C / (k_T J_\Sigma k_Д)$;

$T_{\text{М}}$ – стала часу об'єкту управління, в нашому випадку мала некомпенсована стала часу контуру регулювання швидкості.

Розрахункова схема контуру регулювання показана на рис. 2.

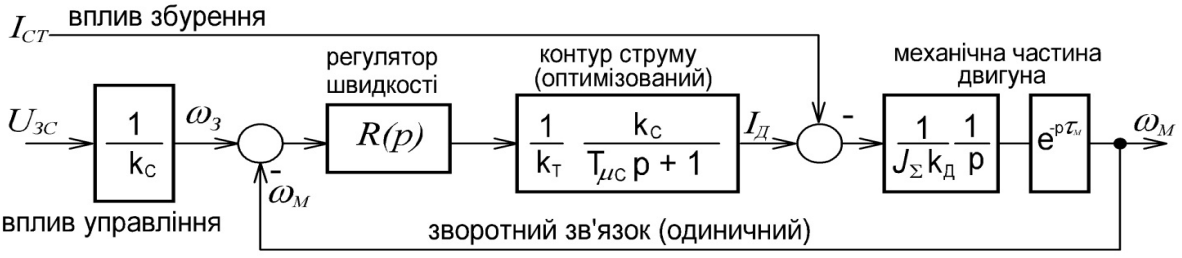


Рис. 1. Вихідна структурна схема контуру регулювання швидкості електромеханічної системи

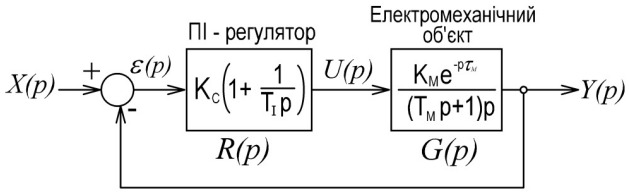


Рис. 2. Розрахункова структурна схема контуру регулювання швидкості електромеханічної системи з ПІ-регулятором

В системі управління пропонується використання пропорційно-інтегрального регулятора (ПІ-регулятора, як різновиду ПІД-регулятора), оскільки такий тип регулятора є найпоширенішим на практиці і простим в реалізації (застосування диференціальної складової вимагає додаткових елементів з метою можливості технічної реалізації регулятора). Крім того, згідно рекомендацій [6] при співвідношенні динамічних характеристик об›єкту $\tau_M / T_M < 0{,}5$ стандартний ПІ-регулятор в цілому здатний реалізувати більшість бажані критерії налаштування.

Передавальна функція аналогового ПІ-регулятору згідно розрахункової структурної схеми:

$$R(p) = K_C \frac{T_I p + 1}{T_I p}. \qquad (2)$$

В роботі [8] для подібного об'єкту управління другого порядку з ланкою чистого запізнювання типу FOLIPD запропоновано більш ніж 70 методів налаштування ПІД-регуляторів за різними критеріями. З представлених для заданих умов розробки контуру регулювання обираємо метод з забезпеченням мінімального значення інтегралу абсолютної похибки (Minimum IAE) Shinskey (1994), згідно з яким розрахункові співвідношення параметрів регулятора наступні.

$$K_R = \frac{K_C}{T_I} = \frac{0{,}15027}{K_M T_M^2 \left(1+k\right)^2}, \quad T_I = 3{,}7 \cdot T_M \left(k+1\right), \qquad (3)$$

де $k = \tau_M / T_M$ – коефіцієнт співвідношення динамічних параметрів об›єкту управління.

Для проведення подальшого комп'ютерного моделювання системи управління, в таблиці 1 представлені розраховані параметри регулятора за умови несталого часу запізнювання τ_M та змінювання приведеного до валу двигуна моменту інерції J_Σ в межах (0,02 … 0,055) кг/м^2.

Інші параметри об'єкту управління вважаються незмінними та визначеними у відповідності з наданими параметрами електромеханічної системи та врахуванням особливостей виконавчого механізму: k_C = 0,0478; k_T = 0,111; J_Σ = 0,02 кг/м^2; K_M = 9,253; $T_{\mu C}$ = T_M = 15 мс. Розрахунки параметрів регулятора проводилися за рівняннями (3).

Результати моделювання – перехідні функції системи за умови різних значень часу запізнювання (на рисунку 3 показані три характеристики для τ_M = 1; 2; 5 мс. у розрахованому діапазоні) демонструють задовільну якість основних параметрів та динамічних властивостей системи з налаштованим регулятором, проте істотним недоліком є значна величина перерегулювання (до 50%), що в більшості випадків є неприпустимим.

Покращення параметрів перехідної функції потребує проведення додаткових заходів, найпростіше з яких – введення додаткового локального зворотного зв'язку між регулятором та електромеханічним об'єктом, як показано на рис. 4. При використані в колі зворотного зв'язку пропорційної ланки $W_{01}(p) = f$ можливе значне зниження перерегулювання при незначному збільшенні часу регулювання, як це демонструють результати моделювання, показані на рис. 5.

Для можливості подальшого аналізу властивостей системи управління визначимо передавальну функцію замкненого контуру регулювання з розробленим ПІ-регулятором:

Таблиця 1

Розраховані параметри комп'ютерної моделі за умови несталих часу запізнювання та моменту інерції

Несталість τ_M, J_Σ = 0,02 кг/м^2			**несталість J_Σ, τ_M = 5 мс**			
τ_M, м/с	K_c	T_I, с	J_Σ, кг/м^2	K_M	K_c	T_I, с
5	3,00443	0,074	0,02	9,25000	3,00541	0,074
4	3,16256	0,0703	0,025	7,40000	3,75676	0,074
3	3,33826	0,0666	0,03	6,16667	4,50811	0,074
2	3,53462	0,0629	0,035	5,28571	5,25946	0,074
1	3,75554	0,0592	0,04	4,62500	6,01081	0,074
0,5	3,87669	0,05735	0,05	3,70000	7,51351	0,074
0,25	3,94024	0,05643	0,055	3,36364	8,26486	0,074

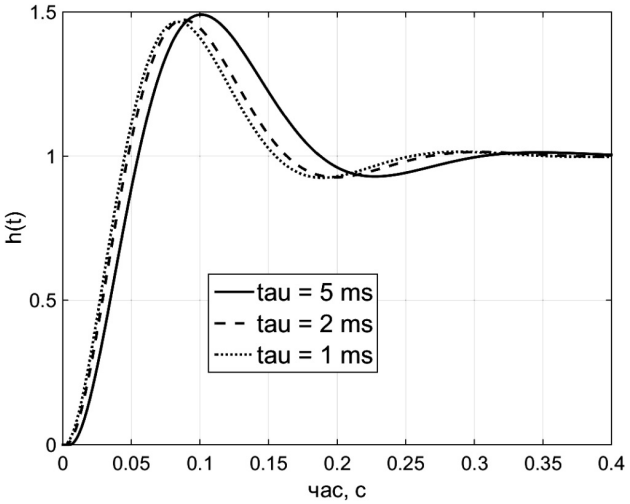


Рис. 3. Перехідні функції контуру регулювання швидкості електромеханічної системи з аналоговим ПІ-регулятором

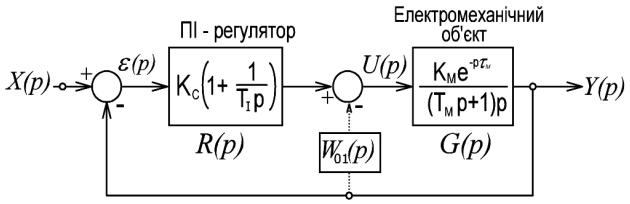


Рис. 4. Структурна схема контуру регулювання швидкості електромеханічної системи з аналоговим ПІ-регулятором та локальним зворотним зв'язком

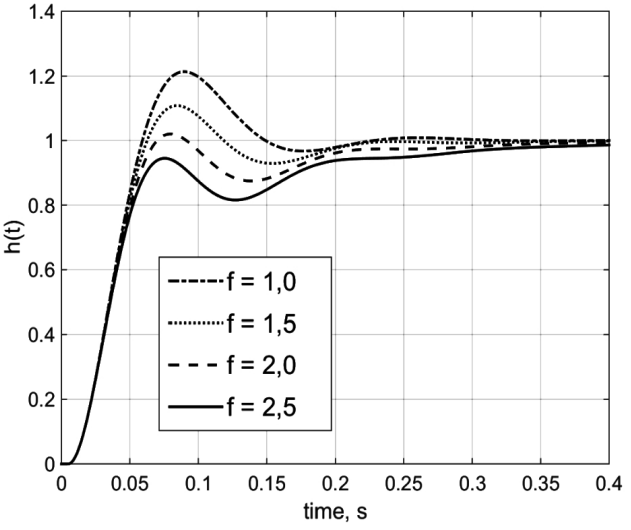


Рис. 5. Перехідні функції контуру регулювання швидкості електромеханічної системи з аналоговим ПІ-регулятором та локальним зворотним зв'язком

$$\Phi(p)=\frac{R(p)G(p)}{1+W_{01}(p)G(p)+R(p)G(p)}, \qquad (4)$$

або після перетворень

$$\Phi(p)=\frac{K_M K_R\left(T_I p+1\right)\exp\left(-p\tau_M\right)}{p^2\left(T_M p+1\right)+K_M\exp\left(-p\tau_M\right)\left(pK_R T_I+pf+K_R\right)}. \qquad (5)$$

Для зручності розрахунків параметрів та характеристик можливо апроксимувати передавальну функцію ланки чистого запізнювання рядом Пада:

$$e^{-p\tau_M}=\frac{a_0p^2-a_1p+1}{a_0p^2+a_1p+1}, \qquad (6)$$

де $a_0=\tau_M^2/12,\ a_1=\tau_M/2$.

Синтез цифрового регулятора контуру швидкості проводимо по методу перетворення безперервного регулятора – аналогу. Використовуємо метод трапецій (Tastin). Час квантування за часом T_0 обираємо меншим кратним часу ланки чистого запізнювання τ_M = 5 мс. Для порівняння визначимо передаточну функцію цифрового регулятора для випадків T_0 = 2,5 мс; T_0 = 1,0 мс; T_0 = 0,5 мс. Отримані передаточні функції:

$$R(z)_{2,5}=\frac{3{,}055z-2{,}954}{z-1}.$$

$$R(z)_1=\frac{3{,}025z-2{,}984}{z-1}.$$

$$R(z)_{0,5}=\frac{3{,}015z-2{,}994}{z-1}.$$

Подальше моделювання системи управління за визначеними параметрами цифрового регулятора (рис. 6) показують цілком задовільну відповідність цифрового регулятору аналоговому прототипу навіть при часі квантування T_0 = 2,5 мс. При цьому коефіцієнт локального зворотного зв'язку прийнятий f = 1,2.

Перерегулювання вихідної величини не перевищує 20%, час встановлення – не перевищує 250 мс. Обраний метод налаштування ПІ-регулятора придатний для швидкого синтезу цифрових регуляторів для розглянутого типу об'єкту управління.

Далі розглянутий вплив на динаміку системи управління з цифровим регулятором зміни часу запізнювання τ_M та змінювання приведеного до валу двигуна моменту інерції J_Σ без переналаштувння параметрів регулятора. Регулятор налаштований на базові параметри механічного навантаження – τ_M = 5 мс, J_Σ = 0,02 кг·м^2 На рис. 7 представлені перехідні характеристики у випадку зміни часу запізнювання τ_M від 1 мс до 7 мс; на рисунку 8 – аналогічні характеристики при варіюванні приведеного моменту інерції в межах 0,02–0,035 кг·м^2.

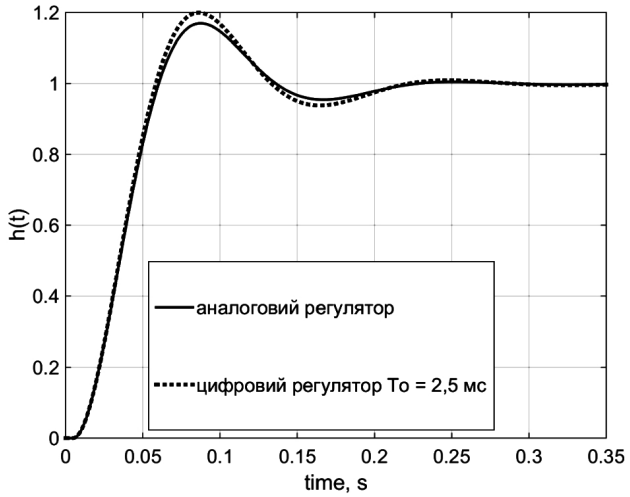


Рис. 6. Перехідні функції контуру регулювання швидкості електромеханічної системи з цифровими та аналоговим регуляторами

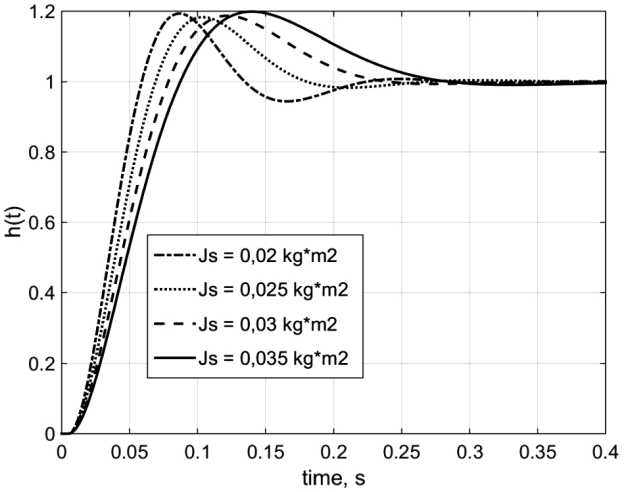


Рис. 8. Перехідні функції контуру регулювання швидкості електромеханічної системи з цифровим регулятором при несталому моменті інерції

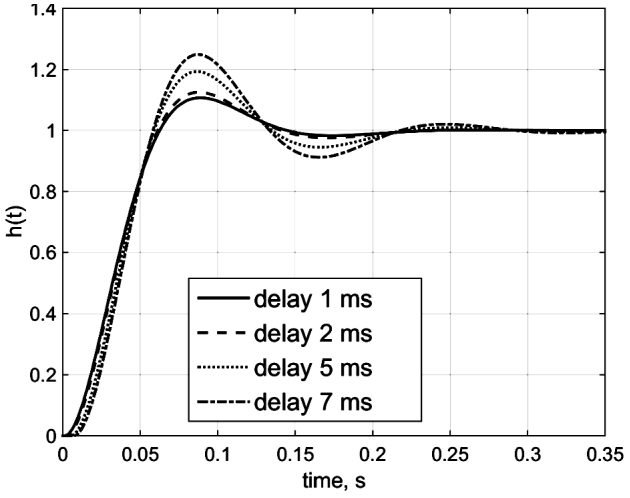


Рис. 7. Перехідні функції контуру регулювання швидкості електромеханічної системи з цифровим регулятором при несталому часі запізнювання

Варіювання часу запізнювання передбачуване пропорційно впливає на величину перерегулювання, та в межах реальної нестабільності часу запізнювання перерегулювання знаходиться в межах (15–22)%. При зростанні моменту інерції зростає час першого узгодження, але перерегулювання та час регулювання залишається практично незмінними. Тобто, розроблений регулятор має низьку чутливість до нестабільності параметрів об'єкту регулювання.

Висновки. В роботі були розглянуті методи розрахунку ПІ-регулятора для роботи в зовнішньому контурі регулювання швидкості електромеханічної системи з двигуном постійного струму за умови наявності оптимізованого внутрішнього контуру, який виступає в якості об'єкта управління. В об'єкті управління врахована наявність ланки чистого запізнювання.

Для вибраного методу Minimum IAE Shinskey (1994) з додатковим введенням локального зворотного зв'язку була визначена передаточна функція замкненого контуру регулювання, отримані модифіковані розрахункові співвідношення для визначення параметрів аналогового регулятора з подальшим його перетворенням у цифровий формат. Проведене комп'ютерне моделювання доводить відповідність вибраного методу поставленому завданню. Отримані результати можуть бути корисними при модернізації існуючих локальних систем управління електроприводами.

Список літератури:

1. Benjamin C. Kuo. Digital Control Systems (The Oxford Series in Electrical and Computer Engineering). Oxford University Press; 2nd edition (June 1, 1995). 784 p.

2. Ромасевич Ю.О., Ловейкін В.С., Ляшко А.П., Макарець В.В. Розроблення методу оптимального налаштування ПІ-регуляторів. Автоматизація виробничих процесів у машинобудуванні та приладобудуванні. 2019. Вип. 53. С. 56–65.

3. Лісовець С.М. Використання цифрового ПІД-регулятора в контурі керування електромеханічним приводом промислового робота. Інформатика, обчислювальна техніка та автоматизація. Вчені записки ТНУ імені В.І. Вернадського. Серія: Технічні науки. 2023. Том 34 (73) № 5 С. 161–165.

4. Ковела І.М., Іванюк О.О. Інженерна методика синтезу автоматичних систем з ПІД-регуляторами. Вісник Національного університету «Львівська політехніка» «Вимірювальна техніка та метрологія». 2008. № 69. С. 43–51.

5. Ковела І.М., Наконечний А.О. Синтез автоматичних систем на основі цифрового ПІД-регулятора з чотирма параметрами настроювання. Вісник Національного університету «Львівська політехніка» «Комп›ютерні науки та інформаційні технології». 2011. № 719. С. 27–36.

6. Бурик М.П. Електромеханічні системи підпорядкованої структури з властивостями слабкої чутливості до параметричних та координатних збурень: монографія. Київ: КПІ ім. Ігоря Сікорського, 2021. 268 с.

7. Лорія М.Г., Поркуян О.В., Ананьєв М.В., Целішев О.Б. Оптимальні настроювання регуляторів промислових систем управління технологічними об'єктами: монографія. Сєвєродонецьк: вид-во СНУ ім. В.Даля, 2019. 192 с.

8. Handbook of PI and PID controller tuning rules (3-nd Edition) / Aidan O*Dwyer/ – Ireland: Imperial College Press, 2009. 623 p.

9. Solving control engineering problems with MATLAB (MATLAB curriculum series) / K. Ogata. – Englewood Cliffs, NJ: Prentice Hall, 2014.5 27 p.

Semenets D.A., Vasylchuk D.P., Arestova T.V. DIGITAL PROPORTIONAL-INTEGRAL CONTROLLER FOR A SECOND-ORDER ELECTROMECHANICAL OBJECT WITH A DELAY

The article deals with the peculiarities of synthesizing a digital proportional-integral controller for operation in the main speed control loop of an electromechanical system. The conditions for modernizing an existing control system with internal control loops optimized according to certain known criteria are taken into account. The possibility of non-stationarity of the mechanical load of the electric drive, namely, the instability of the moment of inertia and the time of the transport delay link, was taken into account. For the actual tuning parameters of the electric drive with a DC motor of the PBV series and a mechanical load identified by the model of a second-order control object of the type with an integral component and time delay (FOLIPD), an analog proportional-integral controller was synthesized according to the criterion of minimum integral absolute error (minimum IAE, Shinskey 1994) with additional local feedback. The parameters of the digital controller were calculated using the analog prototype. The modified design relations for determining the parameters of the proportional-integral controller were obtained, and the dependencies of the control quality indicators on the parameters of the control object were investigated.

The practical value of the proposed methodology is confirmed by the results of computer modeling using the MATLAB Simulink software product. The obtained results showed that the quality of control in the system with a digital controller is comparable to the system with an analog prototype.

The considered method for calculating the parameters of the external speed control loop of an electromechanical system is characterized by simplicity, the ability to quickly adjust to changes in the parameters of the control object, and the low sensitivity of the digital controller to deviations in the parameters of the control object. The results of the study can be useful for the synthesis of digital controllers by continuous analog for electric drives with various mechanical loads.

Key words: *automated electric drive, non-stationarity of the control object, control loop, proportional-integral controller, digital controller, control quality.*

УДК 519.876.2
DOI https://doi.org/10.32782/2663-5941/2025.3.2/46

Ситніков О.В.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Лукінюк М.В.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Фтемов А.О.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

СИНТЕЗ СИСТЕМИ КЕРУВАННЯ ВОДООЧИЩЕННЯ ЗА ДОПОМОГОЮ МОДУЛЬНОГО КОНТЕЙНЕРА СИСТЕМИ ЗВОРОТНОГО ОСМОСУ

У статті приведено основні положення по створенню системи керування та принцип програмування блоку регулятора для модульного контейнера системи зворотного осмосу, що встановлений на південному гірничо-збагачувальному комбінаті. В наш час переважна більшість природних джерел прісної води мають статус непридатних для вживання людиною в якості питної. Виходячи з того, що чиста вода для споживання становить головну потребу людини та, власне, відповідає за можливе існування, то проблема відсутності надійного водопостачання потребує нагальних та кваліфікованих рішень. Сучасні можливості представлені безліччю методик очищення води. В роботі представлено автоматизацію процесу очищення з використанням установки у вигляді модульного контейнера із системою зворотного осмосу, яка забезпечує якісний продукт у вигляді очищеної, пом'якшеної та демінералізованої питної води. Проведений аналіз стану виробництва, виявив, що практично на всіх виробництвах застосовано ПІД-регулятор, а принципи керування із використанням системи з МРС-регулятором не застосовується. У свою чергу, МРС-регулювання – тип регулятора, що функціонує за допомогою передбачення поведінки системи в тих чи інших умовах роботи – використовується в промисловості, коли потрібно застосувати штучний інтелект для передбачення різних ситуацій у перебігу технологічного процесу. В роботі об'єктом дослідження виступає модульний контейнер системи зворотного осмосу, що встановлена на низці вітчизняних та зарубіжних виробничих підприємств. Програмування логічного реле та синтез системи керування із реалізацію МРСрегулювання є основними задачами досліджень. Розглянуто модель мембрани зворотного осмосу типу CSM RE8040-BLF. Після отриманих результатів моделювання було прийнято рішення про застосування в дослідженнях МРСрегулятора. В проведених дослідженнях він використовувався задля точнішої та швидшої реакції системи на зміни під час перехідного процесу роботи. Синтез, програмування та налаштування регулятора проведено із використанням середовища MATLAB Simulink.

Ключові слова: *система керування, мембрана, фільтрація, вимірювальний прилад, клас точності, програмування, програмне середовище.*

Постановка проблеми. Під сучасну пору більшість природних джерел прісної води мають статус непридатними для вживання людиною в якості питної. Деякі з них навіть не можуть мати статусу для використання в технічних потребах. Оскільки чиста вода для споживання становить одну з найважливіших потреб людини, що напряму впливає на здоров'я та благополуччя життєвого циклу людини, то проблема відсутності надійного водопостачання потребує нагальних та кваліфікованих рішень. Сучасні технології передбачають безліч методів очищення води [1]. Для вирішення питання очищення води передбачені санітарно-токсикологічні, органолептичні та фізико-хімічні показники безпечності та якості питної води. Очищення води є однією зі сфер діяльності людини, що відповідно вимагає максимальної автоматизації процесу. В процесі очищення найперспективнішим варіантом виступає установка у вигляді модульного кон-

тейнера із системою зворотного осмосу, яка забезпечує якісний продукт у вигляді очищеної, пом'якшеної та демінералізованої питної води. Головними чинниками, на які потрібно звертати увагу в системі зворотного осмосу, є концентрація домішків ув очищеній воді та продуктивність [2]. Остання визначається витратою пермеату на виході з установки. Для зниження концентрації домішків у пермеаті використовується декілька ступенів фільтрації, основою в яких виступають мембрани зворотного осмосу. Модульний контейнер зворотного осмосу є системою, об'єктами виступають зворотноосмотичні блоки. Відповідно ці блоки підлягають автоматизації.

Аналіз останніх досліджень і публікацій. Стан питання керування системою зворотного осмосу представлений на виробництві із використанням ПІД-регуляторів [3–4]. ПІД-регулятор (*PID-controller*) – це базовий регулятор для будь-якої промисловості. Він дозволяє керувати системою так, щоб вона мала змогу реагувати на будь-яке збурення та утримувати ~~сигнал~~ регульований параметр на відповідному рівні [5]. Виходячи з розглянутих матеріалів, виявлено, що не було розглянуто принципи керування із використанням системи з *MPC*-регулятором [6–7]. *MPC*-регулятор – тип регулятора, що функціонує за допомогою передбачення поведінки системи в тих чи інших умовах роботи та використовується в промисловості, коли потрібно застосовувати штучний інтелект для передбачення різних ситуацій перебігу технологічного процесу [8].

Постановка завдання. Об'єктом дослідження виступає модульний контейнер системи зворотного осмосу, що встановлена на низці вітчизняних виробничих та зарубіжних *виробничих* підприємств. Програмування логічного реле *Siemens LOGO* та синтез системи керування із реалізацію *MPC*-регулювання виступають основними задачами досліджень.

Виклад основного матеріалу. Розрахункову схему зворотноосмотичного блоку представлено на рис. 1.

На схемі представлено наступні складові:

Nv – кількість мембранотримачів;

N – кількість елементів (мембран);

S – площа елемента (мембрани);

*Pif*1,.. *Pif*6 – осмотичний тиск вхідної води для елемента 1–6;

*Pip*1,.. *Pip*6 – осмотичний тиск пермеату для елемента 1–6; у «-» не може бути пробілів

Sel – селективність мембрани;

dPfc – перепад тиску системи по концентрату;

Cfc – концентрація протоку за концентратом;

Cf – концентрація протоку за вхідною водою;

Pp – тиск пермеату;

T – температура води;

Ys – відновлення системи (гідравлічний ККД системи);

FF – фактор забруднення мембрани;

TCF – фактор температурної поправки;

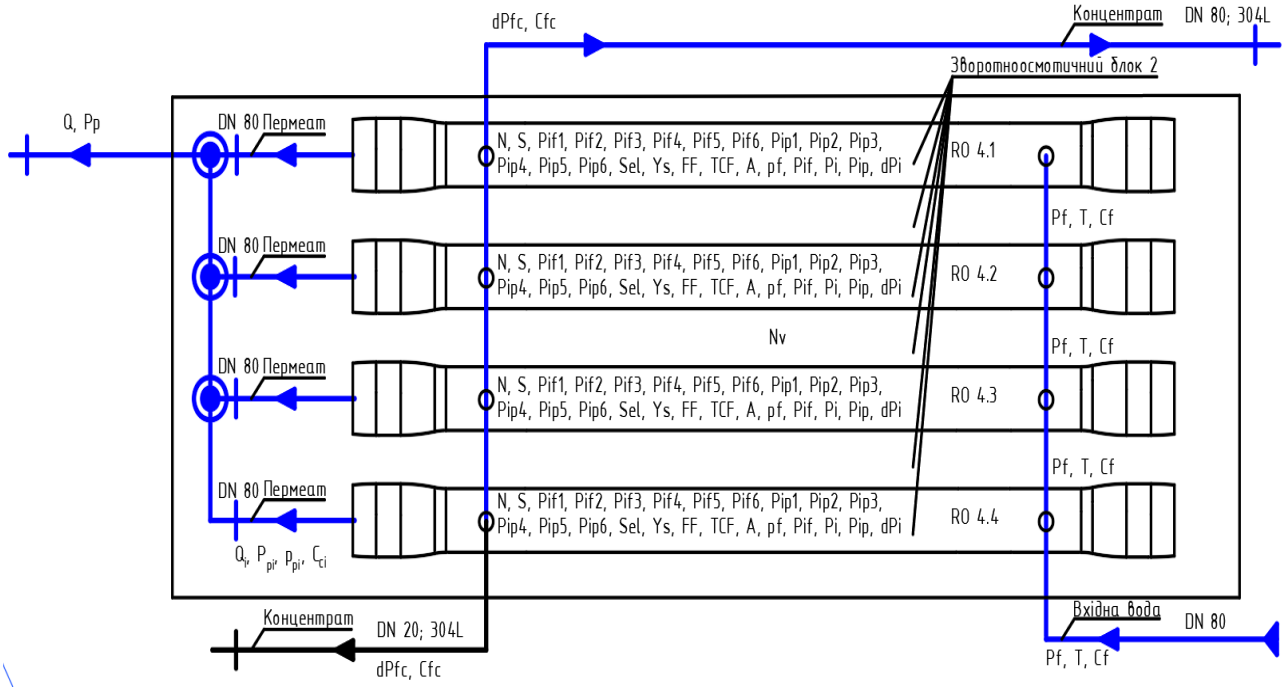


Рис. 1. Розрахункова схема зворотноосмотичного блока

A – проникність мембрани;

pf – концентраційна поляризація;

Pif – осмотичний тиск системи за вхідною водою;

Pi – осмотичний тиск системи за концентратом;

Pip – осмотичний тиск системи за пермеатом;

dPi – осмотичний перепад тиску системи;

Q – витрата пермеату;

Pf – тиск вхідної води.

В роботі представлено модель мембрани зворотного осмосу типу *CSM RE*8040-*BLF*, розрахунки проведено за формулами виробника мембран *CSM*, а фактор температурної поправки за формулою компанії *Dupont*. Оскільки математичні моделі містять у собі необхідні константи, що працюють правильно лише для вимірювальних одиниць тиску у барах і загальноприйнята практика показує, що більшість розрахунків для систем зворотного осмосу ведуться в таких одиницях, їх виконано саме у такий спосіб. Фактор температурної поправки для температури ≤ 25°C прийнято 0,48. Реалізація моделі виконана в програмному середовищі *MATLAB* [9–10]. Модель використана для синтезу системи керування.

На виробництві для системи зворотного осмосу найчастіше використовуються одноконтурні системи керування з ПІД-регулятором. Широкого використання набули регулятори, що використовуються в програмованому логічному контролері з лінійки *Siemens Simatic S*7. Серед вимірювальних приладів можна віднести серію МТР-138 від вітчизняної компанії «Мікрол». В роботі проведено налаштування регулятора за допомогою програмування функцій *Matlab Simulink* у блоці *PID-controller*. Згідно з налаштуваннями програмного середовища можна зробити висновок про параметри налаштування контролера.

Після отримання результатів було прийнято рішення про застосування в дослідженнях *MPC*-регулятора. МРС-регулятор – тип новітнього регулятора, що функціонує за допомогою передбачення поведінки системи в тих чи інших умовах роботи. Часто такий тип регулятора застосовують у автомобільній промисловості, коли потрібно використовувати штучний інтелект для передбачення різних ситуацій на дорозі – від правильного керування авто до передбачення дорожньо-транспортної пригоди. В роботі він буде використовуватися задля точнішої та швидшої реакції системи на зміни під час перехідного процесу роботи. Вони виготовляються такими компаніями, як *Siemens* та *Schneider Electric*.

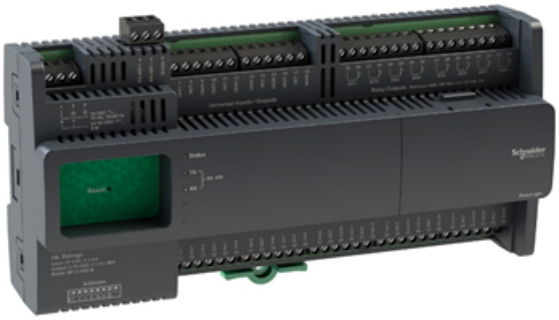


Рис. 2. МРС-регулятор *EasyLogic Schneider Electric*

Слід зазначити, що зазначений пристрій використовується у ПЛК від компанії *Siemens*, таких як *SIMATIC S*7-1500, а також його застосовують у компанії *Tesla*. *MPC*-регулятор для правильного регулювання користується лінійними моделями установки та збурень. Після визначення даних моделі він прогнозує вихідні сигнали та розв'язує задачу оптимальності квадратичного програмування. Працює регулятор таким чином, що сигнал поділяється на певну кількість кроків у часі та під час регулювання обирає на кожному кроці потрібне оптимальне значення, на яке необхідно вийти залежно від поведінки системи у відповідному періоді часу.

До параметрів *MPC*-регулятора входять: час дискретизації, горизонт прогнозування та керування, номінальні значення, масштабні коефіцієнти, ва́ги.

Синтез, програмування та налаштування проведено у середовищі *MATLAB Simulink*.

Проведено визначення структури регулятора за допомогою *MPC Structure* та проведено первинне налаштування.

Проведено налаштування системи. Оскільки система працює в пікових своїх значеннях (24 бари) з допуском зміни на 1 бар, то значного перерегулювання в системі не допускається. Тому для стабільності системи визначено горизонт прогнозу як 20, а горизонт керування – 5.

В роботі значення параметра *Weight* відповідає за підняття графіка до потрібного значення. Зазвичай воно відповідає номінальному значенню системи для вхідного каналу. *Rate weight* потрібен для вирівнювання сигналу до такого значення, яке буде оптимальним для нашої системи. Перерегулювання не потрібне, через що система починає виходити на необхідне значення повільніше, але стабільніше. *Weight* у вихідного сигналу необхідний для прискорення системи, але з присутнім перерегулюванням, через що потрібно зберігати баланс між вхідними та вихідними вагами.

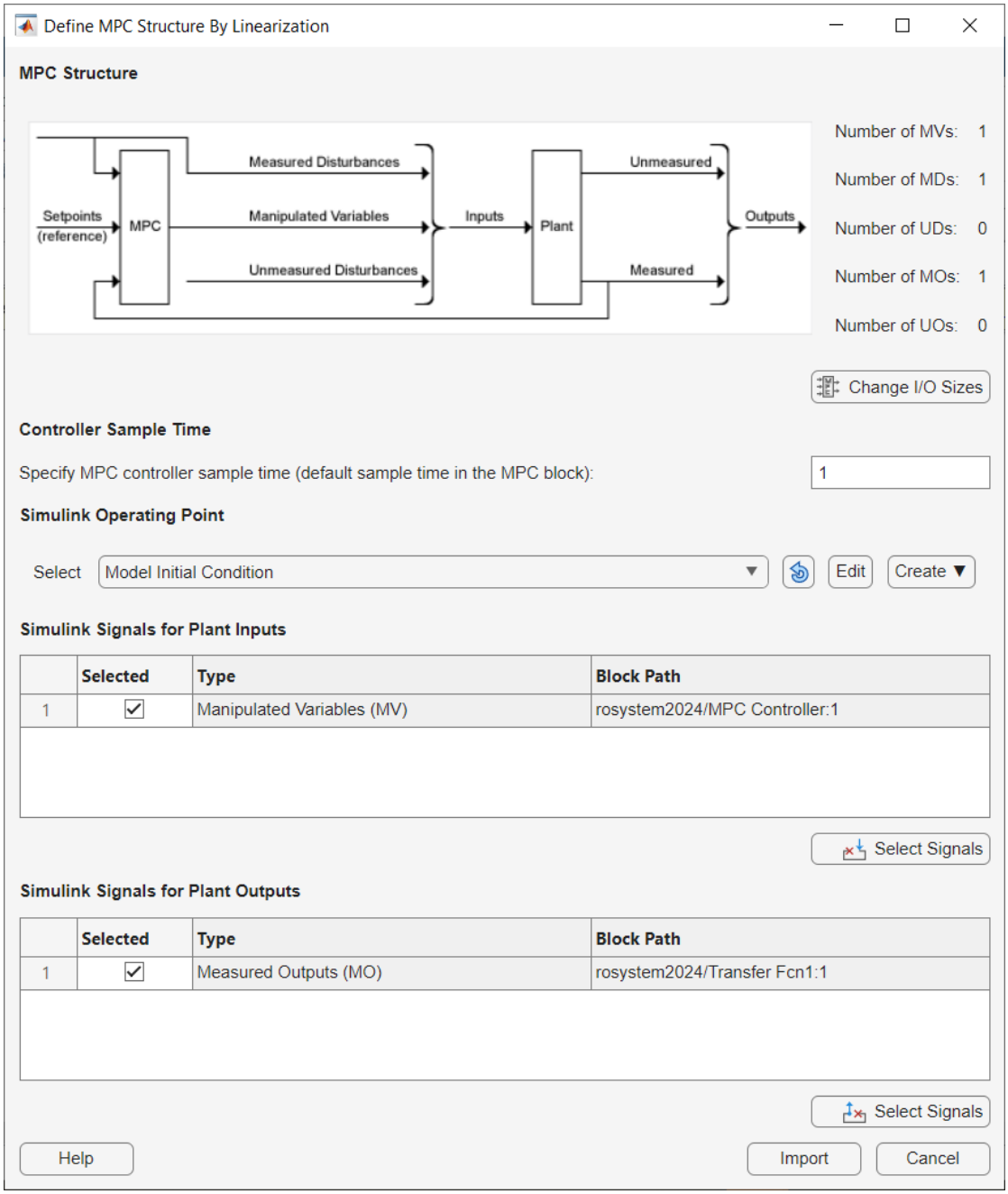


Рис. 3. Первинне налаштування MPC-регулятора в *MATLAB Simulink*

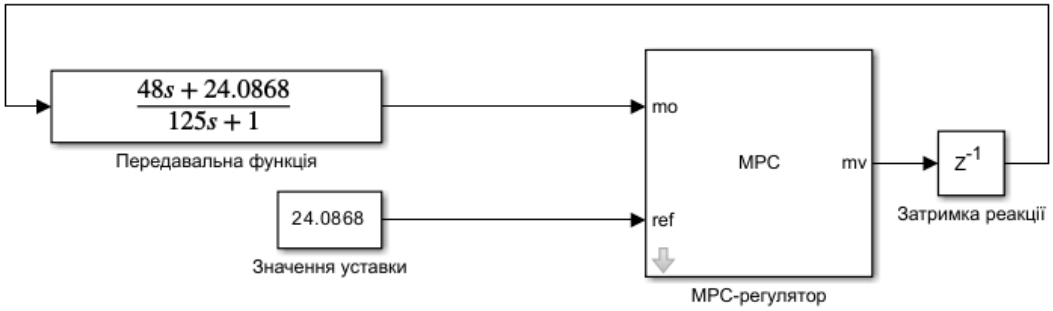


Рис. 4. Система з *MPC*-регулятором

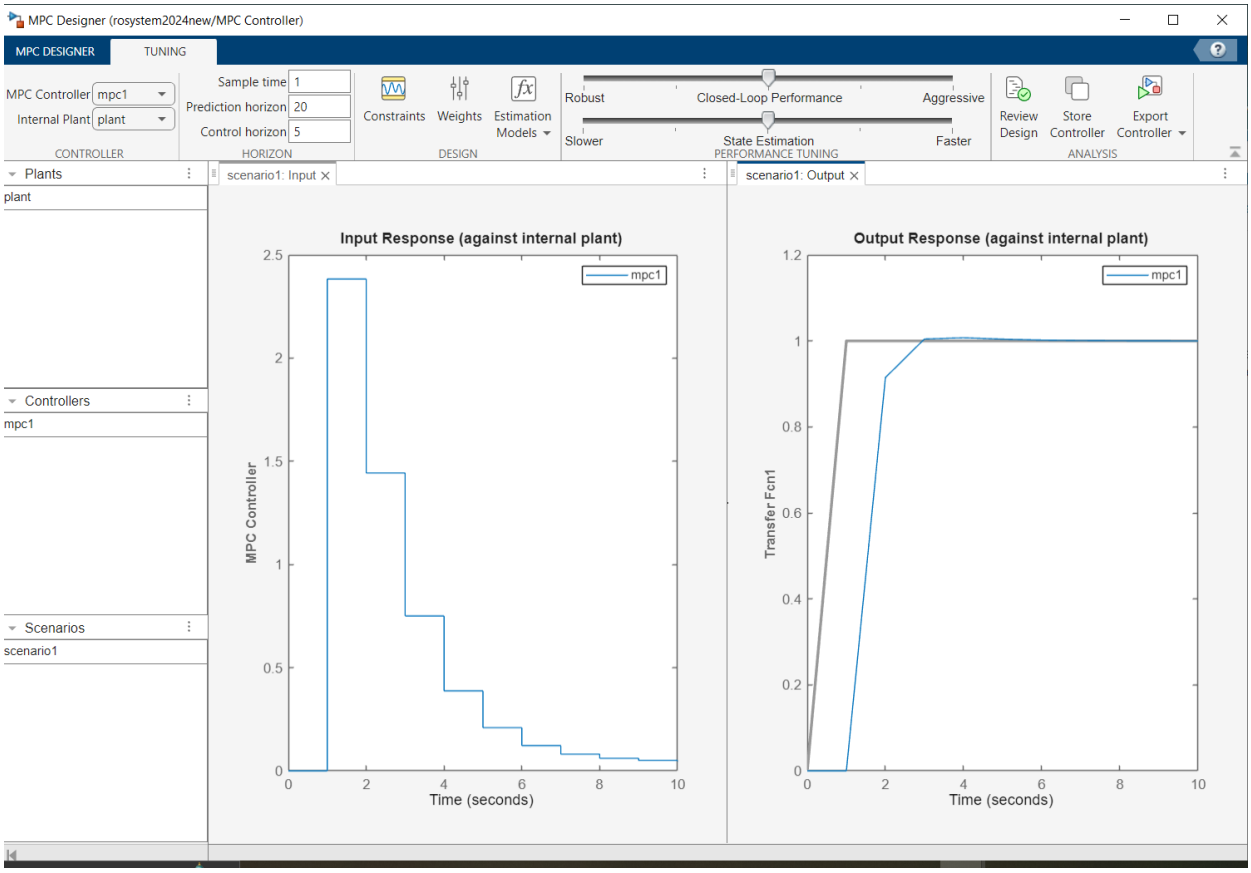


Рис. 5. Налаштування системи та її реакція

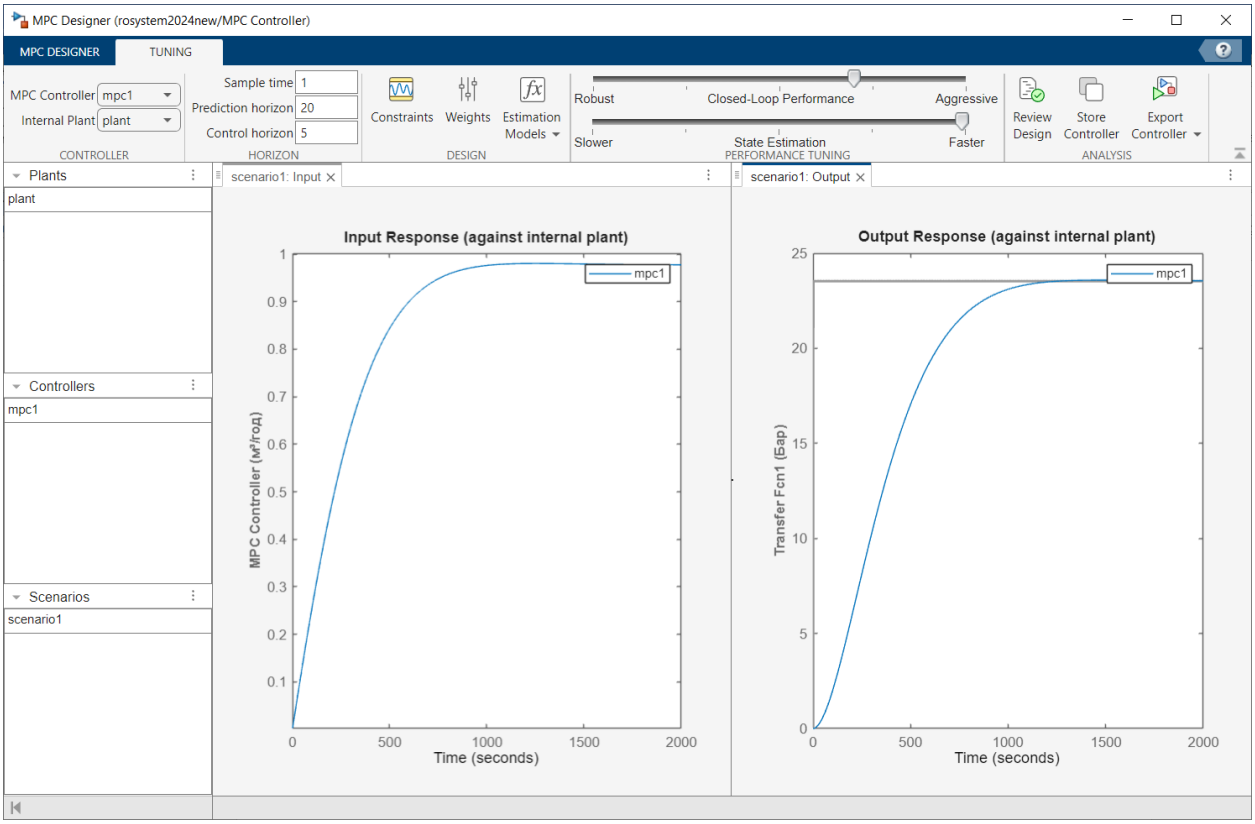


Рис. 6. Перехідна характеристика системи з *MPC*-регулятором

Висновки. Було досліджено модульний контейнер системи зворотного осмосу, розроблений компанією «Нерекс» з імплементацією новітніх технологій програмованого логічного реле *Siemens LOGO*. Синтезовано систему керування із реалізацію *MPC*-регулювання.

Список літератури:

1. Про затвердження Державних санітарних норм і правил «Показники безпечності та окремі показники якості питної води в умовах воєнного стану та надзвичайних ситуаціях іншого характеру» : Наказ МОЗ України від 22.04.2022 № 683. URL: https://zakon.rada.gov.ua/laws/show/z0564-22#Text (дата звернення: 07.05.2025).

2. DuPont de Nemours, Inc. FilmTec™ Reverse Osmosis Membranes Technical Manual, 2023. 211 p.

3. M. S. Islam, A. Sultana, A. H. M. Saadat, M. S. Islam, M. Shammi, M. K. Uddin Desalination Technologies for Developing Countries: A Review URL: https://www.banglajol.info/index.php/JSR/article/view/33179 (дата звернення: 08.05.2025).

4. Jasper van Asselt C.S. Membrane Technology. Sustainable seawater reverse osmosis (SWRO) system design for rural areas of developing countries Engineering & Technology (ET), Enschede: Master Sustainable Energy Technology (SET), June 10, 2021. 27 p.

5. Попович М. Г., Ковальчук О. В. Теорія автоматичного керування : підручник / Рецензент: І. В. Волков, В. І. Сенько, Ю. А. Скрипник. Київ : «Либідь», 2007. 656 с.

6. Lilian Malaeb, George M. Ayoub Reverse osmosis technology for water treatment: State of the art review URL: https://surl.li/fdnwti (дата звернення: 08.05.2025).

7. Ruth Habte Hailemariam**,** Yun Chul Woo, Mekdimu Mezemir Damtie, Bong Chul Kim, Kwang-Duck Park, June-Seok Choi. Reverse osmosis membrane fabrication and modification technologies and future trends: A review. URL: https://surl.li/xeyrcv (дата звернення: 08.05.2025).

8. MPC Signal Types. MathWorks – Maker of MATLAB and Simulink – MATLAB & Simulink. URL: https://surl.li/qcgbig (дата звернення: 7.05.2025).

9. Tune Weights. MathWorks – Maker of MATLAB and Simulink – MATLAB & Simulink. URL: https://www.mathworks.com/help/mpc/ug/tuning-weights.html (дата звернення: 5.05.2025).

10. Фтемов А. О., Ситніков О. В. Синтез системи керування з МРС-регулятором для конденсування продуктів синтезу. // «АКІТ-2024», Київ, 2024. С.71–73

Sytnikov A.V., Lukiniuk M.V., Ftemov A.O. SYNTHESIS OF A WATER TREATMENT CONTROL SYSTEM USING A MODULAR CONTAINER OF A REVERSE OSMOSIS SYSTEM

The article presents the basic provisions for creating a control system and the principle of programming the regulator unit for a modular container of the reverse osmosis system installed at the southern mining and processing plant. Nowadays, the vast majority of natural sources of fresh water have the status of being unsuitable for human consumption as drinking water. Based on the fact that clean water for consumption is the main human need, which is responsible for the very existence, the problem of the lack of reliable water supply requires urgent and qualified solutions. Modern capabilities are represented by a variety of water purification methods. The paper presents the automation of the purification process using a unit in the form of a modular container with a reverse osmosis system, which provides a high-quality product in the form of purified, softened and demineralized drinking water. An analysis of the state of production was conducted, it was found that a PID controller is used in almost all production facilities, and the principles of control using a system with an MPC controller are not applied. In turn, MPC-regulation is a type of regulator that functions by predicting the behavior of the system under certain operating conditions, used in industry when it is necessary to use artificial intelligence to predict various situations of the technological process. In the work, the object of research is a modular container of the reverse osmosis system, which is installed at a number of domestic and foreign manufacturing enterprises. Programming of a logic relay and synthesis of a control system with the implementation of MPC-regulation are the main tasks of the research. The model of the reverse osmosis membrane type CSM RE8040-BLF was considered. After the obtained modeling results, a decision was made to use the MPC-regulator in the research. In the conducted research, it was used for a more accurate and faster response of the system to changes during the transient process of work. The synthesis, programming and tuning of the regulator were carried out using the MATLAB Simulink environment.

Key words: *control system, membrane, filtration, measurement, accuracy class, programming, software environment.*

УДК 004.02
DOI https://doi.org/10.32782/2663-5941/2025.3.2/47

Слабінога М.О.
Івано-Франківський національний технічний університет нафти і газу

Піх В.Я.
Івано-Франківський національний технічний університет нафти і газу

Стрілецький Ю.Й.
Івано-Франківський національний технічний університет нафти і газу

Шекета В.І.
Івано-Франківський національний технічний університет нафти і газу

РОЗПІЗНАВАННЯ ТЕХНІЧНОГО СТАНУ ОБ'ЄКТІВ НА ОСНОВІ ОПРАЦЮВАННЯ ВІБРАЦІЙНИХ СИГНАЛІВ БАГАТОШАРОВИМ ПЕРЦЕПТРОНОМ

У статті розглянуто застосування методів штучного інтелекту, зокрема багатошарових перцептронних мереж, для задачі класифікації технічного стану об'єктів на основі вібраційних сигналів, що є актуальним у контексті цифрової трансформації виробничих процесів. Зазначено, що попереднє оброблення часових рядів даних відіграє важливу роль у підвищенні точності ідентифікації технічного стану. Проаналізовано сучасні підходи, серед яких використання методів опорних векторів, нейронних мереж із генетичними алгоритмами, експертних систем, марківських мереж, а також різних способів попередньої обробки сигналів – вейвлет-перетворення, декомпозиція на емпіричні моди, кепстральний аналіз тощо.

Основна увага приділена практичному експерименту з використанням набору даних, що містить вібраційні сигнали з тестової роторної установки. Дані було кластеризовано методом К-середніх у чотири класи технічного стану: «хороший», «задовільний», «передаварійний» і «аварійний». Побудовано декілька моделей багатошарової перцептронної мережі для розв'язання задачі класифікації із використанням різних методів обробки сигналів: статистичних характеристик (середнє, СКВ, ентропія Шеннона), коефіцієнтів швидкого перетворення Фур'є, кепстрального аналізу та вейвлет-перетворення. Результати показали, що базові статистичні характеристики забезпечують помірну точність (похибка понад 30% для класифікації на 4 класи), тоді як застосування спектральних та вейвлет-перетворень значно підвищує ефективність моделі (похибка знижується до 1–2%). Зокрема, найкращі результати отримано при використанні кепстрального аналізу та вейвлет-перетворення Добеші першого порядку. Продемонстровано, що навіть при зниженні кількості класів (наприклад, зведенні до «нормального» та «аварійного» стану) точність класифікації суттєво зростає, що відкриває можливості для створення надійних систем раннього виявлення несправностей.

У висновках підкреслено доцільність поєднання методів статистичного аналізу з нейромережевими підходами для підвищення точності діагностики технічного стану, а також перспективність дослідження ефективності різних методів попередньої обробки сигналів для зменшення похибки класифікації та оптимізації обчислювальних витрат.

Ключові слова: *перцептрон, опрацювання сигналів, ймовірнісні оцінки, штучні нейронні мережі, ідентифікація, технічний стан.*

Постановка проблеми. У сучасному стані, цифрова трансформація виробничих процесів на промислових об'єктах відкриває нові можливості щодо застосування різноманітних підходів до ідентифікації їх технічного стану. Зокрема, продовжує зростати популярність впровадження методів штучного інтелекту для вирішення таких задач. Задача ідентифікації технічного стану на основі попередніх даних є типовою задачею класифікації. Одним із найпопулярніших алгоритмів на засадах штучного інтелекту є підхід із застосуванням багатошарових перцептронних мереж, який дає можливість виконати ймовірнісну оцінку технічного стану на основі отриманих входів.

Тим не менше, часові ряди даних щодо параметрів технологічного процесу, часто потребують попередньої обробки та застосування різноманітних статистичних методів.

Аналіз останніх досліджень і публікацій. Питанню розпізнавання технічного стану на засадах статистичних методів та алгоритмів штучного інтелекту присвячена значна кількість публікацій. Зокрема, для діагностики технічного стану підшипників в різних процесах застосовувалися методи опорних векторів [1], нечіткої логіки на основі нейромереж [2], марківські мережі [3], нейромережі, оптимізовані на основі генетичних алгоритмів [4], дерева рішень [5], експертні системи [6]. Крім того, важливим фактором було визначення інформації що подавалася на вхід інтелектуалізованих методів, оскільки крім класичних для вирішення задач, пов'язаних із вібраційними процесами, статистичних методів та перетворення Фур'є, були використані такі методи як вейвлет-перетворення [7] та декомпозиція на емпіричні моди [8].

В цілому, варто зазначити, що в залежності від специфіки процесу той чи інший метод статистичного аналізу дає можливість з достатньою точністю розпізнати дані про технічний стан об'єкта.

Постановка завдання. Метою дослідження є дослідження ефективності різних методів опрацювання сигналів для використання результатів опрацювання як входів багатошарових перцептронних мереж у вирішенні задачі розпізнавання технічного стану об'єкта.

Виклад основного матеріалу. Для тестування методів, що досліджують коливні процеси за миттєвими даними, були використані дані зі репозиторію даних для прогнозів NASA, а саме набір вібраційних даних тестової обертової машини Університету Цінціннаті. Ці дані були зібрані з експериментальної установки для вивчення чотирьох підшипників, встановлених на вал обертання з навантаженням близько 2720 кг (6000 фунтів) та зі сталою швидкістю обертання 2000 об/хв. Підшипники Rexnord ZA-2115 були встановлені на валі, як показано на рисунку 1. На основу підшипників були встановлені високочутливі кварцові акселерометри PCB 353B33 (по два на кожен підшипник в експерименті 1, для кожної з осей x та y, по одному на підшипник в експериментах 2 та 3). Розміщення давачів також показано на рис. 1. Всі пошкодження підшипників виникли після перевищення їх експлуатаційного ресурсу, що становив 100 000 000 оборотів.

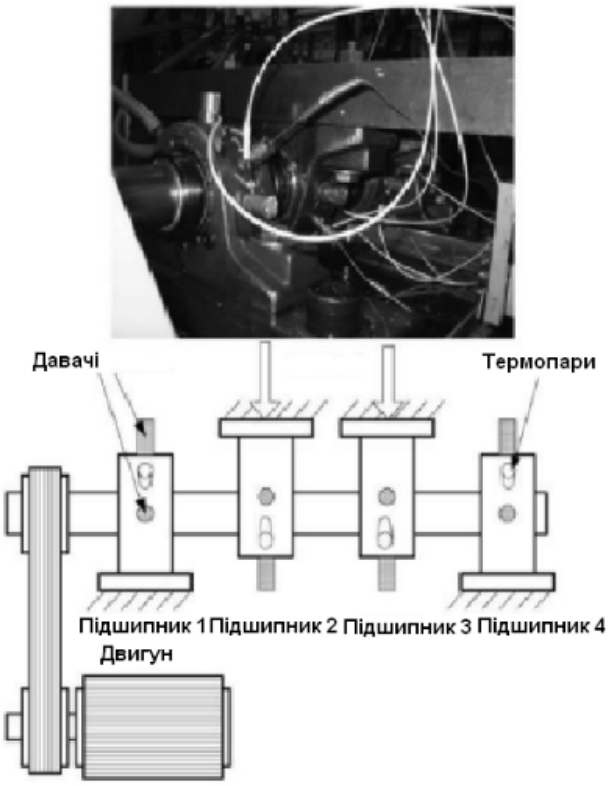


Рис. 1. Розміщення підшипників та вібродавачів

В архіві містилося три набори даних. Кожен із наборів записувався до збою в роботи в системі. Кожен з наборів складається із масиву файлів, які є записом вібраційних даних за одну секунду і записуються з певним інтервалом. Кожен файл складається з 20480 точок із частотою дискретизації 20 кГц. Ім'я файлу визначає також час збору даних.

Для дослідження було вибрано дані експерименту №2, де наприкінці періоду функціонування відбувся збій в роботі роторної машини, пов›язаний з пошкодженням зовнішньої поверхні підшипника 1.

На рис. 2 подано динаміку зміни середньоквадратичного відхилення вібропереміщення в часі, від пуску установки, до відмови.

З рисунка видно, що навіть з такої простої операції, як визначення середньоквадратичного відхилення, може бути отримана інформація для подальшої ідентифікації технічного стану об'єкта.

Для вирішення задачі класифікації початково дані було кластеризовано у 4 групи методом K середніх, де кожен клас відповідає за один із технічних станів: «хороший», «задовільний», «передаварійний» та «аварійний».

Для вирішення задачі розпізнавання технічного стану було вирішено застосовувати тришарові ієрархічні нейронні мережі прямого поши-

Рис. 2. Динаміка зміни середньоквадратичного відхилення віброперемещення підшипника 1

рення, що тренуються за алгоритмом зворотнього поширення похибки.

Навчання методом зворотного поширення помилки такої нейронної мережі припускає два проходження по всіх нейронам мережі – прямий прохід (від вхідного шару до вихідного для отримання результуючого вихідного сигналу) і зворотний прохід (від вихідного шару до вхідного для коригування ваг зв›язків нейронної мережі). Ваги зв›язків між нейронами налаштовуються з метою максимального наближення вихідного сигналу мережі до бажаного в статистичному сенсі.

Нейрони мереж мають сигмоїдну функцію активації. Особливістю сигмоїдної функції є її монотонність і диференційовність у всьому діапазоні значень аргументу. Завдяки цьому сигмоїдна функція часто застосовується при створенні нейронних мереж. Введення функцій сигмоїдального типу було обумовлено обмеженістю нейронних мереж з пороговою функцією активації нейронів – при такій функції активації будь-який з виходів мережі дорівнює або нулю, або одиниці, що обмежує використання мереж не в задачах класифікації. Використання сигмоїдальних функцій дозволило перейти від бінарних виходів нейрона до аналогових.

Крім того, було встановлено наступні правила вибору кількості нейронів в кожному з шарів мережі:

– кількість нейронів у вхідному шарі m рівна кількості вхідних значень, що враховуються одночасно;

– кількість вихідних нейронів n рівна кількості результуючих класів (в даному випадку, кількості можливих станів об’єкта;

– кількість нейронів k у прихованому шарі є числом, що знаходиться в діапазоні між n та m та вибирається експериментальним шляхом, з урахуванням обмежень обчислювальних потужностей ЕОМ.

Відповідно до цього, вхідний шар багатошарового перцептрона містив кількість відповідну кількість входів, що залежала від кількості параметрів для кожного семплу. Проміжний шар перцептронів із сигмоїдною функцією активації опрацьовував отриману з входів інформацію та передавав на вихідний шар із 4 перцептронів (за кількістю поданих розпізнаваних класів. Схема штучної нейронної мережі, що була використана для класифікацією за частотним аналізом (перетворення Фур’є) подана на рис. 3.

Вибірка даних була розбита на набори для навчання (688 значень), валідації (148 значень) та тестування (148 значень).

Загалом, було проведено декілька експериментів із різними методами статистичного аналізу сигналів та навчання перцептронної мережі на основі цих даних. Перелік експериментів подано в таблиці 1.

За результатами першого експерименту (урахування середнього значення, середньоквадратичного відхилення, ентропії Шеннона) було отримано значну похибку класифікації (близько 30%). Розглянемо матрицю змішування класів, подану на рис. 4.

Як видно з рисунку, нейромережа хибно класифікувала вибірки, належні до класу 1 як

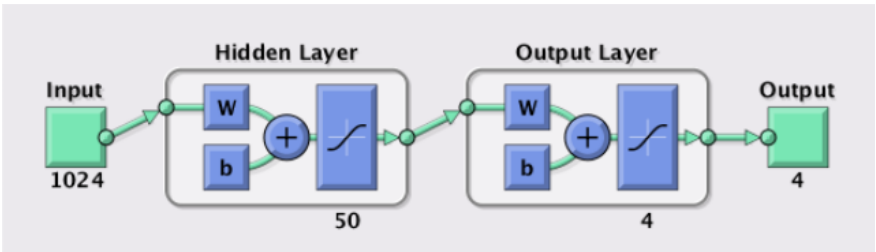


Рис. 3. Схема багатошарової перцептронної мережі для вирішення задачі класифікації

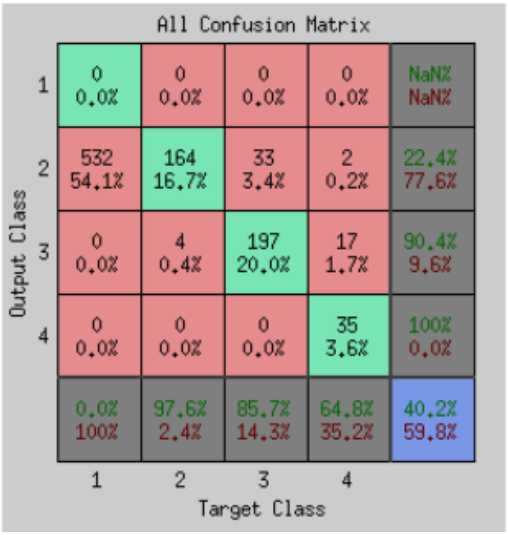


Рис. 4. Матриця змішування класів при класифікації нейромережею на основі статистичних характеристик у експерименті 1

належні до класу 2. Це пов'язано з тим, що статистичні характеристики сигналів у цих двох класах були дуже близькими між собою. Як результат, було проведено додатковий експеримент, де технічний стан об'єкта класифікувався на два класи – нормальний (об'єднання класів 1 та 2) та аварійний (об'єднання класів 3 та 4). Експеримент показав, що така класифікація має похибку близько 5%.

Для розпізнавання технічного стану за коефіцієнтами швидкого перетворення Фур'є застосовувалася штучна нейронна мережа з 1024 входами (для кожної із 1024 компонент), прихованим шаром із 50 нейронів та вихідним шаром із 4 нейронів, що відповідали результуючим класам.

Для розпізнавання технічного стану за коефіцієнтами кепстрального аналізу застосовувалася штучна нейронна мережа з 1024 входами (для кожної із 1024 компонент), прихованим шаром із 50 нейронів та вихідним шаром із 4 нейронів, що відповідали результуючим класам.

Для оцінки технічного стану з використанням коефіцієнтів неперервного вейвлет-перетворення використовувалась штучна нейронна мережа з M входами (M – кількість масштабувань, в даному експерименті – 100), кожен вхід є значенням ентропії за Шенноном коефіцієнтів неперервного вейвлет-перетворення конкретного рівня масштабування. Прихований шар мережі складається із 10 нейронів, вихідний шар – із 4 нейронів.

Експерименти з подачею на вхід багатошарового перцептрона результатів перетворення Фур'є, кепстрального перетворення та вейвлет-перетворення показали, що нейромережа спроможна з достатньою точністю може визначати технічний стан об'єкта в даній задачі за класифікацією по чотирьох класах.

Висновки. Результати дослідження показали, що багатошарові перцептронні мережі добре підходять для розпізнавання технічного стану об'єктів на основі опрацювання вібраційних сигналів. Поєднання обчислення статистичних характеристик сигналів та вирішення задачі класифікації за значеннями цих статистичних характеристик на основі нейронних мереж дає можливість ідентифікувати передаварійні та аварійні стани об'єктів на основі попереднього навчання. Перспективним напрямком є також пріоритезація обчислень статистичних характеристик на основі конкретних методів для мінімізації часу виконання обчислень.

Подяка. *Автори хочуть подякувати Збройним Силам України та всім захисникам України за можливість продовжувати наукову та технічну діяльність під час війни.*

Таблиця 1

Перелік експериментів із застосування багатошарової перцептронної мережі для вирішення задачі розпізнавання технічного стану на основі статистичних характеристик сигналів

Вхідні величини	Похибка навчання	Похибка тестування	Похибка валідації
Середнє значення, середньоквадратичне відхилення, ентропія Шеннона (розпізнавання за 4 класами)	33.1%	33.1%	30.4%
Середнє значення, середньоквадратичне відхилення, ентропія Шеннона (розпізнавання за 2 класами)	6%	2.7%	2.8%
Коефіцієнти швидкого перетворення Фур'є (розклад на 1024 компоненти)	5.6%	6.75%	5.4%
Кепстральний аналіз (розклад на 1024 компоненти)	0%	2%	1.3%
Вейвлет перетворення (розклад на 100 компонент за вейвлетом Добеши першого порядку)	0.7%	1.3%	0.6%

Список літератури:

1. Zhu K., Song X., Xue D. A roller bearing fault diagnosis method based on hierarchical entropy and support vector machine with particle swarm optimization algorithm. *Measurement*. 2014. № 47. С. 669–675.
2. Eristi H. Fault diagnosis system for series compensated transmission line based on wavelet transform and adaptive neuro-fuzzy inference system. *Measurement*. 2013. № 46. С. 393–401.
3. Georgoulas G., Mustafa M.O., Tsoumas I.P., Antonino-Daviu J.A., Climente-Alarcon V., Stylios C.D., Nikolakopoulos G. Principal Component Analysis of the start-up transient and Hidden Markov Modeling for broken rotor bar fault diagnosis in asynchronous machines. *Expert Systems with Applications*. 2013. № 40. С. 7024–7033.
4. Unal M., Onat M., Demetgul M., Kucuk H. Fault diagnosis of rolling bearings using a genetic algorithm optimized neural network. *Measurement*. 2014. № 58. С. 187–196.
5. Muralidharan V., Sugumaran V. Feature extraction using wavelets and classification through decision tree algorithm for fault diagnosis of mono-block centrifugal pump. *Measurement*. 2013. № 46. С. 353–359.
6. Lin H.-C., Su C.-T. A selective Bayes classifier with meta-heuristics for incomplete data. *Neurocomputing*. 2013. № 106. С. 95–102.
7. Wu J.-D., Liu C.-H. An expert system for fault diagnosis in internal combustion engines using wavelet packet transform and neural network. *Expert Systems with Applications*. 2009. № 36. С. 4278–4286.
8. Liu X., Bo L., Luo H. Bearing faults diagnostics based on hybrid LS-SVM and EMD method. *Measurement*. 2015. № 59. С. 145–166.

Slabinoha M.O., Pikh V.Ya., Striletskyi Yu.Y., Sheketa V.I. TECHNICAL CONDITION RECOGNITION BASED ON VIBRATION SIGNAL PROCESSING USING MULTI-LAYER PERCEPTRON

The article explores the application of artificial intelligence methods, particularly multilayer perceptron networks, for the classification of an object's technical condition based on vibration signals, which is highly relevant in the context of the digital transformation of industrial processes. It is noted that the preliminary processing of time series data plays a crucial role in improving the accuracy of technical state identification. The paper analyzes modern approaches, including the use of support vector machines, neural networks optimized with genetic algorithms, expert systems, Markov networks, and various signal preprocessing techniques such as wavelet transform, empirical mode decomposition, and cepstral analysis. The main focus is placed on a practical experiment using a dataset containing vibration signals from a test rotary setup. The data were clustered using the K-means method into four technical condition classes: «good», «satisfactory», «pre-failure», and «failure». Several multilayer perceptron models were constructed to solve the classification task using different signal processing methods: statistical features (mean, RMS, Shannon entropy), Fast Fourier Transform coefficients, cepstral analysis, and wavelet transform.

The results showed that basic statistical features provided moderate accuracy (with over 30% error for four-class classification), whereas the use of spectral and wavelet transforms significantly improved model performance (reducing error to 1–2%). In particular, the best results were achieved using cepstral analysis and Daubechies wavelet transform of the first order. It was demonstrated that even when reducing the number of classes (e.g., merging into "normal" and "failure" states), the classification accuracy increases significantly, offering potential for developing reliable early fault detection systems.

The conclusions highlight the effectiveness of combining statistical analysis methods with neural network approaches to improve the accuracy of technical condition diagnostics, as well as the promising direction of investigating various signal preprocessing methods to reduce classification error and optimize computational costs.

Key words: *perceptron, signal processing, probabilistic estimates, artificial neural networks, identification, technical condition.*

УДК 004.8:004.056.5:621.039.58
DOI https://doi.org/10.32782/2663-5941/2025.3.2/48

Соловій П.С.
Національний університет «Львівська політехніка»

Назаренко В.В.
Національний університет «Львівська політехніка»

ВПРОВАДЖЕННЯ ШІ У СПЕЦІАЛІЗОВАНІ КОМП'ЮТЕРНІ СИСТЕМИ ДЛЯ ПІДТРИМКИ КРИТИЧНОЇ ІНФРАСТРУКТУРИ

У статті розглянуто теоретичні та практичні аспекти впровадження штучного інтелекту у спеціалізовані комп'ютерні системи, орієнтовані на підтримку функціонування об'єктів критичної інфраструктури. Проаналізовано сучасні підходи до інтеграції інтелектуальних алгоритмів у системи моніторингу, прогнозування загроз та управління інформаційними потоками в умовах підвищених ризиків. Особлива увага приділяється використанню багаторівневих нейронних мереж і методів глибинного навчання, які дозволяють підвищити стійкість комп'ютерних систем до динамічних змін середовища та забезпечити високоточне виявлення потенційних технічних і кіберзагроз.

Висвітлено застосування алгоритмів матричної факторизації для зменшення розмірності даних без втрати критичної інформації, що оптимізує процес прийняття рішень у системах реального часу. Описано механізми роботи багатошарового персептрону із зворотним поширенням помилки для адаптивного навчання моделей прогнозування. Проаналізовано роль метаевристичних методів оптимізації у процесі налаштування комп'ютерних систем безпеки, зокрема в умовах неповноти або надлишковості даних.

Окремо розглянуто можливості використання методів оптимізації ансамблів моделей (МОА) для підвищення узагальнюючої здатності спеціалізованих комп'ютерних систем і забезпечення їх самокорекції на основі безперервного зворотного зв'язку. Акцентовано увагу на перспективах застосування інтелектуальних рекомендаційних систем як одного з компонентів підтримки критичної інфраструктури, зокрема для розподілу ресурсів, прогнозування відмов і формування стратегій управління ризиками.

У підсумку підкреслено, що впровадження штучного інтелекту у спеціалізовані комп'ютерні системи є ключовим чинником підвищення ефективності, надійності та стійкості об'єктів критичної інфраструктури. Запропоновані алгоритмічні рішення сприяють розвитку програмних комплексів нового покоління із вбудованими модулями інтелектуальної підтримки прийняття рішень, що забезпечує адаптивність систем у нестабільних умовах та відкриває нові перспективи їх удосконалення.

Ключові слова: *штучний інтелект, спеціалізовані комп'ютерні системи, критична інфраструктура, глибинне навчання, матрична факторизація, метаевристичні методи оптимізації, рекомендаційні системи.*

Постановка проблеми. Впровадження штучного інтелекту у спеціалізовані комп'ютерні системи для підтримки критичної інфраструктури супроводжується низкою комплексних викликів, пов'язаних із забезпеченням надійності, стійкості та адаптивності таких систем в умовах зростання кількості кіберзагроз та мінливості зовнішнього середовища. Незважаючи на активний розвиток технологій глибинного навчання та алгоритмів машинного аналізу даних, інтеграція цих рішень у реальні системи критичного призначення стикається із суттєвими обмеженнями ресурсів, вимогами до роботи в реальному часі, необхідністю оперативного відновлення функціональності після атак або збоїв.

Однією з ключових проблем є адаптація алгоритмів штучного інтелекту до умов експлуатації спеціалізованих комп'ютерних систем, що повинні забезпечувати безперервне функціонування об'єктів енергетики, транспорту, зв'язку та інших критичних галузей. Стандартні методи оптимізації нейронних мереж, такі як стохастичний градієнтний спуск, часто виявляються недостатньо стійкими до нестабільних вхідних даних та змінних режимів роботи системи, що створює загрозу втрати керованості процесами в умовах надзвичайних ситуацій.

Особливе значення має проблема забезпечення балансу між високою точністю прогнозування та обмеженою пропускною здатністю

комп'ютерних систем, які функціонують у реальному часі. У критичних інфраструктурах недопустима затримка у прийнятті рішень або надмірне навантаження на обчислювальні ресурси, що зумовлює необхідність оптимізації моделей за критеріями енергоефективності, швидкодії та стійкості.

Актуальним є також завдання створення самонавчальних та самовідновлюваних інтелектуальних систем, здатних безперервно адаптуватися до змін у середовищі експлуатації без потреби повного перенавчання. Це вимагає впровадження метаевристичних методів оптимізації, які дозволяють підвищити гнучкість налаштування моделей та забезпечити їхню стійкість у нестабільних і непередбачуваних умовах.

Таким чином, основна проблема полягає у розробці ефективних технологій впровадження штучного інтелекту в архітектуру спеціалізованих комп'ютерних систем для критичної інфраструктури, що забезпечують високу точність аналізу, стійкість до атак і технічних відмов, можливість роботи в реальному часі та адаптивність до змін середовища при мінімальних обчислювальних витратах.

Аналіз останніх досліджень і публікацій. Теоретичні аспекти впровадження штучного інтелекту в спеціалізовані комп'ютерні системи для підтримки критичної інфраструктури були досліджені різними вітчизняними науковцями, які зосереджували увагу на питаннях кібербезпеки, діджиталізації економіки та захисту інформаційного середовища.

Зокрема, М. Гуцалюк у своїй праці проаналізував стратегії протидії сучасним кіберзагрозам та забезпечення кіберстійкості об'єктів критичної інфраструктури України. Автор відзначив важливість побудови адаптивних інформаційних систем із використанням інструментів штучного інтелекту, які можуть оперативно реагувати на зміну загрозового середовища [1, с. 170]. У дослідженні Н. Ткачук розглянуто процес діджиталізації фінансово-економічної сфери в Україні, зокрема в аспекті впровадження інноваційних інформаційних технологій для підтримки безперервності фінансових процесів, що має безпосереднє відношення до критичної інфраструктури [2, с. 20]. Як зазначено в праці, застосування алгоритмів штучного інтелекту сприяє підвищенню стійкості та гнучкості систем управління у кризових умовах.

Детальне висвітлення використання штучного інтелекту для забезпечення кібербезпеки об'єктів критичної інфраструктури в умовах воєнного стану наведено у статті С. Казьмірука та ін. Дослідники акцентували увагу на необхідності використання моделей машинного навчання для раннього виявлення загроз та оптимізації процесів захисту [3, с. 2]. С. Цяпа проаналізував правові та організаційні механізми захисту критичної інформаційної інфраструктури від кібератак, звертаючи увагу на необхідність інтеграції інтелектуальних систем в існуючі платформи захисту, що забезпечить своєчасне виявлення і нейтралізацію загроз [4, с. 125].

Окрему увагу слід приділити роботі С. Гнатюка та ін., які запропонували концепцію системи корелювання подій та управління інцидентами кібербезпеки на об'єктах критичної інфраструктури. Запропоновані підходи базуються на використанні технологій штучного інтелекту для обробки великого обсягу інформації у реальному часі [5, с. 180–182]. Питання захисту економічних систем від кіберзагроз у контексті цифрової трансформації економіки було досліджено Г. Михальченком та ін., які акцентували увагу на потребі створення програмних комплексів на основі нейронних мереж для забезпечення кібербезпеки фінансових установ [6, с. 380].

На міжнародному рівні актуальні проблеми застосування штучного інтелекту для захисту критичної інфраструктури висвітлено у праці G. S. Jadoun та співавторів, де проаналізовано загрози, пов'язані із використанням ШІ у кібербезпеці, і запропоновано контрзаходи, що базуються на автоматизованих системах виявлення атак [7, с. 1–13]. O. Okusi зосередив увагу на можливостях застосування штучного інтелекту та машинного навчання для забезпечення захисту національної критичної інфраструктури, підкресливши важливість розвитку адаптивних моделей реагування [8, с. 1–11]. У роботі A. Chehri, I. Fofana та X. Yang розглянуто підходи до моделювання ризиків для розумних електричних мереж у добу великих даних і штучного інтелекту, що є перспективним напрямом для розвитку енергетичної критичної інфраструктури [9, с. 3196]. Крім того, I. R. Kusumasari та співавтори дослідили проблеми й можливості використання штучного інтелекту як інструменту підтримки прийняття бізнес-рішень у цифрову епоху, що є надзвичайно важливим для стабільності функціонування економічно важливих секторів [10, с. 17].

Таким чином, проведений аналіз свідчить про високий рівень наукового інтересу до теми

інтеграції інтелектуальних систем у структури критичної інфраструктури. Проте, незважаючи на наявність численних досліджень, потребує подальшого розвитку питання комплексного використання метаевристичних методів оптимізації в нейронних мережах для підвищення ефективності програмних рішень у сфері критичної інфраструктури.

Постановка завдання. Метою статті є дослідження теоретичних і практичних аспектів впровадження штучного інтелекту в спеціалізовані комп'ютерні системи для підтримки критичної інфраструктури, з акцентом на застосуванні метаевристичних методів оптимізації в архітектурах нейронних мереж з метою підвищення їхньої ефективності, стійкості та адаптивності в умовах реального часу.

Виклад основного матеріалу. Впровадження штучного інтелекту у спеціалізовані комп'ютерні системи підтримки критичної інфраструктури є одним із ключових напрямів розвитку сучасних інформаційних технологій. Особливість даного процесу полягає у необхідності забезпечення безперервної, надійної та адаптивної роботи систем в умовах змінного загрозового середовища. Для досягнення цього рівня функціонування застосовуються алгоритми машинного навчання, глибинного навчання, метаевристичні методи оптимізації та моделі прогнозування [1, с. 170].

Одним із базових підходів до реалізації інтелектуальної обробки даних у спеціалізованих комп'ютерних системах є використання багатошарових нейронних мереж, що дозволяють формувати складні залежності між вхідними параметрами системи. Багатошарові персептрони (Multilayer Perceptrons, MLP) забезпечують обробку великої кількості змінних за рахунок наявності кількох прихованих шарів, де кожен шар виконує роль нелінійного перетворення простору ознак. Адаптивне навчання таких мереж за допомогою алгоритму зворотного поширення помилки (Backpropagation) забезпечує можливість оперативного оновлення вагових коефіцієнтів моделі на основі надходження нових даних без необхідності повного перенавчання [5, с. 180].

Для систем, що працюють в умовах критичної інфраструктури, особливе значення має використання архітектур, які забезпечують швидку конвергенцію і стійкість до змін вхідних потоків інформації. Серед таких архітектур можна виокремити згорткові нейронні мережі (Convolutional Neural Networks, CNN), що застосовуються для обробки структурованих і частково структурованих даних, а також рекурентні нейронні мережі (Recurrent Neural Networks, RNN) та їхні модифікації, такі як довготривала короткочасна пам'ять (Long Short-Term Memory, LSTM), які призначені для роботи з послідовностями подій у часі [8, с. 5].

У випадках, коли необхідно одночасно враховувати як просторові, так і часові залежності, ефективними виявляються гібридні моделі, що поєднують можливості CNN та LSTM. Наприклад, системи раннього виявлення загроз у мережах критичної інфраструктури застосовують комбіновані архітектури CNN-LSTM для швидкої фільтрації вхідних потоків даних і прогнозування аномалій на основі історичних патернів поведінки мережевих пристроїв [9, с. 3196].

Наявність можливості безперервного оновлення моделей без повного перенавчання має вирішальне значення для забезпечення оперативного реагування на нові типи загроз і змін у структурі даних. Це особливо важливо для комп'ютерних систем, що функціонують у критичних умовах, де швидкість і точність прийняття рішень безпосередньо впливають на надійність і безпеку об'єктів інфраструктури. Підвищення гнучкості моделей за рахунок впровадження сучасних архітектур і адаптивних методів навчання є одним із основних шляхів підвищення ефективності спеціалізованих комп'ютерних систем для підтримки критичних процесів.

Для оптимізації навчання та підвищення стійкості моделей у спеціалізованих комп'ютерних системах підтримки критичної інфраструктури широко використовуються метаевристичні методи оптимізації, що демонструють високу ефективність в умовах складних і багатовимірних просторів рішень. Серед найбільш розповсюджених варто відзначити генетичні алгоритми (Genetic Algorithms, GA), алгоритми рою частинок (Particle Swarm Optimization, PSO) та мурашині алгоритми (Ant Colony Optimization, ACO) (Рис. 1).

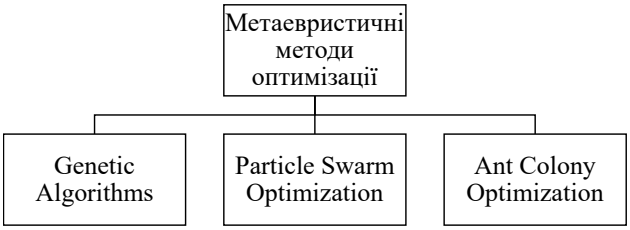


Рис. 1. Метаевристичні методи оптимізації у спеціалізованих комп'ютерних системах підтримки критичної інфраструктури

Генетичні алгоритми (GA) базуються на принципах природного добору та еволюції. Вони імітують біологічні процеси спадковості, мутацій і відбору найкращих рішень для поступового поліпшення характеристик моделей. Застосування GA у спеціалізованих комп'ютерних системах дозволяє ефективно налаштовувати параметри нейронних мереж, забезпечуючи високу адаптивність моделей до нових типів даних без необхідності ручної оптимізації гіперпараметрів.

Алгоритм рою частинок (PSO) моделює колективну поведінку живих істот, наприклад, зграї птахів або косяків риб. У контексті навчання моделей PSO дозволяє координовано досліджувати простір параметрів за рахунок обміну інформацією між «частинками» – можливими рішеннями. Особливістю PSO є висока швидкість збіжності до оптимального рішення, що є критичним фактором у реальному часі для систем підтримки критичної інфраструктури [6, с. 380].

Мурашині алгоритми (ACO) імітують механізми колективної поведінки мурах, які прокладають найкоротші маршрути між джерелом їжі та мурашником, використовуючи феромонні мітки. У комп'ютерних системах ACO дозволяють ефективно вирішувати задачі маршрутизації даних, розподілу навантаження та побудови стійких комунікаційних маршрутів в умовах високої динаміки середовища [10, с. 17].

Застосування зазначених метаевристичних підходів забезпечує подолання обмежень класичних методів оптимізації, таких як стохастичний градієнтний спуск (Stochastic Gradient Descent, SGD), які схильні застрягати у локальних мінімумах, особливо в умовах складних ландшафтів функцій втрат. Метаевристичні алгоритми, завдяки своїй стохастичній природі та можливості глобального пошуку, дозволяють знаходити більш якісні рішення навіть при наявності нерівномірно розподілених або частково зашумлених даних, що є типовим для інформаційних потоків у системах критичної інфраструктури [9, с. 3196].

У процесі впровадження штучного інтелекту у спеціалізовані комп'ютерні системи важливу роль відіграє використання методів матричної факторизації, які дозволяють ефективно зменшити розмірність вхідних даних без втрати критичної інформації. Матрична факторизація (Matrix Factorization) забезпечує поділ великих матриць на декілька менш розмірних матриць, що містять приховані (латентні) особливості вихідних даних. У системах моніторингу критичної інфраструктури це дає змогу підвищити швидкість обробки даних, зменшити обчислювальні навантаження і водночас зберегти релевантність та точність формування рішень [5, с. 179].

Одним із найбільш застосовуваних методів є сингулярне розкладення матриці (Singular Value Decomposition, SVD), яке розкладає матрицю спостережень на добуток трьох менших матриць, що відображають основні напрямки варіацій у даних. Використання SVD дозволяє зосередити аналіз лише на тих компонентах, які несуть максимальну кількість інформації, ігноруючи другорядні шуми або випадкові збурення.

Іншим підходом є метод латентних чинників (Latent Factor Models), який широко використовується у рекомендаційних системах для виявлення прихованих залежностей між об'єктами та ознаками. У контексті підтримки критичної інфраструктури це дозволяє, наприклад, ефективно прогнозувати технічні відмови обладнання на основі неповних або фрагментарних даних, що надходять із сенсорних мереж.

Особливу увагу привертає також метод стохастичної матричної факторизації (Stochastic Matrix Factorization, SMF), що оптимізує процес розкладання для великих потоків даних у режимі реального часу. Використання SMF дозволяє спеціалізованим комп'ютерним системам обробляти великі масиви телеметричної інформації без істотного падіння продуктивності навіть при постійному надходженні нових даних.

Зменшення обсягу оброблюваної інформації без втрати ключових ознак істотно підвищує продуктивність комп'ютерних систем у режимі реального часу, що критично важливо для об'єктів енергетичної, транспортної, фінансової та комунікаційної інфраструктури. В умовах постійних змін і динаміки загроз, оптимізація обчислювальних процесів за допомогою методів матричної факторизації дозволяє не тільки забезпечити швидке прийняття рішень, а й зберегти стабільність та безперервність функціонування критичних об'єктів навіть під час кризових ситуацій.

Прикладом ефективної реалізації такого підходу є системи прогнозування технічних відмов на енергетичних об'єктах, де алгоритми матричної факторизації використовуються для виявлення латентних закономірностей між різними параметрами роботи обладнання [2, с. 22]. За рахунок цього можна своєчасно виявляти аномалії у роботі критичних вузлів і запобігати виникненню серйозних аварій.

Іншою важливою складовою інтеграції штучного інтелекту у спеціалізовані комп'ютерні системи є застосування методів оптимізації ансамблів моделей (MOA). Ансамблювання дозволяє поєднати сильні сторони різних алгоритмів, знижуючи ризик помилкових рішень та підвищуючи стійкість системи до випадкових збурень у даних [4, с. 126]. Застосування MOA особливо ефективне у середовищах із високою динамікою загроз, де окремі моделі можуть не справлятися з новими сценаріями атак.

Практичні приклади впровадження MOA включають системи управління інцидентами в транспортній інфраструктурі, де використання ансамблів моделей дозволяє значно скоротити час виявлення порушень без зниження точності оцінки загроз [5, с. 182]. У результаті досягається підвищення загальної надійності комп'ютерних систем, що обслуговують критичні об'єкти.

З огляду на вимоги роботи у реальному часі, інтеграція глибинного навчання в спеціалізовані комп'ютерні системи передбачає оптимізацію архітектури нейронних мереж. Застосовуються методи редукції глибини мереж без втрати їхньої прогнозної здатності або використання мобільних архітектур, таких як MobileNet, оптимізованих для обмежених обчислювальних ресурсів [6, с. 379]. Це забезпечує можливість використання інтелектуальних функцій навіть на пристроях із середнім рівнем обчислювальної потужності, що характерно для багатьох систем об'єктів критичної інфраструктури.

Окрему увагу приділяють питанням побудови самонавчальних систем, які здатні безперервно адаптувати свої моделі відповідно до змін умов експлуатації. У системах, що підтримують об'єкти критичної інфраструктури, така здатність є ключовою, оскільки дозволяє реагувати на зміну технічних параметрів, поведінкових патернів користувачів чи нові типи загроз без потреби зупиняти роботу або виконувати повне перенавчання моделі. Створення таких систем базується на використанні алгоритмів активного навчання (Active Learning), що дозволяють вибирати найбільш інформативні зразки для донавчання моделей, а також механізмів безперервного контролю за ефективністю моделей у процесі їх функціонування (Continuous Monitoring and Evaluation) [1, с. 172].

Одним із найбільш ефективних підходів є реалізація архітектур з онлайн-навчанням (Online Learning), де модель оновлює свої параметри одразу після надходження кожного нового пакету даних. Такі алгоритми, як Stochastic Gradient Descent (SGD) у режимі онлайн або Adaptive Moment Estimation (Adam) у варіації для потокових даних, дозволяють інтегрувати нову інформацію з мінімальними обчислювальними витратами і без ризику втрати накопиченого досвіду.

Значну роль у побудові самонавчальних систем відіграють алгоритми псевдонагляду (Semi-Supervised Learning) та самоорганізовані карти Кохонена (Self-Organizing Maps, SOM), які дозволяють формувати нові структури класифікації даних на основі непозначених або частково анотованих наборів даних. Це дає можливість системам підтримки критичної інфраструктури своєчасно виявляти нові загрози або аномальні патерни без потреби в постійно актуалізованих вручну еталонних вибірках.

Створення самонавчальних систем також базується на принципах мультиагентних технологій, де окремі підсистеми працюють незалежно і здатні самостійно приймати рішення про необхідність адаптації моделей у відповідь на зміну умов експлуатації. У критичних системах це дозволяє зменшити час реагування на нові типи загроз і забезпечити локальну самовідновлюваність окремих функціональних блоків без втрати загальної працездатності системи. Впровадження таких технологій дозволяє не тільки підвищити точність прогнозування загроз, своєчасно виявляючи нові аномалії у поведінці систем або мереж, але й формує основу для створення адаптивних захисних бар'єрів, які здатні еволюціонувати разом із зміною характеру загроз. Це має особливе значення в умовах функціонування критичних об'єктів, де стабільність, безперервність та оперативність реагування на потенційні ризики є визначальними чинниками загальної безпеки.

Інтеграція штучного інтелекту також суттєво змінює підходи до забезпечення безпеки інформаційних потоків у спеціалізованих комп'ютерних системах, орієнтованих на підтримку критичної інфраструктури. Одним із перспективних напрямів є застосування адаптивних методів шифрування, побудованих на базі нейронних мереж, які використовують механізми глибинного навчання для автоматичної генерації криптографічних ключів та алгоритмів шифрування. Такі підходи дозволяють досягти високого рівня динамічності у захисті даних, забезпечуючи адаптацію криптографічних механізмів до зміни середовища загроз у режимі реального часу [4, с. 127].

Одним із прикладів є використання рекурентних нейронних мереж (Recurrent Neural

Networks, RNN) або їхніх модифікацій, зокрема довготривалої короткочасної пам'яті (Long Short-Term Memory, LSTM), для створення складних шифрувальних схем, що мінімізують ймовірність злому навіть у разі застосування сучасних методів криптоаналізу. Крім того, досліджуються підходи до використання генеративно-змагальних мереж (Generative Adversarial Networks, GAN) для побудови систем шифрування та дешифрування, які унеможливлюють підробку або несанкціоновану модифікацію інформаційних потоків.

Застосування таких технологій забезпечує підвищений рівень захисту даних без надмірного збільшення навантаження на системні ресурси. Завдяки можливості оптимізувати процеси шифрування через машинне навчання, спеціалізовані комп'ютерні системи зберігають високу швидкість обробки інформації, що є критично важливим в умовах обмеженої пропускної здатності або при необхідності роботи у режимі реального часу.

Це особливо важливо для об'єктів критичної інфраструктури, таких як енергетичні системи, транспортні мережі або центри обробки даних, де інформаційні потоки мають високу чутливість, стратегічне значення і потребують максимального рівня захисту як від зовнішніх кібератак, так і від внутрішніх загроз. Використання інтелектуальних механізмів шифрування дозволяє суттєво підвищити рівень кіберстійкості критичних систем, забезпечуючи надійність і безпеку обміну даними навіть у складних умовах інформаційної протидії.

Для систематизації основних напрямків інтеграції штучного інтелекту у спеціалізовані комп'ютерні системи підтримки критичної інфраструктури у таблиці 1 подано узагальнення застосовуваних методів, їх призначення та практичних сфер використання.

Таким чином, впровадження штучного інтелекту в спеціалізовані комп'ютерні системи для підтримки критичної інфраструктури базується на комплексному використанні глибинного навчання, метаевристичних методів оптимізації, матричної факторизації та оптимізації ансамблів моделей. Ефективна інтеграція цих технологій забезпечує підвищення стійкості, надійності та адаптивності об'єктів критичної інфраструктури в умовах зростаючих загроз і нестабільного середовища.

Висновки. У даному дослідженні розглянуто теоретичні та практичні аспекти впровадження штучного інтелекту у спеціалізовані комп'ютерні системи, призначені для підтримки функціонування об'єктів критичної інфраструктури. Проаналізовано можливості застосування глибинного навчання, алгоритмів матричної факторизації, метаевристичних методів оптимізації та технологій ансамблювання моделей для забезпечення стійкості, надійності та адаптивності інформаційних систем у середовищах підвищеного ризику. Доведено, що використання багатошарових нейронних мереж із механізмом зворотного поширення помилки сприяє підвищенню точності прогнозування загроз та оптимізації процесів прийняття рішень у реальному часі. Застосування метаевристичних підходів, таких як генетичні алгоритми та алгоритми рою частинок, забезпечує ефективний пошук оптимальних параметрів моделей у складних та нестабільних середовищах експлуатації. Результати дослідження свідчать про доцільність впровадження матричної факторизації для зменшення розмір-

Таблиця 1

Методи впровадження штучного інтелекту у спеціалізовані комп'ютерні системи критичної інфраструктури

Метод або технологія	Призначення	Приклад застосування
Глибинне навчання	Аналіз великих обсягів даних, виявлення складних залежностей	Системи прогнозування технічних відмов обладнання
Матрична факторизація	Зменшення розмірності даних, оптимізація обчислювань	Оптимізація рішень у системах моніторингу мережевої активності
Метаевристичні методи оптимізації	Пошук глобальних оптимумів у складних просторах рішень	Налаштування параметрів адаптивних нейронних мереж
Оптимізація ансамблів моделей (MOA)	Підвищення точності та стійкості прогнозування	Інтегровані системи виявлення кіберзагроз
Адаптивні методи шифрування на базі ШІ	Захист інформаційних потоків в умовах високої динаміки	Шифрування даних у системах диспетчеризації енергомереж
Самонавчальні системи	Безперервне оновлення моделей без повного перенавчання	Реагування на нові типи атак у критичній інформаційній інфраструктурі

ності вхідних даних без втрати суттєвої інформації, що дозволяє підвищити швидкість роботи спеціалізованих комп'ютерних систем без зниження рівня інформативності результатів. Використання оптимізації ансамблів моделей (МОА) дозволяє суттєво збільшити стійкість систем до випадкових змін у середовищі функціонування та підвищити точність комплексної оцінки загроз. Підкреслено, що створення самонавчальних і самокоригувальних систем на базі ШІ є одним із перспективних напрямів розвитку спеціалізованих комп'ютерних комплексів для критичної інфраструктури. Такі системи здатні безперервно оновлювати свої внутрішні моделі без необхідності повного перенавчання, що є критичним у нестабільних умовах та при обмежених ресурсах. У підсумку впровадження штучного інтелекту у спеціалізовані комп'ютерні системи відкриває нові можливості для забезпечення надійності, безперервності та безпеки функціонування об'єктів критичної інфраструктури. Запропоновані підходи дозволяють підвищити ефективність управління ризиками, оптимізувати використання ресурсів та забезпечити адаптивність систем до постійно змінних умов загрозового середовища.

Список літератури:

1. Гуцалюк М. В. Стратегії протидії сучасним кіберзагрозам та забезпечення кіберстійкості критичної інфраструктури України. *Інформація і право.* 2024. № 2 (49). С. 164–177. DOI: https://doi.org/10.37750/2616-6798.2024.2(49).306199.

2. Ткачук Н. Діджиталізація фінансово-економічної сфери в Україні: стан та перспективи розвитку. *Економічний журнал Волинського національного університету імені Лесі Українки.* 2022. № 3 (31). С. 18–28. DOI: https://doi.org/10.29038/2786-4618-2022-03-18-28.

3. Казьмірук С. Д., Леонов Б. Д., Омельян О. С. Забезпечення кібербезпеки об'єктів критичної інфраструктури на основі використання штучного інтелекту в умовах воєнного стану. *Юридичний науковий електронний журнал.* 2024. № 6. С. 1–5. DOI: https://doi.org/10.32782/2524-0374/2024-6/49.

4. Цяпа С. М. Правове та організаційне забезпечення захисту об'єктів критичної інформаційної інфраструктури від кібератак. *Інформація і право.* 2021. № 4 (39). С. 121–128. DOI: https://doi.org/10.37750/2616-6798.2021.4(39).248832.

5. Гнатюк С., та ін. Система корелювання подій та управління інцидентами кібербезпеки на об'єктах критичної інфраструктури. *Кібербезпека: освіта, наука, техніка.* 2023. № 3 (19). С. 176–196. DOI: https://doi.org/10.28925/2663-4023.2023.19.176196.

6. Михальченко Г. Г., Снітко Ю. М., Іваненко В. О. Кібербезпека в економіці: захист від кіберзагроз у диджиталізованому світі. *Наукові записки Львівського університету бізнесу та права.* 2023. № 38. С. 377–384. DOI: http://dx.doi.org/10.5281/zenodo.10012434.

7. Jadoun G. S., Bhatt D. P., Mathur V., Kaur A. The threat of artificial intelligence in cyber security: risk and countermeasures. *AIP Conference Proceedings.* 2025. Vol. 3191, No. 1. P. 1–13. DOI: https://doi.org/10.1063/5.0248313.

8. Okusi O. Leveraging AI and machine learning for the protection of critical national infrastructure. *Asian Journal of Research in Computer Science.* 2024. Vol. 17, No. 10. P. 1–11. DOI: https://doi.org/10.9734/ajrcos/2024/v17i10505.

9. Chehri A., Fofana I., Yang X. Security risk modeling in smart grid critical infrastructures in the era of big data and artificial intelligence. *Sustainability.* 2021. Vol. 13, No. 6. P. 3196. DOI: https://doi.org/10.3390/su13063196.

10. Kusumasari I. R., Putri P. A., Murdiana N., Puspita D. R. Challenges and opportunities for using artificial intelligence as a supporting tool in business decision making in the digital era. *Jurnal Bisnis dan Komunikasi Digital.* 2025. Vol. 2, No. 2. P. 17–17. DOI: https://doi.org/10.47134/jbkd.v2i2.3469.

Solovii P.S., Nazarenko V.V. IMPLEMENTATION OF ARTIFICIAL INTELLIGENCE INTO SPECIALIZED COMPUTER SYSTEMS FOR CRITICAL INFRASTRUCTURE SUPPORT

The article examines the theoretical and practical aspects of implementing artificial intelligence into specialized computer systems aimed at supporting the functioning of critical infrastructure facilities. Modern approaches to integrating intelligent algorithms into monitoring systems, threat forecasting, and information flow management under conditions of elevated risks are analyzed. Special attention is given to the use of multilayer neural networks and deep learning methods, which enhance the resilience of computer systems to dynamic environmental changes and ensure high-precision detection of potential technical and cyber threats.

The application of matrix factorization algorithms for dimensionality reduction without the loss of critical information is highlighted, optimizing decision-making processes in real-time systems. The mechanisms of

multilayer perceptron operations with backpropagation for adaptive model training are described. The role of metaheuristic optimization methods in configuring security-focused computer systems, particularly in conditions of data incompleteness or redundancy, is analyzed.

The article separately addresses the potential of ensemble model optimization methods (EMO) to enhance the generalization capabilities of specialized computer systems and ensure their self-correction based on continuous feedback. Particular emphasis is placed on the prospects of using intelligent recommendation systems as one of the components supporting critical infrastructure, especially for resource allocation, failure prediction, and risk management strategy development.

In conclusion, it is emphasized that the implementation of artificial intelligence in specialized computer systems is a key factor in improving the efficiency, reliability, and resilience of critical infrastructure facilities. The proposed algorithmic solutions contribute to the development of next-generation software systems with integrated decision support modules, providing system adaptability under unstable conditions and opening new perspectives for their further advancement.

Key words: *artificial intelligence, specialized computer systems, critical infrastructure, deep learning, matrix factorization, metaheuristic optimization methods, recommendation systems.*

УДК 004.415.3:004.738.5
DOI https://doi.org/10.32782/2663-5941/2025.3.2/49

Степанов Д.С.
Національний університет «Львівська політехніка»

Сенів М.М.
Національний університет «Львівська політехніка»

ЗАСТОСУВАННЯ ТЕХНОЛОГІЙ НЕЗМІННОЇ ІНФРАСТРУКТУРИ ДЛЯ ЗАБЕЗПЕЧЕННЯ ПРОЗОРОГО МАСШТАБУВАННЯ БЕЗ ПРОСТОЇВ У ВЕБПОРТАЛАХ

Актуальність дослідження зумовлена необхідністю стабільного функціонування вебпорталів в умовах динамічного зростання навантаження та необхідності безперервного оновлення програмного забезпечення без втрати даних і простоїв. Застосування традиційних архітектурних підходів часто не забезпечує належного рівня надійності, оскільки масштабування призводить до виникнення конфліктів між версіями застосунків, несумісності компонентів та порушення цілісності даних під час оновлень. Відсутність уніфікованих рішень для оперативного відновлення системи у разі збою ускладнює підтримку високих показників доступності вебпорталів, особливо в контексті багатокомпонентних систем із розподіленою архітектурою.

Мета дослідження полягає в розробці методу забезпечення прозорого масштабування вебпорталів на основі інтеграції технологій контейнеризації та незмінної інфраструктури з акцентом на підвищення доступності та відновлюваності в межах концепції надійності програмного забезпечення.

Методологія дослідження охоплює порівняльний аналіз наявних підходів до контейнеризації, моделювання архітектурних схем для інтеграції незмінної інфраструктури у вебпортали, а також структурно-функціональний аналіз систем моніторингу та аварійного відновлення. Виявлено вплив контейнеризації на стабільність роботи компонентів вебпорталів у разі динамічного масштабування.

Наукова новизна дослідження полягає в розробці інтеграційної моделі, яка забезпечує автоматизоване оновлення компонентів без втрати доступності системи шляхом застосування стратегій «blue-green deployment» та «canary deployment».

Встановлено, що впровадження контейнеризації в поєднанні з незмінною інфраструктурою підвищує загальну стійкість системи до збоїв, мінімізуючи вплив людського фактора та забезпечуючи узгодженість версій компонентів. Доведено, що використання централізованих репозиторіїв значно скорочує час відновлення після аварії, а застосування контейнерних сесій із моніторингом забезпечує своєчасне реагування на потенційні загрози.

Перспективи подальших досліджень передбачають адаптацію розробленої моделі до архітектур із розподіленими сервісами та інтеграцію методів прогнозування пікових навантажень із застосуванням інструментів машинного навчання для своєчасного масштабування без зупинки основних сервісів.

Ключові слова: *незмінна інфраструктура, контейнеризація, масштабування без простоїв, надійність програмного забезпечення, відновлюваність, автоматизація оновлень.*

Постановка проблеми. Для сучасних вебпорталів критично важливо забезпечити стабільне функціонування в умовах зростання навантаження та динамічної користувацької активності, що зумовлює необхідність упровадження ефективних технологічних рішень для масштабування без простоїв. Традиційні методи масштабування часто виявляються недостатньо ефективними, оскільки вони можуть призводити до виникнення простоїв під час оновлення інфраструктури, а також до ускладнень у підтримці узгодженості даних між серверами. У відповідь на ці виклики все більшої популярності набувають технології незмінної інфраструктури, які дають можливість створювати повністю оновлені версії програмного забезпечення без втручання у поточні процеси та ризику для стабільності системи. Застосування таких технологій допомагає уникнути ризиків, пов'язаних із людським фактором та зменшити ймовірність виникнення помилок під час розгортання нових версій. Це забезпечує доступність до вебпорталів навіть у періоди активних оновлень та зростання навантаження. Водночас упровадження незмін-

ної інфраструктури потребує врахування особливостей архітектури вебпорталів, оптимізації процесів автоматичного розгортання та інтеграції засобів моніторингу для своєчасного виявлення потенційних загроз стабільності системи. Таким чином, актуальність дослідження полягає у визначенні ефективних моделей застосування технологій незмінної інфраструктури для прозорого масштабування вебпорталів, мінімізації часу простоїв та підвищення загальної стійкості системи до збоїв. З огляду на важливість завдань, що стосуються забезпечення високої доступності вебпорталів у сучасних умовах, доцільно зосередити увагу на розробці інноваційних підходів до реалізації масштабованих архітектур з акцентом на автоматизацію процесів розгортання та моніторингу.

Аналіз останніх досліджень і публікацій. Аналіз сучасних досліджень щодо застосування технологій незмінної інфраструктури для прозорого масштабування без простоїв у вебпорталах свідчить про наявність чотирьох основних змістових напрямів.

Перший напрям присвячено методологічним аспектам побудови незмінної інфраструктури та її впливу на показники надійності програмного забезпечення. Д. Степанов (D. Stepanov) [1] досліджує розвиток методології незмінної інфраструктури та аналізує її вплив на надійність роботи програмних систем, акцентуючи на доступності та швидкому відновленні після збоїв. У подальшій роботі Д. Степанов та М. Сенів [2] розширюють це питання, поєднуючи концепції захищеної інфраструктури, контейнеризації та DevSecOps, що комплексно підвищують стабільність і безпеку вебпорталів. М. Бафана (M. Bafana) та А. Абдулазіз (A. Abdulaziz) [3] розглядають практичні аспекти розгортання незмінної інфраструктури на платформі AWS, надаючи конкретні методологічні рекомендації для забезпечення безперебійної роботи. Н. Давидович (N. Davidovic), Н. Велькович (N. Veljkovic) та Л. Стоіменов (L. Stoimenov) [4] досліджують можливості блокчейн-технологій для гарантування незмінності державних електронних документів, підкреслюючи переваги таких рішень у публічних системах. С. Девіс (C. Davis) [5] описує загальні шаблони проєктування для побудови хмарних рішень із використанням незмінних архітектур, що забезпечують високу стійкість до змін та непередбачуваних навантажень. Перспективними напрямами досліджень є вивчення впливу обраних архітектурних рішень на швидкість масштабування та оцінка їхнього впливу на зниження витрат на обслуговування.

Другий напрям досліджень зосереджено на інтеграції технологій незмінної інфраструктури з блокчейн-рішеннями та розподіленими реєстрами для підвищення прозорості та безпеки масштабування. М. Шайк (M. Shaik) та Г. Бойя (G. Bojja) [6] вивчають інтеграцію блокчейну та сучасних протоколів управління доступом для посилення захисту цифрової інфраструктури. І. Тарханов (I. Tarkhanov), Д. Фомін-Нілов (D. Fomin-Nilov) та М. Фомін (M. Fomin) [7] підтверджують ефективність застосування блокчейну в наукових онлайн-виданнях для збереження незмінності цифрових даних. К. Кнутсен (K. Knutsen) та інші [8] демонструють застосування контейнеризованих систем із використанням блокчейн-технологій для безпечного обміну морськими даними. С. Бхуджел (S. Bhujel) та Й. Рахуламатхаван (Y. Rahulamathavan) [9] вивчають виклики масштабованості, прозорості та безпеки на прикладі NFT-маркетплейсів, наголошуючи на необхідності архітектурних удосконалень. Подальші дослідження доцільно спрямувати на аналіз технічних аспектів реалізації розподілених реєстрів у децентралізованих системах вебпорталів.

Третій напрям охоплює безпекові аспекти впровадження незмінної інфраструктури у вебпортали. Т. Аріф (T. Arif), Б. Джо (B. Jo) та Дж. Пак (J. Park) [10] аналізують комплексний підхід до збереження конфіденційності та безпеки даних у cloud-native середовищах, акцентуючи на необхідності інтеграції систем контролю доступу та верифікації даних. П. Чаухан (P. Chauhan) та П. Бансал (P. Bansal) [11] досліджують можливості незмінної архітектури для збереження цифрових доказів у криміналістичних системах, підкреслюючи її переваги для гарантування незмінності цифрових даних. Р. Нігам (R. Nigam), П. Азеведо де Аморім (P. Azevedo de Amorim) та А. Симпсон (A. Sampson) [12] пропонують модульні архітектурні підходи для апаратного забезпечення з використанням незмінних принципів проєктування. К. Індрасірі (K. Indrasiri) та С. Сухотаян (S. Suhothayan) [13] узагальнюють типові шаблони проєктування cloud-native додатків з незмінною архітектурою, акцентуючи на їхній масштабованості та стійкості до збоїв. Перспективними є дослідження з розробки комплексних моделей захисту даних, які поєднують елементи

незмінної інфраструктури з багаторівневими системами контролю доступу.

Четвертий напрям присвячено аналізу архітектурних рішень та моделей проєктування для підтримки прозорого масштабування незмінних інфраструктур. М. Перрі (M. Perry) [14] здійснює фундаментальний огляд архітектурних принципів побудови незмінних систем, зокрема принципів ідентифікації та автоматизації відновлюваних середовищ. С. Мохаммед (S. Mohammed) та А. Ралеску (A. Ralescu) [15] досліджують перспективи побудови нових архітектурних рішень для системи інтернету майбутнього, підкреслюючи значення незмінної інфраструктури для забезпечення стійкості даних та швидкого масштабування. Перспективи подальших досліджень охоплюють адаптивні моделі проєктування, які дають змогу оптимізувати незмінну інфраструктуру для динамічного масштабування залежно від зміни навантажень та вимог до безпеки.

Отже, аналіз свідчить про актуальність та перспективність технологій незмінної інфраструктури, що потребує подальших поглиблених досліджень щодо методологічних аспектів, інтеграційних можливостей із блокчейном, безпекових рішень та архітектурних моделей для забезпечення високої прозорості та надійності масштабування вебсистем.

Попри досягнення в упровадженні контейнеризації та незмінної інфраструктури, залишаються невирішеними питання інтеграції цих технологій у вебпортали з різними архітектурними моделями без порушення цілісності даних та втрати доступності. Відсутність узагальненої моделі, що поєднує логічне сегментування контейнерів, централізовані репозиторії образів та автоматизовані системи моніторингу, ускладнює масштабування систем без простоїв. Також недостатньо вивчено вплив незмінних контейнерних образів на відновлюваність після аварійних ситуацій, особливо в контексті динамічних навантажень, що є критичним для забезпечення безперервного функціонування вебпорталів. Крім того, потебують глибшого розгляду аспекти впровадження стратегій «blue-green deployment» та «canary deployment» для зниження ризиків під час оновлення системи та забезпечення узгодженості між компонентами. Невизначеність щодо оптимальних підходів до управління ресурсами під час пікових навантажень обмежує можливість реалізації повноцінної моделі резервного копіювання та автоматизованого відновлення. Запропоноване дослідження спрямоване на усунення цих прогалин шляхом розробки інтеграційної моделі, що враховує специфіку різних архітектур та забезпечує стабільність системи в разі збоїв.

Постановка завдання. Мета статті – розробити метод забезпечення прозорого масштабування вебпорталів без простоїв на основі технологій контейнеризації та незмінної інфраструктури з акцентом на підвищення показників доступності та відновлюваності в межах концепції надійності програмного забезпечення.

Завдання статті:

1. Визначити технологічні основи застосування контейнеризації для реалізації незмінної інфраструктури у вебпорталах, проаналізувавши їхній вплив на процеси масштабування та забезпечення узгодженості між компонентами системи.

2. Розробити інтеграційну модель контейнеризації та незмінної інфраструктури для забезпечення прозорого масштабування вебпорталів без простоїв та оцінити її вплив на показники доступності, відновлюваності та архітектурної стійкості системи.

3. Сформулювати рекомендації для адаптації запропонованої моделі до різних архітектур вебпорталів з акцентом на підвищення показників надійності в разі динамічних навантажень з урахуванням ризиків автоматизації процесів оновлення та сумісності з наявними компонентами.

Виклад основного матеріалу. Реалізація незмінної інфраструктури у вебпорталах за допомогою контейнеризації ґрунтується на концепції ізоляції застосунків у незалежних середовищах, що гарантує їхню стабільну роботу незалежно від змін у базовій інфраструктурі. За такого підходу створюються цілковито автономні версії програмного забезпечення, які містять усі необхідні компоненти для виконання застосунку, зокрема операційну систему, бібліотеки та залежності. Вебпортали, що впроваджують контейнеризацію, можуть масштабувати свої компоненти шляхом розгортання додаткових контейнерів без необхідності зміни вихідної архітектури. Це мінімізує час простоїв під час оновлення програмного забезпечення та гарантує його стабільну роботу при збільшенні навантаження. Водночас використання контейнеризації створює можливості для впровадження політик автоматичного відновлення контейнерів у разі збою, підвищуючи загальну надійність системи (табл. 1).

На практиці інтеграція контейнеризації для реалізації незмінної інфраструктури у вебпорталах забезпечує високу гнучкість масштабування

Таблиця 1

Технологічні компоненти контейнеризації для реалізації незмінної інфраструктури у вебпорталах

Технологічний компонент	Функціональна роль	Основні інструменти	Приклади застосування
Контейнерні платформи	Забезпечення ізоляції застосунків та їхньої незалежності від базової інфраструктури	Docker, Kubernetes, Podman	Масштабування вебсервісів шляхом додавання контейнерів під час пікових навантажень
Система оркестрації	Управління життєвим циклом контейнерів, автоматизація процесів розгортання та оновлення	Kubernetes, OpenShift, Docker Swarm	Автоматичний перезапуск контейнерів у разі збоїв
Автоматизоване розгортання	Автоматична генерація нових контейнерних образів на основі оновленого коду	Jenkins, GitLab CI/CD, Ansible	Розгортання нових версій застосунків без простоїв
Моніторинг та логування	Відстеження стану контейнерів та виявлення збоїв у реальному часі	Prometheus, Grafana, ELK Stack	Визначення вузьких місць у масштабованих компонентах системи
Репозиторії контейнерів	Зберігання образів контейнерів для швидкого розгортання нових версій	Docker Hub, Quay, Amazon ECR	Швидке оновлення вебпорталів шляхом заміни старих контейнерів новими версіями

Джерело: сформовано на основі [1, 2, 3, 5, 8]

та автоматизації процесів розгортання нових версій програмного забезпечення. Зокрема, застосування контейнерних платформ, таких як Docker або Kubernetes, дає змогу ізолювати застосунки у відокремлених середовищах, що усуває проблему конфліктів між компонентами різних версій та забезпечує узгодженість середовищ розгортання. Системи оркестрації відіграють вирішальну роль у забезпеченні безперервної роботи вебпорталів під час оновлення або збільшення навантаження, оскільки вони автоматично відстежують стан контейнерів та здійснюють перезапуск у разі збою. Це знижує ризики збоїв під час розгортання нових версій застосунків, оскільки усувається залежність від людського фактора та мінімізується вплив людських помилок.

Автоматизовані процеси розгортання на основі Jenkins або GitLab CI/CD систематично оновлюють застосунки без зупинки основних сервісів, що є критичним для динамічних вебпорталів зі значною кількістю активних користувачів. Водночас моніторингові інструменти на кшталт Prometheus та Grafana дають можливість оперативно виявляти аномалії в роботі контейнерів, що особливо важливо для вебпорталів із критичними бізнес-процесами [4]. Використання централізованих репозиторіїв для зберігання образів контейнерів спрощує управління версіями та сприяє швидкому відновленню системи в разі виникнення непередбачуваних збоїв. Такий підхід є одним з основних чинників підвищення надійності програмного забезпечення у вебпорталах, що функціонують в умовах динамічних навантажень та вимог до високої доступності.

Упровадження незмінної інфраструктури у вебпорталах істотно впливає на показники надійності програмного забезпечення, оскільки забезпечує стабільність його функціонування незалежно від оновлень та змін у системі. Основним принципом такої інфраструктури є створення постійних образів застосунків, що унеможливлює зміни в уже розгорнутих версіях програмного забезпечення. Це дає змогу уникнути непередбачуваних збоїв під час масштабування або розгортання нових компонентів, оскільки всі версії ідентичні та уніфіковані за своїм складом і конфігурацією. Незмінна інфраструктура забезпечує можливість оперативного відновлення системи в разі збою шляхом перезапуску контейнерів із раніше збереженими версіями застосунків. Водночас автоматизація процесів моніторингу сприяє соєчсному виявленню потенційних загроз для доступності системи на ранніх етапах, що знижує ризики простоїв і втрати даних (табл. 2).

Так, незмінна інфраструктура забезпечує високий рівень доступності вебпорталів через

Таблиця 2

Вплив упровадження незмінної інфраструктури на показники надійності вебпорталів

Показник надійності	Функціональна роль	Вплив незмінної інфраструктури	Приклади застосування
Доступність	Забезпечення безперервного функціонування системи під час оновлень	Зменшує ризик простоїв шляхом автоматичного перезапуску контейнерів	Актуально для фінансових платформ зі значною кількістю транзакцій
Відновлюваність	Оперативне відновлення системи після збою	Автоматичне створення резервних копій контейнерів та швидке їхнє відновлення	Застосовується в CRM-системах для збереження даних клієнтів
Взаємозв'язок компонентів	Узгодженість між версіями програмного забезпечення	Застосування ідентичних образів контейнерів у різних середовищах	Упроваджується в системах доставлення контенту для зменшення ризику несумісності
Виявлення аномалій	Моніторинг системи в режимі реального часу	Застосування інструментів моніторингу для ідентифікації потенційних збоїв	Сервіси аналітики для визначення аномалій у трафіку
Автоматизація процесів	Зниження впливу людського фактора на процес оновлення	Автоматичне розгортання контейнерів зі стабільними образами	Інтеграція в мікросервісні архітектури для динамічного масштабування

Джерело: сформовано на основі [1, 2, 4, 6, 7, 10]

можливість безперервного розгортання оновлень без зупинки основних сервісів. Наприклад, у банківських платформах або системах електронної комерції, де кожна секунда простою може призвести до втрати доходів, упровадження незмінної інфраструктури забезпечує стабільність системи під час оновлення програмного забезпечення. У разі виникнення збою система автоматично перезапускає контейнер з останньою стабільною версією, що мінімізує тривалість простою та зберігає цілісність даних.

Крім того, автоматизовані процеси моніторингу дають можливість оперативно виявляти відхилення в роботі системи та запобігати їхньому переростанню в критичні збої. Наприклад, застосування Prometheus у поєднанні з Grafana виявляє аномалії в роботі вебпорталів, аналізуючи метрики продуктивності контейнерів у режимі реального часу. Це забезпечує своєчасне реагування на потенційні загрози та дає змогу підтримувати високу доступність системи навіть за умов підвищеного навантаження.

Модель інтеграції контейнеризації та незмінної інфраструктури для забезпечення прозорого масштабування вебпорталів без простоїв спрямована на уніфікацію процесів розгортання та оновлення компонентів системи з акцентом на архітектурну стійкість та динамічне масштабування. Вона ґрунтується на концепції створення автономних контейнерів із заздалегідь налаштованими образами застосунків, які можуть бути швидко розгорнуті в разі збою або збільшення навантаження. Відмінність цієї моделі від наявних рішень полягає у використанні однотипних образів застосунків незалежно від середовища виконання, що усуває конфлікти версій та забезпечує оперативне відновлення системи без зупинки основних сервісів. Упровадження такої моделі дає змогу централізувати управління компонентами, контролювати їхню продуктивність у режимі реального часу та автоматизувати процеси резервного копіювання, що критично важливо для підтримки доступності та відновлюваності вебпорталів із високим рівнем динамічних навантажень (табл. 3).

У сучасних умовах така модель є особливо актуальною для динамічних вебпорталів, де критичне значення має забезпечення безперервного доступу до сервісів навіть у періоди пікового навантаження. Наприклад, у сфері електронної комерції інтеграція єдиного репозиторію образів дає змогу швидко замінювати застарілі версії застосунків без необхідності зупинки основних сервісів, зберігаючи цілісність транзакційних даних. Логічне сегментування контейнерів із застосуванням Istio або Linkerd дає змогу гнучко керувати мережевими запитами та локалізувати збої в певних компонентах системи, не впливаючи на роботу інших модулів. У разі аварійних ситуацій процес аварійного відновлення автоматично перемикає трафік на резервні контейнери, зберігаючи стабільність системи до завершення відновлювальних робіт.

Упровадження контейнеризації та незмінної інфраструктури у вебпорталах супроводжується низкою проблем, які можуть істотно вплинути на

Таблиця 3

Компоненти інтеграційної моделі для прозорого масштабування вебпорталів

Компонент моделі	Функціональна роль	Технологічна основа	Вплив на архітектурну стійкість
Єдиний репозиторій образів	Зберігання стандартизованих образів застосунків для швидкого розгортання	Amazon ECR, Docker Hub	Забезпечує одноманітність версій та швидке відновлення після збою
Автоматичне оновлення контейнерів	Динамічне оновлення компонентів без зупинки сервісів	Kubernetes, OpenShift	Знижує ризик втрати даних та запобігає простоям
Логічне сегментування контейнерів	Відокремлення основних функціональних модулів для незалежного масштабування	Istio, Linkerd	Підвищує стійкість до локальних збоїв у компонентах
Контейнерні сесії	Відстеження стану застосунків та моніторинг продуктивності	Prometheus, Grafana	Виявляє вузькі місця та прогнозує можливі збої
Процес аварійного відновлення	Автоматичний запуск стабільних версій контейнерів у разі збою	Jenkins, GitLab CI/CD	Забезпечує миттєве перемикання на резервні контейнери

Джерело: власна розробка авторів

стабільність та продуктивність системи. Однією з основних проблем є сумісність контейнерних рішень із наявними архітектурними моделями, особливо у випадках, коли вебпортали побудовані на монолітних системах або використовують застарілі компоненти. У таких умовах інтеграція контейнерних платформ може призвести до конфліктів між мікросервісами, різними версіями програмного забезпечення та базами даних [4]. Це ускладнює забезпечення узгодженості між компонентами системи, що може стати причиною збоїв або втрати даних під час масштабування або оновлення.

Ще однією проблемою є оптимізація ресурсів під час розгортання контейнерів, оскільки кожен контейнер потребує певного обсягу обчислювальних ресурсів та пам'яті [15]. Неправильне налаштування параметрів може призвести до перевантаження системи, надмірного споживання ресурсів або недостатньої потужності для обробки пікових навантажень [8]. Це особливо актуально для вебпорталів, які обслуговують велику кількість користувачів у режимі реального часу, оскільки будь-яка затримка або збій може негативно вплинути на користувацький досвід.

Автоматизація процесів оновлення контейнерів є ще одним критично важливим аспектом, який потенційно може створити ризики для стабільності вебпорталів [9]. У випадку неправильного налаштування системи CI/CD або некоректного тестування оновлених образів контейнерів можливе виникнення ситуацій, коли оновлення запускається одночасно для всіх компонентів системи. Це може призвести до одночасного зупинення кількох основних сервісів, що своєю чергою спричинить значний простій системи. У контексті незмінної інфраструктури проблема ускладнюється через необхідність зберігання попередніх версій контейнерів, що підвищує вимоги до сховищ даних та витрати на обчислювальні ресурси.

Іншою проблемою є моніторинг контейнерів та прогнозування їхньої продуктивності в реальному часі [11]. Упровадження контейнеризації потребує інтеграції додаткових систем моніторингу та логування, що може ускладнити аналіз великих обсягів даних та виявлення аномалій у роботі контейнерів [14]. У разі недостатнього рівня автоматизації моніторингу існує ризик несвоєчасного реагування на потенційні збої, що може спричинити неочікувані простої або втрату даних.

Також актуальним є питання безпеки контейнерів, оскільки використання незмінних образів застосунків передбачає збереження певних конфігурацій, доступів та змінних середовищ, які можуть стати вразливими для зовнішніх атак [6]. Уразливості можуть виникати у разі неправильного налаштування прав доступу до репозиторіїв контейнерів або відсутності перевірок автентичності образів. Це створює додаткові загрози для конфіденційності даних та стабільності вебпорталів, особливо якщо вони обробляють чутливу інформацію або фінансові транзакції.

Адаптація запропонованої моделі інтеграції контейнеризації та незмінної інфраструктури для вебпорталів із різними типами архітектур потребує врахування особливостей кожного типу сис-

теми, оптимізації процесів автоматизованого розгортання та налаштування моніторингових інструментів для забезпечення стабільної роботи під час динамічних навантажень. Для монолітних архітектур доцільно реалізувати поетапну міграцію компонентів до контейнерів із застосуванням стратегій «blue-green deployment» або «canary deployment», що дає змогу тестувати нові версії без ризику для основних сервісів. Застосування контейнеризації для критичних компонентів, таких як бази даних або API-сервіси, забезпечить їхню незалежність від інших модулів та підвищить стійкість до збоїв.

Для мікросервісних архітектур основна увага має бути зосереджена на впровадженні логічного сегментування контейнерів із чітко визначеними зонами відповідальності кожного модуля. Це сприятиме реалізації гнучкої оркестрації на основі Kubernetes або OpenShift, забезпечуючи автоматичне масштабування контейнерів у разі підвищеного навантаження. Рекомендовано налаштувати політики самооновлення контейнерів на основі метрик продуктивності, що мінімізує ризики збоїв та запобігає одночасному оновленню всіх модулів системи.

Для серверлес архітектур ефективність запропонованої моделі досягається шляхом інтеграції подієвих тригерів для активації нових контейнерів у разі різкого збільшення кількості запитів. Застосування автоматичних контейнерних сесій із прогнозуванням пікових навантажень забезпечує безперервність обслуговування навіть за умов раптового збільшення обчислювальних операцій. Доцільно впровадити моніторингові системи з підтримкою машинного навчання для прогнозування пікових навантажень та своєчасного масштабування контейнерів без зупинки основних сервісів.

Для архітектур із розподіленими сервісами доцільно застосовувати концепцію логічних кластерів, які об’єднують функціонально споріднені контейнери для спрощення управління та моніторингу. Такий підхід дає змогу централізовано контролювати стан контейнерів, прогнозувати можливі збої та реалізовувати автоматизовані сценарії відновлення з використанням резервних образів застосунків. Інтеграція політик аварійного відновлення на основі реплікації даних у різних середовищах підвищить відновлюваність системи в разі катастрофічних збоїв або втрати даних.

Таким чином, адаптація моделі для вебпорталів із різними типами архітектур має ґрунтуватися на диференційованому підході до впровадження контейнеризації, сегментування функціональних модулів та інтеграції інструментів моніторингу й аварійного відновлення. Оптимізація політик масштабування та автоматизація оновлень забезпечать стабільність системи навіть під час пікових навантажень, сприяючи підвищенню показників доступності та відновлюваності.

Висновки. Дослідження підтвердило, що інтеграція контейнеризації та незмінної інфраструктури у вебпорталах підвищує надійність програмного забезпечення завдяки ізоляції компонентів та автоматизації процесів розгортання без простоїв. Запропонована модель містить стандартизовані контейнерні образи з логічним сегментуванням для мінімізації впливу локальних збоїв та оперативного відновлення системи. Виявлено проблеми сумісності з монолітними архітектурами та труднощі з управлінням ресурсами під час пікових навантажень, що потребує впровадження централізованих репозиторіїв образів та систем моніторингу. Рекомендовано використовувати поетапну міграцію компонентів із застосуванням стратегій «blue-green deployment» або «canary deployment», що забезпечить безперервність оновлень без зупинки основних сервісів. Подальші дослідження доцільно зосередити на адаптації запропонованої моделі до архітектур із розподіленими сервісами та автоматизації процесів резервного копіювання із застосуванням інструментів CI/CD.

Список літератури:

1. Stepanov D. Research on Development Methodology Utilizing Immutable Infrastructure and Its Impact on Software Reliability. *18th International Conference on Computer Sciences and Information Technologies (CSIT)*. 2023. P. 119–131. DOI: https://doi.org/10.1109/CSIT61576.2023.10324201

2. Степанов Д., Сенів М. Інтеграція захищеної інфраструктури, контейнеризації та DevSecOps для підвищення надійності роботи вебпорталів. *Науковий вісник НЛТУ України*. 2024. Вип. 34, № 5. С. 144–150. DOI: https://doi.org/10.36930/40340519

3. Bafana M., Abdulaziz A. Immutable Infrastructure in Practice: A Comprehensive Guide to AWS Deployment. *Asian American Research Letters Journal*. 2024. Vol. 1, № 1. P. 1–16. URL: https://aarlj.com/index.php/AARLJ/article/download/6/3 (date of access: 09.05.2025).

4. Davidovic N., Veljkovic N., Stoimenov L. Assuring immutability of public e-government documents using blockchain infrastructure. *Proceedings of the 10th International Conference on Information Society and Technology (ICIST 2020)*. 2020. P. 176–180. URL: https://surl.li/ngfjjy (date of access: 09.05.2025).

5. Davis C. *The Art of Immutable Architecture*. Apress: New York, NY, USA. 2020. URL: https://link.springer.com/book/10.1007/979-8-8688-0288-1 (date of access: 09.05.2025).
6. Shaik M., Bojja G. Advanced Identity Access Management and Blockchain Integration: Techniques, Protocols, and Real-World Applications for Enhancing Security, Privacy, and Scalability in Modern Digital Infrastructures. *Libertatem Media Private Limited*. 2022. URL: https://surl.li/uvikti (date of access: 09.05.2025).
7. Tarkhanov I., Fomin-Nilov D., Fomin M. Application of public blockchain to control the immutability of data in online scientific periodicals. *Library Hi Tech*. 2019. Vol. 37, № 4. P. 829–844. DOI: https://doi.org/10.1108/LHT-12-2018-0186
8. Knutsen K. E., et al. Containerized immutable maritime data sharing utilizing Distributed Ledger Technologies. *Journal of Physics: Conference Series*. 2022. Vol. 2311, № 1. Article 012006. DOI: https://doi.org/10.1088/1742-6596/2311/1/012006
9. Bhujel S., Rahulamathavan Y. A survey: Security, transparency, and scalability issues of NFT's and its marketplaces. *Sensors*. 2022. Vol. 22, № 22. Article 8833. DOI: https://doi.org/10.3390/s22228833
10. Arif T., Jo B., Park J. A Comprehensive Survey of Privacy-Enhancing and Trust-Centric Cloud-Native Security Techniques Against Cyber Threats. *Sensors*. 2025. Vol. 25, № 8. Article 2350. DOI: https://doi.org/10.3390/s25082350
11. Chauhan P., Bansal P. Enhancing Trust and Immutability in Cloud Forensics. *Advances in Intelligent Systems and Computing*. 2021. Vol. 1270. P. 891–901. DOI: https://doi.org/10.1007/978-981-15-8289-9_74
12. Nigam R., Azevedo de Amorim P. H., Sampson A. Modular hardware design with timeline types. *Proceedings of the ACM on Programming Languages*. 2023. Vol. 7, № PLDI. P. 343–367. DOI: https://doi.org/10.5281/zenodo.7709916
13. Indrasiri K., Suhothayan S. *Design Patterns for Cloud Native Applications*. O'Reilly Media, Inc. 2021. URL: https://surl.li/ibnkho (date of access: 09.05.2025).
14. Perry M. *Cloud Native Patterns: Designing Change-Tolerant Software*. Simon and Schuster. 2019. URL: https://surl.li/ewjluf (date of access: 09.05.2025).
15. Mohammed S. A., Ralescu A. L. Future internet architectures on an emerging scale – a systematic review. *Future Internet*. 2023. Vol. 15, № 5. Article 166. DOI: https://doi.org/10.3390/fi15050166

Stepanov D.S., Seniv M.M. APPLICATION OF IMMUTABLE INFRASTRUCTURE TECHNOLOGIES TO ENSURE TRANSPARENT SCALING WITHOUT DOWNTIME IN WEB PORTALS

The relevance of the study is determined by the need for stable functioning of web portals in conditions of dynamic load growth and the need for continuous software updates without data loss and downtime. The use of traditional architectural approaches often does not provide an adequate level of reliability, as scaling leads to conflicts between application versions, component incompatibility, and data integrity violations during updates. The lack of unified solutions for rapid system recovery in the event of a failure makes it difficult to maintain high availability of web portals, especially in the context of multi-component systems with a distributed architecture.

The goal of this research is to develop a method for transparent scaling of web portals based on the integration of containerization and immutable infrastructure technologies, with a focus on improving availability and recoverability within the concept of software reliability.

The research methodology includes a comparative analysis of existing approaches to containerization, modeling of architectural schemes for integrating immutable infrastructure into web portals, as well as a structural and functional analysis of monitoring and disaster recovery systems. The impact of containerization on the stability of web portal components in the case of dynamic scaling has been revealed.

The scientific novelty of the research lies in the development of an integration model that provides automated component updates without losing system availability by applying blue-green deployment and canary deployment strategies.

It has been established that the introduction of containerization in combination with an unchanging infrastructure increases the overall stability of the system to failures, minimizing the influence of the human factor and ensuring the consistency of component versions. It has been proven that the use of centralized repositories significantly reduces recovery time after a failure, and the use of container sessions with monitoring ensures timely response to potential threats.

Prospects for further research include adapting the developed model to distributed service architectures and integrating peak load prediction methods using machine learning tools for timely scaling without stopping core services.

Key words: *immutable infrastructure, containerization, scaling without downtime, software reliability, recoverability, update automation.*

УДК 519.1:519.7:007.5
DOI https://doi.org/10.32782/2663-5941/2025.3.2/50

Стьопкін А.В.
ДВНЗ «Донбаський державний педагогічний університет»

РОЗПІЗНАВАННЯ НЕОРІЄНТОВАНИХ ГРАФІВ МУЛЬТИАГЕНТНОЮ СИСТЕМОЮ

Дане дослідження присвячено вирішенню проблеми розпізнавання скінчених зв'язних неорієнтованих графів без петель та кратних ребер мультиагентними системами. Метою дослідження є розробка нового ефективного алгоритму розпізнавання графів мультиагентною системою. В статті використовується наступна методологія досягнення поставленої мети: використати мультиагентну систему, що складається з двох агентів двох різних типів. Перший агент – це агент-дослідник, який може пересуватися по графу, зчитувати та змінювати мітки на відвіданих елементах графу та обмінюватись повідомленнями з другим агентом. Другий агент – це агент-експериментатор – що знаходиться за межами досліджуваного графу, та займається лише побудовою мапи досліджуваного графу на основі повідомлень, отриманих від агента-дослідника. Агент-дослідник має скінчену пам'ять, яка не залежить від розмірності досліджуваного графа. Агент-експериментатор, в свою чергу, має пам'ять, яка залежить від розмірності графа, внаслідок того, що мапа графу зберігається в його пам'яті. В роботі наведено режими функціонування агента-дослідника із зазначенням пріоритетності активації цих режимів в процесі роботи. Також наведено алгоритм функціонування агента-експериментатора з описом процедур обробки отриманих від агента-дослідника повідомлень. В роботі проведено аналіз часової, ємнісної, комунікаційної складностей побудованого алгоритму та кількість переходів по ребрах, які необхідно виконати агенту-досліднику для повного розпізнавання досліджуваного графа. Науковою новизною є отримання ефективного алгоритму розпізнавання скінчених зв'язних неорієнтованих графів без петель та кратних ребер, який дозволяє використовувати для розпізнавання дві фарби різного кольору та один камінь. Таким чином, в роботі запропоновано новий алгоритм розпізнавання графів, який має квадратичні (від числа вершин досліджуваного графа) часову, ємнісну та комунікаційну складності. Кількість переходів по ребрах, які необхідно виконати агенту-досліднику для повного розпізнавання досліджуваного графа оцінюється зверху як $O\left(n^2\right)$*, де* n *– кількість вершин досліджуваного графа. Робота запропонованого алгоритму ґрунтується на методі обходу графа в глибину.*

Ключові слова: *розпізнавання графів, пості графи, складність алгоритму, обхід в глибину, мультиагентна система.*

Постановка проблеми. В наш час досить стрімко розвивається робототехніка, вона набуває все більшої популярності серед молоді та стає все більш затребуваною в різноманітних напрямках діяльності людини. Що зробило її однією з провідних галузей кібернетики. Її витоки пов'язані з прагненням людства полегшити виконання важких або небезпечних для людини завдань, а іноді й повністю замінити людину в таких видах діяльності.

З огляду на те, що залишається велика кількість недосліджених середовищ [1] через їхню недоступність або небезпеку для людини, стає очевидною актуальність задачі їх дослідження за допомогою роботів [2–4]. Зрозуміло, що використовуючи при цьому не одного, а декількох роботів [5–8] підвищується ймовірність успішного виконання поставленого завдання, зокрема у випадках виходу з ладу окремих агентів під час роботи. Використання декількох агентів також може позитивно вплинути на часову чи ємнісну складність побудованих алгоритмів.

Відома значна кількість наукових праць присвячених дослідженню графів за допомогою одного агента, в яких отримано численні результати щодо можливості такого розпізнавання та його складності. Натомість проблема дослідження графів мультиагентною системою, залишається недостатньо вивченою. Саме тому актуальними є дослідження мультиагентних систем, пов'язані з систематичним проведенням експериментів та їх аналізом. До таких задач належать побудова маршрутів переміщення агентів по невідомому графу, маркування його елементів, збирання та обробка локальної інформації про граф, а також побудова мапи графа за цією інформацією, зокрема й з урахуванням нанесених міток. Крім того, важливими залишаються задачі

оптимізації використання ресурсів, витрат часу, навантаження на канали зв'язку, тощо.

Аналіз останніх досліджень і публікацій. Початок досліджень у цьому науковому напрямі зазвичай пов'язують із роботою C. Shannon [9], у якій розглядалася задача пошуку автоматом заданої цілі в лабіринті. Активні дослідження поведінки автоматів у лабіринтах розпочалися у 1971 році після публікації роботи K. Dopp [10], де описувався обхід шахових лабіринтів скінченними автоматами. Згодом умови задачі розширювалися і автомату необхідно вже було відвідати всі вершини та ребра досліджуваного графа (лабіринту). У подальшому, в роботі [11], було проведено дослідження властивостей невідомого середовища залежно від способу взаємодії автомата з середовищем, а також за різної апріорної інформації про це середовище.

Насправді, дослідження з аналізу графів включають низку окремих завдань. Наприклад, задача самолокалізації (визначення вершини графа, в якій агент знаходиться на початку дослідження), задача контролю карти (перевірка ізоморфізму між досліджуваним графом і еталонною картою), задача повного розпізнавання графа (побудова мапи досліджуваного графа).

Для вирішення останньої задачі існує декілька підходів у відповідності до яких розроблено низку алгоритмів обходу графів мобільними агентами та способів маркування елементів графа цими агентами за допомогою різних фарб або камінців, які дозволяють розпізнавати різноманітні графи з точністю до ізоморфізму.

Так, у роботі [12] запропоновано метод, за яким агент під час руху маркує лише інцидентори пройдених ребер. Для дослідження ребер агент проходить по ним, а потім повертається до початкової вершини найкоротшим шляхом. При цьому він запам'ятовує мітки пройдених інцидентрів, які однозначно визначають вершину, до якої агент дійшов під час руху цим ребром. У роботі [13] агенти обмінюються інформацією про виконані завдання з базовою станцією через спеціальну мережу, що має обмеження на обсяг переданих даних. Інакше кажучи, для координації дій між агентами попередньо формується мережа, яка відповідає вказаним умовам. Обмін даними з базовою станцією відбувається у заздалегідь визначених точках. У статті [5] зроблено спробу розробити загальні концепції та вимоги до задачі розпізнавання графа мультиагентними системами. Запропоноване представлення мультиагентної системи орієнтоване на визначення загальних умов завершення процесу розпізнавання графа та забезпечення його завершення за скінченний час. У роботі запропоновано модифікацію матриці інцидентності для організації інформаційного обміну між агентами, а також наведено загальний алгоритм колективної роботи агентів під час дослідження графа. Алгоритм базується на запропонованій узагальненій структурі та модифікованій матриці інцидентності. У роботі [7] розглядається нескінченний в усі сторони граф квадратної решітки. Усі його вершини позначено позначками з деякої відомої множини, але конкретна функція розмітки невідома. Завданням автомата є визначення двох двійок протилежних напрямків на укладанні вищезгаданого графа на цілочисельну решітку, або розпізнавання розмітки. Доведено, що поодинокий автомат не може розпізнати мінімальну детерміновану розмітку нескінченого графа квадратної решітки, але автомат з одним каменем може. Запропоновано новий тип експериментів з графами, призначений для розпізнавання їх розмітки та запропоновано метод побудови та проведення розпізнавального експерименту для нескінченого позначеного графа квадратної решітки.

Постановка завдання. Проаналізувавши дослідження з роботи мультиагентних систем, можна прийти до висновку, що в них мало уваги приділено вивченню процесу обміну інформацією агентами. Що робить актуальними проведення експериментів по розпізнаванню графів мультиагентними системами та оптимізацію витрат часу, ресурсів та навантаження на канал зв'язку між агентами.

Таким чином, метою даного дослідження є створення ефективного алгоритму розпізнавання скінчених зв'язних простих неорієнтованих графів мультиагентною системою, ефективність якого або перевищить раніше розроблені алгоритми, або зменшить кількість використовуваних в них ресурсів [10, 14–17]. Тобто важливою частиною нашої роботи є дослідження часової, ємнісної й комунікаційної складностей побудованого алгоритму та кількості переходів по ребрах, які необхідно виконати агенту-досліднику для повного розпізнавання досліджуваного графа.

Виклад основного матеріалу. В роботі розглядаються зв'язні неорієнтовані прості скінчені графи $G=(V,E)$, де V – множина вершин графа, а E – множина ребер (двоелементних підмножин (u,v), де $u,v\in V$) графа. Трійки $((u,v),v)$ називатимемо інцидентрами

(точкою з'єднання) ребра (u,v) та вершини v. Через I позначимо множину всіх інциденторів графа. Множину $L=V\cup E\cup I$ назвемо множиною всіх елементів графа G. Функцією розфарбування графа G назвемо сюр'єктивне відображення $\mu: L\to\{w,r,b\}$, де w – білий колір, r – червоний, а b – чорний. Пару (G,μ) будемо називати розфарбованим графом. Послідовність $u_1,u_2,...,u_k$ попарно суміжних вершин графа G назвемо шляхом довжини k. Околом $Q(v)$ вершини v будемо називати множину елементів графа, яка складається з вершини v, всіх вершин u, суміжних з v, всіх ребер (v,u) та всіх інциденторів $((v,u),v)$ та $((v,u),u)$. Потужності множини вершин V та множини ребер E позначимо через n та m відповідно. Зрозуміло, що $m\le\frac{n(n-1)}{2}$. Ізоморфізмом графа G та графа H назвемо таку бієкцію $\phi: V_G\to V_H$, що $(v,u)\in E_G$ тоді й тільки тоді, коли $(\phi(v),\phi(u))\in E_H$. Таким чином, ізоморфні графи рівні з точністю до позначення вершин і розфарбування їх елементів.

В статті розглядається мультиагентна система, що складається з двох агентів: один агент-дослідник і один агент-експериментатор. Агент-дослідник може переміщуватися графом і змінювати забарвлення елементів графа. Агент-експериментатор відповідає за побудову мапи графа в своїй пам'яті на основі повідомлень, отриманих від агента-дослідника.

На початку роботи агент-дослідник розміщується в довільній вершині графа G та фарбує її в свій колір. Агент-експериментатор в цей момент додає нову вершину в множину V_H в своїй пам'яті. Агент-дослідник пересувається по графу з вершини v у вершину u по ребру (v,u), може змінювати забарвлення вершин v,u, ребер (v,u), інциденторів $((v,u),v),((v,u),u)$. Перебуваючи у вершині v, агент-дослідник може сприймати мітки всіх елементів околу $Q(v)$ та виходячи з цієї інформації визначає, яким ребром буде далі переміщатися та як фарбуватиме елементи графа. Агент-експериментатор, в свою чергу, передає, приймає та ідентифікує повідомлення, отримані від агента-дослідника та фіксує в своїй пам'яті результат функціонування агента-дослідника на кожному кроці. На основі цієї інформації агент-експериментатор будує мапу досліджуваного графа G, спочатку невідомого агентам, списками ребер і вершин.

Зауважимо, що запропонований алгоритм має декілька особливостей: по-перше, граф G агентам не відомий; по-друге, запропонований метод розпізнавання графа ґрунтується на стратегії пошуку в глибину; по-третє, при обході графа G агенти створюють неявну нумерацію пройдених вершин. На основі побудованої нумерації й відбувається розпізнавання графа G шляхом побудови графа H ізоморфного G. У процесі обходу агент-дослідник будує неявне дерево пошуку в глибину всі ребра якого поділяються на ті, що забарвлюються при першому проходженні в червоний колір та належать самому дереву й ті, що не належать дереву і забарвлюються при першому проходженні в чорний колір – зворотні ребра. Вершина, у якої всі інцидентні їй ребра розпізнані, забарвлюється в чорний колір. Алгоритм закінчує роботу, коли червоний шлях стає порожнім, а всі вершини чорними.

У роботі агента-дослідника є два режими:

Звичайний режим. Агент-дослідник рухається вперед по білих вершинах, забарвлюючи вершини, ребра, що з'єднують ці вершини та дальні інциденти в червоний колір. Якщо в околі вершини в якій знаходиться агент-дослідник немає білих ребер чи вершин, тобто відсутні можливі шляхи переміщення, то агент-дослідник повертається назад, забарвлюючи пройдені вершини, ребра та ближні інциденти у чорний колір. Повернувшись до початкової вершини, агент-дослідник завершує роботу. На кожному кроці агент-дослідник обмінюється даними з агентом-експериментатором.

Режим розпізнавання зворотних ребер. Якщо при русі вперед у вершині v було виявлено зворотне ребро, то агент-дослідник припиняє роботу у звичайному режимі та перемикається у режим розпізнавання зворотних ребер. Агент-дослідник залишає в вершині камінь. Після цього він починає рухатись назад своїм шляхом, до виявлення вершини інцидентної зворотному ребру (тут йдеться мова про біле ребро, у якого дальня вершина пофарбована в червоний колір і в ній знаходиться камінь) та переходить по цьому ребру, фарбуючи його в чорний колір. Далі повертається до попередньої вершини та продовжує рух назад. На цьому етапі можливі два випадки. Випадок, коли агент-дослідник знайде зворотне ребро. У цьому випадку агент-дослідник повторює процедуру розпізнавання зворотного ребра. Та випадок, коли агент-дослідник досягнув початкової вершини, що буде означати, що ним розпізнано всі помічені зворотні ребра. Після чого агент повертається в кінець червоного шляху, де знаходиться камінь, підбирає камінь та продовжує роботу в звичайному режимі.

Алгоритм роботи агента-експериментатора:

Вхід: список список повідомлень M від агента-дослідника.

Вихід: список вершин V_H та ребер E_H графа H, ізоморфного графу G.

Дані: V_H, E_H – списки вершин и ребер графа H, ізоморфного графу G. ct – лічильник кількості відвіданих вершин графа G агентом-дослідником. M – список повідомлень від агента-дослідника. i – лічильник, що використовується для з'ясування номерів дальніх вершин зворотних ребер після того як встановлено камінь у їх спільну вершину. $STOP$ – змінна, що використовується агентом-дослідником, для повідомлення агенту-експериментатору, про завершення обходу графа. $r(1)\ldots r(t)$ – список номерів вершин червоного шляху, де t – довжина цього списку. Mes – змінна, в якій зберігається поточне повідомлення.

Алгоритм роботи агента-експериментатора:

1. $ct := 1, M := \varnothing, i := 0, E_H := \varnothing, STOP := 0$, t:=1, $r(t)$:=ct, $V_H := \{1\}$, $Mes :=$ "";
2. while $STOP = 0$ do
3. if $M \neq \varnothing$,
4. then do
5. зчитуємо в Mes повідомлення і видаляємо його з черги M;
6. LIST_PROCESSING();
7. end do;
8. end do;
9. друк V_H, E_H.

LIST_PROCESSING():

1. if $Mes :=$ "$FORWARD$"; then FORWARD();
2. if $Mes :=$ "$BACK$" then BACK();
3. if $Mes :=$ "$RETURN$" then RETURN();
4. if $Mes :=$ "$FORWARD_IE$" then FORWARD_IE();
5. if $Mes :=$ "$RECOGNIZED_IE$" then RECOGNIZED_IE();
6. if $Mes :=$ "$STOP$" then STOP();

FORWARD(): $ct := ct + 1$; $t := t + 1$; $r(t) := ct$; $V_H := V_H \cup \{ct\}$; $E_H := E_H \cup \{r(t-1), r(t)\}$;

BACK(): Зі списку $r(1)\ldots r(t)$ видаляється елемент $r(t)$; $t := t - 1$;

RETURN(): $i := i + 1$;

FORWARD_IE(): $E_H := E_H \cup \{r(t), r(t-i)\}$;

RECOGNIZED_IE(): $i := 0$;

STOP(): $STOP := 1$;

На початку роботи алгоритму розпізнавання, при $n \geq 3$, як мінімум два рази, коли агент-дослідник відвідує білі вершини досліджуваного графа G, виконується процедура FORWARD(v). При цьому процедурою FORWARD() створюється нова вершина графа H. Таким чином, процес виконання даного алгоритму індукує відображення $\varphi : V_G \to V_H$ вершин графа G в вершини графа H. Причому рівність $\varphi(v) = ct$ встановлюється, коли вершина v фарбується в червоний колір.

Зазначене відображення природнім чином встановлює неявну нумерацію вершин графу G. Більш того, відображення φ є бієкцією, оскільки розглядаються виключно зв'язні графи, а в таких графах всі вершини досяжні з будь-яких початкових вершин. Тому всі вершини досягаються агентом-дослідником і при першому ж відвідуванні вершини, агенти надають їй неявний номер.

З опису алгоритму видно, що агент-дослідник проходить абсолютно всі ребра графу G, оскільки після виконання алгоритму всі ребра графу зафарбовані в чорний колір.

Виконуючи процедуру FORWARD() агент-експериментатор розпізнає деревне ребро (v,u) і так нумерує вершину u, що ребру (v,u) графа G однозначно відповідає ребро $(\varphi(v), \varphi(u))$ графа H. Виконуючи процедуру FORWARD_IE(), агент-експериментатор розпізнає зворотне ребро (v,u) графа G і ставить йому в однозначну відповідність ребро $(\varphi(v), \varphi(u))$ графа H. А так як всі ребра графа поділяються виключно на деревні та зворотні, то можна зробити висновок, що відображення φ є ізоморфізмом графа G на граф H.

Теорема. *Два агенти, виконавши алгоритм розпізнавання графа, розпізнають будь-який граф з точністю до ізоморфізму.*

Розглянемо докладніше властивості червоного шляху. З опису алгоритму слідує, що на кожному кроці алгоритму червоний шлях – це простий шлях, що бере початок вершині v з номером $\varphi(v) = 1$, що закінчується у вершині u з номером $\varphi(u) = ct$. Отже, довжина червоного шляху не перевищує n. Щоразу при виконанні процедур FORWARD(v) та BACK(v) агент-дослідник проходить одне ребро. При одноразовому виконанні процедури з режиму розпізнавання зворотних ребер агент-дослідник розпізнає не більше $n-2$ зворотних ребер в процесі чого: витрачає один хід на встановлення каменя у вершину, по два рази (прохід назад та повернення в останню червону вершину) проходить не більше $n-1$ ребер червоного шляху, а також по два рази проходить не більше $n-2$ зворотних ребер при їх фарбуванні.

При підрахунку часової складності алгоритму вважатимемо, що ініціалізація алгоритму і вибір однієї з можливих процедур займають деяку

постійну кількість одиниць часу. Також будемо вважати, що прохід по ребру агента-дослідника та обробка команди агентом-експериментатором, отриманої на даному етапі від агента-дослідника, здійснюється за час, що дорівнює деякій константі. Приймемо, що час пересилання одного повідомлення дорівнює константі, яка не перевищує часу проходу по ребру агента-дослідника. Причому загальний час, що витрачається на аналіз околу $Q(v)$ вершини v та вибір необхідних ребер оцінюється зверху як $O(n^2)$.

Тоді часова складність алгоритму визначається наступними співвідношеннями:

1. Процедури FORWARD(v) і BACK(v) виконуються не більше ніж $2\times(n-1)$ разів та загальний час їх виконання оцінюється як $O(n)$.

2. Час, що витрачається на розпізнавання зворотних ребер оцінюється як $(n-2)\times(2\times(n-1)+2\times(n-2))$ разів, тобто по суті $O(n)\times O(n)$ разів, а отже загальний час виконання процедури розпізнавання зворотних ребер оцінюється як $O(n^2)$.

Таким чином, верхня оцінка числа переходів по ребрах, що здійснюються агентом-дослідником оцінюється зверху як $O(n)+O(n^2)=O(n^2)$, а сумарна часова складність $T(n)$ алгоритму задовольняє співвідношенню: $T(n)=O(n^2)$.

Ємнісна складність $S(n)$ алгоритму визначається сумарною складністю списків E_H, V_H та $r(1),...,r(t)$ складність яких визначається величинами $O(n^2)$, $O(n)$, $O(n)$ відповідно.

А отже верхня оцінка ємнісної складності алгоритму задовольняє співвідношенню: $S(n)=O(n^2)$.

При підрахунку комунікаційної складності алгоритму, з урахуванням того, що агент-дослідник може відправляти агенту-екпериметатору лише одне з шести можливих повідомлень (*FORWARD,BACK,RETURN*,FORWARD_IE, RECOGNIZED_IE,*STOP*), вважатимемо, що одне повідомлення має обсяг, рівний одиниці пам'яті.

Так як при роботі в звичайному режимі, на кожному кроці каналом зв'язку передається одне повідомлення (*FORWARD*,*BACK*,*STOP*), то загальний обсяг переданої інформації в цьому режимі буде оцінюватись зверху як $2\times O(n)$. При роботі в режимі розпізнавання зворотних ребер, агент-дослідник при русі назад по червоному шляху, покроково відправляє одне сповіщення (*RETURN*), тобто загальний обсяг переданої інформації в цьому режимі також буде оцінюватись зверху як $n\times O(n)$. При першому переході по зворотних ребрах відправляється одне повідомлення (FORWARD_IE) тобто загальний обсяг переданої інформації в цьому режимі також буде оцінюватись зверху як $n\times O(n)$. При переході назад по зворотному ребру ніяких повідомлень не надсилається. Також буде надіслано одне повідомлення (FORWARD_IE), при поверненні в вершину, в якій знаходиться камінь, тобто обсяг переданої інформації в цьому режимі буде оцінюватись зверху як $n\times O(1)$.

Таким чином загальний обсяг інформації, що передається агентами оцінюється як $K(n)=2\times O(n)+n\times(O(n)+O(n)+O(1))=O(n^2)$.

Теорема. *Часова, ємнісна, комунікаційна складності алгоритму розпізнавання та кількість переходів по ребрах, що здійснюються агентом-дослідником, оцінюються зверху як* $O(n^2)$*, де* n *– кількість вершин в графі. Для розпізнавання достатньо двох фарб різного кольору.*

Висновки. Дослідження присвячено проблемі розпізнавання графів мультиагентною системою. Запропоновано новий алгоритм розпізнавання скінчених зв'язних неорієнтованих графів без петель та кратних ребер, квадратичної часової, ємнісної та комунікаційної складностей. Верхня оцінка числа переходів по ребрах, що здійснює агент-дослідник для повного розпізнавання графа дорівнює $O(n^2)$. Агент-дослідник має скінчену пам'ять, яка не залежить від розмірності досліджуваного графа. Агент-експериментатор, в свою чергу, має пам'ять, яка залежить від розмірності графа, внаслідок того, що мапа графу зберігається в його пам'яті. Для виконання алгоритму розпізнавання необхідно дві фарби різного кольору і один камінь.

Список літератури:

1. Goth G. Exploring new frontiers. *Communications of the ACM*. 2009. 52 (11). P. 21–23.

2. Sapunov S.V. Experiments on Recognition of Infinite Grid Graph Labelling. *Праці ІПММ НАН України*, 2021. 35 (1), 67–78. URL: https://doi.org/10.37069/1683-4720-2021-35-6

3. Wang H., Jenkin M., Dymond P. Enhancing exploration in graph-like worlds. *Proceedings of the Canadian Conference on Computer and Robot Vision (CRV)*. 2008. P. 53–60.

4. Stopkin A. V. Finite graph exploration by a mobile agent. *Mathematical Modeling and Computing*. 2025. Vol. 12, No. 1, P. 75–82.

5. Nagavarapu S.C., Vachhani L., Sinha A. et al. Generalizing Multi-agent Graph Exploration Techniques. *International Journal of Control, Automation and Systems*. 2020. URL: https://doi.org/10.1007/s12555-019-0067-8

6. Sapunov S.V. Minimal Deterministic Traversable Vertex Labelling of Infinite Square Grid Graph. Праці ІПММ НАН України, 2020. 34, 118–132. URL: https://doi.org/10.37069/1683-4720-2020-34-12

7. Sapunov, S.V. Directional Movement of a Collective of Compassless Automata on a Square Lattice of Width 2. Cybernetics and Systems Analysis, 2024. vol. 60, is. 6, P. 899–906. URL: https://doi.org/10.1007/s10559-024-00727-x

8. Стьопкін А. В. Алгоритм розпізнавання простих неорієнтованих графів колективом агентів. Вчені записки таврійського національного університету імені В.І. Вернадського Серія: Технічні науки. Київ, 2024. Т. 35 (74) № 5. С. 303–309.

9. Shannon C.E. Presentation of a maze-solving machine. *Cybernetics Trans, of the 8 th Conf. of the Josiah-Macy Jr. Found* / Editor: H. Foerster. 1951. P. 173–180.

10. Dopp K. Automaten in labirinthen. *Electronische Informationsverarbeitung und Kybernetik*. 1971. V.7, № 2. P. 79–94.

11. Dudek G., Jenkin M. Computational principles of mobile robotics. Cambridge Univ. press. 2000. 280 p.

12. Dudek G., Jenkin M., Milios E., Wilkes D. Map validation in a graphlike world. *Proceedings of the 13th International Joint Conference on Artifical Intelligence.* San Fransisco: Morgan Kaufmann Publishers Inc., 1993. P. 1648–1653.

13. Banfi, J., Quattrini Li, A., Rekleitis, I. et al. Strategies for coordinated multirobot exploration with recurrent connectivity constraints. *Autonomous Robots*. 2018. Vol. 42. P. 875–894. URL: https://doi.org/10.1007/s10514-017-9652-y

14. Stepkin A. Using a Collective of Agents for Exploration of Undirected Graphs. *Cybernetics and Systems Analysis*. 2015. V.51, № 2. P. 223–233. URL: https://doi.org/10.1007/s10559-015-9715-z

15. Стьопкін А. В. Мультиагентна система розпізнавання неорієнтованих графів. Інформаційні технології та суспільство. Київ, 2024. Вип. 3 (14). 38–43 с.

16. Стьопкін А.В. Мультиагентна система розпізнавання графів. Штучний інтелект, 2025, №2, P. 84–93. URL: https://doi.org/10.15407/jai2025.01.027

17. Dudek G., Jenkin M., Milios E., Wilkes D. (1998) Topological exploration with multiple robots. Robotics with Application (ISORA) : Proc. 7th International Symposium. Alaska.

Stopkin A.V. UNDIRECTED GRAPH EXPLORATION USING A MULTI-AGENT SYSTEM

This study is devoted to solving the problem of graph exploration by multi-agent systems in the class of finite, connected, undirected graphs without loops and multiple edges. The goal of the research is to develop a new efficient algorithm for graph exploration by a multi-agent system.

The proposed methodology involves the use of a multi-agent system consisting of two agents of different types. The first is an agent-researcher, which can move through the graph, read and modify labels on visited elements of the graph, and exchange messages with the second agent. The second is an agent-experimenter, which is located outside the graph and is solely responsible for constructing a map of the explored graph based on the messages received from the agent-researcher.

The agent-researcher is equipped with finite memory, independent of the size of the explored graph. The agent-experimenter, in turn, has memory that depends on the size of the graph, since the map is stored in its memory.

The paper outlines the operating modes of the agent-researcher, indicating the priority for their activation during exploration. It also presents the functioning algorithm of the agent-experimenter, including procedures for processing the messages received from the agent-researcher.

The paper analyzes the time, space, and communication complexities of the proposed algorithm, as well as the number of edge transitions that must be performed by the agent-researcher to completely explore the graph.

The scientific novelty lies in obtaining an efficient algorithm for the exploration of finite, connected, undirected graphs without loops and multiple edges. The algorithm uses two distinct colors and one pebble to exploration.

Thus, the paper proposes a new graph exploration algorithm with quadratic time, space, and communication complexities (with respect to the number of nodes in the graph). The upper bound on the number of edge transitions required by the agent-researcher to fully explore the graph is estimated as $O(n^2)$, where n is the number of nodes in the explored graph. The operation of the proposed algorithm is based on the depth-first search (DFS) strategy.

Key words: *graph exploration, finite graphs, algorithm complexity, depth-first search, multi-agent system.*

УДК 004.932.72
DOI https://doi.org/10.32782/2663-5941/2025.3.2/51

Танчук В.С.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Колобродов В.Г.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

МЕТОД МОДЕЛЮВАННЯ ТЕПЛОВІЗІЙНИХ ЗОБРАЖЕНЬ НА ОСНОВІ СИСТЕМНОГО ПІДХОДУ

В даній роботі запропоновано метод моделювання тепловізійних зображень об'єктів, що базується на системному підході до формування тепловізійного інфрачервоного сигналу. Головною метою цього дослідження є розробка методики моделювання синтетичних тепловізійних зображень з урахуванням фізичних і технічних характеристик реального тепловізора. Відповідна методика є особливо актуальною для задач комп'ютерного зору, де критично необхідно мати якісний та правдивий набір даних, який впливає на ефективність навчання моделей штучного інтелекту (ШІ) для автоматизованого виявлення та розпізнавання цілей та полегшує розробку алгоритмів автоматичних систем спостереження. Також використання моделювання тепловізійних зображень дає змогу адаптувати та підготувати систему до змін, які можуть виникати в процесі експлуатації. Тест-об'єктом спостереження, для якого виконується модуляція в цій роботі, є безпілотний літальний апарат (БПЛА).

Метод складається з двох основних етапів. На першому етапі виконується рендер тепловізійного зображення сцени, враховуючи закон Планка та випромінювальну здатність матеріалів об'єкта, за допомогою програмного пакета 3D моделювання, результатом якого є псевдо тепловізійне зображення з відповідним об'єктом спостереження. Другим етапом є цифрова постобробка, яка імітує вплив оптичної системи (ОС) тепловізора на якість зображення, враховуючи параметри тепловізора та його особливості. Під час пост обробки розраховується вплив об'єктива та мікроболометричної матриці (МБМ) ОС тепловізора, що описується через модуляційну передавальну функцію (МПФ). Для обчислення МПФ враховуються реальні параметри тепловізора, такі як фокусна відстань, розмір вхідної зіниці, розмір пікселя, довжина хвилі випромінювання та частота кадрів. В результаті отримуємо двовимірний частотний фільтр, який застосовується до ідеального зображення у частотній області. Після цього використовуються ще кілька функцій пост-обробки.

В результаті було змодельовані тепловізійні зображення з урахуванням фізичних властивостей теплобачення та характеристик реального тепловізора та проведено їх порівняння з реальними тепловізійними зображеннями.

Ключові слова: *моделювання, тепловізор, пост-обробка, безпілотний літальний апарат, дрон, закон Планка, модуляційна передавальна функція, штучний інтелект.*

Постановка проблеми. Тепловізійні системи спостереження (ТСС) все частіше починають використовуватись у системах автоматичного виявлення та розпізнавання об'єктів. Експлуатація таких систем дає змогу виявити та розпізнати об'єкти в таких умовах, в яких системи зі звичайними камерами видимого діапазону не мають змоги надати відповідну точність. Для прикладу, в нічний час або під час опадів, що спричиняє обмежену видимість об'єктів спостереження. Проте автоматичні ТСС мають ряд недоліків, які можуть значно знизити ефективність роботи такої системи. Одним із таких недоліків є обмежена кількість тепловізійних даних та сцен, які можна використати для навчання моделей розпізнавання, розробки алгоритмів та адаптації системи до постійних змін.

У задачі виявлення та розпізнавання БПЛА ефективне навчання моделей ШІ ускладняється через обмежену кількість відкритих тепловізійних датасетів. Це зумовлено як високою вартістю тепловізійного обладнання, так і технічними труднощами його використання в системах спостереження, що, у свою чергу, обмежує масштабованість польових досліджень та доступність якісних даних. Окрім цього, моделі комп'ютерного зору, які були навчені на даних з одного типу тепловізора, можуть втратити точ-

ність у разі заміни приладу на інший через відмінності оптичних параметрів. Ще однією проблемою є поява нових типів БПЛА. ТСС, яка раніше не мала подібних об'єктів в датасеті під час навчання, може не розпізнати їх під час роботи, а створення нового навчального набору під кожну нову модель дрона є ресурсномістким завданням і вимагає часу, що суттєво знижує адаптивність системи до нових загроз.

У зв'язку з вище зазначеними проблемами, виникає потреба у знаходженні методу модуляції тепловізійних зображень, в якому можна легко змінювати об'єкти розпізнавання. Модуляція тепловізійних зображень повинна враховувати фізичні властивості теплобачення та адаптувати зображення під відповідні параметри тепловізора, який використовується в ТСС.

Аналіз останніх досліджень і публікацій. Останні декілька років усе частіше починають набувати розвитку дослідження систем спостереження з використанням тепловізійних систем та ШІ. Більшість цих досліджень зіштовхуються з проблемою обмеженої кількості тренувальних даних, які можна використовувати для навчання моделей та розробки алгоритмів виявлення та розпізнавання. Ця проблема впливає на точність та ефективність таких систем.

У роботі [1] розглядається ця проблема і рішення, яке пропонує автор, полягає в використанні зображень видимого діапазону, до яких застосовують цифрову обробку зображень для їх перетворення у псевдо тепловізійні. Такий підхід є доволі обмеженим через те, що цифрова обробка складається лише з декількох простих етапів, таких як переведення кольорового зображення у відтінки сірого, переведення у HSV формат та інверсія. Як результат, зображення приймають вигляд псевдо інфрачервоного (ІЧ) діапазону, проте не було враховано фізичні властивості теплобачення та особливості обробки ІЧ сигналів тепловізором. Згідно з результатами дослідження, які були викладені автором, точність виявлення моделей зростає приблизно на 10%, але лише в тому випадку, коли кількість тепловізійних зображень мало. Коли кількість тепловізійних зображень збільшується, то великого приросту в точності не має. Це свідчить про те, що для досягнення високої точності виявлення необхідно мати багато справжніх тепловізійних зображень.

У своєму дослідженні [2], для порівняння різних типів існуючих моделей виявлення об'єктів для задач виявлення БПЛА, було використано вище зазначений метод. До Anti-UAV Challenge датасету [3] було додано зображення дронів з Drone vs Bird Challenge датасету [4] та перетворено їх у псевдо тепловізійні зазначеним вище методом. Цей підхід дозволив збільшити розмір датасету майже у два рази та збільшив точність моделей. Проте точність моделей, які досліджувались, була недостатньою. Для досягнення кращого результату необхідно було збільшити кількість реальних тепловізійних зображень або знайти метод модуляції зображень, які враховують фізичні властивості теплобачення.

У дослідженні [5] запропоновано систему ThermalSynth для створення синтетичних теплових зображень людей, згенерованих за допомогою ігрового рушія Unity, з реальними тепловими фонами. Для апроксимації зовнішнього вигляду синтетичних персонажів у тепловому діапазоні було використано шейдер, заснований на законі Стефана-Больцмана [6]. Також застосовується постобробка отриманих зображень для їх наближення до реальних. Однак закон Стефана-Больцмана описує повне теплове випромінювання тіла та охоплює весь спектр електромагнітного випромінювання, тоді як тепловізійні системи можуть працювати тільки у певному діапазоні довжин хвиль. Тому для більш точної модуляції зображень під певний тип тепловізора краще використати закон Планка [7], який розраховує спектральну густину енергії для певного діапазону довжин хвиль. Проте закон Планка є дещо складнішим для імплементації, ніж закон Стефана-Больцмана. Також в цій роботі не враховувались характеристики ТСС, хоча це є доволі важливим фактором через те, що від типу тепловізора залежить якість отриманих зображень.

В роботі [8] описано новий метод розрахунку ймовірності розрізнення цілі контрастно обмеженою системою БПЛА. Розпізнавання проводиться оператором. Важливим аспектом є те, як враховані всі фізичні властивості теплобачення. Детально описано розрахунок МПФ тепловізійної системи та її вплив на ймовірність розпізнавання оператором цілі. Також завдяки цій роботі можна гарно побачити вплив оптичних параметрів тепловізора на МПФ. Цей метод можна адаптувати під систему виявлення об'єктів з використанням ТСС та ШІ та на основі цього знайти оптимальні параметри тепловізора, при яких ймовірність розпізнавання буде достатньою. Для цього необхідно розрахувати пороговий контраст моделі виявлення об'єктів. Це можна виконати, використавши модуляцію тепловізійних зображень.

Підбиваючи підсумки аналізу останніх досліджень та публікацій, можна сказати, що задача розробки методу моделювання тепловізійних зображень з урахуванням всіх властивостей теплобачення є актуальною.

Постановка завдання. Необхідно розробити метод моделювання тепловізійних зображень об'єктів, зокрема БПЛА, з урахуванням фізичних властивостей теплобачення та характеристик ТСС. Формування 3D сцен та заміна об'єктів спостереження повинна бути швидкою та простою. Змодельовані тепловізійні зображення необхідно порівняти із реальними тепловізійними зображеннями.

Виклад основного матеріалу. *Методика моделювання.* В даній роботі пропонується метод моделювання тепловізійних зображень на основі системного підходу, який складається з двох основних етапів. Першим етапом цього методу є створення 3D сцени, всі об'єкти якої будуть мати теплові характеристики, розраховані відповідно до фізичних властивостей теплобачення. В результаті буде створено ідеальні синтетичні тепловізійні зображення сцени. Основними вимогами до цього етапу є простота в додаванні та заміні об'єктів в 3D сцені та правдивість їх теплових характеристик. Другий етап полягає в пост-обробці отриманих синтетичних тепловізійних зображень для їх адаптації під реальні тепловізійні зображення. Для цього основною задачею є врахування властивостей тепловізора, його оптичних параметрів та характеристик, щоб максимально наблизити синтетичні зображення до реальних. Загальна схема запропонованого методу зображена на рис. 1.

Спочатку на першому етапі виконується розробка 3D моделі об'єкта дослідження, з використанням спеціального програмного забезпечення для створення тривимірної комп'ютерної графіки. У даній роботі в якості об'єкта використовується БПЛА. Після розробки моделі об'єкту присвоюються теплові характеристики, де необхідно враховувати матеріал окремих елементів об'єкта та їх температуру. Також необхідно надати теплові характеристики всій сцені та всім фоновим об'єктам. Останнім кроком першого етапу є синтез ідеального тепловізійного зображення.

Першим кроком другого етапу є пост-обробка. Спочатку необхідно врахувати усі характеристики тепловізора та створити фільтр, використавши який зображення буде максимально наближене до реального. В результаті пост-обробки отримаємо синтезоване тепловізійне зображення, що максимально наближене до реального.

Створення синтетичних тепловізійних зображень. Як зазначалось раніше, першим етапом є синтез ідеальних тепловізійних зображень. В даному дослідженні для виконання задач на першому етапі в якості програмного пакета для створення 3D графіки використовується Blender версії 4.4.

Blender – це програмний пакет з відкритим кодом для створення якісного 3D-моделювання, анімації та графіки високої роздільної здатності. Дане ПЗ має достатню кількість функціоналу для реалізації поставлених задач. Також Blender має низький поріг входження для початку розробки 3D сцен у ньому [9].

Оскільки об'єктом дослідження в цій роботі є БПЛА, спочатку необхідно розробити 3D моделі необхідних нам дронів. У зв'язку з цим, було розроблено дві моделі дронів: DJI Mavic 3T та DJI Phantom 4 RTK. На рис. 2 зображені розроблені 3D моделі цих дронів.

Після цього необхідно присвоїти теплові характеристики для всіх елементів дронів, враховуючи температуру та матеріал. В Blender є робоче середовище під назвою «Shading» яке можна використати для присвоєння температурних профілів. Для цього треба створити спеціальну схему яка складається з декількох елементів. Розроблена схема зі середовища Shading зображена на рис. 3.

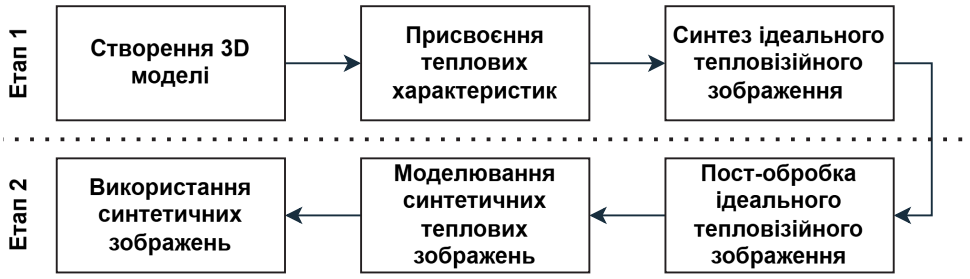


Рис. 1. Схема запропонованого методу моделювання

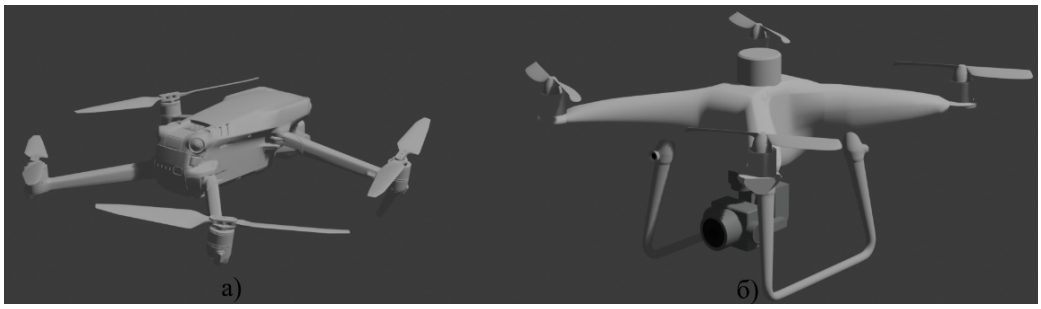


Рис. 2. 3D модель: а) DJI Mavic 3T; б) DJI Phantom 4 RTK

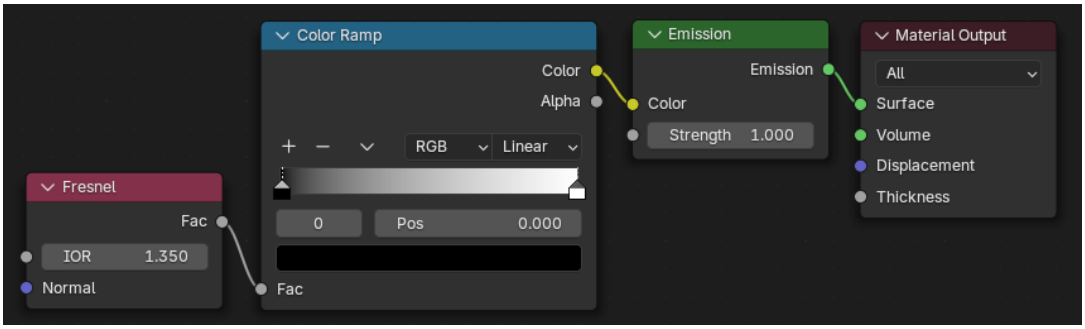


Рис. 3. Схема Shading теплового текстурування елементів об'єкта

Для надання теплового випромінювання використовується властивість «Emission», де необхідно встановити параметр «Emission», який напряму залежить від енергетичної яскравості окремим елементом об'єкта. Синтезовані тепловізійні зображення будуть згенеровані у градаціях сірого, через що встановлюємо градієнт кольору від 0.0 до 1.0 (від чорного до білого), використовуючи параметр «Color Ramp». Аби тепловізійне випромінювання не здавалося занадто різким та залежало від кута огляду, можна додати ефект Френеля, який згладжує випромінювання на краях об'єкта. Для цього використовується параметр «Frensel» з коефіцієнтом 1,35.

Для розрахунку енергетичної яскравості об'єктів можна використати закон Планка для певного спектрального діапазону довжин хвиль [10, с. 2]:

$$B_{\lambda}(\lambda,T)=\varepsilon\cdot\int_{\lambda_1}^{\lambda_2}\frac{2hc^2}{\lambda^5}\frac{1}{e^{hc/\left(\lambda k_B T\right)}-1} \qquad (1)$$

де λ – довжина хвилі, $\lambda_1 - \lambda_2$ – діапазон довжин хвиль (у нашому випадку 8–14 мкм), c – швидкість світла, $h=6{,}626\cdot10^{-34}$ – стала планка, $k_B=1{,}38\cdot10^{-23}$ – стала Больцмана, T – температура тіла в кельвінах, ε – спектральна випромінювальна здатність (тепловіддача) тіла.

Спектральна випромінювальна здатність залежить від діапазону довжин хвиль та матеріалу об'єкта. Проте, вже відомі коефіцієнти для певних довжин хвиль та матеріалів, які були знайдені експериментальним шляхом. Як результат, знаючи температуру тіла та його матеріал, можна розрахувати енергетичну яскравість, яка є еквівалентом Emission Strength.

Для нас важливими є інформація про такі елементи дрона як двигуни, батареї, камери та корпусу. Усі ці елементи нагріваються найбільше. Основні параметри окремих елементів дрона наведені у табл. 1.

Після присвоєння теплових характеристик до окремих деталей дрона можна почати рендер ідеальних синтетичних тепловізійних зображень дронів. Фінальні зображення першого етапу можна побачити на рис. 5б.

Розрахунок МПФ тепловізора. Наступний етап складається з пост-обробки синтетичних тепловізійних зображень для максимального наближення їх до реальних. Для цього, як вже зазначалось раніше, необхідно врахувати оптичні параметри реального тепловізора. У зв'язку з цим, найкраще для цього підходить модуляційна передавальна функція.

Модуляційна передавальна функція (MTF – Modulation Transfer Function) – це характеристика оптичної системи, яка описує, наскільки добре вона передає контраст деталей різних просторових частот з об'єкта на зображення.

В даній роботі буде проводитись визначення МПФ тепловізора, який складається з об'єктива

Таблиця 1

Параметри елементів дрона

Елемент	t◦, ◦C	Матеріал	ε	$B_\lambda(\lambda,T)$	Джерело
Мотори	60 – 80	Анодований алюміній	0,856	94,107	[11, с. 2], [12, с. 10]
Батарея	25 – 100	Пластмасовий корпус	0.93	80,534	[13, с. 1], [14, с. 4]
Корпус	0 – 40	Пластмасовий корпус	0.93	53,342	[15], [14, с. 4]
Гвинти	30 – 40	Карбон	0.95	54,489	[16], [17, с. 10]
Камера	0 – 40	Пластмасовий корпус	0.93	45,706	[14, с. 4]

та мікроболометричної матриці (МБМ). Математичну модель ТСС розглядаємо в області просторових і часових частот та вважаємо, що дана система є лінійною інваріантною. Згідно з цими умовами, МПФ такої ТСС можна знайти добутком МПФ окремих її елементів [8, с. 25]:

$$M_s(\nu) = M_o(\nu)\cdot M_{Ds}(\nu)\cdot M_{Dt}(t) \qquad (2)$$

де, v – просторова частота, $M_o(\nu)$ – МПФ об'єктива, $M_{Ds}(\nu)$ – просторова МПФ МБМ, $M_{Dt}(t)$ – часова МПФ МБМ.

МПФ дифракційно обмеженого об'єктиви з діаметром вхідної зіниці D_{po} можна знайти за формулою [8, с. 25]:

$$M_o(\nu) = \begin{cases} \frac{2}{\pi}\left(\arccos x - x\sqrt{1-x^2}\right), & 0 \le x \le 1 \\ 0, & x > 1 \end{cases} \qquad (3)$$

де $x = \frac{v}{v_c}$ – співвідношення просторової частоти до граничної частоти об'єктива.

У свою чергу, граничну частоту об'єктива можна представити у вигляді диску Ейрі:

$$v_c = 1.22\frac{f_o'\lambda}{D} \qquad (4)$$

Проте функція (3) є доволі складною, але вона наближена до лінійної функції, у зв'язку з чим, для спрощення математичних перетворень, апроксимуємо функцію (3) лінійною функцією та підставимо формулу (4) у співвідношення. Як результат, отримаємо наступну апроксимовану МПФ для об'єктива:

$$M_o(\nu) = \begin{cases} 1 - 1,22\cdot\lambda\frac{\nu}{D_{po}}, & 0 \le \nu \le 0,82\cdot\frac{D_{po}}{\lambda} \\ 0, & x > 0,82\cdot\frac{D_{po}}{\lambda} \end{cases} \qquad (5)$$

Одновимірну просторову МПФ МБМ можна апроксимувати добутком сігнум функцій [18]:

$$M_{Ds}(\nu) = \mathrm{sinc}\left(\frac{W_D}{f_o'}\cdot\nu\right)\mathrm{sinc}\left(\frac{w_D}{f_o'}\cdot\nu\right) \qquad (6)$$

де W_D та w_D – період матричної структури та розмір чутливої площадки пікселя відповідно.

У свою чергу, часова МПФ МБМ є просторовим фільтром низьких частот. Тому M_{Dt} можна представити в такому вигляді:

$$M_{Dt} = \frac{1}{\sqrt{1+\left(2\pi t_D f\right)^2}} \qquad (7)$$

де t_D – постійна часу МБМ, f – часова частота оновлення МБМ.

Для того, щоб показати вплив часової МПФ МБМ, необхідно знайти зв'язок між просторовою частотою ν та часовою частотою f. Зв'язок між ними визначають співвідношенням [8, с. 26]:

$$f = \frac{\beta_D\nu}{t_0} \qquad (8)$$

де β_D – кутовий розмір пікселя матриці, а t_0 – час формування сигналу одним пікселем.

Кутовий розмір пікселя матриці β_D можна знайти співвідношенням розміру одного пікселя МБМ до фокусної відстані об'єктива. Тому, підставивши формулу (8) у часову МПФ МБМ (7), маємо:

$$M_{Dt}(\nu) = \frac{1}{\sqrt{1+\left(2\pi\frac{t_D W_D}{t_0 f_o'}\nu\right)^2}} \qquad (9)$$

Маючи всі МПФ всіх елементів ТСС, ми маємо змогу розрахувати загальну МПФ ТСС як добуток МПФ всіх елементів.

Для прикладу, у даному досліджені було задано такі параметри ТСС, для яких розраховується МПФ: фокусна відстань об'єктива $f_o' = 70$ мм, діаметр вхідної зіниці об'єктива $D_{po} = 70$ мм, розмір пікселя МБМ $W_D = 17$ мкм, розмір чутливої площадки $w_D = 14$ мкм, розмір МБМ $X_D = 6,8$ мм, постійна часу МБМ $t_D = 10$ мс, частота оновлення $f = 50$ Гц.

Використовуючи вище розглянуті функції, ми розрахували МПФ окремих елементів та результуючу МПФ ТСС. Розраховані МПФ можна знайти на рис. 4.

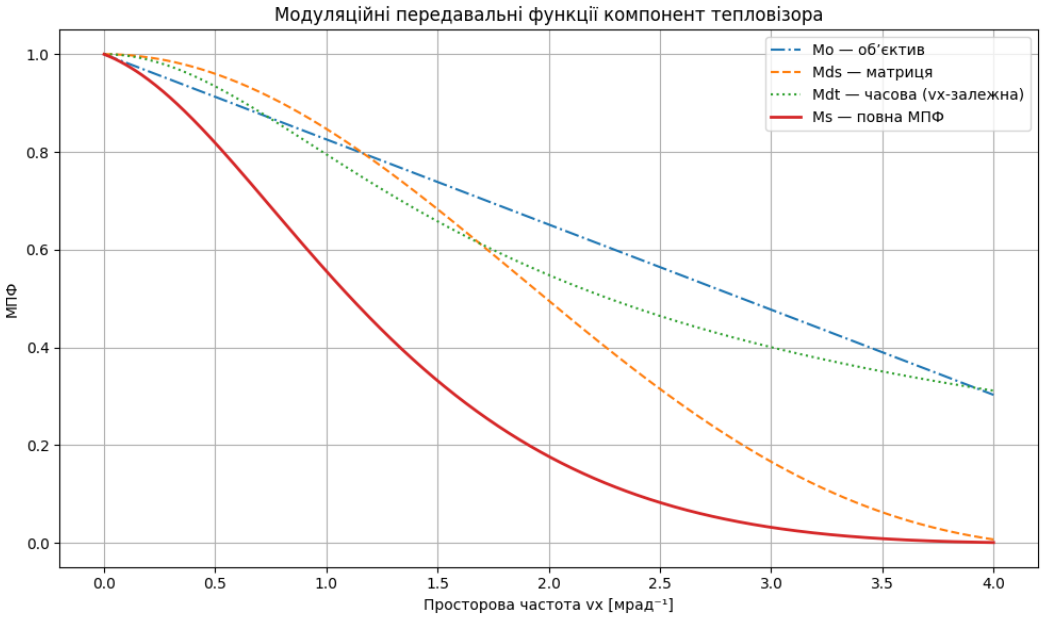


Рис. 4. Модуляційні передавальні функції МПФ

Пост-обробка синтетичних тепловізійних зображень. Перед тим як виконувати постобробку ідеальних тепловізійних зображень, отриманих з Blender, необхідно провести нормалізацію зображення. Тобто вхідне 16-бітне зображення нормалізується до діапазону від 0 до 1:

$$I_{norm} = \frac{I_{raw}}{2^n - 1} \tag{10}$$

де $n = 16$ – розрядність зображення.

Наступним етапом буде застосування МПФ фільтра, використовуючи розраховану МПФ заданої ТСС. Для цього необхідно створити двовимірний МПФ фільтр, використавши розраховану функцію. Вплив тепловізора на зображення можна описати як згортку в просторовій області або як множення на МПФ у частотній області. Тому використання МПФ фільтру на вхідне зображення можна описати наступною формулою:

$$I_{mtf} = \Phi^{-1}\left[\Phi(I_{norm}) \cdot M_s\right] \tag{11}$$

де Ф – оператор прямого перетворення Фур'є, а Φ^{-1} – оператор зворотного перетворення.

Завдяки цьому підходу можна врахувати фізичне згладжування та втрату контрастності зображення через обмеження оптичної системи тепловізора.

Далі можна застосувати радіальну функцію віньєтування $V(x,y)$ для моделювання оптичних втрат по краях зображення. Ця функція зменшує інтенсивність пікселів від центру зображення [19]. У більш сучасних тепловізорах дана абераційна відсутня, тому ця функція повинна застосовуватись при необхідності:

$$I_v(x,y) = I_{norm}(x,y) \cdot V(x,y) \tag{12}$$

Наступним етапом є підсилення теплих об'єктів. Для цього до пікселів з інтенсивністю вище 0,95 застосовується нелінійне підсилення:

$$I_b(x,y) = \begin{cases} \min\left(1; 1{,}5 \cdot I_v(x,y)\right), & I_v > 0{,}95 \\ I_v(x,y) \end{cases} \tag{13}$$

Відразу після підсилення необхідно перевести нормалізовані значення у фізичний температурний діапазон:

$$T(x,y) = T_{\min} + I_b(x,y) \cdot \left(T_{\max} - T_{\min}\right) \tag{14}$$

Також важливим етапом буде модуляція теплового шуму через те, що синтезовані зображення взагалі його не мають. Для цього застосовується Гаусівський шум зі стандартним відхиленням, пропорційним шумово-еквівалентній температурі (NETD) [20]:

$$\begin{aligned} T_n(x,y) &= T(x,y) + f(0,\sigma^2) \\ \sigma &= NETD \cdot \left(T_{\max} - T_{\min}\right) \end{aligned} \tag{15}$$

де $f(0,\sigma^2)$ – це функція випадкових чисел за нормальним розподілом.

Наступним кроком буде імітація розмиття, з використанням Гаусівського згладжування [21]:

$$T_{blur} = G_\sigma \cdot T_n \tag{16}$$

де $\sigma = 1.5$, а $G_\sigma = \begin{bmatrix} 3 & 3 \end{bmatrix}$ – Гаусівське ядро.

Завершальним етапом пост-обробки є квантування, тобто перетворення зображення назад у n-розрядне зображення:

$$Q(x,y)=\left(\frac{T_{blur}(x,y)-T_{\min}}{T_{\max}-T_{\min}}\cdot\left(2^n-1\right)\right) \quad (17)$$

Результати пост-обробки та фінальні зображення розглянутого методу модуляції синтетичних тепловізійних зображень дронів зображені на рис. 5в.

Результати моделювання. Для порівняння та оцінки отриманих синтетичних тепловізійних зображень, використовуючи запропонований в цій роботі метод, було підібрано реальні тепловізійні зображення з дронами Mavic 3T та Phantom 4 RTK. Дані зображення було взято із набору даних Anti-UAV [3] з тих зображень, які не використовувались для навчання нейронної мережі. 3D сцени змодельованих зображень розроблялися максимально наближено до реальних. Одним нюансом реальних тепловізійних зображень є те, що відсутня інформація про тип та характеристики тепловізора, який їх створив. Тому для розрахунку МПФ ТСС використовувались найбільш розповсюджені параметри тепловізора.

На рис. 5 представлені (а) реальні зображення, (б) ідеальні синтезовані зображення, створенні застосунком Blender, та (в) фінальні змодельовані синтетичні тепловізійні зображення після пост-обробки.

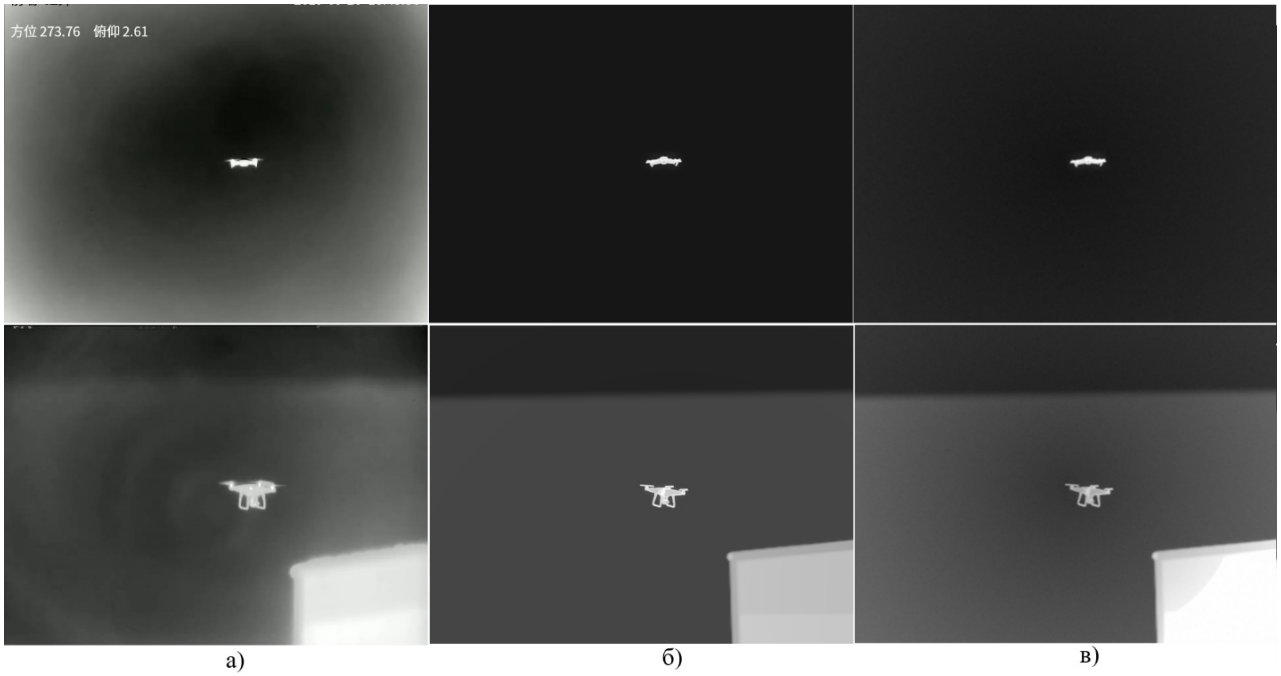
Для порівняння реальних зображень з синтетичними було розраховано локальну контрастність Вебера та середньоквадратичну (RMS) контрастність. Локальна контрастність Вебера використовується для порівняння яскравості дрібних елементів на великому однорідному фоні та розраховується за формулою:

$$C_W=\frac{I_o-I_b}{I_b} \quad (18)$$

Середньоквадратичну (RMS) контрастність розраховується по формулі:

$$C_{RMS}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(I_i-\mu)^2} \quad (19)$$

Також в минулому дослідженні [2] проводились аналіз та порівняння існуючих моделей виявлення об'єктів, в результаті чого було навчено моделі виявлення дронів на тепловізійних зображеннях. Тому для порівняння реальних та синтетичних тепловізійних зображень можна використати ці зображення та порівняти точність, з якою розпізнається дрон. В якості моделі буде використано YOLOv8m яка була навчена лише на тепловізійних зображеннях з датасету Anti-UAV. Результати порівняння наведені у таблиці 2. Також на рис. 6 зображені результати виявлення дронів на реальних і синтетичних зображеннях.



а) б) в)

Рис. 5. Результати моделювання: а) реальне, б) синтезоване в Blender, в) синтезоване після пост-обробки

Провівши візуальне порівняння, можна дійти висновку, що запропонований метод модуляції дозволяє отримувати зображення, які є дуже близькими до справжніх. Завдяки присвоєнню різних теплових характеристик до різних елементів дрона, можна помітити, що такі елементи як двигун, батарея та камера виділяються із загального зображення. Це добре видно на зображеннях Phantom, що є важливо, адже для моделі необхідно мати точну інформацію про окремі шаблони, які зустрічаються на зображені. Також можна помітити як відстань впливає на кількість корисної інформації як на реальному, так і на синтезованому зображеннях. Це помітно на зображеннях з Mavic 3T.

Також завдяки врахуванням фізичних властивостей теплобачення та характеристик тепловізора вдалося максимально наблизити синтезовані зображення до реальних. Це можна помітити, порівнявши контрастність та точність виявлення. Всі параметри доволі близькі один до одного.

Висновки. В даній роботі було запропоновано та реалізовано метод модуляції тепловізійних зображень на основі системного підходу. Для створення 3D сцен використовувався програмний застосунок Blender, в якому було створено 3D об'єкти та присвоєно теплові характеристики до окремих елементів дрона. Враховуючи випромінювальну здатність матеріалу та температуру, було розраховано Emission Strength, використовуючи закон Планка. Після створення ідеальних синтетичних тепловізійних зображень було проведено пост-обробку, враховуючи пара-

Таблиця 2

Результат порівняння

Тип дрону	**Mavic 3T**		**Phantom 4 RTK**	
Тип зображення	Реальне	Синтетичне	Реальне	Синтетичне
C_W	1,992	1,896	0.615	0.667
C_{RMS}	0,165	0,111	0,129	0,112
Точність виявлення	0,77	0,74	0,82	0,84

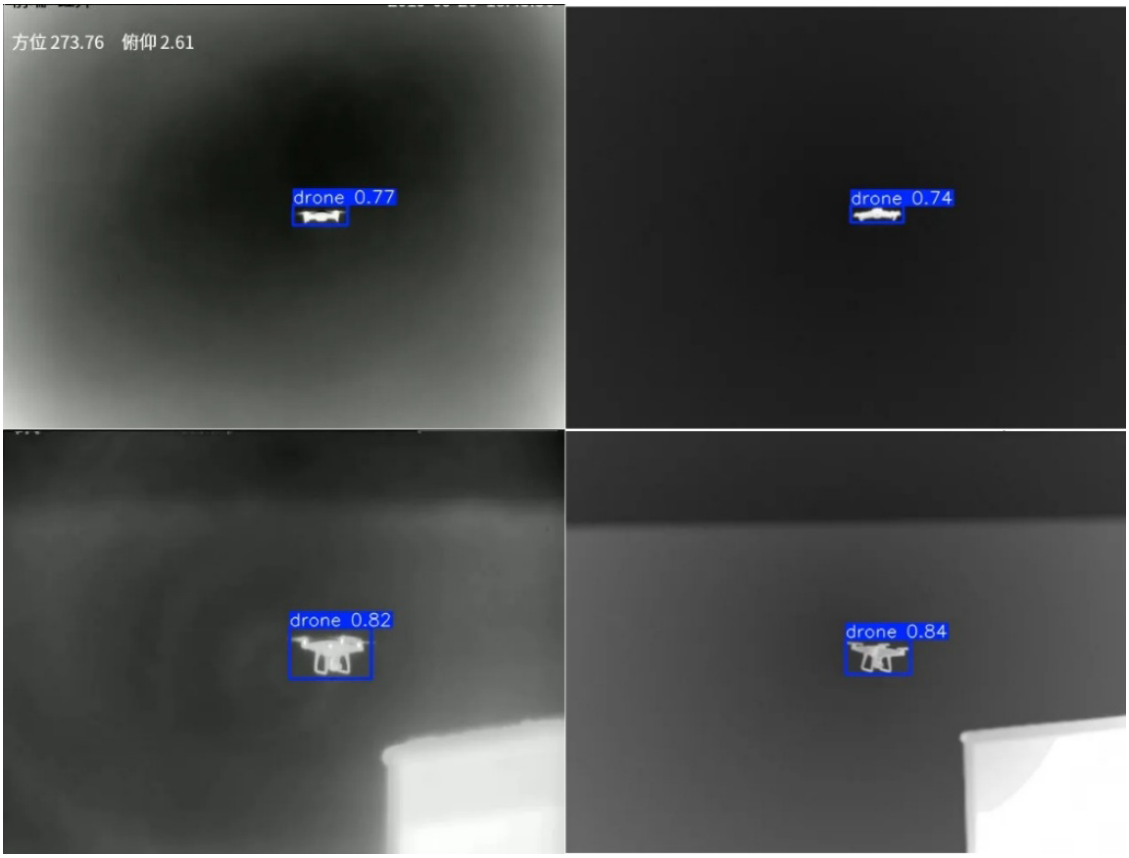


Рис. 6. Результат виявлення

метри тепловізора та втрати контрастності. Для цього була розрахована МПФ тепловізора та використана у вигляді фільтру використовуючи Фур'є перетворення. Також було додано тепловий шум та розмиття до зображення в рамках пост-обробки.

Результати показали, що синтезовані зображення наближені до реальних. Навчена модель на реальних тепловізійних зображеннях розпізнала та видала однакові результати виявлення як для реальних, так і для синтетичних зображень.

Запропонований метод є дієвим, простим в реалізації та експлуатації, тому що для створення синтезованих зображень з новим типом БПЛА, достатньо просто змінити 3D модель.

В подальших дослідженнях можна розробити окремий аддон для Blender, який автоматично буде розраховувати Emission Strength, базуючись на матеріалі та температурі, та задавати відповідні теплові характеристики для об'єктів. Також, використовуючи даний метод, можна збільшити навчальний датасет для підвищення точності моделей виявлення БПЛА. На додаток, можна розрахувати порогову контрастність моделі ШІ для розпізнавання дронів, щоб розрахувати оптимальні параметри тепловізора для ТСС виявлення дронів.

Список літератури:

1. Sommer L., Schumann A. Deep learning-based drone detection in infrared imagery with limited training data. *SPIE, Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies IV*. 2020. Vol. 11542, P. 1154204. DOI: https://doi.org/10.1117/12.2574171

2. Танчук В.С., Колобродов В.Г. Аналіз існуючих моделей виявлення БПЛА на тепловізійних зображеннях. *Вісник КПІ. Серія Приладобудування: збірник наукових праць*. Вип. 2024. Вип. 68. Ч. 2. С. 5–10. DOI: https://doi.org/10.20535/1970.68(2).2024.318080

3. Anti-UAV Challenge Organizers. The 4th anti-uav workshop and challenge. URL: https://anti-uav.github.io (дата звернення: 10.05.2024)

4. Coluccia A., Fascista A., Sommer L., Schumann A., Dimou A., Zarpalas D. The Drone-vs-Bird Detection Grand Challenge at ICASSP 2023: A Review of Methods and Results. *IEEE Open Journal of Signal Processing*. 2024.

5. Madan N., et al. Thermalsynth: A novel approach for generating synthetic thermal human scenarios. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2023. P. 130–139.

6. Wellons M. The Stefan-Boltzmann Law. *Physics Department, The College of Wooster, Wooster, Ohio*. 2007. 44691: 25.

7. Plank M. The theory of heat radiation. *Entropie*. 1900, 144.190: 164.

8. Колобродов В.Г., Лихоліт М., Тягур В., Васильковська І., Колобродов М. Ймовірність розрізнення цілі контрастно обмеженою тепловізійною системою безпілотника. *Космічна наука і технологія*. 2023. Вип. 29, Ч. 5, С. 20–32.

9. Fu J., Xia N., Lu Z., Shen D., Zhang J. Rapid Establishment of 3D Geological Model Based on Blender Software. *2024 International Seminar on Artificial Intelligence, Computer Technology and Control Engineering (ACTCE)*. Wuhan, China. 2024. P. 46–50. DOI: 10.1109/ACTCE65085.2024.00017.

10. Marr Jonathan M, Wilkin Francis P. A better presentation of Planck's radiation law. *American Journal of Physics*. 2012. Vol. 80(5), P. 399–405.

11. Mazur M., Skarka W., Kobielski M., Kądzielawa D., Kubica R., Haas C., Unterberger H. Heat Exchange Analysis of Brushless Direct Current Motors. *Energies*,. 2024. Vol. 17(24), 6469.

12. Gustavsen A., Berdahl P. Spectral emissivity of anodized aluminum and the thermal transmittance of aluminum window frames. *Nordic Journal of Building Physics*. 2003. Vol: 3.1970, P. 1–12.

13. Lalinde I., Berrueta A., Arza J., Sanchis P., Ursúa A. On the characterization of lithium-ion batteries under overtemperature and overcharge conditions: Identification of abuse areas and experimental validation. *Applied Energy*. 2024. Vol: 354, 122205.

14. Goddijn-Murphy L., Williamson. B. On thermal infrared remote sensing of plastic pollution in natural waters. *Remote Sens*. 2019. Vol: 11, 2159.

15. What Temperature Is Too Hot To Fly a Drone. 2022. URL: https://droneflyingpro.com/what-temperature-is-too-hot-to-fly-a-drone (дата звернення 15.04.2025)

16. Mukti S., Megawanto F. C., Arthaa H. Exhaust temperature analysis of UAV propeller materials. *Majalah Ilmiah Pengkajian Industri; Journal of Industrial Research and Innovation*. 2020. Vol: 14 (2), P. 93–98.

17. Adibekyan A., Kononogova E., Monte C., Hollandt J. Review of PTB measurements on emissivity, reflectivity and transmissivity of semitransparent fiber-reinforced plastic composites. *International Journal of Thermophysics*. 2019. Vol: 40, P. 1–14.

18. Kolobrodov V. H. Modulation transfer function of the thermal imaging monocular. *Вісник Національного технічного університету України Київський політехнічний інститут. Серія: Радіотехніка. Радіоапаратобудування*. 2019. Vol: 78, P. 74–78.

19. Zheng Y., Yu J., Kang S.B., Lin S., Kambhamettu C. Single-image vignetting correction using radial gradient symmetry. *In 2008 IEEE Conference on Computer Vision and Pattern Recognition*. 2008. P. 1–8.

20. Kennedy H.V. Modeling noise in thermal imaging systems. *SPIE, Infrared Imaging Systems: Design, Analysis, Modeling, and Testing IV*. 1993. Vol. 1969.

21. Wink A.M., Roerdink, J.B. Denoising functional MR images: a comparison of wavelet denoising and Gaussian smoothing. *IEEE transactions on medical imaging*. 2004. Vol. 23 (3), P. 374–387.

Tanchuk V.S., Kolobrodov V.G. THERMAL IMAGE MODELING METHOD BASED ON A SYSTEMATIC APPROACH

This paper proposes a method for modeling thermal images of objects based on a systematic approach to the formation of a thermal imaging infrared signal. The main goal of this research is to develop a methodology for modeling synthetic thermal images, taking into account the physical and technical characteristics of a real thermal imager. The corresponding methodology is especially relevant for computer vision tasks, where it is critical to have a high-quality and reliable dataset, which affects the effectiveness of training artificial intelligence models for automated target detection and recognition, and facilitates the development of algorithms for automatic surveillance systems. Additionally, the use of thermal imaging modeling allows you to adapt and prepare the system for changes that may occur during operation. The test surveillance object for which modulation is performed in this work is an unmanned aerial vehicle (UAV).

The method consists of two main stages. The first stage is the rendering of a thermal image of the scene, taking into account Planck's law and the emissivity of the object's materials, using a 3D modeling software package, which results in a pseudo thermal image with the corresponding observation object. The second stage is digital post-processing, which simulates the influence of the optical system (OS) of the thermal imager on the image quality, taking into account the parameters of the thermal imager and its features. During post-processing, the influence of the lens and microbolometric matrix (MBM) on the OS of the thermal imager is calculated, which is described through the modulation transfer function (MTF). To calculate the MTF, real parameters of the thermal imager are considered, such as focal length, entrance pupil size, pixel size, radiation wavelength, and frame rate. As a result, we obtain a two-dimensional frequency filter, which is applied to the ideal image in the frequency domain. After that, several more post-processing functions are used.

As a result, thermal images were simulated, taking into account the physical properties of thermal vision and the characteristics of a real thermal imager, and their comparison with real thermal images was carried out.

Key words: *modeling, thermal imager, post-processing, unmanned aerial vehicle, drone, Planck's law, modulation transfer function, artificial intelligence.*

УДК 004.8
DOI https://doi.org/10.32782/2663-5941/2025.3.2/52

Терейковський І.А.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Шевченко О.В.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

АНАЛІЗ НЕЙРОМЕРЕЖЕВИХ РІШЕНЬ УПРАВЛІННЯ БЕЗПІЛОТНИМ ПРИСТРОЄМ

У статті здійснено аналіз нейромережевих рішень для управління безпілотними пристроями з акцентом на сучасні підходи та перспективи використання штучного інтелекту у сфері інформатики, обчислювальної техніки та автоматизації. Визначено актуальність дослідження, яка зумовлена необхідністю ефективного та надійного управління безпілотними апаратами в умовах швидкої зміни навколишнього середовища, обмеженої інформації та високих вимог до точності й автономності системи. Аналізуються ключові нейромережеві архітектури та моделі, такі як згорткові нейромережі (CNN), рекурентні нейромережі (RNN), глибоке підкріплене навчання (Deep Reinforcement Learning), та гібридні моделі, які дозволяють реалізовувати складні сценарії поведінки безпілотних систем. Автором проведено огляд сучасних досліджень і публікацій, де розглядаються різні підходи до побудови нейромережевих систем управління, їх переваги та недоліки, що забезпечує формування всебічного розуміння поточного стану наукових досягнень у цій галузі. Встановлено, що серед головних невирішених проблем є питання стабільності роботи нейромереж у складних умовах, недостатня пояснюваність моделей та труднощі забезпечення повної автономності у реальних застосуваннях. Мета статті полягає у здійсненні комплексного аналізу нейромережевих рішень для управління безпілотними пристроями, визначенні сильних та слабких сторін існуючих моделей, а також виявленні потенційних напрямків удосконалення та інноваційного розвитку нейромережевого управління. Основний матеріал статті охоплює теоретичні основи функціонування нейромережевих моделей, детальний опис архітектур та алгоритмів навчання, а також результати експериментальних досліджень з використанням моделювання роботи безпілотних пристроїв у різних сценаріях. Проведений аналіз дозволяє обґрунтувати вибір оптимальних підходів до створення ефективних нейромережевих систем управління. У висновках статті автором узагальнено отримані результати, підкреслено актуальність та перспективність використання нейромережевих технологій у сфері управління безпілотними пристроями. Визначено основні напрямки майбутніх досліджень, які включають покращення стабільності й надійності моделей, підвищення їх прозорості та пояснюваності, а також подальше вдосконалення методів глибокого підкріпленого навчання для розширення меж автономності безпілотних апаратів.

Ключові слова: *нейромережі, управління безпілотними пристроями, штучний інтелект, автономність, глибоке підкріплене навчання, інформатика, автоматизація.*

Постановка проблеми. Управління безпілотними пристроями є одним з найбільш динамічно розвинених напрямків сучасної інформатики та автоматизації. Впровадження таких систем у різних галузях економіки, промисловості, аграрному секторі, транспортній інфраструктурі, а також у військовій та цивільній сферах демонструє стрімкий прогрес і значні переваги. Зокрема, безпілотні пристрої здатні виконувати широкий спектр завдань, які для людини є небезпечними, важкодоступними або ресурсомісткими. Важливість теми підтверджується результатами досліджень [1, с. 352; 3, с. 1]. Однак їх ефективне та безпечне функціонування залишається значним науково-практичним викликом. Основною проблемою управління безпілотними пристроями є забезпечення їх надійної та автономної роботи у реальних умовах, які характеризуються значною невизначеністю, нестабільністю та динамічністю змін зовнішнього середовища [2, с. 2]. Традиційні алгоритми управління, що базуються на жорстко заданих правилах і моделях, часто виявляються недостатньо адаптивними для ефективної роботи в складних сце-

наріях з великою кількістю непередбачуваних факторів. Водночас, розвиток штучного інтелекту, зокрема нейромережевих технологій, відкриває нові перспективи для вирішення цієї проблеми. Завдяки своїй здатності до навчання та узагальнення, нейромережеві моделі здатні адаптуватися до нових умов, розпізнавати шаблони, ухвалювати рішення у складних ситуаціях, забезпечуючи високий рівень автономності. Однак, незважаючи на активні дослідження та численні успіхи, нейромережеві підходи також мають низку невирішених завдань, серед яких стабільність їх роботи за умов обмежених ресурсів та непередбачуваних ситуацій, недостатня прозорість ухвалюваних рішень і складність забезпечення повної автономності без втручання людини. Таким чином, виникає нагальна необхідність комплексного аналізу нейромережевих рішень з метою виявлення оптимальних підходів для побудови ефективних і надійних систем управління безпілотними пристроями. Це зумовлює актуальність даного дослідження, результати якого можуть суттєво вплинути на подальший розвиток сфери автономних систем і розширити межі їх практичного застосування.

Аналіз останніх досліджень і публікацій. В останні рокипроблематика управління безпілотними пристроями активно досліджується в численних наукових працях. Особливу увагу вчені приділяють використанню нейромережевих технологій, оскільки саме вони демонструють потенціал ефективного вирішення завдань, пов›язаних з адаптивністю та автономністю систем управління. Дослідники в роботах часто розглядають можливості використання глибоких нейромереж, зокрема згорткових та рекурентних нейромереж, для розпізнавання та аналізу навколишнього середовища, ухвалення оперативних рішень у складних ситуаціях [7, с. 1]. Також активно вивчається метод глибокого підкріпленого навчання, який дозволяє системі самостійно адаптувати свої стратегії управління залежно від зміни умов навколишнього середовища. Проте, незважаючи на значний прогрес у цій сфері, дослідження демонструють наявність ряду невирішених питань. Зокрема, багато публікацій підкреслюють проблему недостатньої стабільності роботи нейромереж за умов неповної або шумної інформації, складність забезпечення прозорості рішень та їх пояснюваності. Окремі автори наголошують на тому, що для практичного впровадження нейромережевих систем управління важливим є вирішення питань ефективності їх роботи в реальних умовах, де суттєву роль відіграють обмеженість ресурсів та необхідність мінімізації людського втручання.

Постановка завдання. Метою статті є здійснення всебічного аналізу існуючих нейромережевих рішень управління безпілотними пристроями, оцінювання їх ефективності, переваг та недоліків, а також визначення шляхів і підходів для подальшого удосконалення таких систем. Завдання дослідження полягає в ідентифікації основних факторів, які впливають на надійність, адаптивність і автономність нейромережевих систем управління, та формуванні рекомендацій щодо напрямків подальших досліджень і практичного впровадження.

Виклад основного матеріалу. Управління безпілотними пристроями з використанням нейромережевих технологій передбачає застосування складних архітектур і алгоритмів машинного навчання, що дозволяють реалізувати адаптивні, надійні та автономні системи. Для глибокого розуміння переваг і недоліків таких підходів необхідно розглянути теоретичні основи, архітектури та методи навчання нейромереж, що застосовуються у сфері управління безпілотними пристроями. Згорткові нейромережі (CNN) [5, с. 331] є одним з найпоширеніших інструментів у завданнях візуального сприйняття та аналізу навколишнього середовища безпілотних систем. На рисунку 1 зображено структуру згорткової нейромережі, що використовується для обробки візуальної інформації безпілотними системами.

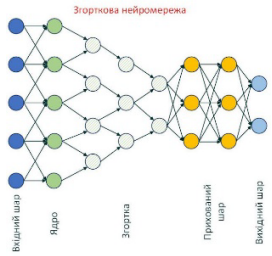


Рис. 1. Структурна схема згорткової нейромережі з виділеними фазами: вхідний шар, ядро, згортка, приховані та вихідні шари

Завдяки своїй здатності автоматично витягувати ознаки зображень, згорткові мережі активно використовуються для навігації пристроїв у складних умовах, розпізнавання перешкод, оцінки дистанції до об›єктів та інших важливих завдань. Перевагою CNN є їх висока точність розпізнавання, що підтверджується численними

експериментами, проте недоліками залишаються висока обчислювальна складність і значні витрати на навчання мереж. Рекурентні нейромережі (RNN) [5, с. 375], зокрема їх різновид – довга короткострокова пам›ять (LSTM) [5, с. 408], ефективно застосовуються для аналізу послідовностей даних, таких як часові ряди з сенсорів та телеметрична інформація. На рисунку 2 зображено архітектуру рекурентної нейронної мережі, що може бути використана для обробки візуальної інформації безпілотними системами.

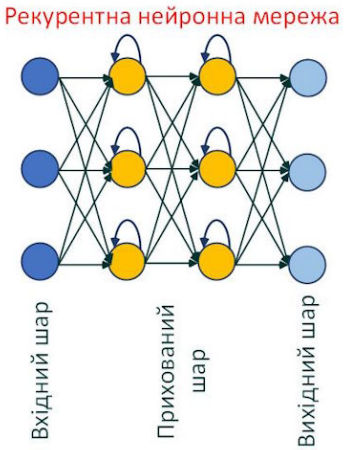


Рис. 2. Архітектура рекурентної нейронної мережі (RNN) з циклічними зв'язками у прихованому шарі, що дозволяє моделі зберігати контекст у часових послідовностях

Особливо цінними ці мережі є для передбачення поведінки безпілотних пристроїв у часі, прогнозування траєкторій руху та адаптації до динамічно змінюваних умов. Однак рекурентні нейромережі можуть мати проблеми з обробкою довгих послідовностей через поступову втрату інформації, що створює додаткові виклики в їх використанні. Глибоке підкріплене навчання (Deep Reinforcement Learning, DRL) [4, с. 142] набуло широкого поширення для створення систем управління безпілотними пристроями завдяки своїй здатності навчати агента ухвалювати рішення в умовах невизначеності та обмежених ресурсів. Основною ідеєю методу є навчання нейромережевого агента шляхом взаємодії з навколишнім середовищем та отримання зворотного зв›язку у вигляді винагород чи штрафів. DRL успішно використовується для вирішення завдань автономної навігації, обходу перешкод та виконання специфічних місій. Проте, важливо зазначити, що DRL характеризується складністю стабілізації процесу навчання та високими обчислювальними витратами. Гібридні нейромережеві моделі поєднують переваги різних архітектур і використовуються для створення комплексних систем управління. Такі моделі здатні ефективніше вирішувати завдання управління безпілотними пристроями, поєднуючи переваги CNN для аналізу навколишнього середовища та RNN для прогнозування поведінки системи.

Одним із найактуальніших завдань у застосуванні нейромережевих технологій для управління безпілотними пристроями є забезпечення їх ефективного функціонування в умовах обмежених ресурсів, таких як обмежена обчислювальна потужність, оперативна пам›ять і енергоспоживання [6, с. 5]. Особливо це стосується малих безпілотних пристроїв, таких як дрони, роботи або автономні датчики, які не мають можливості використовувати потужні процесори або графічні прискорювачі. Для вирішення цієї задачі активно застосовуються методи оптимізації нейромереж, які дозволяють зберегти високий рівень продуктивності й точності моделей при мінімальних ресурсних витратах [9, с. 1]. Серед найпоширеніших підходів варто відзначити такі методи, як квантування параметрів нейромережі [8, с. 3], обрізання (pruning) [10, с. 5], дистиляція знань [11, с. 2] та застосування архітектур зниженої складності. Квантування нейромережевих моделей передбачає зменшення точності представлення чисел з плаваючою точкою до чисел з фіксованою точкою або навіть до двійкових представлень. Це дозволяє суттєво скоротити вимоги до обчислювальних ресурсів і пам›яті, що особливо важливо для малопотужних пристроїв. Процес квантування, показаний на рисунку 3, зменшує обсяг моделі та дозволяє виконувати обчислення з меншими витратами енергії.

Експерименти з квантуванням нейромереж для безпілотних систем показали, що хоча точність моделей дещо знижується, вона залишається прийнятною для більшості завдань управління й навігації. При цьому швидкість виконання обчислень та автономність пристроїв істотно зростають за рахунок зменшення споживання енергії. Обрізання нейромереж (pruning) полягає у видаленні менш важливих зв›язків і нейронів, які мають незначний вплив на результати роботи моделі. В результаті такої оптимізації можна суттєво зменшити кількість обчислювальних операцій, що необхідні для роботи нейромережі, при цьому точність роботи моделі залишається на високому рівні. На рисунку 4 схематично подано алгоритм обрізання нейронної мережі з подальшим донавчанням.

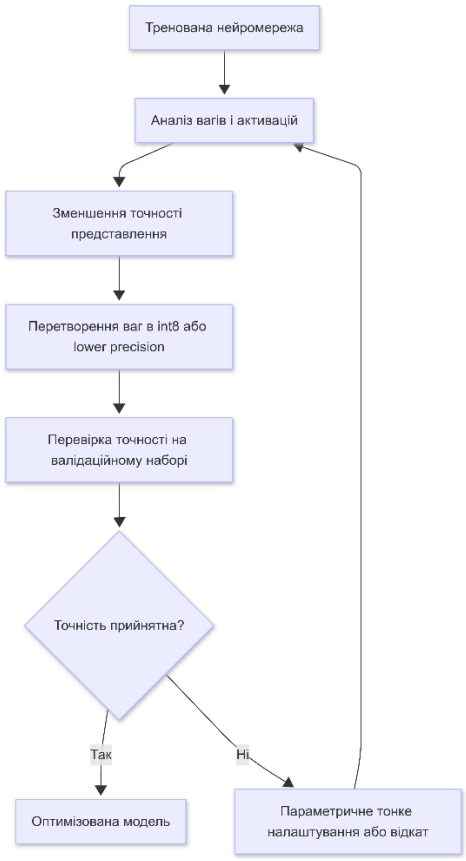


Рис. 3. Автоматизований процес квантування моделі з перевіркою збереження точності та циклічним донавчанням при потребі

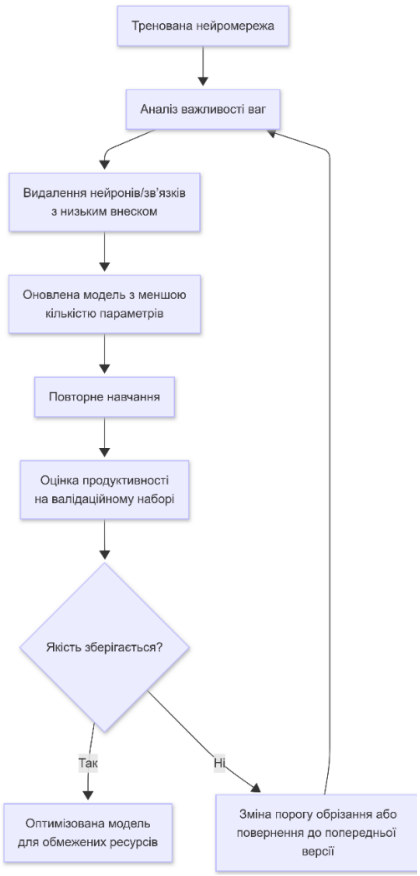


Рис. 4. Алгоритм зменшення моделі за допомогою аналізу вагів, редукції архітектури та повторного навчання

Дослідження свідчать, що нейромережі після обрізання можуть демонструвати навіть вищу продуктивність у режимі реального часу, порівняно з оригінальними моделями, завдяки зменшенню часу обчислень та підвищенню енергоефективності пристроїв. Обрізання знижує обчислювальні витрати, водночас зберігаючи продуктивність (рис. 4). Метод дистиляції знань полягає у навчанні компактної нейромережі («студента») за допомогою великої і складної мережі («вчителя»). На рисунку 5 представлено типову архітектуру передачі знань від великої моделі («вчителя») до малої («студента»).

Менша мережа переймає найбільш важливі патерни поведінки складної моделі, таким чином зберігаючи високу точність і функціональність, але суттєво зменшуючи обчислювальні витрати. Процес дистиляції знань забезпечує ефективну адаптацію моделей до пристроїв з обмеженими ресурсами (рис. 5). Практичні експерименти з використанням цього методу на безпілотних апаратах засвідчили, що «студентські» нейромережі демонструють високу ефективність та швидкість роботи навіть за умов суттєвої обме-

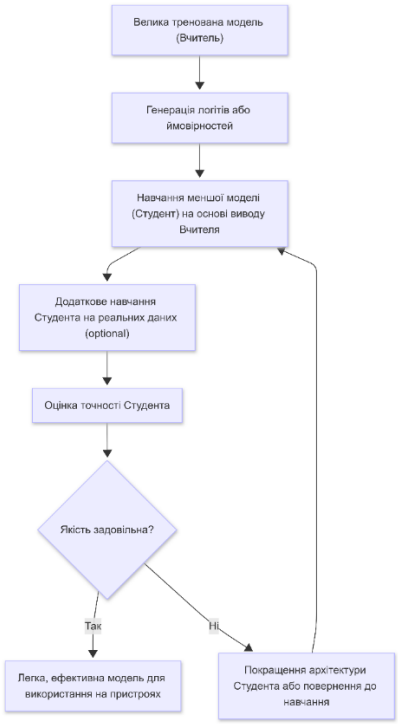


Рис. 5. Схема дистиляції знань: навчання компактної моделі («студента») за допомогою виводу потужної моделі («вчителя»)

женості апаратних ресурсів. Ще одним важливим підходом є розроблення спеціалізованих архітектур нейромереж, таких як MobileNet, ShuffleNet та SqueezeNet, які спочатку створюються з орієнтацією на пристрої з низькою продуктивністю. Ці моделі містять спеціальні блоки та операції, що значно зменшують складність нейромережі без суттєвої втрати її точності. Досвід застосування таких нейромереж у безпілотних системах показує, що вони забезпечують прийнятний рівень автономності та ефективно виконують завдання управління в реальному часі.

Аналіз впливу оптимізації на точність, швидкість та автономність безпілотних систем свідчить, що оптимізовані нейромережі демонструють хорошу компромісну продуктивність. Як видно з рисунка 6, оптимізація нейромережі значно знижує споживання ресурсів, зокрема пам'яті та енергії. Зменшення розміру моделі та її ресурсних вимог дозволяє значно підвищити швидкість обробки даних і реакції системи, що критично важливо в умовах реального часу. Як видно з рисунка 7, після оптимізації зросли стабільність та швидкість, а точність знизилася незначно. На рисунку 8 зображено динаміку навчання моделей до та після оптимізації: оптимізовані моделі демонструють повільніше зростання точності. Хоча точність роботи моделей може незначно знижуватись, на практиці ця втрата часто компенсується швидкістю та енергоефективністю, що відкриває ширші можливості для застосування нейромережевих рішень у безпілотних пристроях з обмеженими ресурсами. Таким чином, використання оптимізованих нейромережевих моделей є перспективним напрямком розвитку управління автономними системами, особливо в умовах жорстких апаратних обмежень, і заслуговує на подальше активне дослідження та практичне застосування.

Рис. 6. Порівняння споживання енергії, оперативної пам'яті та часу обчислень нейромережею до і після застосування оптимізаційних методів

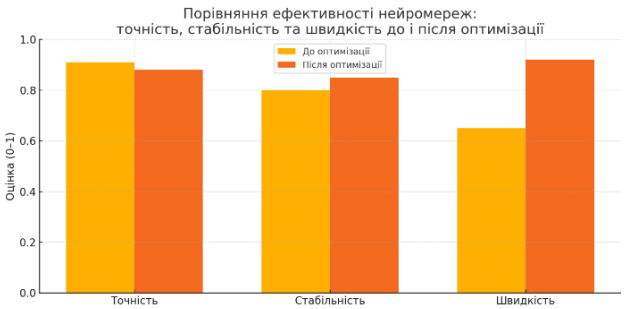


Рис. 7. Вплив оптимізації на основні характеристики нейромереж: зростання швидкості обчислень і стабільності з мінімальною втратою точності

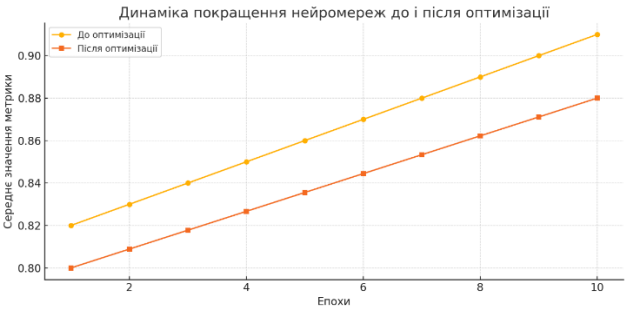


Рис. 8. Порівняння результатів навчання нейромережевої моделі на базовій та оптимізованій архітектурах

Під час проведення експериментальних досліджень нейромережевих моделей було виявлено, що одним із критичних факторів ефективності є якість даних, які використовуються для навчання. Недостатня якість або неадекватна репрезентативність даних суттєво знижує продуктивність нейромережевих систем управління. Це підкреслює важливість ретельного збору та підготовки навчальних наборів, що включають різноманітні сценарії поведінки та умови експлуатації. Ще однією важливою проблемою, яку виявили під час досліджень, є пояснюваність нейромережевих рішень. Оскільки нейромережі часто ухвалюють рішення за допомогою «чорної скриньки», для практичного застосування критично важливим є розроблення методів, які дозволяють пояснювати та інтерпретувати отримані результати. Застосування методів пояснюваності може значно підвищити довіру до нейромережевих систем та сприяти їх широкому впровадженню у реальні умови. Проведені дослідження також засвідчили, що існуючі нейромережеві моделі часто стикаються з труднощами у забезпеченні стабільності роботи в складних умовах з високим рівнем шуму або недостатньою інформацією. Вирішення цієї проблеми можливе за допомогою розроблення нових алго-

ритмів навчання, які здатні ефективніше адаптуватися до реальних умов експлуатації. На основі аналізу результатів досліджень можна стверджувати, що подальший розвиток нейромережевих систем управління безпілотними пристроями має спрямовуватись на підвищення їх стабільності, адаптивності та автономності. Зокрема, перспективними напрямками є вдосконалення алгоритмів глибокого підкріпленого навчання, інтеграція пояснюваних моделей, а також створення гібридних систем, які поєднують переваги різних архітектур нейромереж. Таким чином, нейромережеві технології мають значний потенціал для вирішення актуальних проблем управління безпілотними пристроями та відкривають широкі можливості для подальших досліджень і практичних застосувань у різних галузях.

Висновки. Проведений аналіз нейромережевих рішень управління безпілотними пристроями свідчить про значний потенціал таких технологій у вирішенні складних завдань автоматизації та автономності систем. Нейромережеві моделі демонструють високу здатність до адаптації в умовах змінного і непередбачуваного середовища, а також мають кращі показники точності, стабільності й оперативності, ніж традиційні алгоритмічні підходи.

Однак, попри очевидні переваги, нейромережеві системи все ще мають суттєві недоліки, серед яких найбільш актуальними залишаються низька пояснюваність рішень, складність забезпечення стабільності в умовах обмежених ресурсів та недостатня відмовостійкість. Для подолання цих проблем перспективними є використання ансамблевого навчання, резервування компонентів нейромереж, а також розроблення спеціалізованих алгоритмів навчання, які враховують можливі збої та обмеження реальних умов експлуатації.

Подальші дослідження в цьому напрямку мають бути спрямовані на розроблення більш прозорих і зрозумілих моделей, підвищення надійності й автономності систем та оптимізацію нейромереж для їх ефективного використання на пристроях з обмеженими ресурсами. Реалізація цих завдань дозволить значно розширити застосування нейромережевих технологій у сфері управління безпілотними пристроями та підвищити їхню конкурентоспроможність у практичних застосуваннях.що ...

Список літератури:

1. Інноваційні методи протидії повітряній розвідці рф: розробка та впровадження українських технологій боротьби з БпЛА / І. Мельніков та ін. InterConf. 2024. № 48 (213). С. 350–364. URL: https://doi.org/10.51582/interconf.19-20.08.2024.030 (дата звернення: 02.09.2024).

2. Семенов С. Г., Волошин Д. Г. Перспективи розвитку системи управління безпілотними літальними апаратами : thesis. 2018. URL: http://repository.kpi.kharkov.ua/handle/KhPI-Press/39918 (дата звернення: 23.12.2024).

3. Ященко К. А. Спеціалізовані програмні засоби БПЛА : thesis. 2020. URL: https://er.nau.edu.ua/handle/NAU/51040 (дата звернення: 17.11.2024).

4. Awad M., Khanna R. Deep neural networks. Efficient learning machines. Berkeley, CA, 2015. P. 127–147. URL: https://doi.org/10.1007/978-1-4302-5990-9_7 (date of access: 24.11.2024).

5. Bengio Y., Courville A., Goodfellow I. Deep learning. MIT Press, 2016. 800 p.

6. Efficient densely connected convolutional neural networks / G. Li et al. Pattern recognition. 2021. Vol. 109. P. 107610. URL: https://doi.org/10.1016/j.patcog.2020.107610 (date of access: 02.12.2024).

7. Efficient processing of deep neural networks: a tutorial and survey / V. Sze et al. Proceedings of the IEEE. 2017. Vol. 105, No. 12. P. 2295–2329. URL: https://doi.org/10.1109/jproc.2017.2761740 (date of access: 16.12.2024).

8. Keyboard dynamic analysis by Alexnet type neural network / S. Toliupa et al. 2020 IEEE 15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET), Lviv-Slavske, Ukraine, 25–29 February 2020. 2020. URL: https://doi.org/10.1109/tcset49122.2020.235466 (date of access: 07.01.2025).

9. Krizhevsky A., Sutskever I., Hinton G. E. ImageNet classification with deep convolutional neural networks. Communications of the ACM. 2017. Vol. 60, No. 6. P. 84–90. URL: https://doi.org/10.1145/3065386 (date of access: 07.04.2025).

10. The method of semantic image segmentation using neural networks / I. Tereikovskyi et al. International journal of image, graphics and signal processing. 2022. Vol. 14, No. 6. P. 1–14. URL: https://doi.org/10.5815/ijigsp.2022.06.01 (date of access: 08.02.2025).

11. Understanding deep learning (still) requires rethinking generalization / C. Zhang et al. Communications of the ACM. 2021. Vol. 64, No. 3. P. 107–115. URL: https://doi.org/10.1145/3446776 (date of access: 07.04.2025).

Tereikovskiy I.A., Shevchenko O.V. **ANALYSIS OF NEURAL NETWORK SOLUTIONS FOR UNMANNED VEHICLE CONTROL**

The article analyzes neural network solutions for controlling unmanned devices with an emphasis on modern approaches and prospects for the use of artificial intelligence in the field of informatics, computing, and automation. The relevance of the study is determined, which is due to the need for effective and reliable control of unmanned devices in conditions of rapid environmental change, limited information, and high requirements for system accuracy and autonomy. Key neural network architectures and models are analyzed, such as convolutional neural networks (CNN), recurrent neural networks (RNN), deep reinforcement learning (Deep Reinforcement Learning), and hybrid models that allow implementing complex scenarios of behavior of unmanned systems. The author reviews modern research and publications, which consider various approaches to building neural network control systems, their advantages and disadvantages, which provides a comprehensive understanding of the current state of scientific achievements in this field. It was established that among the main unresolved problems are the issues of the stability of neural networks in complex conditions, insufficient explainability of models and difficulties in ensuring full autonomy in real applications. The purpose of the article is to carry out a comprehensive analysis of neural network solutions for controlling unmanned devices, determine the strengths and weaknesses of existing models, as well as identify potential areas for improvement and innovative development of neural network control. The main material of the article covers the theoretical foundations of the functioning of neural network models, a detailed description of architectures and learning algorithms, as well as the results of experimental studies using modeling of the operation of unmanned devices in various scenarios. The analysis allows us to justify the choice of optimal approaches to creating effective neural network control systems. In the conclusions of the article, the author summarizes the results obtained, emphasizes the relevance and prospects of using neural network technologies in the field of controlling unmanned devices. The main directions of future research are identified, which include improving the stability and reliability of models, increasing their transparency and explainability, as well as further improving deep reinforcement learning methods to expand the limits of autonomy of unmanned vehicles.

Key words: *neural networks, control of unmanned devices, artificial intelligence, autonomy, deep reinforcement learning, computer science, automation.*

УДК 681.5:621.867
DOI https://doi.org/10.32782/2663-5941/2025.3.2/53

Тиханська А.М.
Криворізький національний університет

Тиханський М.П.
Криворізький національний університет

КЕРУВАННЯ ВАЖКИМ СТРІЧКОВИМ КОНВЕЄРОМ З УРАХУВАННЯМ ТЕХНІЧНОГО СТАНУ ЙОГО ОБЛАДНАННЯ

У статті представлено комплексний підхід до керування важким стрічковим конвеєром з урахуванням технічного стану його основних вузлів та механізмів. Актуальність дослідження зумовлена потребою підвищення надійності та ефективності роботи конвеєрних систем на гірничо-видобувних підприємствах, де нерівномірне зношування стрічки, роликів і барабанів призводить до аварійних простоїв та збільшує витрати на ремонт. Для запобігання передчасним відмовам обладнання та забезпечення безпечної експлуатації необхідне керування роботою конвеєра, що адаптується до поточного технічного стану його компонентів. У дослідженні проведено аналіз сучасних рішень і наукових праць у галузі конвеєрних технологій, а також теоретичне моделювання динамічних навантажень і зношування елементів конвеєра під час пуску та роботи під навантаженням. У рамках запропонованого підходу використано частотно-регульований електропривод для реалізації плавного запуску конвеєра та розроблено алгоритми моніторингу технічного стану, інтегровані в систему керування. Отримані результати демонструють, що впровадження інтелектуальної системи керування дозволяє істотно знизити динамічні навантаження на стрічку і механізми. Розроблено алгоритм автоматизованого плавного пуску конвеєра, який регулює прискорення та тягове зусилля з урахуванням маси вантажу на стрічці і ступеня її зносу. Це дозволяє уникнути різких ривків та пробуксовок під час запуску, зменшуючи пікові навантаження, що раніше вдвічі перевищували усталений режим при традиційних методах пуску. Також запропоновано алгоритм регулювання швидкості руху стрічки та її натягу в процесі роботи конвеєра на основі безперервного моніторингу стану обладнання. Система контролю визначає, чи відбувається прослизання стрічки по барабану, відстежує перегрів підшипників і появу пошкоджень полотна стрічки та автоматично знижує швидкість або навантаження, чи навіть зупиняє конвеєр при наближенні параметрів до критичних. Новизна отриманих результатів полягає в інтегрованому підході до керування: вперше запропоновано поєднати алгоритми плавного запуску і робочого руху конвеєра з безперервним відстеженням технічного стану стрічки та інших вузлів. Результати дослідження можуть бути впроваджені на підприємствах гірничої промисловості та інших галузях, де використовуються важкі стрічкові конвеєри, з метою підвищення їхньої ефективності та безпечності експлуатації.

Ключові слова: *стрічковий конвеєр, технічний стан, автоматизоване керування, пусковий режим, інтелектуальні системи, зношування стрічки, натяг стрічки, прослизання, частотно-регульований привід, моніторинг обладнання.*

Постановка проблеми. Принципи керування стрічковим конвеєром на основі технічного стану його електромеханічного обладнання визначаються з урахуванням поточного стану ключових вузлів і механізмів. Оскільки окремі елементи конструкції мають похибки під час виробництва і монтажу, в процесі експлуатації вони зношуються нерівномірно. Це може призвести до аварійних відмов або скорочення міжремонтних інтервалів.

Відомо, що коригування режимів роботи конвеєра дає змогу знизити рівень пошкоджень його елементів. Регулювання таких параметрів, як швидкість транспортування, ступінь завантаження, натяг стрічки і тягове зусилля під час запуску навантаженого конвеєра, сприяє зменшенню навантажень, знижує знос стрічки і роликів. Цілеспрямована оптимізація режимних характеристик допомагає запобігти критичним перевантаженням.

У разі поздовжнього або поперечного розриву стрічки, перегріву підшипникових вузлів двигуна або редуктора, а також часткового руйнування футерування футерування барабана зниження швидкості роботи конвеєра дасть змогу продовжити експлуатацію до завершення зміни та під

час профілактичного ремонту ухвалити економічно обґрунтоване рішення. Тому розробка принципів керування важким стрічковим конвеєром з урахуванням технічного стану його обладнання є актуальним завданням.

Аналіз останніх досліджень і публікацій. Проблема ефективного управління стрічковими конвеєрами із врахуванням технічного стану обладнання набуває особливої актуальності останніми роками, з огляду на необхідність підвищення надійності, зниження експлуатаційних витрат та покращення умов роботи промислових підприємств. У сучасних наукових дослідженнях особливий акцент робиться на розробці методик моніторингу та діагностики технічного стану ключових вузлів конвеєрних систем.

Деякі науковці зосередились на керуванні швидкістю конвеєра. У роботі [1] запропоновані способи плавного запуску і зупинки за допомогою частотно-регульованих приводів, що дозволяє зменшити динамічні навантаження і знос стрічки. Схожі підходи розглядаються й у вітчизняних роботах [2], де аналізується, як зміна швидкості впливає на технічний стан і надійність системи.

Інша частина досліджень присвячена прогнозному керуванню. Наприклад, у [3] для системи «конвеєр–дробарка» застосовується метод прогнозуючого керування (Model Predictive Control), який дозволяє уникати перевантаження. А в [4] описано використання алгоритмів машинного навчання (random forest) для технічного обслуговування приводу і редуктора, що дає змогу виявляти критичні зміни заздалегідь і запобігати аваріям.

Для оцінки зношення стрічки у [5] використовується мультиспектральне зображення в поєднанні з глибоким навчанням – така система працює навіть у складних умовах. Щодо роликів, у [6] продемонстровано ефективність алгоритмів градієнтного бустингу для виявлення пошкоджень оболонок та підшипників, а в [7] запропоновано метод на основі аналізу вібрацій, який адаптується до конкретних умов роботи.

Окремі дослідження зосереджені на більш комплексних підходах. У [8] розглядається застосування інтелектуальних систем контролю для своєчасного виявлення дефектів та регулювання робочих параметрів, що допомагає зменшити зношення. А в роботі [9] пропонується система, яка автоматично регулює роботу багаторівневих конвеєрних систем, орієнтуючись на їхній технічний стан, що підвищує енергоефективність і знижує ризики аварій.

Таким чином, аналіз сучасної літератури свідчить про наявність численних досліджень у сфері управління стрічковими конвеєрами з урахуванням технічного стану обладнання. Проте, комплексні системи, які в єдиній архітектурі враховують одночасно параметри керування, технічного стану і навантаження, залишаються перспективним напрямом для подальших досліджень і практичної реалізації.

Постановка завдання. Мета дослідження полягає у розробленні системи автоматизованого керування важким стрічковим конвеєром на основі комплексного врахування технічного стану його обладнання для підвищення надійності та ефективності роботи. Науковий фокус роботи спрямований на інтеграцію методів моніторингу технічного стану конвеєра з алгоритмами керування його режимами, що дозволяє знизити знос основних елементів і запобігти аварійним відмовам.

Виклад основного матеріалу. Оптимальне керування конвеєром на основі технічного стану його вузлів передбачає автоматичний вибір пріоритетного компонента, залежно від якого коригуються швидкість транспортування, натяг стрічки, рівень завантаження, а також перерозподіляється тягове зусилля між барабанами.

Відомо, що найдорожчою і водночас найменш надійною частиною системи є конвеєрна стрічка, термін служби якої визначається характером діючих навантажень і типами зносу. Вид зносу, своєю чергою, залежить від фракційного складу матеріалу, що транспортується.

Знос робочої обкладки конвеєрної стрічки під час транспортування вантажу мілких фракцій відбувається переважно через функціональну втому поверхневого шару, спричинену впливом ковзних виступів матеріалу. Згідно з теорією фрикційно-втомного зносу при пружному контакті, розробленою І.В. Крагельським, лінійна інтенсивність зносу визначається таким виразом:

$$J_l = \varepsilon \cdot H_{max} \cdot \eta_a / (V+1) \cdot n_f \cdot d_{av}$$

де J_l – лінійна інтенсивність зносу, що відображає співвідношення висоти зношеного шару до пройденого шляху тертя; ε – відносне зближення поверхні обкладки з виступами вантажу; H_{max} – максимальна висота виступів вантажу, мкм; η_a – фактична площа контакту, віднос. єдиний; V – показник шорсткості вантажу, який транспортується; n_f – число циклів до руйнування; d_{av} – середній діаметр плями контакту окремого виступу з поверхнею обкладки. Рівняння інтенсивності зносу для залізної руди набуває такого вигляду [10]:

$$J_l = 0{,}92 \cdot P_{nom}^{\,1+0{,}11t_d} \cdot E^{0{,}89t_d^{-1}} \left(f / \sigma_0\right)^{t_d}$$

У цьому виразі f – коефіцієнт тертя вантажу по обкладці; E – модуль пружності обкладки, Па; t_d – коефіцієнт динамічної витривалості обкладки; P_{nom} – номінальний тиск (навантаження).

З рівняння випливає, що знос стрічки залежить від навантаження нелінійним чином. Отже, зменшення навантаження, що не погіршує економічних показників, збільшує працездатність стрічки.

Наступною за рівнем надійності слабкою ланкою є ролики, до того ж найбільш вразливою їхньою частиною є підшипники.

Таким чином, керування режимами роботи конвеєра повинно насамперед забезпечувати захист стрічки і роликів від надлишкових навантажень, які можна мінімізувати за допомогою автоматизованої системи керування.

Особливо інтенсивний знос футерування барабана, неробочої обкладки стрічки й оболонок роликів спостерігається під час запуску конвеєра. Аналіз пускових режимів свідчить, що запуск стрічкових конвеєрів зазвичай відбувається або з підтримкою майже постійного моменту приводного двигуна, або шляхом контролю прискорення приводного барабана. Однак обидва ці методи супроводжуються значними динамічними зусиллями в стрічці, які можуть удвічі й більше разів перевищувати навантаження в усталеному режимі.

Запуск конвеєра під навантаженням часто викликає ривки та пробуксовки, що спричиняє руйнування стикових з'єднань стрічки та прискорений знос футерування барабана і нижнього шару обкладки стрічки. Основна причина цього – неефективність наявних пускових режимів і відсутність системи контролю натягу стрічки.

У початковий момент пуску приводний барабан швидко набирає окружну швидкість, яка незабаром зрівнюється зі швидкістю руху стрічки. Однак до того, як хвиля розтягування досягне приводного барабана по нижній гілці, прискорення барабана припиняється, а швидкість стрічки, навпаки, починає знижуватися. Різниця між цими швидкостями призводить до прослизання (пробуксовки) стрічки по поверхні барабана. Пробуксовка виникає в момент, коли окружне зусилля барабана $P = S_{in} - S_{out}$ досягає максимального значення, а натяг стрічки S_{out} у точці збігання з барабана різко зменшується.

Розглянемо наявні методи пуску стрічкового конвеєра з погляду впливу на знос стрічки і барабанів. Запуск конвеєра складається з двох етапів: початкового рушання і подальшого розгону. Для оцінки умов виникнення прослизання проаналізуємо перший спосіб запуску, за якого окружне зусилля приводного барабана у фазі рушання визначається таким виразом:

$$P(t) = E_0 du(0,t) / dx =$$

$$= \left(P_d + m_d g x / (1+x) \cdot \left[1 - \exp\left(-\sqrt{E_0 \rho_e} / m_d\right)\right]\right).$$

При цьому, через швидке зменшення експоненціальної складової, для довгих і важких стрічкових конвеєрів до кінця фази рушання зусилля досягає максимального значення:

$$P_{max} \approx m_d g x / (1+x)$$

де x – швидкість поширення пружної хвилі (для конвеєрів гірничих підприємств $0{,}7 < x < 1$, а отже, $0{,}4 < x / (1 + x) < 0{,}5$); q – інтенсивність сил опору, яка визначається як відношення сили опору руху стрічки на одиницю її довжини до приведеної погонної маси стрічки; m_d – зосереджена маса приводу: E_0 – агрегатна динамічна жорсткість стрічки; ρ_e – еквівалентна погонна маса моделі конвеєрної стрічки, яка визначається за формулами:

$$\rho_e = E_0 / a_e^2,$$

$$a_e = 2a_1 a_2 / (a_1 + a_2)$$

де a_1, a_2 – швидкості поширення хвилі деформації по навантаженій і порожній гілках.

У фазі розгону окружне зусилля приводного барабана визначається рівнянням руху приведеної маси приводу: $P_{circ} = P_d - m_d j_{mov}$, де j_{mov} – прискорення рухомих мас конвеєра. Максимальне зусилля P_{max} значно перевищує зусилля у фазі розгону, оскільки у фазі рушання сили опору руху підключаються поступово, що призводить до нерівномірних навантажень на систему.

Для режиму з постійним прискоренням приводного барабана конвеєра застосовують такі формули:

$$P_{\text{circ}} = P_0 + 2\rho_e l j = 2\rho_e l (q_e + j)$$

$$P_{max} = 2\rho_e l \left(q_e + j + \sqrt{j(q_e + j)}\right)$$

Максимальне окружне зусилля у фазі розгону значно перевищує P_{circ} при цьому в момент переходу від стадії рушання до стадії розгону спостерігається різкий стрибок навантаження.

Таким чином, до завершення фази рушання окружні зусилля і натяг стрічки в точці її входу на приводний барабан досягають граничних значень, які істотно перевищують аналогічні показники у фазі розгону і в сталому режимі роботи.

Подібні перевантаження призводять не тільки до прискореного зносу стрічки і футерування барабана, а й спричиняють деформацію стикувальних з'єднань, а також інтенсивні коливання тягового органу, що негативно позначається на стані роликів і підшипників, прискорюючи їхнє руйнування.

Пробуксовки, ривки та значні динамічні зусилля в стрічці під час запуску конвеєра зумовили необхідність розроблення алгоритмів автоматичного пуску, що враховують технічний стан стрічки та наявність вантажу на ній. Такі алгоритми мають забезпечувати плавний старт, знижуючи рівень навантажень у критичні моменти.

Як перший етап реалізації цього підходу необхідно визначити тип електроприводу, який дасть змогу забезпечити плавний запуск конвеєра й ефективне регулювання швидкості.

Найповніше вимогам, що висуваються, задовольняє частотно регульований електропривод.

З урахуванням поставлених завдань було сформульовано основні принципи, на основі яких розроблено алгоритм керування тяговим зусиллям. При цьому величина прискорення на різних ділянках розраховується з урахуванням наявності матеріалу на стрічці, гранулометричного складу вантажу (включно з умістом великих фракцій) та інтенсивності зносу конвеєрної стрічки. Моменти двигуна на етапах рушання і розгону також залежать від цих параметрів.

Оптимальним рішенням у фазі рушання є зниження пускового моменту приводного двигуна порівняно з фазою розгону. Це дасть змогу зменшити навантаження на стрічку і, отже, знизити її знос, особливо в області нижньої обкладки.

Діаграму рекомендованої залежності моменту і швидкості електродвигунів приводу в період запуску наведено на рис. 1.

Функціонування алгоритму (рис. 2) проходить у чотири послідовні етапи: підготовчий, розрахунок моменту двигуна у фазі розгону M_{eng2}, розрахунок моменту у фазі рушання M_{eng1}, запуск двигуна за керуючим сигналом.

На першому етапі вводяться вихідні дані, необхідні для визначення пускового прискорення і крутного моменту двигуна. Також оцінюється кількість вантажу на стрічці перед зупинкою і можлива маса найбільшого фрагмента матеріалу. Потім обчислюється середня лінійна $P_{пот}$ (блок 3), на її основі – інтенсивність зносу обкладки стрічки (блок 4) і прискорення стрічки на другій ділянці пускової діаграми (блок 5):

$$J = \left(\left[P_{пот}\right] - P_{\Sigma}\right)\left(q_n k_p\left[\left(1 + 0{,}486 m_p q_n^{-1}\right) g\right]^{-1} + m_k \sin\Delta\theta\right),$$

де $[P_{пот}]$ – допустиме навантаження на стрічку, *н*; P_{Σ} – сумарне навантаження на стрічку:

$$P_{\Sigma} = \left(q_l + q_L + q_r\right) \cdot g \ell_p; \qquad \Delta\theta = \theta_2 - \theta_1.$$

де q_l, q_L, q_r – лінійна маса, відповідно вантажу, стрічки, роликів верхньої гілки, кг/м.

На наступному етапі проводиться розрахунок відсутніх параметрів, необхідних для визначення моменту двигуна на основній ділянці розгону.

Зокрема, обчислюють приведену масу вантажу і рухомих частин конвеєра G_e (блок 6):

$$G_e = k_e\left(m_0 + k_g m_{rot} + m_L\right),$$

де k_e – коефіцієнт, що враховує подовження стрічки; m_0 – маса ходової частини конвеєра; m_{rot} – маса частин, що обертаються; m_L – маса вантажу на стрічці, кг; k_g = 0,5 ÷ 0,9 – коефіцієнт, який враховує, що колова швидкість мас, що обертаються, менша за швидкість стрічки.

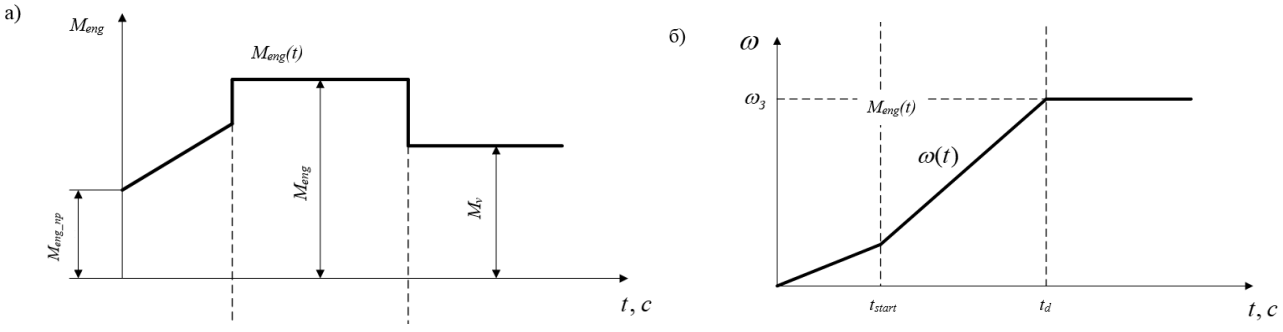


Рис. 1. Зміна режимів роботи двигуна під час пуску конвеєра: а) зміна моменту двигуна; б) зміна швидкості двигуна

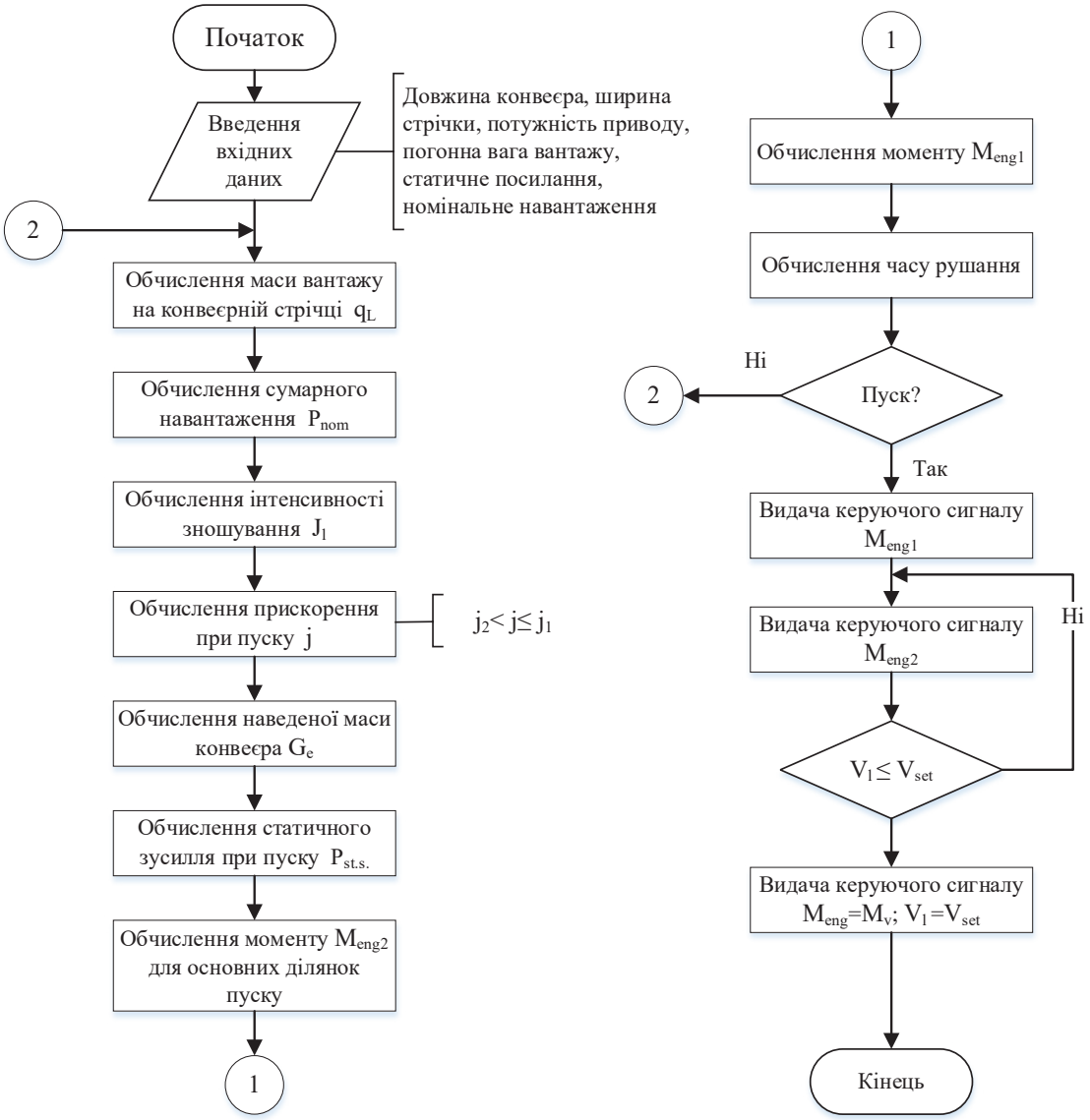


Рис. 2. Алгоритм формування закону керування пуском конвеєра

У блоці 7 визначається статичне зусилля в період пуску Pst.s:

$$P_{st.s.} = q_l H + \omega_{st}\left(P - q_l H\right),$$

де q_l – лінійне навантаження конвеєра; P – тягове зусилля під час усталеного руху, н; H – висота підйому вантажу, м; ω_{st} – коефіцієнт статичних опорів тертя.

У блоці 9 проводиться остаточний розрахунок крутного моменту двигуна на основній ділянці розгону M_{eng2}:

$$M_{eng2} = j[D_0^{-1}\left(G_e D_0^2 + C_i\left(G_i D_i^2\right)\cdot i_0^2\eta_0 + P_{st.s.}D_0\right]\cdot(2i_0\eta_0)^{-1},$$

причому $M_{eng2} \leq 1{,}3$ М.

Після цього приступають до обчислення параметрів першої ділянки діаграми, а саме, моменту і тривалості першої фази пуску (рушання). Для обчислення часу рушання використовуємо низку залежностей, наведених до досліджуваної задачі (блок 10):

$$x = 1 - 2\rho_e q_e l\,(M_{eng2}/D_0 - P_{circ,0})^{-1}$$

$$a_e = (E_0/\rho_e)^{-1/2}; \qquad t_{start} = 2La_e^{-1}\left(1 + 1/x\right),$$

де x – відносна швидкість пружної хвилі; ρ_e, Q_e – еквівалентні інтенсивність сил опору двигуну і погонна маса; $P_{circ,0}$ – початкове окружне зусилля; E_0 – динамічна агрегатна жорсткість стрічки; L – довжина конвеєра.

Потім обчислюють момент двигуна на ділянці рушання за відомою залежністю:

$$M_{eng1} = D_0(i_0\eta_0)^{-1}\left(m_d kxa_e + P_{circ,0} + kE_0 t\right),$$

де $k = q_e x / a_e\left(1 - x^2\right)$

Після цього запускають двигун і подають керуючий сигнал для реалізації фази рушання

(M_{eng2}, t_{start}). Після завершення цієї фази формується сигнал на розгін двигуна M_{eng2}. Паралельно контролюють процес розгону $V_b \leq V_{set}$ і забезпечують роботу приводу у встановлених режимах.

Такий поступовий запуск дає змогу не тільки захистити стрічку від надмірних навантажень, а й знизити знос роликів, барабанів і опорних конструкцій.

Швидкість руху стрічки, особливо якщо вона вже має певний знос або пошкодження, істотно впливає на подальший розвиток дефектів і прискорення зношування. Зниження швидкості транспортування в моменти підвищеного навантаження (наприклад, під час перевезення великогабаритного вантажу) або в разі прогресуючих ушкоджень – таких як руйнування кромки або поява «пробоїн» у покритті – допомагає продовжити термін служби стрічки та зберегти її працездатність.

На рисунку 3 наведено алгоритм керування швидкістю конвеєра і натягом стрічки відповідно до її технічного стану. Оскільки стан стрічки тісно пов'язаний зі станом приводних і від відхиляючих барабанів, наявністю прослизання і технічним станом роликів, ці елементи розглядаються як єдина взаємопов'язана система [12, 13].

На початковому етапі здійснюють введення вихідних параметрів, необхідних для побудови алгоритму запуску конвеєра і подальшого регулювання його роботи. Зокрема, задають номінальну швидкість, швидкість обертання барабана і натяг стрічки.

Після цього перевіряють, чи відбувається прослизання стрічки на барабані. Для цього порівнюють швидкість руху стрічки на барабані та на робочій гілці. Якщо прослизання зафіксовано, оцінюють натяг стрічки і, за потреби, його коригують.

Далі контролюють технічний стан підшипникових вузлів барабана і оболонки. Окрему увагу приділяють самій стрічці – перевіряють наявність поздовжніх і поперечних розривів, а також розбіжність швів. Виявлені пошкодження порівнюють з допустимими або аварійними межами. Залежно від результатів, система формує команди на зниження швидкості, зменшення лінійного навантаження або повне відключення конвеєра.

Ступінь зносу стрічки визначається за спеціальним алгоритмом. Якщо знос досягає рівня, який ще дозволяє експлуатацію ($\delta_l + \Delta$), навантаження на конвеєр знижують для полегшення умов роботи. Якщо ж знос наближається до критичного ($\delta_l + \Delta_d$), але є необхідність продовжити роботу до найближчого ремонту, формується додатковий сигнал на зниження швидкості транспортування.

Наступним важливим елементом є опорні ролики [13]. Їхній стан оцінюють за станом підшипників і зношеністю оболонки. Якщо кількість роликів з несправними підшипниками або сильно зношеними оболонками не перевищує допустиму норму, переходять на щадний режим експлуатації. Якщо ж пошкоджень більше – конвеєр переводять в аварійний режим або зупиняють.

Висновки. У результаті проведеного дослідження запропоновано підхід до автоматизованого керування важким стрічковим конвеєром з урахуванням технічного стану його обладнання. Розроблено алгоритми плавного пуску конвеєра з урахуванням зносу стрічки і навантаження, що дозволяють значно знизити динамічні навантаження на елементи системи. Це забезпечує зменшення інтенсивності зношування стрічки та роликів, знижує ризик аварійних відмов і підвищує загальну надійність конвеєрної лінії. Додатково, інтегроване регулювання швидкості та натягу стрічки на основі моніторингу її технічного стану дозволяє своєчасно виявляти погіршення стану обладнання та оперативно коригувати режими роботи. Досягнуто підвищення рівня автоматизації керування конвеєром, оскільки система самостійно адаптується до поточного стану вузлів без втручання оператора. Впровадження запропонованих рішень сприяє подовженню терміну служби основних компонентів та покращує техніко-економічні показники роботи транспортної системи. Втім, отримані результати є лише одним кроком у вирішенні проблеми, і окремі аспекти потребують подальших досліджень. Зокрема, актуальним є удосконалення алгоритмів пуску для різних умов експлуатації та типів конвеєрів, щоб забезпечити оптимальні режими розгону без перевантажень. Перспективним напрямом є впровадження нових матеріалів з підвищеною зносостійкістю для стрічок та роликів, що дозволить зменшити зношення і продовжити ресурс обладнання. Подальшого розвитку потребують інтегровані системи моніторингу технічного стану, здатні у реальному часі оцінювати параметри роботи конвеєра і прогнозувати виникнення несправностей. Поєднання таких систем з автоматизованим керуванням створить передумови для максимально надійної та ефективної роботи важких стрічкових конвеєрів.

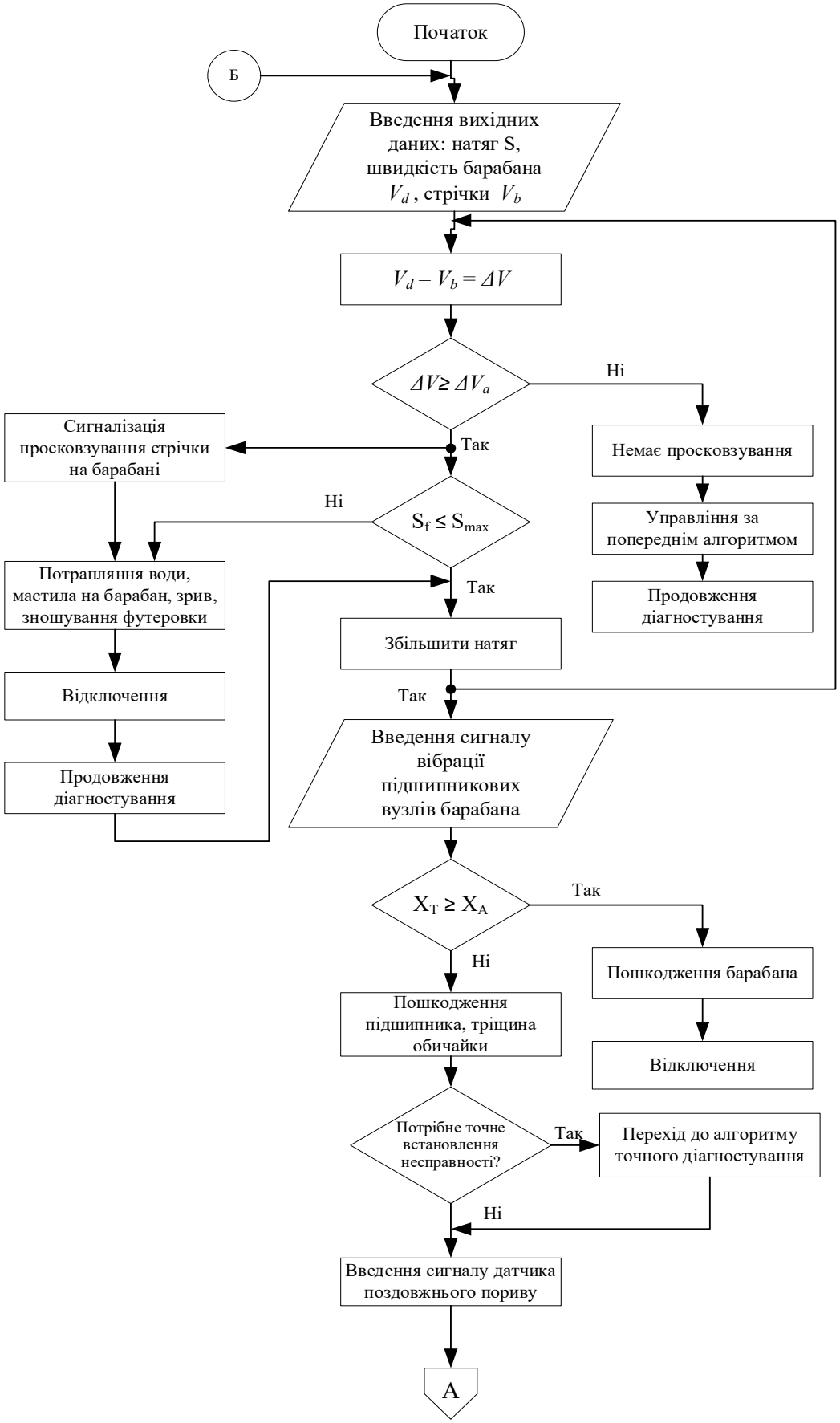


Рис. 3. Алгоритм формування керуючих сигналів за технічним станом елементів стрічкового конвеєра

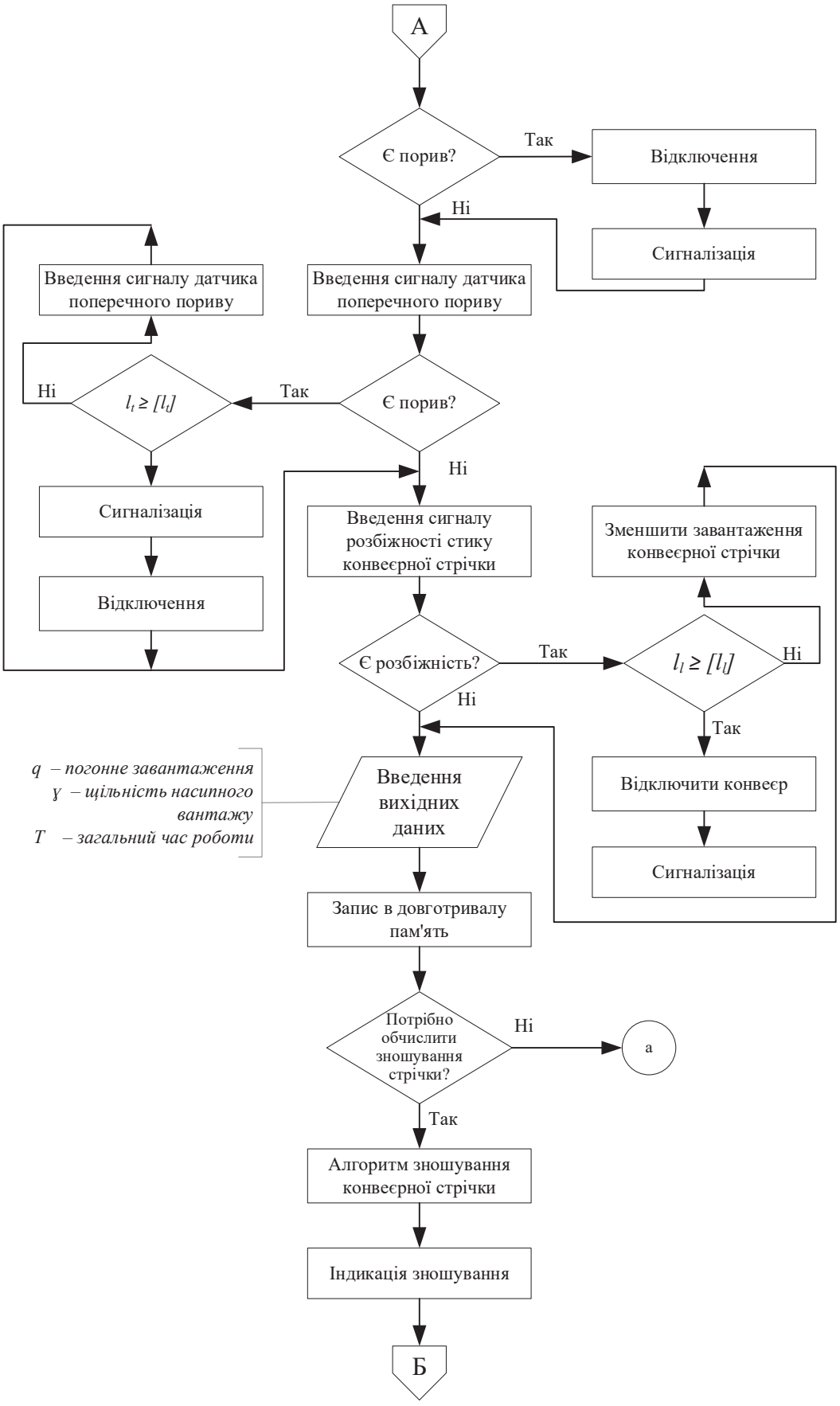


Рис. 3. Алгоритм формування керуючих сигналів за технічним станом елементів стрічкового конвеєра (продовження)

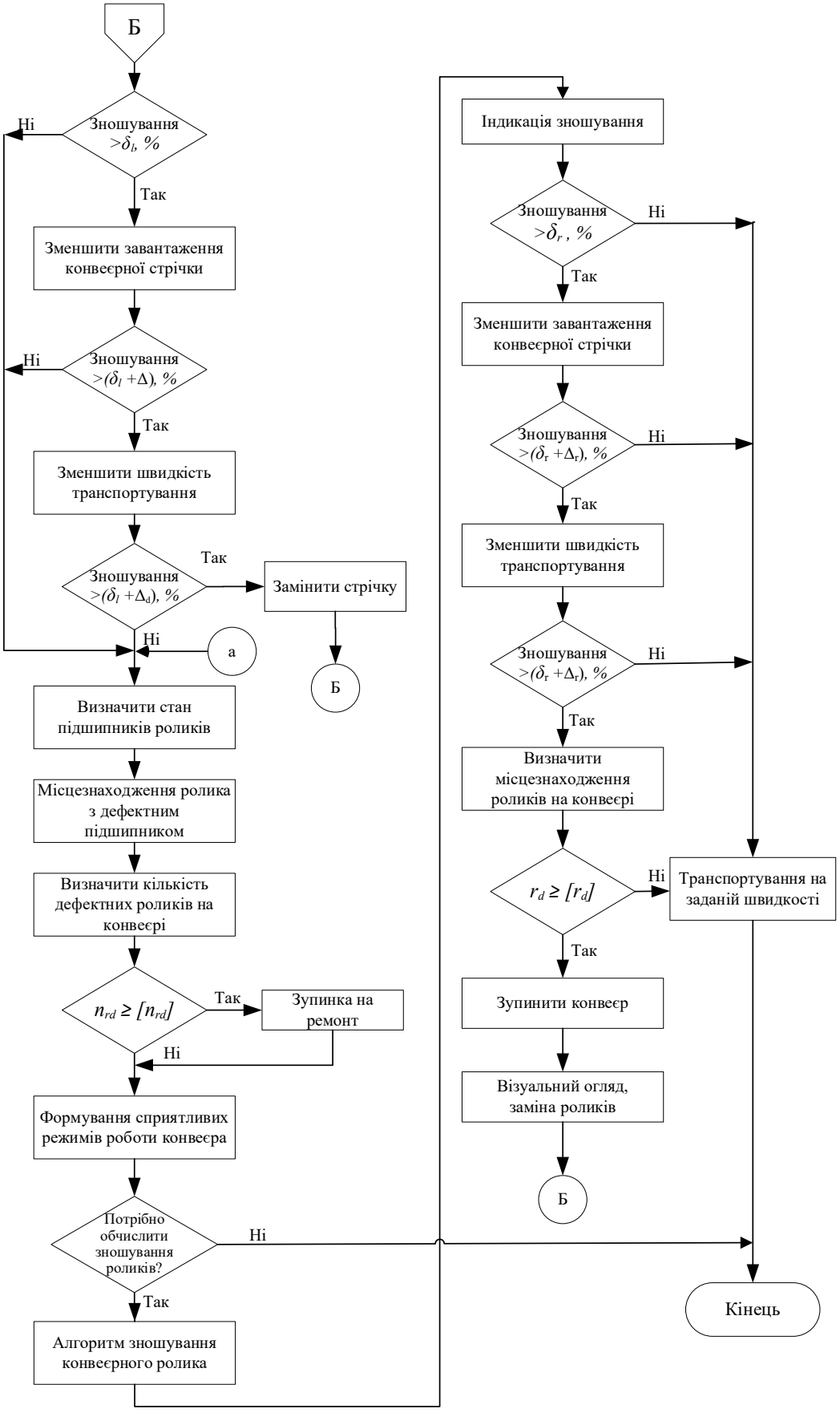


Рис. 3. Алгоритм формування керуючих сигналів за технічним станом елементів стрічкового конвеєра (продовження)

Список літератури:

1. He D., Pang Y., Lodewijks G. Speed control of belt conveyors during transient operation. *Powder Technology.* 2016. Vol. 301, P. 622–631. DOI: 10.1016/j.powtec.2016.07.004.
2. Печеник М.В., Бур'ян С.О., Маліборський С.О. Аналіз характеру відпрацювання динамічної похибки за швидкістю електромеханічної системи, спричиненої штучною зміною навантаження в автоматизованому комплексі металообробки. *Вісник ВПІ*, 2021. Вип. 5, С. 103–107. DOI: https://doi.org/10.31649/1997-9266-2021-158-5-103-107
3. Zhang S., Mao W. Optimal operation of coal conveying systems assembled with crushers using model predictive control methodology. *Applied Energy*. 2017. Vol. 198, P. 65–76. DOI: 10.1016/j.apenergy.2017.04.037.
4. Wu M., Goh K.W., Chaw K.H., et al. An intelligent predictive maintenance system based on random forest for addressing industrial conveyor belt challenges. *Frontiers in Mechanical Engineering*. 2024. Vol. 10. DOI: 10.3389/fmech.2024.1383202.
5. Zhou M., Chen Y., Hu F., Lai W., Gao L. A deep learning approach for accurate assessment of conveyor belt wear state based on multispectral imaging. *Optics & Laser Technology*. 2025. Vol. 181. Part B. 111782. URL: https://doi.org/10.1016/j.optlastec.2024.111782.
6. Soares J.L.L., Costa T.B., Moura L.S., et al. Fault diagnosis of belt conveyor idlers based on gradient boosting decision tree. *The International Journal of Advanced Manufacturing Technology.* 2024. Vol. 132. P. 3479–3488. URL: https://link.springer.com/article/10.1007/s00170-024-13549-0.
7. Liu X., He D., Lodewijks G., Pang Y., Mei J. Integrated decision making for predictive maintenance of belt conveyor systems. *Reliability Engineering & System Safety.* 2019. Vol. 188. P. 347–351. DOI: 10.1016/j.ress.2019.03.047.
8. Mahmood M. S., Shareef I. R. Applications of Artificial Intelligence for Smart Conveyor Belt Monitoring Systems: A Comprehensive Review. *Journal Européen des Systèmes Automatisés.* 2024. Vol. 57 (4). P. 1195–1206. DOI: 10.18280/jesa.570426.
9. Fei Zeng; Zhe-yang Xie. Coordinated Control System of Multi-Level Belt Conveyors for Promotion the Energy Efficiency Based on IoT-Technology. *5th International Conference on Information Science and Control Engineering, Zhengzhou*, China. 2018. DOI: 10.1109/ICISCE.2018.00145
10. Лобов В.Й., Єфіменко Л.І., Тиханський М.П., Рубан С.А. Автоматизовані системи керування конвеєрними установками : монографія. Кривий Ріг: Видавничий центр ДВНЗ "КНУ", 2015. 450 с.
11. Тиханський М. П., Єфіменко Л. І., Тиханська А. М. Узгодження системи керування режимами роботи й системи діагностування технічного стану конвеєра. *Гірничий вісник*. 2017. Вип. 102. С. 129–134.
12. Єфіменко Л.І., Тиханський М.П., Тиханська А.М. Автоматичне керування конвеєрними установками з адаптацією за технічним станом його обладнання. *Вісник Криворізького національного університету.* 2020. Вип. 50. С.122–126. DOI: 10.31721/2306-5451-2020-1-50-122-126
13. Єфіменко Л.І., Тиханський М.П., Тиханська А.М. Структура блоку визначення технічного стану роликів стрічкового конвеєра оснащеного автоматичною системою керування. *Вісник Криворізького національного університету.* 2019. Вип. 49. С. 96–101.

Tykhanska A.M., Tykhanskyi M.P. CONTROL OF A HEAVY BELT CONVEYOR TAKING INTO ACCOUNT THE TECHNICAL CONDITION OF ITS EQUIPMENT

The article presents a comprehensive approach to the control of a heavy belt conveyor that takes into account the technical condition of its main units and mechanisms. The relevance of the study is due to the need to increase the reliability and efficiency of conveyor systems at mining enterprises, where uneven wear of the belt, rollers, and drums leads to unplanned downtimes and increased maintenance costs. To prevent premature equipment failures and ensure safe operation, the conveyor must be controlled in a way that adapts to the current technical condition of its components. The study includes an analysis of modern solutions and scientific works in the field of conveyor technologies, as well as theoretical modeling of dynamic loads and wear of conveyor elements during start-up and operation under load. As part of the proposed approach, a variable-frequency drive is used to enable smooth conveyor start-up, and algorithms for monitoring technical condition are developed and integrated into the control system.

The results show that the implementation of an intelligent control system significantly reduces dynamic loads on the belt and mechanical components. An algorithm for automated smooth conveyor start-up has been developed, which adjusts acceleration and traction force based on the load mass on the belt and its degree of wear. This avoids sharp jerks and slippage during start-up, reducing peak loads that previously exceeded

steady-state levels by up to two times under traditional start-up methods. An algorithm is also proposed for regulating belt speed and tension during conveyor operation, based on continuous monitoring of equipment condition. The control system detects whether slippage of the belt on the drum occurs, monitors overheating of bearings and the appearance of belt damage, and automatically reduces speed or load, or even stops the conveyor as critical thresholds are approached.

The novelty of the obtained results lies in the integrated control approach: for the first time, algorithms for both smooth start-up and steady-state operation of the conveyor are combined with continuous monitoring of the technical condition of the belt and other components. The research results can be implemented at mining enterprises and in other industries that use heavy belt conveyors, with the aim of improving their efficiency and operational safety.

Key words: *belt conveyor, technical condition, automated control, start-up mode, intelligent systems, belt wear, belt tension, slippage, variable-frequency drive, equipment monitoring.*

УДК 004.891.3:621.398
DOI https://doi.org/10.32782/2663-5941/2025.3.2/54

Усов О.С.
Національний університет «Одеська політехніка»

МЕТОДИ ВІЗУАЛЬНОЇ SLAM ДЛЯ НАВІГАЦІЇ У ВНУТРІШНЬОМУ СЕРЕДОВИЩІ БЕЗ ДОСТУПУ ДО ГЛОБАЛЬНОГО ПОЗИЦІОНУВАННЯ

Актуальність дослідження зумовлена потребою підвищення точності одночасної локалізації та картографування (Simultaneous Localization and Mapping, SLAM) у внутрішніх просторах без доступу до глобальної навігації, що є критично важливим для забезпечення стабільної навігації в середовищах зі складною структурою та динамічними об'єктами.

Метою дослідження є здійснення комплексного аналізу методів реалізації візуальної одночасної локалізації та картографування у внутрішніх умовах з акцентом на визначення оптимальних підходів для підвищення точності локалізації та створення ефективних алгоритмів картографування.

Методологія дослідження охоплює порівняльний аналіз наявних методів SLAM, структурно-функціональний аналіз сенсорних систем (монокулярних, стерео- і RGB-D камер, а також інерціальних вимірювальних одиниць), а також моделювання впливу алгоритмів глибинного навчання на сегментацію динамічних об'єктів та зменшення похибок у побудові карт простору.

Результати дослідження. Встановлено, що комбіноване використання сенсорів різних типів дає змогу зменшити похибки локалізації за рахунок взаємної компенсації недоліків кожної системи. Доведено, що інтеграція алгоритмів глибинного навчання сприяє підвищенню точності ідентифікації динамічних об'єктів, що допомагає зменшити помилки картографування в середовищах зі складною архітектурою. Виявлено, що системи SLAM, що зорієнтовані на семантичну сегментацію, забезпечують більш стабільну локалізацію в середовищах із повторюваними структурами, проте вимагають значних обчислювальних ресурсів.

Висновки. Обґрунтовано доцільність використання гібридних SLAM-систем, які поєднують різні типи сенсорів та алгоритми глибинного навчання для компенсації похибок і стабілізації локалізації в динамічних середовищах. Рекомендовано впровадження технологій інфрачервоного підсвічування та прогнозування траєкторій рухомих об'єктів для зменшення впливу динамічних факторів.

Перспективи дослідження пов'язані з розробленням адаптивних SLAM-архітектур з автоматичним налаштуванням параметрів сенсорів та алгоритмів для забезпечення стабільної локалізації та картографування в середовищах зі складною структурою та високою динамічністю.

Ключові слова: *навігаційні системи, динамічні об'єкти, семантична сегментація, нейромережі, локалізація, картографування, сенсори.*

Постановка проблеми. Сучасні системи навігації у внутрішньому середовищі зіштовхуються з низкою викликів, зумовлених відсутністю доступу до глобальних навігаційних систем, зокрема глобальної системи позиціонування (Global Positioning System, GPS), що обмежує можливості орієнтації та локалізації в умовах замкнених просторів. У будівлях зі складною архітектурною структурою, насичених такими однотипними елементами, як коридори та кімнати, навігація ускладнюється через відсутність чітких орієнтирів і проблеми повторюваності структур. Візуальні системи SLAM, що використовують дані з камер та інших сенсорів, надають можливість одночасно будувати карту середовища та визначати місцезнаходження об'єкта в просторі. Проте точність та ефективність цих методів суттєво залежать від типу сенсора, з-поміж яких монокулярні, стерео- та RGB-D камери (Red, Green, Blue – Depth, тобто камери, що поєднують зображення з глибоким каналом) або інерціальні вимірювальні одиниці (Inertial Measurement Unit, IMU). Зокрема використання глибинних камер може значно підвищити точність локалізації в складних умовах, проте збільшує обчислювальне навантаження на систему. Впровадження алгоритмів глибинного навчання для семантичної сегментації та фільтрації динамічних об'єктів розширює можливості традиційних методів одночасної локалізації та картографування, забезпечуючи більш точне розпізнавання навколишніх об'єктів і підвищуючи

стійкість до змін середовища. Водночас зростає необхідність дослідження ефективності таких методів у середовищах зі значними перепонами для виявлення головних обмежень та розроблення нових підходів до підвищення точності картографування та локалізації без доступу до глобальних навігаційних систем.

Аналіз останніх досліджень і публікацій. Аналіз наукових досліджень щодо методів візуального SLAM для навігації у внутрішньому середовищі дає змогу виокремити чотири головні напрями, що відображають сучасні технічні рішення для побудови надійних навігаційних систем у приміщеннях.

Перший напрям зосереджується на використанні різних сенсорів для підвищення точності SLAM-систем. Так, А. Мерзляков (A. Merzlyakov) та С. Масенскі (S. Macenski) у своєму дослідженні аналізують ефективність монокулярних, стерео- та RGB-D камер для SLAM, підкреслюючи значення коректного вибору сенсорів для підвищення точності в складних середовищах [1]. Дослідники З. С. Хашеміфар (Z. S. Hashemifar), С. Адіварахан (C. Adhivarahan), А. Балакрішнан (A. Balakrishnan) та К. Данту (K. Dantu) пропонують інтеграцію Wi-Fi сенсорів у SLAM-системи для підвищення точності локалізації в приміщеннях із низьким рівнем освітлення [2]. Автори С. Сюй (C. Xu), Ч. Лю (Z. Liu) та Ч. Лі (Z. Li) досліджують поєднання стереокамер з інерційними вимірювальними одиницями (IMU) для підвищення стабільності позиціонування в динамічних умовах [3]. Науковці Ц. Чен (C. Chen), Х. Чжу (H. Zhu), Л. Ван (L. Wang) та Ю. Лю (Y. Liu) пропонують алгоритм стерео-візуально-інерціального SLAM, що забезпечує високу точність локалізації за відсутності оклюзій [4]. Натомість М. Хельмберґер (M. Helmberger), К. Морін (K. Morin), Б. Бернер (B. Berner), Н. Кумар (N. Kumar), Дж. Чіоффі (G. Cioffi) та Д. Скарамуцца (D. Scaramuzza) представляють Hilti SLAM Challenge Dataset як тестову платформу для оцінки SLAM-систем з урахуванням різних типів сенсорів [5]. Подальші дослідження в цьому напрямі мають бути зосереджені на розробленні адаптивних алгоритмів, що здатні ефективно комбінувати дані з різних типів сенсорів для компенсації зовнішніх впливів.

Другий напрям присвячений інтеграції глибинного навчання у SLAM-системи для покращення обробки динамічних об'єктів. У своєму дослідженні Д. Дам'янович (D. Damjanović), П. Біочич (P. Biočić), С. Праклячич (S. Prakljačić), Д. Чинчурак (D. Činčurak) та Й. Бален (J. Balen) аналізують використання згорткових нейронних мереж (CNN) для семантичної сегментації рухомих об'єктів у внутрішніх середовищах [6]. Автори Й. Чарновскі (J. Czarnowski), Т. Лейдлоу (T. Laidlow), Р. Кларк (R. Clark) та А. Дж. Девісон (A. J. Davison) представляють алгоритм DeepFactors для щільного монокулярного SLAM у реальному часі, що дає змогу зменшити обчислювальне навантаження під час оброблення великих об'ємів даних [7]. Науковці Б. Бескос (B. Bescos), С. Кампос (C. Campos), Х. Д. Тардос (J. D. Tardós) та Х. Неіра (J. Neira) пропонують DynaSLAM II, що поєднує мультиоб'єктний трекінг із SLAM для одночасної обробки статичних і рухомих об'єктів [8]. У своєму дослідженні К. Дж. Сінґх (K. J. Singh), Д. С. Капур (D. S. Kapoor), К. Тхакур (K. Thakur), А. Шарма (A. Sharma), А. Найяр (A. Nayyar) та С. Махаджан (S. Mahajan) досліджують активний SLAM у соціально-орієнтованих середовищах, що передбачає побудову карт у реальному часі з використанням даних глибинного навчання [9]. Для подальших досліджень необхідно зосередитися на інтеграції глибинного навчання з методами класифікації та виокремлення рухомих об'єктів для підвищення точності навігації.

Третій напрям охоплює адаптацію SLAM-систем для людей із порушеннями зору. У своєму огляді М. Бамдад (M. Bamdad), Д. Скарамуцца (D. Scaramuzza) та А. Дарвіші (A. Darvishy) аналізують переваги візуального SLAM для просторової орієнтації, наголошуючи на важливості коректної адаптації алгоритмів для забезпечення стабільності позиціонування [10]. Так, Р. Н. Кандалан (R. N. Kandalan) та К. Намудурі (K. Namuduri) розглядають методи корекції траєкторії для користувачів із вадами зору, акцентуючи на компенсації неточностей у системах із монокулярними камерами [11]. Автори Д. Хан (D. Khan), Ч. Чен (Z. Cheng), Х. Учійяма (H. Uchiyama), С. Алі (S. Ali), М. Асшад (M. Asshad) та К. Кійокава (K. Kiyokawa) наголошують на необхідності врахування специфіки просторової орієнтації у SLAM-системах для людей із сенсорними обмеженнями [12]. Натомість С. Томажич (S. Tomažič) розглядає внутрішнє позиціонування для людей із порушеннями зору з акцентом на стабільності систем у приміщеннях зі змінним освітленням [13]. Подальші розробки мають бути спрямовані на інтеграцію SLAM із тактильними і аудіовізу-

альними сенсорами для підвищення точності та безпеки.

Четвертий напрям стосується адаптації SLAM для мобільної робототехніки з обмеженими обчислювальними ресурсами. У своєму дослідженні Ю. Чжан (Y. Zhang), М. Сяо (M. Hsiao), Ю. Чжао (Y. Zhao), Цз. Донг (J. Dong) та Д. Дж. Енгель (J. J. Engel) пропонують клієнт-серверну оптимізацію для SLAM-систем з обмеженими ресурсами, що дає змогу зменшити обчислювальні витрати [14]. Науковці П. Ю. Цзен (P. Y. Tseng), Цз. Дж. Лінь (J. J. Lin), Ю. Ч. Чан (Y. C. Chan) та А. Ю. Чен (A. Y. Chen) представляють систему для швидкої локалізації в кризових умовах із використанням візуального SLAM, що оптимізована для роботи з малопотужними пристроями [15]. Для підвищення ефективності систем у подальших дослідженнях необхідно розробити методи компресії даних та оптимізації алгоритмів для роботи на мобільних платформах.

Аналіз літератури засвідчує зацікавленість у використанні візуального SLAM для навігації у внутрішніх середовищах без доступу до глобального позиціонування. Головні напрями досліджень передбачають використання мультисенсорних систем, інтеграцію глибинного навчання, адаптацію технологій для людей із вадами зору та оптимізацію алгоритмів для мобільних платформ. Для подальших досліджень доцільно зосередитися на інтеграції різних сенсорних систем, розробленні гібридних моделей SLAM та адаптації алгоритмів до умов динамічних середовищ.

Попри значний прогрес у розробленні систем SLAM, невирішеними залишаються питання, що пов'язані з точністю локалізації у внутрішніх умовах за відсутності глобальної навігації. Зокрема недостатньо досліджено вплив різних типів сенсорів на якість картографування в середовищах зі складною структурою, обмеженою освітленістю та наявністю динамічних об'єктів. Використання алгоритмів глибинного навчання також демонструє перспективи у фільтрації динамічних об'єктів, проте потребує подальшого аналізу для забезпечення стабільної локалізації в складних умовах.

Запропоноване дослідження спрямоване на заповнення цих прогалин шляхом комплексного аналізу ефективності різних сенсорних систем у SLAM, розроблення алгоритмів сегментації динамічних об'єктів на основі глибинного навчання та оптимізації методів картографування для підвищення точності локалізації. Це дасть змогу створити адаптивну SLAM-архітектуру, що здатна інтегрувати різнотипні сенсорні дані та зменшувати системні похибки в динамічних середовищах.

Постановка завдання. Метою статті є здійснення комплексного аналізу методів реалізації візуальної одночасної локалізації та картографування (SLAM) у внутрішніх умовах за відсутності глобальної навігації для визначення оптимальних підходів до підвищення точності локалізації та картографування.

Завдання статті:

Визначити особливості використання різних типів сенсорів (монокулярні, стерео- та RGB-D камери, інерціальні вимірювальні одиниці) у візуальних SLAM-системах та оцінити їхній вплив на точність локалізації у внутрішньому середовищі.

Дослідити роль алгоритмів глибинного навчання в покращенні результатів візуального SLAM, зокрема в контексті сегментації динамічних об'єктів і зниження похибок у процесі побудови карт у середовищах із повторюваними структурами.

Розробити рекомендації щодо оптимізації SLAM-систем для внутрішнього середовища з урахуванням обмеженої освітленості, динамічних об'єктів і нестабільних сенсорних даних.

Виклад основного матеріалу. Візуальна одночасна локалізація та картографування (SLAM) у внутрішніх умовах є складним завданням, оскільки відсутність глобальної навігації обмежує можливості точного визначення місцезнаходження та побудови карти простору. У таких умовах значну роль відіграє тип сенсорів, які використовуються для збору візуальних даних. Монокулярні камери забезпечують спрощену реалізацію SLAM, проте вони не здатні надавати глибинну інформацію, що обмежує їхню точність у складних середовищах із багаторазовими перешкодами. Стереокамери, що складаються з двох синхронізованих камер, дають змогу отримувати просторову інформацію, але потребують більш високої обчислювальної потужності для обробки даних. Камери RGB-D інтегрують глибинний канал разом із кольоровим зображенням, що сприяє більш точній локалізації, проте вони мають обмежений діапазон дії та чутливі до змін освітлення. Інерціальні вимірювальні одиниці (IMU) забезпечують додаткову інформацію про прискорення та кутову швидкість, підвищуючи стійкість системи до динамічних змін, однак їх

використання вимагає складних алгоритмів для компенсації дрейфу даних (табл. 1).

На практиці інтеграція різних типів сенсорів у системах SLAM спрямована на подолання обмежень кожного окремого сенсора та забезпечення комплексного збору даних для точного моделювання середовища. Монокулярні камери, що складаються з одного об'єктива, зазвичай використовуються в мобільних роботах з обмеженими обчислювальними ресурсами, де пріоритетом є мінімізація витрат на апаратне забезпечення [9]. Проте відсутність глибинної інформації зумовлює необхідність застосування складних алгоритмів оцінки масштабу, що підвищує ризик накопичення помилок під час руху в середовищах із повторюваними структурами.

Стереокамери, що складаються з двох синхронізованих камер, завдяки здатності до стереоскопічного зчитування забезпечують більш точне позиціонування в умовах обмеженого простору та дають змогу зменшити похибки, спричинені динамічними змінами. Водночас високі вимоги до калібрування та збільшення обчислювальних навантажень можуть ускладнювати використання таких камер у реальному часі.

Камери RGB-D набули поширення в задачах семантичного картографування, де глибинні дані поєднуються з візуальними ознаками для створення тривимірної моделі простору [11]. Проте їхня ефективність значно залежить від стабільного освітлення та відстані до об'єктів, що ускладнює роботу в умовах низької освітленості або великих приміщень.

Інерціальні вимірювальні одиниці (Inertial Measurement Unit, IMU), що складаються з акселерометра та гіроскопа, здатні реєструвати зміни прискорення та обертання. Їх часто використовують для компенсації втрати позиційних даних у складних умовах. Однак накопичення похибок у даних IMU вимагає застосування алгоритмів корекції дрейфу, зокрема розширеного фільтра Калмана (Extended Kalman Filter, EKF) або некалманівського фільтра (Unscented Kalman Filter, UKF) [16]. Так, у сучасних практичних реалізаціях одночасної локалізації та картографування актуальності набуває підхід до комбінованого використання сенсорів і методів машинного навчання для одночасного підвищення точності картографування та зменшення впливу системних похибок.

Системи SLAM у середовищах із повторюваними структурами стикаються з низкою викликів, зумовлених подібністю об'єктів та їхньою динамічністю. У таких умовах традиційні методи SLAM, які переважно базуються на аналізі статичних структур і фіксованих орієнтирів, мають обмеження в здатності до коректного розпізнавання об'єктів. Серед найбільш поширених методів виділяють алгоритми, зорієнтовані на обробку візуальних даних, зокрема ORB-SLAM (Oriented FAST and Rotated BRIEF), LSD-SLAM (Large Scale Direct Monocular SLAM) та DSO (Direct Sparse Odometry). В умовах середовищ із повторюваними структурами (коридори, офісні приміщення або склади) ці алгоритми стикаються з проблемами правильної ідентифікації ключових точок через схожість елементів, що призводить до помилок у побудові карт і визначенні місцезнаходження [5]. Динамічні об'єкти (люди, транспортні засоби або рухомі меблі) також суттєво впливають на точність SLAM-систем, оскільки їх переміщення може бути помилково інтерпретовано як зміни в структурі середовища (табл. 2).

Таблиця 1

Методи нейромережевої обробки медичних даних у контексті сумаризації та релевантного висновування

Тип сенсора	Головні характеристики	Переваги	Обмеження
Монокулярна камера	Один об'єктив, 2D-зображення	Простота реалізації, низька вартість	Відсутність глибинної інформації, залежність від текстур
Стереокамера	Два об'єктиви, обчислення глибини	Отримання 3D-даних, покращення точності	Висока обчислювальна вартість, чутливість до калібрування
RGB-D камера	Кольорове зображення та глибинний канал	Висока точність у ближній зоні, доступ до глибинних даних	Обмежений діапазон дії, чутливість до освітлення
IMU	Акселерометр та гіроскоп	Виявлення руху, компенсація дрейфу	Накопичення похибок, необхідність в алгоритмах корекції

Джерело: сформовано автором на основі [1; 4; 5; 6; 7; 14]

Таблиця 2

Ефективність методів SLAM у середовищах із повторюваними структурами та динамічними об'єктами

Метод SLAM	Головний принцип роботи	Стійкість до повторюваних структур	Виявлення динамічних об'єктів	Обмеження
ORB-SLAM	Використання візуальних ключових точок із фіксованими ознаками	Середня	Обмежена	Залежність від статичних об'єктів
LSD-SLAM	Пряме відстеження зміни пікселів	Низька	Відсутня	Складнощі з виявленням руху
DSO	Спарсне одометріювання з використанням інтенсивності пікселів	Висока	Часткова	Складність обробки динамічних сцен
Semantic SLAM	Інтеграція семантичної сегментації	Висока	Висока	Високі обчислювальні витрати
Dynamic SLAM	Виявлення рухомих об'єктів та їхня фільтрація	Висока	Висока	Вимоги до комплексної обробки даних

Джерело: сформовано автором на основі [2; 3; 8; 9; 10; 12]

У середовищах із повторюваними структурами (коридори офісних будівель, склади з однаковими стелажами або паркінги) головною проблемою для алгоритмів одночасної локалізації та картографування є розрізнення об'єктів зі схожими ознаками. Наприклад, у системах, що базуються на таких візуальних ключових точках, як ORB-SLAM, часто спостерігається помилкова ідентифікація місця розташування через невідповідність між фактичною структурою середовища і збереженими орієнтирами. Це може призвести до втрати локалізації чи до побудови некоректної карти. Метод LSD-SLAM, що зорієнтований на аналіз інтенсивності пікселів, демонструє вищу стійкість до повторюваних структур, проте зазнає труднощів із відстеженням об'єктів, які рухаються незалежно від основної сцени [10].

У динамічних середовищах, де присутні рухомі об'єкти, наприклад, на заводах або в торгових центрах, відсутність механізмів виявлення та ігнорування динамічних елементів може стати критичним чинником, що впливає на загальну точність SLAM-систем. У випадках, коли такі об'єкти сприймаються системою як постійні орієнтири, можливі значні помилки в оцінюванні місцезнаходження. Застосування підходів, що використовують відокремлення статичних і динамічних елементів, дає змогу не лише усунути ці помилки, а й підвищити загальну надійність побудованих карт. Dynamic SLAM, зосереджений на виявленні та фільтрації динамічних об'єктів, виявляє ефективність у сценаріях із високою щільністю руху, проте вимагає значних обчислювальних ресурсів і може працювати зі зниженням швидкості в середовищах із великою кількістю рухомих об'єктів.

Алгоритми глибинного навчання суттєво змінили підходи до реалізації SLAM, відкриваючи можливості для покращення точності ідентифікації та класифікації об'єктів у середовищі. Традиційні методи SLAM базуються на аналізі ключових точок і візуальних ознак, однак у динамічних середовищах ці підходи часто виявляються недостатньо ефективними через наявність рухомих об'єктів, які спричиняють помилки в побудові карти [13]. Застосування алгоритмів глибинного навчання дає змогу інтегрувати семантичну інформацію про об'єкти та простір, що сприяє більш точному визначенню контурів об'єктів, їхніх розмірів і категорій.

Семантична сегментація, яка передбачає розподіл сцени на класи об'єктів, є одним із ключових підходів до покращення результатів SLAM у складних середовищах. Використання нейронних мереж для сегментації дає змогу розпізнавати динамічні об'єкти та виключати їх із процесу побудови карти, зосереджуючись на статичних структурах. Наприклад, у середовищі з великою кількістю рухомих об'єктів (транспортні засоби чи люди) семантична сегментація здатна автоматично відокремити ці об'єкти та зменшити кількість помилкових орієнтирів. Такий підхід допомагає зберегти цілісність карти та підвищити точність локалізації без впливу тимчасових змін у сцені.

Іншим важливим напрямом застосування алгоритмів глибинного навчання є фільтрація динамічних об'єктів. Алгоритми, що зорієнтовані

на відокремлення рухомих об'єктів, здатні розпізнавати зміни в сцені, пов'язані з переміщенням об'єктів, та ігнорувати їх під час побудови карти. Це особливо актуально для сценаріїв із високою щільністю руху (торгові зали або аеропорти), де постійний рух може спричинити накопичення помилок у системі SLAM [6]. Використання глибинного навчання для прогнозування траєкторій рухомих об'єктів дає змогу знижувати вплив таких змін на кінцевий результат картографування.

На практиці застосування алгоритмів глибинного навчання в SLAM реалізується через комбіновані архітектури, що інтегрують нейронні мережі для семантичної сегментації та модулі фільтрації динамічних об'єктів. Це забезпечує не лише підвищення точності карти, а й можливість оновлення даних у реальному часі, що є критично важливим для динамічних середовищ.

Реалізація системи SLAM у внутрішніх умовах супроводжується низкою технічних проблем, що впливають на точність і стабільність роботи алгоритмів. Однією з головних проблем є обмежена освітленість, яка суттєво знижує ефективність візуальних сенсорів [4]. У темних або слабо освітлених приміщеннях камери генерують недостатню кількість даних для формування чітких орієнтирів, що призводить до втрати точок відстеження та збільшення кількості помилкових спостережень. Використання технологій активного освітлення, наприклад, інфрачервоних сенсорів, частково розв'язує цю проблему, але збільшує енергоспоживання системи та може спричинити конфлікти у разі наявності декількох пристроїв одночасно.

Другою важливою проблемою є наявність динамічних об'єктів у середовищі, які порушують структуру сцени та ускладнюють процес відстеження. Рухомі об'єкти (люди, транспортні засоби або меблі) можуть бути інтерпретовані системою як статичні орієнтири, що призводить до помилкового визначення траєкторії [9]. Для мінімізації впливу динамічних об'єктів використовують алгоритми семантичної сегментації, які дають змогу ідентифікувати рухомі елементи та виключити їх із процесу картографування. Проте такі алгоритми вимагають значних обчислювальних ресурсів і можуть знижувати швидкість роботи системи в режимі реального часу.

Третьою проблемою є нестабільність даних із сенсорів, що зумовлена зовнішніми факторами та апаратними обмеженнями. Наприклад, IMU зазнають накопичення похибок у процесі вимірювання прискорення та кутової швидкості, що призводить до накопичення систематичних помилок і дрейфу траєкторії [17]. Подібні проблеми спостерігаються і в камерах у разі раптових змін освітлення чи перевантаження обчислювальної системи. Використання алгоритмів корекції даних, зокрема розширеного фільтра EKF, дає змогу зменшити вплив сенсорного шуму, але не усуває повністю проблему втрати орієнтирів у складних середовищах. Таким чином, ефективність реалізації SLAM у внутрішніх умовах визначається здатністю системи до адаптації до змін у середовищі, компенсації динамічних об'єктів і стабілізації даних у режимі реального часу.

Оптимізація SLAM у внутрішніх умовах передбачає інтеграцію алгоритмів глибинного навчання разом із комбінованими сенсорними системами для підвищення точності локалізації. В умовах обмеженої освітленості доцільним є використання камер з інфрачервоним підсвічуванням у поєднанні з інерціальними вимірювальними одиницями. Такий підхід знижує вплив втрати візуальних орієнтирів і забезпечує стабільність локалізації навіть за недостатнього освітлення чи відсутності текстурованих поверхонь.

У середовищах із високою динамікою застосування семантичної сегментації дає змогу відокремлювати рухомі об'єкти від статичних структур, зберігаючи цілісність карти. Використання попередньо навчених нейронних мереж для розпізнавання типових об'єктів, наприклад, людей чи транспортних засобів, мінімізує вплив тимчасових елементів на загальну структуру сцени. Інтеграція алгоритмів прогнозування траєкторій руху допомагає системі заздалегідь визначати потенційні загрози для точності картографування, знижуючи ризик накопичення помилок.

Для стабілізації даних у системах SLAM доцільним є впровадження алгоритмів об'єднання сенсорної інформації, таких як розширений фільтр Калмана чи некалманівський фільтр. Вони дають змогу компенсувати шум у даних та узгодити інформацію з камер, IMU і глибинних сенсорів, що особливо актуально в динамічних середовищах. Застосування багатошарових нейронних мереж для сегментації сцени в режимі реального часу забезпечує коректне відстеження як статичних, так і динамічних об'єктів, що підвищує загальну стійкість SLAM до раптових змін у середовищі. Такий підхід

допомагає зберегти стабільність локалізації та підвищити точність картографування навіть за умов нестабільних сенсорних даних.

Висновки. Ефективність SLAM у внутрішніх умовах без глобальної навігації значною мірою залежить від типу сенсорів, умов освітленості та динаміки середовища. Встановлено, що інтеграція алгоритмів глибинного навчання підвищує точність сегментації динамічних об'єктів і зменшує похибки картографування. Комбіновані сенсорні системи, що охоплюють монокулярні, стерео- і RGB-D камери, а також інерціальні вимірювальні одиниці, забезпечують вищу стабільність локалізації за рахунок компенсації похибок. Головними проблемами є зниження якості візуальних даних у разі слабкого освітлення, нестабільність сенсорних показників і помилкова ідентифікація рухомих об'єктів як статичних. Застосування традиційних методів SLAM у таких умовах призводить до накопичення похибок і втрати локалізації. Рекомендовано впроваджувати гібридні SLAM-системи з інтеграцією глибинного навчання для прогнозування траєкторій рухомих об'єктів і стабілізації зображення в умовах обмеженого освітлення. Перспективи майбутніх досліджень полягають у розробленні адаптивних архітектур SLAM, здатних інтегрувати різнотипні сенсорні дані та автоматично коригувати системні похибки в реальному часі.

Список літератури:

1. Merzlyakov A., Macenski S. A Comparison of Modern General-Purpose Visual SLAM Approaches. *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2021. P. 9190–9197. DOI: https://doi.org/10.1109/IROS51168.2021.9636615

2. Hashemifar Z. S., Adhivarahan C., Balakrishnan A., Dantu K. Augmenting visual SLAM with Wi-Fi sensing for indoor applications. *Autonomous Robots*. 2019. Vol. 43, № 6. P. 2245–2260. DOI: https://doi.org/10.1007/s10514-019-09874-z

3. Xu C., Liu Z., Li Z. Robust visual-inertial navigation system for low precision sensors under indoor and outdoor environments. *Remote Sensing*. 2021. Vol. 13, № 4. P. 772. DOI: https://doi.org/10.3390/rs13040772

4. Chen C., Zhu H., Wang L., Liu Y. A Stereo Visual-Inertial SLAM Approach for Indoor Mobile Robots in Unknown Environments Without Occlusions. *IEEE Access*. 2019. Vol. 7. P. 185408–185421. DOI: https://doi.org/10.1109/ACCESS.2019.2961266

5. Helmberger M., Morin K., Berner B., Kumar N., Cioffi G., Scaramuzza D. The Hilti SLAM Challenge Dataset. *IEEE Robotics and Automation Letters*. 2022. Vol. 7, № 3. P. 7518–7525. DOI: https://doi.org/10.1109/LRA.2022.3183759 (date of access: 17.05.2025).

6. Damjanović D., Biočić P., Prakljačić S., Činčurak D., Balen J. A comprehensive survey on SLAM and machine learning approaches for indoor autonomous navigation of mobile robots. *Machine Vision and Applications*. 2025. Vol. 36, № 1. Article 55. DOI: https://doi.org/10.1007/s00138-025-01673-0

7. Czarnowski J., Laidlow T., Clark R., Davison A. J. DeepFactors: Real-Time Probabilistic Dense Monocular SLAM. *IEEE Robotics and Automation Letters*. 2020. Vol. 5, № 2. P. 721–728. DOI: https://doi.org/10.1109/LRA.2020.2965415

8. Bescos B., Campos C., Tardós J. D., Neira J. DynaSLAM II: Tightly-Coupled Multi-Object Tracking and SLAM. *IEEE Robotics and Automation Letters*. 2021. Vol. 6, № 3. P. 5191–5198. DOI: https://doi.org/10.1109/LRA.2021.3068640

9. Singh K. J., Kapoor D. S., Thakur K., Sharma A., Nayyar A., Mahajan S. Map Making in Social Indoor Environment Through Robot Navigation Using Active SLAM. *IEEE Access*. 2022. Vol. 10. P. 134455–134465. DOI: https://doi.org/10.1109/ACCESS.2022.3230989

10. Bamdad M., Scaramuzza D., Darvishy A. SLAM for Visually Impaired People: A Survey. *IEEE Access*. 2024. Vol. 12. P. 130165–130211. DOI: https://doi.org/10.1109/ACCESS.2024.3454571

11. Kandalan R. N., Namuduri K. Techniques for Constructing Indoor Navigation Systems for the Visually Impaired: A Review. *IEEE Transactions on Human-Machine Systems*. 2020. Vol. 50, № 6. P. 492–506. DOI: https://doi.org/10.1109/THMS.2020.3016051

12. Khan D., Cheng Z., Uchiyama H., Ali S., Asshad M., Kiyokawa K. Recent advances in vision-based indoor navigation: A systematic literature review. *Computers & Graphics*. 2022. Vol. 104. P. 24–45. DOI: https://doi.org/10.1016/j.cag.2022.03.005

13. Tomažič S. Indoor positioning and navigation. *Sensors*. 2021. Vol. 21, № 14. P. 4793. DOI: https://doi.org/10.3390/s21144793

14. Zhang Y., Hsiao M., Zhao Y., Dong J., Engel J. J. Distributed Client-Server Optimization for SLAM with Limited On-Device Resources. *2021 IEEE International Conference on Robotics and Automation (ICRA)*. 2021. P. 5336–5342. DOI: https://doi.org/10.1109/ICRA48506.2021.9561638

15. Tseng P. Y., Lin J. J., Chan Y. C., Chen A. Y. Real-time indoor localization with visual SLAM for in-building emergency response. *Automation in Construction*. 2022. Vol. 140. Article 104319. DOI: https://doi.org/10.1016/j.autcon.2022.104319

16. Zhang H., Ye C. A Visual Positioning System for Indoor Blind Navigation. *2020 IEEE International Conference on Robotics and Automation (ICRA)*. 2020. P. 9079–9085. DOI: https://doi.org/10.1109/ICRA40945.2020.9196782

17. Sahili A. R., Hassan S., Sakhrieh S. M., Mounsef J., Maalouf N., Arain B. A Survey of Visual SLAM Methods. *IEEE Access*. 2023. Vol. 11. P. 139643–139677. DOI: https://doi.org/10.1109/ACCESS.2023.3341489

Usov O.S. VISUAL SLAM METHODS FOR NAVIGATION IN INDOOR ENVIRONMENTS WITHOUT ACCESS TO GLOBAL POSITIONING

The relevance of the research is determined by the need to improve the accuracy of simultaneous localization and mapping (SLAM) in indoor environments without access to global navigation, which is critical for ensuring stable navigation in environments with complex structures and dynamic objects.

The aim of the study is to conduct a comprehensive analysis of methods for implementing visual simultaneous localization and mapping in indoor environments, with an emphasis on identifying optimal approaches for improving localization accuracy and creating effective mapping algorithms.

The research methodology includes a comparative analysis of existing SLAM methods, a structural and functional analysis of sensor systems (monocular, stereo, and RGB-D cameras, as well as inertial measurement units), as well as modeling the impact of deep learning algorithms on the segmentation of dynamic objects and reducing errors in the construction of spatial maps.

Research results. It has been established that the combined use of different types of sensors makes it possible to reduce localization errors by mutually compensating for the shortcomings of each system. It has been proven that the integration of deep learning algorithms improves the accuracy of dynamic object identification, which helps reduce mapping errors in environments with complex architecture. It has been found that SLAM systems focused on semantic segmentation provide more stable localization in environments with repetitive structures, but require significant computing resources.

Conclusions. The feasibility of using hybrid SLAM systems that combine different types of sensors and deep learning algorithms to compensate for errors and stabilize localization in dynamic environments has been justified. The implementation of infrared illumination and trajectory prediction technologies for moving objects is recommended to reduce the influence of dynamic factors.

Prospects for further research are related to the development of adaptive SLAM architectures with automatic adjustment of sensor parameters and algorithms to ensure stable localization and mapping in environments with complex structures and high dynamics.

Key words: *navigation systems, dynamic objects, semantic segmentation, neural networks, localization, mapping, sensors.*

UDC 004.13.8
DOI https://doi.org/10.32782/2663-5941/2025.3.2/55

Ushakova I.O.
Simon Kuznets Kharkiv National University of Economics

Bondarenko D.O.
Simon Kuznets Kharkiv National University of Economics

Chirva Yu.Ye.
Simon Kuznets Kharkiv National University of Economics

Znakhur L.V.
Simon Kuznets Kharkiv National University of Economics

COMPUTER MODELING OF COMPANY EMPLOYEE CHURN USING MACHINE LEARNING AND PREDICTIVE ANALYTICS METHODS

The study proposes an approach to management employee churn using machine learning, predictive analytics methods and with the support of information technologies. It systematically explores the theoretical underpinnings of employee turnover, categorizing it into various types and identifying key components for measurement, as well as factors influencing churn rates and their potential regulation. The discourse highlights the importance of employee retention strategies, emphasizing elements such as compensation, work planning, performance evaluations, training programs, and opportunities for career advancement. The authors propose utilizing decision tree and logistic regression methodologies to predict employee churn, selecting a binary classification criterion that distinguishes between employees who remain with the company and those who depart. Two predictive models are developed, showcasing significant accuracy metrics: the decision tree model demonstrates an impressive 91.3% accuracy on training data and 74.19% on test data, while the logistic regression model indicates 88.41% accuracy on training data and 90.32% on test data. These findings underscore the reliability of the proposed models in forecasting employee turnover. Furthermore, the article outlines practical applications for these models, providing actionable insights for organizations aiming to implement data-driven strategies for improving employee retention. By leveraging advanced analytical techniques, the study contributes valuable methodologies and frameworks for companies grappling with the challenges of employee churn, positioning itself as a critical resource for human resource professionals and organizational leaders seeking to enhance workforce stability and performance. Overall, the research underscores the potential of predictive analytics in informing strategic decisions regarding employee retention and maintaining a healthy climate in the company.

Key words: *computer modeling, machine learning, logistic regression, decision tree, employee churn, predictive analytics.*

Statement of the problem. One of the main tasks of modern companies is the search and selection of personnel, their retention in the company. HR departments of companies in the current conditions have to respond quickly to all changes in the specialist market and constantly process a large amount of information. In these conditions, conducting scientific research on computer simulation of employee churn is an extremely urgent task. Determining staff churn allows the management of an company to determine which part of the staff remains loyal to the company, and which part is able to go to competitors, thereby reducing profits. The company's profit depends on the quality and ease of forecasting the churn of personnel.

Timely forecasting of the churn of employee is the key to the successful and effective work of HR managers in the field of search, selection and retention of personnel, and therefore to the effective work of the company. In today's world, there are many mathematical methods and software tools that can be used to build predictive models that allow timely forecasting of staff churn and simplify business processes.

Analysis of recent research and publications. The problem of modeling the predictive analytics of the churn of company employee in economic and mathematical science was raised by a number of researchers. For example, scientific articles [1, 2] propose a method of substantiating the innovative

strategy of a company and analyzing employee management mechanisms in the context of innovative business activities.

Researches [3–5] offer methods for optimizing employee management of companies to improve control processes and evaluate the effectiveness of web technologies. The issue of optimizing the interface of predictive analytics information systems is considered in scientific articles [6, 7].

Predictive analytics is also important in the analysis of employee churn. Predictive models based on classification and machine learning methods allow you to determine which class an employee belongs to: will be fired or will continue to work. This allows companies to provide decision support for creating sound and effective HR strategies [8–11]. Various methods are used to create predictive models, such as Logistic Regression, Decision Trees, Decision List Algorithm etc. [12–15].

In order to take competitive positions in the market, companies are increasingly using predictive analytics and information technology in HR management it in their activities. Innovative methods and technologies help companies identify factors and trends in employee churn in advance and develop appropriate measures to reduce employee churn. Therefore, there is a need to develop a methodology for predicting the churn of company employee using predictive analytical models, based on a reasonable set of factors affecting the outflow.

Task statement. The purpose of this article is to substantiate the technology of using mathematical methods and computer modeling tools for building and analyzing predictive models of employee churn.

When analyzing employee churn, it is important to assess which employee leave and which remain in the organization, whether the trend of changing the quality of employee coincides with the strategic settings defined by the management. Without this assessment, it is impossible to understand whether the existing level of variability is a positive or negative phenomenon. If it is precisely those employee who need to be got rid of, it means that the organization is on the right track. If it loses the best employee, then the issue of churn must be quickly resolved. Also, the situation when the changeability and rotation of personnel, both large and small, does not contribute to improving the quality of work is also not encouraging.

Outline of the main material of the study. It should be noted that the statement that the lower the churn rate, the better is false. It is not true. Extremely low churn rates indicate that the employee of the organization is not updated much and even ineffective employees do not leave it. The main issue of any company is the provision of uninterrupted work, avoiding problems related to the unexpected dismissal of employee. In modern theories of employee churn management, there are methods that allow you to calculate with a certain percentage of probability how much a person will work. They make it possible to determine this already in the process of negotiations and interviews, as well as based on the analysis of information about the dates of admission and dismissal in the candidate's employment book. It is useful to process them on a computer and get the result in graphic form. The degree of loyalty of the candidate to the previous employers and the tendency to change the attitude immediately becomes obvious. The HR manager can ask to comment on the graphs and observe the sincerity and logic of the answers, as well as the applicant's attitude, constructiveness, intellectual abilities and communication style. According to the form of manifestation, physical and hidden employee churn should be distinguished. Physical churn includes those employees who, due to various reasons, are fired and leave the organization. Hidden (psychological) staff churn occurs among those employees who outwardly do not leave the organization, but actually leave it, disengage from organizational activities.

The next division of employee churn is churn in relation to the organization – intra-organizational and inter-organizational churn. Internal organizational churn is related to labor movements within the organization, and external churn is between organizations, industries and spheres of the economy.

Depending on the size, liquidity is divided into natural and excess. Natural churn (3–5% per year) contributes to the timely renewal of the team and does not require special measures on the part of the management and employee service. High levels of staff churn lead to significant economic losses. Companies get organizational, personnel, technological and psychological difficulties as a result.

Almost all large organizations measure employee churn. At the same time, you should pay attention to the following components:

the general level of churn and its comparison with churn in other companies on the market (benchmarking);

dynamics of churn rate over time;

changing the structure of reasons for churn;

change in churn of different categories of personnel.

Companies participate in specially organized studies to obtain benchmarks or churn rates in the market or industry. In the process of research, employee-initiated churn and employer-initiated churn are often measured separately. And also the churn of new employee: during the first 3 months or during the first year of work.

To analyze employee churn, first of all, it is necessary to consider the factors affecting employee churn and their regulation. These include the following factors:

material (uncompetitive pay rates, unfair pay structure, unstable earnings);

organizational (schedule, regime, working conditions do not meet the employee's expectations, lack of opportunity for career growth, professional development);

interpersonal (relationships with management and colleagues);

the employee's age (the most risky age of transition to another job is under 25);

employee qualification (the lower the qualification, the more often they change jobs);

the employee's place of residence (the further the employee lives from work, the greater the risk of his leaving);

length of service at the enterprise (after three years of experience there is a sharp decrease in churn, which is explained both by the age factor and adaptation problems).

Retention measures should include each direction where adherence problems and dissatisfaction may appear such as:

payment;

work planning;

assessment of activity results;

training;

career development;

commitment and engagement;

team feeling;

conflicts.

To reduce staff churn, leading companies are now widely using predictive analytics, which is based on statistical methods, methods of intelligent data analysis, game theory, analysis of current and historical facts to make predictions about future events. Predictive analytics relies heavily on advanced information technologies for intelligent data analysis.

Logistic regression and decision trees are the most popular methods for predicting employees churn. The features of their possible application are analyzed below Logistic regression allows you to determine the predicted outcome for two conditions, using several factors. To determine employees churn, you need to divide them into two groups: employees either leave or stay in the company.

Decision trees are also an important tool in the work of every specialist involved in data analysis. This method has several algorithms for tree construction and is widely used to analyze staff churn. A decision tree allows you to build rules based on tracking with a certain probability all paths from the root node to the leaves of the tree. Both methods have certain advantages and disadvantages, therefore, the third section will implement models based on these methods with further comparative analysis of their obtained results.

Currently, the methods of predictive analytics are quite popular for solving the problem of employee churn. Predictive analytics is a data analysis method. The main task of these methods is to determine the state or direction of behavior of any entities in the future. The results obtained are used to justify management decisions [1–2, 6–7, 9, 14]. For forecasting, the following methods of data mining, statistical analysis, game theory, etc. can be used. Predictive analytics compares the current situation with historical data analytics, which makes it possible to determine predictions about future events. Thus, predictive analytics models in real business conditions determine the relationships between several indicators that influence decision-making, and allow assessing the risks associated with a certain set of factors. Today, predictive analytics relies heavily on advanced machine learning technologies. Data scientists use deep learning and complex algorithms to analyze multiple variables to create predictive models that can predict typical behavior based on big data. Next, we will consider methods that are widely used for predictive modeling, such as logistic regression and decision trees.

1. Logistic regression.

Let's define the features of using the logistic regression method to predict employees churn. [14]. This technique is one way to answer questions like "why do people leave a company?", "how can we influence turnover rates?", and "who is most likely to leave in the next year?" The goal of any type of regression is to predict an outcome using some factors. To predict churn (the dependent variable) employee data (the independent variables) is use. Logistic regression, simply put, is a form of regression that is used when the predicted outcome is 1 or 0. This is exactly what is needed to pre-

dict turnover: people either leave or stay with the company.

Logistic regression determines the probability of an event occurring based on the set of attribute values *y*. For this purpose, the so-called dependent variable y is introduced, which can take only one of two values. As a rule, these are the numbers 0 (event did not happen) and 1 (event did happen), and a set of independent variables (also called attributes, predictors or regressors) that are real values $x_1, x_2, \ldots, x_n$. They are based on the values needed to calculate the probability of assuming a certain value of the dependent variable. Let's introduce a fictitious attribute $x_0 = 1$ to simplify the writing later.

It is assumed that the probability of the event occurring *y* = 1:

$$P\{y = 1|x\} = f(z),$$

where $z = \theta^T x = \theta_0 + \theta_1 x_1 + \ldots + \theta_n x_n$, where x and θ are column vectors of values of the independent variable $1, x_1, ..., x_n$ and parameters (regression coefficients) are real numbers θ_0, ..., , and $f(z)$ is the so-called logistic function (sometimes also called sigmoid or logit function):

$$f(z) = \frac{1}{1 + e^{-z}}.$$

Since *y* takes only the values 0 and 1, the probability of taking the value 0 is:

$$P\{y = 0|x\} = 1 - f(z) = 1 - f(\theta^T x).$$

In general, the distribution function *y* for a given *x* can be written as:

$$P\{y|x\} = f(\theta^T x)^y \left(1 - f(\theta^T x)\right)^{1-y}, y \in \{0,1\}.$$

2. Decision tree.

Decision tree methods are a family of algorithms based on creating a hierarchical structure based on the answer to a set of "Yes" or "No" questions. The create of decision tree methods for data analysis is associated with the C4.5 and C5.0 algorithms, the CART (Classification and Regression Trees) algorithm and the random forest method [14].

The basis of most popular decision tree training algorithms is the principle of data partitioning. This principle is implemented by the following algorithm. Consider a training set S containing n examples, each of which is given a class label $Ci(i = 1..k)$ and n attributes $Aj(j = 1..m)$, which are supposed to determine whether an object belongs to a particular class. Then three cases are possible:

1. All examples of the set S have the same class label *Ci*, that is, all training examples belong to only one class. Obviously, training in this case does not make sense, since all examples the presented models will be of the same class, which will "learn" to recognize the model. The decision tree itself in this case will be a leaf associated with the class *Ci*. The practical use of such a tree is meaningless, since it will attribute any new object only to this class.

2. The set S does not contain any examples at all, i.e. is an empty set. In this case, a leaf will also be created for it (applying the rule to create a node to an empty set is meaningless), the class of which will be selected from another set (for example, the class that occurs most often in the parent set).

3. The set S contains training examples of all classes Ck. In this case, it is necessary to divide the set S into subsets associated with classes. To do this, one of the attributes Aj of the set S is selected, which contains two or more unique values $(a1, a2, ..., ap)$, where p is the number of unique values of the attribute. Then the set S is partitioned into p subsets $(S1, S2, ..., Sp)$, each of which includes examples containing the corresponding attribute value. Then the next attribute is selected and the partitioning is repeated. This procedure is repeated recursively until all examples in the resulting subsets are of the same class.

In the process of constructing a decision tree, it is necessary to select a partitioning attribute. The best partition attribute will be the one that provides the maximum information gain of the resulting node relative to the original:

$$Gain(A) = Info(S) - Info(S_A),$$

where $Info(S)$ is the information associated with the subset S before partitioning,

$Info(S_A)$ is the information associated with the subset obtained by partitioning by attribute A.

Thus, the task of choosing a partition attribute in a node is to maximize the value of $Gain(A)$, which is called the information gain.

Experimental part. For the practical implementation of predictive methods (the decisions tree and logistic regression) we will use the modern package of IBM SPSS Modeler application programs [9]. Modeling employee churn is a description and analysis to predict the category to which a particular employee belongs. In our case, the simulation of employee churn is based on signs of churn: work (True), left (False). We estimate churn using the constructed flow and the company's personnel database. First, we estimate the importance of the predictor, our model determined that the most important for forecasting is work experience at the start of cooperation, the value of which = 0.15 (Fig. 3).

After completing the construction of the flow, the final result was the model of building a decision tree, which is presented on the Fig. 2.

Using a decision tree and analyzing data on an individual worker, it is possible to simulate the situations under which he will stay in the company, and under which he will leave the company.

The CART algorithm has two children at each node of the decision tree. A rule is formed at each node. It allows you to divide the set of examples of a node (the training set) into two parts. The first part contains examples in which the rule is fulfilled (usually the child is on the right), and the second part in which the rule is not fulfilled (the child is on the left). The partition quality evaluation function is used to determine the optimal rules. For example, based on decision trees, it can be assumed that workers who have had overtime, the number of companies they have worked for >2, will leave the company with a probability of 100%, instead, workers, who worked for our company in the same company, or had no work experience at all, will remain. This may be due to the fact that a worker who already has experience working in another company/companies without overtime with a standardized work schedule, while working in a new com-

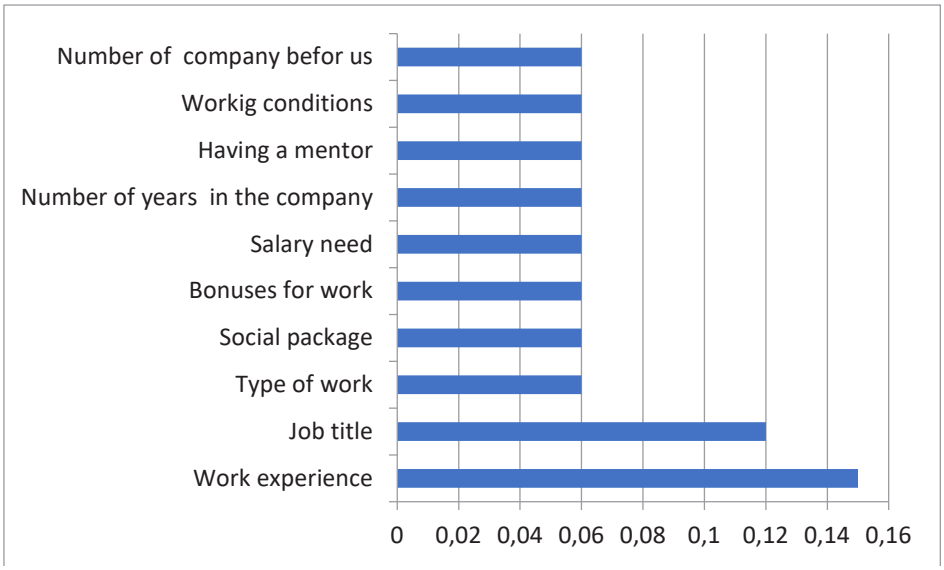


Fig. 1. Estimate the importance of the predictor

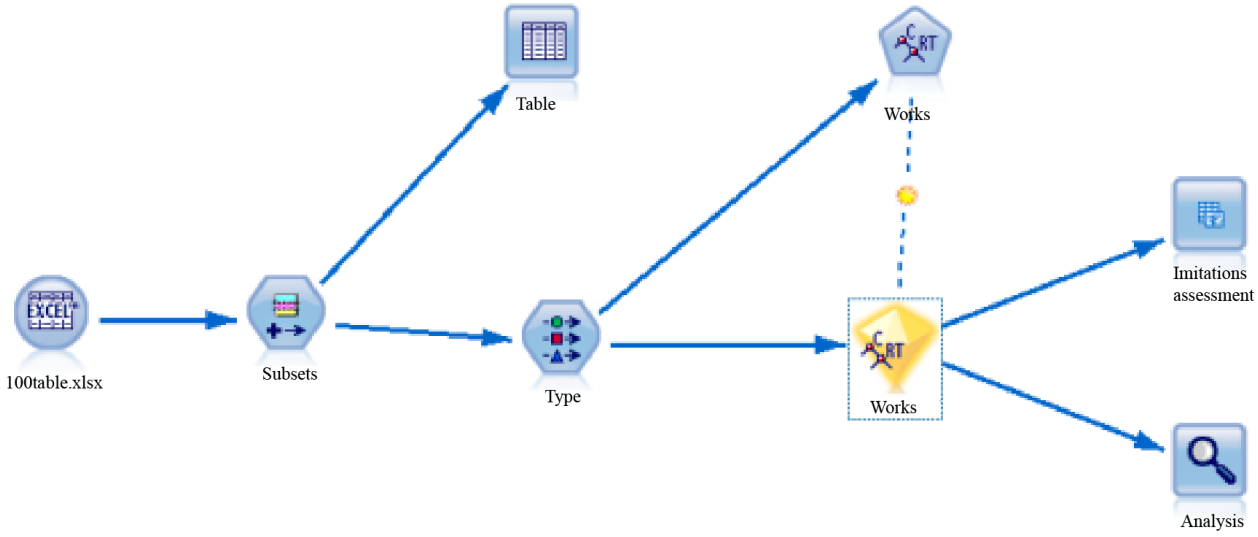


Fig. 2. Decision tree construction flow

pany with overtime, can compare the conditions and change the place of work.

Let's consider another branch of our tree. An employee who did not have overtime may leave because he does not have a social package, and for him this is more important than anything.

Let's also consider the option of a worker who does not have overtime. According to the analyzed data, a person who has work experience <=2.8 years with a 100% probability will remain working for the company. An employee who has work experience in the company >2.8 years, work experience in our company <3.7 years is highly likely to leave the company.

Therefore, it can be concluded that the main factors that affect the churn of employee are: the presence of overtime, the position they hold, the social package, work experience and the number of companies in which the employee worked before starting work in our company.

The goal of the whole process of building a decision tree is to create a model that can be used to classify cases and decide what values the objective function can take, given several variables as inputs. Thus, the classification of each new case occurs when moving down to the letter, which will tell us the value of the objective function in each specific case.

In our case, when it is necessary to make a decision about the worker (the objective function can take the values "working" and "left") based on information about the worker (for example: work experience at the time of the start of cooperation, availability of overtime, marital status, and so on).

To build a model using logistic regression, we perform the same settings as for the decision tree, namely adding nodes: source, subset, table and types. As a result of construction, we get a logistic regression model, which is presented on the Fig. 3.

After running the logistic regression model, a classification table was obtained, which shows how correctly the built model predicts (Table 3). A total of 100 records were processed. The percentage of correct predictions for employees with the attribute "work" is 94.2%, and among those who "left" – 83.3%. The average percentage is 89.0%.

Table 3

Classification table

Sign	Predicted		
	Works	Left	Percentage of correct
Works	49	3	94,2
Left	8	40	83,3
The overall percentage			89,0

For employed persons, 49 correct predictions were obtained, and 40 correct predictions for persons who are not employed. When training the model, 8 types of false classifications for those who are employed, and 3 false negatives for those who are not employed, must be omitted.

Next, we will analyze the coefficients of the so-called chances, which show the statistical

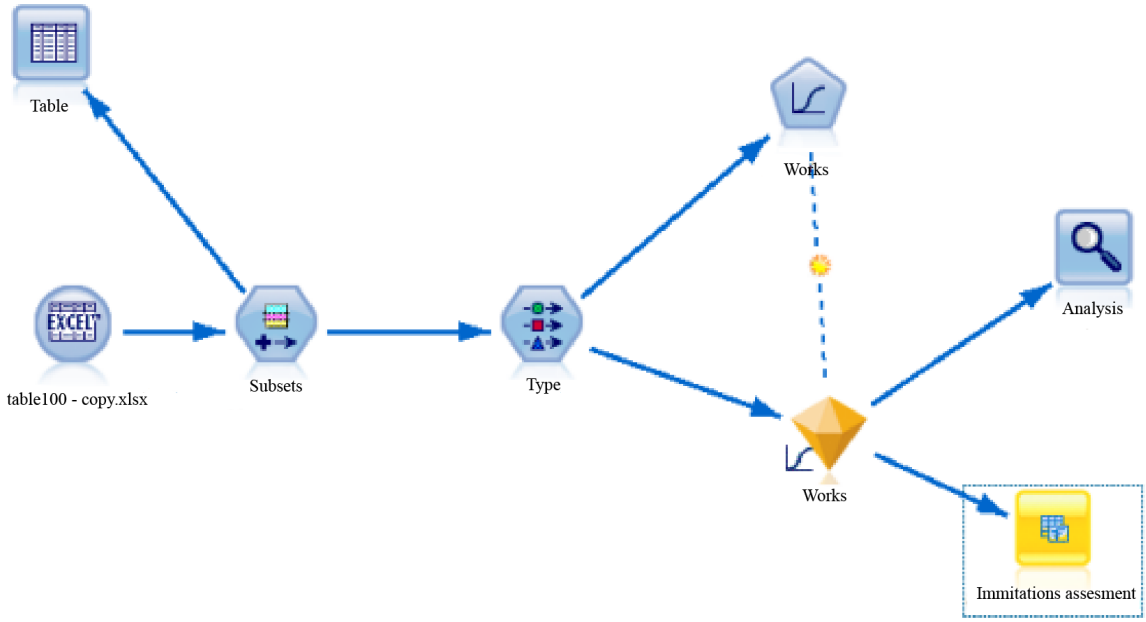


Fig. 3. Logistic regression model

influence of variables on the churn of employee (Fig. 4). As we can see, these are the variables included in the model. In this table, it is worth paying attention to the evaluations of the variables. When we build a model based on these estimates, these are the odds ratios. The table shows the confidence interval for the odds ratios. Indicators in column B are the actual beta coefficients on which the model is built. Odds ratios determine the influence of variables, the higher this ratio (greater than 1), the more likely this variable will affect the dismissal of the worker or his further work in the company.

Analysis of the simulation results shows that the significance level for the model is < 0.05, which means that we have a significant model and will predict well.

The analysis of the simulation results also shows how many variants of the dependent variable are explained by the predictors. In the built model, this is about 73% of the result, or 73 percent of the options in the result will be predicted by our predictors (or affected by our predictor). Thus, it can be concluded that a significant percentage of the results will be predicted by the previously selected predictors.

Therefore, it can be concluded that the result was correctly predicted by the model for 89%. This is much better than the null hypothesis before model training, as it was only 52%.

Comparing the obtained results and works based on the comparison tables, where the number and percentage of correct and incorrect, trained and tested stream entries are presented, the decision tree model showed results of 91.3% and 74.19% correct for trained and tested, respectively, and the logistic regression model – 88.41% and 90.32%. Thus for the decision tree there were better results for the trained set, and for regression the tested set.

Also evaluating the performance by the indicators trained in the decision tree 0.514 and 0.698 respectively (True, False) and logistic regression 0.463 and 0.691 respectively. According to the indicators of tested decision trees 0.458 and 0.333 respectively (True, False) and logistic regression 0.601 and 0.601 respectively.

To obtain the general results of the simulation for two models, it is necessary to run the model analysis node for execution. This node makes it possible to assess the adequacy of the constructed model. After its execution, the system displays the results of the model evaluation for the decision tree (Fig. 5).

Let us consider the values of the model evaluation indicators, namely the values of the AUC indicator and the Gina coefficient. AUC is area under curve ROC (receiver operating characteristic) and the Gina coefficient is quantitative indicator that shows the degree of inequality of the distribution options. The values of these indicators are always

Variables in the equation

	B	Root mean squared error	Wald	d.f.	Value	Exp (B)
Work experience by the employment(in years)	,948	,359	6,973	1	,948	2,580
How many companies have worked before us	-,074	,471	0,025	1	-,074	,929
Years in company	,597	,427	1,953	1	,597	1,816
Education	6,178	3,530	3,063	1	6,178	481,843
Social package	-26,258	40193,018	,000	1	-26,258	,000
Working conditions	-502	724	482	1	-502	605

Fig. 4. Statistical characteristics of model variables

'Subsets'	1_Training		2_Testing	
Model	AUC	Jeanie	AUC	Jeanie
$R-Works	0,951	0,902	0,817	0,634

Fig. 5. Results of decision tree model analysis

'Subsets'	1_Training		2_Testing	
Model	AUC	Jeanie	AUC	Jeanie
$L-Works		0,881	0,908	0,815

Fig. 6. The results of the analysis by logistic regression

between 1 and 0, the larger the value, the better the classifier.

Next, we perform regression analysis. The results of the logistic regression analysis are presented in Fig. 6.

The AUC and Gina scores are almost the same, but the decision tree is several hundredths better in terms of AUC than logistic regression with 0.951 and 0.817, and logistic regression with 0.941 and 0.815. The Gina coefficient also shows that the decision tree is better, 0.902 and 0.634, respectively, and the logistic regression is 0.881 and 0.815, respectively.

Thus, both models showed satisfactory results. But according to the AUC and Gina indicators, the preference can be given to the decision tree.

Conclusions. Thus, the study substantiated the approach to employee turnover management using predictive analytics methods. The theoretical foundations of employee turnover in a company are the types of employee turnover, such as natural turnover and intra-organizational turnover. An important element of employee turnover management are the components of its measurement, namely, the overall level of turnover and its comparison with turnover in other companies on the market (benchmarking); the dynamics of the turnover coefficient over time; changes in the structure of the causes of turnover; changes in the turnover of different categories of personnel. In addition, it is necessary to take into account the factors that influence employee turnover and their regulation: material, organizational, interpersonal.

For the practical implementation of employee turnover management, the use of mathematical methods and information technologies was proposed, which allowed building and analyzing predictive models of employee turnover.

Before modeling, the choice of indicators that influence employee turnover was justified. Next, flows were constructed for modeling the outflow using the decision tree and logistic regression methods. The experimental results showed that the accuracy of the predictive model by both methods is quite high. Having evaluated the data obtained on the analysis of the constructed models by the decision tree and logistic regression methods, a conclusion was made about the adequacy of both models.

Comparing the results of the work, which presents the number and percentage of correct and incorrect trained and tested records in the flow, the decision tree model showed results of 91.3% and 74.19% correct for trained and tested, respectively, and the logistic regression model – 88.41% and 90.32%. These results indicate a fairly high accuracy of the forecast

The scientific significance of the provisions and recommendations developed in the article is that they allow solving the scientific and practical task of developing methodological foundations and practical recommendations for the use of computer modeling methods in predicting the outflow of enterprise personnel.

The validity of the scientific results of the study is determined by a multi-faceted consideration of the problem, research and experimental confirmation of the adequacy of the developed predictive model.

The practical significance of the results obtained lies in the development of recommendations for the construction and analysis of models for predicting employee turnover for the enterprise.

Bibliography:

1. Catlett F. E. Management Strategies to Improve Employee Retention and Organizational Performance. Dissertations of Doctor of Business Administration Walden*:* 2024, 136 p. URL: https://scholarworks.waldenu.edu/dissertations/15867

2. Paruzel A., Schmidt L., Maier G. W. Corporate social responsibility and employee innovative behaviors: A meta-analysis*. Journal of Cleaner Production.* 2023. № 393. URL: https://doi.org/10.1016/j.jclepro.2023.136189

3. Starkova O., Bondarenko D., Hrabovskyi Y. Providing software support for economic analysis. *Technology Audit and Production Reserves.* 2023. № 5/2 (73). P. 34–39. URL: http://repository.hneu.edu.ua/handle/123456789/30456

4. Ushakova I., Skorin Y., Shcherbakov A. Methods of quality assurance of software development based on a systems approach. *III International Scientific And Practical Conference "Information*

Security And Information Technologies", September 13–19, 2021, Odesa, Ukraine. P. 158–168. URL: https://ceur-ws.org/Vol-3200/paper22.pdf

5. Ushakova I., Hrabovskyi Ye. Methodology for developing an information site with Workflow support for publishing articles. *Development Management.* 2022. Vol. 21. No. 3. P. 20–28. URL: https://doi.org/10.57111/devt.20(3).2022.20-28.

6. Hrabovskyi Ye., Bondarenko D., Kobzev I. Improving the technology for constructing a software tool to determine the similarity of raster graphic images. *Eastern-European Journal of Enterprise Technologies.* 2024. № 1/2(127). P. 16–35. URL: https://journals.uran.ua/eejet/article/view/298744/291778

7. Hrabovskyi Y., Bondarenko D., Starkova O. Development of a methodology for creating a website for online healthcare services. *EUREKA: Physics and Engineering.* 2024. № 5. P. 183–200. URL: https://journal.eu-jr.eu/engineering/article/view/3440/2496.

8. Erden S. B., Erişen M., Sarıgül Y. N., Eken B., Özcan T. HR Analytics in Retail: Predicting Employee Churn with Machine Learning. *Intelligent Industrial Informatics and Efficient Networks Proceedings of the INFUS 2024 Conference.* 2024. Vol. 3. P. 109–116. URL: https://link.springer.com/book/10.1007/978-3-031-67192-0

9. Meddeb E. The Human Resource Management challenge of predicting employee turnover using machine learning and system dynamics. *CEUR-WS.* 2021. Vol. 2991. P. 184–196. URL: https://ceur-ws.org/Vol-2991/paper16.pdf

10. Jiaxuan Zh., Zinan C., Haichen Q., Meng W. Management and prediction of employee turnover in enterprises based on big data analytics and machine learning. *Proceedings of the 2nd International Conference on Software Engineering and Machine Learning.* 2024. P. 103–108. URL: http://dx.doi.org/10.54254/2755-2721/76/20240573

11. Ushakova I., Hrabovskyi Ye. Modeling and Selection of a Distance Learning System for a Higher Education Institution Based on the Method of Hierarchy Analysis Using the DSS. Вчені записки ТНУ імені *В.І.* Вернадського. 2023. №. 34 (73). P. 246–254. URL: https://doi.org/10.32782/2663-5941/2023.2.1/39

12. Bandyopadhyay N., Jadhav A. Churn Prediction of Employees Using Machine Learning Techniques. *TEHNIČKI GLASNIK*. 2021. Vol. 15. № 1. P. 51–59. URL: https://doi.org/10.31803/tg-20210204181812

13. Ushakova I. Simulation of the impact social media on promoting education services. Proceedings of the 12th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Kyiv, Ukraine, June 21–24, 2016. Vol. 1614. P. 575–590. URL: https://ceur-ws.org/Vol-1614/paper_60.pdf

14. Ushakova I., Dorohov O., Dorohova L., Malyarets L. Customer churn predictive modeling by classification method. *Bulletin of the Transilvania University of Brasov. Series III: Mathematics, Informatics, Physics.* 2020. Vol. 13 (62). № 1. P. 347–362. URL: https://doi.org/10.31926/but.mif.2020.13.62.1.26

15. Ushakova I., Hrabovskyi Ye., Szymczyk K. Analytical Decision List Algorithm for Managing Customer Reaction to a Marketing Campaign. *Polish Journal of Management Studies.* 2023. Vol. 28. № 1. P. 360–379. URL: https://pjms.zim.pcz.pl/article/541885/en

Ушакова І.О., Бондаренко Д.О., Чирва Ю.Є., Знахур Л.В. КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ ВІДТОКУ СПІВРОБІТНИКІВ КОМПАНІЇ ЗА ДОПОМОГОЮ МЕТОДІВ МАШИННОГО НАВЧАННЯ ТА ПРОГНОЗНОЇ АНАЛІТИКИ

У дослідженні запропоновано підхід до управління відтоком співробітників за допомогою методів машинного навчання, прогнозної аналітики та за підтримки інформаційних технологій. В рамках досліджування було визначено теоретичні основи плинності кадрів, проведено її класифікацію за різними типами та визначено ключові компоненти для її вимірювання. Крім того були обґрунтовані основні фактори, що впливають на показники відтоку кадрів та їх потенційне регулювання. Дискурс підкреслює важливість стратегій утримання співробітників, наголошуючи на таких елементах, як винагорода, планування роботи, оцінка ефективності, програми навчання та можливості кар'єрного зростання. Автори пропонують використовувати методології дерева рішень і логістичної регресії для прогнозування відтоку співробітників, вибираючи двійковий критерій класифікації, який розрізняє співробітників, які залишаються в компанії, і тих, хто йде. Розроблено дві прогностичні моделі, які демонструють значні показники точності: модель дерева рішень демонструє вражаючу точність 91,3% щодо даних навчання та 74,19% даних тестування, тоді як модель логістичної регресії вказує на точність 88,41% даних навчання та 90,32% даних тестування. Отримані результати підкреслюють надійність запропонованих моделей у прогнозуванні відтоку кадрів.

Крім того, у статті описано практичне застосування цих моделей, що надає важливу інформацію для організацій, які прагнуть реалізувати керовані даними стратегії в напрямку утримання співробітників в компанії. Завдяки застосуванню передових аналітичних методів дослідження

створює цінні методології та рамки для компаній, які борються з проблемами відтоку працівників, позиціонуючи себе як найважливіший ресурс для спеціалістів із відділу людських ресурсів та керівників організацій, які прагнуть підвищити стабільність та продуктивність співробітників. Загалом дослідження підкреслює потенціал застосування методів машинного навчання, прогнозної аналітики та комп'ютерного моделювання для прийняття обґрунтованих рішень щодо утримання співробітників і підтримки здорового клімату в компанії.

Ключові слова: *комп'ютерне моделювання, машинне навчання, логістична регресія, дерево рішень, відтік співробітників, прогнозна аналітика.*

УДК 004.942
DOI https://doi.org/10.32782/2663-5941/2025.3.2/56

Філімончук Т.В.
Харківський національний університет радіоелектроніки

Колтун Ю.М.
Харківський національний університет радіоелектроніки

Климова І.М.
Харківський національний університет радіоелектроніки

Холобок В.І.
Харківський національний університет радіоелектроніки

МОДЕЛЬ ПІДХОДУ РОЗРОБКИ МОБІЛЬНИХ ЗАСТОСУНКІВ

У статті досліджується актуальність розробки мобільних застосунків у сучасних умовах цифрової трансформації з акцентом на вдосконалення моделі вибору між нативним та кросплатформним підходами до їх створення. Аналізуються ключові аспекти, що впливають на успішність мобільних рішень: архітектура застосунку, вибір операційної системи (Android, iOS), аудиторія та терміни реалізації проєкту. Дослідження базується на вивченні наукових публікацій, які розглядають інструменти розробки, принципи проєктування та порівняльний аналіз існуючих технологій. В статті наголошено на важливості застосування сучасних інструментів для масштабованості, надійності, зручності та ефективності мобільного застосунку. На основі проведеного аналізу існуючого стану запропонована оновлена модель розробки мобільних застосунків, яка включає додаткові елементи, що націлені на створення перспективного рішення для користувачів. Доцільність кожної складової в моделі детально аналізується з точці зору переваг та недоліків. Такий підхід дозволяє проєктувати більш гнучкі, функціональні та адаптовані до ринкових вимог програмні рішення.

Впровадження моделі в процес розробки мобільних застосунків сприяє створенню стабільних, масштабованих та користувацько-орієнтованих рішень, що відповідають сучасним вимогам цифрового ринку за рахунок підтримки стандартів розробки. Результати впровадження моделі наголошують, що нативна розробка забезпечує високу продуктивність та доступ до всіх можливостей пристрою, але потребує більших ресурсів, кросплатформні рішення економлять час та кошти, хоча мають обмеження в інтеграції з API, використання гібридних та PWA-рішень більш виправдане для веборієнтованих продуктів із меншими вимогами до продуктивності. Результати тестування використання моделі свідчать про необхідність гнучкого підходу до вибору технологій залежно від цілей проєкту та потреб аудиторії.

Ключові слова: *розробка мобільних застосунків, нативний та кросплатформний підхід, Android, iOS, UX, UI, безпека даних, API, хмарні сервіси, бази даних.*

Постановка проблеми. На даний час однією із найбільш динамічних галузей, що просувається, є розробка мобільних застосунків, які відкривають нові можливості для надання послуг та взаємодії з користувачами в реальному часі. Проте ключ до успіху мобільного рішення полягає не лише в його інноваційних функціях, а залежить його архітектури.

Дана робота присвячена вибору напряму розробки мобільних застосунків (нативного або кросплатформного), який допоможе розробнику прийняти рішення з урахуванням важливих компромісів. Дослідження включає перелік чітких критеріїв, яким слід приділити увагу при розгортанні застосунку.

У вік цифрових технологій мобільні застосунки виступають інструментом для багатьох індустрій, таких як транспорт, електронна комерція, онлайн-банкінг, подорожі, роздрібна торгівля та корпоративні послуги. Застосунки можуть бути орієнтовані на споживачів (B2C) або на бізнес (B2B). B2C-застосунки спрямовані на задоволення індивідуальних потреб користувачів та поділяються на контент-орієнтовані, маркетингові та сервісні. Контент-орієнтовані рішення надають користувачам інформацію, розваги, можливості для спілкування або підвищення продуктивності. Маркетингові, як правило, використовуються для реклами бренду, а сервісні дозволяють виконувати певні задачі, такі

як бронювання квитків або перевірка розкладу поїздів.

B2B-застосунки підтримують внутрішні бізнес-процеси компаній, наприклад, управління складом або автоматизацію продажів. Вони зазвичай мають обмежений доступ та призначені для використання лише певними співробітниками. Вибір операційної системи, яку слід підтримувати, є одним з ключових рішень у процесі розробки таких застосунків.

Ринок смартфонів значно виріс за останнє десятиріччя. Протягом цього часу використовувалися різні мобільні ОС, такі як Android, Blackberry, iOS, Symbian та Windows. Сьогодні на ринку домінують дві платформи – Android та iOS, але у майбутньому ситуація може змінитися. Компанії повинні враховувати ці невизначені обставини, приймаючи рішення про стратегію розробки. Вибір ОС для розробки мобільних застосунків залежить не тільки від частки ринку або орієнтованої індустрії. Виділяють чотири фактори для визначення, яка ОС найкраще підходить для вимог підприємства:

– аудиторія: Android має більшу частку ринку, і за 2023 рік Google Play Store отримав 110 мільярдів завантажень застосунків по всьому світу, в той час як Apple Application Store отримав лише 41,5 мільярдів;

– монетизація: хоча Android має більше завантажень, Apple майже в два рази випереджає за валовими витратами споживачів;

– термін реалізації проєкту: розробка застосунків для Android займає в середньому на 30–40% більше часу, ніж для iOS;

– бюджет: вартість розробки мобільного застосунку залежить від обсягу та складності проєкту, а також від підтримки пристроїв та версій ОС, маркетингових зусиль, участі команди та міждепартаментної взаємодії, а також підтримки та оновлення.

Аналіз останніх досліджень і публікацій. Під час вивчення сучасних підходів до розробки мобільних застосунків було проаналізовано низку наукових публікацій, присвячених архітектурі, інструментам та принципам проєктування мобільних застосунків.

У роботі [1] розглянуто модель мобільного застосунку, яка орієнтована на обробку даних. Основна увага у роботі приділяється взаємодії ключових компонентів, таких як інтерфейс користувача (UI), клієнтська та серверна частини, API, заходи безпеки, модулі аналітика та моніторингу, хмарні сервіси та модулі тестування. Автори натякають, що ретельно продумана модель мобільного застосунку є ключем до його успішної реалізації та конкурентоспроможності на ринку, забезпечує ефективну взаємодію компонентів, безпеку та продуктивність, що робить її придатною для сучасних технологічних вимог.

У роботі [2] розглянуто основні аспекти створення мобільних застосунків та вибір відповідних інструментів для їх розробки. Автори аналізують переваги та недоліки нативного та кросплатформного підходів. Використання першого забезпечує високу продуктивність та оптимізований UX, але їх розробка значно дорожча. Застосунки, які розроблено з орієнтацією на другий підхід, є більш економічними, проте можуть мати обмеження у продуктивності та гнучкості. Загалом в статті підкреслюється важливість усвідомленого підходу до вибору інструментів розробки, що сприяє створенню якісних, продуктивних та доступних мобільних рішень.

Метою роботи [3] є порівняльний аналіз трьох популярних крос-платформних рішень для створення мобільних застосунків: React Native, Flutter та Kotlin Multiplatform. Вибір між цими технологіями, як правило, залежить від вимог проєкту, кожна із розглянутих мов та технологій може бути використана тільки з точці зору аналізу завдань та сценаріїв використання. Фреймворк React Native підходить для реалізації швидких прототипів та веборієнтованих команд, Flutter є оптимальним для застосунків із складним UI та необхідністю однакового вигляду на всіх платформах, а Kotlin Multiplatform найбільше підходить для проєктів, які потребують високої продуктивності та інтеграції з нативними рішеннями. Таким чином, використання кросплатформних рішень дозволить розробнику зменшити витрати на розробку мобільних застосунків, проте кожне з рішень має свої переваги та обмеження, які необхідно враховувати при обиранні технології.

В роботі [4] розглянуто архітектуру платформи Android та структуру мобільного застосунку. Автори докладно розкривають сутність та призначення основних програмних компонентів: Activity, Service, Content Provider та Broatcast Resiver, наводять методи, які слід використовувати для побудови інтуїтивного інтерфейсу користувача. Особлива увага приділяється формам збереження та обробки даних на мобільному пристрої.

Метою роботи [5] є ретельний аналіз підходів та інструментальних засобів кросплат-

формної розробки мобільних застосунків, які дозволяють прискорити процес написання програмного коду. Автор проводить аналіз популярних фреймворків, які можливо використовувати для розробки кінцевого рішення за певними критеріями. При обиранні технології слід керуватися потребами застосунку, бюджетом, наявністю команди розробників та іншими чинниками. Автор пропонує використовувати фреймворк React Native, який задовольняє більшість потреб.

У роботі [6] проведено аналіз вимог до проєкту, проєктуванню, подальшої реалізації з використанням відповідного інструментарію, тестуванню та подальшої підтримки. Автор розглядає процес побудови мобільних застосунків, як набір окремих модулів, що дозволяє забезпечити гнучкість системи. Головними складовими мобільного застосунку є UX та UI, моделювання яких повинно розпочинатись на етапі проєктування. Також автор пропонує окрему увагу приділити зберіганню інформації та наводить переваги та недоліки використання існуючих на даний час БД.

Розглянуті роботи демонструють динамічний розвиток мобільних технологій та різноманітність підходів до створення застосунків. Вони висвітлюють ключові аспекти, починаючи від вибору моделі розробки до використання сучасних технологій, таких як React Native, Flutter, Kotlin Multiplatform, а також хмарних сервісів та локальних баз даних. Таким чином, розробка мобільних застосунків – це не лише актуальна, а й стратегічно важлива сфера розвитку сучасних цифрових технологій, яка формує майбутній досвід користувача, змінює способи ведення бізнесу та створює нові можливості для взаємодії у цифровому світі.

Постановка завдання. Метою даного дослідження є модифікація існуючої моделі, яка орієнтована на розробку мобільних застосунків, за рахунок додавання та розширення її складових. Запропонована модель дозволить створювати якісні, продуктивні та зручні продукти, що відповідають потребам користувачів.

Виклад основного матеріалу. Проведений аналіз виявив, що існуюча на даний момент модель, яка орієнтована на розробку мобільних застосунків має наступні складові (1): FE (front-end) – клієнтська частина; BE (back-end) – серверна частина; API (application programming interface) – інтерфейс програмування застосунків, який забезпечує взаємодію між клієнтською та серверною частинами; DS (data storage) – методи зберігання даних застосунку.

$$M=\{FE, BE, API, DS\}. \quad (1)$$

Наведена модель (1) підтримує просту взаємодію між кінцевим користувачем та мобільним застосунком, та має ряд недоліків:

– вона може бути менш не точною, оскільки не враховує деякі аспекти та взаємодії між окремими компонентами, що в свою чергу впливає на результати роботи застосунку та його ефективність;

– просту модель легше проєктувати та впроваджувати, але це обмежує її гнучкість та подальше удосконалення;

– просту модель не можливо використовувати в задачах, де потрібна складка або спеціалізована обробка.

Для усунення перелічених недоліків було вирішено модифікувати модель (1) за рахунок додавання та розширення її окремих складових. Запропонована модель (2) буде забезпечувати більш ефективну інтеграцію front-end та back-end компонентів, більш високий рівень продуктивності, безпеки та легкість масштабування за рахунок врахування ключових аспектів сучасної розробки:

$$M = \{OS, FE, BE, DS, UX, UI, PS, G, SDE, PN, TT, MAT\}, \quad (2)$$

де: OS (operating system) – нативна розробка на iOS або Andoid; UI (user interface) – інтерфейс користувача; UX (user experience) – досвід користувача; PS (payment systems) – інтеграція з платіжними системами; G (geolocation) – технологія визначення місцезнаходження; SDE (secure data exchange) – модуль безпечного обміну даними; PN (push notifications) – система повідомлень; TT (testing tools) – модуль інструментів тестування; MAT (monitoring and analytics tools) – інструменти для моніторингу та аналітики.

На даний час на ринку розробки мобільних застосунків конкурують дві платформи: iOS та Android. Головна їх відмінність полягає в використанні екосистеми. Як правило, платформа iOS має неповторну ідентичність, яка підтримується суворим контролем якості та дизайном. Платформа Android, надає розробникам більш гнучке середовище для розробки з використанням широкого спектру пристроїв. Кожна із платформ має свої переваги та недоліки, але орієнтуватись потрібно на цілі проєкту, доступні людські та грошові ресурси та очікувану продуктивність.

Front-end програмування відповідає за роботу клієнтської частини застосунку (веб або мобільний). На даному етапі розробки особлива увага приділяється розробці інтерфейсу між кінцевим користувачем та серверною частиною мобільного рішення. Складова FE запропонованої моделі відповідає за введення інформації від користувача, її первинну обробку, а також подальше її відправку на сервер за допомогою відповідного API.

Back-end програмування орієнтоване на розробку серверної частини програми, яка відповідає за передачу даних між користувачами. Складова BE моделі, що запропонована містить наступні інструменти (3):

$$BE = \{SA, API, AP, M\}, \quad (3)$$

де SA (server architecture) – серверна архітектура, яка містить алгоритми завантаження даних, методи авторизації, кешування, бази даних та інше; API (application programming interface) – інтерфейс прикладного програмування, який визначає функціональність серверної логіки; AP (admin panel) – адміністративна панель застосунку, функціонал якої розробляється виходячи з цілей та задач проєкту; M (metrics) – метрики, які оцінюють ефективність проєкту (статистичні дані щодо приросту користувачів, активності, щоденна відвідуваності проєкту в цілому та його окремих розділів, демографічні дані).

Складова DS (4) орієнтована на використання низки методів, які дозволяють здійснювати зберігання даних проєкту. Слід розуміти, що база даних є ключовим компонентом будь-якого сучасного застосунку, який забезпечує надійне зберігання та управління інформацією. У вебзастосунках база даних використовується, як правило, для зберігання контенту, даних користувача, а також журналів транзакцій. У мобільних застосунках вона часто виступає у ролі локального сховища для тимчасового зберігання даних або кешування, що забезпечує швидкий доступ до інформації навіть без підключення до мережі.

$$DS = \{DB, CS, FS, CM\}, \quad (4)$$

де DB (database) – організована структура для зберігання, зміни та обробки взаємозалежної інформації, переважно великих обсягів; CS (cloud storage) – хмарні сховища, орієнтовані на зберігання даних на віддалених серверах, доступ до яких здійснюється через інтернет; FS (file system) – файлова система, орієнтована на зберігання та управління даними на пристрої зберігання; C (caching mechanisms) – набір механізмів, що кешують.

В якості складової DB можливо використовувати:

– реляційні бази даних, які організовують дані у таблиці, що з'єднані між собою за допомогою ключів (MySQL, PostgreSQL, SQLite, MariaDB);

– NoSQL бази даних, які призначені для роботи з неструктурованими або слабо структурованими даними. NoSQL включають документоорієнтовані, колоночні, графові та інші типи баз даних (MongoDB, CouchDB, Firebase Firestore, Apache Cassandra, HBase, ScyllaDB, Neo4j, OrientDB);

– розподілені бази даних, де дані зберігаються на кількох серверах, що забезпечує високу доступність, відмовостійкість та масштабованість (CockroachDB, Google Cloud Spanner, TiDB).

Хмарні сховища (CS) – це сервіси, які дозволяють зберігати та керувати даними у віддалених дата-центрах, забезпечуючи доступність, безпеку та масштабованість. Перевагами використання хмарних сховищ є можливість резервного копіювання та захист даних, а також спільна робота з ними з будь-якого пристрою. Більшість хмарних сховищ, орієнтовані під використання конкретних цілей: для зберігання файлів (Google Drive, Dropbox, OneDrive), для застосунків (Amazon S3, Firebase, Cloud Storage), для баз даних (Cloud SQL, Firestore), для віртуальних машин (Amazon EBS, Azure Disks).

Файлова система (FS) – це структура, яка використовується операційною системою для розміщення та упорядкування даних, управління доступом, відстеження використання простору, веде процес журналювання (NTFS, ext4, XFS), здійснює захист від збоїв та відновлення даних збоїв (Btrfs, ZFS).

Файлові системи різняться за будовою та функціонуванням: локальні використовуються на жорстких дисках, SSD та флеш носіях (FAT32, exFAT, NTFS, APFS, ext4, XFS, Btrfs), мережні – для спільного доступу до мережі (NFS, SMB, AFP, WebDAV), розподілені – в кластерах та хмарних системах (GFS, HDFS, CephFS, GlusterFS). Вибір файлової системи залежить від конкретних потреб, сумісності з ОС та типу носія інформації.

Кешування – тимчасове зберігання даних, які часто використовуються у швидкому сховищі (оперативна пам'ять, SSD, диск) для прискорення

доступу та зниження навантаження на основне джерело даних.

Існують різні види кешування (CM):

– апаратне: кеш-пам'ять процесора, кеш-контролери жорстких дисків;

– програмне: в оперативній пам'яті, на диску;

– мережне: кешування у CDN, DNS-кеш.

Прикладами популярних інструментів кешування є Redis, Memcached, Varnish, NGINX FastCGI Cache, Cloudflare CDN, їх використання значно покращує швидкодію застосунків.

Інтерфейс користувача (UI) – це головний елемент мобільного застосунку. Використання існуючих на даний час фреймворків та спеціалізованих бібліотек спрощують його розробку.

Кросплатформні UI-фреймворки дозволяють писати один код та запускати його на платформах iOS та Android. Наприклад, Flutter дозволяє розробляти UI за допомогою віджетів (Material Design, Cupertino), має гнучку систему кастомізації інтерфейсу та високу продуктивність завдяки використання рушия Skia. Фреймворк React Native має компоненти для створення UI у стилі iOS та Android, підтримує сторонні UI-бібліотеки та дозволяє використовувати нативні модулі кастомізації.

Нативні UI-фреймворки використовуються для розробки під конкретну платформу. Наприклад, для розробки під Android можливо використовувати зв'язку XML та View System для класичного способу розмітки UI або Jetpack Compose для написання UI в декларативному стилі. Для iOS в якості основного фреймворку можливо використовувати UIKit або SwiftUI, які забезпечують декларативний підхід для організації швидкого прототипування.

Досвід користувача (UX) – це емоційні враження людини від взаємодії з кінцевим рішенням і при розробці слід дотримуватися принципів UX-дизайну:

– простота: мінімалізм в інтерфейсі та зрозуміла навігація;

– швидкодія: швидкість завантаження та чуйність критичні;

– інтуїтивність: логічна архітектура, зрозумілі жести та елементи управління;

– доступність: підтримка різних екранів, шрифтів, режимів;

– зворотній зв'язок: анімації, вібрації, підказки.

Складова PS моделі, що пропонується, забезпечує зручні та безпечні онлайн-платежі для мобільного застосунку та надає інтеграцію з платіжними системами, такими як Stripe, PayPal, Apple Pay та Google Pay. Ці системи дозволяють приймати платежі від клієнтів по всьому світу, використовуючи різні методи оплати. Варто зазначити, що доступність та умови використання платіжних систем можуть відрізнятися залежно від країни. Наприклад, у деяких певні платіжні системи можуть бути недоступні або мають обмеження.

Технологія визначення місцезнаходження користувача або пристрою (складова G) на основі GPS, Wi-Fi, мобільних мереж або IP-адреси використовується у мобільних застосунках, вебсайтах та корпоративних рішеннях для доставки персоналізованого контенту, логістики, картографії та навігації. На даний час технологія геолокації застосовується у:

– логістиці та доставці: трекінг посилок, моніторинг кур'єрів;

– навігації: GPS-карти для автомобілів, туристичні маршрути;

– локальних сервісах: пошук найближчих закладів, ресторанів та ін.;

– безпеці: відстеження викрадених гаджетів.

Модуль, який відповідає за безпечний обмін даних (SDE) – це комплекс заходів, що включає механізми шифрування, захищені канали зв'язку та автентифікацію (5). Чим більш чутливі дані передаються, тим більше рівнів захисту слід застосовувати при розробці застосунку. Впровадження ефективних практик безпеки є критичним для забезпечення конфіденційності, цілісності та доступності даних користувачів.

$$SDE = \{E, SCE, AV, SL\}, \quad (5)$$

де E (encryption) – механізми шифрування; SCC (secure communication channels) – використання захищених каналів зв'язку; AV (authentication + verification) – аутентифікація та авторизація; SL (social login) – аутентифікація через акаунти соціальних мереж.

Механізми шифрування (E) – це набір засобів, які дозволяють здійснювати перетворення даних у захищений формат, що можна прочитати лише за допомогою спеціального ключа. Шифрування використовується для захисту конфіденційної інформації під час її зберігання та передачі. На даний час існує декілька типів шифрування, які розрізняються за алгоритмами та областями застосування:

– симетричне шифрування: AES, DES, 3DES, Blowfish/Twofish;

– асиметричне шифрування: RSA, ECC, Diffie-Hellman;

– хешування: SHA, MD5, bcrypt / scrypt / Argon2;

– шифрування на транспортному рівні: TLS, SSL, End-to-End (E2EE);

– стеганографія.

Вибір типу шифрування залежить від задачі, яка вирішується. Наприклад, для захисту файлів слід обрати AES-256, для веббезпеки – TLS + RSA/ECC, для зберігання паролів – bcrypt або Argon2, для обміну ключами – Diffie-Hellman, для зашифрованих месенджерів – End-to-End (E2EE).

Використання надійних протоколів передачі даних (SCE) при розробці мобільних застосунків дозволить зробити кінцеве рішення надійним, стабільним та затребуваним у користувачів. Вибір протоколу передачі даних також слід орієнтувати на вирішення конкретних задач. Наприклад, для організації безпечних вебзастосунків та API слід обирати HTTPS або TLS, для фінансових транзакцій орієнтуватись на TLS, HMAC, OAuth.

Аутентифікація та авторизація (AV) – це важливі процеси у сфері інформаційної безпеки, які використовують разом, але вони виконують різні функції. Аутентифікація – перевіряє особу або систему, яка намагається отримати доступ до ресурсу (2FA/MFA). Аутентифікація підтверджує, що користувач є тим, за кого себе видає, за допомогою облікових даних, таких як логін та пароль, біометричні дані або інші методи. Авторизація – це визначення прав та привілеїв аутентифікованого користувача щодо доступу до певних ресурсів або виконання дій (OAuth, JWT, 2FA). Після успішної аутентифікації система перевіряє, які дії дозволені цьому користувачу.

Можливість входу через акаунти соціальних мереж (SL) дозволяє користувачам здійснювати аутентифікацію на вебсайтах або в мобільних застосунках, використовуючи облікові записи із соціальних мереж, таких як Facebook, Google, Twitter, GitHub, LinkedIn. Така аутентифікація спрощує процес реєстрації та входу для користувачів, оскільки їм не потрібно створювати новий обліковий запис або запам'ятовувати ще один пароль. Перевагами входу через акаунти соціальних мереж є швидка реєстрація (без паролю), зручність для користувачів (авто-вхід), безпека (OAuth 2.0 + шифрування) та підтримка різних соцмереж (Google, Facebook, Apple).

Технологія push-сповіщень (PN) дозволяє надсилати користувачам повідомлення навіть тоді, коли застосунок не активний або працює у фоновому режимі. Такі повідомлення використовуються для інформування про важливі оновлення, новини, акції чи події, залучаючи користувачів до взаємодії із застосунком. Push-сповіщення доставляються в режимі реального часу; можуть містити текст, зображення, кнопки та інші інтерактивні елементи; працюють через сервери повідомлень (наприклад, Firebase Cloud Messaging для Android та APNs для iOS); використовуються для підвищення залученості користувачів.

Процес тестування є обов'язковою складовою розробки кінцевого продукту, який забезпечує стабільність, надійність та якість програмного забезпечення. Тестування дозволяє виявляти та усувати помилки на ранніх етапах, що мінімізує ризик збоїв після впровадження змін чи масштабування застосунку, а також гарантує, що функціонал працює відповідно до заданих специфікацій, забезпечуючи якісний досвід користувача. У розробці слід обов'язково застосовувати різні види тестування (6):

$$TT = \{F, UA/UX, P, S, C, A\}, \quad (6)$$

де F (functional) – функціональне тестування, яке перевіряє роботу застосунку в цілому; UA/UX – тестування зручності та інтуїтивності інтерфейсу, роботи анімації, зрозумілості; P (productivity) – тестування продуктивності (швидкість відповіді, навантаження); S (security) – тестування безпеки (захист даних при пересиланні та зберіганні); C (compatibility) – тестування сумісності (робота на різних пристроях, із різними версіями ОС); A (automated) – автоматизоване тестування (з використанням скриптів).

Наявність інструментів аналітики та моніторингу (MAT) є важливими складовими сучасного застосунку, які забезпечують глибоке розуміння поведінки користувачів та продуктивності системи. Інтеграція інструментів аналітики дозволяє відстежувати активність користувачів, аналізувати взаємодію з різними елементами інтерфейсу та збирати ключові метрики, такі як конверсії, час сесій та залученість.

Моніторинг продуктивності системи забезпечує можливість своєчасного виявлення та вирішення потенційних проблем, таких як високий час відгуку сервера, перевантаження бази даних чи збої в роботі API. Використання таких інструментів, як Google Analytics, Firebase

Analytics, Mixpanel або Sentry, дозволяє автоматизувати процес збору та аналізу даних, забезпечуючи актуальну інформацію для прийняття відповідних рішень. Такий підхід підвищує якість досвіду користувача, дозволяє своєчасно реагувати на проблеми та адаптувати функціонал відповідно до потреб цільової аудиторії. Інтеграція аналітики та моніторингу уніфікує процес управління даними, сприяючи підвищенню ефективності розробки та підтримки застосунків.

Висновки. В ході аналізу підходів для проєктування та розгортання мобільних застосунків було розглянуто наукові публікацій, які пропонують використовувати різноманітні мобільні платформ та SDK з вказанням на їх переваги та недоліки. Наявність великої кількості інструментів розробки створюють унікальні проблеми, такі як вибір SDK, досвід користувача, стабільність фреймворка, легкість оновлень, вартість розробки для кількох платформ та час виходу застосунку на ринок.

Проведений аналіз показав, що вибір напряму розробки мобільних застосунків, як правило, залежить від цілей проєкту, доступних ресурсів та очікуваної продуктивності: використовувати нативні застосунки слід для складних рішень із високими вимогами до продуктивності, кросплатформні підходи для ситуацій, коли є потреба створювати мобільні рішення для широкої аудиторії, а гібридні та PWA-рішення для веборієнтованих додатків з меншими вимогами до продуктивності. Також слід зазначити, що нативні рішення розробляються окремо для кожної платформи, що потребує більше зусиль: як грошових, так і людських. В протилежність нативній розробці кросплатформна дозволяє створювати застосунок з використанням єдиного коду для кількох платформ.

Вибір кросплатформного підходу може здатися очевидним, оскільки можна мати єдину базу коду для кількох ОС, що робить розробку та підтримку швидшими, простішими та дешевшими. Однак це не завжди так. Кросплатформні фреймворки мають обмеження, які роблять їх менш здатними для розробки застосунків у порівнянні з нативним підходом і важко інтегруються з API пристрою. Також слід зазначити, що інструменти для кросплатформної розробки складніші для тестування та налагодження, ніж нативні.

При розробці мобільних застосунків можливо використовувати низку інструментів дизайну та прототипування для створення макетів, прототипів та інтерактивних інтерфейсів (Figma, Adobe XD, Sketch, InVision). Каркасні макети можуть допомогти з проєктуванням логіки взаємодії користувачів із продуктом (Miro, Whimsical, Balsamiq). Дизайн-системи та UI-кити – це готові компоненти, які прискорюють роботу та роблять дизайн одноманітним (Material Design, Apple HIG, Figma Community). Використання при розробці застосунків інструментів для дослідження та тестування UX допоможе проаналізувати поведінку користувачів, протестувати прототипи та зібрати зворотний зв'язок (Hotjar, Google Analytics, Maze, UserTesting, Lookback). Інструменти тестування доступності допоможуть перевірити, наскільки інтерфейс зручний для людей з обмеженими можливостями (Stark, Axe DevTools, WAVE).

Модель вибору підходу розробки мобільного застосунку має низку складових, які орієнтовано на використання різноманітних інструментів розробки. В залежності від потреб застосунку обирається інструментарій, який дозволить збалансувати продуктивність, швидкість розробки та витрати. Для складних і ресурсомістких застосунків доцільним є використання нативної розробки, яка забезпечує найвищу продуктивність і доступ до всіх можливостей пристрою. Кросплатформні фреймворки дозволяють пришвидшити розробку та зменшити витрати, проте можуть мати обмеження щодо інтеграції з API пристрою та продуктивності. Використання гібридних та PWA-рішень виправдане для веборієнтованих продуктів із меншими вимогами до продуктивності. Крім того, інструменти для дизайну, UX-досліджень і тестування забезпечують створення зручного, доступного та функціонального інтерфейсу. Запропонована модель підвищує консистентність коду, оскільки всі частини проєкту розроблятимуться в єдиному стилі та з використанням єдиних парадигм. Кожен із цих компонентів впливає на загальну структуру застосунку, створюючи стабільну, продуктивну та легко масштабовану архітектуру.

Список літератури:

1. Бешта В.С., Комаричев А.В., Філімончук Т.В., Баранєй Д.І. Модель мобільного додатку, яка орієнтована на обробку даних. *Системи управління, навігації та зв'язку*. Полтава: ПНТУ, 2024. Випуск. 3 (77). С. 80–83. DOI: 10.26906/SUNZ.2024.3.080.

2. Якимчук В.С., Гаврилюк В.С. Порівняння операційних систем iOS та Android. Переваги та недоліки розробки мобільних додатків для кожної з систем. *Біомедична інженерія і технологія*. Київ: КПІ ім. Ігоря Сікорського, 2019. Вип. 2. С 86–94.

3. Ічанська Н.В., Улько С.І. Основні аспекти створення мобільних додатків та вибір інструментів їх розробки. *Системи управління, навігації та зв'язку*. Полтава: ПНТУ, 2020. Вип. 1(59). С. 74–78. DOI: 10.26906/SUNZ.2020.1.074.

4. Шматко О.В., Поляков А.О., Федорченко В.М. Аналіз методів і технологій розробки мобільних додатків для платформи Android: навч. посібник. Харків: НТУ «ХПІ», 2018. 284 с.

5. Кислий О.І. Порівняльний аналіз інструментальних засобів для розробки мобільних додатків. *Наукові записки молодих учених*. Кропивницький: РВВ ЦДУ ім В. Винниченка, 2021. Вип. 8. URL: https://phm.cuspu.edu.ua/ojs/index.php/SNYS/article/view/1863/pdf (дата звернення 24.04.2025).

6. Сідельнікова Д.С. Етапи створення дизайну мобільного застосунку. *Мультимедійні технології в освіті та інших сферах діяльності:* матеріали XIII наук.-практ. конф. (Київ, 2 листопада 2021 р). Київ: НАУ, 2022. Секція № 2. С. 105–109.

Filimonchuk T.V., Koltun Yu.M., Klymova I.M., Kholobok V.I. MODEL OF THE MOBILE APPLICATION DEVELOPMENT APPROACH

The article explores the relevance of mobile application development in the current conditions of digital transformation, with a focus on improving the model for choosing between native and cross-platform approaches to app creation. It analyzes key aspects that influence the success of mobile solutions: application architecture, operating system selection (Android, iOS), target audience, and project implementation timelines. The research is based on a study of scientific publications that review development tools, design principles, and comparative analysis of existing technologies. The article emphasizes the importance of using modern tools for scalability, reliability, convenience, and efficiency of mobile applications. Based on the analysis of the current state, an updated model of mobile app development is proposed, which includes additional elements aimed at creating a forward-looking solution for users. The relevance of each component in the model is thoroughly analyzed in terms of its advantages and disadvantages. This approach allows for designing more flexible, functional, and market-adapted software solutions. Implementing the model in the mobile app development process contributes to the creation of stable, scalable, and user-oriented solutions that meet the modern demands of the digital market through compliance with development standards. The results of implementing the model highlight that native development provides high performance and full access to device capabilities but requires more resources; cross-platform solutions save time and money, though they have limitations in API integration; hybrid and PWA solutions are more suitable for web-oriented products with lower performance requirements. Testing results indicate the need for a flexible approach in choosing technologies based on the project's goals and the needs of the target audience.

Key words: *mobile app development, native and cross-platform approach, Android, iOS, UX, UI, data security, API, cloud services, databases.*

UDC 623.746:004.89
DOI https://doi.org/10.32782/2663-5941/2025.3.2/57

Tsybulnyk S.O.
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"

Shelemakha V.V.
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"

OVERVIEW OF METHODS AND MEANS OF AUTOMATED DETERMINATION OF THE CURRENT STATE OF BIOLOGICAL OBJECTS

The beginning of the fourth industrial revolution and the rapid development of technologies – in particular artificial intelligence, cloud computing and the Internet of Things – have given rise to a new era in agriculture, which has been called "smart" or "digital" agriculture. Modern agriculture faces numerous challenges that prevent it from obtaining the desired harvest. This paper overviews scientific works related to cultivated plants grown around the world. It is shown that climate change, which causes droughts, floods, frosts or other adverse weather conditions, soil depletion, water shortages and increased costs for fertilizers and pesticides, plant diseases and the spread of pests – all these factors lead to significant crop losses. Traditional methods of monitoring the condition of cultivated plants and soils do not allow for quick results and forecasting changes. On the other hand, modern automated diagnostic systems provide farmers with the opportunity to quickly obtain a wide range of necessary information about the current condition of cultivated plants. The rapid development of digital technologies leads to the need to carry out a comprehensive review of existing methods and means of determining the current state of biological objects, analyze their advantages and disadvantages, degree of automation, accuracy, adaptability to different types of crops and growing conditions. The results of this analysis will allow us to form substantiated recommendations for improving existing or developing new automated systems for monitoring cultivated plants that will meet modern requirements for accuracy, speed and scalability. Agriculture plays a key role in the production of food, fodder and industrial crops. However, traditional farming methods are largely dependent on the workforce, which can seriously affect the efficiency and sustainability of field production. In this regard, the introduction of field robots that can replace humans in performing routine and labor-intensive tasks such as plowing, sowing, spraying, fertilizing, harvesting and transportation is becoming extremely relevant. This will contribute to increasing the level of automation of agricultural processes and help combat global food shortages. Modern agricultural robots combine innovative technologies, including advanced robotics, sensors, artificial intelligence and big data analytics. Due to this, they provide high-precision observation, autonomous decision-making, intelligent process management and efficient task execution, which opens the way to a fully autonomous agriculture of the future. That is why, to solve the existing problems of agriculture, it is necessary to create a comprehensive automated system that would include modern solutions in the field of robotics and artificial intelligence.

Key words: *automated collection and processing, artificial intelligence, neural networks, unmanned aerial vehicles, biological objects.*

Formulation of the problem. In the current conditions of rapid development of biotechnology and growing challenges associated with climate change, monitoring of biological objects is becoming particularly relevant. Biological objects are all forms of living organisms, covering a wide range: from microorganisms to highly organized animals and plants. Their systematic study, observation and analysis allow for timely detection of violations in the functioning of biosystems and prevent potential threats to both individual species and entire ecosystems [1].

Among the diversity of biological objects, plants occupy a special place, which are the basis of life on Earth. They not only form the basis of food chains, but also provide oxygen, stabilize the soil, and regulate climatic conditions. In agriculture, plants are the main sources of food, fodder and raw materials, and that is why effective management of their condition is a strategic task for the agricultural sector.

However, plants, like all biological organisms, are susceptible to harmful factors, including diseases, pests, and adverse weather conditions. These

factors can significantly reduce yields, degrade product quality, and cause economic losses. That is why systematic monitoring of plant health and early diagnosis of pathologies are becoming key elements of modern crop production. Early detection of signs of diseases or pest infestation [2] allows for timely preventive or curative measures, avoiding large-scale spread of the problem, and maintaining the productivity of agricultural systems.

The application of innovative technologies in this area, such as remote sensing, computer modeling, machine learning, and automated pathology detection systems, opens new opportunities for increasing the accuracy and efficiency of diagnostics. Thus, the transition from a general understanding of biological objects to a targeted study of the state of plants as key elements of agrobiosystems is not only logical, but also an extremely important step in ensuring food security and sustainable development of agriculture.

In the context of modern technologies, the use of unmanned aerial vehicles (UAVs) is of particular importance for automated monitoring of the condition of plants, as they are an effective source of data for analysis and diagnostics. Data obtained from UAVs can be processed using automated diagnostic systems based on artificial intelligence, computer vision, and machine learning. Such systems can detect signs of diseases, pest infestation, nutrient deficiencies, or environmental stress factors in real time [2]. This significantly increases the speed and accuracy of agronomic decision-making, helps optimize the use of plant protection products, and minimizes costs. The use of UAVs in combination with analytical platforms creates intelligent decision-making support systems that are able not only to identify problems but also to predict their development. As a result, the integration of UAVs into the agricultural monitoring system ensures the transition to precision agriculture, which is based on the operational analysis of large amounts of data and allows achieving high efficiency with minimal environmental impact.

Analysis of recent research and publications. Crop problems caused by diseases, pests and adverse weather conditions are key challenges for modern agriculture. Climate change and rising global temperatures, these factors tend to intensify, which negatively affects the productivity of agricultural systems. Foreign researchers are actively working to study these problems and develop effective approaches to solving them.

In particular, the authors of work [3] note that climate change significantly affects the spread and intensity of plant diseases. Changing temperature regimes and uneven distribution of precipitation create favorable conditions for the development of new pathogens. The authors emphasize the need to integrate modern methods of monitoring and forecasting plant diseases for timely detection and control.

In work [4], the relationship between climate change and the spread of crop pests is examined. The authors note that an increase in average temperature contributes to the expansion of the ranges of some pests that were previously distributed only in warm regions. Researchers recommend, in addition to traditional methods and remote sensing technologies, to implement adaptive plant protection strategies, in particular, breeding varieties resistant to specific pests.

Another approach to solving the problem of climate change and pests is presented in [5]. Climate change significantly affects the biological characteristics, distribution and probability of pest outbreaks in various crops and on all types of land and agricultural landscapes. Already now, up to 40% of the world's food resources are lost due to the activity of pests. In this regard, reducing their negative impact is an extremely important task for ensuring global food security, reducing the use of agricultural resources and reducing greenhouse gas emissions. That is why the authors proposed a climate-smart plant protection system – an integrated interdisciplinary approach aimed at reducing crop losses from pests, improving ecosystem functions, reducing the intensity of emissions per unit of output and increasing the resilience of agricultural systems to climate challenges.

The issue of the impact of weather conditions on crop yields is considered in the study [6]. The authors note that to counteract existing problems, it is necessary to implement adaptation and mitigation strategies. These include both traditional and agroecological practices that contribute to improving soil health, efficient water use and carbon sequestration. Climate-smart technologies and improved irrigation systems will increase the resilience of agricultural systems to change. Educational programs for farmers and local initiatives play an important role. Therefore, to ensure food security in the face of climate change, it is necessary to implement holistic approaches to adaptation and sustainable development of agriculture.

Plant diseases seriously threaten food security and biodiversity, so timely and accurate diagnostics are critically important. As the authors of [7] argue, CRISPR/Cas systems have proven themselves as

effective tools for detecting pathogens due to their high accuracy and speed. However, there is a lack of a comprehensive overview of their application in phytopathology, which limits their practical implementation. In the article, the authors considered the principles of CRISPR/Cas biosensors, examples of detecting viruses, bacteria and fungi in plants, as well as the prospects for the development of these technologies for precision agriculture and effective crop protection.

Climate change also contributes to the spread of new diseases. In the study [8], an analysis of the spread of wheat diseases under conditions of increasing temperature was conducted. The authors found that certain types of fungal diseases, such as powdery mildew, become more aggressive under increased humidity and warm weather. They recommend the use of combined protection methods, including biological preparations and monitoring using UAVs.

In the modern scientific community, much attention is paid to the application of **artificial neural networks and machine learning** algorithms for the classification of biological objects. Due to the ability to automatically learn from large amounts of data and detect hidden patterns, these methods demonstrate high efficiency in recognizing biological samples, such as cells, tissues, or plant and animal species.

Recent advances in the use of convolutional (CNN) and graph convolutional networks have significantly improved the classification of hyperspectral images. However, the limited amount of data, noise, high spectral variability, and complex spatial structures remain serious obstacles, especially in agriculture. To overcome these problems, the authors of [9] proposed a new model – DRFG (dimensionality reduction fuzzy graph network), which combines the advantages of other networks. The system works in two stages: first, pre-classification is performed using CNN, after which the data is refined using lightweight graph convolutional networks and clustering. The authors argue that DRFG provides effective dimensionality reduction and high classification accuracy of hyperspectral images, making it a promising technology for precision agriculture.

The intensive modernization of agriculture requires effective solutions for accurate and operational monitoring of pests. Traditional observation methods are inferior in efficiency, and existing deep learning models are often unstable in complex conditions. In response to these challenges, the authors of [10] proposed an innovative monitoring system that combines an advanced convolutional neural network (SAO-CNN) with swarm intelligence for controlling drones. The system uses adaptive convolutional layers, self-supervised learning, and ConvLSTM for effective video data analysis, and the ACO and PSO algorithms optimize UAV trajectories and task distribution. As a result, a classification accuracy of 91.2% was achieved, flight time was reduced by 29.2%, and energy consumption was reduced by 32%. The solution outperforms popular models (YOLO, ResNet, etc.) and provides high real-time performance, making it an effective tool for precision agriculture and sustainable crop management.

In the study [11], a forest image segmentation model based on CNN with U-Net architecture was implemented and evaluated. The algorithm involved preprocessing satellite images and corresponding masks: resizing, normalization, and division into training and test sets. The model consisted of an encoder, decoder, and pass-through connections, and was trained using the binary cross-entropy loss function and the Adam optimizer, with early stopping and checkpoint retention mechanisms. To assess the quality of segmentation, the authors used IoU, Dice, accuracy, completeness, specificity, and F1-score metrics. The study confirmed the effectiveness of U-Net in segmenting forest areas and the importance of high-quality data selection for training.

In the study [12], deep learning models VGG16, Inception v3, ResNet50 and a specially designed CNN were compared for the detection of cauliflower diseases common in countries such as Bangladesh and India. The models were trained using transfer learning on the VegNet dataset and evaluated using the following metrics: accuracy, loss and F1 score. ResNet50 showed the highest accuracy of 90.85%, while the special model achieved 89.04%. Based on the results, the authors conclude that deep learning, in particular ResNet50, is an effective tool for automatic disease detection, which can improve crop yields and strengthen food security.

Pine wilt disease threatens ecosystems and causes economic losses, so its early detection is critically important. The authors of [13] proposed a PWDViTNet model for detecting early signs of infection. It is based on ShuffleNetV2 (as an efficient CNN architecture) and Vision Transformer, which combine local and global feature modeling. The model also uses multi-scale feature fusion to more accurately focus on infected areas. The authors tested the model on images collected by UAVs in Laoshan National Park. It achieved an accuracy

rate of 72.6%, which is 4.1% higher than the baseline ShuffleNetV2 model, with a slight increase in computational cost and size. The authors claim that PWDViTNet demonstrates effectiveness in early detection of pine wilt disease and can potentially be used for broader monitoring in agriculture.

Weeds significantly reduce crop yields, and the emergence of herbicide-resistant species complicates their control. Effective weed management requires targeted, species-specific strategies that can be implemented using machine learning technologies. However, due to the biological diversity of weeds and changing environmental conditions, their accurate detection remains challenging. The aim of the study [14] was to create an annotated image database of five weed species (common nettle, dandelion, water hemp, Palmer's amaranth, lamb's quarter) and evaluate the performance of the YOLOv8–YOLOv11 models and fast R-CNN. For this purpose, 2348 field images were collected, which underwent pre-processing and annotation. As a result, the authors concluded that the YOLO models demonstrate better suitability for accurate and fast real-time weed detection in agriculture.

The use of UAVs in agriculture is one of the key tools of modern precision agriculture, which allows to significantly increase the efficiency of agricultural land management. However, despite numerous advantages, the use of UAVs in this area is accompanied by several problems that need to be solved. This overview considers only the main challenges described in modern scientific works.

Due to the increase in the cost of labor, agricultural robots are becoming a key element of "smart" agricultural production, replacing people in complex tasks such as harvesting, weeding, pruning and pollination. An important component of such robots is hand-eye coordination, which allows them to act in real time based on visual data. Unlike industrial robots, agricultural robots operate in an unstable natural environment, which makes it difficult to accurately recognize targets and avoid obstacles. It is important to achieve a balance between accuracy and speed of task performance. Despite significant progress, robots still face challenges, especially in object recognition and adaptation to changing conditions. Therefore, the authors [15] concluded that further research should focus on imitating human movements and vision, improving self-learning capabilities, detecting errors in real-world environments, and improving systems with multiple sensors and manipulators.

In [16], an efficient method for accurate detection and localization of banana clusters in natural environments is proposed, aimed at applications in harvesting robots. Existing methods have the following limitations: many parameters and insufficient performance. To address these problems, a lightweight Slim-Banana model based on the improved YOLOv8l architecture is developed. For 3D banana localization, the model is combined with a RealSense depth sensor and TOF (time of flight) technology. It is deployed on an Nvidia Orin NX device. The model achieves 94.7% accuracy, 94.8% recall, and 98% mAP with an inference time of 113.6 ms. The average localization error is approximately 13 mm across all coordinates. The authors claim that this is the first known solution implemented on peripheral devices that demonstrates high accuracy and efficiency even in difficult horticultural conditions.

Remote control of multiple agricultural robots increases the efficiency of field work, but remains a challenging task due to the remoteness of the equipment and changing terrain conditions. Estimating the progress of work is complicated by the difference in task completion times, and the issue of synchronizing the speed of multiple robots in unstable field conditions are still insufficiently studied. In [17], a remote monitoring system is proposed that allows assessing the progress of work and managing field work. The key feature is the calculation of the remaining task completion time in real time and automatic adjustment of the robot speed depending on the current progress. The authors tested the system on real machines: in online simulation, the delay was ~100 ms (enough to safely control up to 50 robots), and in real conditions, it was about 2 s. Thanks to automatic speed adjustment, deviations from the schedule decreased from 26–33 s to only 4–13 s, which reduced the error in task performance.

The authors of the study [18] proposed an autonomous navigation method for agricultural robots designed to work on high beds. The method combines two approaches: movement along set route points and orientation along the beds themselves, which allows the robot to move efficiently without complex trajectory planning, even in environments with minimal landmarks. The robot uses LiDAR data for navigation. The method was tested in a virtual environment and on a real strawberry farm, where the robot demonstrated stability with a deviation of no more than ±0.05 m and ±5° relative to the bed. The results confirm the high accuracy and reliability of the proposed system. The authors of the study also emphasize the importance of preliminary modeling, which allows optimizing the characteris-

tics of the robots and navigation algorithms before real implementation. It was shown that even parameters such as the LiDAR range significantly affect the stability of movement. The authors' future research will focus on modeling more complex conditions–with uneven surfaces and variability of farm landscapes–to improve the accuracy and practical use of field robots.

With the development of autonomous navigation, agricultural robots are increasingly being used in the agricultural sector. One of the key challenges is dynamic obstacle avoidance in complex field conditions. The popular DWA (Dynamic Window Approach) algorithm allows local obstacle avoidance, but its effectiveness is limited by fixed weight coefficients, which reduces adaptability to environmental changes. To solve this problem, the authors of [19] integrated the TD3 (Twin Delayed Deep Deterministic Policy Gradient) deep learning method into DWA. Thanks to this, the weights of the evaluation function adapt to the situation in real time, improving the flexibility and accuracy of navigation. Simulations and field tests have shown that the new TD3-DWA approach successfully avoids obstacles in more than 90% of cases, outperforming the classic DWA. This makes it a promising solution for safe and efficient navigation of agricultural robots.

Agricultural multi-robot task assignment (AMRTA) is a key direction for improving the efficiency of robotic agriculture. In the study [20], the AMRTA task was first presented as a traveling salesman problem with workload constraints (NWC-MTSP), which allows to minimize the maximum working time of individual robots and evenly distribute the load. The authors proposed a new method NWC-APONet, which combines graph neural networks and reinforcement learning for optimal task assignment. Experimental verification on real and synthetic agricultural data showed high efficiency of the model, confirming its practical value for multi-robot control systems in the agricultural sector.

Task statement. Due to the urgency of the problem of ensuring the stable development of the agro-industrial complex and the growing risks associated with diseases of cultivated plants, the impact of pests and adverse weather conditions, there is an objective need to develop effective means of monitoring biological objects, in particular plants. Timely diagnostics of changes in the physiological state of plants is the key to preserving the harvest, increasing the efficiency of agrotechnical measures and minimizing financial losses.

The rapid development of digital technologies leads to the need to carry out a comprehensive overview of existing methods and means of determining the current state of biological objects, analyze their advantages and disadvantages, the degree of automation, accuracy, adaptability to different types of crops and growing conditions. The results of this analysis will allow us to form substantiated recommendations for improving existing or developing new automated systems for monitoring cultivated plants that will meet modern requirements for accuracy, speed and scalability.

Outline of the main material of the study. Determining the current state of biological objects, in particular plants, is critically important for increasing the efficiency and sustainability of agricultural production. Modern agricultural production requires operational and accurate monitoring, which allows for timely detection of problems, optimization of resource use and increase in yield.

Although traditional monitoring methods, in particular visual inspection of fields, remain widespread, they are not effective in large areas due to the high time consumption, the need for significant labor resources and low accuracy. In contrast, automated solutions, based on UAVs, sensor systems and artificial intelligence tools, open new opportunities for farmers.

Today, machine learning and neural networks play a key role in improving the analysis of plant health. The overview showed that one of the most widespread is the convolutional neural network and its variations. Thanks to their ability to process large amounts of data, these mathematical models can accurately (from 70% to 99% depending on the specific type of neural network, the training data set and environmental conditions) recognize disease symptoms, detect pests, analyze vegetation changes and assess stress factors (such as lack of water or nutrients). Combining these technologies with UAV images allows you to create detailed field maps with precise positioning of problem areas. This allows for localized applications of fertilizers, water or plant protection products, which reduces costs and minimizes damage to the environment.

However, despite significant advantages, an overview of scientific works showed that there are several technical barriers that complicate the effective use of UAVs for precise plant monitoring. Among the main problems are navigation limitations. In particular, drones can lose GPS signal in areas with poor coverage or in conditions of obstacles (dense plantings, buildings, complex terrain). In such situ-

ations, alternative navigation methods are needed, such as visual SLAM navigation or inertial systems, which are complex and expensive to implement. In addition, the limited flight time due to battery capacity limits the size of the area that can be covered in a single flight, which creates difficulties when covering large areas.

Among the existing problems, a significant challenge is the delay in data transmission between the UAV and the base station, especially when performing complex tasks in real time. Even small delays in receiving or processing commands can lead to inaccuracies in the flight path, missed target areas, or errors in image collection. The problem is exacerbated when using wireless communication channels in conditions of interference or over long distances, which reduces the responsiveness of the control system. This limits the ability of the drone to effectively operate in the online correction mode, which is critical for precision agricultural operations.

In addition, the accuracy of plant classification depends largely on the quality of image collection. Factors such as lighting variations, shadows, leaf movement due to wind, weather conditions (fog, dust), as well as differences in crop growth phases can complicate recognition. Even modern neural networks can make mistakes in conditions of excessive visual similarity between species or when training samples were limited.

Another major challenge remains the accessibility of modern solutions for small and medium-sized farms. The high cost of equipment (in particular, UAVs, sensors and software) and the need for specialized skills to operate them hinder widespread implementation. This highlights the need for government support, subsidies and training programs that will help farmers integrate innovations into their operations.

Thus, the combination of high technology, UAVs or ground robots, artificial intelligence and automated monitoring systems has the potential to significantly transform modern agriculture, making it more accurate, environmentally sustainable and economically profitable. The development of such automated system is an urgent task today, therefore, in further work, the design and development of an automated system for determining the current state of biological objects will be carried out.

Conclusions. Agriculture is the basis of the production of food, fodder and industrial crops. However, traditional approaches to its management largely depend on the workforce, and its shortage can negatively affect the efficiency and stability of agricultural production. An overview of scientific works has shown that traditional methods of monitoring cultivated plants and soil quality do not provide the necessary information in a short period of time, which can be critical for the survival of the crop.

The use of robotic systems contributes to an increase in the level of automation of agricultural processes and is a promising tool for overcoming global challenges of food security. It has been shown that modern robots integrate sensors, artificial intelligence and data analytics, which allows achieving high accuracy in monitoring, autonomy in decision-making and efficiency in performing tasks. This creates prerequisites for the development of smart, fully autonomous agriculture.

An overview of scientific works has also shown that the use of UAVs is limited by technical and economic factors. Loss of GPS signal in dense vegetation, short flight time due to limited battery capacity, and communication delays with the base reduce the accuracy of data collection. Additionally, changing lighting, wind, and different phases of crop growth complicate image analysis, and the high cost of equipment and the need for specialized skills limit implementation for small farms. To overcome the existing challenges of the industry, it is advisable to develop an integrated automated system that combines advanced solutions in the field of artificial intelligence and robotics, ensuring adaptability, scalability, and efficient management of agricultural production.

Bibliography:

1. Soreanu G., Tanase C., Mardari C., Cretescu I. Assessment of air purification biosystems for environmental applications. Process Safety and Environmental Protection, Volume 195, 2025, 106742. https://doi.org/10.1016/j.psep.2024.12.123.

2. Canicattì M., Vallone M. Drones in vegetable crops: A systematic literature review. Smart Agricultural Technology, Volume 7, 2024, 100396. https://doi.org/10.1016/j.atech.2024.100396.

3. Sishodia R. P., Ray R. L., Singh S. K. Applications of Remote Sensing in Precision Agriculture: A Review. Remote Sensing, 12 (19), 2020, 3136. https://doi.org/10.3390/rs12193136

4. Subedi B., Poudel A., Aryal S. The impact of climate change on insect pest biology and ecology: Implications for pest management strategies, crop production, and food security. Journal of Agriculture and Food Research, Volume 14, 2023, 100733. https://doi.org/10.1016/j.jafr.2023.100733.

5. Heeb L., Jenner E., Cock M.J.W. Climate-smart pest management: building resilience of farms and landscapes to changing pest threats. J Pest Sci, 92, 2019, P. 951–969. https://doi.org/10.1007/s10340-019-01083-y

6. Imtiaz S. et al Impact of Climate Change on Agriculture. In: Fahad, S., Saud, S., Nawaz, T., Gu, L., Ahmad, M., Zhou, R. (eds) Environment, Climate, Plant and Vegetation Growth. Springer, Cham, 2024, P. 285–305. https://doi.org/10.1007/978-3-031-69417-2_10

7. Gong X.-Y. et al Recent Application of CRISPR/Cas in Plant Disease Detection. TrAC Trends in Analytical Chemistry, 2025, 118251. https://doi.org/10.1016/j.trac.2025.118251.

8. Sharma B. et al Drought stress and its management in wheat (Triticum aestivum L.): a review. Agricultural science and technology,. Vol. 14, No. 1, 2022. P. 3–14. https://doi.org/10.15547/10.15547/ast.2022.01.001

9. Min Peng et al Advanced image segmentation for precision agriculture using CNN-GAT fusion and fuzzy C-means clustering. Computers and Electronics in Agriculture, Volume 226, 2024, 109431. https://doi.org/10.1016/j.compag.2024.109431.

10. Chu S., Bao W. Dynamic Agricultural Pest Classification Using Enhanced SAO-CNN and Swarm Intelligence Optimization for UAVs. International Journal of Cognitive Computing in Engineering, 2025. https://doi.org/10.1016/j.ijcce.2025.04.002.

11. Bhatti M. A. et al Utilizing convolutional neural networks (CNN) and U-Net architecture for precise crop and weed segmentation in agricultural imagery: A deep learning approach. Big Data Research, Volume 36, 2024, 100465. https://doi.org/10.1016/j.bdr.2024.100465.

12. Arnob A. S. et al Comparative result analysis of cauliflower disease classification based on deep learning approach VGG16, inception v3, ResNet, and a custom CNN model. Hybrid Advances, Volume 10, 2025, 100440. https://doi.org/10.1016/j.hybadv.2025.100440.

13. Chen Z. et al PWDViTNet: A lightweight early pine wilt disease detection model based on the fusion of ViT and CNN. Computers and Electronics in Agriculture, Volume 230, 2025, 109910. https://doi.org/10.1016/j.compag.2025.109910.

14. Sharma A., Kumar V., Longchamps L. Comparative performance of YOLOv8, YOLOv9, YOLOv10, YOLOv11 and Faster R-CNN models for detection of multiple weed species. Smart Agricultural Technology, Volume 9, 2024, 100648. https://doi.org/10.1016/j.atech.2024.100648.

15. He L. et al Advance on Agricultural Robot Hand–Eye Coordination for Agronomic Task: A Review. Engineering, 2025. https://doi.org/10.1016/j.eng.2025.01.022.

16. Zheng Z. et al An efficient and lightweight banana detection and localization system based on deep CNNs for agricultural robots. Smart Agricultural Technology, Volume 9, 2024, 100550. https://doi.org/10.1016/j.atech.2024.100550.

17. Ospina R., Noguchi N. Real-time work progress estimation based on GIS remote monitoring system for agricultural robot vehicles. Computers and Electronics in Agriculture, Volume 234, 2025, 110313. https://doi.org/10.1016/j.compag.2025.110313.

18. Fujinaga T. Autonomous navigation method for agricultural robots in high-bed cultivation environments. Computers and Electronics in Agriculture, Volume 231, 2025, 110001. https://doi.org/10.1016/j.compag.2025.110001.

19. Wang J. et al Dynamic obstacle avoidance control based on a novel dynamic window approach for agricultural robots. Computers in Industry, Volume 167, 2025, 104272. https://doi.org/10.1016/j.compind.2025.104272.

20. Lu Z. A reinforcement learning-based optimization method for task allocation of agricultural multi-robots clusters. Computers and Electrical Engineering, Volume 120, Part B, 2024, 109752. https://doi.org/10.1016/j.compeleceng.2024.109752.

Цибульник С.О., Шелемаха В.В. ОГЛЯД МЕТОДІВ ТА ЗАСОБІВ АВТОМАТИЗОВАНОГО ВИЗНАЧЕННЯ ПОТОЧНОГО СТАНУ БІОЛОГІЧНИХ ОБ'ЄКТІВ

Початок четвертої промислової революції та стрімкий розвиток технологій – зокрема штучного інтелекту, хмарних обчислень і Інтернету речей – стали поштовхом до нової ери в сільському господарстві, що отримала назву «розумне» або «цифрове» землеробство. Сучасне сільське господарство стикається з численними викликами, які перешкоджають отриманню бажаного врожаю. У даній роботі проведено огляд наукових робіт, пов'язаних з культурними рослинами, які вирощують у всьому світі. Показано, що кліматичні зміни, які спричиняють засухи, повені, заморозки чи інші несприятливі погодні умови, виснаження ґрунтів, нестача водних ресурсів і зростання витрат на добрива та пестициди, рослинні хвороби та поширення шкідників – усі ці фактори призводять до значної втрати врожаю. Традиційні методи моніторингу стану культурних рослин та ґрунтів не дають змогу швидко отримати результати та прогнозувати зміни. З іншого боку, сучасні автоматизовані діагностичні системи надають аграріям можливість швидкого отримання

широкого кола необхідної інформації про поточний стан культурних рослин. З огляду на стрімкий розвиток цифрових технологій, необхідно здійснити комплексний огляд існуючих методів і засобів визначення поточного стану біологічних об'єктів, проаналізувати їх переваги та недоліки, ступінь автоматизації, точність, адаптивність до різних типів культур і умов вирощування. Результати цього аналізу дозволять сформувати обґрунтовані рекомендації щодо вдосконалення існуючих або розроблення нових автоматизованих систем моніторингу культурних рослин, що відповідатимуть сучасним вимогам до точності, швидкодії та масштабованості. Сільське господарство відіграє ключову роль у вирощуванні продуктів харчування, кормових та технічних культур. Проте традиційні методи ведення господарства значною мірою залежать від великої кількості робочої сили, нестача якої може серйозно вплинути на ефективність і стабільність польового виробництва. У зв'язку з цим надзвичайно актуальним стає впровадження польових роботів, здатних замінити людину у виконанні рутинних і трудомістких завдань, таких як оранка, сівба, обприскування, внесення добрив, збирання врожаю та транспортування. Це сприятиме підвищенню рівня автоматизації сільськогосподарських процесів та допоможе боротися з глобальним дефіцитом продовольства. Сучасні аграрні роботи поєднують інноваційні технології, включаючи передову робототехніку, сенсори, штучний інтелект і аналітику великих даних. Завдяки цьому вони забезпечують високоточне спостереження, автономне прийняття рішень, розумне управління процесами та ефективне виконання завдань, що відкриває шлях до повністю автономного сільського господарства майбутнього. Саме тому для вирішення наявних проблем сільського господарства необхідно створити комплексну автоматизовану систему, яка включала б в себе сучасні рішення в області робототехніки та штучного інтелекту.

Ключові слова: *автоматизований збір та обробка, штучний інтелект, нейронні мережі, безпілотні літальні апарати, біологічний об'єкт.*

УДК 622.24:681.3
DOI https://doi.org/10.32782/2663-5941/2025.3.2/58

Чигур І.І.
Івано-Франківський національний технічний університет нафти і газу

Лагойда А.І.
Івано-Франківський національний технічний університет нафти і газу

Чигур Л.Я.
Івано-Франківський національний технічний університет нафти і газу

ІНФОРМАТИВНІ ОЗНАКИ В ЗАВДАННЯХ АВТОМАТИЗОВАНОГО РОЗПІЗНАВАННЯ ТЕХНОЛОГІЧНИХ СИТУАЦІЙ ПРИ БУРІННІ СВЕРДЛОВИН

Стаття присвячена актуальній проблемі автоматизованого розпізнавання та класифікації технологічних ситуацій, що виникають у процесі буріння свердловин, як важливого етапу для прийняття обґрунтованих рішень при проектуванні та управлінні цим складним технологічним процесом.

Зазначено, що в процесі буріння виникає необхідність ідентифікувати різноманітні фактори, такі як зміни буримості та тріщинуватості гірських порід, ступінь зношення бурового інструменту, а також геологічні ускладнення та аварії. Інформація, отримана в результаті розпізнавання, є критично важливою для ефективного керування процесом буріння, визначення оптимального моменту для підйому бурового інструменту та, що особливо важливо, для попередження аварійних ситуацій.

Запропоновано завдання розпізнавання образів звести до виділення елементів, що належать до конкретного класу, серед множини елементів, які можуть належати до різних класів за допомогою дивергенції. Підкреслюється, що наявність певних спільних властивостей або ознак у елементів одного класу є ключовим критерієм для їх ідентифікації та віднесення до цього класу.

Розглядається важлива проблема попередньої обробки та стиснення вихідних даних у задачах розпізнавання технологічних ситуацій з метою зменшення розмірності вектору вимірювань.

Запропоновано два основні підходи до зменшення числа ознак: селекція найбільш інформативних ознак шляхом відсікання менш значущих та створення невеликої кількості нових, комбінованих ознак на основі наявних. Підкреслено, що другий підхід, хоча й більш трудомісткий, потенційно забезпечує вищу інформативність новостворених ознак за рахунок комбінування вихідних.

Наголошено на необхідності мінімізації втрати інформативності при відборі або комбінуванні ознак, пропонуючи використовувати поняття дивергенції як міри інформативності. Запропоновано подальше застосування дивергенції для скорочення кількості ознак на прикладі двох важливих випадків розподілу даних: нормального розподілу та розподілу Бернуллі, з аналізом обох способів зменшення розмірності (відсікання та лінійне перетворення ознак) для кожного з них.

Таким чином, стаття є вступом до важливої науково-технічної проблеми розпізнавання та класифікації технологічних ситуацій у бурінні, окреслює її актуальність, аналізує існуючі підходи та пропонує потенційні інформаційні основи для подальших досліджень у цій галузі.

Ключові слова: *автоматизоване розпізнавання, буріння, інформативні ознаки, дивергенція, розподіл Бернуллі.*

Постановка проблеми. Прийняттю рішень при проектуванні або управлінні процесом буріння завжди передує процес розпізнавання конкретних ознак, що характеризують умови буріння і визначають вибір найбільш раціональної техніки, технології та організації бурових робіт [1]. Особливо актуальним є розпізнавання технологічних ситуацій при бурінні свердловин. До недавнього часу таке розпізнавання здійснювалося людиною на основі досвіду, знань, інтуїції.

У процесі буріння необхідно розпізнавати гірські породи різної буримості і тріщинуватості, стан зношення породоруйнівного інструменту, геологічні ускладнення і аварії. Результати розпізнавання використовуються для керування процесом буріння, визначення моменту підйому породоруйнівного інструменту, попередження аварій [2]. Природа розпізнавання об'єктів істотно впливає на процес розпізнавання і вирішальні схеми (алгоритми), що дозволяють на

основі аналізу значень ознак об'єкта (ситуацій) зробити висновок про те, до якого класу належить даний об'єкт (ситуація).

Традиційні заходи попередження аварій і ускладнень зводилися в основному до контролю стану обладнання, інструменту, бурового розчину і виконання заданих технологічних параметрів ведення процесу. Попередження аварійних ситуацій в процесі буріння ґрунтувалося на досвіді бурильника, узагальнити і формалізувати який у вигляді рекомендацій без відповідних засобів фіксації процесу досить важко.

Аналіз останніх досліджень і публікацій. Першими дослідженнями по формалізації і систематизації, інформаційних моделей у вигляді аналогових записів витрат потужності при алмазному бурінні з'явилися роботи О. І. Шерстюка, Н. Н. Міхеєва та ін. Інформаційною основою для розпізнавання технологічних ситуацій служили спеціальні форми запису витрат потужності для різних вибійних процесів (припалювання коронки, заполірування алмазів і ін.). Однак аналіз аналогових записів самописного ваттметра показав велику різноманітність і, що особливо важливо, неможливість однозначного тлумачення форм запису, а отже, і розпізнавання ситуацій.

Дослідженнями інформаційних ознак процесу буріння в тому числі з метою ідентифікації його станів займалися такі вчені як Г. Н. Семенцов, М. І. Горбійчук, М. А. Мислюк, А. І. Волобуєв, В. Б. Кропивницька, А. А. Погарський, Р. М. Ейгелес, Е. А. Козловський, Г. Д. Бревдо, Н. Б. Ситніков, Ю. А. Пєшалов, Ю. Ф. Алексєєв, А. І. Співак, В. А. Бражніков, В. К. Маурер (США), І. М. Галлі (США), Г. Б. Вудс (США), А. Т. Боургойн (США), Ф. С. Йоунг (США) та ін.

На основі проведеного аналізу публікацій встановлено, що інформаційною основою розпізнавання можуть служити параметри і показники технологічного процесу буріння свердловин (осьове навантаження, крутний момент, витрати потужності, витрата очисного агента, швидкість буріння, інтенсивність зміни цих параметрів і ін.), Енергоємність руйнування породи, параметри коливань бурильної колони (частота, амплітуда і ін.), властивості промивальної рідини (щільність, в'язкість, ступінь аерації) на вході і виході зі свердловини, статистичні характеристики перерахованих ознак (дисперсія, спектральна густина і ін.) [3].

Постановка завдання. Метою статті є розробка методики оцінки інформативності ознак, що характеризують процес буріння свердловин для забезпечення ефективного розпізнавання технологічних ситуацій, які виникають у процесі роботи.

Виклад основного матеріалу. Загальне завдання розпізнавання полягає у виділенні елементів, що належать конкретному класу, серед безлічі елементів, що відносяться до декількох класів. Елементи, що відносяться до одного класу, мають ряд загальних властивостей або ознак. Наявність цих ознак і служить тим критерієм, який дозволяє відносити конкретні елементи до даного класу.

Однак об'єкти, що підлягають розпізнаванню, можуть мати дуже велику кількість ознак, інформативність яких може бути різна. Для формалізованої оцінки інформативності ознак використовують характеристики розсіювання випадкових величин і різні ентропійні заходи, отримані із статистичної теорії інформації. У таких завданнях доцільно використати дивергенція в якості міри «відстані» або відмінності між класами. Дивергенція – це повна середня інформація для розрізнення класів. У разі неперервних розподілів випадкової величини x вона задається формулою

$$I = \int_{-\infty}^{+\infty} \left(p_1(x) - p_2(x)\right) \ln \frac{p_1(x)}{p_2(x)} dx, \qquad (1)$$

де $p_1(x)$ і $p_2(x)$ – щільності першого і другого розподілів.

Для дискретних розподілів існує аналогічна формула:

$$I = \sum_a \left(p_1(a) - p_2(a)\right) \ln \frac{p_1(a)}{p_2(a)}, \qquad (2)$$

де випадкова величина x може приймати N різних значень a з вірогідністю, рівними $p_1(a)$ і $p_2(a)$ для першого і другого розподілів відповідно.

Надалі для спрощення викладу вважатимемо, що є два класи об'єктів, перехід до більшої кількості класів здійснюється по індукції. Для кожного з класів розподіл об'єктів за ознаками є деяким імовірнісним розподілом. Нехай загальна кількість ознак становить n, а $x_i^{(j)}$ – значення i-ї ознаки у j -го об'єкта, $i= 1,2, ...,n$. Тоді об'єкт можна представити як n-вимірний вектор $x^{(j)} = \left(x_1^{(j)}, ..., x_n^{(j)}\right)$ або точку в n-вимірному просторі R^n, а сукупність всіх ознак утворює n-вимірний простір R^n.

Вираз для визначення дивергенції залежить від закону розподілу ознак. У бурінні вимірювані ознаки і показники мають, як правило, неперерв-

ний розподіл. Багато з них підпорядковуються нормальному або близькому до нього закону розподілу.

У деяких задачах розпізнавання аварійних ситуацій зручніше розглядати ознаки, які приймають лише два значення: +1 при перевищенні порогового значення деяким параметром і -1 при неперевищенні цього значення. Оцінки ймовірностей перевищення порогу, тобто того, що ознаки приймуть значення +1, визначаються на основі збору статистичних даних в конкретних умовах. У цьому випадку розподіл ознак є дискретним, причому ознаки x_i, приймають лише два значення. У цьому випадку кількість різних векторів із цих ознак становить 2^n, де n – кількість ознак. Якщо ж ознаки незалежні, то випадкова величина х має розподіл Бернуллі.

При виведенні формул дивергенції для двох найбільш важливих для буріння законів розподілу використовувалася властивість адитивності дивергенції:

$$I\left(\xi^{(1)},\xi^{(2)}\right)=\sum_{i=1}^{n} I\left(\xi^{(1)},\xi^{(2)}\right). \qquad (3)$$

Зауважимо, що ознаки $\xi^{(1)},\xi^{(2)},i=1,...,n;$ $j=1,...,n$ також є векторними величинами.

Знайдемо спочатку дивергенцію в разі нормального розподілу. Нехай є два нормальних розподіли з однаковими матрицями коваріацій $C^{(1)}=C^{(2)}=C$. У цьому випадку щільності нормальних розподілів представляють собою наступні функції:

$$p_1(x)=\frac{1}{\sqrt{(2\pi)^n \det C}}\exp\left\{-\frac{1}{2}\left[C^{-1}\left(x-m^{(1)}\right)\right]\right\}, \qquad (4)$$

$$p_2(x)=\frac{1}{\sqrt{(2\pi)^n \det C}}\exp\left\{-\frac{1}{2}\left[C^{-1}\left(x-m^{(2)}\right)\right]\right\}, \qquad (5)$$

де $x=\left(x_1,...,x_n\right)$ – вектор ознак; $m^{(1)}=\left(m_1^{(1)},...,m_n^{(1)}\right)$ і $m^{(2)}=\left(m_1^{(2)},...,m_n^{(2)}\right)$ – вектори математичних очікувань першого і другого розподілу відповідно.

В роботі [6] доведено, що

$$I=\left(C^{-1}\delta,\delta\right),\delta=m^{(1)}-m^{(2)}. \qquad (6)$$

Візьмемо в якості базису сукупність власних векторів матриці С. У цій системі координат матриця С має вигляд:

$$C=\begin{pmatrix}\sigma_1^2 & ... & 0\\ ... & ... & ...\\ 0 & ... & \sigma_n^2\end{pmatrix}.$$

У зв'язку з цим вираз (6) перетворюється в наступний:

$$I=\sum_{i=1}^{n}\frac{\delta_i^2}{\sigma_i^2}\ . \qquad (7)$$

Зауважимо, що умова рівності коваріаційних матриць в задачах буріння не є жорсткою. Дисперсія параметрів і показників буріння протягом рейсу визначається в основному властивостями гірських порід – зміна твердості, тріщинуватості і ін. Статистична обробка матеріалів показує, що для багатьох гірських порід (монолітні, мономінеральні і ін.) характерною особливістю є однакова в межах точності вимірювань дисперсія. Тому правильним відбором породоруйнівного інструменту і параметрів режиму буріння можна зменшити дисперсію показників в породах які різко відрізняються властивостями, наблизивши її до дисперсії показників в породах з однаковими властивостями.

Тепер знайдемо дивергенцію між двома розподілами Бернуллі. Ці розподіли визначаються параметрами $p_1^{(1)},...,p_n^{(1)}$ і $p_1^{(2)},...,p_n^{(2)}$, де

$$1,...p_i^{(1)}=p\left\{\xi_i^{(1)}=1\right\},p_i^{(2)}=p\left\{\xi_i^{(2)}=1\right\},i=1,...,n;$$

$$1-p_i^{(1)}=p\left\{\xi_i^{(1)}=-1\right\},1-p_i^{(2)}=p\left\{\xi_i^{(2)}=-1\right\},i=1,...,n;$$

$$\xi^{(1)}=\left(\xi_1^{(1)},...,\xi_n^{(1)}\right),\xi^{(2)}=\left(\xi_1^{(2)},...,\xi_n^{(2)}\right).$$

Величини $\xi_i^{(1)},\xi_j^{(1)}$ можуть приймати лише два значення: 1 або -1; *i=1,...,n; j=1,...,n.*

Нехай спочатку *n=1*. Тоді з визначення дивергенції (2) випливає, що

$$I\left(\xi^{(1)},\xi^{(2)}\right)=\sum_{a=1,-1}\left(p^{(1)}(a)-p^{(2)}(a)\right)\ln\frac{p^{(1)}(a)}{p^{(2)}(a)}=$$

$$=\left(p^{(1)}-p^{(2)}\right)\ln\frac{p^{(1)}}{p^{(2)}}+\left(1-p^{(1)}-1+p^{(2)}\right)\ln\frac{1-p^{(1)}}{1-p^{(2)}}=$$

$$=\left(p^{(1)}-p^{(2)}\right)\ln\frac{p^{(1)}\left(1-p^{(2)}\right)}{p^{(2)}\left(1-p^{(1)}\right)}\ . \qquad (8)$$

Звідси, користуючись властивістю адитивності дивергенції і виразом (8), отримуємо формулу для дивергенції між двома розподілами Бернуллі для будь-якої кількості ознак *n*:

$$I\left(\xi^{(1)},\xi^{(2)}\right)=\sum_{i=1}^{n}\left(p_i^{(1)}-p_i^{(2)}\right)\ln\frac{p_i^{(1)}\left(1-p_i^{(2)}\right)}{p_i^{(2)}\left(1-p_i^{(1)}\right)}\ . \qquad (9)$$

Таким чином, при визначенні за допомогою дивергенції інформативності ознак, які приймають неперервні значення (швидкість буріння, витрати потужності і ін.), слід користу-

ватися виразом (7). У той же час в деяких задачах розпізнавання можна вважати, що будь-яка з ознак приймає всього два значення, тобто при їх незалежності – підпорядковується розподілу Бернуллі. У цьому випадку для визначення інформативності ознак слід користуватися виразом (9). Очевидно, що результати оцінки інформативності одних і тих же ознак будуть різними.

Оскільки виміряти всі ознаки не завжди можливо і не завжди доцільно з технічних або економічних причин, то необхідно зі всієї множини ознак вибрати найбільш інформативні. Іншими словами, перш ніж приступити до вирішення завдання розпізнавання ситуацій, необхідно провести попередню обробку і стиснення вихідних даних з метою зниження розмірності вектору вимірювань. Завдання зменшення числа ознак можна вирішувати двома способами. Перший полягає у виборі найбільш інформативних ознак і відсікання інших. Інший підхід полягає у створенні з великої кількості наявних ознак невеликої кількості комбінованих ознак. При цьому кожна нова ознака залежить від початкових, наприклад, є їх лінійною комбінацією або іншою функцією. При другому підході інформативність новостворених ознак є вищою оскільки комбінуючи вихідні ознаки можна обмежувати зниження інформативності, хоча цей підхід є більш трудомістким.

Відбираючи ознаки способом відсікання або комбінування, слід обмежити зменшення інформативності ознак, що залишилися тобто дивергенцію, якщо вона використовується в якості міри інформативності.

Перейдемо до використання поняття дивергенції для скорочення кількості ознак при вирішенні завдань з розпізнавання ситуацій. З врахуванням вищенаведеного, необхідно прагнути щоб зменшення дивергенції, тобто інформативності було мінімальним.

Розглянемо, як найважливіші, два випадки: нормальний розподіл і розподіл Бернуллі. Для кожного з них розберемо обидва способи скорочення кількості ознак: відсікання і лінійне перетворення ознак.

Якщо в разі нормальних розподілів з однаковою матрицею коваріацій $C = C^{(1)} = C^{(2)}$ скористатися методом відсікання ознак, то очевидно, що нова матриця коваріацій C^* буде являти собою підматрицю колишньої коваріаційної матриці С. Розподіл при цьому як і раніше залишиться нормальним, і тому вираз для дивергенції «усічених» розподілів залишиться аналогічною виразу (6):

$$I^*\left(\xi^{(1)},\xi^{(2)}\right) = I\left(C^*,\delta^*\right) = \left(C^{*-1}\delta^*,\delta^*\right), \qquad (10)$$

де δ^* вектор-стовбець різниці математичних очікувань

$$\delta^* = \begin{pmatrix} \delta_{i1} \\ \vdots \\ \delta_{im} \end{pmatrix} = \begin{pmatrix} m_{i1}^{(1)} - m_{i1}^{(2)} \\ \vdots \\ m_{im}^{(1)} - m_{im}^{(2)} \end{pmatrix};$$

$i = 1,...,m$ – номера змінних, які не брали участі у процесі відсікання.

Емпірично нескладно вибрати ті m невідсічених ознак, при яких дивергенція (10) буде максимальна.

Перейдемо тепер до вибору комбінованих ознак. Зробимо лінійне відображення простору ознак $\left(x_1,...,x_m\right)$ в простір ознак $\left(y_1,...,y_m\right)$:

$$y_i = \sum_{j=1}^{n} a_{ij}x_j, i = 1,...,m,$$

або $y=AX$, где X – це n-вимірний вектор; у – це m-вимірний вектор; А – матриця розмірності $m \times n$.

Обчислимо дивергенцію I^* для цих нових ознак:

$$I^* = I^*\left(C,\delta,A\right) = \left(C^{*-1}\delta^*,\delta^*\right), \qquad (11)$$

де $\delta^* = A\delta$ – це m-вимірний вектор; $C^* = ACA'$ – це m-вимірна матриця.

При будь-якому m і будь-якій матриці A виконується наступна нерівність:

$$I^*\left(C,\delta,A\right) \le I\left(C,\delta\right),$$

в якому існує така матриця А, для якої

$$I^*\left(C,\delta,A\right) = I\left(C,\delta\right).$$

Слід зауважити насамперед, що якщо $m=n$ і А – невироджена матриця (тобто простір всіх ознак залишається незмінним, змінюються тільки самі ознаки), то з урахуванням виразу (11):

$$I^*\left(C,\delta,A\right) = \left(C^{*-1}\delta^*,\delta^*\right) = \left(A'^{-1}C^{-1}A^{-1}A\delta,A\delta\right) =$$

$$= \left(C^{-1}\delta,\delta\right) = I\left(C,\delta\right). \qquad (12)$$

Це означає, що дивергенція залежить лише від простору ознак, але не залежить від вибору самих ознак. Звідси, очевидним є те, що і для m-вимірних ознак дивергенція також залежить тільки від багатовимірного простору $a_i = \left(a_{i1},...,a_{in}\right)$, $i=1,...,m$, але не від самих цих векторів.

Доведено, що існує таке лінійне перетворення А простору ознак, яке дозволяє отримати вектор нових ознак меншої розмірності, причому

нові ознаки є лінійною комбінацією попередніх, а дивергенція між двома новими розподілами, тобто міра відмінності наявних двох класів об'єктів, дорівнює дивергенції, що базується на попередніх ознаках.

У разі розподілу Бернуллі функціональне відображення ознак є недоцільним, оскільки якщо при відображенні 2^n точок $a=(a_1,...,a_n)$, $a_i=\pm 1$ переходять в різні точки, то дивергенція не змінюється. У цьому випадку доцільним є здійснити відсічення ознак.

Розглянемо вираз для дивергенції (9) у для розподілу Бернуллі. У цьому випадку для будь-яких значень $P_i^{(1)}, P_i^{(2)}$, $i=1,...,n$ кожен доданок суми є додатнім. Звідси випливає, що якщо в разі розподілу Бернуллі зменшувати кількість ознак за допомогою відсікання, то відсікати слід ті ознаки, внесок яких в вираз (9) для дивергенції мінімальний. Тоді при фіксованій кількості нових ознак дивергенція зменшиться мінімально.

Нехай з n ознак з номерами $i=1,...,n$ необхідно залишити лише 5 найбільш інформативних ознак. Оскільки мірою інформативності ми вважаємо дивергенцію, то залишити слід ті ознаки з номерами <<Eqn056.wmf>> для яких вираз

$$I=\left(p_{i_m}^{(1)}-p_{i_m}^{(2)}\right)\ln\frac{p_{i_m}^{(1)}\left(1-p_{i_m}^{(2)}\right)}{p_{i_m}^{(2)}\left(1-p_{i_m}^{(1)}\right)},\ m=1,...,s\ , \quad (13)$$

тобто вклад в дивергенцію приймає найбільші значення. При цьому нова випадкова величина $y=\left(x_{i_1},...,x_{i_s}\right)$ як і раніше буде мати розподіл Бернуллі з тими ж параметрами $p_{i_1}^{(1)},...,p_{i_s}^{(1)}, p_{i_1}^{(2)},....,p_{i_s}^{(2)}$. Нова дивергенція I' буде наступною:

$$I'=\sum_{m=1}^{s}\left(p_{i_m}^{(1)}-p_{i_m}^{(2)}\right)\ln\frac{p_{i_m}^{(1)}\left(1-p_{i_m}^{(2)}\right)}{p_{i_m}^{(2)}\left(1-p_{i_m}^{(1)}\right)}. \quad (14)$$

Оскільки всі складові в виразі (9) є додатними, то існує нерівність $I'\le I$. Таким чином можна зробити висновок, що у випадку розподілу Бернуллі для скорочення кількості ознак краще користуватися методом відсікання, а в разі нормального розподілу – методом лінійного комбінування ознак , який дозволяє зберегти відмінність між класами (дивергенцію) на початковому рівні.

Розглянуті ймовірнісні методи оцінки інформативності не можуть дати однозначну для всіх умов оцінку тих чи інших параметрів і показників технологічного процесу. Залежно від співвідношення різноманітних факторів (геологічний розріз, властивості розбурюваних порід, що впливають на розподіл показників, конструкція свердловини, спосіб буріння та інші), можуть бути більш інформативними ті чи інші ознаки в різних завданнях розпізнавання, тому інформативність ознак необхідно оцінювати для кожної ситуації окремо.

Висновки. Розроблено методику оцінки інформативності ознак за допомогою дивергенції. Сформовано залежності, які дозволяють оцінити інформативність ознак, що підпорядковуються нормальному закону розподілу і закону розподілу Бернуллі.

З метою скорочення кількості ознак при вирішенні завдань розпізнавання ситуацій, запропоновано методику, що ґрунтується на методах відсікання і лінійного перетворення ознак.

Список літератури:

1. Чигур І.І. Особливості контролю технічного стану алмазних доліт в процесі буріння свердловин на нафту і газ. Вимірювальна та обчислювальна техніка в технологічних процесах. 2000. №4. С. 182–184.

2. Чигур І.І. Застосування нейро-фаззі моделювання для розпізнавання нештатних ситуацій, що виникають в процесі буріння свердловин на нафту і газ. Академічний вісник Криворізького територіального відділення Міжнародної Академії комп'ю-терних наук і систем. Кривий ріг. 2004 № 13. С. 78–80.

3. Семенцов Г.Н., Чигур І.І. Основні концепції створення автоматизованої системи контролю за технічним станом породоруйнівного інструменту. Нафтогазова енергетика. 2007. № 1. С. 61–63.

4. Molinari A, Moufki A. The Merchant's model of orthogonal cutting revisited: A new insight into the modeling of chip formation. International Journal of Mechanical Sciences. 2008. № 50. P. 124–131.

Chyhur I.I., Lahoida A.I., Chyhur L.Ya. INFORMATIVE SIGNS IN THE TASKS OF AUTOMATED RECOGNITION OF TECHNOLOGICAL SITUATIONS DURING WELL DRILLING

The article is devoted to the urgent problem of automated recognition and classification of technological situations that arise during the drilling process, as an important stage for making informed decisions in the design and management of this complex technological process.

It is noted that during the drilling process there is a need to identify various factors, such as changes in the drillability and fracturing of rocks, the degree of wear of the drilling tool, as well as geological complications

and accidents. The information obtained as a result of recognition is critically important for effective control of the drilling process, determining the optimal moment for lifting the drilling tool and, what is especially important, for preventing emergency situations.

It is proposed to reduce the task of pattern recognition to the selection of elements belonging to a specific class among a set of elements that may belong to different classes using divergence. It is emphasized that the presence of certain common properties or features in elements of the same class is a key criterion for their identification and assignment to this class.

The important problem of pre-processing and compression of output data in the tasks of recognizing technological situations in order to reduce the dimensionality of the measurement vector is considered.

Two main approaches to reducing the number of features are proposed: selection of the most informative features by cutting off less significant ones and creating a small number of new, combined features based on existing ones. It is emphasized that the second approach, although more laborious, potentially provides higher informativeness of newly created features due to combining the original ones.

The need to minimize the loss of informativeness when selecting or combining features is emphasized, proposing to use the concept of divergence as a measure of informativeness. Further application of divergence to reduce the number of features is proposed using the example of two important cases of data distribution: normal distribution and Bernoulli distribution, with an analysis of both methods of dimensionality reduction (cut-off and linear transformation of features) for each of them.

Thus, the article is an introduction to the important scientific and technical problem of recognition and classification of technological situations in drilling, outlines its relevance, analyzes existing approaches and offers potential information bases for further research in this area.

Key words: *automated recognition, drilling, informative features, divergence, Bernoulli distribution.*

УДК 004.4
DOI https://doi.org/10.32782/2663-5941/2025.3.2/59

Чижмотря О.В.
Державний університет «Житомирська політехніка»

Вакалюк Т.А.
Державний університет «Житомирська політехніка»

Клименко Д.О.
Державний університет «Житомирська політехніка»

Чижмотря О.Г.
Державний університет «Житомирська політехніка»

Гордієнко І.В.
Дрогобицький державний педагогічний університет імені Івана Франка

ДОСЛІДЖЕННЯ СТРУКТУРИ ВЕБ-ТРАФІКУ ТА ВИЗНАЧЕННЯ ВИРІШАЛЬНИХ ХАРАКТЕРИСТИК ЗАПИТІВ

У статті представлено результати комплексного дослідження структури веб-трафіку з метою визначення найбільш інформативних характеристик для ефективного виявлення кіберзагроз та аномальної мережевої активності. В умовах експоненціального зростання обсягів веб-трафіку та еволюції методів кібератак традиційні сигнатурні підходи до забезпечення безпеки стають недостатніми.

Розглянуто багаторівневу архітектуру HTTP/HTTPS запитів, включаючи детальний аналіз заголовків, параметрів URL, тіла запитів, типів методів та протоколів передачі даних. Особливу увагу приділено дослідженню заголовків запитів, відповідей та загальних заголовків, які містять критично важливу інформацію про середовище клієнта та параметри авторизації.

Проведено систематизацію ключових характеристик мережевих потоків, що охоплює часові параметри (тривалість сесій, інтервали між пакетами, швидкість передачі), розмірні характеристики (статистичні показники розмірів пакетів) та контрольні біти TCP. Детально досліджено роль прапорців PSH та URG у контексті ідентифікації трафіку реального часу та виявлення специфічних типів атак.

Особливу увагу приділено аналізу ролі сесійних даних та cookie-файлів у контексті поведінкового моніторингу користувачів та виявлення аномальних змін у поведінкових патернах.

Досліджено дискримінативну здатність різних типів ознак та запропоновано їх трирівневу класифікацію за ступенем важливості для задач машинного навчання. Встановлено, що критичними ознаками є заголовки User-Agent та Authorization, характеристики тіла запитів та частотні параметри HTTP-методів.

Результати створили міцну теоретичну основу для розробки ефективних алгоритмів машинного навчання в галузі кіберзахисту. Практичне значення полягає у можливості оптимізації систем IDS/IPS та створення адаптивних систем кіберзахисту.

Ключові слова: *веб-трафік, HTTP-запити, виявлення аномалій, кіберзагрози, мережева безпека, машинне навчання, аналіз заголовків, класифікація трафіку.*

Постановка проблеми. У сучасному світі використання інтернет-ресурсів є невід›ємною частиною повноцінного існування як окремих користувачів, так і організацій. За останні десятиліття спостерігається експоненціальне зростання обсягів веб-трафіку, що зумовлено стрімким розвитком цифрових технологій, впровадженням хмарних сервісів та загальною цифровізацією суспільства.

Згідно з даними провідних телекомунікаційних компаній, щомісячний глобальний IP-трафік зростає з темпами понад 20% на рік, що створює як нові можливості, так і серйозні виклики для забезпечення кібербезпеки. Паралельно з легі-

тимним трафіком зростає і кількість кібератак: від простих спроб несанкціонованого доступу до складних багатоетапних атак, включаючи DDoS-атаки, SQL-ін'єкції, крос-сайтовий скриптинг (XSS) та атаки типу «людина посередині».

Традиційні методи виявлення загроз, що базуються на сигнатурному аналізі та правилах, вже не можуть ефективно справлятися з сучасними викликами кібербезпеки. Це пов'язано з тим, що зловмисники постійно удосконалюють свої методи, використовують поліморфні техніки та застосовують штучний інтелект для обходу систем захисту. У таких умовах особливої актуальності набувають методи поведінкового аналізу та машинного навчання, які дозволяють виявляти раніше невідомі типи атак на основі аномалій у структурі та характеристиках веб-трафіку.

Аналіз структури веб-трафіку передбачає детальне вивчення HTTP/HTTPS запитів, включаючи заголовки, параметри, тіло запитів, часові характеристики, сесійні дані та метадані мережевих потоків. Кожен з цих компонентів містить цінну інформацію, яка може свідчити про нормальну або аномальну поведінку користувачів. Виявлення та класифікація таких аномалій є критично важливим завданням для своєчасного реагування на потенційні загрози та мінімізації можливих збитків.

Сучасні дослідження в галузі кіберзахисту показують, що ефективність систем виявлення вторгнень значно підвищується при використанні комплексного підходу до аналізу веб-трафіку, який поєднує статистичний аналіз, методи машинного навчання та глибокого навчання. Особливу увагу приділяють виділенню найбільш інформативних ознак (feature selection), що дозволяє підвищити точність класифікації та зменшити обчислювальні витрати.

Аналіз останніх досліджень і публікацій. Сучасні дослідження в галузі аналізу веб-трафіку та виявлення аномалій демонструють стрімкий розвиток методів машинного навчання та глибокого навчання для забезпечення кібербезпеки.

Так, у [1] підкреслюється критичну важливість аналізу структури даних для ідентифікації потенційних загроз. Автори систематизували основні виклики, пов›язані з обробкою великих обсягів гетерогенних даних, та запропонували комплексний підхід до виявлення аномальної поведінки на основі статистичних методів та машинного навчання.

Особливо цінним є дослідження [2], в якому розроблено програмну систему глибокої інспекції пакетів для аналізу мережевого трафіку та виявлення аномалій. Запропонована система використовує багаторівневий підхід до аналізу, що включає детальне вивчення заголовків пакетів, часових характеристик потоків та статистичних параметрів мережевої активності.

Автори [3] зосередилися на застосуванні методів глибокого навчання для прогнозування мережевого трафіку. Їхнє дослідження продемонструвало ефективність рекурентних нейронних мереж та згорткових нейронних мереж у задачах класифікації та прогнозування аномального трафіку.

Науковці [4] запропонували комплексний підхід до проектування та впровадження систем виявлення вторгнень на основі глибоких нейронних мереж. Їхнє дослідження включає детальний аналіз архітектурних рішень та оптимізацію гіперпараметрів для досягнення максимальної ефективності виявлення загроз.

У [5] досліджено вразливості систем виявлення вторгнень на основі глибоких нейронних мереж до адверсаріальних атак. Їхня робота розкриває критичні аспекти безпеки самих систем виявлення та пропонує методи підвищення стійкості до цілеспрямованих спроб обходу захисту.

У [6] представлено інноваційний підхід до розуміння внутрішньої структури глибоких нейронних мереж через концепцію «машини Джейнса». Їхнє дослідження надає теоретичну основу для оптимізації архітектури нейронних мереж у задачах класифікації мережевого трафіку.

У [7] і детально розглянуто математичні основи гауссівського розподілу та його застосування в задачах аналізу даних. Хоча дослідження носить більш теоретичний характер, воно надає важливі дані щодо статистичних методів аналізу аномалій у великих масивах даних.

Аналіз досліджень показує, що сучасні підходи до виявлення аномалій у веб-трафіку характеризуються переходом від традиційних систем до складних гібридних архітектур, що поєднують статистичні методи, машинне навчання та глибоке навчання. Однак залишається недостатньо дослідженим питання систематизації та ранжування характеристик веб-запитів.

Постановка завдання. Мета роботи – дослідити структуру веб-трафіку та визначити найбільш вирішальні характеристики HTTP/HTTPS запитів, які дозволяють ефективно ідентифікувати аномальну поведінку та потенційні кіберзагрози в мережевому трафіку.

Виклад основного матеріалу. Ефективним із способів виявлення потенційних загроз є аналіз структури запитів та вивчення активності користувачів, що дозволяє виявляти відхилення від нормальної поведінки. Це ж і допоможе у навчання моделей та класифікації трафіку.

Веб-запити, які здійснюють користувачі, містять у собі значну кількість інформації, яка може бути використана для визначення нормальної чи аномальної поведінки. Структура цих запитів, типи HTTP запитів, параметри, заголовки та інші метадані можуть мати значний вплив на класифікацію дії. Також під час запитів важливо мати можливість збору та аналізу даних, які містяться у сесійних даних, журналах веб-серверу та cookie-файлів [1].

HTTP-запити є основним елементом, який використовують клієнти для комунікації із серверами [2]. Кожен запит містить важливу інформацію, яка може бути використана для аналізу взаємодії користувачів з вебсайтом. Розберемо структуру заголовків та склад запитів (рис. 1).

Одним із елементів запиту є протокол. Він забезпечує набір правил і стандартів для здійснення комунікації між пристроями в мережі, а саме визначає яким чином передаються дані. У дослідженні роботи використовуються протоколи для інтернет передачі даних, а саме TCP та UPD [2].

Використання HTTP або HTTPS теж важливо для ідентифікації, хоча і в сучасних системах налаштовується переадресація на захищений протокол. Це більш категоріальна змінна, яка не може впливати на те чи буде запит аномальним чи ні, але при потребі може виступати елементом фільтрації, якщо моделі готуються для окремих протоколів.

Посилання містить у собі інформацію для ідентифікацію ресурсу направлення запиту. Саме посилання ідентифікують ресурс, до якого йшов запит. Таким чином при ідентифікації аномалії можна буде визначити що було ціллю зловмисних дій. На аномалію можуть вказувати підозрілі параметри запитів, або часті спроби доступу до конкретного посилання. Збереження даного елементу є важливим для ідентифікації та є частиною моніторингу.

Аналогічна важливість і збереження типу запиту. Якщо це RESTful API, то типи запиту можуть бути GET, POST, PUT, PATCH, DELETE, HEAD, OPTIONS. При виявленні аномалій можна звернути увагу на частоту запитів за певним типом, це може вказувати на DDoS або бутрфорс атаки.

Аналіз заголовків є надзвичайно важливим для вивчення при визначенні аномалій. Хоч вони і не містять дані, але визначають тип оброблення. Заголовки можна поділити на декілька категорій [3]. Заголовки запиту, які клієнт надсилає разом із запитом. Містять усобі інформацію про середовище клієнта, інтерфейс користувача, авторизацію. Заголовки відповіді, які сервер надсилає клієнту разом із відповіддю. Містять у собі

TCP Segment Header Format

Bit #	0 – 7	8 – 15	16 – 23	24 – 31
0	Source Port		Destination Port	
32	Sequence Number			
64	Acknowledgment Number			
96	Data Offset	Res	Flags	Window Size
128	Header and Data Checksum		Urgent Pointer	
160...	Options			

UDP Datagram Header Format

Bit #	0 – 7	8 – 15	16 – 23	24 – 31
0	Source Port		Destination Port	
32	Length		Header and Data Checksum	

Рис. 1. Структура заголовків запиту

інформацію про тип контенту відповіді, статус, кешування. Загальні заголовки можуть бути присутні у запитах та відповідях та містять у собі загальну мета-інформацію.

При дослідженні аномальних значень заголовки можуть вказувати на спроби обходу безпеки, наприклад фальшиві заголовки User-Agent для маскування клієнта або спроби підробити Authorization заголовки для отримання авторизованого доступу. Також про аномалію може свідчити відсутність стандартних заголовків, що може вказувати на використання ботів або заскиртованих сценаріїв. Також на аномалію може вказувати і наявність непотрібних заголовків, що може свідчити про спроби втручання в сесії.

Тіло запиту є надзвичайно важливим для визначення аномальної поведінки, адже саме в ньому зберігається інформація у складних форматах, яку сервер використовує для обробки та збереження. Для аналізу тіла, з метою забезпечення безпеки, звертають увагу на розмір, перевищення або надмірне зменшення його. Також важливим є формат, який передбачений або не передбачений системою. Саме ж наповнення тіла також може мати у собі загрози, якщо містить у собі елементи виконання коду або передаються дані, невідповідні до очікуваних.

У рамках дослідження трафіку важливо звертати увагу не лише на тіло запиту та відповіді, а і на сам потік, який може включати у себе декілька запитів [4]. Потік – це послідовність пакетів, які передаються у мережі між пристроями у межах одного безперервного з'єднання. При аналізі потоку з метою визначення важливих характеристик для ідентифікації аномалій важливо звернути увагу на такі особливості [2].

Основні характеристики потоку, що включає тривалість потоку, загальну кількість пакетів відправлення та отримання, а також загальний розмір пакетів відправлення та отримання.

Часові показники потоку, що включають швидкість передачі байтів та пакетів при відправленні, отриманні та в загальному; часові інтервали між потоками (середній, максимальний, мінімальний, відхилення від середнього), а також часові інтервали між пакетами в обох напрямках окремо та в загальному.

Також при дослідженні трафіку важливо звернути увагу на прапорці, такі як PSH або URG [3]. Ці контрольовані біти впливають на чергу оброблення всіх або окремих пакетів у TCP запитах відповідно. Зазвичай вони використовуються при обміні даних у реальному часі, що може включати відео або голосові дзвінки, а також передачі сигналів на термінове виконання певних команд. Таким чином ці дані можуть мати одне із вирішальних значень при класифікації аномальної поведінки.

Сесійні дані також є важливою складовою для відстеження взаємодії користувача з веб-сайтом. За допомогою них можна зберігати стан користувача впродовж певного часу в різних ситуаціях, таких як авторизація, кошик, налаштування браузера на сайті. Завдяки моніторингу компонентів сесії можна визначити послідовність змін, які можуть впливати на класифікацію дій користувача. Для визначення аномальної поведінки можна слідкувати за такими даними сесії: ідентифікатор сесії, тривалість, активність користувача на сторінках [3; 6].

Аналогічно до сесій також можна слідкувати за зміною файлів cookies. Ці невеликі файли зберігають у собі інформацію про стан сесії, персональні налаштування браузера, часто використовуються для ідентифікації користувачів. Таким чином різка зміна файлів cookies між запитами може слідчити про аномальну активність.

Використовуючи сесії та файли cookies можна зберігати інформацію для моніторингу для подальшого аналізу. Для прикладу можна додати додаткові змінні, які будуть визначати поведінкові характеристики користувача. Це може бути час перебування на сайті, частота входів, певні дії з інтерфейсом вебсайту, які можуть бути вирішальними.

Аналіз структури веб-трафіку, зокрема зазначених вище елементів, можуть дати повну інформацію для ідентифікації класу активності у веб-системі.

Висновки. У результаті проведеного дослідження структури веб-трафіку та визначення вирішальних характеристик для ідентифікації аномальної поведінки було встановлено, що сучасні веб-запити містять багаторівневу ієрархію інформативних елементів, кожен з яких має різну цінність для виявлення потенційних кіберзагроз.

Детальний аналіз структури HTTP/HTTPS запитів показав, що найбільш значущими компонентами для виявлення аномалій є заголовки запитів, які містять критично важливу інформацію про клієнтське середовище, авторизацію та параметри сесії. URL-параметри та тіло запитів дозволяють ефективно виявляти спроби ін›єкцій та несанкціонованої передачі даних, тоді як ано-

мальна частота певних типів HTTP-методів може свідчити про DDoS або брутфорс атаки.

Дослідження ключових характеристик мережевих потоків виявило, що для ефективної класифікації трафіку необхідно враховувати комплекс часових та розмірних параметрів. Часові характеристики, такі як тривалість потоків, інтервали між пакетами та швидкість передачі даних, виявилися особливо інформативними для розпізнавання аномальної активності. Розмірні характеристики пакетів, включаючи їх максимальні, мінімальні та середні значення разом зі статистичними відхиленнями, також надають цінну інформацію для класифікації. Контрольні біти TCP, зокрема PSH та URG, виявилися критично важливими для ідентифікації трафіку реального часу та можуть слугувати індикаторами специфічних типів атак.

Аналіз ролі сесійних даних та cookie-файлів продемонстрував їх високу цінність для поведінкового аналізу користувачів. Моніторинг ідентифікаторів сесій дозволяє ефективно виявляти спроби перехоплення сесій, а аналіз тривалості та активності користувачів на сторінках надає можливість створення детальних поведінкових профілів. Різкі зміни в cookie-файлах між запитами часто індикують аномальну активність та можуть свідчити про компрометацію облікового запису або спроби несанкціонованого доступу.

Список літератури:

1. Min Yang and Jiajie Zhang. Data Anomaly Detection in the Internet of Things: A Review of Current Trends and Research Challenges. International Journal of Advanced Computer Science and Applications (IJACSA), 14 (9), 2023. http://dx.doi.org/10.14569/IJACSA.2023.0140901

2. Song, W., Beshley, M., Przystupa, K., Beshley, H., Kochan, O., Pryslupskyi, A., Pieniak, D., & Su, J. A Software Deep Packet Inspection System for Network Traffic Analysis and Anomaly Detection. *Sensors*, 20 (6), 2020, 1637. https://doi.org/10.3390/s20061637

3. R. Vinayakumar, K. P. Soman and P. Poornachandran. Applying deep learning approaches for network traffic prediction. 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, India, 2017, P. 2353–2358, DOI: 10.1109/ICACCI.2017.8126198.

4. Edosa Osa, Patience E. Orukpe, Usiholo Iruansi. Design and implementation of a deep neural network approach for intrusion detection systems, e-Prime – Advances in Electrical Engineering, Electronics and Energy, Volume 7, 2024, 100434, https://doi.org/10.1016/j.prime.2024.100434

5. Xingwei Zhang, Xiaolong Zheng, Desheng Dash Wu. Attacking DNN-based Intrusion Detection Models. IFAC-PapersOnLine, Volume 53, Issue 5, 2020, P. 415–419, https://doi.org/10.1016/j.ifacol.2021.04.118.

6. Venkat Venkatasubramanian, N. Sanjeevrajan, Manasi Khandekar, Abhishek Sivaram, Collin Szczepanski. Jaynes machine: The universal microstructure of deep neural networks. Computers & Chemical Engineering, Volume 192, 2025, 108908, https://doi.org/10.1016/j.compchemeng.2024.108908.

7. Chen,S. The Gaussian distribution: Derivation method and its applications in selected examples. Theoretical and Natural Science, 38, 2024, P. 196–202. 10.54254/2753-8818/38/20240595

Chizhmotria O.V., Vakaliuk T.A., Klymenko D.O., Chizhmotria O.H., Hordienko I.V. RESEARCH OF WEB TRAFFIC STRUCTURE AND DETERMINATION OF DECISIVE CHARACTERISTICS OF REQUESTS

The article presents the results of a comprehensive study of web traffic structure to identify the most informative characteristics for effectively detecting cyber threats and abnormal network activity. With the exponential growth of web traffic and the evolution of cyberattack methods, traditional signature-based approaches to security are becoming insufficient.

The multi-level architecture of HTTP/HTTPS requests is considered, including a detailed analysis of headers, URL parameters, request bodies, method types, and data transfer protocols. Particular attention is paid to the study of request headers, responses, and general headers that contain critical information about the client environment and authorisation parameters.

The key characteristics of network flows are systematised, covering time parameters (session duration, packet intervals, transmission speed), dimensional characteristics (statistical indicators of packet sizes) and TCP control bits. The role of PSH and URG flags in real-time traffic identification and detection of specific types of attacks has been studied in detail.

Particular attention has been paid to analysing the role of session data and cookie files in monitoring user behaviour and detecting abnormal changes in behaviour patterns.

The discriminative ability of different types of features has been investigated, and a three-level classification of their importance for machine learning tasks has been proposed. It has been established that the critical features are User-Agent and Authorization headers, request body characteristics and HTTP method frequency parameters.

The results have created a solid theoretical basis for the development of effective machine learning algorithms in cyber security. The practical significance lies in the possibility of optimising IDS/IPS systems and creating adaptive cyber security systems.

Key words: *web traffic, HTTP requests, anomaly detection, cyber threats, network security, machine learning, header analysis, traffic classification.*

УДК 378.147:004.9:51
DOI https://doi.org/10.32782/2663-5941/2025.3.2/60

Ярецька Н.О.
Хмельницький національний університет

Рамський А.О.
Хмельницький національний університет

Поплавська О.А.
Хмельницький національний університет

ОСОБЛИВОСТІ ВИКЛАДАННЯ ВИЩОЇ МАТЕМАТИКИ СТУДЕНТАМ ГАЛУЗІ ЗНАНЬ «ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ» В УМОВАХ ЦИФРОВІЗАЦІЇ ОСВІТИ

Стаття присвячена питанням удосконалення методики викладання вищої математики для студентів галузі знань «Інформаційні технології» в умовах цифровізації освіти, що супроводжує трансформацію освітнього простору України. Виявлено, що зменшення кількості аудиторних годин, передбачених на вивчення вищої математики, в поєднанні зі складністю абстрактного змісту навчального матеріалу, формує виклик для викладачів щодо ефективної організації освітнього процесу, а для студентів – ускладнює глибоке опанування базових математичних понять. Проаналізовано вітчизняний та зарубіжний досвід використання цифрових освітніх технологій, який свідчить про зростаюче значення адаптивного навчання, автоматизованих систем оцінювання знань, мультимедійних платформ та масових відкритих онлайн-курсів, зокрема для викладання дисциплін з високим рівнем теоретичної абстракції.

Оцінено потенціал використання сучасних програмних продуктів (GeoGebra, Mathematica, Maple, MathCAD), що дозволяють в інтерактивному форматі здійснювати математичні побудови, моделювання та аналітичні обчислення. Обґрунтовано ефективність змішаних моделей навчання, які поєднують переваги дистанційного і традиційного форматів, сприяючи індивідуалізації освітньої траєкторії, підвищенню мотивації та засвоєнню складного математичного матеріалу. Водночас акцентовано на необхідності подолання викликів цифровізації, зокрема низького рівня цифрової грамотності частини учасників освітнього процесу, нестачі технічних ресурсів, недостатньої адаптації навчальних платформ до інклюзивного середовища, а також зменшення емоційної взаємодії у процесі онлайн-навчання. Розроблено модель інтерактивного навчання з використанням мультимедійних презентацій, генераторів завдань та хмароорієнтованих сервісів, що дозволяє реалізувати концепцію цифрової трансформації математичної освіти у технічному університеті. Запропоновано комплексні заходи щодо інтеграції цифрових платформ у навчальний процес, включаючи підготовку педагогів, забезпечення технічної підтримки, організацію супровідного навчання для здобувачів та постійний моніторинг результативності використаних технологій.

Ключові слова: *вища математика, цифровізація освіти, інформаційні технології, змішане навчання, цифрові платформи, математична підготовка.*

Постановка проблеми. В реаліях трансформації освітнього простору України, що визначається активною фазою цифровізації всіх ланок вищої освіти, особливої актуальності набуває питання удосконалення методики викладання базових дисциплін, зокрема вищої математики, для студентів галузі знань «Інформаційні технології». З огляду на стрімкий розвиток цифрових технологій, змінюється не лише структура навчального процесу, але й самі підходи до формування змісту та організації освітніх програм, що орієнтовані на підготовку фахівців ІТ-сфери. У цьому контексті варто зазначити, що зменшення кількості аудиторних годин, відведених на вивчення вищої математики, при збереженні значного обсягу абстрактного, складного за змістом матеріалу, створює серйозні виклики як для викладачів, так і для здобувачів освіти [8].

В умовах цифрової трансформації української вищої школи особливо важливим є не лише забезпечення технічної готовності університетів до впровадження інновацій, але й методичне

переосмислення традиційної структури лекційних і практичних занять з вищої математики [6]. Дослідницький інтерес викликає розробка ефективних педагогічних моделей, які б ураховували як специфіку математичних знань, так і потреби сучасних студентів, що навчаються в динамічному, цифрово орієнтованому середовищі. Саме тому поточне дослідження присвячене визначенню ключових особливостей викладання вищої математики студентам ІТ-спеціальностей в Україні в умовах цифровізації освіти, з акцентом на інтеграцію сучасних інформаційних технологій у навчальний процес як чинника підвищення якості математичної підготовки.

Аналіз останніх досліджень і публікацій. У працях Гуржій А. М., Глазунова О. Г., Волошина Т. В. [2], Суліма О. [13] та Чака К. [17] акцентовано на тому, що цифрові освітні технології суттєво розширюють можливості персоналізованого навчання, надаючи як викладачам, так і студентам доступ до адаптивних навчальних ресурсів, інтерактивних інструментів та онлайн-лабораторій, що особливо актуально для дисциплін з високим рівнем абстракції, таких як вища математика. Проте автори справедливо вказують і на певні обмеження цифровізації, серед яких ключовими залишаються потреба у підвищенні цифрової грамотності всіх учасників освітнього процесу, а також технічна готовність інфраструктури закладів до ефективного функціонування цифрових середовищ.

Значну увагу у роботах Клєопа І., Тютюнник О., Крупський Я., Добранюк Ю. [4], Сиваш С., Соколовська Г. [12], приділено аналізу переваг онлайн-курсів з математичних дисциплін, акцентуючи на важливості побудови адаптивних освітніх траєкторій з використанням штучного інтелекту та елементів машинного навчання, які здатні автоматично коригувати складність матеріалу залежно від індивідуального прогресу студента. На їхню думку, саме таке поєднання технологій забезпечує якісно новий рівень викладання складних теоретичних курсів, включаючи вищу математику.

Так Насонова С. [9], у своєму дослідженні розглянула інклюзивний потенціал масових відкритих онлайн-курсів, що відкривають нові можливості для студентів з особливими освітніми потребами. Автори стверджують, що цифрові платформи дозволяють здійснювати повноцінне математичне навчання без потреби фізичної присутності в аудиторії, однак вказують і на недоліки веб-доступності, які потребують вдосконалення інтерфейсів і підходів до подачі контенту. Інше дослідження, представлені авторством Ковальчук М. [5] та Сікора Я., Скоробагатська О., Лиходєєва Г., Максименко А., Цехмістер Ю. [19], підтверджують, що для успішної підготовки майбутніх ІТ-фахівців надзвичайно важливою є не лише технічна, але й методична готовність до використання цифрових інструментів, які, своєю чергою, сприяють формуванню математичного мислення, навичок дистанційної взаємодії та самостійного здобуття знань. Загалом аналіз літератури дає підстави стверджувати, що цифровізація навчального процесу у вищій освіті, зокрема в частині математичної підготовки студентів галузі знань «Інформаційні технології».

Постановка завдання. Метою статті є дослідження цифрового інструментарію забезпечення навчального курсу вищої математики для студентів галузі знань «Інформаційні технології» та розробка заходів направлених на поєднання цифрових платформ та навчального процесу.

Виклад основного матеріалу. У контексті цифрової трансформації вищої освіти України, що зумовлює необхідність глибокого переосмислення традиційних підходів до викладання вищої математики для студентів спеціальності «Інформаційні технології», принципового значення набуває впровадження сучасних цифрових інструментів, які забезпечують ефективну візуалізацію та інтенсифікацію навчального процесу. Одним із напрямів реалізації цього підходу є використання спеціалізованого програмного забезпечення для створення мультимедійного супроводу лекційних курсів, серед якого найбільш поширеним і доступним для вищих навчальних закладів залишається Microsoft PowerPoint, що входить до складу пакета Microsoft Office та забезпечує можливість інтеграції графічних, анімаційних і відеоелементів з метою наочного представлення абстрактних математичних понять [10, с. 135]. Використання таких інструментів як відеофрагменти, анімації формул та динамічні ілюстрації дозволяє не лише підвищити зацікавленість студентів у змісті навчального матеріалу, але й значно спрощує сприйняття складних логіко-математичних побудов, з якими традиційно виникають труднощі в умовах скорочення аудиторного часу та слабкої математичної підготовки, характерної для частини студентів ІТ-спеціальностей (рис. 1) [11].

Що ж стосується практичних занять, то в умовах цифровізації освіти дедалі більшої популярності набувають програмні продукти

освітнього призначення, зокрема GeoGebra, MathCAD, Mathematica, Maple, Cabri Geometry, GEONExT, C.a.R., які дозволяють не лише виконувати складні обчислення та графічні побудови, а й інтерактивно моделювати процеси, що лежать в основі математичних залежностей. Серед них особливо виділяється безкоштовна й інтуїтивно зрозуміла програма GeoGebra, яка дозволяє візуалізувати зміну геометричних об›єктів у реальному часі при варіюванні параметрів, що суттєво сприяє глибшому розумінню структури та властивостей математичних моделей, особливо у розділах аналітичної геометрії, математичного аналізу та лінійної алгебри (рис. 2) [3, с. 15–16].

Інтеграція цифрових платформ у викладання вищої математики дозволяє суттєво покращити якість засвоєння складного абстрактного матеріалу, водночас сприяючи активізації самостійної роботи студентів, розширенню доступу до навчальних ресурсів і розвитку цифрових навичок, що є критично важливими для майбутніх ІТ-фахівців [1]. Застосування онлайн-курсів, електронних навчальних модулів, хмарних сервісів та мультимедійних елементів забезпечує

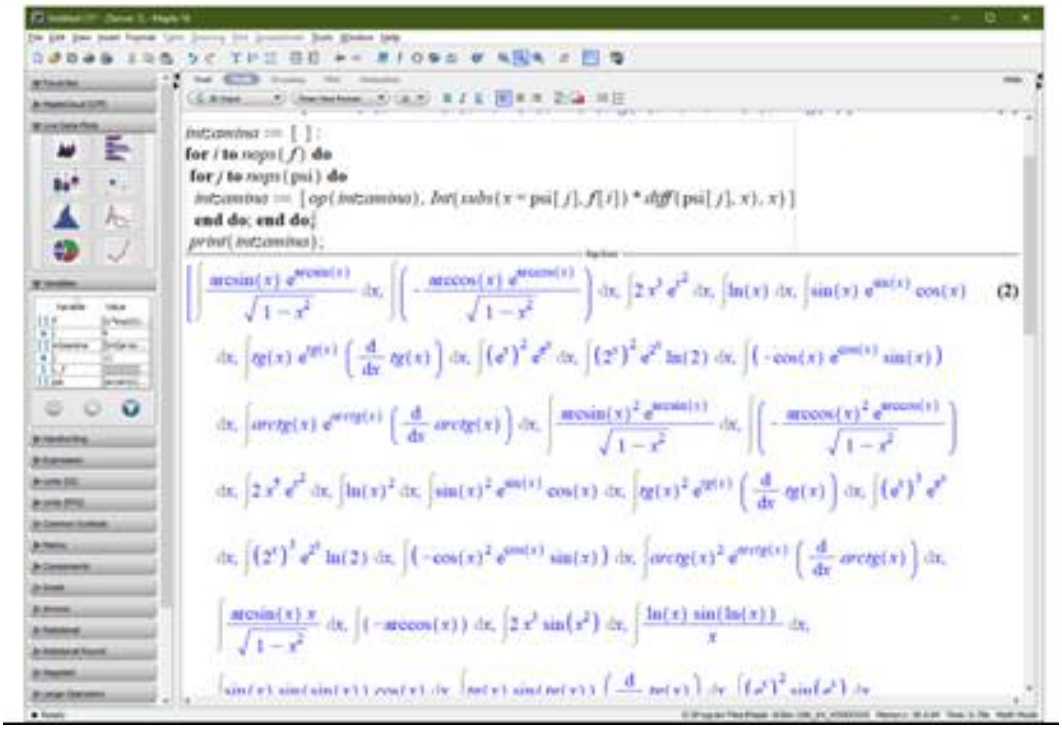


Рис. 1. Вікно роботи генератора завдань з вищої математики (сформовано авторами)

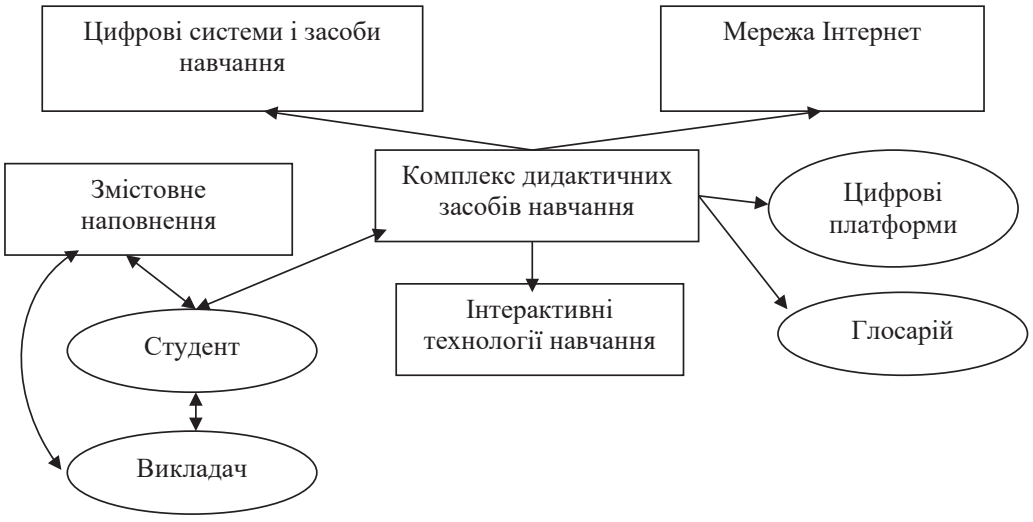


Рис. 2. Модель інтерактивного навчання вищої математики з використанням цифрових технологій (розроблено авторами)

не лише візуалізацію математичних понять, а й адаптацію навчального контенту до індивідуального рівня підготовки кожного студента (рис. 3) [18, с. 51].

Особливо ефективним виявляється використання змішаних моделей навчання, у яких цифрові інструменти органічно доповнюють класичні форми проведення лекцій та практичних занять, але всі вони мають свої переваги і недоліки (табл. 1) [15, с. 45].

Разом з тим, успішна реалізація потенціалу цифрового навчання зумовлена необхідністю комплексного підходу, який враховує не лише технічне забезпечення й організаційні аспекти освітнього середовища, а й рівень цифрової компетентності як викладачів, так і студентів. Одним із найважливіших завдань є адаптація існуючих методичних розробок до нових форматів подачі матеріалу, а також забезпечення належного рівня підготовки педагогів до роботи з цифровими платформами (табл. 2) [7, с. 51].

Незважаючи на численні переваги цифровізації, такі як персоналізація навчального процесу, автоматизований контроль знань, широке охоплення ресурсів та інтерактивність, існують і певні обмеження, які слід брати до уваги. До таких належать, зокрема, обмежений доступ до якісного інтернету або цифрової техніки в окремих регіонах України, недостатній рівень цифрової грамотності частини педагогічного складу, а також потреба у формуванні нових педагогічних стратегій, які б відповідали викликам цифрового навчального середовища. Крім того, у процесі цифрового навчання часто фіксується зниження рівня емоційної залученості та міжособистісної взаємодії між викладачем і студентом, що може негативно впливати на глибину опанування складних тем, особливо таких як математичний аналіз, теорія ймовірностей чи дискретна математика [16]. Отже, ключовим завданням у впровадженні цифрових освітніх технологій у викладання вищої математики є пошук балансу між технологічними інноваціями та збереженням педагогічної ефективності живого спілкування.

Водночас не слід повністю відмовлятися від традиційних технологій навчання, які, попри відсутність цифрових елементів, мають низку безперечних переваг, зокрема системність і логічну послідовність подачі матеріалу, чітко організовану структуру занять, емоційний вплив особистості викладача на навчальну мотивацію студентів, а також економічну ефективність у контексті масового навчання [14, с. 63]. Тому оптимальним підходом до викладання вищої математики студентам ІТ-спеціальностей в умовах цифровізації освіти є гармонійне поєднання новітніх цифрових інструментів із традиційними методами педагогічного впливу, що дозволяє максимально ефективно формувати математичну компетентність майбутніх фахівців у сфері інформаційних технологій.

Висновки. Вивчено ключові фактори, що впливають на ефективність цифрового навчання, зокрема застосування гібридних форм освітнього процесу та виділено сильні та слабкі сторони застосування цифрових технологій у викладанні вищої математики для студентів спеціальностей у сфері інформаційних технологій. Досліджено ефективність використання цифрових платформ у викладанні математичних дисциплін та підтверджено їхню здатність адаптувати матеріали до рівня підготовки кожного студента, забезпечувати оперативний зворотний зв'язок і сприяти реалізації принципів інклюзивної освіти. Обґрунтовано переваги змішаних (гібридних) моделей навчання, де цифрові інструменти поєднуються з традиційними методами. Такий підхід дозволяє зберігати важливу роль педагога, одночасно

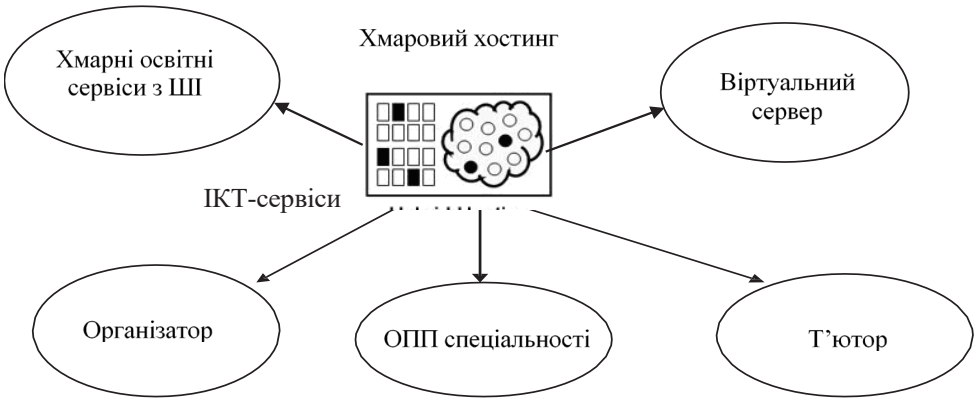


Рис. 3. Інтегрована гібридна модель хмароорієнтованого навчально-дослідного середовища закладу вищої освіти в умовах цифровізації (розроблено авторами)

Таблиця 1

Переваги та недоліки використання цифрових платформ у викладанні вищої математики спеціальності «Інформаційні технології» (сформовано авторами)

Переваги	Недоліки
Забезпечення доступності навчальних ресурсів: Цифрові платформи надають студентам ІТ-напрямів доступ до матеріалів з вищої математики в будь-який час і з будь-якого місця, що сприяє ефективному дистанційному навчанню	Обмеження прямого спілкування: Онлайн-формат може ускладнювати взаємодію між викладачем і студентами, що впливає на засвоєння складних математичних концепцій
Гнучкість у плануванні навчання: Студенти можуть самостійно регулювати темп вивчення, повторювати складні розділи вищої математики або пропускати вже засвоєні теми	Технічні виклики: Стабільне інтернет-з'єднання та сучасне обладнання є обов'язковими; можливі збої в роботі платформ
Персоналізація навчального процесу: Адаптивні системи враховують рівень знань студентів ІТ-спеціальностей, пропонуючи індивідуальні завдання з вищої математики	Високі вимоги до самодисципліни: Студенти повинні бути мотивованими та організованими для самостійного вивчення математичних дисциплін
Інтерактивність та залученість: Використання відеолекцій, симуляцій і тестів підвищує інтерес до вищої математики та мотивацію студентів ІТ-напрямів	Необхідність освоєння нових технологій: Викладачі та студенти можуть стикатися з труднощами при вивченні цифрових інструментів для навчання математики

Таблиця 2

Рекомендації щодо ефективного впровадження цифрових платформ у викладанні вищої математики спеціальності «Інформаційні технології» (сформовано авторами)

Рекомендація	Опис	Приклад цифрової платформи
1. Застосування змішаних моделей навчання	Поєднання традиційних методів викладання вищої математики з цифровими інструментами для студентів ІТ-напрямів забезпечує гнучкість, персоналізацію та оптимальні результати. Змішане навчання дозволяє інтегрувати очні заняття з дистанційними форматами.	Khan Academy: Некомерційна освітня платформа, що пропонує безкоштовні курси з вищої математики, інтерактивні вправи та відеоуроки. Підтримує персоналізоване навчання та адаптацію до рівня студента
2. Забезпечення технічної підтримки	Надання студентам і викладачам доступу до обладнання, програмного забезпечення та оперативної технічної підтримки для безперебійного використання цифрових платформ у навчанні вищої математики.	Coursera: Міжнародна платформа з курсами від провідних університетів світу, включаючи спеціалізовані програми з математики для ІТ-напрямів. Надає доступ до лекцій, завдань і сертифікатів
3. Навчання викладачів і студентів	Організація тренінгів і семінарів з використання цифрових інструментів для підвищення цифрової грамотності викладачів і студентів ІТ-напрямів, що сприяє ефективному освоєнню вищої математики.	edX: Платформа, створена MIT і Гарвардом, що пропонує курси з вищої математики, програмування та інших дисциплін. Підтримує інтерактивні лабораторії та інструменти для змішаного навчання.
4. Постійний моніторинг і оцінка ефективності	Регулярний аналіз прогресу студентів і коригування навчального процесу з вищої математики на основі даних цифрових платформ для підвищення якості освіти в умовах цифровізації.	Wolfram Alpha: Обчислювальна платформа, що дозволяє розв'язувати складні математичні задачі, створювати візуалізації та перевіряти рішення. Широко використовується студентами ІТ-напрямів для роботи з алгеброю, аналізом і статистикою

використовуючи потенціал сучасних технологій для підвищення якості освіти.

Розроблено загальну модель інтерактивного цифрового навчання вищої математики, яка передбачає застосування мультимедійних презентацій, комп'ютерних симуляцій, адаптивних навчальних систем та віртуальних лабораторій разом із класичними формами проведення лекцій і практичних занять. Запропоновано конкретні рекомендації щодо ефективного впровадження цифрових технологій у процес викладання математики студентам ІТ-спеціальностей: організація спецкурсів для педагогів з оволодіння цифровими інструментами; створення служб технічної підтримки у закладах вищої освіти; регулярний аналіз результативності використовуваних цифрових рішень; адаптація навчального контенту для студентів з особливими освітніми потребами; розвиток цифрової інфраструктури закладів.

Список літератури:

1. Геревенко А. М., Ільїна Т. В., Ібрагімова Л. А. Використання цифрових платформ для підвищення якості професійної освіти. *Академічні візії.* 2024. № 31. DOI: https://doi.org/10.5281/zenodo.11442893
2. Гуржій А. М., Глазунова О. Г., Волошина Т. В. Цифровий навчальний контент для системи відкритої освіти: Сучасні інформаційні технології та інноваційні методики навчання у підготовці фахівців: методологія, теорія, досвід, проблеми: Київ-Вінниця: ТОВ фірма «Планер», 2020. 268 с.
3. Кіяновська Н. М., Рашевська Н. В. Аналіз, особливості та вимоги до деяких методів навчання вищої математики студентів технічних ВНЗ. *Теорія та методика навчання математики, фізики, інформатики.* 2015. Т. 13, № 2. С. 9–22. DOI: https://doi.org/10.55056/tmn.v13i2.762
4. Клєопа І. А., Тютюнник О. І., Крупський Я. В., Добранюк Ю. В. Особливості використання сучасних інформаційно-комунікаційних технологій у вищій математичній освіті. Modern Information Technologies and Innovation *Methodologies of Education in Professional Training Methodology Theory Experience Problems.* № 72. С. 113–124. https://doi.org/10.31652/2412-1142-2024-72-113-124
5. Ковальчук М. Б. Професійна спрямованість навчання математики як інтеграційна основа фахової підготовки студентів інженерних спеціальностей : монографія. Вінниця : ВНТУ, 2020. 348 с.
6. Македон В. В., Валіков В. П., Кошляк Є. Є. Світовий ринок праці в координатах цифрової економіки. *Академічний огляд.* 2020. № 1 (52). С. 91-107. DOI: 10.32342/2074-5354-2020-1-52-9.
7. Мар'єнко М., Коваленко В. Штучний інтелект та відкрита наука в освіті. *Фізико-математична освіта.* 2023. Том 38. № 1. С. 48–53. DOI: 10.31110/2413- 1571-2023-038-1-007.
8. МОН України. Концепція цифрової трансформації освіти і науки: МОН запрошує до громадського обговорення. URL: https://mon.gov.ua/ua/news/koncepciya-cifrovoyi-transformaciyi-osviti-i-nauki-monzaproshuye-dogromadskogo-obgovorennya
9. Насонова С. Сучасний підхід до викладання математичних дисциплін у вищих навчальних закладах. *Перспективи Та Інновації Науки.* 2022. № 1 (6). URL: https://doi.org/10.52058/2786-4952-2022-1(6)- 284-293
10. Новицкая Л. М. Математична підготовка майбутніх бакалаврів із комп'ютерних наук в аграрному університеті. *Молодь і ринок.* 2024. № 9 (229). С. 133–137. DOI: https://doi.org/10.24919/2308-4634.2024.300096.
11. Потапова О. М., Олексенко В. М., Петрина Я. Д. Сучасні підходи до викладання математичних дисциплін у вищій школі. *Інноваційна педагогіка.* 2025. Вип. 79, Т. 2. С. 154–160. DOI: https://doi.org/10.32782/2663-6085/2025/79.2.30
12. Сиваш С. В., Соколовська Г. В. Особливості викладання вищої математики у технічному університеті в сучасних умовах. *Наука і техніка сьогодні.* 2023. № 9 (23). С. 437–450. DOI: https://doi.org/10.52058/2786-6025-2023-9(23)-437-450
13. Суліма О. В. Особливості викладання вищої математики для студентів інженерних спеціальностей в технічному університеті в умовах війни. *Матеріали Міжнародної науково-практичної конференції «Актуальні проблеми сучасної науки, освіти та технологій».* Рига: Baltija Publishing, 2025. С. 203–208. DOI: https://doi.org/10.30525/978-9934-26-522-8-48 .
14. Чорна І., Мельничук Л., Татарин О. Використання інтерактивних технологій в освітньому процесі під час викладання природничо-математичних дисциплін. *Молодий Вчений.* 2022. № 6 (106). С. 61–64. URL: https://doi.org/10.32839/2304-5809/2022-6- 106-13
15. Шишенко І. Деякі аспекти впливу цифрових технологій на освітній процес закладів вищої освіти: огляд проблем та викликів. *Освіта. Інноватика. Практика.* 2022. № 10 (5). С. 42–47. DOI: 10.31110/2616-650X-vol10i5-006
16. Bobro N. The use of artificial intelligence in the organization of the educational process in a digital educational environment. *Social Science and Humanities Journal.* 2024. № 08 (02). P. 34586–34589. DOI: 10.18535/sshj.v8i03.945.
17. Chaka C. Fourth industrial revolution–a review of applications, prospects, and challenges for artificial intelligence, robotics and blockchain in higher education. *Research and Practice in Technology Enhanced Learning (RPTEL).* 2023. 18 (2). URL: http://rptel.apsce.net/index.php/RPTEL/article/view/2023-18002.
18. Makedon V., Myachin V., Plakhotnik O., Fisunenko N., Mykhailenko O. Construction of a model for evaluating the efficiency of technology transfer process based on a fuzzy logic approach. *Eastern-European Journal of Enterprise Technologies.* 2024. No 2 (13 (128)). P. 47–57. https://doi.org/10.15587/1729-4061.2024.300796.
19. Sikora Ya., Skorobahatska O., Lykhodieieva H., Maksymenko A., Tsekhmister Y. Informatization and digitization of the educational process in higher education: main directions, challenges of the time. *Revista Eduweb.* 2023. Vol. 17. No 2. P. 244–256. DOI: https://doi.org/10.46502/issn.1856-7576/2023.17.02.21.

Yaretska N.O., Ramskyi A.O., Poplavska O.A. **FEATURES OF TEACHING HIGHER MATHEMATICS TO STUDENTS OF THE FIELD OF KNOWLEDGE "INFORMATION TECHNOLOGY" IN THE CONDITIONS OF DIGITALIZATION OF EDUCATION**

The article is devoted to the issues of improving the methodology of teaching higher mathematics for students of the field of knowledge "Information Technologies" in the conditions of digitalization of education, which accompanies the transformation of the educational space of Ukraine. It was found that the reduction in the number of classroom hours provided for the study of higher mathematics, combined with the complexity of the abstract content of the educational material, creates a challenge for teachers to effectively organize the educational process, and for students – complicates the in-depth mastery of basic mathematical concepts. The omestic and foreign experience of using digital educational technologies is analyzed, which indicates the growing importance of adaptive learning, automated knowledge assessment systems, multimedia platforms and massive open online courses, in particular for teaching disciplines with a high level of theoretical abstraction. The potential of using modern software products (GeoGebra, Mathematica, Maple, MathCAD), which allow for interactive mathematical constructions, modeling and analytical calculations, is assessed. The effectiveness of blended learning models that combine the advantages of distance and traditional formats is substantiated, contributing to the individualization of the educational trajectory, increasing motivation and mastering complex mathematical material. At the same time, the need to overcome the challenges of digitalization is emphasized, in particular, the low level of digital literacy of some participants in the educational process, the lack of technical resources, insufficient adaptation of educational platforms to an inclusive environment, as well as the reduction of emotional interaction in the process of online learning. A model of interactive learning using multimedia presentations, task generators and cloud-based services has been developed, which allows implementing the concept of digital transformation of mathematical education at a technical university. Comprehensive measures are proposed to integrate digital platforms into the educational process, including teacher training, providing technical support, organizing accompanying training for applicants and constant monitoring of the effectiveness of the technologies used.

Key words: *higher mathematics, digitalization of education, information technologies, blended learning, digital platforms, mathematical training.*

БУДІВНИЦТВО

УДК 624.131.32:622.235.8
DOI https://doi.org/10.32782/2663-5941/2025.3.2/61

Деревягіна Н.І.
Національний технічний університет «Дніпровська політехніка»

Садовенко І.О.
Національний технічний університет «Дніпровська політехніка»

ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ ЗМІН ПАРАМЕТРІВ СТІЙКОСТІ ЛЬОСОВИХ МАСИВІВ ПРИ ВИБУХОВИХ НАВАНТАЖЕННЯХ

Мета. Встановити закономірності змін фізико-механічних властивостей льосових суглинків Придніпров'я під впливом вибухового навантаження в умовах техногенно зміненого геологічного середовища для оцінки стійкості та прогнозування поведінки ґрунтів при подальшого інженерному освоєнні територій.

Методика. Проведено серію лабораторних експериментів на зразках льосових суглинків, відібраних з правобережної частини міста Дніпро. Зразки проаналізовано за їх фізичними параметрами (щільність, вологість, пористість), а також за гранулометричним та хімічним складом. Випробування здійснювались з використанням стабілометра для визначення параметрів міцності до та після вибухового навантаження. Аналізувалися також тиксотропні зміни ґрунтів протягом 35 діб після вибуху.

Результати. Встановлено закономірності зміни міцності у зоні, ущільненої вибухом, де у зоні максимального впливу спостерігалося зростання питомого зчеплення у 2–2,5 рази, яке поступово зменшувалося до природних значень на відстані 13–14 м, а потім – знижувалося нижче за початковий рівень. Протягом 6–7 днів після вибуху відбувалося зростання зчеплення на 25–30%, що підтверджує прояв тиксотропії. Кут внутрішнього тертя змінювався у перші 5–7 діб, після чого стабілізувався.

Наукова новизна. Вперше експериментально встановлено зміни фізико-механічних характеристик льосових суглинків Придніпров'я внаслідок деградації під дією вибухового навантаження. Визначено залежність зміни зчеплення та кута внутрішнього тертя від часу після навантаження, що дозволяє обґрунтувати механізм деградації масиву в умовах техногенного впливу.

Практична значимість. Одержані експериментальні результати можуть бути використані при проектуванні, будівництві та експлуатації інженерних споруд на територіях, що знаходяться під впливом вибухових навантажень. Врахування змін фізико-механічних властивостей грунтового масиву дозволяє підвищити безпеку та довговічність об'єктів будівництва, зокрема на специфічних льосових ґрунтах.

Ключові слова: *льосові суглинки, вибухове навантаження, зчеплення, кут внутрішнього тертя, ущільнення, техногенез, Придніпров'я.*

Постановка проблеми. Необхідність забезпечення стійкості схилів, споруд, оцінки міцності та несучої здатності масивів порід, порушених вибуховим навантаженням, пов'язані зі змінами параметрів грунтів у часі та просторі. Особливої актуальності проблема забезпечення стійкості геологічного середовища набуває саме у регіонах із широким розповсюдженням специфічних за своїми властивостями льосових порід, наприклад у Придніпров'ї. Під дією динамічного навантаження структура льосових порід зазнає значних змін, іноді – до повного руйнування їх мікро- та макропористої структури, нерівномірного зниження фільтраційних властивостей, маючи специфічну форму деградації в цілому. Наслідком цього може бути активізація негативних інженерно-геологічних процесів, таких як зсуви, осідання, просідання та в цілому зниження довготривалої стабільності інженерних споруд. Деякі теоретичні та емпіричні передумови до прове-

дення серії лабораторних досліджень льосових порід, характерних для Придніпров'я, висвітлені в роботах [1–4], де позначено суттєво новий підхід у розумінні механізму геодинаміки цих ґрунтів за сучасних умов техногенного навантаження.

Аналіз останніх досліджень і публікацій. Проблематика зміни властивостей порід при дії динамічних навантажень (а саме – вибухового генезису), висвітлена у низці наукових робіт, однак в більшості з них увага зосереджена на піщаних та частково глинистих породах, а об'ємне динамічне стиснення тих же льосових порід мало описане у таких умовах. В попередніх дослідженнях [5] недостатньо уваги приділено вивченню фізико-механічних параметрів льосових грунтів різних типів та встановлення кореляційних зв'язків процесів об'ємного деформування з відповідними параметрами ґрунтів, а саме речовинним та гранулометричним складом, пористістю або ступенем водонасичення. Були розглянуті окремі питання порушеності льосових порід [4–6], запропоновано індуктивні моделі деградації, їх районування, однак без врахування впливу вибухового імпульсу та розподілу його у масиві. Сучасні дослідження в галузі геології та геомеханіки [5, 7] демонструють, що короткочасні імпульсні навантаження, якими по суті є вибухові, викликають локальні зони руйнування структури в пористих масивах, що істотно впливає на їхню несучу здатність. На території правобережної України, багато прикладних завдань, пов'язаних із цим, реалізується саме у льосових грунтах [3, 8–9]. Враховуючи вищезазначене, дослідження комплексу гідрогеодинамічних, енергетичних і геомеханічних критеріїв деградації льосових структур з урахуванням їх техногенезу в умовах динамічного навантаження, дозволить отримати принципово нові наукові результати щодо їх стійкості.

Постановка завдання. Тому основною метою цього дослідження є встановлення закономірностей, що характеризують діапазон змін параметрів деградації грунтового льосового масиву внаслідок динамічних (а саме – вибухових) навантажень з урахуванням їх техногенезу в різних умовах, що забезпечить комплексний підхід у вирішенні цієї актуальної проблеми і дозволить отримати нові наукові результати щодо механізма їх руйнування.

Виклад основного матеріалу. Територія досліджень розташована на терасованій правобережній частині міста Дніпро, у його південній частині, поблизу Старих Кодаків, і представлена осадовою товщою антропогенових та верхньопліоценових відкладів, що залягають на кристалічних породах та їх корі вивітрювання, а саме каолінах та дресві, та складаються з пісків полтавської серії, строкатих та червоно-бурих глин та льосової серії.

Верхня частина розрізу складається з буровато-пальових та світло-бурих крупнопильоватих льосових суглинків із включеннями карбонатів. Їх потужність змінюється в межах 3–5 м. Нижче залягає жовтувато-пальовий крупнопилуватий льосовий суглинок з пропластками викопних грунтів, що вказують також на ранні стадії формування ландшафту на цій території. Ці породи карбонатизовані, слабоструктуровані та мають тиксотропні властивості при зволоженні.

На глибині 6–9 м фіксується жовтувато-пальовий легкий суглинок, що у нижній частині переходить у супісок. Цей перехідний горизонт має підвищену водопроникність і знижені міцнісні властивості, фактично створюючи ослаблену зону. Нижче за розрізом залягає прошарок глин потужністю до 3 м та пісок із включенням каоліну. Ці породи алювіального походження, мають середню щільність та знижений вміст органічних домішок. Під алювіальними відкладами залягає тріщинувата кора вивітрювання кислих кристалічних порід, яка створює нижній структурний ярус розрізу.

В цілому, в геологічній будові району досліджень спостерігається значна літологічна неоднорідність, строкате перешарування водопроникних та водотривких горизонтів, де особливої уваги потребують просадні льосові суглинки, оцінка стійкості яких актуальна для прогнозу деформаційних процесів в умовах вибухового впливу.

У якості зразків для експериментів було відібрано різновиди льосового суглинку з наступними параметрами: об'ємна вага 1,8–1,95 г/см^3, об'ємна вага скелету 1,7–1,74 г/см^3, пористість 35–36%, вологість (вагова) 13–15,5, також було проведено гранулометричний аналіз відібраної проби (рис. 1).

Ці льосові різновиди мають переважно кремнеземисто-глинистий склад з домішками карбонатів та оксидів металів. Основним компонентом є кремнезем (SiO_2), вміст якого змінюється від 55 до 65%. Значна частка оксидів алюмінію (Al_2O_3) становить 10–15%, а оксиди заліза (Fe_2O_3) – 4–7%. Карбонати, представлені в основному оксидом кальцію (CaO) та оксидом магнію (MgO), присутні у кількості 3–7% і 1–3%

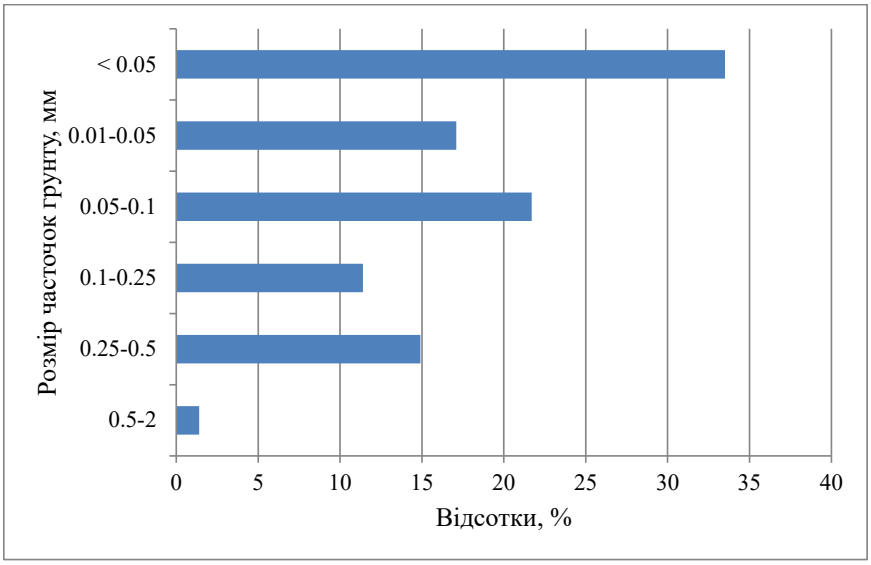


Рис. 1. Результати гранулометричного аналізу відібраних зразків льосового суглинку

відповідно. Вміст лужних елементів, таких як оксид калію (K_2O) та оксид натрію (Na_2O), становить 2–4% і 1–2% відповідно. Втрата при прожарюванні, що включає карбонати, органічні речовини та гігроскопічну воду, може досягати 8–15% [3, 8–9]. Наявність карбонатів свідчить про помірну стійкість до зволоження льосових суглинків.

Вологість грунтів є одним з основних факторів, які визначають їх стан при навантаженні, особливо це стосується льосових порід, специфічні властивості яких ускладнюють прогноз їх поведінки як у природних умовах, так і при техногенному навантаженні. В залежності від характеру навантаження – статичного або динамічного, вплив вологості на деформування грунту відрізняється у кількісному та якісному відношенні, що пояснюється різницею у механізмі деформації. В умовах статичного навантаження деформація пов'язана зі зміною співвідношення між компонентами грунта за рахунок видалення вільної вологості при стабілізації деформацій (процесі консолідації), в умовах динамічного навантаження (наприклад внаслідок вибухів або ударів), за рахунок короткочасної дії надлишкового тиску, деформування таких грунтів відбувається в більшості випадків, за умов зменшення вільної пористості, і частково внаслідок пружного стиснення води та мінерального скелету. Таким чином при динамічному навантаженні є поняття оптимальної вологості [5, 9–10], при якій незворотні деформації грунтів будуть найбільшими, і для кожного окремого типу грунту цей показник не буде постійним, знаходячись в залежності від навантаження та швидкості його дії.

Тут треба зазначити також, що вологість повинна розглядатись як регіональний фактор впливу, бо є відповідні результати досліджень (за Вовком О.О.), які показують суттєві відмінності впливу вибухів у літній час у південних районах України при ваговій вологості грунтів від 4 до 9–10% і поблизу Києва, при ваговій вологості грунтів 12–16%. Вони показують суттєво різний характер деформування і зміни механічних властивостей грунтів різної вологості, речовинного і гранулометричного складу. Це доводить, що вологість повинна враховуватись не суто як змінна при лабораторних або польових дослідженнях, а як регіональний параметр техногенного впливу, що визначає геодинамічну характеристику льосових грунтів під дією вибухових навантажень. Таким чином, географічна неоднорідність зволоження повинна враховуватись при розробці моделі прогнозу техногенних деформацій і оцінки інженерно-геологічних ризиків при експлуатації та реконструкції інженерних об'єктів різного призначення, цьому буде присвячений наступний етап наших досліджень.

При впливі вибухової хвилі на масив грунту у межах зони непружних деформацій із падінням напруг на фронті одночасно змінюється режим навантаження, тобто знижується його швидкість, наближаючись до квазістатичних умов. На основі аналізу відібраних зразків та відповідних лабораторних досліджень встановлено, що у зв'язних грунтах найбільші зміни щільності відбуваються в зоні, що безпосередньо прилягає до виїмки. Відповідно при збільшенні вологості грунту ущільнення значно вище (18–20%), а ось при значеннях, близьких до повного водонасичення,

розміри ущільненої зони зростають, а ось зміна самої щільності в цій зоні вже незначне. Схожі дані були отримані у дослідженнях [10–11].

На основі відібраних зразків, було проведено дослідження міцнісних параметрів льосових суглинків, в першу чергу їх зміни в ущільненій вибухом зоні, а також тиксотропних змін у грунті при впливі вибухового навантаження. Встановлені значення питомого зчеплення та коефіцієнта тертя методом тривісного стиснення, а саме – використанням TriSCAN (VJTech, Великобританія). Його перевага саме для випробувань льосових порід полягає в максимальному наближенні гідрогеомеханічних режимів випробувань зразків порід до реальних умов стану грунтів в масиві. Наведений пристрій дозволяє проводити автоматизований контроль тиску в камері навантаження, лінійних та об›ємних деформацій зразка, а також витрати води, що проходить через зразок у режимі реального часу. Серія випробувань фільтраційних параметрів порід проведена в умовах тривісного стиску при різних значеннях навантажень, геостатичний тиск приймався в діапазонах, що відповідають глибинам відібраних зразків (5–7 м), з природною вологістю, що відповідало реальним умовам льосового масиву. Час випробування складав 24 години. Перед проведенням досліду проводилося консолідаційне доущільнення зразків при навантаженнях, що відповідали геостатичному тиску, до умовної стабілізації деформацій за чинними нормативами. Зразки були відібрані у проміжку часу 6–7 днів після вибухів на досліджуваній території.

Обробка результатів експериментів дозволила встановити наступне. Величина зчеплення грунтів безпосередньо поблизу від стінок вибухової порожнини та далі вглиб масиву, зростає у порівнянні з тим, яку грунт мав у природному недеформованому стані, в 2–2,5 разів.

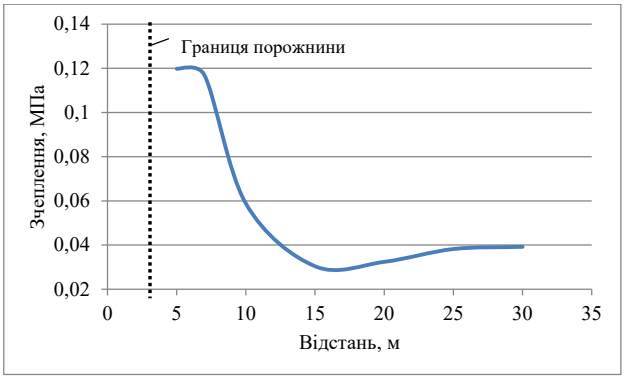


Рис. 2. Зміна значення зчеплення льосових суглинків в залежності від відстані від границі порожнини, утвореної вибухом

За результатами дослідження встановлено, що зміна величини зчеплення (рис. 2), по мірі віддалення від центру вибуху, різко зменшується і приблизно на відстані 13–14 м стає рівною тій, яку грунт мав у недеформованому стані. Далі, у глиб масиву, зчеплення зменшується до величин в 1,5–2 рази менше початкових, а з наближенням до границь зони деформацій знов збільшується до природних величин.

Кут внутрішнього тертя в ущільненій зоні змінюється наступним чином: у контура порожнини (у зоні ущільнення) його величина збільшується на 3–4,5 градуси у порівнянні з початковим до вибуху (рис. 3). Зі збільшенням відстані від цен-

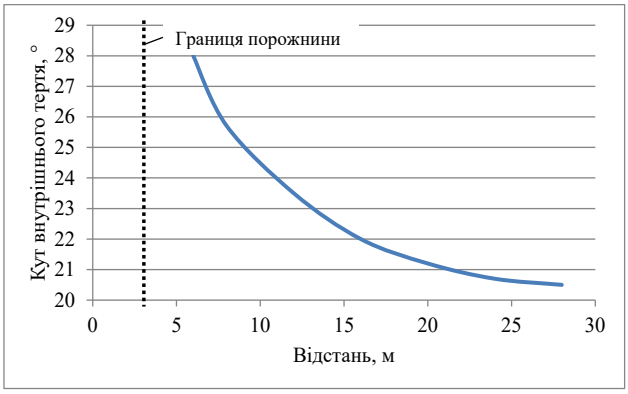


Рис. 3. Зміна значення кута внутрішнього тертя льосових суглинків в залежності від відстані від границі порожнини, утвореної вибухом

тру вибуху величина φ зменшується до початкових значень, причому розміри зони його зміни значно менше розмірів зон зміни зчеплення.

Таким чином, зміни міцнісних характеристик льосових суглинків в ущільненій вибухом зоні знакозмінно: грунт в деяких зонах тільки набуває додаткову міцність, але й у відповідних зонах втрачає її у порівнянні з початковою, тобто розущільнюється.

Є дані досліджень [11–12], які свідчать про те, що величина зчеплення грунту з часом збільшується, причому основний її приріст спостерігається у період часу близько 6–7 днів після вибуху (саме у цей проміжок було відібрано зразки для даного дослідження). За абсолютною величиною приріст зчеплення складає приблизно 25–30% від значень, які спостерігаються одразу після вибухового навантаження, причому найбільш інтенсивно зчеплення збільшується в зоні максимальних деформацій, а менше – в периферійних областях. Через приблизно 30–35 діб після вибухового навантаження може спостерігатись деяке збільшення величини зчеплення, однак швид-

кість цих змін невелика, тому процес тиксотропії можна вважати завершеним. З часом величини кутів внутрішнього тертя ущільненої вибухом зони збільшуються, причому ця зміна зареєстрована лише в період часу перших 5–7 днів, далі зміни φ практично не спостерігаються.

Аналіз результатів досліджень показує, що в деформованих вибухом суглинистих грунтах відбуваються тиксотропні зміни, що підтверджується змінами у часі величин зчеплення на ділянках ущільненої зони з незначними об'ємними деформаціями за її межами, де зміни щільності (об'ємної ваги) не відмічено. Ще більш показово описані особливості поведінки грунтів проявляються при вибухах зарядів в глинистих грунтах та обводнених льосових суглинках [5, 7, 10].

Невелика кількість експериментів дає можливість якісно підтвердити описані вище результати, а також відмітити, що вказані особливості в глинах та обводнених льосових суглинках простежуються більш відчутно. Одночасно зони зі зниженими величинами зчеплення як за абсолютною величиною, так і за розмірами зон розущільнення перевищують отримані для суглинків.

Висновки. 1. Експериментально підтверджено, що вибухове навантаження суттєво впливає на фізико-механічні характеристики льосових суглинків Придніпров'я, викликаючи зміну їхньої щільності, зчеплення та кута внутрішнього тертя. Встановлено, що в зоні безпосереднього впливу вибуху (поблизу вибухової порожнини) спостерігається зростання питомого зчеплення ґрунту в 2–2,5 рази у порівнянні з його природним станом, далі ця величина поступово зменшується з віддаленням від епіцентру вибуху, досягаючи природних значень на відстані 13–14 м, а потім знижується у певних зонах до рівня 1,5–2 рази менше від початкових значень.

2. Також встановлені знакозмінні закономірності поведінки міцнісних характеристик льосових суглинків в ущільненій вибухом зоні: у деяких ділянках відбувається ущільнення і збільшення міцності, а в інших – розущільнення та зниження міцності. Тиксотропні зміни у льосових суглинках призводять до поступового зростання зчеплення у перші 6–7 днів після вибуху, причому цей приріст становить 25–30% від початкових значень. У подальші 30–35 днів швидкість цих змін значно зменшується, що свідчить про завершення тиксотропного процесу. Зміна кута внутрішнього тертя відбувається в перші 5–7 днів після вибуху, після чого цей параметр залишається практично стабільним, що вказує на незворотність деформаційних процесів у досліджених породах.

3. Одержані результати свідчать про необхідність врахування змін у фізико-механічних параметрах льосових суглинків при проєктуванні та експлуатації інженерних споруд на територіях, що зазнають вибухового навантаження, з метою забезпечення їхньої стійкості та довговічності.

Список літератури:

1. Sadovenko I. A., Derevyagina N. I., Podvigina E. O., Zagricenko A. N. Dynamics of loess mass deformation due to technogenic load. *Збірник наукових праць Національного гірничого університету*. 2014. Вип. 45. С. 76–81.

2. Мокрицька Т. П. Формування та еволюція геологічного середовища Придніпровського промислового регіону: монографія. Дніпропетровськ : Ліра, 2013. 278 с.

3. Манюк В. В. Особливості будови четвертинних відкладів у типових розрізах Середнього Придніпров'я. *Вісник Дніпропетровського університету*. Серія: Геологія. Географія. 2014. Вип. 15. С. 2–9.

4. Мокрицька Т. П., Євграшкіна Г. П., Манюк В. В., Бондар О. В., Баранов П. Н., Самойлич К. О. Небезпечні стани та явища у геологічному середовищі на прикладі міст Середнього Придніпров'я: монографія. Дніпро: ПП Акцент, 2019. 205 с.

5. Ремез Н. С. Взаємодія вибухових хвиль з ґрунтами і елементами техноурбоекосистем. Київ: Центр учбової літератури, 2019. 232 с.

6. Мокрицька Т. П. Індуктивні моделі деградації просадності. Інженерна геологія. 2013. № 5. С. 34–38.

7. He L., Jiao Y., Zhang Y., Zheng F., Peng H., Ranjith P. G. Innovative geological–geotechnical zoning framework for urban planning: Wuhan's experience. Geomechanics and Geophysics for Geo-Energy and Geo-Resources. 2024. Vol. 10. Article 176. https://doi.org/10.1007/s40948-024-00830-5

8. Герасимов І. П., Федорчук В. О. Лесові породи України та їхні геологічні особливості. Київ: Наукова думка, 2001. 208 с.

9. Шульга П. О. Четвертинні відклади Лісостепу України: генезис та особливості. Харків: Видавництво ХНУ, 2010. 184 с.

10. Ambrosini D., Luccioni B. Effects of underground explosions on soil and structures. Underground Space. 2019. Vol. 5, No. 4. P. 241–251. DOI: https://doi.org/10.1016/j.undsp.2019.09.002.

11. González-Nicieza C., Álvarez-Fernández M. I., Álvarez-Vigil A. E. та ін. Influence of depth and geological structure on the transmission of blast vibrations. Bulletin of Engineering Geology and the Environment. 2014. Vol. 73. P. 1211–1223. DOI: https://doi.org/10.1007/s10064-014-0595-7.

12. Maheshwari P., Murmu S., Verma H. K. Modeling of blast induced damage distance for underground tunnels. Proceedings of the Indian Geotechnical Conference 2019 / S. Patel [et al.] (Eds.). Singapore : Springer, 2021. Vol. 137. P. 545–553. (Lecture Notes in Civil Engineering). DOI: https://doi.org/10.1007/978-981-33-6466-0_57.

Dereviahina N.I., Sadovenko I.O. **EXPERIMENTAL RESEARCH OF THE CHANGES IN THE STABILITY PARAMETERS OF LOESS MASSIFS UNDER EXPLOSIVE LOADS**

Purpose. Establishing the patterns of changes in physical and mechanical properties of loess loams of the Dnipro region under the influence of explosive loading in conditions of a technogenically altered geological environment for assessment of stability and prediction of soil behavior during further engineering development of the territories.

Methods. A series of laboratory experiments is conducted on samples of loess loams collected from the right-bank part of the city of Dnipro, specifically from an area damaged by an explosion. The samples are analyzed for their physical parameters (density, moisture content, porosity), as well as for their granulometric and chemical composition. Tests are carried out using the TriSCAN (VJTech) software to determine strength parameters before and after explosive loading. Thixotropic changes in the soils are also analyzed over a period of 35 days following the explosion.

Results. The patterns of strength changes in the zone compacted by the explosion are established, with a 2–2.5-fold increase in specific cohesion observed in the area of maximum impact. This increase gradually decreased to natural values at a distance of 13–14 meters, and then dropped below the initial level. Over the course of 6–7 days following the explosion, cohesion increased by 25–30%, confirming the manifestation of thixotropy. The internal friction angle changed during the first 5–7 days, after which it stabilized.

Scientific novelty. For the first time, changes in physical and mechanical characteristics of loess loams of the Dnipro region due to degradation under explosive loading have been experimentally established. The dependency of cohesion and internal friction angle changes on time after loading is determined, which makes it possible to justify the mechanism of massif degradation under technogenic impact conditions.

Practical significance. The obtained experimental results can be used in the design, construction, and operation of engineering structures in areas affected by explosive loading. Taking into account the changes in physical and mechanical properties of the soil massif allows improving safety and durability of construction projects, particularly on specific loess soils.

Key words: *loess loams, explosive loading, cohesion, internal friction angle, compaction, technogenesis, Dnipro region.*

Відомості про авторів

Антонюк Д.С. – кандидат педагогічних наук, доцент, доцент кафедри інженерії програмного забезпечення Державного університету «Житомирська політехніка», докторант Інститут цифровізації освіти Національної академії педагогічних наук України

Арестова Т.В. – асистент кафедри електромеханічних та комп'ютерних систем Навчально-наукового професійно-педагогічного інституту Харківського національного університету імені В.Н. Каразіна

Арпентій С.П. – провідний науковий співробітник Українського науково-дослідного інституту спеціальної техніки та судових експертиз Служби безпеки України

Астістова Т.І. – кандидат технічних наук, доцент кафедри комп'ютерної інженерії та електромеханіки Київського національного університету технологій та дизайну

Бараннік В.В. – доктор технічних наук, професор, професор кафедри математичного моделювання та аналізу даних Харківського національного університету імені В.Н. Каразіна

Бараннік Д.В. – кандидат технічних наук, викладач кафедри інформаційно-мережної інженерії Харківського національного університету радіоелектроніки

Батюк С.Г. – кандидат технічних наук, доцент кафедри автоматизації енергетичних процесів Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Безкровний М.М. – аспірант Тернопільського національного технічного університету імені Івана Пулюя

Березюк І.А. – кандидат технічних наук, доцент Центральноукраїнського національного технічного університету

Берчанов А.А. – аспірант кафедри математичного моделювання та аналізу даних Харківського національного університету імені В.Н. Каразіна

Богдан Г.А. – кандидат технічних наук, доцент кафедри автоматизації та систем неруйнівного контролю Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Бойко О.В. – кандидат технічних наук, доцент, доцент Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Бондаренко Д.О. – кандидат економічних наук, доцент, завідувач кафедри інформаційних систем Харківського національного економічного університету імені Семена Кузнеця

Бочок В.О. – аспірант Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Бурлуцький О.В. – кандидат технічних наук, старший викладач кафедри механіки та конструювання машин Українського державного університету залізничного транспорту

Вакалюк Т.А. – доктор педагогічних наук, професор, завідувач кафедри інженерії програмного забезпечення Державного університету «Житомирська політехніка»

Вакалюк Т.А. – доктор педагогічних наук, професор, завідувач кафедри інженерії програмного забезпечення Державного університету «Житомирська політехніка»

Васильчук Д.П. – старший викладач кафедри електромеханічних та комп'ютерних систем Харківського національного університету імені В.Н. Каразіна

Вернік М.О. – аспірант кафедри програмного забезпечення комп'ютерних систем Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Войтко В.В. – кандидат технічних наук, доцент, доцент кафедри програмного забезпечення Вінницького національного технічного університету

Воронюк М.В. – викладач циклової комісії автоматизації технологічних процесів Київського індустріального фахового коледжу Київського національного університету будівництва та архітектури

Глухова Н.В. – кандидат технічних наук, доцент, доцент кафедри кіберфізичних та інформаційно-вимірювальних систем Національного технічного університету «Дніпровська політехніка»

Голод І.В – аспірант кафедри комп'ютерно-інтегрованих технологій, автоматизації та робототехніки Харківського національного університету радіоелектроніки

Голуб Г.М. – кандидат технічних наук, доцент кафедри автоматизації та комп'ютерно-інтегрованих технологій транспорту Національного транспортного університету

Гордієнко І.В. – кандидат педагогічних наук, доцент, доцент кафедри математики та економіки Дрогобицького державного педагогічного університету імені Івана Франка

Григор'єва Т.І. – кандидат технічних наук, доцент, завідувач кафедри інформаційних технологій Міжнародного гуманітарного університету

Губар В.Г. – старший викладач кафедри конструювання електронно-обчислювальної апаратури Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Гузь Г.М. – асистент кафедри спеціалізованих комп'ютерних систем Українського державного університету науки і технологій

Гуржій С.В. – старший науковий співробітник Українського науково-дослідного інституту спеціальної техніки та судових експертиз Служби безпеки України

Дем'янчук Т.М. – студент Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Деревягіна Н.І. – кандидат технічних наук, доцент, завідувач кафедри гідрогеології та інженерної геології Національного технічного університету «Дніпровська політехніка»

Добровольський Г.А. – кандидат технічних наук, доцент кафедри комп'ютерних наук Запорізького національного університету

Довгалюк І.С. – здобувач Державного університету «Житомирська політехніка»

Дроменко В.Б. – кандидат технічних наук, доцент кафедри інженерних систем та технологій Таврійського національного університету імені В.І. Вернадського

Дубовик Т.М. – старший викладач кафедри спеціалізованих комп'ютерних систем Українського державного університету науки і технологій

Заварін В.О. – магістрант факультету прикладної математики кафедри системного програмування і спеціалізованих комп'ютерних систем Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Заячук Я.І. – кандидат технічних наук, доцент, доцент кафедри комп'ютерних систем і мереж Івано-Франківського національного технічного університету нафти і газу

Здоренко С.С. – аспірант кафедри автоматизації та систем неруйнівного контролю Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Зилевіч М.О. – доктор філософії, асистент кафедри конструювання електронно-обчислювальної апаратури Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Зіменков Д.А. – студент кафедри системного програмування і спеціалізованих комп'ютерних систем факультету прикладної математики Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Знахур Л.В. – старший викладач кафедри інформаційних систем Харківського національного економічного університету імені Семена Кузнеця

Зубенко В.О. – кандидат технічних наук, доцент Херсонського державного аграрно-економічного університету

Іванишин В.В. – старший викладач Національного авіаційного університету

Ізовіта О.Л. – аспірант кафедри автоматизації виробничих процесів Центральноукраїнського національного технічного університету

Калашник В.Ю. – кандидат технічних наук, асистент кафедри комп'ютерної інженерії та електромеханіки Київського національного університету технологій та дизайну

Карпенко М.І. – аспірант кафедри автоматизації та комп'ютерних технологій систем управління ім. проф. А.П. Ладанюка Національного університету харчових технологій

Клименко Д.О. – студент Державного університету «Житомирська політехніка»

Климова І.М. – асистент кафедри електронних обчислювальних машин Харківського національного університету радіоелектроніки

Колобродов В.Г. – доктор технічних наук, професор Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Коломієць У.Д. – магістр Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Колтун Ю.М. – кандидат технічних наук, доцент, доцент кафедри електронних обчислювальних машин Харківського національного університету радіоелектроніки

Кордюк Р.І. – аспірант кафедри комп'ютерних технологій у видавничо-поліграфічних процесах Національного університету «Львівська політехніка»

Кропивницька В.Б. – кандидат технічних наук, доцент, доцент кафедри комп'ютерних систем і мереж Івано-Франківського національного технічного університету нафти і газу

Кропивницький Д.Р. – доктор філософії, доцент кафедри комп'ютерних систем і мереж Івано-Франківського національного технічного університету нафти і газу

Крохмаль М.В. – аспірант кафедри інформаційних технологій Міжнародного гуманітарного університету

Кубрак А.В. – аспірант кафедри автоматизації та комп'ютерно-інтегрованих технологій транспорту Національного транспортного університету

Лагойда А.І. – кандидат технічних наук, завідувач кафедри автоматизації та комп'ютерно-інтегрованих технологій Івано-Франківського національного технічного університету нафти і газу

Левчук І.Л. – кандидат технічних наук, доцент кафедри комп'ютерно-інтегрованих технологій та автоматизації Українського державного університету науки і технологій

Лиходєєв О.С. – кандидат технічних наук, доцент, доцент кафедри інженерно-авіаційного забезпечення Харківського національного університету Повітряних Сил

Лукінюк М.В. – старший викладач кафедри технічних та програмних засобів автоматизації Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Майданюк В.П. – кандидат технічних наук, доцент кафедри інженерія програмного забезпечення Вінницького національного технічного університету

Мацуй А.М. – доктор технічних наук, професор, доцент кафедри автоматизації виробничих процесів Центральноукраїнського національного технічного університету

Мінтій І.С. – кандидат педагогічних наук, доцент, старший дослідник, ад'юнкт кафедри інформатики факультету фізики та прикладної інформатики Лодзького університету у Польщі

Мойсеєнко О.В. – кандидат технічних наук, доцент, доцент кафедри комп'ютерних систем і мереж Івано-Франківського національного технічного університету нафти і газу

Мошенский А.О. – доцент Таврійського національного університету імені В.І. Вернадського

Мошенський А.О. – кандидат технічних наук, доцент, доцент кафедри інформаційних технологій штучного інтелекту і кіберзахисту Національного університету харчових технологій

Назаренко В.В. – аспірант кафедри спеціалізованих комп'ютерних систем Національного університету «Львівська політехніка»

Новіцька І.В. – кандидат педагогічних наук, доцент, завідувач відділу аспірантури та докторантури, доцент кафедри професійно-педагогічної, спеціальної освіти, андрагогіки та управління Житомирського державного університету імені Івана Франка

Олещенко Л.М. – кандидат технічних наук, доцент кафедри програмного забезпечення комп'ютерних систем Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Олійник Р.С. – студент кафедри конструювання електронно-обчислювальної апаратури Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Парінов І.С. – директор і власник компанії Inspire Median LLC, Бока-Ратон, Флорида

Пасічник В.В. – аспірант кафедри комп'ютерних наук Запорізького національного університету

Піх В.Я. – кандидат технічних наук, доцент, декан факультету інформаційних технологій Івано-Франківського національного технічного університету нафти і газу

Позур М.Ю. – аспірант кафедри програмного забезпечення Вінницького національного технічного університету

Поліщук С.М. – начальник відділу Українського науково-дослідного інституту спеціальної техніки та судових експертиз Служби безпеки України

Положаєнко С.А. – доктор технічних наук, професор, завідувач кафедри Національного університету «Одеська політехніка»

Полтавський Д.А. – керівник інженерної групи Uplandme Inc Чорноморського національного університету імені Петра Могили

Помазуєв В.В. – старший викладач кафедри тактики авіації Харківського національного університету Повітряних Сил

Поплавська О.А. – старший викладач кафедри вищої математики та комп'ютерних застосувань Хмельницького національного університету

Прокофьєв А.Ю. – аспірант Національного університету «Одеська політехніка»

Рамський А.О. – кандидат фізико-математичних наук, доцент, завідувач кафедри вищої математики та комп'ютерних застосувань Хмельницького національного університету

Регенель Т.Ю. – здобувач вищої освіти спеціальності «Веб-технології» Державного університету «Житомирська політехніка»

Рихальський О.Ю. – викладач кафедри комп'ютерних наук та програмної інженерії факультету інформаційних систем та технологій ПВНЗ «Європейський університет»

Розенвассер Д.М. – кандидат технічних наук, доцент кафедри комп'ютерних наук Міжнародного гуманітарного університету

Романчук О.О. – асистент кафедри спеціалізованих комп'ютерних систем Українського державного університету науки і технологій

Савіцький Р.С. – старший викладач кафедри інженерії програмного забезпечення Державного університету «Житомирська політехніка»

Садовенко І.О. – доктор технічних наук, професор, професор кафедри гідрогеології та інженерної геології Національного технічного університету «Дніпровська політехніка»

Семенець Д.А. – кандидат технічних наук, доцент кафедри електромеханічних та комп'ютерних систем Харківського національного університету імені В.Н. Каразіна

Сенів М.М. – кандидат технічних наук, доцент, доцент кафедри програмного забезпечення Національного університету «Львівська політехніка»

Сербул О.М. – кандидат технічних наук, доцент, доцент кафедри автоматизації виробничих процесів Центральноукраїнського національного технічного університету

Ситніков О.В. – кандидат технічних наук, доцент кафедри технічних та програмних засобів автоматизації Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Сідченко Є.С. – аспірант кафедри математичного моделювання та аналізу даних Харківського національного університету імені В.Н. Каразіна

Складанюк О.О. – аспірант кафедри програмного забезпечення Вінницького національного технічного університету

Слабінога М.О. – кандидат технічних наук, доцент кафедри комп'ютерних систем і мереж Івано-Франківського національного технічного університету нафти і газу

Соловій П.С. – аспірант кафедри спеціалізованих комп'ютерних систем Інституту комп'ютерних технологій, автоматики та метрології Національного університету «Львівська політехніка»

Сташевський В.М. – аспірант Інституту механіки та автоматики агропромислового виробництва Національної академії аграрних наук України

Степанов Д.С. – асистент Національного університету «Львівська політехніка»

Стрілецький Ю.Й. – доктор технічних наук, професор, професор кафедри інформаційно-телекомунікаційних технологій та систем Івано-Франківського національного технічного університету нафти і газу

Стьопкін А.В. – кандидат фізико-математичних наук, доцент, доцент кафедри математики та інформатики ДВНЗ «Донбаський державний педагогічний університет»

Танчук В.С. – аспірант Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Терейковський І.А. – доктор технічних наук, професор, професор кафедри системного програмування і спеціалізованих комп'ютерних систем факультету прикладної математики Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Терейковський І.А. – кандидат технічних наук, старший науковий співробітник, професор кафедри системного програмування і спеціалізованих комп'ютерних систем Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Терейковський І.А. – кандидат технічних наук, старший науковий співробітник, професор кафедри системного програмування і спеціалізованих комп'ютерних систем Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Тиханська А.М. – старший викладач кафедри автоматизації, комп'ютерних наук і технологій Криворізького національного університету

Тиханський М.П. – кандидат технічних наук, доцент, доцент кафедри автоматизації, комп'ютерних наук і технологій Криворізького національного університету

Топчій Н.В. – старший науковий співробітник Українського науково-дослідного Інституту спеціальної техніки та судових експертиз Служби безпеки України

Ульяницька К.О. – кандидат технічних наук, доцент Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Усов О.С. – аспірант Національного університету «Одеська політехніка»

Ушакова І.О. – кандидат економічних наук, доцент, професор кафедри інформаційних систем Харківського національного економічного університету імені Семена Кузнеця

Фант М.О. – кандидат філологічних наук, доцент кафедри інженерії програмного забезпечення Державного університету«Житомирська політехніка»

Федорова Н.В. – доктор технічних наук, професор, професор кафедри інженерії програмного забезпечення в енергетиці Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Федоровський О.В. – старший викладач кафедри тактики авіації Харківського національного університету Повітряних Сил

Феоклістов Д.О. – аспірант кафедри автоматизації та систем неруйнівного контролю Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Філімончук Т.В. – кандидат технічних наук, доцент, доцент кафедри електронних обчислювальних машин Харківського національного університету радіоелектроніки

Фомін О.В. – доктор технічних наук, професор кафедри вагонів та управління вагонами Національного транспортного університету

Фтемов А.О. – магістрант кафедри технічних та програмних засобів автоматизації Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Фурсина А.Д. – кандидат технічних наук, доцент кафедри механіки Національного університету «Запорізька політехніка»

Хлєбніков М.В. – аспірант кафедри автоматизації виробничих процесів Центральноукраїнського національного технічного університету

Хнюнін С.Г. – кандидат технічних наук, доцент кафедри інформаційних технологій Міжнародного гуманітарного університету

Холобок В.І. – магістрант кафедри електронних обчислювальних машин Харківського національного університету радіоелектроніки

Цибульник С.О. – кандидат технічних наук, доцент, доцент кафедри комп'ютерно-інтегрованих оптичних та навігаційних систем Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Чайковський С.Ю. – старший науковий співробітник Українського науково-дослідного інституту спеціальної техніки та судових експертиз Служби безпеки України

Черепанська І.Ю. – доктор технічних наук, професор, професор кафедри автоматизації та систем неруйнівного контролю Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Чигур І.І. – кандидат технічних наук, доцент кафедри автоматизації та комп'ютерно-інтегрованих технологій Івано-Франківського національного технічного університету нафти і газу

Чигур Л.Я. – кандидатка технічних наук, доцентка кафедри автоматизації та комп'ютерно-інтегрованих технологій Івано-Франківського національного технічного університету нафти і газу

Чижмотря О.В. – старший викладач кафедри інженерії програмного забезпечення Державного університету «Житомирська політехніка»

Чижмотря О.Г. – старший викладач кафедри інженерії програмного забезпечення Державного університету «Житомирська політехніка»

Чирва Ю.Є. – кандидат економічних наук, доцент, доцент кафедри інформаційних систем Харківського національного економічного університету імені Семена Кузнеця

Шевченко О.В. – студент кафедри системного програмування і спеціалізованих комп'ютерних систем Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Шекета В.І. – доктор технічних наук, професор, професор кафедри інженерії програмного забезпечення Івано-Франківського н. –аціонального технічного університету нафти і газу

Шелемаха В.В. – аспірант кафедри комп'ютерно-інтегрованих оптичних та навігаційних систем Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Шорнікова С.В. – начальник відділу Українського науково-дослідного інституту спеціальної техніки та судових експертиз Служби безпеки України

Ярецька Н.О. – кандидат фізико-математичних наук, доцент, доцент кафедри вищої математики та комп'ютерних застосувань Хмельницького національного університету

Науковий журнал

ВЧЕНІ ЗАПИСКИ
ТАВРІЙСЬКОГО НАЦІОНАЛЬНОГО УНІВЕРСИТЕТУ ІМЕНІ В.І. ВЕРНАДСЬКОГО

Серія: Технічні науки

Том 36 (75) № 3 2025

Частина 2

Коректура • *Н. Славогородська*

Комп'ютерна верстка • *Т. Клименко*

Адреса редакції:
Таврійський національний університет імені В. І. Вернадського
м. Київ, вул. Джона Маккейна, 33
Електронна пошта: editor@tech.vernadskyjournals.in.ua
Сторінка журналу: www.tech.vernadskyjournals.in.ua

Формат 60 × 84/8. Гарнітура Times New Roman.
Папір офсетний. Цифровий друк. Обл.-вид. арк. 44,48. Ум. друк. арк. 56,50. Зам. № 0925/720
Підписано до друку 30.06.2025. Наклад 150 прим.

Видавництво і друкарня – Видавничий дім «Гельветика»
65101, м. Одеса, вул. Інглезі, 6/1
Телефони: +38 (095) 934 48 28, +38 (097) 723 06 08
E-mail: mailbox@helvetica.ua
Свідоцтво суб'єкта видавничої справи
ДК № 7623 від 22.06.2022 р.