

**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 37 (76) № 2 2026

Частина 1



Видавничий дім
«Гельветика»
2026

Головний редактор:

Кисельов Володимир Борисович, доктор технічних наук, професор, директор навчально-наукового інституту муніципального управління та міського господарства, Таврійський національний університет імені В. І. Вернадського, Україна

Члени редакційної колегії:

Бронін Сергій Вадимович, кандидат технічних наук, доцент кафедри кібербезпеки, Національний технічний університет «Харківський політехнічний університет», Україна;

Гуйда Олександр Григорович, кандидат наук з державного управління, професор, завідувач кафедри комп'ютерних та інформаційних технологій, Таврійський національний університет імені В. І. Вернадського, Україна;

Лисенко Олександр Іванович, доктор технічних наук, професор, професор кафедри телекомунікацій Навчально-наукового інституту телекомунікаційних систем, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Україна;

Лісовець Сергій Миколайович, кандидат технічних наук, доцент, доцент кафедри інженерних систем та технологій Навчально-наукового інституту муніципального управління та міського господарства, Таврійський національний університет імені В. І. Вернадського, Україна;

Медведєв Микола Георгійович, доктор технічних наук, професор, професор кафедри інженерних систем та технологій, Таврійський національний університет імені В. І. Вернадського, Україна;

Олещенко Любов Михайлівна, кандидат технічних наук, доцент, доцент кафедри програмного забезпечення комп'ютерних систем, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Україна;

Скуратовський Руслан Вячеславович, кандидат технічних наук, доцент, докторант, Інститут математики НАН України, Україна;

Alexandr Neftissov, PhD, Associate Professor, Vice-Rector for Science and Innovation, Kazakh National University of Sports, Kazakhstan;

Andreas Pester, Dr.phil.habil., Professor, Professor at the Faculty of Informatics and Computer Science, British University of Egypt, Egypt;

Danielienė Renata, Dr., Associate Professor, Associate Professor of the Institute of Social Sciences and Applied Information, Vilnius University, Lithuania.

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 № 10 від 16 квітня 2026 року).**

Науковий журнал «Вчені записки ТНУ імені В.І. Вернадського. Серія: Технічні науки»
zareestrovano 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, 2026

ЗМІСТ

ІНФОРМАТИКА, ОБЧИСЛЮВАЛЬНА ТЕХНІКА ТА АВТОМАТИЗАЦІЯ

Ахмедов М.Ш., Федорова Н.В.

ДИНАМІЧНЕ БАЛАНСУВАННЯ НАВАНТАЖЕННЯ В РЕАЛЬНОМУ ЧАСІ ТА ПРОБЛЕМИ
ЙОГО ВПРОВАДЖЕННЯ В ЕНЕРГОМЕРЕЖІ SMART GRID.....1

Баглай О.П.

ВИКОРИСТАННЯ СУЧАСНИХ СТАТИСТИЧНИХ МЕТОДІВ У ЗВАРЮВАННІ
МЕТАЛОКОНСТРУКЦІЙ..... 9

Bilyi M.O., Krylov Ie.V.

MEDIAN CHANGE POINT DETECTION: A SEMANTIC TEXT SEGMENTATION METHOD
FOR INFORMATION SYSTEMS WITH VECTOR DATABASES.....17

Вакалюк Т.А., Фант М.О., Дунєв С.С., Власенко О.В., Коломієць В.В.

ОГЛЯД МЕТОДІВ ПРОГНОЗУВАННЯ ДИНАМІКИ ПОПИТУ ТОВАРІВ.....25

Волк М.О., Ковтун Є.І.

МОДЕЛІ ДИНАМІЧНОГО РОЗПОДІЛЕННЯ ОБЧИСЛЮВАЛЬНИХ ЗАДАЧ
У ГЕТЕРОГЕННИХ КОМП'ЮТЕРНИХ СИСТЕМАХ.....32

Гейко О.О., Варава І.А., Пуха Г.С., Дембіцький В.В., Вдовина А.В.

АРХІТЕКТУРА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ВАЛІДАЦІЇ МАТЕМАТИЧНИХ МОДЕЛЕЙ. 38

Glukhova N.V.

SPECIFICS OF METHODS FOR ASSESSING THE ACCURACY OF PARAMETER
DETERMINATION WHEN IMPROVING COMPUTER-INTEGRATED TECHNOLOGY
FOR ANALYSIS OF GAS-DISCHARGE RADIATION IMAGES..... 44

Громський О.О., Нерода С.І., Яцишин А.В.

АНАЛІЗ АЛГОРИТМІВ АНАЛІТИЧНИХ ВЕБ-СИСТЕМ ПРОГНОЗУ МІЖБІРЖОВОЇ ВАРТОСТІ
ЦИФРОВИХ КРИПТОАКТИВІВ.....49

Денисенко І.В., Терейковський І.А.

ПАРАМЕТРИ ОЦІНКИ ЕФЕКТИВНОСТІ ЗАСОБІВ РОЗПІЗНАВАННЯ ЕМОЦІЙНОГО СТАНУ
В СИСТЕМАХ ДИСТАНЦІЙНОГО НАВЧАННЯ.....54

Діденко Т.В., Кунгурцев О.Б.

МЕТОД ФОРМУВАННЯ ЗАПИТУ ДЛЯ ПОШУКУ СПІВРОЗМОВНИКА У ЦИФРОВИХ
КОМУНІКАЦІЙНИХ СИСТЕМАХ З УРАХУВАННЯМ ЕМОЦІЙНИХ
ХАРАКТЕРИСТИК КОРИСТУВАЧА.....60

Єременко В.С., Монченко О.В., Монченко Т.М.

ІНФОРМАЦІЙНЕ ПРЕДСТАВЛЕННЯ ОБ'ЄКТА В ЕНДОКРИНОЛОГІЧНИХ ДОСЛІДЖЕННЯХ.68

Засадна Х.О., Коцун В.І.

ГЕНЕРУВАННЯ КЛЮЧІВ В КРИПТОСИСТЕМІ RSA ТА ЇХ ЗАСТОСУВАННЯ.....74

Здоренко С.С., Черепанська І.Ю.

СИСТЕМА МРС-КЕРУВАННЯ ТЕМПЕРАТУРОЮ ПОВІТРЯ В БУДІВЛЯХ З АДАПТАЦІЄЮ
МАТЕМАТИЧНОЇ МОДЕЛІ.....83

Зилевіч М.О., Сорокін Д.А.

СИСТЕМАТИЗАЦІЯ ДЖЕРЕЛ ПОХИБОК ПРИ ВИРІШЕННІ ЗАДАЧ ГЕОЛОКАЦІЇ
З ВИКОРИСТАННЯМ ТЕХНОЛОГІЇ LORA..... 90

Ілуща А.С.

ВЕБ-ОРІЄНТОВАНІ ПРОГРАМНІ ЗАСОБИ ІНТЕГРАЦІЇ PAAS І SAAS ДЛЯ ПІДВИЩЕННЯ
ПРОДУКТИВНОСТІ ДОСЛІДНИЦЬКОЇ ТА ОСВІТНЬОЇ ДІЯЛЬНОСТІ.....97

Іляш Ю.Ю., Горєлов В.О., Антоняк В.С.

АЛГОРИТМИ КЛІТИННИХ АВТОМАТІВ ДЛЯ ГЕНЕРАЦІЇ РІВНІВ
У КОМП'ЮТЕРНИХ ІГРАХ.....108

Іносов С.В., Гаврюков О.В., Бондарчук О.В., Луценко В.Ю., Самойленко М.І. СИНТЕЗ АЛГОРИТМУ ПРОГРАМНОГО РЕГУЛЮВАННЯ ДЛЯ СИЛЬНО НЕЛІНІЙНОГО ТЕПЛОВОГО ОБ'ЄКТУ (НА ПРИКЛАДІ ПРОПАРІЮВАЛЬНОЇ КАМЕРИ).....	114
Кандиба І.О., Гончаров Д.С. ІНТЕГРАЦІЯ ДАНИХ СЕРВІСУ GOOGLE FIT ДО СИСТЕМИ МОНІТОРИНГУ СТАНУ ЗДОРОВ'Я....	120
Карплєвський Б.В. ТЕХНОЛОГІЇ ІНТЕГРАЦІЇ ВІРТУАЛЬНИХ СЕРЕДОВИЩ У СИСТЕМИ УПРАВЛІННЯ РЕАЛЬНИМ СЦЕНІЧНИМ ОБЛАДНАННЯМ.....	126
Козак Р.Р., Дудник О.О., Субін А.А. ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ У ВІЗУАЛІЗАЦІЇ СКЛАДНИХ НАУКОВИХ ДАНИХ.....	135
Корабльов М.М., Дикий С.А. ДІАГНОСТИКА ЕМОЦІЙНОГО СТАНУ ДИТИНИ НА ОСНОВІ АНАЛІЗУ ЇЇ МАЛЮНКІВ З ВИКОРИСТАННЯМ НЕЙРОМЕРЕЖЕВОГО ПІДХОДУ	142
Kuznetsov D.I., Muzyka I.O., Ivashchenko O.R. A MACHINE VISION SYSTEM FOR QUALITY-BASED GRAIN CLASSIFICATION USING MACHINE LEARNING.....	149
Купін А.І., Духов Д.Ю. ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ АВТОМАТИЗОВАНОГО ГЕНЕРУВАННЯ ПРОГРАМНОГО КОДУ ЗАСОБАМИ ШТУЧНОГО ІНТЕЛЕКТУ.....	157
Суховій С.І., Лакоза С.Л., Мацюк Д.М. АНАЛІЗ СУЧАСНИХ ТЕНДЕНЦІЙ ВДОСКОНАЛЕННЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ СИСТЕМ КЕРУВАННЯ ГАЗОТУРБІННИХ ДВИГУНІВ.....	164
Lomovatskyi A.A., Basyuk T.M. METHODS FOR SENTIMENT ANALYSIS OF UKRAINIAN-LANGUAGE CONTENT: RULE-BASED, MACHINE LEARNING AND TRANSFORMER APPROACHES.....	173
Михалюк А.П., Міркевич Р.М. РОЗШИРЕНА КІЛЬКІСНА ОЦІНКА КІБЕРФІЗИЧНИХ РИЗИКІВ: ІНТЕГРАЦІЯ БЕЗПЕКИ, КОНТЕКСТУ ТА ОПЕРАЦІЙНИХ МОЖЛИВОСТЕЙ У СЕРЕДОВИЩАХ ПОТ ТА АВТОМАТИЗАЦІЇ	181
Міхав В.В., Мелешко Є.В., Якименко М.С., Босько В.В., Березюк І.А. МОДЕЛІ ДЕЦЕНТРАЛІЗОВАНИХ Q&A СИСТЕМ В PEER-TO-PEER КОМП'ЮТЕРНИХ МЕРЕЖАХ....	189
Могильний Г.А., Смагіна О.О., Переяславська С.О. АНАЛІЗ ВПЛИВУ КОНТЕЙНЕРИЗАЦІЇ НА ЖИТТЄВИЙ ЦИКЛ РОЗРОБКИ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ (SDLC) В ПАРАДИГМІ DEVOPS.....	195
Нальотов Д.Р., Ляшенко О.С. МОДЕЛЬ КЕРУВАННЯ СКЛАДНИМ ТЕХНОЛОГІЧНИМ ОБ'ЄКТОМ НА ОСНОВІ МЕТОДІВ НЕЧІТКОЇ ЛОГІКИ.....	200
Невлюдов І.Ш., Новоселов С.П., Сичова О.В., Теслюк С.І. РОЗПОДІЛЕНА СИСТЕМА ПОЗИЦІОНУВАННЯ МОБІЛЬНИХ РОБОТІВ НА ОСНОВІ МЕРЕЖІ ARUCO-СЕНСОРІВ ТА АРХІТЕКТУРИ ROS ДЛЯ КІБЕРФІЗИЧНОГО ВИРОБНИЦТВА.....	205
Oleshchenko L.M., Vovchenko D.S. SOFTWARE-BASED LOAD BALANCING METHOD FOR BIG DATA PROCESSING IN DISTRIBUTED INFORMATION SYSTEM.....	213
Пархоменко Ю.М., Пархоменко М.Д., Мацуї А.М., Бокій А.Р., Комаров Д.Ю. ВИЗНАЧЕННЯ МІСЦЯ ВСТАНОВЛЕННЯ ТА ТОЧНОСТІ РЕЄСТРАЦІЇ ДАТЧИКОМ ЗЕРНОВОГО ПОТОКУ, ЩО ФОРМУЄТЬСЯ КОТУШКОВИМ ВИСІВНИМ АПАРАТОМ.....	220
Пелих С.М., Ложечников В.Ф., Зісров О.Р. АЛГОРИТМІЧНО-ОРІЄНТОВАНА ЦИФРОВА СИСТЕМА КЕРУВАННЯ МІКРОКЛІМАТОМ ВИРОБНИЧИХ ПРИМІЩЕНЬ.....	227
Писаренко Ю.В., Кармазін К.В. ЦИФРОВИЙ ДВІЙНИК МЕРЕЖІ ТА ФЕДЕРАТИВНЕ МУЛЬТИАГЕНТНЕ НАВЧАННЯ ДЛЯ РЕЗИЛЬЄНТНОГО КЕРУВАННЯ ПОТОКОВИМ ВІДЕО І ТЕЛЕМЕТРІЄЮ В СИТУАЦІЙНИХ ЦЕНТРАХ.....	234

Piotrovskiy Ye.Ye. THE ROLE OF AUTOMATED FRONTEND-BACKEND SYNCHRONIZATION IN REDUCING DIGITAL PRODUCT TIME TO MARKET.....	241
Прокопович-Ткаченко Д.І., Торстенссон О., Черкаський Д.О., Переметчик Д.О., Берт С.Д. СОЦІОТЕХНІЧНІ ЧИННИКИ ВИНИКНЕННЯ ВРАЗЛИВОСТЕЙ МЕРЕЖЕВИХ ПРОТОКОЛІВ ЗАХИЩЕНОЇ ПЕРЕДАЧІ ДАНИХ.....	249
Пупена О.М. ВИКОРИСТАННЯ AUTOMATIONML В ЖИТТЄВОМУ ЦИКЛІ СИСТЕМ КЕРУВАННЯ ЗА УМОВ ОБМЕЖЕНИХ РЕСУРСІВ.....	262
Пухалевич А.В., Макарова Л.М., Коваленко В.В. МАТЕМАТИЧНА МОДЕЛЬ ДЛЯ ОЦІНЮВАННЯ ЯКОСТІ ЗАСТОСУНКІВ З ВІДКРИТИМ КОДОМ НА JAVA, ЩО СТВОРЮЮТЬСЯ З ВИКОРИСТАННЯМ ФРЕЙМВОРКУ HIBERNATE...	271
Рубель Ю.Б., Грицюк Ю.І. ГЕНЕРУВАННЯ СИНТЕТИЧНОГО БАЛІСТИЧНОГО НАБОРУ ДАНИХ НА ПІДСТАВІ ЧИСЕЛЬНОГО МОДЕЛЮВАННЯ ТА РАНДОМІЗАЦІЇ ДОМЕНІВ ДЛЯ СИСТЕМИ УПРАВЛІННЯ АРТИЛЕРІЙСЬКИМ ВОГНЕМ.....	278
Руденко Ю.О., Агаджанова С.В., В'юненко О.Б., Ситник Л.Г., Агаджанов-Гонсалес К.Х. ОЦІНЮВАННЯ РЕЗУЛЬТАТИВНОСТІ ОПАНУВАННЯ LOW-CODE ТЕХНОЛОГІЙ У МЕЖАХ ПІДГОТОВКИ ФАХІВЦІВ ДО АНАЛІТИКИ ВЕЛИКИХ ДАНИХ.....	289
Samoilov O.O., Doshchenko H.G. ARTIFICIAL INTELLIGENCE TECHNIQUES FOR SHIP ELECTRICAL SYSTEM CONDITION MONITORING.....	295
Ushakova I.O. MODERNIZATION OF THE BACKEND OF THE WEB APPLICATION BASED ON MICROSERVICE ARCHITECTURE.....	303
Янко А.С., Гончаренко С.О. ПОРІВНЯЛЬНИЙ АНАЛІЗ АРХІТЕКТУРНИХ ТА КРИПТОГРАФІЧНИХ ІННОВАЦІЙ У СПЕЦИФІКАЦІЯХ LORAWAN 1.0.4 ТА 1.1.....	311
ВІДОМОСТІ ПРО АВТОРІВ.....	318

CONTENTS

INFORMATICS, COMPUTER ENGINEERING AND AUTOMATION

Akhmedov M.Sh., Fedorova N.V.

DYNAMIC LOAD BALANCING IN REAL TIME AND IMPLEMENTATION CHALLENGES
IN SMART GRID ENERGY NETWORKS.....1

Baglai O.P.

APPLICATION OF MODERN STATISTICAL METHODS IN WELDING
OF METAL STRUCTURES..... 9

Bilyi M.O., Krylov Ie.V.

MEDIAN CHANGE POINT DETECTION: A SEMANTIC TEXT SEGMENTATION METHOD
FOR INFORMATION SYSTEMS WITH VECTOR DATABASES.....17

Vakaliuk T.A., Fant M.O., Duniev S.S., Vlasenko O.V., Kolomiets V.V.

OVERVIEW OF METHODS FOR FORECASTING THE DYNAMICS OF DEMAND FOR GOODS...25

Volk M.O., Kovtun Ye.I.

MODELS OF DYNAMIC DISTRIBUTION OF COMPUTING TASKS
IN HETEROGENEOUS COMPUTER SYSTEMS.....32

Heiko O.O., Varava I.A., Pukha H. S., Dembitsky V.V., Vdovyna A.V.

SOFTWARE ARCHITECTURE FOR VALIDATING MATHEMATICAL MODELS.....38

Glukhova N.V.

SPECIFICS OF METHODS FOR ASSESSING THE ACCURACY OF PARAMETER
DETERMINATION WHEN IMPROVING COMPUTER-INTEGRATED TECHNOLOGY
FOR ANALYSIS OF GAS-DISCHARGE RADIATION IMAGES..... 44

Hromskiy O.O., Neroda S.I., Iatsyshyn A.V.

ANALYSIS OF ALGORITHMS OF ANALYTICAL WEB SYSTEMS FOR FORECASTING
THE INTEREXCHANGE VALUE OF DIGITAL CRYPTOASSETS..... 49

Denysenko I.V., Tereikovskiy I.A.

PARAMETERS FOR ASSESSING THE EFFICIENCY OF EMOTION RECOGNITION TOOLS
IN DISTANCE LEARNING SYSTEMS.....54

Didenko T.V., Kunhurtsev O.B.

METHOD FOR FORMING A QUERY TO SEARCH FOR AN INTERLOCUTOR IN DIGITAL
COMMUNICATION SYSTEMS CONSIDERING USERS' EMOTIONAL CHARACTERISTICS.....60

Yeremenko V.S., Monchenko O.V., Monchenko T.M.

INFORMATION REPRESENTATION OF AN OBJECT IN ENDOCRINOLOGICAL RESEARCH.....68

Zasadna Kh. O., Kotsun V. I.

KEY GENERATION IN THE RSA CRYPTOSYSTEM RSA AND THEIR APPLICATION..... 74

Zdorenko S.S., Cherepanska I.Yu.

MODEL PREDICTIVE CONTROL SYSTEM OF AIR TEMPERATURE IN BUILDINGS
WITH A MATHEMATICAL MODEL ADAPTATION MECHANISM..... 83

Zylevich M.O., Sorokin D.A.

SYSTEMATIZATION OF ERROR SOURCES IN GEOLOCATION USING LORA TECHNOLOGY...90

Ilutsa A.S.

WEB-ORIENTED SOFTWARE INTEGRATION OF PAAS AND SAAS TO INCREASE
THE PRODUCTIVITY OF RESEARCH AND EDUCATIONAL ACTIVITIES.....97

Iliash Yu.Yu., Horielov V.O., Antoniuk V.S.

CELLULAR AUTOMATION ALGORITHMS FOR LEVEL GENERATION IN COMPUTER GAMES.108

Inosov S.V., Gavryukov O.V., Bondarchuk O.V., Lutsenko V.Yu., Samoylenko M.I.

SYNTHESIS OF AN ALGORITHM OF PROGRAMMED CONTROL FOR A STRONGLY
NONLINEAR PLANT (BY THE EXAMPLE OF A CURING PIT).....114

Kandyba I.O. Honcharov D.S.

INTEGRATION OF DATA FROM THE GOOGLE FIT SERVICE INTO A HEALTH
MONITORING SYSTEM.....120

Karplevskiy B.V. TECHNOLOGIES FOR INTEGRATING VIRTUAL ENVIRONMENTS INTO REAL STAGE EQUIPMENT CONTROL SYSTEMS.....	126
Kozak R.R., Dudnyk O.O., Subin A.A. INFORMATION TECHNOLOGIES IN THE VISUALIZATION OF COMPLEX SCIENTIFIC DATA ..	135
Korablyov M.M., Dykyi S.A. DIAGNOSIS OF A CHILD'S EMOTIONAL STATE BASED ON THE ANALYSIS OF HIS DRAWINGS USING A NEURAL NETWORK APPROACH.....	142
Kuznietsov D.I., Muzyka I.O., Ivashchenko O.R. A MACHINE VISION SYSTEM FOR QUALITY-BASED GRAIN CLASSIFICATION USING MACHINE LEARNING.....	149
Kupin A.I., Dukhov D.Yu. RESEARCH INTO THE EFFICIENCY OF AUTOMATED SOFTWARE CODE GENERATION USING ARTIFICIAL INTELLIGENCE	157
Sukhovii S.I., Lakoza S.L., Matsiuk D.M. MODERN TRENDS ANALYSIS OF GAS-TURBINE ENGINE CONTROL SYSTEMS SOFTWARE IMPROVING	164
Lomovatskyi A.A., Basyuk T.M. METHODS FOR SENTIMENT ANALYSIS OF UKRAINIAN-LANGUAGE CONTENT: RULE-BASED, MACHINE LEARNING AND TRANSFORMER APPROACHES.....	173
Mykhalyuk A.P., Mirkevich R.M. ADVANCED CYBER-PHYSICAL RISK QUANTIFICATION: INTEGRATING SECURITY, CONTEXT AND OPERATIONAL CAPABILITIES IN IIOT AND AUTOMATION ENVIORNMENTS	181
Mikhav V.V., Meleshko Ye.V., Yakymenko M.S., Bosko V.V., Bereziuk I.A. MODELS OF DECENTRALIZED Q&A SYSTEMS IN PEER-TO-PEER COMPUTER NETWORKS..	189
Mohylnyi H.A., Smahina O.O., Pereiaslavskya S.O. ANALYSIS OF THE IMPACT OF CONTAINERIZATION ON THE SOFTWARE DEVELOPMENT LIFE CYCLE (SDLC) IN THE DEVOPS PARADIGM.....	195
Nalotov D.R., Liashenko O.S. COMPLEX TECHNOLOGICAL OBJECT CONTROL MODEL BASED ON FUZZY LOGIC METHODS...	200
Nevliudov I.Sh., Novoselov S.P., Sychova O.V., Tesliuk S.I. DISTRIBUTED POSITIONING SYSTEM FOR MOBILE ROBOTS BASED ON ARUCO SENSOR NETWORK AND ROS ARCHITECTURE FOR CYBER-PHYSICAL MANUFACTURING.....	205
Oleshchenko L.M., Vovchenko D.S. SOFTWARE-BASED LOAD BALANCING METHOD FOR BIG DATA PROCESSING IN DISTRIBUTED INFORMATION SYSTEM.....	213
Parkhomenko Yu.M., Parkhomenko M.D., Matsui A.M., Bokii A.R., Komarov D.Yu. DETERMINATION OF THE INSTALLATION PLACE AND THE ACCURACY OF REGISTRATION BY THE GRAIN FLOW SENSOR FORMED BY A REEL SEEDER.....	220
Pelykh S.M., Lozhechnikov V.F., Ziierov O.R. ALGORITHM-ORIENTED DIGITAL SYSTEM FOR CONTROLLING THE MICROCLIMATE OF PRODUCTION PREMISES.....	227
Pisarenko J.V., Karmazin K.V. NETWORK DIGITAL TWIN AND FEDERATED MULTI-AGENT LEARNING FOR RESILIENT VIDEO STREAMING AND TELEMETRY MANAGEMENT IN SITUATION CENTERS.....	234
Piotrovskiy Ye.Ye. THE ROLE OF AUTOMATED FRONTEND-BACKEND SYNCHRONIZATION IN REDUCING DIGITAL PRODUCT TIME TO MARKET.....	241
Prokopovych-Tkachenko D.I., Torstensson O., Cherkaskiy D.O., Peremetchyk D.O., Bert S.D. SOCIOTECHNICAL FACTORS IN THE EMERGENCE OF VULNERABILITIES IN NETWORK PROTOCOLS FOR SECURE DATA TRANSMISSION.....	249

Pupena O.M. USE OF AUTOMATIONML IN THE LIFECYCLE OF CONTROL SYSTEMS UNDER LIMITED RESOURCES.....	262
Pukhalevych A.V., Makarova L.M., Kovalenko V.V. A MAHEMATICAL MODEL FOR ASSESSING THE QUALITY OF OPEN SOURCE JAVA APPLICATIONS CREATED USING THE HIBERNATE FRAMEWORK.....	271
Rubel Yu.B., Hrytsiuk Yu.I. GENERATION OF SYNTHETIC BALLISTIC DATASET BASED ON NUMERICAL MODELLING AND DOMAIN RANDOMISATION FOR ARTILLERY FIRE CONTROL SYSTEM.....	278
Rudenko Yu.O., Ahadzhanova S.V., Vyunenko O. B., Sytnyk L.H., Ahadzhanov-Honsales K.H. EVALUATION OF THE EFFECTIVENESS OF MASTERING LOW-CODE TECHNOLOGIES IN THE TRAINING OF SPECIALISTS FOR BIG DATA ANALYTICS.....	289
Samoilov O.O., Doshchenko H.G. ARTIFICIAL INTELLIGENCE TECHNIQUES FOR SHIP ELECTRICAL SYSTEM CONDITION MONITORING.....	295
Ushakova I.O. MODERNIZATION OF THE BACKEND OF THE WEB APPLICATION BASED ON MICROSERVICE ARCHITECTURE.....	303
Yanko A.S., Goncharenko S.O. COMPARATIVE ANALYSIS OF ARCHITECTURAL AND CRYPTOGRAPHIC INNOVATIONS IN LORAWAN 1.0.4 AND 1.1 SPECIFICATIONS.....	311
INFORMATION ABOUT AUTHORS.....	318

ІНФОРМАТИКА, ОБЧИСЛЮВАЛЬНА ТЕХНІКА ТА АВТОМАТИЗАЦІЯ

УДК 621.3

DOI <https://doi.org/10.32782/2663-5941/2026.2.1/01>

Ахмедов М.Ш.

<https://orcid.org/0009-0000-7291-4328>

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Федорова Н.В.

<https://orcid.org/0000-0002-4548-4198>

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

ДИНАМІЧНЕ БАЛАНСУВАННЯ НАВАНТАЖЕННЯ В РЕАЛЬНОМУ ЧАСІ ТА ПРОБЛЕМИ ЙОГО ВПРОВАДЖЕННЯ В ЕНЕРГОМЕРЕЖІ SMART GRID

Статтю присвячено розв'язання науково-практичної задачі – підвищенню ефективності балансування навантаження в сучасних енергомережах Smart Grid шляхом розроблення методу динамічного регулювання в реальному часі. Інтенсивне зростання динамічності електроспоживання, поява розподілених джерел генерації та накопичувачів енергії докорінно змінюють характер функціонування енергосистем. Встановлено, що локальні перевантаження й короткочасні дисбаланси виникають значно частіше, ніж це передбачалося для традиційних централізованих схем регулювання, що робить швидкість реакції системи визначальним фактором.

У роботі проаналізовано особливості формування електричного навантаження та визначено функціональні вимоги до систем балансування на базі розподілених мікроконтролерних вузлів. Було обґрунтовано перехід від централізованих моделей до децентралізованого керування на периферійному рівні (edge computing), де рішення приймаються безпосередньо в точці виникнення відхилення. Запропонований метод інтегрує використання мікроконтролерів (STM32, ESP32), буферизацію енергії за допомогою суперконденсаторів та застосування низькорівневих протоколів комунікації (SPI, I2C, UART).

Особливу увагу приділено програмній реалізації на базі операційної системи реального часу FreeRTOS. Досліджено механізми планувальника завдань, пріоритизації переривань та використання черг для забезпечення детермінованості керування в умовах обмежених обчислювальних ресурсів. Використання суперконденсаторів у поєднанні з DC-DC перетворювачами дозволяє фізично компенсувати пікові імпульси споживання протягом мілісекунд, запобігаючи просіданню напруги в локальних сегментах мережі. Результати дослідження формують підґрунтя для створення стабільних, адаптивних мікромереж, здатних підтримувати якість електроенергії без жорсткої ієрархічної координації.

Ключові слова: мікроконтролерне керування, суперконденсатори, embedded-системи, операційні системи реального часу, FreeRTOS, буферизація енергії, розподілені енергетичні системи, детерміноване керування, протоколи низького рівня, енергоефективність систем.

Постановка проблеми. Smart Grid змінює не лише технічну архітектуру енергосистеми, а й спосіб її поведінки. Коли генерація концентрувалася на великих електростанціях, режим мережі був відносно інерційним: навантаження прогнозу-

валося, відхилення коригувалися централізовано. Проте вже з появою розподіленої генерації ситуація стала менш однорідною. Частина потужності формується на рівні домогосподарств або малих об'єктів, частина акумулюється в накопичувачах,

а частина повертається в мережу. Відповідно, потоки більше не мають одного напрямку.

Нестабільність виникає не лише через погодну залежність відновлюваних джерел. Навіть за стабільного виробітку локальні вузли можуть швидко змінювати режим споживання. У певні інтервали виникає надлишок, в інші – дефіцит. Ці коливання не обов'язково критичні окремо, але вони взаємодіють. Частота, напруга й навантаження трансформаторів реагують синхронно, тому мережа перестає поводитися як лінійна система.

Централізоване балансування передбачає, що стан мережі можна описати агрегованими показниками. Однак варто зазначити, що у децентралізованій конфігурації цього недостатньо. Дисбаланс формується на рівні конкретного вузла, але його наслідки поширюються далі, іноді непередбачувано. Реакція із запізненням навіть у кілька секунд може змінити розподіл потоків у сусідніх гілках. Тут важливий не стільки масштаб відхилення, скільки швидкість його розвитку.

Динамічне балансування в реальному часі пов'язане з безперервним аналізом телеметрії та локальним прийняттям рішень. Частина регулювання переноситься ближче до джерела зміни режиму. Це не скасовує центральної координації, але змінює її роль: замість прямого втручання – узгодження автономних дій елементів мережі. За високої частки розподілених ресурсів така модель дозволяє зменшити перевантаження, обмежити втрати та підтримувати допустимі параметри якості електроенергії без жорсткої ієрархії керування.

Питання балансування дедалі частіше розглядається крізь призму стійкості інфраструктури, денавантаження на промислові комплекси, транспорт, цифрові платформи робить мережу чутливою до короткочасних відхилень. У цьому контексті механізми адаптивного регулювання стають складовою загальної архітектури безпеки енергосистеми.

Аналіз останніх досліджень і публікацій. Аналіз сучасних публікацій свідчить, що проблематика балансування навантаження в Smart Grid розвивається нерівномірно – різні автори фокусуються на різних рівнях системи. Частина досліджень концентрується на алгоритмічному аспекті управління. Так, Ф. Р. Албогами (F. R. Albogamy) та співавтори запропонували підхід до керування енергоспоживанням із врахуванням генерації відновлюваних джерел, орієнтований на перерозподіл навантаження в пікові періоди [1]. У наступній роботі ці ж автори зміщують акцент у бік динамічного планування, де рішення коригуються залежно

від змін виробітку та споживання [2]. У цих дослідженнях помітна спроба поєднати оптимізаційні процедури з режимом реального часу.

Інший підхід демонструють Д. Канеллопулос (D. Kanelloropoulos) та В. К. Шарма (V. K. Sharma), які розглядають балансування в контексті IoT-середовища, де ключовим стає масштаб підключених пристроїв і необхідність розподіленої координації [3]. У роботі Г. Григораша (G. Grigoras) та співавторів увага зміщується до конкретного технічного завдання – вирівнювання фаз навантаження на основі даних інтелектуальних лічильників [4]. Тут балансування постає не як абстрактна оптимізаційна проблема, а як інженерна задача з чітко визначеними параметрами мережі.

Паралельно формується напрям, пов'язаний із архітектурними рішеннями. Л. Хагнегахдар (L. Haghnegahdar) та співавтори пропонують мультиагентну cloud-fog модель, у якій частина обчислень перенесена на периферійний рівень для зменшення затримок реагування [5]. Подібну логіку розвивають Д. Ю (D. Yu) та співавтори, обґрунтовуючи доцільність поєднання fog- та cloud-обчислень для підтримки механізмів балансування [6]. У дослідженні Дж. Уддоха (J. Uddoh) та співавторів ця архітектурна модель доповнюється цифровими двійниками на основі штучного інтелекту, що дозволяє моделювати сценарії розподілу ресурсів до їх реалізації в реальній мережі [7]. К. Н. Куреші (K. N. Qureshi) та співавтори зосереджуються на програмно-керованій моделі Smart Grid, підкреслюючи гнучкість і масштабованість такого підходу [8].

Окрему групу становлять роботи, у яких балансування розглядається через децентралізовану взаємодію учасників енергоринку. В. Тушар (W. Tushar) та співавтори аналізують peer-to-peer механізми обміну енергією, де координація відбувається без жорсткої централізації [9]. У дослідженні О. М. Батта (O. M. Butt) та співавторів цифрові технології розглядаються як базовий чинник трансформації управління енергомережами [10]. А в свою чергу М. Інчі (M. Inci) та співавтори акцентують увагу на ролі електромобілів, які здатні виступати не лише споживачами, а й елементами балансування [11]. Робота Т. Мазхара (T. Mazhar) та співавторів розширює цей контекст, пов'язуючи розвиток Smart Grid із концепцією Industry 5.0 та інтелектуалізацією енергетичних процесів [12].

Питання оптимізації функціонування енергосистем узагальнено у дослідженні А. Джошана (A. Joshan), де розглядаються сучасні методи про-

гнозування та адаптивного управління ресурсами в умовах високої динаміки мережових режимів [13]. У цьому підході балансування розглядається як частина ширшої проблеми ефективного управління енергетичною інфраструктурою.

Як бачимо, у публікаціях про Smart Grid переважають алгоритмічні та хмарні рішення. Натомість периферійний рівень, де працюють мікроконтролери з обмеженими ресурсами, описаний значно менше. Саме там виникають практичні труднощі: обмежена пам'ять, нестача обчислювальної потужності, жорсткі часові вимоги. Динамічне балансування в таких умовах не зводиться до оптимізаційної моделі – воно впирається в таймінги, синхронізацію задач і стабільність живлення.

Архітектури з використанням суперконденсаторів і низькорівневих протоколів комунікації також залишаються недостатньо формалізованими. Реалізація детермінованого керування на FreeRTOS залежить від конфігурації планувальника, пріоритетів і апаратних обмежень. Тому невеликі затримки передачі даних або конфлікти задач можуть змінювати поведінку системи при змінному навантаженні. Дослідження зосереджується на цьому рівні реалізації: формуються функціональні вимоги, проектується мікроконтролерна архітектура із суперконденсаторами, розробляється метод динамічного балансування під FreeRTOS з урахуванням ресурсних обмежень і часової детермінованості.

Постановка завдання. Метою статті є підвищення ефективності балансування навантаження в енергомережах Smart Grid шляхом розроблення запропонованого методу динамічного балансування в реальному часі на основі мікроконтролерного керування, використання суперконденсаторів, низькорівневих протоколів комунікації та операційної системи реального часу FreeRTOS.

Для досягнення поставленої мети визначено такі завдання:

1. Проаналізувати особливості формування електричного навантаження в енергомережах Smart Grid і визначити вимоги до систем балансування навантаження в реальному часі на базі розподілених мікроконтролерних вузлів.

2. Обґрунтувати архітектуру та принципи реалізації системи динамічного балансування навантаження із застосуванням мікроконтролерів, суперконденсаторів, низькорівневих протоколів комунікації та FreeRTOS.

3. Розробити запропонований метод динамічного балансування навантаження, що забезпечує

адаптивне та детерміноване керування енергетичними процесами в реальному часі.

Виклад основного матеріалу. Навантаження в Smart Grid формується інакше, ніж у традиційних мережах: генерація розподілена, частина споживачів може тимчасово виступати джерелами, режими роботи змінюються швидко [4]. У результаті виникають локальні піки потужності та нерівномірність розподілу енергії, які не завжди встигає компенсувати централізоване регулювання. За таких умов балансування переноситься ближче до вузлів мережі. Система має не лише фіксувати параметри, а й реагувати без суттєвої затримки. Розподілені керувальні вузли на базі мікроконтролерів забезпечують локальний контроль і скорочують час формування керувального впливу, що особливо помітно при короткочасних перевантаженнях (табл. 1).

У сегментах Smart Grid керування навантаженням фактично відбувається на рівні плати. Мікроконтролери STM32 чи ESP32 постійно працюють із сирими сигналами датчиків струму й напруги, і рішення приймаються там само – без звернення до віддаленого центру [5]. Обмеження пам'яті й тактової частоти відразу накладають рамки на складність алгоритму.

При короткочасних стрибках навантаження активується буферна ланка з суперконденсатором. Підключення здійснюється через керований перетворювач; це типовий сценарій під час запуску двигунів або раптового ввімкнення потужних споживачів. У такі моменти вирішальною стає не оптимізаційна модель, а швидкість перемикання й узгодженість дій вузлів.

Обмін між контролерами відбувається через SPI, UART або I2C – короткі пакети, мінімальна затримка, синхронізація без надлишкових рівнів протоколів. Поведінка системи значною мірою визначається конфігурацією задач у FreeRTOS: порядок виконання, пріоритети, часові інтервали [2]. У мікромережах або зарядних станціях електромобілів ця логіка відчутна буквально – балансування не «планується», воно відбувається в межах кількох мілісекунд.

Коли навантаження в локальному сегменті різко змінюється, система не має часу на централізовану координацію. Затримка в обробці даних навіть у кілька десятків мілісекунд змінює розподіл струмів у сусідніх гілках. Саме тому регулювання зміщується до периферії: мікроконтролер працює безпосередньо з вимірними параметрами, а буферна ланка з суперконденсатором приймає на себе короткий енергетичний імпульс, не перевантажуючи основну лінію.

Архітектура при цьому визначається не теоретичною оптимальністю, а часовими обмеженнями. Обмін між вузлами відбувається через прості інтерфейси з мінімальною затримкою; складність протоколу тут менш важлива, ніж швидкість реакції. У змінному режимі саме локальна синхронізація і послідовність виконання задач утримують параметри мережі в допустимих межах (табл. 2).

У локальному сегменті мережі зміна навантаження відчувається миттєво. Шина живлення просідає ще до того, як це фіксується на рівні всієї системи. Мікроконтролер у цей момент не «аналізує архітектуру» – він працює з конкретними значеннями струму і напруги, отриманими через SPI або I2C. Обчислення миттєвої потужності і швидкості її зростання відбувається в циклі, де кожна затримка має значення. Підключення зарядного пристрою або пуск обладнання створює короткий, але різкий імпульс споживання. Якщо реакція запізнюється, то навантаження переходить на основне джерело. У реальній конфігурації сигнал керування надходить на DC-DC перетворювач (рис. 1), а суперконденсатор підхоплює цей імпульс [8]. Це не оптимізація на рівні моделі, а фізичне перерозподілення енергії в межах мілі-

секунд. Вузли обмінюються короткими повідомленнями через UART або SPI. Ніякої складної мережевої логіки – лише стан і часові мітки.

У мікромережі з кількома джерелами саме така взаємодія формує баланс: кожен модуль повідомляє про свій режим і навантаження розподіляється без централізованого «перерахунку». Конфігурація FreeRTOS тут визначає все: порядок задач, пріоритети, допустимі інтервали реакції [1]. У резервних системах телекомунікацій це відчувається буквально: відсутність навіть короткого провалу напруги означає відсутність перезавантаження обладнання.

У системі балансування рахунок іде на мілісекунди. Якщо обробка сигналу затримується, стрибок навантаження вже встигає вплинути на шину живлення. Тому питання не в тому, чи є багатозадачність, а в тому, чи виконається потрібна задача саме тоді, коли це потрібно. FreeRTOS використовується тут не як формальна надбудова, а як інструмент жорсткого розподілу часу.

Задача вимірювання не чекає завершення другорядних обчислень. Обробник перевантаження має вищий пріоритет, ніж фоновий моніторинг. Планувальник визначає порядок, а не випадко-

Таблиця 1

Функціональні вимоги до систем балансування навантаження в енергомережах Smart Grid

Функціональна вимога	Технічна реалізація	Значення для балансування навантаження
Моніторинг параметрів мережі	Датчики струму і напруги, мікроконтролери	Своєчасне виявлення змін навантаження
Локальна обробка даних	Обчислення параметрів навантаження на мікроконтролері	Скорочення часу реакції системи
Обмін даними між вузлами	Протоколи I2C, UART, SPI	Координація керувальних впливів
Буферизація енергії	Суперконденсатори і силові перетворювачі	Компенсація пікових навантажень
Керування в реальному часі	FreeRTOS і механізми планування завдань	Забезпечення стабільності роботи
Адаптивне регулювання	Алгоритми керування на мікроконтролері	Підвищення ефективності балансування

Джерело: сформовано автором на основі [1; 2; 3; 6; 9]

Таблиця 2

Структурно-функціональні компоненти системи динамічного балансування навантаження

Компонент системи	Функціональне призначення	Технічна реалізація
Сенсорний модуль	Вимірювання електричних параметрів мережі	Датчики струму і напруги, аналогово-цифрові перетворювачі
Мікроконтролерний модуль керування	Обробка даних і формування керувальних сигналів	Мікроконтролери STM32, ESP32, AVR
Модуль накопичення енергії	Буферизація і компенсація пікових навантажень	Суперконденсатори з керованими перетворювачами
Комунікаційний модуль	Передача даних і синхронізація роботи компонентів	Інтерфейси SPI, UART, I2C
Силовий керувальний модуль	Регулювання потоків енергії	DC-DC перетворювачі, електронні ключі
Програмний модуль реального часу	Координація роботи системи	FreeRTOS або інша RTOS

Джерело: сформовано автором на основі [3; 8; 10; 12]

вість. У результаті реакція на зміну режиму не залежить від того, що ще виконується на контролері. Для мікроконтролерної системи принципово що детермінованість стає частиною самої архітектури, а не лише програмною опцією (табл. 3).

У контролері балансування кілька процесів відбуваються одночасно і саме це створює основну складність. Опитування АЦП триває з фіксованим інтервалом, обчислення параметрів навантаження накладається на нього, а формування керувального імпульсу не може чекати завершення фонових процедур. FreeRTOS у цій конфігурації не просто розподіляє задачі, ця операційна система реального часу задає часову дисципліну. Різкий стрибок струму ініціює апаратне переривання, а обробник із вищим пріоритетом витісняє менш критичні процеси. Черги передають вимірювальні

дані без втрат, семафори обмежують доступ до спільних ресурсів. Якщо система налаштована коректно, реакція залишається стабільною навіть тоді, коли загальне обчислювальне навантаження зростає. У контролерах інверторів або автономних енергетичних модулів саме така поведінка визначає чи збережеться напруга в допустимому діапазоні під час різкої зміни режиму [9].

Мікроконтролер у системі балансування не працює в комфортних умовах. Поки він рахує миттєву потужність, у черзі вже накопичуються нові вимірювання. Якщо частоту оновлення підняти занадто високо, процесор починає втрачати темп, але не через складність алгоритму, а через банальний дефіцит часу і пам'яті на ядрі/ядрах процесора. Те, що в моделі виглядає як «оновлення в реальному часі», на платі перетворює-

Таблиця 3

Механізми FreeRTOS для реалізації керування навантаженням у режимі реального часу

Механізм FreeRTOS	Функціональне призначення	Значення для керування навантаженням
Планувальник завдань (Scheduler)	Визначення порядку виконання завдань відповідно до пріоритетів	Забезпечує своєчасне виконання критичних керувальних операцій
Завдання (Tasks)	Реалізація окремих функціональних процесів	Дозволяє розділити вимірювання, аналіз і керування
Переривання (Interrupts)	Обробка асинхронних подій	Забезпечує негайну реакцію на зміну параметрів навантаження
Черги (Queues)	Передача даних між завданнями	Забезпечує узгоджений обмін вимірювальною і керувальною інформацією
Семафори і м'ютекси	Синхронізація доступу до ресурсів	Запобігає конфліктам при одночасному виконанні процесів
Таймери реального часу	Формування періодичних подій	Забезпечує регулярне виконання керувальних алгоритмів

Джерело: сформовано автором на основі [1; 3; 10; 13]

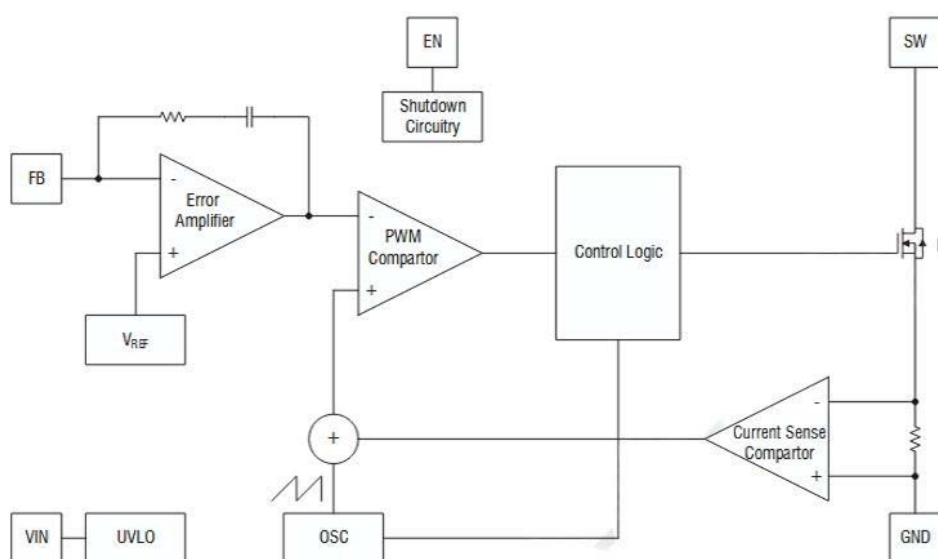


Рис. 1. Функціональна діаграма DC-DC перетворювача MT3608

Джерело: [14]

ється на конкуренцію за процесорний час. Обмін даними додає ще один рівень напруги. Інтерфейс передає пакет із невеликою затримкою, а оцінка стану вже не може вважатись актуальною [7]. Якщо кілька вузлів реагують асинхронно, їхні керувальні імпульси можуть накладатися. У силовій частині це проявляється перехідними процесами, які не згасають миттєво [3]. Зовні система ніби працює, але в піковий момент запас стійкості зменшується. Навіть програмна багатозадачність не гарантує стабільності, так як один невчасно виконаний обробник переривання – і критична операція може зміститися в часі. У таких умовах стабільність не задається декларативно, її доводиться буквально «виборювати» конфігурацією таймерів, черг і пріоритетів. Запропонований метод динамічного балансування побудований із припущення, що реакція має формуватися там, де виникає дисбаланс, а не на рівні централізованого контролера. Тому керувальний вузол розглядається як замкнений контур: вимірювання, аналіз і регулювання об'єднані в межах одного мікроконтролера. Сенсорні дані струму та напруги надходять через I2C або SPI безпосередньо в обчислювальний модуль. Оцінюється не лише миттєве значення потужності, а й швидкість її зміни. Саме цей параметр визначає момент втручання. Якщо навантаження зростає різко, мікроконтролер формує сигнал для DC-DC перетворювача, а суперконденсатор короткочасно підключається до шини. Буфер працює в імпульсному режимі, згладжуючи провал або перевищення до того, як вони вплинуть на інші вузли. FreeRTOS використовується як механізм часової організації. Задачі

вимірювання, аналізу та формування керувального сигналу мають різні пріоритети, а їх виконання прив'язане до заданих інтервалів. Важливо не саме багатозадачне виконання, а гарантія того, що обробка перевантаження не буде витіснена другорядними процесами. За обмежених ресурсів мікроконтролера саме така конфігурація зберігає детермінованість керування (рис. 2).

Метод реалізується не як окремий модуль «поза системою», а як частина конкретного сегмента. Наприклад, у вузлі живлення автономного об'єкта або в інверторному блоці мікромережі. Мікроконтролер постійно отримує дані про струм і напругу, оскільки режим мережі змінюється, іноді різко. Якщо споживання зростає швидше, ніж основне джерело здатне відреагувати, формується сигнал на DC-DC перетворювач, де суперконденсатор короткочасно підключається до шини. Буфер працює імпульсно: він не замінює джерело живлення, а лише згладжує провал або піковий стрибок. Коли параметри повертаються до робочого діапазону, накопичувач знову переходить у режим заряджання. Уся послідовність – від вимірювання до формування керувального сигналу, – організована через FreeRTOS. Пріоритети задач і часові інтервали задають порядок реакції, де система не чекає завершення другорядних процесів. У такій конфігурації балансування відбувається локально, без передачі запиту на вищий рівень керування. Навантаження на силові компоненти не накопичується в пікові моменти, а напруга не встигає вийти за допустимі межі.

Висновки. В умовах швидких змін навантаження реакція централізованого регулювання



Рис. 2. Структура розробленого методу динамічного балансування навантаження

Джерело: власна розробка автора

часто виявляється запізнілою. Перенесення обробки даних і формування керувального сигналу на рівень локального вузла змінює часову структуру процесу: вимірювання, оцінка стану і компенсація відбуваються в одному циклі без передачі рішення на вищий рівень.

Мікроконтролер, суперконденсатор і FreeRTOS утворюють замкнений контур. Сенсорні дані обробляються безпосередньо на платі, короточасні піки навантаження гасяться підключенням буферної ємності через DC-DC перетворювач. Детермінованість забезпечується конфігурацією задач і їх пріоритетів, а не додатковими обчислювальними ресурсами.

Разом із цим обмеження апаратної платформи, затримки комунікації та вимоги до синхронізації між вузлами залишаються критичними факторами. За змінного режиму навіть незначне зміщення у часі впливає на поведінку всієї конфігурації. Тому архітектурні рішення, алгоритмічна оптимізація і організація обміну даними розглядаються як єдиний комплекс.

Подальший розвиток методу може бути пов'язаний із підвищенням адаптивності системи, уточненням механізмів прогнозування навантаження та розширенням масштабованості розподіленої структури.

Список літератури:

1. Albogamy F. R., Khan S. A., Hafeez G., Murawwat S., Khan S., Haider S. I., Thoben K. D. Real-time energy management and load scheduling with renewable energy integration in smart grid. *Sustainability*. 2022. Vol. 14, № 3. DOI: <https://doi.org/10.3390/su14031792>.
2. Albogamy F. R., Paracha M. Y. I., Hafeez G., Khan I., Murawwat S., Rukh G., Khan M. U. A. Real-time scheduling for optimal energy optimization in smart grid integrated with renewable energy sources. *IEEE Access*. 2022. Vol. 10. DOI: <https://doi.org/10.1109/ACCESS.2022.3161845>.
3. Kanellopoulos D., Sharma V. K. Dynamic load balancing techniques in the IoT: A review. *Symmetry*. 2022. Vol. 14, № 12. DOI: <https://doi.org/10.3390/sym14122554>.
4. Grigoraş G., Neagu B. C., Gavrilas M., Triştiu I., Bulac C. Optimal phase load balancing in low voltage distribution networks using a smart meter data-based algorithm. *Mathematics*. 2020. Vol. 8, № 4. DOI: <https://doi.org/10.3390/math8040549>.
5. Haghnegahdar L., Chen Y., Wang Y. Enhancing dynamic energy network management using a multiagent cloud-fog structure. *Renewable and Sustainable Energy Reviews*. 2022. Vol. 162. DOI: <https://doi.org/10.1016/j.rser.2022.112439>.
6. Yu D., Ma Z., Wang R. Efficient smart grid load balancing via fog and cloud computing. *Mathematical Problems in Engineering*. 2022. Vol. 2022, № 1. DOI: <https://doi.org/10.1155/2022/3151249>.
7. Uddoh J., Ajiga D., Okare B. P., Aduloju T. D. Developing AI optimized digital twins for smart grid resource allocation and forecasting. *Journal of Frontiers in Multidisciplinary Research*. 2021. Vol. 2, № 2. URL: <https://doi.org/10.54660/IJFMR.2021.2.2.55-60> (date of access: 05.02.2026).
8. Qureshi K. N., Hussain R., Jeon G. A distributed software defined networking model to improve the scalability and quality of services for flexible green energy internet for smart grid systems. *Computers & Electrical Engineering*. 2020. Vol. 84. DOI: <https://doi.org/10.1016/j.compeleceng.2020.106634>.
9. Tushar W., Saha T. K., Yuen C., Smith D., Poor H. V. Peer-to-peer trading in electricity networks: An overview. *IEEE Transactions on Smart Grid*. 2020. Vol. 11, № 4. DOI: <https://doi.org/10.1109/TSG.2020.2969657>.
10. Butt O. M., Zulqarnain M., Butt T. M. Recent advancement in smart grid technology: Future prospects in the electrical power network. *Ain Shams Engineering Journal*. 2021. Vol. 12, № 1. DOI: <https://doi.org/10.1016/j.asej.2020.05.004>.
11. Inci M., Çelik Ö., Lashab A., Bayındır K. Ç., Vasquez J. C., Guerrero J. M. Power system integration of electric vehicles: A review on impacts and contributions to the smart grid. *Applied Sciences*. 2024. Vol. 14, № 6. DOI: <https://doi.org/10.3390/app14062246>.
12. Mazhar T., Shahzad T., Rehman A. U., Hamam H. Integration of smart grid with industry 5.0: applications, challenges and solutions. *Measurement: Energy*. 2025. Vol. 5. DOI: <https://doi.org/10.1016/j.meane.2024.100031>.
13. Joshan A. Emerging trends and advanced techniques in power system optimization for future smart grids. *Power, Control, and Data Processing Systems*. 2025. Vol. 2, № 2. DOI: 10.30511/pcdp.2025.2060096.1022.
14. MT3608 2A, High Efficiency, 1.2MHz Step-Up Converter. AEROSEMI. [Електронний ресурс]. – Режим доступу: https://e2e.ti.com/cfs-file/__key/communityserver-discussions-components-files/196/4012002220_2D004753D358_IC_5B00_MT3608_2C00_SMD_2C00_SOT23_2D00_6L_2C00_AEROSEMI_5D00_.pdf (дата звернення: 15.02.2026).

**Akhmedov M.Sh., Fedorova N.V. DYNAMIC LOAD BALANCING IN REAL TIME
AND IMPLEMENTATION CHALLENGES IN SMART GRID ENERGY NETWORKS**

The article is devoted to solving a scientific and practical problem – increasing the efficiency of load balancing in modern Smart Grid power networks by developing a method of dynamic regulation in real time. The intensive growth of the dynamics of electricity consumption, the emergence of distributed sources of generation and energy storage fundamentally change the nature of the functioning of power systems. It was established that local overloads and short-term imbalances occur much more often than was expected for traditional centralized control schemes, which makes the speed of the system's response a determining factor.

The paper analyzes the features of the formation of electrical load and determines the functional requirements for balancing systems based on distributed microcontroller nodes. The transition from centralized models to decentralized control at the peripheral level (edge computing) was substantiated, where decisions are made directly at the point of deviation. The proposed method integrates the use of microcontrollers (STM32, ESP32), energy buffering using supercapacitors and the use of low-level communication protocols (SPI, I2C, UART).

Particular attention is paid to the software implementation based on the real-time operating system FreeRTOS. The mechanisms of the task scheduler, interrupt prioritization and the use of queues to ensure deterministic control under conditions of limited computing resources are investigated. The use of supercapacitors in combination with DC-DC converters allows for physical compensation of peak consumption pulses within milliseconds, preventing voltage sags in local network segments. The research results form the basis for the creation of stable, adaptive microgrids capable of maintaining electricity quality without rigid hierarchical coordination.

Keywords: *microcontroller-based control, supercapacitors, embedded systems, real-time operating systems, FreeRTOS, energy buffering, distributed energy systems, deterministic control, low-level communication protocols, system energy efficiency.*

Дата першого надходження статті до видання: 19.02.2026

Дата прийняття статті до друку після рецензування: 16.03.2026

Дата публікації (оприлюднення) статті 11.05.2026

Баглай О.П.<https://orcid.org/0009-0003-7814-0206>

Український державний університет залізничного транспорту

ВИКОРИСТАННЯ СУЧАСНИХ СТАТИСТИЧНИХ МЕТОДІВ У ЗВАРЮВАННІ МЕТАЛОКОНСТРУКЦІЙ

Статтю присвячено аналізу сучасних статистичних методів та інформаційних технологій у задачах контролю якості, прогнозування дефектів і оптимізації процесів зварювання металоконструкцій. Показано обмеженість традиційних емпіричних підходів у використанні великих масивів технологічних даних реального часу. У роботі розглянуто перехід від суб'єктивного контролю до даних-орієнтованого аналізу процесів зварювання в контексті Індустрії 4.0. Особливу увагу приділено методам описової та математичної статистики для виявлення закономірностей між вхідними параметрами процесу зварювання (струм, напруга) та вихідними характеристиками зварного шва (стабільність процесу, ймовірність утворення дефектів).

На основі датасету *Metal Arc Welding – Predictive Quality Data* [1] виконано попередню обробку та стандартизацію часових рядів, візуалізацію даних, обчислення описової статистики, кореляційний аналіз (коефіцієнт Пірсона) [2, с. 3], перевірку нормальності розподілів за критерієм Шاپіро-Уїлка [3, с. 92] та порівняння класів якості за допомогою непараметричного тесту Манна-Уїтні.

Візуалізація показала, що дефектні шви мають нестабільну дугу з глибокими провалами струму та піками напруги, тоді як якісні – плавну динаміку. Статистичний аналіз підтвердив значущі відмінності ($p < 0.001$): середній струм у дефектних швах вищий на 0.492 стандартних відхилення, дисперсія струму більша на 11,7 %, а кореляція між струмом і напругою змінюється з негативної ($r = -0.143$) на помірну позитивну ($r = 0.286$). Розподіли ненормальні ($p < 0.001$). Отримані статистичні характеристики були використані як ознаки для побудови моделей класифікації. Застосування логістичної регресії як параметричного статистичного методу дозволило досягти точності 70.1 % ($\text{ROC-AUC} = 0.79$), тоді як непараметричний метод k найближчих сусідів ($k\text{-NN}$) продемонстрував вищу ефективність з точністю 76.4 % та $\text{ROC-AUC} = 0.85$, що підтверджує інформативність статистичних ознак для прогнозування якості зварного шва. Отримані дані можуть слугувати основою для розробки інтелектуальних систем моніторингу та автоматизованого управління процесами зварювання в реальному часі, що відповідає завданням цифрової трансформації промисловості та підвищення конкурентоспроможності українського машинобудування.

Ключові слова: зварювання металоконструкцій, статистичні методи, аналіз часових рядів, контроль якості, Індустрія 4.0, якість зварювання.

Постановка проблеми. Незважаючи на прогрес у технологіях зварювання, забезпечення стабільної якості зварних з'єднань у металоконструкціях залишається актуальною проблемою машинобудування та будівництва. Дефектні шви (тріщини, пористість, геометричні спотворення) є однією з основних причин передчасного руйнування конструкцій під дією навантажень, корозії та втоми матеріалу. За оцінками, дефекти зварювання спричиняють 20-30 % аварій на критичних об'єктах, що призводить до значних економічних втрат і загрози безпеці. [4, с. 1].

Традиційні методи контролю якості зварних швів базуються переважно на візуальному огляді,

вибіркових неруйнівних випробуваннях (ультразвук, рентген, магнітно-порошкова дефектоскопія) та досвіді зварювальника. Такі підходи мають суттєві обмеження:

- вони є післяпроцесними (виявляють дефекти вже після завершення зварювання);
- не враховують стохастичний характер процесу (випадкові коливання струму, напруги, швидкості подачі дроту, зовнішніх умов);
- не дозволяють прогнозувати ймовірність дефектів і оперативно коригувати параметри процесу [5, с. 406].

Унаслідок цього значна частина дефектів виявляється лише на етапі експлуатації, що зни-

жує надійність конструкцій і збільшує витрати на ремонт. Водночас розвиток сенсорних систем, цифрових технологій збору даних та концепцій Індустрії 4.0 створює передумови для переходу до проактивного, даних-орієнтованого управління якістю зварювання [6, с. 21]. Основною проблемою залишається відсутність системного застосування сучасних статистичних методів для аналізу й прогнозування параметрів зварювального процесу в реальному часі, що обмежує можливості раннього виявлення нестабільності дуги та автоматизованого коригування режимів зварювання.

Аналіз останніх досліджень і публікацій. Аналіз наукових публікацій свідчить про поступову еволюцію підходів до забезпечення якості та надійності зварних з'єднань у металоконструкціях – від емпіричних методів підбору параметрів до комплексного використання статистичних, чисельних та інформаційних технологій аналізу даних. Сучасні дослідження зосереджені на оптимізації параметрів процесу зварювання, прогнозуванні експлуатаційних характеристик і автоматизованому контролю якості.

У роботі Abatta-Jacome та ін. (2025) [7, с. 13] для оцінки механічних властивостей зварних і незварних арматурних прутків зі сталі ASTM A706 застосовано класичні статистичні методи (критерії Шапіро-Уїлка, Андерсона-Дарлінга та Манна-Уїтні). Отримані результати підтвердили ефективність статистичних тестів для виявлення прихованих змін міцнісних характеристик і обґрунтували доцільність їх використання для стандартизації процесів зварювання, що узгоджується з підходами, використаними в даній роботі.

У публікації Muhammed та ін. (2020) [8, с. 22] автори показали можливість прогнозування корозійної стійкості та міцнісних характеристик на основі статистичних моделей, що ілюструє перехід від якісного аналізу до формалізованих, даних-орієнтованих підходів, подібних до застосованих у цій роботі.

У роботі Zhang та ін. (2020) [9, с. 13] проаналізовано комп'ютеризовані методи виявлення дефектів зварювання за рентгенографічними зображеннями та показано обмеження традиційних неруйнівних методів, зокрема їх післяпроцесний характер і залежність від оператора. Автори обґрунтовують доцільність переходу до статистичних і даних-орієнтованих підходів прогнозування дефектів у реальному часі.

Дослідження Sumesh та ін. (2021) [10, с. 8] підтвердило ефективність використання машинного навчання (SVM) для прогнозування якості зварю-

вання на основі сенсорних даних, що підкреслює перспективність статистичних і даних-орієнтованих методів контролю процесу.

Незважаючи на це, системний аналіз часових рядів електричних параметрів у реальному часі та їх інтеграція в автоматизовані системи моніторингу залишаються недостатньо вивченими.

Постановка завдання. Метою статті є аналіз сучасних статистичних методів обробки даних зварювальних процесів, виявлення закономірностей між параметрами процесу (струм, напруга) та стабільністю зварного шва, а також обґрунтування перспективності статистичного підходу для автоматизованого контролю якості в умовах Індустрії 4.0.

Для досягнення цієї мети необхідно виконати такі завдання:

1. Проаналізувати сучасні підходи до застосування статистичних методів і даних-орієнтованих моделей у задачах аналізу та контролю якості зварних з'єднань;
2. Виконати попередню обробку та стандартизацію часових рядів електричних параметрів процесу зварювання на основі відкритого датасету;
3. Здійснити візуалізацію даних, обчислення описових статистичних характеристик і кореляційний аналіз для виявлення відмінностей між якісними та дефектними швами;
4. Перевірити статистичні гіпотези щодо відмінностей між класами якості з використанням непараметричних критеріїв, а також оцінити можливість застосування моделей логістичної регресії та методу k найближчих сусідів для класифікації якості зварного шва;
5. Узагальнити результати та оцінити ефективність статистичних методів і моделей для автоматизованого контролю якості зварювання.

Виклад основного матеріалу. Аналіз сучасних підходів до застосування статистичних методів у задачах контролю якості зварних з'єднань, проведений у попередньому розділі, показав, що більшість досліджень зосереджена на постфактум-оцінці механічних властивостей матеріалів або на обмеженому використанні статистики для окремих параметрів процесу. Водночас системний аналіз часових рядів електричних параметрів (струм, напруга) у реальному часі залишається недостатньо вивченим.

Для практичної перевірки потенціалу статистичних методів у виявленні закономірностей між параметрами процесу та стабільністю шва використано відкритий датасет Metal Arc Welding – Predictive Quality Data [1]. Даний датасет містить

часові ряди значень струму та напруги під час зварювання, а також позначки якості зварного шва («добре» або «погано»), що дозволяє поєднувати статистичний аналіз параметрів процесу з оцінкою стабільності результатів.

Дослідження проведено за допомогою мови програмування Python з використанням бібліотек pandas для обробки даних, matplotlib та seaborn [11, с.1] для візуалізації, scipy для статистичних тестів [12, с.267], а також а також scikit-learn для побудови моделей класифікації.

На першому етапі дослідження було проведено попередню обробку даних. Початковий датасет містив 200 тис. зварювальних циклів із часовими рядами струму та напруги. Розподіл класів якості мав такий вигляд:

- клас «-1» (без мітки) – 112777 записів (53,9 %);
- клас «0» (погано) – 49742 записи (23,8 %);
- клас «1» (добре) – 46666 записів (22,3 %).

Після фільтрації залишено лише розмічені приклади (класи 0 та 1) загальною кількістю 96408 записів. Пропущених значень не виявлено. Виявлені поодинокі аномалії (0,01 %), що виходили за фізично обґрунтовані межі, були скориговані лінійною інтерполяцією. Усі параметри струму та напруги стандартизовано для забезпечення коректності подальшого статистичного аналізу [13, с. 119].

Після попередньої обробки та стандартизації даних розпочато аналіз агрегованих характеристик часових рядів. Для швидкості розрахунків

та зручності візуалізації спочатку використано частину датасету (nrows = 15000 записів). У цій вибірці розподіл класів якості склався так: 11681 запис класу «погано» (0) та 3319 записів класу «добре» (1).

Візуальний аналіз часових рядів струму та напруги (рис. 1) показав чіткі відмінності між якісними та дефектними швами: для класу «добре» характерна стабільна динаміка параметрів, тоді як для класу «погано» спостерігається хаотична поведінка з провалами струму та нестабільністю напруги.

Гістограма середнього струму (рис. 2) підтверджує ці відмінності, демонструючи ширший і зміщений розподіл для дефектних швів порівняно з компактним розподілом для якісних.

Діаграма розсіювання середнього струму та напруги (рис. 3) показує чітке розділення класів: якісні шви утворюють компактний кластер зі стабільними значеннями, тоді як дефектні характеризуються значним розкидом і екстремальними відхиленнями.

Візуальні спостереження підтвердили суттєві відмінності в поведінці параметрів процесу залежно від якості шва. Для кількісної оцінки цих відмінностей проведено статистичний аналіз агрегованих характеристик часових рядів за класами «добре» (1) та «погано» (0). Описова статистика представлена в (табл. 1,2). В таблиці 1 представлена описова статистика агрегованих ознак за класами якості, усі значення стандар-

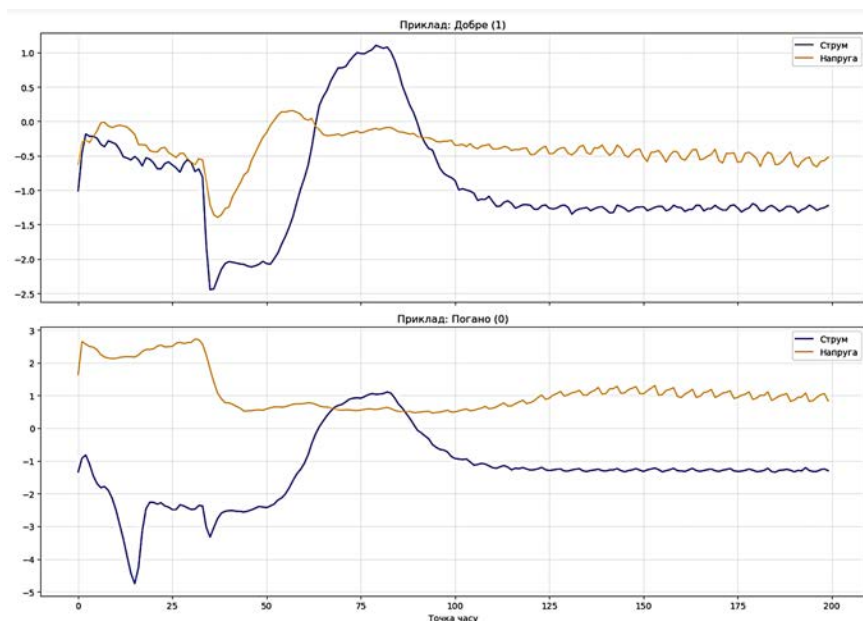


Рис. 1. Часові ряди

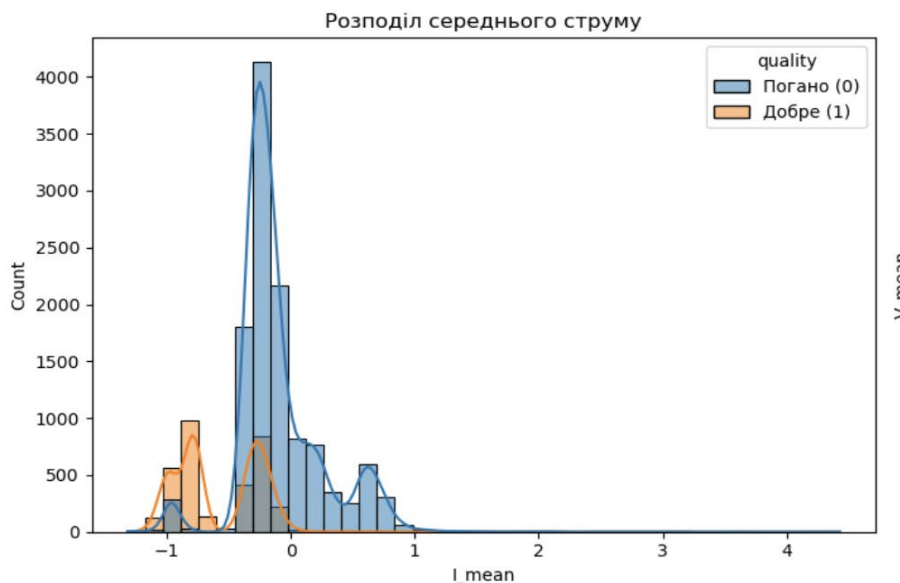


Рис. 2. Гістограма середнього струму

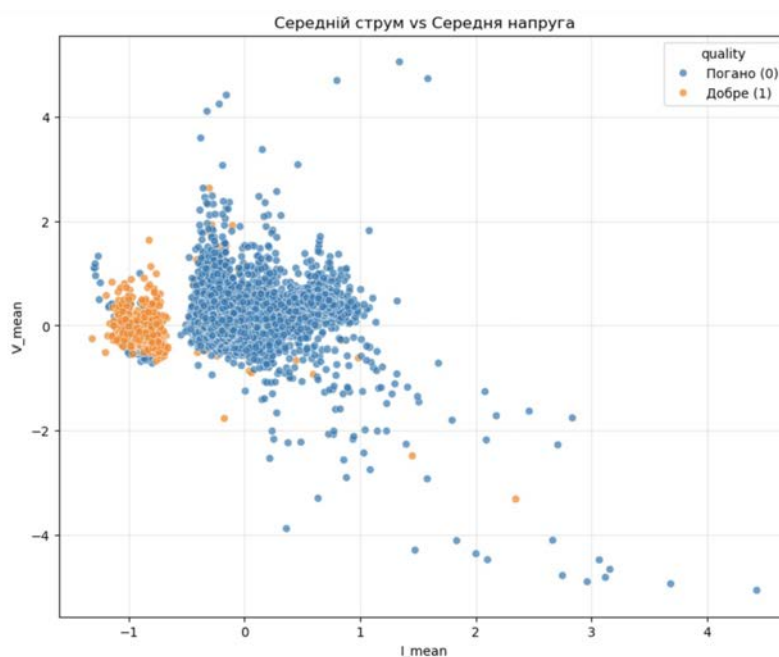


Рис. 3. Діаграма розсіювання

тизовани, тому середнє ≈ 0 , а стандартне відхилення ≈ 1 . Проаналізувавши отримані результати, можна побачити, що середній струм у дефектних швах вищий на 0.492 стандартних відхилення. Водночас дисперсія струму та напруги вища саме в якісних швах: на 11.7 % для струму та на 20.8 % для напруги. Це свідчить про те, що якісне зварювання характеризується контролю-

ваними коливаннями, тоді як дефектне – зміщенням середнього значення струму при меншій загальній варіативності.

Отримані дані свідчать, що основною ознакою дефектності є не стільки дисперсія, скільки зміщення середнього струму, що може бути використано для автоматизованого виявлення відхилень.

Таблиця 1

Описова статистика агрегованих ознак
за класами якості

Параметр	Клас	Середнє	Стандартне відхилення
Середній струм	Добре (1)	-0,254	0,325
Середній струм	Погано (0)	0,238	0,539
Середня напруга	Добре (1)	-0,238	0,612
Середня напруга	Погано (0)	0,223	0,539
Дисперсія струму	Добре (1)	0,651	0,401
Дисперсія струму	Погано (0)	0,583	0,301
Дисперсія напруги	Добре (1)	0,645	0,581
Дисперсія напруги	Погано (0)	0,534	0,455

Для оцінки взаємозв'язку між параметрами проведено кореляційний аналіз. Кореляційний аналіз показав слабку позитивну залежність між середніми значеннями струму та напруги в усіх даних ($r = 0.207$). Однак у якісних швах кореляція стає слабко негативною ($r = -0.143$), що відповідає стабільній роботі дуги (зростання струму супроводжується зменшенням напруги). У дефектних швах кореляція помірна позитивна ($r = 0.286$), що свідчить про порушення стабільності процесу – струм і напруга змінюються синхронно в одному напрямку, що характерно для неконтрольованої дуги.

Перевірка нормальності за критерієм Шапіро-Уїлка показала ненормальний характер розподілів ($p < 0.001$), у зв'язку з чим для порівняння класів застосовано непараметричний тест Манна-Уїтні (табл. 2). Результати підтвердили статистично значущі відмінності між класами за всіма параметрами, що узгоджується з результатами візуального та кореляційного аналізу.

Отримані результати описового, кореляційного та непараметричного аналізу підтвердили наявність стійких відмінностей між якісними та дефектними зварними швами за ключовими параметрами процесу, що створює передумови для переходу від інтерпретаційного аналізу до формалізованого статистичного моделювання.

Таблиця 2

Порівняння класів якості
за тестом Манна-Уїтні

Параметр	Статистика тесту	p-value	Висновок
Середній струм	61557111172	$< 0,001$	Значущі відмінності
Середня напруга	515983654	$< 0,001$	Значущі відмінності
Дисперсія струму	123368111714	$1,20e-69$	Значущі відмінності
Дисперсія напруги	130908084229	$5,65e-259$	Значущі відмінності

На наступному етапі дослідження застосовано сучасні статистичні методи класифікації – логістичну регресію та метод k найближчих сусідів (k-NN), які, на відміну від класичних статистичних тестів, дозволяють здійснювати прогноз якості зварного шва на основі сукупності параметрів процесу. Логістична регресія використовується для моделювання ймовірності належності до класу, тоді як метод k-NN як непараметричний підхід не потребує припущень щодо форми розподілів, що є доцільним для зварювальних процесів із ненормальними статистичними характеристиками.

Для оцінки якості побудованих моделей використано стандартні метрики класифікації: accuracy, precision, recall, F1-score та площу під ROC-кривою (ROC-AUC). Оцінювання проводилося на тестовій вибірці, що не використовувалась під час навчання моделей.

Результати логістичної регресії свідчать про задовільну здатність моделі до розрізнення якісних і дефектних швів: точність класифікації становить 70,1 %, а значення ROC-AUC дорівнює 0,79, що вказує на добру дискримінаційну здатність моделі. Повнота для класу «добре» становить 74,2, що свідчить про ефективне виявлення якісних швів.

Метод k найближчих сусідів продемонстрував вищі показники ефективності порівняно з логістичною регресією: точність – 76,6 %, ROC-AUC – 0,85. Це свідчить про кращу адаптацію непараметричної статистичної моделі до складної структури даних та підтверджує перспективність її використання для автоматизованого контролю якості зварювання. Узагальнені результати оцінювання подано в табл. 3.

Таблиця 3

Порівняння ефективності
статистичних моделей

Модель	Accuracy, %	Precision	Recall	F1-score	ROC-AUC
Логістична регресія	70,1	67,4	74,2	70,6	0,79
k-NN	76,6	76,0	75,0	75,5	0,85

Матриці помилок для логістичної регресії та методу k-NN (рис. 4) демонструють, що модель k-NN забезпечує більш збалансовану класифікацію обох класів, зменшуючи кількість хибних класифікацій як для дефектних, так і для якісних швів. Це підтверджує її вищу чутливість до локальних особливостей даних.

ROC-криві для обох моделей (рис. 5) підтверджують перевагу методу k-NN, площа під кривою

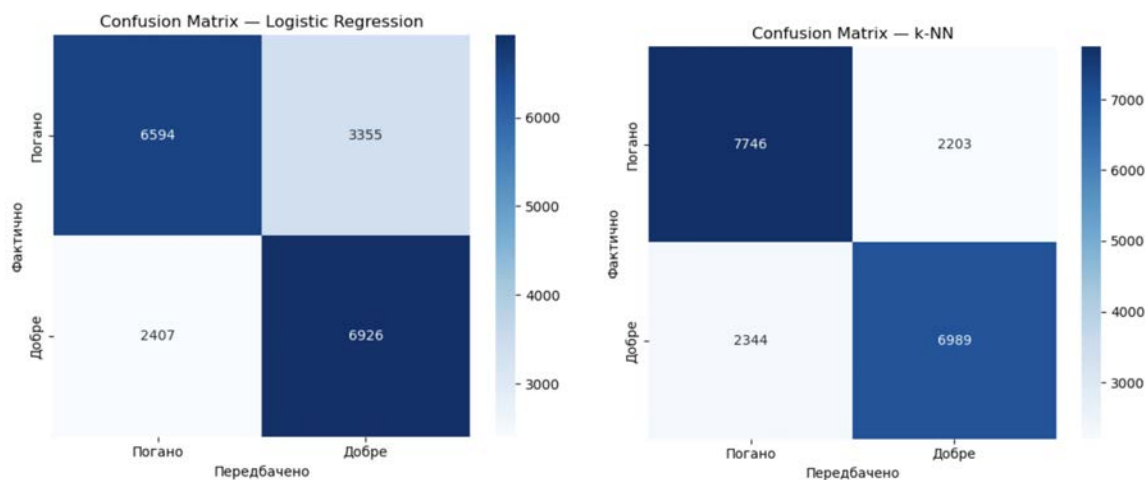


Рис. 4. Матриці помилок для логістичної регресії та k-NN

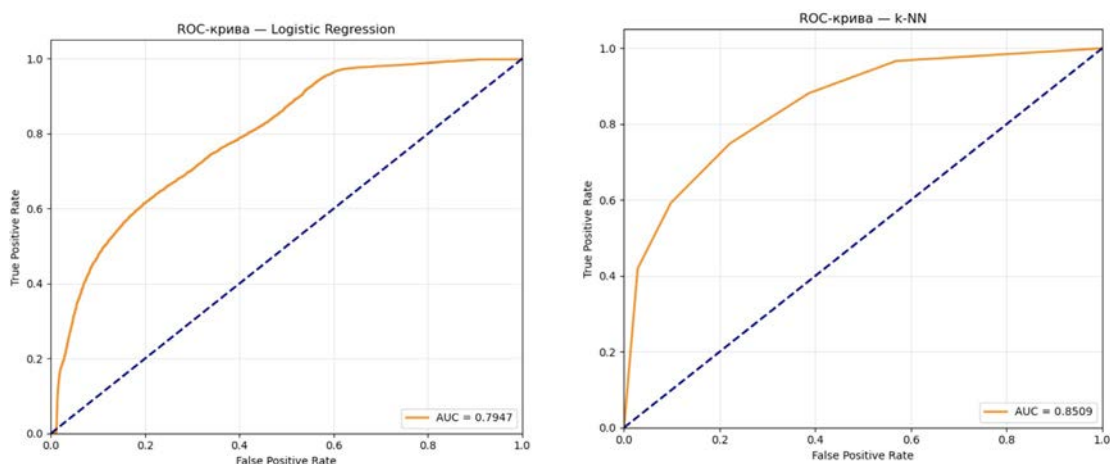


Рис. 5. ROC-криві статистичних моделей класифікації

якого становить 0,85, порівняно з 0,79 для логістичної регресії. Це свідчить про вищу здатність непараметричної статистичної моделі коректно розрізняти класи за різних порогів прийняття рішення, що є критично важливим для задач автоматизованого контролю якості зварювання.

Отримані результати підтверджують, що сучасні статистичні методи класифікації, зокрема логістична регресія та метод k найближчих сусідів, здатні ефективно використовувати агреговані статистичні характеристики часових рядів для прогнозування якості зварних швів. Особливо ефективним виявився непараметричний підхід k-NN, що узгоджується з ненормальним характером розподілів параметрів процесу.

Висновки. У результаті проведеного дослідження виконано поставлені завдання та отримано теоретично обґрунтовані й кількісно підтверджені результати щодо можливостей застосування статистичних методів для аналізу й прогнозування

якості зварних швів на основі часових рядів електричних параметрів процесу.

Проаналізовано сучасні підходи до застосування статистичних і даних-орієнтованих методів у задачах контролю якості зварювання. Встановлено, що більшість досліджень зосереджена на постпроцесному аналізі або використанні окремих статистичних тестів, тоді як системний аналіз часових рядів електричних параметрів у реальному часі залишається недостатньо розвиненим.

Виконано попередню обробку та стандартизацію даних відкритого датасету (200 тис. зварювальних циклів). Після фільтрації залишено 96408 розмічених записів (23,8 % – «погано», 22,3 % – «добре»). Аномальні значення становили 0,01 % і були скориговані. Усі ознаки стандартизовано ($\mu \approx 0$, $\sigma \approx 1$), що забезпечило коректність подальшого статистичного аналізу та моделювання.

Виявлено кількісні відмінності між класами якості. Середній струм у дефектних швах вищий на

0,492 стандартних відхилення. Дисперсія струму у якісних швах більша на 11,7 %, а напруги – на 20,8 %. Кореляція між середніми значеннями струму та напруги для якісних швів є слабко негативною ($r = -0,143$), тоді як для дефектних – помірно позитивною ($r = 0,286$). Непараметричний тест Манна-Уїтні підтвердив статистично значущі відмінності за всіма параметрами ($p < 0,001$).

Оцінено ефективність статистичних моделей класифікації. Логістична регресія продемонструвала асигурацію 70,1 % та ROC-AUC 0,79, тоді як

метод k-NN – 76,6 % та 0,85 відповідно. Приріст точності становить 6,5 %, що свідчить про кращу адаптацію непараметричного підходу до складної структури даних.

Підтверджено практичну доцільність інтеграції статистичних методів у системи автоматизованого контролю якості. Отримані моделі забезпечують дискримінаційну здатність до 85 % (ROC-AUC) та можуть бути використані як основа для реалізації проактивного моніторингу процесу зварювання в рамках концепції Індустрії 4.0.

Список літератури:

1. Hahn Y. Metal Arc Welding – Predictive Quality Arc Welding Dataset (Version v2) [Electronic resource]. Zenodo, 2023. DOI: 10.5281/zenodo.10017718.
2. Šverko Z., Vrankić M., Vlahinić S., Rogelj P. Complex Pearson correlation coefficient for EEG connectivity analysis. *Sensors*. 2022. Vol. 22, no. 4. Article 1477. DOI: 10.3390/s22041477.
3. Hanusz Z., Tarasinska J., Zielinski W. Shapiro-Wilk test with known mean. *SHILAP Revista de Lepidopterologia*. 2022. DOI: 10.57805/revstat.v14i1.180.
4. Kowalczyk M., Andruszko J., Stefanek P., Mazur R. Failure and degradation mechanisms of steel pipelines: analysis and development of effective preventive strategies. *Materials*. 2025. Vol. 18, no. 1. Article 134. DOI: 10.3390/ma18010134.
5. Zhang Y. та ін. Real-time monitoring of welding process using deep learning. *Journal of Manufacturing Processes*. 2023. DOI: 10.1016/j.jmapro.2023.01.012.
6. Баглай О.П., Тимофеева М.А. Застосування інформаційних технологій для контролю якості зварних з'єднань металоконструкцій // Вісник Херсонського національного технічного університету. – Херсон : ХНТУ, 2023. – Вип. 2(77). – С. 18-25. – URL: https://journals.kntu.kherson.ua/index.php/visnyk_kntu/article/view/1019 (дата звернення: 08.02.2026).
7. Abatta-Jacome L., Rosero-Pazmiño D., Rosero-Vivas J., Chávez-Guerrero B. F., Barrionuevo G. O. Statistical evaluation of the mechanical properties of welded and unwelded ASTM A706 reinforcing steel bars of different commercial brands. *Metals*. 2025. Vol. 15, no. 12. Article 1307. DOI: 10.3390/met15121307.
8. Muhammed M., Mustapha M., Ginta T. L., Ali A. M. Statistical review of microstructure-property correlation of stainless steel: implication for pre- and post-weld treatment. *Processes*. 2020. Vol. 8, no. 7. Article 811. DOI: 10.3390/pr8070811.
9. Hou W., Zhang D., Wei Y., Guo J., Zhang X. Review on computer aided weld defect detection from radiography images. *Applied Sciences*. 2020. Vol. 10, no. 5. Article 1878. DOI: 10.3390/app10051878.
10. Sumesh A. та ін. Real-time weld quality prediction using a support vector machine based on optical sensor data. *Journal of Manufacturing Processes*. 2021.
11. Waskom M. L. Seaborn: statistical data visualization. *Journal of Open Source Software*. 2021. Vol. 6, no. 60. Article 3021. DOI: 10.21105/joss.03021.
12. Virtanen P. та ін. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*. 2020. Vol. 17. P. 261–272. DOI: 10.1038/s41592-019-0686-2.
13. Zikopoulos P., Eaton C., deRoos D. та ін. Understanding Big Data. New York : McGraw-Hill, 2020.

Baglai O.P. APPLICATION OF MODERN STATISTICAL METHODS IN WELDING OF METAL STRUCTURES

The article is devoted to the analysis of modern statistical methods and information technologies in quality control, defect prediction, and optimization of welding processes for metal structures. The limitations of traditional empirical approaches in processing large volumes of real-time technological data are demonstrated. The study considers the transition from subjective control to data-driven analysis of welding processes within the framework of Industry 4.0. Particular attention is paid to descriptive and mathematical statistical methods for identifying relationships between input welding parameters (current, voltage) and output weld characteristics (process stability, probability of defect formation).

Based on the Metal Arc Welding – Predictive Quality Data dataset [1], preprocessing and standardization of time series were performed, including data visualization, descriptive statistical analysis, correlation analysis (Pearson correlation coefficient) [2, p. 3], normality testing using the Shapiro-Wilk test [3, p. 92], and comparison of quality classes using the nonparametric Mann-Whitney test.

Visualization revealed that defective welds are characterized by arc instability with deep current drops and voltage peaks, whereas high-quality welds exhibit smoother dynamics. Statistical analysis confirmed significant differences ($p < 0.001$): the mean current in defective welds is higher by 0.492 standard deviations, current variance is 11.7% higher, and the correlation between current and voltage changes from weak negative ($r = -0.143$) to moderately positive ($r = 0.286$). The distributions were found to be non-normal ($p < 0.001$). The obtained statistical characteristics were used as features for classification modeling.

The application of logistic regression as a parametric statistical method achieved an accuracy of 70.1% (ROC-AUC = 0.79), while the nonparametric k-nearest neighbors (k-NN) method demonstrated higher performance with an accuracy of 76.4% and ROC-AUC = 0.85, confirming the informativeness of the statistical features for weld quality prediction. The results may serve as a basis for developing intelligent monitoring systems and automated real-time welding process control, contributing to the digital transformation of industry and enhancing the competitiveness of Ukrainian mechanical engineering.

Keywords: *welding of metal structures, statistical methods, time series analysis, quality control, Industry 4.0, weld quality.*

Дата першого надходження статті до видання: 27.02.2026

Дата прийняття статті до друку після рецензування: 24.03.2026

Дата публікації (оприлюднення) статті 11.05.2026

Bilyi M.O.<https://orcid.org/0009-0004-0639-3658>

National Technical University of Ukraine

"Igor Sikorsky Kyiv Polytechnic Institute"

Krylov Ye.V.<https://orcid.org/0000-0003-4313-938X>

National Technical University of Ukraine

"Igor Sikorsky Kyiv Polytechnic Institute"

MEDIAN CHANGE POINT DETECTION: A SEMANTIC TEXT SEGMENTATION METHOD FOR INFORMATION SYSTEMS WITH VECTOR DATABASES

This article addresses the pressing issue of improving the efficiency of information systems with vector databases by advancing the methods of data preprocessing and semantic segmentation of unstructured text. The operational quality of such systems, particularly those built on the Retrieval-Augmented Generation (RAG) architecture, critically depends on how accurately the input text is divided into logically complete and coherent fragments. Existing heuristic methods (such as fixed-length chunking or paragraph-based splitting) often sever semantic connections and disrupt coreference, while semantic approaches based on encoder models are computationally prohibitive. Alternative statistical methods that utilize the perplexity (PPL) metric of Large Language Models (LLMs) to detect topic boundaries demonstrate high processing speed but suffer from a significant fundamental flaw. They are highly sensitive to local linguistic fluctuations: the occurrence of specific terminology, abbreviations, or rare proper nouns causes sharp spikes in Cross-Entropy Loss, which baseline algorithms falsely identify as semantic block boundaries, leading to excessive and unnatural text fragmentation.

To overcome this challenge, this study proposes and experimentally validates a novel semantic segmentation algorithm named Median Change Point Detection (MCPD). The proposed method radically shifts the mathematical paradigm of boundary detection. Firstly, instead of using the arithmetic mean, it employs the median of token losses within a sentence, which acts as a natural low-pass filter and allows the algorithm to ignore isolated lexical anomalies, identifying the true baseline of syntactic complexity. Secondly, the algorithm abandons unstable local cut-off thresholds in favor of global optimization using the Pruned Exact Linear Time (PELT) algorithm. PELT analyzes the variance of complexity across broad text windows and mathematically pinpoints points of sustained topic change in linear time $O(N)$, effectively ignoring temporary bursts of linguistic "noise". Furthermore, MCPD implements a new segment formation strategy that strictly prioritizes the preservation of semantic integrity over rigid chunk length constraints.

The experimental evaluation of the algorithm was conducted on three diverse datasets: WikiSection, Qasper, and 2WikiMultihopQA, utilizing the Qwen2-1.5B-Instruct model. The research results conclusively demonstrated that the baseline PPL-based method is fundamentally unstable, requiring manual threshold tuning (ranging from 0.9 to 0.05) for each specific domain. In contrast, the developed MCPD method proved to be entirely universal and operated effectively with a single set of parameters across all datasets. On the task of identifying authorial section boundaries (WikiSection), the MCPD algorithm achieved a Boundary F1 score of 23.81%, outperforming the best possible result of the baseline method by 67%. The preservation of logical continuity in the generated fragments also ensured an increase in the quality of final answer generation (QA F1) on complex scientific papers and multi-hop reasoning tasks. The obtained results confirm that the transition to statistically grounded global optimization and the use of robust signals enables the creation of a highly efficient data preparation tool for modern information retrieval systems.

Keywords: Retrieval-Augmented Generation, chunking, perplexity, vector databases, information systems.

Formulation of the problem. Modern information systems, particularly those built on the Retrieval-Augmented Generation (RAG) architecture [10], utilize vector databases as an external knowledge source to overcome issues such as the generation of incorrect information ("hallucinations") [7] and the lack of access to up-to-date or proprietary corporate data.

The fundamental operational process of a RAG system involves vectorizing the user's query, retrieving the most relevant information fragments from a vector database, and subsequently generating a response by a language model based on the retrieved context [10]. Consequently, the accuracy of the information retrieval stage is a critical factor: the quality, completeness, and relevance of the retrieved information directly determine the reliability of the system's final response.

For efficient information retrieval, modern systems rely on dense semantic search methods [8]. Textual data is transformed into dense vectors of real numbers, called embeddings, using specialized neural network models. This approach allows semantically similar texts to be mapped close to each other in a multidimensional feature space, enabling search by meaning rather than merely by exact lexical keyword matches. However, the effectiveness of semantic search is significantly constrained by the architectural and mathematical properties of the embedding models.

Firstly, most modern embedding models impose a strict limit on the length of the input token sequence, typically ranging from 512 to 8192 tokens, due to the quadratic complexity of the attention mechanism [14]. Secondly, there is the fundamental problem of the "information bottleneck" [11]. If an excessively large text fragment containing numerous heterogeneous ideas and facts is fed into the model, the representation of these features is compressed and "averaged" into a single vector. As a result, specific semantic details are lost, critically degrading the vector's ability to be retrieved in response to a narrow, highly specific search query. Given this, the data preprocessing stage – segmenting large documents into smaller fragments prior to vectorization, becomes of paramount importance.

Simple mechanical splitting of text into fixed-length fragments or dividing it exclusively at sentence boundaries often leads to the severance of logical connections. A classic consequence of such an approach is the disruption of coreference: if a subject, such as a person's name, is mentioned in the first segment, and the subsequent segment refers to it only via a pronoun (e.g., "he" or "she"), the seman-

tic content of the second segment becomes isolated and incomplete [6]. During a search query regarding this person's actions, the second fragment will not be retrieved by the system, leading to the loss of critically important context.

Therefore, ensuring the semantic integrity of information during its vectorization is a significant challenge. The development and investigation of novel, semantically grounded methods for adaptive text segmentation is a highly relevant scientific and practical task. Solving this problem will significantly reduce the level of information noise, eliminate the loss of local context, and consequently enhance the overall efficiency of information systems with vector databases by improving the precision and recall metrics of intelligent search.

Analysis of recent research and publications.

The effectiveness of modern information systems with vector databases depends on two key stages: the quality of search algorithms and result aggregation, and the quality of data preprocessing. Although significant research efforts are directed toward improving hybrid search methods and rank fusion, particularly by modifying Reciprocal Rank Fusion (RRF) algorithms to mitigate the impact of irrelevant results [1], the fundamental problem remains the quality of the vector representations themselves. No matter how advanced the search algorithms are, their accuracy will be low if the input text was incorrectly divided into segments during the indexing stage.

Existing approaches to effective text preparation and segmentation for information systems can be broadly classified into three main groups: heuristic (rule-based), semantic, and statistical-model-based.

Heuristic approaches, particularly segmenting text into fixed-length fragments or segmenting with partial overlap, are the most common due to their simplicity and low computational complexity [13]. However, as noted by the authors in [12], such methods are "blind" to context: they frequently sever logical connections, paragraphs, and sentences at unnatural boundaries, leading to the loss of critically important information during vector search. A better form of the heuristic approach is structural segmentation – dividing text by natural paragraphs or Markdown structure. Although this partially avoids breaking logical chains, such a method generates fragments of highly uneven length. Furthermore, long paragraphs often contain several distinct semantic blocks, the information density of which is averaged out when transformed into a single vector.

To overcome these drawbacks, semantic segmentation methods have been proposed. Their essence lies in using encoder models to generate vector rep-

representations of adjacent sentences, followed by calculating their cosine similarity. A drop in similarity below a predefined threshold is interpreted as a topic change. While such approaches are convenient from an engineering perspective and demonstrate a higher quality of baseline context preservation compared to heuristics, they have significant limitations. Firstly, studies [12] indicate their excessive computational cost, as they require the generation of a large number of embeddings during the preprocessing stage. Secondly, as noted in [15], semantic similarity models are often unable to capture subtle logical dependencies between sentences. For instance, a sequence of sentences with a clear logical progression (from general to specific or cause-and-effect) may be incorrectly severed by the algorithm due to the low cosine similarity of their lexical composition, ultimately causing the vector search results to deviate from the core semantic unit.

A distinct modern direction in semantic segmentation is the use of Large Language Models as analytical agents (LLM-as-a-Judge). In such approaches, the text division process is delegated directly to the LLM via specially crafted instructions (prompts), tasking the model with analyzing the text and autonomously determining the logical boundaries of paragraphs based on a deep understanding of the context [4]. This method ensures the highest quality of semantic connection preservation and can account for complex narrative structures within a document. However, its practical application in large-scale RAG systems is severely limited by high costs: processing each document requires full response generation from the language model, making this approach unsuitable for the mass indexing of large knowledge bases during preprocessing.

Given these limitations, an alternative and less resource-intensive direction is the use of internal statistical metrics of Large Language Models. Specifically, recent studies [15] have proposed using Perplexity (PPL) and information entropy to determine chunk boundaries. These metrics are based on logits – raw, unnormalized numerical values that the neural network generates for each word (token) from its vocabulary. The magnitude of a logit reflects the model's degree of confidence that this specific token should be the next in the text. After passing through a Softmax function, logits are converted into probabilities. Based on these probabilities, perplexity is calculated, conceptually reflecting the model's overall level of "surprise": the lower it is, the easier it is for the model to predict the next word. The core hypothesis is that local minima of perplexity – moments

when the model is most confident in predicting the next token – coincide with the boundaries of logically complete semantic blocks or standard linguistic constructions. This approach is computationally efficient, as it allows topic boundaries to be identified "on the fly" during a single pass of the text through the model.

Despite the promise of using perplexity for text segmentation, the existing approach has a significant flaw that limits its effectiveness in real-world information systems. The problem lies in the high sensitivity of the PPL metric to local linguistic fluctuations. A conventional search for PPL minima reacts to any drop in the complexity of token generation. Such a drop is often caused not by a change in the thematic block, but by the appearance of common phrases, predictable syntactic constructions, or highly frequent proper nouns in the text.

To illustrate this, consider the following example: "The local government implemented strict environmental standards for all industrial enterprises in the region. This initiative received widespread support. The level of atmospheric carbon dioxide pollution decreased to a historic minimum." From a semantic perspective, these three sentences form a single, indivisible logical block describing a cause and its direct consequence. However, the language model evaluates their complexity differently. The first and third sentences are saturated with specific terminology, resulting in relatively higher perplexity scores (2.11 and 3.06, respectively). In contrast, the second sentence is a common journalistic cliché with high predictability, causing its perplexity to plummet (to 1.58), forming a local minimum as depicted in Figure 1. The baseline algorithm falsely identifies this drop as a chunk boundary and mistakenly cuts the text, separating the implemented standards from the results of their action.

As a result, such "noise" in the perplexity signal disrupts the integrity of semantic blocks, as the algorithm is unable to distinguish a minor predictable phrase from a genuine structural transition between document topics. Thus, the problem of developing a mathematically grounded criterion that would allow filtering out minor perplexity fluctuations and accurately identifying the true boundaries of semantic blocks – without resorting to resource-intensive semantic comparison methods or costly iterative calls to Large Language Models for direct text analysis – remains unresolved.

Task statement. Considering the identified problem of excessive sensitivity of existing statistical segmentation methods to local linguistic fluctuations, the main objective of this study is to improve the quality

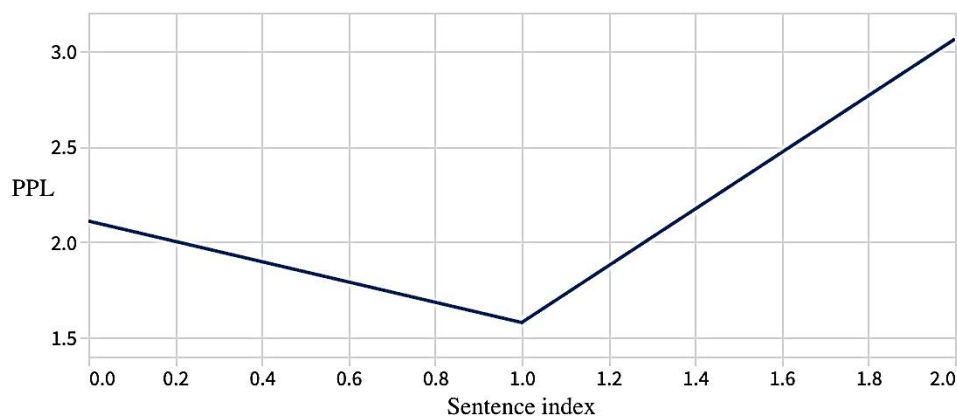


Fig. 1. Local perplexity minimum

of semantic segmentation of unstructured texts for information systems with vector databases by developing a mathematically grounded criterion for filtering "noise" in perplexity metrics.

To achieve this goal, the following scientific and practical tasks are to be solved in this work: to analyze the nature of the occurrence of local perplexity minima during the passage of text through a large language model and to identify patterns that distinguish genuine thematic transitions from minor syntactic or lexical predictabilities; to develop a new metric criterion for identifying the boundaries of semantic blocks, which will reduce the impact of local information noise and increase the accuracy of determining text division points; to build a segmentation algorithm based on the proposed criterion that will ensure the preservation of the semantic integrity of fragments for subsequent vectorization; to conduct an experimental validation of the developed method on benchmark datasets and perform a comparative analysis of its effectiveness against existing segmentation approaches using precision, recall, and F1-score metrics.

Outline of the main material of the study. Within the scope of this study, a novel algorithm for semantic text segmentation for retrieval-augmented generation systems, called Median Change Point Detection (MCPD), was developed and experimentally validated. The proposed method conceptually advances the idea of utilizing the internal "uncertainty" of large language models to detect logical boundaries in text, a concept initially described in the Meta-Chunking framework [15]. This method resolves the fundamental structural flaws of the existing approach by altering the mathematical paradigm of boundary detection and the segment formation strategy.

The theoretical foundation of the MCPD algorithm is based on generating a robust signal of text complex-

ity. Let the input text be pre-divided into a sequence of sentences $S = (s_1, s_2, \dots, s_n)$. After tokenization, each sentence s_i is represented as a sequence of K_i tokens: $s_i = (t_1^i, t_2^i, \dots, t_{K_i}^i)$. For each token t_k^i , the large language model computes the probability of its occurrence given the entire preceding context $t_{<k}^i$ (all tokens preceding it in the current sentence) and $t_{<i}$ (all tokens from previous sentences). In the context of language models, this probability essentially reflects the degree of predictability of the next word: a high probability, close to 1, indicates that the word is a logical and grammatical continuation of the thought, whereas a low probability, close to 0, signifies a topic change, the introduction of a new term, or an unexpected syntactic turn. To convert these probabilities into a convenient additive metric, the Cross-Entropy Loss is calculated, which reflects the degree of the model's "error" in prediction. For a specific token t_{ik} , it is defined as the negative logarithm of this probability:

$$L(t_k^i) = -\log P_M(t_k^i | t_{<k}^i, t_{<i}) \quad (1)$$

where P_M – is the probability distribution generated by model M .

In the baseline Meta-Chunking method, the semantic complexity of a sentence s_i is calculated as the arithmetic mean of the losses of all its tokens. However, this approach has proven to be highly sensitive to statistical outliers. The presence of a single specific term or an unusual proper noun artificially inflates the average score of the entire sentence, generating lexical "noise".

To illustrate this, consider the sentence: "The weather is beautiful today, the sun is shining, and we are walking around the city of Tlaquepaque." Most words in this construction are common and easily predictable for a language model. The model generates them with high probability (e.g., $P(t_k) = 0.9$), accordingly, their loss metric is minimal: $-\log(0.9) \approx 0.046$. However, the last word – "Tlaquepaque"

(the name of a relatively obscure Mexican city) – is semantically unexpected in this context. For this single token, the model will generate an extremely low probability (e.g., $P(t_k) = 0.0001$), leading to a sharp spike in the loss metric: $-\log(0.0001) \approx 4.0$. Consequently, due to one specific token, the entire sentence receives an artificially inflated complexity score.

In contrast, the proposed MCPD method forms a set of losses for all tokens in the sentence $L_i = \{L(t_1), L(t_2), \dots, L(t_K)\}$, and the sentence signal is defined as the median of this set. Mathematically, if the elements of set L_i are sorted in ascending order ($L_{(1)} \leq L_{(2)} \leq \dots \leq L_{(K_i)}$), the median is calculated as:

$$\text{MedianLoss}(s_i) = \begin{cases} L_{((K_i+1)/2)}, & \text{if } K_i \text{ is odd} \\ (L_{(K_i/2)} + L_{(K_i/2+1)}) / 2, & \text{if } K_i \text{ is even} \end{cases} \quad (2)$$

The use of the median acts as a natural low-pass filter: it allows the algorithm to ignore lexical anomalies and analyze exclusively the "background" syntactic and semantic complexity.

Once each sentence has received its robust signal, the task arises to find the optimal cut points that will divide the sequence into logically coherent fragments. Existing approaches use "greedy" algorithms, where a cut is made when the difference between the signals of two adjacent sentences exceeds a rigidly defined threshold, for example, $\Delta\text{PPL} \geq 0.3$. This approach is unstable because it reacts to local noise and fails to "see" the overall picture of the text, leading to false fragmentation and high domain dependency.

To overcome this problem, in the MCPD method, the segmentation task is formalized as a classical change point detection problem in a time series, solved using the Pruned Exact Linear Time (PELT) algorithm [9]. The goal of PELT is to divide the text into segments such that each segment is as homogeneous as possible in terms of complexity, without creating too many small fragments. Mathematically, this reduces to minimizing the objective function:

$$\min_{\tau} m \left(\sum_{i=1}^{m+1} C(y_{(\tau_{i-1}+1):\tau_i}) + \beta f(m) \right) \quad (3)$$

where, m - is the number of cut points dividing the text into $m + 1$ segments; τ_i - is the index of the i -th cut point (with $\tau_0 = 0$, and $\tau_{m+1} = N$, where N - is the total number of sentences); $y_{(\tau_{i-1}+1):\tau_i}$ - is the sequence of sentence perplexity medians within the i -th segment; C - is a cost function measuring the level of internal variance within a single segment; $\beta f(m)$ is a penalty function for creating new cuts, which prevents excessive text fragmentation and false triggering of the algorithm on local information noise.

In this work, the sum of squared deviations from the mean is chosen as the cost function C , and the penalty function is linear ($f(m) = m$), where each new cut increases the total cost by a fixed constant β .

The proposed objective function implements a strict balancing mechanism. The algorithm executes a text cut only if the reduction in internal variance resulting from the division exceeds the fixed cost of the new cut β . Due to this, PELT ignores minor local perplexity fluctuations and reacts exclusively to significant, global shifts in the baseline level of text complexity.

Calculating all possible combinations of cuts would take exponential time. However, the innovativeness of PELT lies in its mathematically proven pruning mechanism. The algorithm sequentially evaluates each sentence and, if it determines that a certain point in the past can mathematically never become part of the optimal set of cuts, it permanently removes it from memory. Thanks to this, PELT guarantees finding the absolute exact global optimum at a speed linearly dependent on the text length – $O(N)$.

The effectiveness of this approach is clearly demonstrated by the dynamics of perplexity changes depending on the text topic, as illustrated in Figure 2. The graph clearly traces the structural blocks of the document: sentences 0 to 35 describe historical events and have their own baseline complexity level; the subsequent block, sentences 35-73, is dedicated to demographics, accompanied by a change in the overall signal variance. This is followed by transitions to the economy block, sentences 73-91, and the government description, sentences 92-113. As seen in the figure, local "spikes" (noise) are present within each thematic block, yet the PELT algorithm successfully ignores them, fixing cuts only at the points of global change in the statistical characteristics of the signal between these macro-blocks.

The physical essence of this process can be illustrated through an analogy with driving a car, where the text acts as the road, and token perplexity acts as the "tension" level of the language model. Local lexical noise (e.g., the appearance of a rare word) resembles a pothole on a smooth highway: the system registers a sharp jump in the PPL metric, but immediately after passing this word, the complexity returns to the initial baseline level. The PELT algorithm, analyzing the variance over a broader window of sentences, recognizes this spike as a temporary anomaly and does not create a cut. Conversely, a genuine topic change (e.g., a transition from a general description to complex terminology) acts like turning onto a winding mountain road: the complexity level rises and remains consistently high throughout the subsequent fragment. In this case, the algorithm detects a sustained shift in the baseline perplexity level (e.g., a change in the mean value from 12.0 to 17.5) and mathematically justifies dividing the semantic blocks at exactly this point.

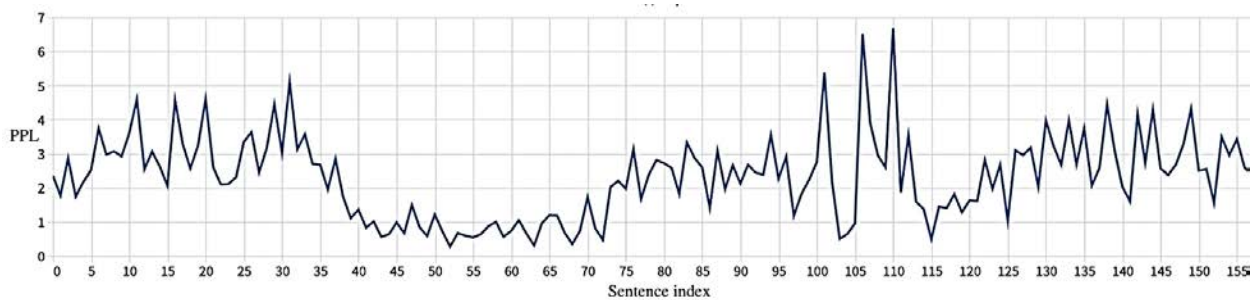


Fig. 2. Dynamics of perplexity changes across structural thematic blocks of the document

An important aspect of the practical application of the PELT algorithm is rethinking the approach to chunk length and the need to adapt the mathematical model to the engineering constraints of vector databases. Since classical PELT optimizes only variance and does not account for the physical size of the blocks, an adaptive buffering strategy was developed in this study. The algorithm accumulates input sentences into a buffer whose size is twice the target maximum segment length, for example, 2×256 tokens. This window size was not chosen randomly: it is the minimum sufficient size for the algorithm to "see" a broader context and compare the variance of the current block with the next potential block, while maintaining low RAM consumption and high computational speed compared to analyzing the entire document at once. Subsequently, PELT analyzes the accumulated array and generates a set of optimal cut points. From this set, the algorithm selects the point that forms the largest semantically coherent segment not exceeding the specified limit.

However, traditional systems tend to artificially and strictly limit segment length, often resorting to forced cutting of long paragraphs or mechanical merging of small fragments if a cut point is not found. In contrast, the MCPD method relies on the logic that semantic continuity of the text is the absolute priority. In cases where PELT mathematically proves that an entire text array within the limit is absolutely homogeneous and contains no logical breaks (e.g., a block of 400 or 500 tokens), a "relaxed limit" strategy is applied. Instead of "blindly" cutting the text in half, the algorithm ignores the strict limit and leaves such a block undivided, executing a cut only at the first detected semantic boundary. It is significantly better to form a larger segment that retains all necessary interconnected information than to artificially tear it apart. Despite the increased text volume, modern vectorization models are capable of effectively capturing the meaning of long but semantically homogeneous passages without losing semantic focus in the generated vector. Similarly, generative models in RAG systems operate much more accurately when they receive the complete authorial context without losing logical connections.

To objectively confirm the theoretical advantages of the method, a series of experiments was conducted on three diverse datasets. The algorithm's ability to accurately determine authorial section boundaries was tested on Wikipedia articles from the WikiSection dataset [2] using the Boundary F1 metric. The impact of segmentation on the quality of answer generation in RAG systems was evaluated using the QA F1 metric on complex scientific articles from the Qasper dataset [3], as well as on multi-hop logical reasoning tasks from the 2WikiMultihopQA dataset [5]. The Qwen2-1.5B-Instruct model was used as the baseline model for both generating logits and formulating final answers.

The F1 Score, which is the harmonic mean of precision and recall, was chosen as the primary criterion for evaluating generation quality. In the context of RAG systems, a high F1 score indicates that the segmentation algorithm forms text fragments that are so high-quality and semantically clean that the language model can find all the necessary information for an answer within them, without being distracted by irrelevant "noise" from adjacent paragraphs.

The obtained results are presented in Table 1. They demonstrate that the use of rigid local thresholds makes the baseline method fundamentally unstable. During testing on the WikiSection dataset, it achieved its maximum result (Boundary F1 = 14.21%) at a high cut threshold of $\theta=0.9$, whereas on the Qasper scientific texts, the threshold had to be reduced 18-fold – to $\theta=0.05$ – to achieve the maximum (QA F1 = 18.82%). Such adjustment makes it impossible to use the baseline method in production systems without prior fine-tuning for each text type. In contrast, the developed MCPD method proved to be universal and demonstrated high effectiveness without additional adaptation, using a single set of baseline settings across all datasets.

The experiment on the WikiSection dataset confirmed the undeniable superiority of global optimization over local optimization: the MCPD algorithm achieved a Boundary F1 score of 23.81%, which outperforms the best possible result of the baseline PPL Chunking method by 67%. The preservation of semantic integrity without artificial fragmentation had a direct positive impact on the quality of final

answer generation. On the Qasper scientific article dataset, the MCPD method provided an F1 Score of 19.29%, bypassing the baseline PPL Chunking method. Similar stability (F1 = 18.12%) was observed on the 2WikiMultihopQA dataset compared to 18.06% for the baseline approach. It is worth noting that the increase in the overall score in the MCPD method is achieved primarily due to an increase in precision, which confirms the algorithm's ability to filter out irrelevant context.

Table 1

Comparison of the effectiveness of the PPL Chunking method and the MCPD method

Datset	Metric	PPL Chunking	MCPD
WikiSection	Boundary F1	14.21% (at $\theta = 0.90$)	23.81%
2WikiMultihopQA	QA F1	18.82% (at $\theta = 0.05$)	19.29%
Qasper	QA F1	18.06% (at $\theta = 0.30$)	18.12%

Conclusions. This study solves the relevant scientific and practical task of improving the quality of semantic segmentation of unstructured texts for information systems with vector databases. Based on the conducted analysis, it was revealed that existing statistical methods relying on the search for local perplexity minima are excessively sensitive to lexical "noise", inevitably leading to false and excessive text fragmentation. To overcome this problem, a novel algorithm, Median Change Point Detection (MCPD), was developed. Its key innovation is the transition from local heuristic thresholds to statistically

grounded global optimization using the PELT algorithm, as well as the use of the median of token losses as a robust signal. This allowed for the effective filtering of minor linguistic fluctuations and the accurate identification of genuine thematic transitions.

Experimental validation on three diverse datasets confirmed the high effectiveness of the proposed approach. The results proved that the MCPD algorithm is universal and operates stably with a single set of settings, eliminating the need to tune thresholds for specific text types, unlike the baseline method. Specifically, on the task of determining authorial section boundaries (WikiSection), the developed method achieved a Boundary F1 score of 23.81%, which outperforms the best possible result of the baseline approach by 67%. Furthermore, preserving the semantic integrity of fragments without artificial cutting or mechanical merging ensured an increase in the quality of final answer generation in RAG systems on complex scientific and multi-hop texts.

Prospects for further research in this direction encompass several key tasks. Firstly, there are plans to develop a mechanism for the dynamic, adaptive calculation of the penalty for the PELT algorithm depending on the overall length and structural density of the input document. Secondly, it is expedient to investigate the impact of the language model's size and architecture – particularly the use of more compact and faster LLMs – on the quality of logit generation and the ultimate segmentation accuracy. Finally, an important step will be expanding the experimental base to evaluate the effectiveness of the MCPD algorithm on multilingual text corpora and integrating the method into real-time data streaming pipelines.

Bibliography:

- Білий М., Крилов Є. Модифікація методу Reciprocal Rank Fusion для поліпшення результатів гібридного пошуку у інформаційних системах з векторними базами даних. Інформаційні технології та суспільство. 2025. № 2 (17). С. 14–19. DOI: <https://doi.org/10.32689/maup.it.2025.2.2>.
- Arnold S., Schneider R., Cudré-Mauroux P., Gers F. A., Löser A. SECTOR: A Neural Model for Coherent Topic Segmentation and Classification. arXiv preprint arXiv:1902.04793. 2019. P. 1–16. DOI: <https://doi.org/10.48550/arXiv.1902.04793>.
- Dasigi P., Lo K., Beltagy I., Cohan A., Smith N. A., Gardner M. A Dataset of Information-Seeking Questions and Answers Anchored in Research Papers. arXiv preprint arXiv:2105.03011. 2021. P. 1–18. DOI: <https://doi.org/10.48550/arXiv.2105.03011>.
- Duarte A. V., Marques J., Graça M., Freire M., Li L., Oliveira A. L. LumberChunker: Long-Form Narrative Document Segmentation. arXiv preprint arXiv:2406.17526. 2024. P. 1–16. DOI: <https://doi.org/10.48550/arXiv.2406.17526>.
- Ho X., Nguyen A. K., Sugawara S., Aizawa A. Constructing A Multi-hop QA Dataset for Comprehensive Evaluation of Reasoning Steps. arXiv preprint arXiv:2011.01060. 2020. P. 1–17. DOI: <https://doi.org/10.48550/arXiv.2011.01060>.
- Jang Y., Hong S., Son J., Park S., Park C., Lim H. From Ambiguity to Accuracy: The Transformative Effect of Coreference Resolution on Retrieval-Augmented Generation systems. arXiv preprint arXiv:2507.07847. 2025. P. 1–15. DOI: <https://doi.org/10.48550/arXiv.2507.07847>.
- Ji Z., Lee N., Frieske R., Yu T., Su D., Xu Y., Ishii E., Bang Y. J., Madotto A., Fung P. Survey of Hallucination in Natural Language Generation. arXiv preprint arXiv:2202.03629. 2022. P. 1–43. DOI: <https://doi.org/10.48550/arXiv.2202.03629>.
- Karpukhin V., Oguz B., Min S., Lewis P., Wu L., Edunov S., Chen D., Yih W. Dense Passage Retrieval for Open-Domain Question Answering. arXiv preprint arXiv:2004.04906. 2020. P. 1–14. DOI: <https://doi.org/10.48550/arXiv.2004.04906>.
- Killick R., Fearnhead P., Eckley I. A. Optimal Detection of Change-points With a Linear Computational

Cost. arXiv preprint arXiv:1101.1438. 2011. P. 1–25. DOI: <https://doi.org/10.48550/arXiv.1101.1438>.

10. Lewis P., Perez E., Piktus A., Petroni F., Karpukhin V., Goyal N., Küttler H., Lewis M., Yih W., Rocktäschel T., Riedel S., Kiela D. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. arXiv preprint arXiv:2005.11401. 2020. P. 1–19. DOI: <https://doi.org/10.48550/arXiv.2005.11401>.

11. Liu N. F., Lin K., Hewitt J., Paranjape A., Bevilacqua M., Petroni F., Liang P. Lost in the Middle: How Language Models Use Long Contexts. arXiv preprint arXiv:2307.03172. 2023. P. 1–20. DOI: <https://doi.org/10.48550/arXiv.2307.03172>.

12. Qu R., Tu R., Bao F. Is Semantic Chunking Worth the Computational Cost? arXiv preprint arXiv:2410.13070. 2024. P. 1–23. DOI: <https://doi.org/10.48550/arXiv.2410.13070>.

13. Shaukat M. A., Adnan M., Kuhn C. C. N. A Systematic Investigation of Document Chunking Strategies and Embedding Sensitivity. arXiv preprint arXiv:2603.06976. 2026. P. 1–25. DOI: <https://doi.org/10.48550/arXiv.2603.06976>.

14. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser Ł., Polosukhin I. Attention Is All You Need. arXiv preprint arXiv:1706.03762. 2017. P. 1–15. DOI: <https://doi.org/10.48550/arXiv.1706.03762>.

15. Zhao J., Ji Z., Feng Y., Qi P., Niu S., Tang B., Xiong F., Li Z. Meta-Chunking: Learning Text Segmentation and Semantic Completion via Logical Perception. arXiv preprint arXiv:2410.12788. 2024. P. 1–18. DOI: <https://doi.org/10.48550/arXiv.2410.12788>.

Білий М. О., Крилов Є. В. MEDIAN CHANGE POINT DETECTION: МЕТОД СЕМАНТИЧНОЇ СЕГМЕНТАЦІЇ ТЕКСТІВ ДЛЯ ІНФОРМАЦІЙНИХ СИСТЕМ З ВЕКТОРНИМИ БАЗАМИ ДАНИХ

У статті розглядається актуальна проблема підвищення ефективності інформаційних систем із векторними базами даних шляхом вдосконалення методів попередньої підготовки та семантичної сегментації неструктурованих текстових даних. Якість функціонування таких систем, зокрема тих, що побудовані на архітектурі генерації, доповненої пошуком (Retrieval-Augmented Generation, RAG), критично залежить від того, наскільки точно вхідний текст поділяється на логічно завершені фрагменти. Існуючі евристичні методи (фіксована довжина, поділ за абзацами) часто розривають семантичні зв'язки та порушують кореферентність, тоді як семантичні підходи на базі моделей-енкодерів є обчислювально надто ресурсоемними. Альтернативні статистичні методи, що використовують показник перплексії (Perplexity, PPL) великих мовних моделей для пошуку меж тем, демонструють високу швидкість, проте мають суттєвий фундаментальний недолік. Вони є вкрай чутливими до локальних мовних флуктуацій: поява специфічних термінів, аббревіатур чи рідкісних власних назв викликає різкі стрибки показника втрат (Cross-Entropy Loss), що хибно ідентифікуються базовими алгоритмами як межі смислових блоків, призводячи до надмірної та неприродної фрагментації тексту.

Для вирішення цієї проблеми у дослідженні розроблено та експериментально обґрунтовано новий алгоритм семантичного сегментування – Median Change Point Detection (MCPD). Запропонований метод кардинально змінює математичну парадигму виявлення меж. По-перше, замість арифметичного середнього використовується медіана втрат токенів речення, що діє як природний фільтр низьких частот і дозволяє ігнорувати точкові лексичні аномалії, виокремлюючи справжній базовий рівень синтаксичної складності. По-друге, алгоритм відмовляється від нестабільних локальних порогів розрізу на користь глобальної оптимізації за допомогою алгоритму Pruned Exact Linear Time (PELT). PELT аналізує дисперсію складності на широких вікнах тексту та математично точно визначає точки стійкої зміни теми за лінійний час $O(N)$, ігноруючи тимчасові сплески "шуму". Крім того, MCPD реалізує нову стратегію формування сегментів, яка надає абсолютний пріоритет збереженню семантичної цілісності тексту над жорсткими обмеженнями довжини чанків.

Експериментальна перевірка алгоритму проводилася на трьох різнопланових наборах даних: WikiSection, Qasper та 2WikiMultiHopQA з використанням моделі Qwen2-1.5B-Instruct. Результати дослідження переконливо довели, що базовий метод на основі PPL є фундаментально нестабільним і потребує ручного налаштування порогів (від 0.9 до 0.05) під кожен окремий домен. Натомість розроблений метод MCPD виявився повністю універсальним і працював з єдиними налаштуваннями на всіх датасетах. На завданні визначення авторських меж розділів (WikiSection) алгоритм MCPD досяг показника Boundary F1 на рівні 23.81%, що на 67% перевершує найкращий можливий результат базового методу. Збереження логічної неперервності фрагментів також забезпечило підвищення якості фінальної генерації відповідей (QA F1) на складних наукових статтях та завданнях із багатокроковим логічним виведенням. Здобуті результати підтверджують, що перехід до статистично обґрунтованої глобальної оптимізації та використання робастних сигналів дозволяє створити високоефективний інструмент підготовки даних для сучасних інформаційно-пошукових систем.

Ключові слова: генерація доповнена пошуком, чанкінг, перплексія, векторні бази даних, інформаційні системи.

Дата першого надходження статті до видання: 07.03.2026

Дата прийняття статті до друку після рецензування: 03.03.2026

Дата публікації (оприлюднення) статті 11.05.2026

УДК 004.855.5:004.032.26:338.5

DOI <https://doi.org/10.32782/2663-5941/2026.2.1/04>**Вакалюк Т.А.**<https://orcid.org/0000-0001-6825-4697>

Державний університет "Житомирська політехніка"

Фант М.О.<https://orcid.org/0000-0002-4994-8009>

Державний університет "Житомирська політехніка"

Дунєв С.С.<https://orcid.org/0009-0006-2041-0885>

Державний університет «Житомирська політехніка»

Власенко О.В.<https://orcid.org/0000-0001-6697-2150>

Державний університет "Житомирська політехніка"

Коломієць В.В.<https://orcid.org/0009-0002-8106-5502>

Державний університет "Житомирська політехніка"

ОГЛЯД МЕТОДІВ ПРОГНОЗУВАННЯ ДИНАМІКИ ПОПИТУ ТОВАРІВ

У статті проведено комплексний огляд сучасних методів і підходів до прогнозування динаміки попиту товарів, що застосовуються в електронній комерції та роздрібній торгівлі. Розглянуто три основні категорії методів, зокрема статистичні моделі, методи машинного навчання та аналітика великих даних. У межах статистичних моделей проаналізовано методи лінійної та множинної регресії, експоненціального згладжування, а також моделі часових рядів ARIMA і SARIMA. Показано, що ці підходи ефективні для прогнозування в умовах стабільного попиту, але мають обмеження при роботі з нелінійними залежностями та великими масивами різнорідних даних.

Для подолання цих обмежень розглянуто методи машинного навчання, зокрема логістичну регресію, дерева рішень, ансамблеві методи (Random Forest і Gradient Boosting), а також рекурентні нейронні мережі типу LSTM (Long Short-Term Memory). Проведено порівняльний аналіз зазначених підходів за ключовими критеріями, серед яких точність прогнозування, масштабованість, здатність моделювати нелінійні залежності, вимоги до обсягу тренувальних даних та обчислювальна складність. Доведено, що нейронні мережі LSTM мають суттєві переваги для прогнозування часових рядів завдяки здатності фіксувати довгострокові залежності, моделювати нелінійні взаємозв'язки та адаптуватися до нових ринкових умов через перенавчання.

Окремо висвітлено роль аналітики великих даних у сучасних системах прогнозування. Розглянуто можливості платформ Hadoop і Apache Spark для розподіленої обробки структурованих і неструктурованих даних з різнорідних джерел у режимі реального часу. Обґрунтовано доцільність використання багатошарової архітектури LSTM з dropout-регуляризациєю та оптимізатором ADAM для побудови аналітично-інформаційної системи прогнозування попиту. Запропонована система передбачає опрацювання вхідних даних про попередні обсяги попиту, сезонні фактори та додаткові параметри, їх обробку через приховані шари LSTM та формування прогнозованих значень на вихідному шарі.

Ключові слова: прогнозування попиту, машинне навчання, нейронні мережі, LSTM, часові ряди, ARIMA, SARIMA, аналітика великих даних, Random Forest, Gradient Boosting.

Постановка проблеми. У сучасному конкурентному середовищі точне прогнозування попиту на продукцію є вирішальним фактором для успіху бізнесу. В умовах зростаючих обсягів даних, доступних у реальному часі, компанії, що працюють в електронній комерції, як і в інших галузях, усе частіше застосовують прогресивні аналітичні підходи для побудови ефективних систем прогнозування попиту [1, 10]. Використання таких систем дозволяє бізнесам мінімізувати ризики, пов'язані з надлишком або нестачею товарів, оптимізувати ланцюги постачання, знижувати витрати та підвищувати рівень задоволеності клієнтів.

На сучасному етапі існує безліч методів і програмних рішень, які дозволяють прогнозувати попит на основі аналізу великих даних, обробки часових рядів і машинного навчання. Поширені інструменти прогнозування, які використовують компанії, варіюються від простих статистичних моделей до складних моделей на основі штучного інтелекту, здатних враховувати численні зовнішні фактори, такі як сезонні зміни, погодні умови, поведінка споживачів і навіть дані соціальних мереж тощо.

Інтернет-магазини, такі як Amazon, Walmart, Alibaba та інші, вже використовують передові методи прогнозування. Вони не лише оптимізують запаси, але й забезпечують більш персоналізований досвід для покупців. Алгоритми, що лежать в основі їхніх систем, часто поєднують різні підходи, зокрема, аналіз часових рядів, регресійні моделі, нейронні мережі, кластеризацію та асоціативний аналіз, що дозволяє враховувати як довгострокові тенденції, так і короткострокові зміни в попиті [5].

Прогнозування попиту є складним процесом, що потребує обробки великих обсягів даних, оскільки він враховує не лише історичні показники, але й безліч зовнішніх факторів, які можуть впливати на динаміку продажів. З цієї причини найбільш ефективні рішення у цій сфері зазвичай базуються на поєднанні кількох методів і алгоритмів [9]. Наприклад, аналітичні системи можуть використовувати методи регресії для визначення основних трендів, алгоритми машинного навчання для виявлення прихованих патернів, а також нейронні мережі для аналізу та передбачення сезонних коливань. Окрім цього, все більшого поширення набувають підходи, що базуються на аналітиці великих даних, яка обробляє і аналізує великий масив інформації, включаючи взаємодію клієнтів із сайтом та зовнішні економічні показники.

Аналіз останніх досліджень і публікацій. Проблема прогнозування попиту досліджується у численних наукових роботах, присвячених статистичним методам, машинному навчанню та великим даним. Класичні підходи до побудови моделей прогнозування розглянуто у монографії McKinney [1], де представлено практичні інструменти аналізу даних засобами Python. Фундаментальні засади статистичного навчання та регресійного аналізу ґрунтовно описано у роботі Hastie, Tibshirani і Friedman [2], яка залишається одним із найавторитетніших посібників у галузі статистичного моделювання. Класичний підхід до аналізу часових рядів, зокрема моделі ARIMA та SARIMA, детально викладено у монографії Box, Jenkins, Reinsel і Ljung [3], яка стала стандартом для дослідників у цій сфері. Питання побудови моделей часових рядів для задач прогнозування розглянуто також у роботі Brockwell і Davis [4], де наведено теоретичне обґрунтування та практичні приклади застосування відповідних методів.

Методи розпізнавання образів та машинного навчання, включаючи дерева рішень, нейронні мережі та методи ансамблю, систематизовано у фундаментальній праці Bishop [5]. Теоретичні основи глибинного навчання, у тому числі архітектура рекурентних нейронних мереж та LSTM, детально описано у монографії Goodfellow, Bengio і Courville [6]. Практичні аспекти застосування LSTM для задач прогнозування часових рядів викладено у роботі Brownlee [11], де наведено приклади реалізації відповідних архітектур та оцінку їх точності. Принципи прогнозування та їх практичне застосування в бізнесі розглянуто у роботі Hyndman і Athanasopoulos [8], яка охоплює як класичні статистичні методи, так і сучасні підходи на основі машинного навчання.

Питання бізнес-аналітики та управління ланцюгами постачання на основі великих даних досліджено у статті Chen, Chiang і Storey [7], де висвітлено трансформацію від традиційного аналізу до аналітики на основі великих масивів даних. Стратегічні аспекти управління ланцюгами постачання та методи прогнозування в контексті логістики розглянуто у роботі Chopra і Meindl [10]. Технологічна база для обробки великих даних описана у монографіях White [12] та Karau, Konwinski, Zaharia і Wendell [13], де висвітлено платформи Hadoop і Apache Spark відповідно. Таким чином, незважаючи на значну кількість публікацій у цій галузі, комплексний порівняльний аналіз методів прогнозування попиту з обґрунтуванням переваг архітектури LSTM для систем електронної комерції залишається актуальним завданням.

Постановка завдання. Метою статті є огляд сучасних систем прогнозування попиту, які використовуються на ринку, і методів, що лежать в основі цих систем.

Виклад основного матеріалу. Розглянемо найпоширеніші підходи до прогнозування, а також приклади їх застосування в інтернет-магазинах і великих роздрібних мережах.

Статистичні моделі є основою багатьох аналітичних систем, призначених для прогнозування попиту. Ці моделі ґрунтуються на припущенні, що історичні тенденції та закономірності можуть бути використані для передбачення майбутнього. Використовуючи різні математичні підходи, статистичні моделі дозволяють аналізувати взаємозв'язки між змінними, виявляти тренди, сезонність і циклічність у даних. Найчастіше застосовуються для довгострокового прогнозування в умовах стабільного попиту та коли потрібно отримати швидкі оцінки для планування [1].

Один з найпоширеніших методів у статистичних моделях – *лінійна регресія*, яка дозволяє визначити лінійну залежність між змінною попиту Y та іншими факторами (наприклад, ціна, маркетингові витрати). Модель описується рівнянням:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$$

де:

- Y – прогнозована величина попиту,
- $X_1, X_2, \dots, X_n$ – незалежні змінні (фактори, що впливають на попит),
- $\beta_0, \beta_1, \dots, \beta_n$ – коефіцієнти регресії, що показують вплив кожної змінної на попит,
- ϵ – випадкова похибка, яка враховує непередбачувані фактори.

Лінійна регресія є ефективною, коли між змінними існує чітка лінійна залежність. Це дозволяє легко інтерпретувати вплив кожного фактора на попит, але модель обмежена, якщо дані мають нелінійну природу [2].

Множинна регресія – це розширення лінійної регресії, яке дозволяє враховувати кілька змінних одночасно. Цей метод корисний для складніших прогнозів, де попит залежить від багатьох факторів. Наприклад, для прогнозування продажів комп'ютерної техніки можна використовувати множинну регресію, враховуючи сезонність, економічні умови, рівень конкуренції тощо.

Метод *експоненціального згладжування* є більш гнучким і застосовується для даних, які мають сильні тенденції або сезонні коливання. Він передбачає зниження ваги для старих значень у часі, що дозволяє краще враховувати останні

зміни у попиті. Основні формули експоненціального згладжування:

- *просте експоненціальне згладжування* для випадків без тренду і сезонності:

$$S_t = \alpha Y_t + (1 - \alpha) S_{t-1}$$

де: S_t – згладжене значення на момент t , Y_t – фактичне значення попиту, α – параметр згладжування (від 0 до 1).

- *експоненціальне згладжування з трендом (Holt's Method)* для випадків, коли дані мають тренд:

$$S_t = \alpha Y_t + (1 - \alpha)(S_{t-1} + T_{t-1})$$

$$T_t = \beta(S_t - S_{t-1}) + (1 - \beta)T_{t-1}$$

де: T_t – компонент тренду, β – параметр згладжування для тренду.

Методи експоненціального згладжування є ефективними, коли потрібно швидко реагувати на останні зміни у даних. Вони добре працюють із сезонними даними та дозволяють швидко адаптувати прогноз до нових умов [3].

Варто також розглянути моделі часових рядів, які включають моделі ARIMA та SARIMA.

Модель *ARIMA* (Auto-Regressive Integrated Moving Average) є однією з найпопулярніших моделей для аналізу часових рядів. ARIMA враховує автокореляцію між спостереженнями, що робить її ефективною для довгострокового прогнозування.

Модель складається з трьох компонентів: AR (Auto-Regressive), що враховує автокореляцію, коли поточне значення залежить від попередніх; I (Integrated), яка використовується для усунення тренду; та MA (Moving Average), що враховує випадкові зміни й згладжує дані.

ARIMA ефективно працює для даних із сезонними коливаннями та трендами, де є складні залежності між попередніми значеннями та поточним попитом [4].

Для даних із вираженою сезонністю застосовують модель SARIMA (Seasonal ARIMA), яка додає до ARIMA додаткові сезонні компоненти. SARIMA ефективна для складних часових рядів із сезонними коливаннями, що може бути корисним для прогнозування попиту, наприклад, на техніку у святкові сезони [5].

Статистичні моделі є основою для аналізу та прогнозування попиту завдяки своїй інтерпретованості та здатності обробляти сезонні та трендові компоненти. Методи, такі як лінійна та множинна регресія, експоненціальне згладжування, ARIMA та SARIMA, забезпечують широкий інструментарій для точного прогнозування залежно від особливостей даних.

Вони дозволяють відстежувати тенденції, сезонні коливання та вплив різних факторів на попит, що робить їх ідеальними для бізнесів, які прагнуть стабільності у постачанні та задоволенні попиту.

Для прогнозування попиту також застосовуються методи машинного навчання, які забезпечують точність і ефективність аналізу. Вони допомагають моделювати складні залежності та виявляти закономірності в даних.

Машинне навчання (ML) є сучасним підходом до прогнозування попиту, який здатен обробляти великі обсяги даних і виявляти приховані патерни у складних багатовимірних наборах даних. Алгоритми машинного навчання самостійно навчаються на історичних даних і можуть адаптуватися до нових трендів у міру надходження нових даних, що робить їх особливо ефективними для задач прогнозування в умовах швидкозмінного ринку. Методи ML широко використовуються в електронній комерції та роздрібній торгівлі, де вони дозволяють підвищити точність прогнозів і знизити витрати на управління запасами [1].

Для задач прогнозування динаміки попиту лінійні моделі, такі як *лінійна регресія та регресія Лассо*, можуть використовуватися для оцінки та аналізу змінних, що впливають на попит. Логістична регресія використовується для передбачення ймовірності настання певних подій, наприклад, високого чи низького попиту, тоді як регресія Лассо допомагає враховувати великий набір змінних та обирати лише найважливіші з них.

Дерева рішень є простими і потужними моделями для прогнозування, які ієрархічно розбивають дані на основі певних умов. Це дозволяє класифікувати або прогнозувати дані на основі багатьох факторів. Для задач прогнозування попиту дерева рішень є цінними через можливість враховувати взаємодії між різними змінними.

Формула для дерева рішень будується шляхом мінімізації похибки, де кожне розгалуження збільшує точність прогнозу. Проте, окремі дерева рішень можуть бути чутливими до викидів, тому популярними є *методи ансамблю*, такі як: *Random Forest та Gradient Boosting*.

Метод ансамблю *Random Forest* створює багато дерев рішень та об'єднує їх прогнози для підвищення точності. Метод ансамблю *Gradient Boosting* поетапно покращує прогнозування, додаючи нові дерева, які мінімізують похибку попередніх прогнозів. Методи ансамблю підходять для складних наборів даних і забезпечують точні прогнози, оскільки знижують ризик переобучення за рахунок усереднення [4].

Нейронні мережі є одним із найбільш гнучких і потужних методів для прогнозування попиту, особливо при роботі з великими наборами даних і складними залежностями між змінними. Модель нейронної мережі складається з багатьох шарів (вхідний, приховані, вихідний), де кожен шар містить нейрони, які з'єднані між собою. Мережа навчається на історичних даних, адаптуючи свої параметри для мінімізації похибки між прогнозованим і фактичним попитом.

Рекурентні нейронні мережі (RNN) і їхня вдосконалена версія – *LSTM (Long Short-Term Memory)* – спеціально розроблені для роботи з послідовними даними, як-от часові ряди. У прогнозуванні попиту LSTM є ефективними завдяки здатності запам'ятовувати важливу інформацію на довгі проміжки часу та ігнорувати менш важливі дані. Архітектура LSTM містить осередки пам'яті, які контролюються за допомогою "ворот" (forget, input і output). LSTM є ідеальним вибором для задач прогнозування попиту, оскільки вони здатні враховувати як короткострокові, так і довгострокові залежності.

Кластеризація допомагає сегментувати дані про попит на групи, що мають подібні характеристики, що дозволяє створювати прогнози для кожного кластера окремо. Наприклад, K-Means кластеризація розподіляє точки даних на k груп, зменшуючи внутрішню варіативність у кожному кластері. Це корисно для рітейлерів, які можуть використовувати прогнози попиту для різних клієнтських сегментів або товарних категорій.

Методи машинного навчання характеризуються високою точністю при роботі з великими обсягами даних, здатністю обробляти складні та нелінійні взаємозв'язки, а також адаптивністю до змінних ринкових умов. Водночас їх використання має певні обмеження, зокрема потребу у великих обсягах даних для тренування моделей, високі обчислювальні витрати та труднощі в інтерпретації складних моделей, таких як нейронні мережі.

Аналітика великих даних стала важливою складовою сучасних систем прогнозування, особливо у великих компаніях та інтернет-магазинах, таких як Amazon, Walmart, Alibaba та інші. Завдяки технологіям Big Data аналітики можуть обробляти та аналізувати надзвичайно великі обсяги інформації, отримуючи важливі інсайти про поведінку клієнтів, ринкові тенденції, сезонні коливання та інші фактори, які впливають на попит. На відміну від традиційних методів аналізу даних, аналітика великих даних дозволяє швидко обробляти і зберігати інформацію, включаючи неструктуровані

дані (соціальні медіа, відгуки клієнтів, погодні умови тощо) [1].

Великі дані характеризуються трьома основними аспектами. Обсяг (Volume) стосується величезних масивів даних, які перевищують можливості традиційних систем обробки. Швидкість (Velocity) відображає високі темпи генерації та оновлення даних, що вимагає майже миттєвої обробки. Різноманітність (Variety) включає різні джерела та формати даних, зокрема структуровані, неструктуровані та напівструктуровані.

Для прогнозування попиту застосовують алгоритми машинного навчання, що дозволяють виявляти закономірності у великих масивах даних, моделювати взаємозв'язки між змінними та сегментувати клієнтів за типами попиту. Аналіз часових рядів на основі технологій Big Data забезпечує виявлення трендів, сезонності та циклічності, а обробка даних у реальному часі дає змогу оперативного адаптувати прогнози до ринкових змін. Додатково враховується неструктурована інформація – відгуки клієнтів, дані соціальних мереж, погодні звіти, що розширює можливості моделювання поведінки споживачів.

Серед інструментів аналітики великих даних вирізняється Hadoop – відкрита платформа розподіленого зберігання та обробки інформації. Модель MapReduce забезпечує паралельні обчислення на кластерах серверів, а HDFS гарантує надійне зберігання даних через реплікацію. Горизонтальна масштабованість платформи дозволяє нарощувати обчислювальні потужності без істотних витрат на продуктивність [12].

Серед популярних інструментів варто відзначити Apache Spark – платформа для розподіленої обробки даних, яка значно прискорює операції завдяки виконанню обчислень в пам'яті. Spark підтримує широкий спектр завдань, таких як обробка поточкових даних, аналітика великих обсягів інформації, а також машинне навчання через бібліотеку MLlib. Платформа інтегрується з різними джерелами даних, включаючи HDFS, Apache Cassandra та Amazon S3, що робить її універсальним інструментом для роботи з різними форматами та типами даних. Вона широко використовується в компаніях, які потребують обробки даних у реальному часі, наприклад, для аналізу логів або обробки фінансових транзакцій [13].

MongoDB і Cassandra є двома основними представниками NoSQL баз даних, що спеціалізуються на обробці неструктурованих та напівструктурованих даних. MongoDB відома своєю гнучкістю завдяки зберіганню даних у форматі

BSON (Binary JSON), що дозволяє зберігати складні структури і легко інтегрувати їх у сучасні веб-додатки. Cassandra, в свою чергу, відрізняється високою доступністю та масштабованістю, що робить її ідеальним вибором для розподілених систем, що потребують роботи з великими обсягами даних в різних локаціях. Її архітектура без єдиного вузла відмови забезпечує надійність і безперервну роботу [12].

Для обробки поточкових даних популярною платформою є Kafka. Вона дозволяє здійснювати швидку передачу інформації між різними додатками та обробляти велику кількість подій у реальному часі. Kafka часто використовується для обробки даних з IoT-пристроїв або аналітики користувацької активності на вебсайтах. Крім того, Kafka інтегрується з іншими платформами, такими як Hadoop і Spark, утворюючи потужну екосистему для обробки та аналізу великих даних.

Аналітика великих даних відкриває нові можливості для прогнозування попиту, оскільки дозволяє враховувати як структуровані, так і неструктуровані дані. Вона забезпечує високу швидкість обробки інформації, що є надзвичайно важливим для ритейлерів, які повинні швидко реагувати на змінні умови ринку. Завдяки можливості обробляти різноманітні джерела даних, Big Data дозволяє створювати більш точні моделі прогнозування, що базуються на актуальних даних про ринок і споживчі уподобання.

Для прогнозування попиту, що враховує сезонність, тренди та раптові коливання, архітектура LSTM має низку ключових переваг. По-перше, модель ефективно обробляє довгострокові залежності в часових рядах, зберігаючи інформацію про попередні значення на тривалих інтервалах. По-друге, LSTM здатна моделювати нелінійні взаємозв'язки між змінними, що особливо актуально для електронної комерції, де попит залежить від маркетингових кампаній, зовнішніх подій і нових трендів. По-третє, модель легко інтегрує різноманітні джерела даних – від історії продажів до погодних умов і регіональних особливостей. Нарешті, можливість перенавчання на нових даних забезпечує адаптивність системи до змін ринкового середовища.

Нейронні мережі LSTM мають певні переваги в порівнянні з іншими методами прогнозування попиту. Статистичні моделі, такі як ARIMA і SARIMA, хоча і ефективні для прогнозування часових рядів, мають обмеження у роботі з довгостроковими нелінійними залежностями. Вони добре підходять для моделювання лінійних трен-

дів, але не здатні забезпечити точність прогнозів при складних або змінних взаємозв'язках. На відміну від них, LSTM краще справляється з складними і довгостроковими трендами, забезпечуючи більш точні прогнози для таких даних.

Методи експоненціального згладжування часто використовуються для короткострокових прогнозів, однак вони мають свої обмеження при роботі з великими обсягами даних. Вони можуть бути ефективними для спрощених задач, але не здатні враховувати складні фактори, які впливають на попит, такі як сезонність чи маркетингові кампанії. Крім того, ці методи не завжди дають змогу коректно прогнозувати зміни в попиті в динамічних умовах.

Інструменти аналітики великих даних, або Big Data, здатні збирати і обробляти великі обсяги інформації, але для точного прогнозування конкретних змін у попиті, особливо на короткий термін, необхідні спеціалізовані моделі, такі як LSTM. Ці моделі здатні враховувати складні взаємозв'язки між змінними і адаптуватися до нових даних, що робить їх більш ефективними для точного прогнозування.

Для AI системи пропонується використати архітектуру *багатошарової LSTM* із декількома прихованими шарами, що дозволить обробляти як сезонні, так і короткострокові коливання попиту. Додатково, структура включатиме параметри для регуляризації (dropout) і оптимізацію (наприклад, ADAM), що дозволить уникнути перенавчання та знизити похибку прогнозування.

Вхідний шар отримує дані про попередні періоди попиту, сезонні фактори та додаткові параметри, такі як місяць, пора року або маркетингові активності. Ці дані використовуються для подальшої обробки в моделі. Приховані шари LSTM відповідають за обробку інформації про довгострокові залежності, що дозволяє забезпечити високу точність прогнозування навіть у випадку сезонних змін попиту. Останній етап моделювання – вихідний шар, який на основі поточного стану передає прогнозовані значення попиту для подальшого аналізу.

Враховуючи складність і багатофакторність процесу прогнозування попиту, метод LSTM надає найточніші і стабільні результати. Його здатність працювати з часовими рядами, зберігати довгострокову пам'ять і адаптуватися до нових умов

робить його ідеальним вибором для нашої аналітично-інформаційної системи. Крім того, LSTM дозволяє враховувати складні залежності між змінними, що підвищує точність прогнозування, особливо в умовах мінливої ринкової ситуації.

Отже, для розробки аналітично-інформаційної системи прогнозування попиту методом вибору є модель нейронних мереж (LSTM). Вона не лише відповідає потребам прогнозування на основі часових рядів, але й забезпечує гнучкість і адаптивність, необхідні для роботи в умовах динамічного ринку.

Висновки. Статистичні моделі (регресійний аналіз, експоненціальне згладжування, ARIMA та SARIMA) ефективні в умовах стабільної ринкової кон'юнктури та вирізняються інтерпретованістю результатів, проте мають обмежені можливості при роботі з нелінійними залежностями та великими неструктурованими масивами даних. Методи машинного навчання, зокрема ансамблеві алгоритми Random Forest і Gradient Boosting, забезпечують вищу точність прогнозування завдяки здатності моделювати складні взаємозв'язки, хоча й потребують значних обчислювальних ресурсів і великих тренувальних вибірок. Платформи аналітики великих даних – Hadoop, Apache Spark, Kafka – розширюють можливості опрацювання різномірних джерел у реальному часі, однак їх впровадження вимагає суттєвих інфраструктурних витрат.

На основі порівняльного аналізу обґрунтовано вибір архітектури LSTM, яка перевершує статистичні моделі у роботі з довгостроковими залежностями та нелінійними паттернами, а в умовах часових рядів демонструє кращу адаптивність порівняно з ансамблевими методами. Багатошарова LSTM із dropout-регуляризацією та оптимізатором ADAM слугує надійною основою для подальшого практичного впровадження системи прогнозування попиту.

Перспективами подальших досліджень є практична реалізація та тестування запропонованої системи на реальних наборах даних електронної комерції, а також дослідження можливостей гібридних підходів, що поєднують LSTM із статистичними методами або ансамблевими алгоритмами для підвищення точності прогнозування в умовах різних сценаріїв попиту.

Список літератури:

1. McKinney, W. Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter. 3rd ed. O'Reilly Media, 2022. 579 p.
2. Hastie, T., Tibshirani, R., & Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. Springer, 2009. 745 p. DOI: 10.1007/978-0-387-84858-7.

3. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. Time Series Analysis: Forecasting and Control. 5th ed. Wiley, 2015. 712 p. DOI: 10.1111/jtsa.12194
4. Brockwell, P. J., & Davis, R. A. Introduction to Time Series and Forecasting. 3rd ed. Springer, 2016. 425 p. ISBN 978-3-319-29852-8. DOI: 10.1007/978-3-319-29854-2.
5. Bishop, C. M. Pattern Recognition and Machine Learning. Springer, 2006. 738 p. ISBN 978-0-387-31073-2.
6. Goodfellow, I., Bengio, Y., & Courville, A. Deep Learning. MIT Press, 2016. 775 p. ISBN 978-0-262-03561-3. URL: <https://www.deeplearningbook.org>.
7. Chen, H., Chiang, R. H. L., & Storey, V. C. Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly. 2012. Vol. 36, No. 4. P. 1165–1188. DOI: 10.2307/41703503.
8. Hyndman, R. J., & Athanasopoulos, G. Forecasting: Principles and Practice. 3rd ed. OTexts, 2021. 442 p. URL: <https://otexts.com/fpp3>.
9. Hochreiter, S., & Schmidhuber, J. Long Short-Term Memory. Neural Computation. 1997. Vol. 9, No. 8. P. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.
10. Chopra, S. Supply Chain Management: Strategy, Planning, and Operation. 7th ed. Pearson, 2018. 528 p.
11. Brownlee, J. Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs and LSTMs in Python. Machine Learning Mastery, 2018. 596 p. URL: <https://machinelearningmastery.com>.
12. White, T. Hadoop: The Definitive Guide: Storage and Analysis at Internet Scale. 4th ed. O'Reilly Media, 2015. 756 p.
13. Karau, H., Konwinski, A., Zaharia, M., & Wendell, P. Learning Spark: Lightning-Fast Data Analysis. O'Reilly Media, 2015. 276 p.

Vakaliuk T. A., Fant M. O., Duniev S. S., Vlasenko O. V., Kolomiets V. V. OVERVIEW OF METHODS FOR FORECASTING THE DYNAMICS OF DEMAND FOR GOODS

The article provides a comprehensive review of modern methods and approaches to forecasting the dynamics of demand for goods used in e-commerce and retail trade. Three main categories of methods are considered, namely statistical models, machine learning methods, and big data analytics. Within the framework of statistical models, the methods of linear and multiple regression, exponential smoothing, as well as ARIMA and SARIMA time series models are analysed. It is shown that these approaches are effective for forecasting in conditions of stable demand, but have limitations when working with non-linear dependencies and large arrays of heterogeneous data.

To overcome these limitations, machine learning methods are considered, in particular logistic regression, decision trees, ensemble methods (Random Forest and Gradient Boosting), as well as recurrent neural networks of the LSTM (Long Short-Term Memory) type. A comparative analysis of these approaches is carried out according to key criteria, including forecasting accuracy, scalability, ability to model nonlinear dependencies, training data volume requirements, and computational complexity. It has been proven that LSTM neural networks have significant advantages for time series forecasting due to their ability to capture long-term dependencies, model nonlinear relationships, and adapt to new market conditions through retraining.

The role of big data analytics in modern forecasting systems is highlighted separately. The capabilities of the Hadoop and Apache Spark platforms for distributed processing of structured and unstructured data from heterogeneous sources in real time are considered. The feasibility of using a multi-layer LSTM architecture with dropout regularisation and an ADAM optimiser to build an analytical and information system for demand forecasting is justified. The proposed system involves processing input data on previous demand volumes, seasonal factors and additional parameters, processing them through hidden LSTM layers and forming predicted values on the output layer.

Keywords: demand forecasting, machine learning, neural networks, LSTM, time series, ARIMA, SARIMA, big data analytics, Random Forest, Gradient Boosting.

Дата першого надходження статті до видання: 05.02.2026

Дата прийняття статті до друку після рецензування: 06.03.2026

Дата публікації (оприлюднення) статті 11.05.2026

Волк М.О.

<https://orcid.org/0000-0003-4229-9904>

Харківський національний університет радіоелектроніки

Ковтун Є.І.

<https://orcid.org/0009-0006-4775-1292>

Харківський національний університет радіоелектроніки

МОДЕЛІ ДИНАМІЧНОГО РОЗПОДІЛЕННЯ ОБЧИСЛЮВАЛЬНИХ ЗАДАЧ У ГЕТЕРОГЕННИХ КОМП'ЮТЕРНИХ СИСТЕМАХ

Стаття присвячена проблемі динамічного розподілення обчислювальних задач у гетерогенних комп'ютерних системах. Такі системи мають ресурси з відмінними характеристиками продуктивності та енергоспоживання. Актуальність цього дослідження зумовлена зростанням навантажень у хмарних і розподілених середовищах, а також необхідністю забезпечувати баланс між енергоефективністю та дотриманням вимог до якості обслуговування (QoS). Існуючі підходи до планування часто орієнтуються на окремі критерії, що знижує їх ефективність в умовах високої динаміки. Тому перспективними є дослідження, пов'язані з врахуванням значної кількості характеристик таких систем. В роботі запропоновано математичну модель багатокритеріальної оптимізації, у якій функція мети формулюється як зважена сума трьох складових: сумарного енергоспоживання обчислювальних вузлів, штрафу за перевантаження ресурсів та вартості міграції віртуальних машин. Енергоспоживання описується лінійною залежністю. Для мінімізації ризику деградації продуктивності застосовано квадратичну штрафну функцію за перевищення допустимого порогу завантаження. Це сприяє рівномірному розподілу навантаження. Додатково враховано параметр вартості міграції, що дозволяє обмежувати частоту перенесення задач і підвищувати стабільність системи. Експериментальна перевірка показала зниження середнього енергоспоживання до 18% порівняно з базовими моделями та суттєве скорочення часу перебування вузлів у перевантаженому стані. Найкращі інтегральні показники досягнуто у збалансованому режимі налаштування вагових коефіцієнтів. Запропоновані моделі демонструють адаптивність до змін навантаження та придатність для застосування у хмарних і високопродуктивних обчислювальних системах. Перспективи подальших досліджень пов'язані з розширенням моделі на багаторесурсні середовища та інтеграцією гібридних метаевристичних алгоритмів глобальної оптимізації.

Ключові слова: гетерогенні комп'ютерні системи, динамічне розподілення задач, планування ресурсів, енергоефективність, QoS, міграція віртуальних машин, багатокритеріальна оптимізація, функція мети, хмарні обчислення.

Постановка проблеми. Швидке зростання обчислювальних навантажень разом із різноманітністю апаратного забезпечення істотно ускладнює задачі планування та розподілення ресурсів у сучасних гетерогенних комп'ютерних системах. У практичних конфігураціях такі системи, як правило, поєднують високопродуктивні сервери, звичайні CPU-вузли, графічні прискорювачі, а також спеціалізовані обчислювальні модулі, які функціонують у спільному інформаційному середовищі. Водночас, попри наявність значної кількості підходів до балансування навантаження, забезпечення стабільної продуктивності й досі залишається проблемним, особливо за динамічних змін

вхідних потоків, енергетичних обмежень та змінної доступності обчислювальних ресурсів[1].

Надійність та ефективність функціонування гетерогенних обчислювальних платформ значною мірою визначаються здатністю системи оперативно адаптуватися до змін характеристик навантаження та доступності ресурсів. Традиційні підходи до розподілу обчислювальних задач, як правило, ґрунтуються на наперед заданих правилах або евристичних алгоритмах, що обмежує їхню гнучкість в умовах високої динаміки та варіативності робочих сценаріїв.

Але значна частина існуючих оптимізаційних моделей динамічного планування не враховує

повністю складність взаємодії між компонентами, що мають різномірний характер. Особливо це виникає за наявності вимог з декількома критеріями, які пов'язані із якістю обслуговування, часовими обмеженнями та необхідністю мінімізації енергоспоживання [2].

Тому вважаємо актуальним обговорення математичних моделей і методів, які здатні забезпечувати збалансоване керування обчислювальними задачами та ресурсами у гетерогенних середовищах. При цьому важливо одночасно враховувати продуктивність, затримки в передачі даних та обробки, енергетичні параметри і вартість балансування та перерозподілу. Додаткової складності додає багатопараметричність самих цільових функцій, необхідність врахування ймовірностного характеру навантажень та особливості роботи окремих модулів.

У науковій літературі багато уваги приділяється еволюційним та метаевристичним алгоритмам, які надають можливість підвищити ефективність процесів динамічного розподілення задач. Досягається це за рахунок адаптивного пошуку рішень у просторах з декількома вимірами. Перспективними напрямками є алгоритми, які моделюють біологічні або фізичні процеси та демонструють здатність до швидкої самоадаптації.

У статті запропоновано математичні моделі динамічного розподілення задач у гетерогенних комп'ютерних системах. Запропонований підхід розрахований на методи оптимізації з адаптивними евристичними, що дозволяє врахувати як апаратні особливості ресурсів, так і мінливий характер обчислювальних потоків. Розроблені рішення спрямовані на підвищення стабільності системи за динамічних умов і можуть бути застосовані у хмарних середовищах, на НРС-платформах та в інфраструктурах обробки великих даних.

Аналіз останніх досліджень і публікацій. Проблема динамічного розподілення обчислювальних задач у гетерогенних системах привертає значну увагу дослідників, насамперед через збільшення варіативності апаратної платформи (CPU, GPU, FPGA) та непередбачуваність навантажень у реальному часі. У класичному підході до віртуалізації і балансування навантаження застосовуються прості евристичні та правила (наприклад, round-robin, least-loaded), що, однак, часто не дозволяють ефективно поєднати енергетичну ефективність і суворі SLA у сучасних датацентрах. Комплексні багатокритеріальні підходи, які враховують одночасно енергію, час відгуку та вартість міграцій, все активніше розглядаються

як більш привабливі для практичних систем, але вимагають високоефективних моделей для методів пошуку [1, с.638].

За останні роки помітно зросла кількість робіт, що застосовують сучасні метаевристичні методи до задач розміщення віртуальних машин та консолідації ресурсів. Зокрема, поліпшені версії Slime Mould Algorithm (SMA) демонструють високі результати в оптимізаційних задачах, а в роботах 2023–2024 років запропоновано низку модифікацій, які підвищують збіжність та зменшують ризик застрягання у локальних мінімумах [1, с.641; 2, с.3]. Ці вдосконалення роблять SMA перспективною базою для задач управління ресурсами.

Окремий напрям становлять метаевристики, що використовують чисельні схеми на основі методів Рунге–Кутти. У роботах 2023 року показано ефективність таких підходів у задачах з високою розмірністю параметричного простору, де класичні еволюційні алгоритми втрачають стабільність [3, с.124]. Це робить RUN-підходи привабливими як компонент локального вдосконалення всередині комбінованих оптимізаторів.

У практичних системах хмарних обчислень 2023–2024 років спостерігається тенденція до одночасного врахування енергоспоживання, часу відповіді та вартості переміщення віртуальних машин. Автори робіт [4, с.276; 5, с. 76] підкреслюють, що оптимальні стратегії балансування навантаження базуються на гібридних методах: швидкі евристичні забезпечують оперативність рішень, а метаевристики – точність на довгих часових горизонтах.

Ще одним активно досліджуваним напрямом є використання адаптивних популяційних алгоритмів, зокрема модифікованих Harris Hawks Optimization (ННО), для задач розподілу навантаження у хмарних та контейнерних середовищах [6, с. 3097; 7, с. 5]. Результати цих робіт демонструють, що такі методи забезпечують високий рівень стабільності навіть при обмежених обчислювальних ресурсах. Водночас автори наголошують на потребі точнішого визначення штрафних функцій і динамічних критеріїв міграції, що і визначає наукові розриви у сучасних дослідженнях [8, с. 14].

Підсумовуючи, у наявній літературі бракує узгоджених моделей, які одночасно:

- формалізують багатокритеріальну ціль із урахуванням енергоспоживання, QoS та вартості міграцій;
- забезпечують локальний критерій прийняття рішень, придатний для онлайн-орієнтованих оркестраторів;

– інтегрують комбіновані метаевристики у єдину узагальнену процедуру пошуку.

Запропоновані у статті моделі спрямовані саме на усунення цих прогалин.

Постановка завдання. У гетерогенних обчислювальних системах, що поєднують різноманітні апаратні ресурси (CPU, GPU, високошвидкісні прискорювачі), виникає потреба у розподіленні задач таким чином, щоб одночасно досягати високої продуктивності та енергетичної ефективності. Динамічне колювання навантажень, відмінності в архітектурі вузлів і різна інтенсивність обчислень у задачах створюють багатокритеріальну проблему оптимізації, яка ускладнюється необхідністю реагувати в реальному часі. Існуючі методи балансування навантаження часто орієнтуються на один критерій (наприклад, мінімізацію вартості обчислень, часу виконання або енергії), а тому не забезпечують збалансованості рішень у сучасних середовищах із високою динамікою стану.

У цьому дослідженні ставиться задача формалізувати модель, що дозволяє описати процес динамічного розподілення обчислювальних задач між вузлами гетерогенної системи, де кожний вузол характеризується обмеженою ємністю, а задачі – змінними у часі вимогами. Ключовою особливістю моделі є одночасне врахування трьох компонентів:

– енергоспоживання обчислювальних вузлів, яке залежить від їх активного стану і рівня завантаження;

– відхилення від допустимого рівня продуктивності (QoS), що відображає ризик перевантаження та затримок;

– вартості міграції між вузлами, яка включає часові та енергетичні втрати, пов'язані з перенесенням задач.

Модель передбачає, що система керування отримує дані моніторингу в дискретні моменти часу з інтервалом T , у межах якого навантаження задач вважається кусково-сталім. Кожному вузлу приписується індивідуальна характеристика енергоспоживання, яка дає змогу описати як базові витрати, так і динамічну складову, пропорційну його завантаженню. Це дозволяє враховувати ключові відмінності між типами апаратури, що є критичним для гетерогенних систем.

Проблема розподілення формулюється як оптимізація комбінованої цільової функції, яка включає енергетичний складник, штраф за перевантаження та вартість міграцій. Таким чином, модель забезпечує рівновагу між мінімізацією енергоспоживання та дотриманням вимог щодо продуктивності.

Структура цільової функції забезпечує гладкість штрафів за QoS-порушення та дозволяє використовувати як аналітичні методи, так і метаевристичні процедури пошуку.

Особливістю запропонованої постановки є її придатність до інтеграції з онлайн-орієнтованими алгоритмами управління: у межах кожного інтервалу T контролер може оцінити стан вузлів, спрогнозувати короткострокові зміни навантаження та ініціювати локальні або глобальні перенесення задач. Обмеження міграцій та можливість переведення малозавантажених вузлів у режим standby дозволяють моделі відображати реальні процеси в сучасних датацентрах.

Таким чином, сформульована модель задає формальний апарат для подальшого розроблення методів оптимального або наближеного керування ресурсами, зокрема гібридних алгоритмів, що поєднують швидкі локальні рішення та глобальну оптимізацію на основі сучасних метаевристик.

Нехай існують N віртуальних машин VM_i ($i = 1..N$) та M фізичних вузлів PM_j ($j = 1..M$). Нехай для кожної VM відомі її поточні потреби в ресурсах (агрегований показник) $r_i \geq 0$. Ємність кожного фізичного вузла позначимо $C_j > 0$. Нехай бінарна змінна x_{ij} означає, що VM_i розміщена на вузлі j або ні ($x_{ij}=0$). Обмеження цілісності розміщення: для кожної VM має виконуватись $\sum_{j=1}^M x_{ij} = 1$.

Завантаження вузла j (у відносних одиницях) позначимо U_j , його розрахунок:

$$U_j = \frac{1}{C_j} \sum_{i=1}^N r_i x_{ij}.$$

Енергоспоживання вузла j нехай залежить від його завантаження й складатиметься з базової витрати P_j^{idle} (споживання у режимі простою) та динамічної частини, пропорційної завантаженню: для простоти прийемо модель

$$P_j = \begin{cases} 0, & \text{якщо вузол вимкнено} \\ P_j^{idle} + \gamma_j U_j, & \text{якщо вузол увімкнено} \end{cases}$$

де γ_j – коефіцієнт, що відображає додаткове споживання при завантаженні. У моделі враховуємо, що при низькому сумарному навантаженні деякі вузли можуть бути поставлені у стан standby (вимкнені) з нульовим активним споживанням (спрощено).

Функція мети. Оскільки потрібно врахувати дві цілі – мінімізацію енергоспоживання та підтримку продуктивності – формулюємо комбіновану функцію мети у вигляді зваженої суми:

$$\min(F) = \alpha \cdot E + \beta \cdot Q,$$

де $E = \sum_{j=1}^M P_j$ – сумарне енергоспоживання кластера, Q – скоригований показник «неякості» (quality penalty), що вимірює відхилення від бажаних QoS-

параметрів (наприклад, перевантаження вузлів або перевищення допустимого часу відповіді), а α , β – невід’ємні вагові коефіцієнти, які задає оператор або система управління відповідно до пріоритетів (баланс між енергією та продуктивністю). Важливо: вибір α та β дає змогу налаштувати політику – від енергоорієнтованої до performance-oriented.

Пропонуємо просту інтерпретацію Q : нехай перевантаження вузла j вважається критичним, якщо $U_j > U^{thr}$ (наприклад, $U^{thr} = 0.8$). Тоді

$$Q = \sum_{j=1}^M \max(0, U_j - U^{thr})^2.$$

Це означає, що невелике перевищення порогу штрафується слабо, а значні перевантаження отримують квадратичне покарання – це спонукає до більш рівномірного розподілу навантаження. Квадратична форма дає гладкість і ускладнює концентрацію багатьох VM на одному вузлі.

Обмеження. Модель доповнюється стандартними обмеженнями:

- ємнісні обмеження: для кожного j виконується $\sum_{i=1}^N r_{ij} x_{ij} \leq C_j$;
- у кожній VM – одне розміщення: $\sum_{j=1}^M x_{ij} = 1$;
- бінарність: $x_{ij} \in \{0,1\}$.

Вартість міграції. Окремо вводимо штраф за міграцію VM, оскільки часті переміщення знижують QoS і витрачають ресурси. Нехай m_i – вартість переміщення VM_i (у енергетичних або часових одиницях), а y_i – індикатор того, чи була VM переміщена у поточному кроці (1 – так, 0 – ні). Тоді додаємо до функції мети терм $\delta \sum_{i=1}^M m_i y_i$, де δ – ваговий коефіцієнт для зменшення частоти міграцій. У підсумку повна функція мети:

$$F = \alpha \sum_{j=1}^M P_j + \beta \sum_{j=1}^M \max(0, U_j - U^{thr})^2 + \delta \sum_{i=1}^N m_i y_i.$$

Зауваження: в умовах реального датацентру параметри $P_j^{idle}, \gamma_j, C_j, m_i$ вимірюються або оцінюються на підставі моніторингу. Для цілей роботи їх можна задати експериментально у моделюванні (симуляції).

Формальна модель локальної оптимізації

Нехай поточна конфігурація системи описується матрицею $X = [x_{ij}]$, а зміна стану після переміщення однієї віртуальної машини VM_i з вузла a на вузол b позначається $X' = X + \Delta_{i,a,b}$. Визначимо градієнтну апроксимацію зміни функції мети (Ідейна формула локальної вигоди переносу $i:a \rightarrow b$):

$$\begin{aligned} \Delta F_{i,a,b} &= F(X') - F(X) \approx \alpha \cdot \Delta E_{a,b} + \beta \cdot \Delta Q_{a,b} + \delta \cdot m_i, \\ \Delta E_{a,b} &= \gamma_a (U'_a - U_a) + \gamma_b (U'_b - U_b), \\ \Delta Q_{a,b} &= \max(0, (U'_a - U^{thr})^2) - \max(0, (U_a - U^{thr})^2) + \\ &\quad \max(0, (U'_b - U^{thr})^2) - \max(0, (U_b - U^{thr})^2), \end{aligned}$$

де U'_a, U'_b – нові завантаження після переносу обчислюються:

$$U'_a = \frac{\sum_i (r - r_i)}{c_a}, U'_b = \frac{\sum_i (r - r_i)}{c_b}.$$

Рішення вважається доцільним, якщо $\Delta F_{i,a,b} < 0$ та після переносу не порушуються ємнісні обмеження: $\sum_i r_i x_{ib} \leq C_b$.

Пояснення вибору компонентів моделі та інтерпретація параметрів

Математична структура моделі (див. рис. 1) обрана так, щоб відобразити два взаємопротилежні прагнення: мінімізувати енергоспоживання (через консолідацію та вимикання вузлів) та уникати перевантаження ресурсів (через квадратичну штрафну функцію).

Structural Model of Dynamic Task Allocation

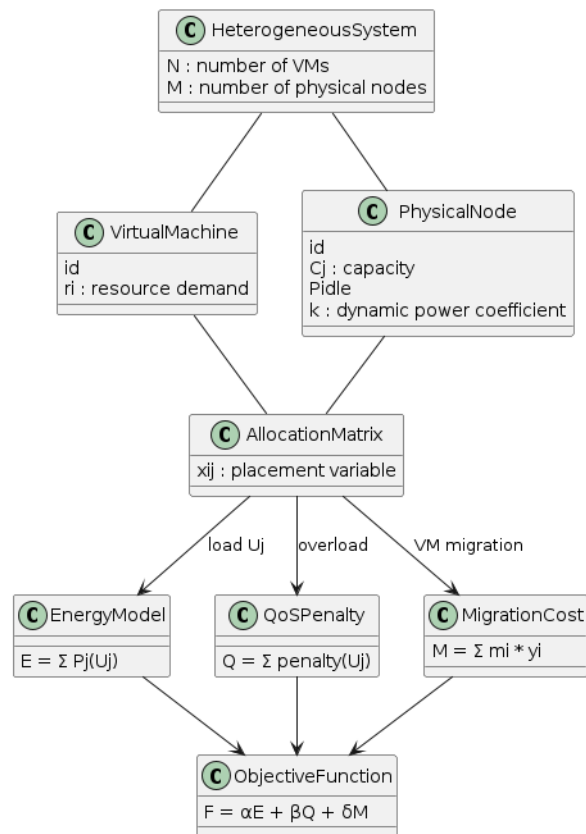


Рис. 1. Структурна UML діаграма взаємозв'язків моделі

Квадратичний штраф за перевантаження стимулює більш рівномірний розподіл навантаження, ніж лінійна форма, і робить стратегію більш «консервативною» щодо локальних перевищень порогу. Коефіцієнти α і β – інструмент налаштування політики. Якщо оператор ставить пріоритет на енергію (економія витрат), α збільшують відносно β ; якщо ж на першому місці QoS – навпаки.

Параметр δ контролює агресивність міграцій: велике δ зменшує кількість переміщень і робить систему більш «стабільною» у плані QoS. Поріг U^{thr} та допоміжні U^{upper} , U^{target} задаються експериментально: наприклад, $U^{thr} = 0.8, U^{upper} = 0.95, U^{target} = 0.65$. Така схема дозволяє уникати частих осциляцій (hysteresis) – вузол вважається перевантаженим тільки при досягненні вищого порогу, і після перерозподілу його завантаження приводиться до цільового значення нижчого рівня. Модель міграції спрощена: вартість m_i може залежати від об'єму пам'яті VM та доступної пропускної здатності мережі. У реальних умовах можна використовувати більш складні оцінки (включаючи час зупинки сервісу), але спрощена модель достатня для демонстрації поведінки алгоритму.

Експериментальна перевірка пропонує моделей динамічного розподілення задач виконувалася з використанням симуляційного середовища GridSim з відкритим кодом [9, с. 1190]. Воно дозволяє моделювати розподілені обчислювальні системи різномірної природи, які можуть включати ресурси різної продуктивності, мережевої затримки та врахувати динамічні потоки задач. Використовувалась Java-платформа із розширенням стандартного планувальника шляхом інтеграції запропонованої багатокритеріальної функції мети. Протягом експерименту було змодельовано кластер із 10 гетерогенних фізичних вузлів трьох типів із різними характеристиками продуктивності та енергоспоживання. У систему надходило 60 віртуальних машин із змінними вимогами до ресурсів, що формували квазіперіодичні пікові навантаження. Тривалість кожного експериментального прогону становила одну годину модельного часу, а результати усереднювалися за десятьма незалежними ітераціями для зменшення впливу стохастичних факторів. Для порівняння було реалізовано три режими керування: базовий евристичний алгоритм типу round-robin, запроповану модель у збалансованому режимі вагових коефіцієнтів та енергоорієнтований режим із підвищеним значенням коефіцієнта

при енергетичній складовій функції мети. Оцінювання проводилося за такими показниками: середнє енергоспоживання кластера, частка часу перевантаження вузлів, середня кількість міграцій за годину та інтегральне значення функції мети. Узагальнені кількісні результати наведено в таблиці 1.

Отримані результати свідчать про те, що застосування запропонованої моделі дозволяє зменшити середнє енергоспоживання на 13–18 % порівняно з базовим алгоритмом. При цьому скорочується частка часу перебування вузлів у перевантаженому стані. У збалансованому режимі досягнуто найменшого інтегрального значення функції мети. Це підтверджує доцільність використання багатокритеріального підходу. З іншого боку, енергоорієнтований режим забезпечує мінімальне енергоспоживання, однак супроводжується більшим рівнем перевантаження та відбувається зростання кількості міграцій. Збільшення кількості міграцій у запропонованій моделі можна пояснити активною реакцією алгоритму на зміни навантаження. Введення вагового коефіцієнта при вартості міграцій дозволяє контролювати їх інтенсивність та уникати надмірних переміщень у короткочасних флуктуаціях.

Висновки. Проведене експериментальне дослідження підтвердило ефективність запропонованої моделі динамічного розподілення обчислювальних задач у гетерогенних комп'ютерних системах. Багатокритеріальна функція мети, яка поєднує енергетичну складову, штраф за перевантаження та вартість міграцій дозволила забезпечити збалансоване керування ресурсами за умов динамічної зміни навантаження. Результати моделювання демонструють можливість зниження енергоспоживання кластера до 18 % порівняно з традиційними евристичними підходами. Одночасно зменшення рівня порушень QoS сталося більш ніж у два рази. На нашу думку, найкращий інтегральний результат було отримано у режимі збалансованих вагових коефіцієнтів. Це підтверджує доцільність адаптивного налаштування параметрів моделі залежно від

Таблиця 1

Порівняння результатів моделювання

Режим керування	Середнє енергоспоживання, Вт	Частка перевантаження, %	Кількість міграцій (за годину)	Значення функції мети F
Round-robin	1245	12.8	6	1.00 (нормоване)
Запропонована модель (збаланс.)	1082	4.6	14	0.71
Запропонована модель (енергоорієнт.)	1015	7.9	18	0.76

стратегії управління. Отримані результати свідчать про практичну придатність запропонованих моделей для застосування у хмарних та гетерогенних обчислювальних середовищах. Подальші дослідження доцільно спрямувати на розширення

моделі до багаторесурсного випадку з окремим урахуванням CPU, пам'яті та мережних параметрів, а також на інтеграцію гібридних метаевристичних алгоритмів для глобальної оптимізації розміщення задач.

Список літератури:

1. Verma P., Maurya A. K., Yadav R. A survey on energy-efficient workflow scheduling algorithms in cloud computing. *Software: Practice and Experience*. 2023. vol. 54. pp. 637–682. DOI: 10.1002/spe.3292
2. Bano F, Ahmad F, Shahid M, Alam M, Hasan F, Sajid M. A Levelized Multiple Workflow Heterogeneous Earliest Finish Time Allocation Model for Infrastructure as a Service (IaaS) Cloud Environment. *Algorithms*. 2025. 18(2):99. <https://doi.org/10.3390/a18020099>
3. Mamchych O., Volk M. A unified model and method for forecasting energy consumption in distributed computing systems based on stationary and mobile devices. *Radioelectronic and Computer Systems*. [S.l.], v. 2024, n. 2. p. 120-135. DOI: <https://doi.org/10.32620/reks.2024.2.10>.
4. Senthilkumar G., Anandamurugan S. Energy and time-aware scheduling in diverse virtualized cloud computing environments using optimized self-attention progressive generative adversarial network. *Network: Computation in Neural Systems*. 2025. vol. 36. no. 2. p. 274–293. DOI: 10.1080/0954898X.2024.2391401
5. Волк М.О., Курочкін В. С., Запороженко А.П., Паронікян П.А. Гібридний метод розподілу ресурсів в хмарних системах. *Системи управління, навігації та зв'язку*. 2025. випуск 2(76). с. 70-83. doi:10.26906/SUNZ.2024.2.070.
6. Ahmed W., Gautam G., Alam B. et al. An analytical review and performance measures of state-of-art scheduling algorithms in heterogeneous computing environment. *Archives of Computational Methods in Engineering*. 2024. vol. 31. pp. 3091–3113. DOI: 10.1007/s11831-024-10069-8
7. Chandrasiri S., Meedeniya D. Energy-efficient dynamic workflow scheduling in cloud environments using deep learning. *Sensors*. 2025, vol. 25, art. 1428. DOI: 10.3390/s25051428
8. Miao Z., Shao C., Li H., Tang Z. Review of task-scheduling methods for heterogeneous chips. *Electronics*. 2025. vol. 14, art. 1191. DOI: 10.3390/electronics14061191
9. Buyya R., Murshed M. GridSim: a toolkit for the modeling and simulation of distributed resource management and scheduling for Grid computing. *Concurrency and Computation: Practice and Experience*, 2002, vol. 14, no. 13–15, pp. 1175–1220. DOI: 10.1002/cpe.710

Volk M.O., Kovtun Ye.I. MODELS OF DYNAMIC DISTRIBUTION OF COMPUTING TASKS IN HETEROGENEOUS COMPUTER SYSTEMS

The article is devoted to the dynamic distribution of computational tasks in heterogeneous computer systems. Such systems have resources with different performance and energy consumption characteristics. The relevance of this study is due to the growth of loads in cloud and distributed environments, as well as the need to ensure a balance between energy efficiency and compliance with quality of service (QoS) requirements. Existing planning approaches are often focused on individual criteria, which reduces their effectiveness in conditions of high dynamics. Therefore, research related to taking into account a significant number of characteristics of such systems is promising. The paper proposes a mathematical model of multi-criteria optimisation, in which the objective function is formulated as a weighted sum of three components: the total energy consumption of computing nodes, the penalty for resource overload, and the cost of virtual machine migration. Energy consumption is described by a linear dependence. To minimise the risk of performance degradation, a quadratic penalty function is applied for exceeding the permissible load threshold. This contributes to uniform load distribution. Additionally, the migration cost parameter is taken into account, which allows limiting the frequency of task transfer and increasing system stability. Experimental verification showed a decrease in average energy consumption of up to 18% compared to the baseline models and a significant reduction in the time nodes spend in an overloaded state. The best integral indicators were achieved in the balanced mode of setting weight coefficients. The proposed models demonstrate adaptability to load changes and suitability for use in cloud and high-performance computing systems. Prospects for further research are related to the expansion of the model to multi-resource environments and the integration of hybrid metaheuristic algorithms for global optimisation.

Keywords: heterogeneous computer systems, dynamic task scheduling, resource scheduling, energy efficiency, QoS, virtual machine migration, multi-criteria optimization, objective function, cloud computing.

Дата першого надходження статті до видання: 06.02.2026

Дата прийняття статті до друку після рецензування: 04.03.2026

Дата публікації (оприлюднення) статті 11.05.2026

УДК 519.876.5

DOI <https://doi.org/10.32782/2663-5941/2026.2.1/06>

Гейко О.О.

<https://orcid.org/0009-0006-3279-8274>

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Варава І.А.

<https://orcid.org/0000-0001-9874-016X>

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Пуха Г.С.

<https://orcid.org/0000-0001-5728-1577>

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Дембіцький В.В.

<https://orcid.org/0009-0004-4119-4742>

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Вдовина А.В.

<https://orcid.org/0009-0001-1116-708X>

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

АРХІТЕКТУРА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ВАЛІДАЦІЇ МАТЕМАТИЧНИХ МОДЕЛЕЙ

У статті описано підхід до створення програмної платформи для верифікації та валідації (V&V) математичних і чисельних моделей, яка автоматизує порівняння двох залежностей - отриманої з експерименту та сформованої в результаті моделювання [3,10]. Запропоноване програмне забезпечення орієнтоване на інженерні задачі, де надійність висновків суттєво залежить від об'єктивності процедури порівняння, а «візуальна оцінка» є недостатньою через суб'єктивність та різну інтерпретацію форм сигналів [5,7]. Платформа реалізує типову послідовність V&V: імпорт даних, керувану попередню обробку (фільтрацію, узгодження дискретизації, синхронізацію в часі та обрізку), обчислення набору детерміністичних метрик подібності форми та подання результатів у графічному інтерфейсі. Окремо висвітлено архітектуру програмного комплексу валідації, що включає підсистему взаємодії з сенсорними даними (розбір пакетів, відокремлення телеметрії та акустичних компонент, калібрування AHRs), підсистему моделювання та керування експериментами, контур аналізу параметрів і формування області параметрів, підсистему валідації з переходом у частотну область та розрахунком метрик, а також модулі візуалізації часових і спектральних представлень [11]. Для забезпечення відтворюваності процедура порівняння супроводжується фіксацією налаштувань попередньої обробки, параметрів моделі та умов експерименту, що дає змогу повторювати серії запусків і коректно зіставляти результати між різними сценаріями. Оцінювання подібності здійснюється як у часовій області (для контролю форми й синхронності сигналів), так і у спектральній (для аналізу частотного складу), що підвищує інформативність висновків у випадках наявності шумів або нестационарності процесів [5,7]. Результати обчислень зберігаються у форматах, придатних для звітності та подальшого аналізу (табличні дані та растрові графіки), а також можуть бути використані для порівняння альтернативних конфігурацій моделі та налаштувань обчислювального експерименту. Запропонований підхід зменшує ризик людських помилок, прискорює повторювані цикли V&V і підвищує довіру до моделей у процесі ухвалення інженерних рішень [10].

Ключові слова: верифікація, валідація, чисельне моделювання, порівняння часових залежностей, метрики подібності, попередня обробка сигналів, частотний аналіз, архітектура ПЗ.



Постановка проблеми. У багатьох прикладних галузях інженерії якість рішень залежить від того, наскільки точно чисельна модель відтворює реальну поведінку об'єкта чи процесу [3,10]. Типовим способом перевірки адекватності є порівняння часових рядів або сигналів: експериментального та розрахункового. Проте на практиці часто застосовують візуальну оцінку - аналіз піків, фазових зсувів, форми коливань [5,7]. Такий підхід не гарантує відтворюваності: різні фахівці можуть робити різні висновки за однакових даних, а формалізувати критерії прийнятності результату складно.

Отже, актуальною є задача розробки програмного інструмента, який переводить порівняння залежностей у площину формалізованих, кількісних показників, забезпечує повторюваність процедури та зручну інтерпретацію результатів у контексті V&V [7,9]. Концептуальні засади еталонної архітектури програмної платформи для такого класу задач наведено в [13].

Аналіз останніх досліджень і публікацій. У сучасних підходах V&V розрізняють два взаємодоповнювальні процеси: верифікацію (перевірку правильності реалізації математичної постановки та її чисельного розв'язання) і валідацію (оцінювання того, наскільки модель відповідає реальному світу в межах запланованого застосування). У рекомендаціях професійних спільнот ці поняття визначаються як базові етапи життєвого циклу комп'ютерного моделювання і є критичними для підвищення довіри до результатів, особливо в задачах із високими вимогами до безпеки та відповідальності рішень [9,10].

Окрема гілка досліджень присвячена метрикам порівняння залежностей, які дають числову оцінку подібності. Описані підходи включають метрики, що розділяють вплив амплітуди (величини) та фази/часового зсуву; інтегральні та точкові показники розбіжності; статистичні тести на основі аналізу варіацій, які інтерпретують залишок як шум за умови «однаковості» події [6,7].

Разом із тим практичне застосування таких метрик вимагає коректної підготовки даних: часто експеримент і модель мають різні частоти дискретизації, стартові моменти, довжини записів, а також різний рівень шуму. Тому програмні рішення в цій сфері мають поєднувати модулі попередньої обробки сигналів, обчислення метрик і засоби інтерпретації/візуалізації результатів. Для інженерних задач, зокрема у гідроакустичних вимірюваннях, це проявляється як принципова неузгодженість дискретизації: сенсор

фіксує сигнал з апаратно заданою частотою, тоді як часовий крок моделювання визначається умовами стійкості, параметрами розрахункової сітки та властивостями середовища. Тому ключовою є процедура узгодження дискретизації та часової прив'язки (ресемплінг/синхронізація) без втрати фізичного змісту сигналів.

Постановка завдання. Метою роботи є розроблення та реалізація програмної платформи для параметричного аналізу моделей із можливістю: (1) імпорту/зберігання та попередньої обробки даних моделювання й експерименту (синхронізація, ресемплінг, фільтрація); (2) виконання процедур верифікації та валідації з використанням метрик відповідності; (3) організації параметричних досліджень у вигляді black-box оптимізації для пошуку області параметрів Ω , що забезпечує прийнятну узгодженість моделі та експерименту; (4) формування звітів та візуалізації результатів.

Виклад основного матеріалу. Робочий процес будується як послідовність кроків: (а) завантаження даних; (б) узгодження та очищення сигналів; (в) розрахунок метрик; (г) інтерпретація і збереження результатів. Після введення двох кривих система автоматично виконує підготовчі операції, а потім формує набір метрик подібності і наочні графіки, що дозволяє швидко оцінити, наскільки модель відтворює експеримент.

Щоб метрики працювали коректно, криві потрібно зробити порівнюваними [7].

Фільтрація. Для зменшення впливу шумів застосовується цифрова низькочастотна фільтрація. Як типове рішення може використовуватися фільтр Баттерворта 4-го порядку з двопрорідною схемою (прямий та зворотний прохід), що дає близький до нульового фазовий зсув і відповідає підходам на кшталт SAE J211 [12] для обробки вимірювань у динамічних тестах.

Узгодження дискретизації (ресемплінг). Багато метрик ґрунтуються на порівнянні за відліками, тому обидва сигнали мають бути представлені з однаковою частотою дискретизації. Якщо частоти різні, дані з нижчою частотою перераховуються до вищої (наприклад, лінійною інтерполяцією).

Синхронізація в часі. Експериментальні та модельні дані можуть бути неузгоджені за часовою прив'язкою (містити часовий зсув), зумовлений, зокрема, запуском реєстрації, тригерними умовами або різними системами відліку часу. Для компенсації зсуву застосовуються методи пошуку такого часового зсуву, який мінімізує різницю між кривими – наприклад, мінімізація площі залишків

або суми квадратів залишків. За потреби користувач може повторити синхронізацію з іншим стартовим наближенням або іншим критерієм.

Після узгодження частоти та синхронізації система приводить криві до однакової тривалості: довгий сигнал обрізається до розміру коротшого.

Метрики порівняння залежностей. Для кількісної оцінки узгодженості між експериментальними та модельними даними використовується набір детерміністичних показників і критеріїв, що забезпечують відтворюване порівняння часових залежностей (сигналів) за умов можливої різної дискретизації, часових зсувів та обмежених обчислювальних ресурсів. Після приведення сигналів до порівнюваного вигляду (зокрема ресемплінгу, нормування та синхронізації) узгодженість у часовій області доцільно оцінювати похибковими метриками, які відображають відмінність за величиною та інтегральну розбіжність, зокрема абсолютним відхиленням, відносним коефіцієнтом та середньоквадратичною помилкою (MSE), що може застосовуватись як базовий критерій проходження/непроходження заданого порога. Для підвищення стійкості оцінювання у випадку широкосмугових завад та для фокусування на

фізично значущій складовій сигналу доцільно доповнювати аналіз переходом у частотну область, контролюючи наявність вираженої інформативної компоненти (спектрального піку на цільовій частоті або у вузькій смузі) та узгодженість фазових/часових характеристик відповідних компонент. У задачах параметричних досліджень коректність вибраної конфігурації доцільно оцінювати не лише за близькістю до експерименту, а і за критеріями чисельної здійсненності та стабільності, зокрема з урахуванням обмежень дискретизації й умов коректності типу CFL, а також за ресурсними показниками (час виконання та пікове споживання пам'яті) і, за потреби, за критерієм самоузгодженості результатів при контрольному прогоні з підвищеною точністю. Практичний досвід показує, що надмірна кількість метрик часто не підвищує інформативність через дублювання змісту, тому доцільно використовувати компактний набір показників, який покриває відмінність за величиною (абсолютна/відносна), інтегральну помилку (MSE) та часово-частотну узгодженість, доповнену критеріями коректності й ресурсної здійсненності для параметричного аналізу.

Функціональна модель (use-case)

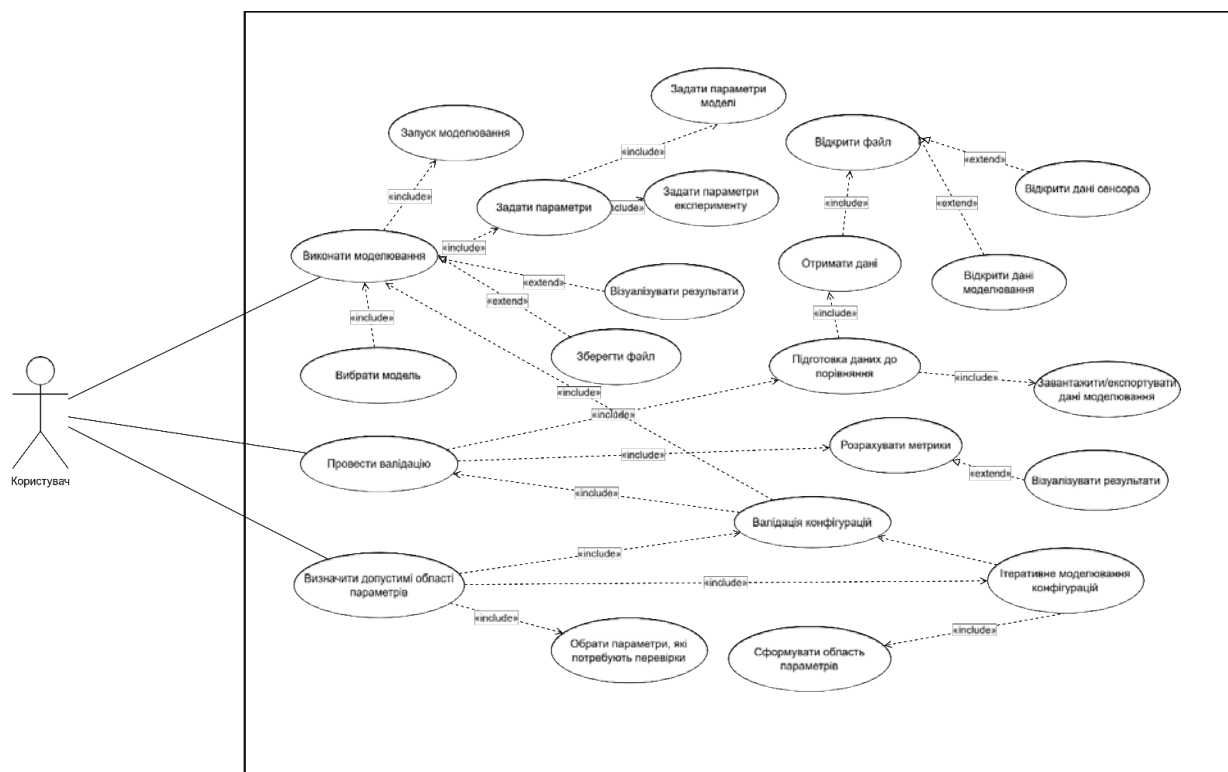


Рис. 1. Діаграма прецедентів програмного комплексу

Взаємодія з системою описується сценаріями, де користувач:

1. Завантажує/відкриває дані моделювання та експерименту/сенсора.
2. Задає параметри моделі та параметри експерименту (зокрема геометрію, координати, частотні та амплітудні характеристики, параметри середовища, профілі тощо).
3. Запускає моделювання, обирає модель/конфігурацію, виконує ітераційне уточнення налаштувань.
4. Проводить валідацію: готує дані до порівняння, обчислює метрики, переглядає результати.
5. Формує/уточнює допустимі області параметрів і визначає параметри, що потребують перевірки.
6. Візуалізує та зберігає підсумки у файли (таблиці/графіки).

Модульна архітектура. Архітектура організована як набір підсистем, що покривають весь цикл «дані → моделювання → V&V → звіт» [11].

Підсистема зв'язку із сенсором відповідає за приведення «сирих» вимірювань до придатного для подальшої роботи вигляду. У її складі реалі-

зовано читання бітової структури пакетів даних, відокремлення телеметрійної складової від акустичної, а також калібрування AHRS. Завдяки цьому на виході формується узгоджений і структурований набір сенсорних даних, який можна коректно обробляти та порівнювати з результатами моделювання.

Модуль FFT забезпечує виконання спектральних перетворень. Його призначення – переведення сигналів у частотну область для аналізу спектральних характеристик, а також для тих метрик і процедур порівняння, які чутливі до частотного складу, домінуючих гармонік та енергетичного розподілу.

Підсистема моделювання включає модуль завантаження моделей і призначена для підключення, вибору та підготовки конкретної чисельної моделі до запуску. Вона забезпечує коректну ініціалізацію модельних компонентів і формує основу для подальшого виконання розрахунків відповідно до заданої конфігурації.

Підсистема управління експериментами виконує роль керівного контуру, який збирає параметри, запускає моделювання та дає змогу вико-

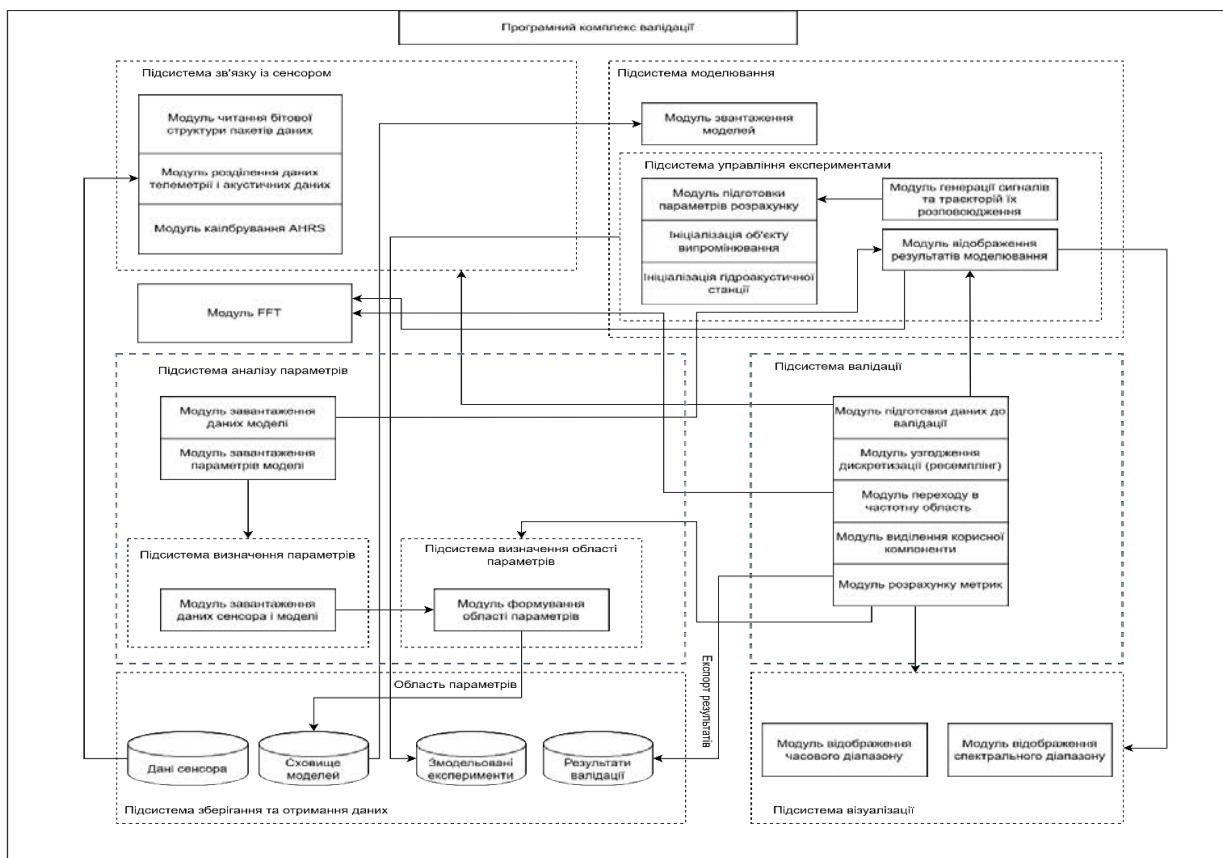


Рис. 2. Архітектура програмного забезпечення для валідації математичних моделей

нати первинний перегляд отриманих результатів. У межах цієї підсистеми здійснюється підготовка параметрів розрахунку, ініціалізація об'єкта випромінювання та гідроакустичної станції, генерація сигналів і траєкторії розповсюдження, а також відображення результатів моделювання для швидкої оцінки коректності виконання експерименту.

Підсистема аналізу параметрів і формування області параметрів орієнтована на підтримку параметричних досліджень та ітераційного V&V. У загальному випадку модель можна розглядати як чорний ящик $y = F(p)$, де p - вектор параметрів; необхідно сформулювати область Ω у параметричному просторі, в якій одночасно виконуються: (1) вимоги точності за обраними метриками порівняння, (2) чисельна коректність та стабільність розрахунку, (3) ресурсна здійсненість (час/пам'ять). Пошук такої області доцільно реалізовувати як поетапний відбір конфігурацій з відсівом за кожним з критеріїв. Вона охоплює завантаження даних і параметрів моделі, спільне використання даних сенсора та моделі, а також формування області параметрів. Завдяки цьому стає можливим визначати допустимі діапазони, системно відбирати параметри, які потребують перевірки, і організовувати серії запусків із контрольованими змінами налаштувань.

Підсистема валідації є центральною частиною комплексу, оскільки саме тут реалізовано основну логіку порівняння сигналів і формування числових показників адекватності моделі. У ній виконується підготовка даних до порівняння, узгодження дискретизації шляхом ресемплінгу, перехід у частотну область, виділення корисної компоненти та розрахунок метрик. Результатом роботи підсистеми є набір кількісних оцінок, що відображають ступінь відповідності моделі експерименту.

Підсистема візуалізації забезпечує інтерпретацію отриманих результатів у наочній формі. Вона реалізує відображення сигналів у часовому діапазоні та у спектральному представленні, що дозволяє інженеру швидко побачити ключові відмінності й перевірити, чи узгоджується метрико-оцінка з візуальною картиною порівняння.

Підсистема зберігання та отримання даних організовує роботу зі сховищами, у яких накопичуються дані сенсора, моделі, змодельовані експерименти та результати валідації. Така структура підтримує повторні запуски, дає змогу порівню-

вати серії експериментів і суттєво спрощує підготовку звітних матеріалів.

Експорт і документування результатів забезпечують збереження підсумків роботи платформи у формах, зручних для інженерної практики. Числові значення метрик можуть зберігатися як табличні дані, а графічні порівняння - як зображення.

Висновки. У роботі розглянуто підхід до побудови програмної платформи верифікації та валідації математичних і чисельних моделей на основі автоматизованого порівняння експериментальних та модельних часових залежностей. Запропонований процес орієнтований на підвищення відтворюваності оцінювання та зменшення суб'єктивності, що притаманна суто візуальному аналізу, і реалізує послідовність етапів імпорту даних, їх підготовки, кількісного оцінювання та формування звітних матеріалів.

Сформульовано вимоги до приведення експериментальних і модельних даних до порівнюваного вигляду з урахуванням можливої різної частоти дискретизації, часових зсувів і шумів вимірювань. Це забезпечує коректне обчислення метрик узгодженості та можливість повторюваного застосування процедури V&V для серійних експериментів. Для кількісної оцінки узгодженості застосовано детерміністичні метрики, зокрема абсолютне відхилення, відносний коефіцієнт та середньоквадратичну помилку, а також використано аналіз у частотній області для контролю інформативної складової сигналу та узгодженості фазових/часових характеристик.

Постановку параметричних досліджень сформульовано як задачу визначення області параметрів Ω , що забезпечує виконання заданих критеріїв узгодженості за метриками V&V, з урахуванням чисельної коректності та обмежених ресурсів. Запропоновано враховувати додаткові умови здійсненності, зокрема обмеження дискретизації та критерії стабільності типу CFL, а також ресурсні показники (час виконання та пікове споживання пам'яті), що є критичним для практичного застосування платформи в умовах обмежених обчислювальних ресурсів.

Отримані результати підтверджують доцільність використання запропонованої архітектури та контуру обробки даних для прискорення і стандартизації циклів V&V, а також для підвищення обґрунтованості рішень, що приймаються на основі результатів моделювання.

Список літератури:

1. Bruce A. McCarl. Model Validation: An Overview with some Emphasis on Risk Models. The World's Largest Open Access Agricultural & Applied Economics Digital Library, 1984. 1–5.
2. D. Sornette, A. B. Davis, K. Ide, K. R. Vixie, V. Pisarenko, and J. R. Kamm. Algorithm for model validation: Theory and applications, 2007. 1–2. <https://doi.org/10.1073/pnas.0611677104>

3. D.J. Murray-Smith. Testing and Validation of Computer Simulation Models, Simulation Foundations, Methods and Applications. Springer International Publishing Switzerland, 2015. 1–20.
4. Gonçalves S. N., Albuquerque D. M., Pereira J. C. Modelling and energy efficiency analysis of the microwave continuous processing of limestone. Journal of Cleaner Production, 142912, 2024. <https://doi.org/10.1016/j.jclepro.2024.142912>
5. Mongiardini M. Development of a Computer Program for the Verification and Validation of Numerical Simulations in Roadside Safety. Worcester Polytechnic Institute, 2010. <https://digitalcommons.wpi.edu/etd-dissertations/276/>
6. Mongiardini M., Ray M. H., Anghileri M. Acceptance criteria for validation metrics in roadside safety based on repeated full-scale crash tests. International Journal of Reliability and Safety, 4(1), 2010. 69. <https://doi.org/10.1504/ijrs.2010.029565>
7. Mongiardini M., Ray M. H., Anghileri M. Development of a software for the comparison of curves during the verification and validation of numerical models. 7th European LS-DYNA Conference, 2009. 1–12. <https://www.dynalook.com/european-conf-2009/K-I-03.pdf>
8. Pal N., Yadav D. K. Modeling and verification of software evolution using bigraphical reactive system. Cluster Computing, 2024. <https://doi.org/10.1007/s10586-024-04597-y>
9. Ray M. H., Mongiardini M., Atahan A. O., Anghileri M. Recommended procedures for verification and validation of computer simulations used for roadside safety. ResearchGate, 2008. https://www.researchgate.net/publication/313094419_Recommended_procedures_for_verification_and_validation_of_computer_simulations_used_for_roadside_safety_applications
10. Schwer L. E. An overview of the PTC 60/V&V 10: guide for verification and validation in computational solid mechanics. Engineering With Computers, 23(4), 2007. 245–252. <https://doi.org/10.1007/s00366-007-0072-z>
11. Tao F., Sun X., Cheng J., Zhu Y., Liu W., Wang Y., Xu H., Hu T., Liu X., Liu T., Sun Z., Xu J., Bao J., Xiang F., Jin X. makeTwin: A reference architecture for digital twin software platform. Chinese Journal of Aeronautics, 37(1), 2024. 1–18. <https://doi.org/10.1016/j.cja.2023.05.002>
12. SAE International. SAE J211/1_202208 Instrumentation for Impact Test Part I – Electronic Instrumentation. SAE International, 2022. doi:10.4271/J211/1_202208.
13. Гейко О., Варава І. Еталонна архітектура для програмної платформи верифікації математичних моделей. Information Technology and Society, №2(13), 2024. DOI: 10.32689/maup.it.2024.2.3.

Heiko O.O., Varava I.A., Pukha H. S., Dembitsky V.V., Vdovyna A.V. SOFTWARE ARCHITECTURE FOR VALIDATING MATHEMATICAL MODELS

This paper describes a software platform that supports verification and validation (V&V) of mathematical and numerical models by automating the comparison of experimental and simulated signals. The platform is intended for engineering applications where the credibility of conclusions depends on an objective and traceable comparison procedure, and where purely visual curve assessment is insufficient due to subjectivity, varying interpretation of waveform features, and sensitivity to scaling, offsets, or noise. To address these limitations, the platform provides a repeatable end-to-end workflow that integrates data acquisition/import, controlled pre-processing, quantitative metric evaluation, visualization, and reporting in a single toolchain.

In the proposed approach, experimental recordings and simulation outputs are imported from heterogeneous sources and aligned through a unified pre-processing stage. This stage explicitly accounts for differences in sampling rates, time shifts, segment length, and measurement noise, and may include filtering, resampling/interpolation, trimming to comparable windows, and time synchronization to ensure consistent signal matching for subsequent analysis. All processing steps and their settings are recorded to improve reproducibility and enable fair comparisons across multiple experiments, model revisions, and hardware configurations.

V&V is performed using a set of deterministic similarity metrics that quantify agreement in waveform shape and signal characteristics, complemented by visualization in both the time and frequency domains. The combination of numerical scores and graphical representations provides interpretable evidence of agreement (or mismatch), supports diagnostics of error sources (e.g., phase misalignment versus spectral discrepancy), and facilitates consistent decision making based on predefined acceptance criteria. The platform outputs results in reporting-friendly formats (e.g., tables and raster plots) to support documentation and further offline analysis.

Keywords: verification, validation, numerical simulation, curve comparison, similarity metrics, signal preprocessing, frequency analysis, software architecture.

Дата першого надходження статті до видання: 26.02.2026

Дата прийняття статті до друку після рецензування: 20.03.2026

Дата публікації (оприлюднення) статті 11.05.2026

Glukhova N.V.

<https://orcid.org/0000-0003-0817-5465>

Dnipro University of Technology

SPECIFICS OF METHODS FOR ASSESSING THE ACCURACY OF PARAMETER DETERMINATION WHEN IMPROVING COMPUTER-INTEGRATED TECHNOLOGY FOR ANALYSIS OF GAS-DISCHARGE RADIATION IMAGES

The article is devoted to the consideration of the features of methods for assessing the accuracy of determining image parameters. Existing problems associated with the specifics of methods for obtaining measurement information based on image processing and assessing the accuracy of determining geometric and brightness parameters of images are analyzed.

The task of improving modern computer-integrated technologies, an integral part of which is the image registration and processing subsystem, is associated with ensuring increased accuracy and reliability of measurement results.

The implementation of computer-integrated technologies that include image processing procedures requires an analysis of the components of measurement uncertainty that arise at different stages of the implementation of information measuring technology.

The article provides a comparative analysis of the specifics of image registration and processing methods, which allowed us to determine the specifics of methods for assessing the accuracy of parameter determination when improving computer-integrated technology for analyzing gas-discharge radiation images. Compared with other measurement conditions and models used in measurement methods with visual measurement information, the gas-discharge radiation method can be characterized by two specific features: the absence of fluctuations in the external illumination level, which is provided at the hardware level in accordance with the requirements of the measurement experiment; the presence of requirements for determining and setting the parameters of the hardware and software complex for image registration.

To compare different modifications of hardware and software complexes for recording images of gas-discharge radiation, it is proposed to use a test object and the corresponding methodology for experimentally assessing the random dispersion of geometric parameters of images based on the assessment of the mean square deviation for individual series of observations made at different time intervals. To substantiate the possibility of combining individual series of observations into a sample when processing experimental data, the Student and Fisher criteria were used, based on which the hypothesis of equal accuracy of the performed series of observations was confirmed.

Keywords: computer-integrated technology, image processing, gas-discharge radiation, measurement uncertainty

Formulation of the problem. When developing modern computer-integrated technologies, it is necessary to pay attention to the improvement of hardware and software. The corresponding improvement consists both in the positive dynamics of qualitative changes in the technology of obtaining and processing experimental data, and in ensuring the improvement of quantitative indicators. Qualitative changes may consist in expanding the list of studied parameters and improving the methods of processing experimental data used in computer-integrated technology.

Improving quantitative indicators, which together determine the quality of modern computer-integrated technology, consists primarily in increasing the accuracy and reliability of the results obtained during the registration and processing of experimental data.

Analysis of recent research and publications.

The vast majority of modern automated systems and computer-integrated technologies imply the registration of data in quantitative form. In this case, methods are used both for measuring individual parameters of research objects, and more complex methods for an



integral assessment of the state of various objects or processes, which may be based on obtaining images. In both the first and second cases, quantitative values are obtained as a result of the application of certain measurement methods and information-measuring technologies.

In accordance with modern requirements for the quality of measurements, their accuracy should be assessed in the form of measurement uncertainty. If the uncertainty is calculated when measuring individual physical quantities obtained using appropriate measuring instruments or transducers, then today the methods have been developed that are regulated in the relevant standards [1].

In cases where the computer-integrated technology includes components (subsystems) designed to record, convert, process and analyze images, their use requires appropriate procedures for assessing the accuracy of parameters determined on the basis of image analysis. Due to the development of computer-integrated technologies, images are increasingly used as measurement information, including for measurements in laboratory conditions, in scientific research and in industry.

It should be taken into account that the general statement of the measurement task, the measurement goal and the implementation of the measurement experiment assume the presence of a model of the measurement object. In the case of using images as measurement information, it is first necessary to determine which features of the images will be used as measurement results, whether it is planned to determine their quantitative value and perform an accuracy assessment. When solving this issue, problems may arise that depend on the specifics of the object under study.

In [2] it is noted that there are no general models for the analysis of images or their parts, characteristics, and relationships. Variants of such models are known, but they are created for individual image processing tasks or objects of a certain class. Also, according to the conclusions made in [2], there are a number of problems associated with the development of universal methods for assessing the accuracy of measurements, which are implemented on the basis of image analysis.

The process of obtaining measurement information in the form of images is significantly different from measurement procedures and methods for performing measurements of one individual physical quantity, so it always includes several stages, in particular, registration, processing and analysis of the image. Thus, the sources of measurement uncer-

tainty and the corresponding uncertainty components should be analyzed and calculated separately for each of these stages. The authors of the article [3] proposed to allocate three separate steps at which the measurement uncertainty components should be analyzed. In the article [4], the use of calibration is proposed to assess the accuracy of image parameters, which provides an estimate of the uncertainty of the dimensions in the image.

From the analysis of literary sources, it can be concluded that the vast majority of scientific research related to the assessment of the uncertainty of quantitative characteristics of images is aimed at tasks that arise when studying images of real objects in industry, in particular in the fields of machine vision systems and robotic systems. For these industries, the tasks of assessing the uncertainty components caused by changes in the illumination of the object of research; changes in the geometric dimensions, shapes, contours of objects; changes in the image background are relevant. But not all tasks of image registration and processing are concentrated in these areas.

When using the gas-discharge radiation method, the image acquisition and registration are significantly different from similar procedures common in machine vision systems. The gas-discharge radiation method involves performing an active measuring experiment under the influence of a pulsed electromagnetic field on the liquid sample under study. As a result, an image of a corona gas discharge is formed, which is a superposition of a spatial structure consisting of a set of gas-discharge tracks [5, 6].

Task statement. Since the procedure for generating gas-discharge radiation images and possible methods for their registration is characterized by significant differences from the objects of research in machine vision systems and the specifics of the measurement model, it requires the development or modification of existing methods for assessing the uncertainty of image parameters.

The task of the research is to determine the specifics of methods for assessing the accuracy of image parameters when improving computer-integrated technology for analyzing gas-discharge radiation images.

Outline of the main material of the study. When studying the properties of liquid-phase objects by the method of gas-discharge radiation, the stage of formation of a corona gas discharge in the form of a set of gas-discharge tracks should be considered as a probabilistic process. The complex physical processes of the emergence and existence of gas discharges underlie the stochastic dynamic model of the object

of research. Thus, the model of the measured object turns out to be quite complex, so the determination of its parameters should be based on the methods of probability theory and mathematical statistics.

The hardware and software complex used to perform experimental studies involves the formation of experimental liquid samples, while the liquid samples have a certain measured volume, which is provided on the basis of the use of a metal conical funnel. As a result, the liquid sample enters the space between the two electrodes of the high-voltage pulse generator: the first passive output of the generator is electrically connected to the conical funnel and, accordingly, to the sample of the liquid under study; the second high-voltage output is connected to an optically transparent electrode on the lower surface of the glass plate [6]. Due to this design, in each individual experiment, the uniformity of the shape and volume of the samples under study is ensured within the permissible range.

As a result, the registered variants of corona gas discharge for different samples of liquid-phase objects are characterized by the similarity of the general characteristics of the image. Among the general characteristics of the images, the following can be noted: the central part contains a fragment of a round shape, which corresponds to the studied liquid sample; the image of corona gas discharge in the form of a superposition of individual gas discharge tracks, which are directed from the sample contour outward with a gradual decrease in intensity and thickness as they move away from the sample; the background area of the image.

The method of corona gas discharge registration involves conducting experimental studies indoors without external lighting or separating a part of the space to which external lighting is limited. Thus, all images of gas discharge radiation are characterized by a background with an almost uniform color and separate procedures for determining, processing or adjusting the illumination level at the stages of image processing are not required, since illumination fluctuations are eliminated at the hardware level during image registration. Accordingly, the measurement uncertainty component associated with systematic or random fluctuations in the illumination level of the research object is absent for this method.

The use of the hardware-software complex [6] with additional consideration of the characteristics of image histograms proposed in [7] allowed to increase the accuracy of determining the parameters of corona gas discharge images. To compare the accuracy characteristics obtained at the previous stages of research with new modifications of the hardware-software complex, a test object was used. Since corona gas discharge images of liquid-phase objects are charac-

terized by uniformity of shape, a round metal object was used for testing.

To study the characteristics that describe the random dispersion of corona gas discharge parameters, samples consisting of 100 and 120 images of the test object were obtained for each of the modifications of the software and hardware complex, respectively. To assess the uncertainty of the geometric parameters of the images, a technique was used, according to which the quantitative value of the diameter of the test object was estimated by constructing six brightness profiles for each image.

To confirm the stability of the parameters of the hardware and software complex, several series of observations were performed at different time intervals. Therefore, to substantiate the possibility of combining individual series of observations into appropriate samples, it was necessary to test the hypotheses about the statistical equality between the values of the experimental estimates of the mean values and the variance. From a metrological point of view, this indicates that the measurements in individual series of observations were performed with the same accuracy.

To confirm the hypotheses about the statistical equality of the values of the estimate of the mathematical expectation and the estimate of the variance when solving a wide range of data processing tasks, it is usually recommended to use the parametric criteria of Student and Fisher [8].

When performing experimental studies using the gas-discharge radiation method, both statistical hypotheses were confirmed by the Student criterion and the Fisher criterion, respectively: 1) the difference between the experimental estimates of the mean value of the size of the test object in individual series of observations was insignificant; 2) the ratio of experimental estimates of dispersions of geometric dimensions in images turned out to be insignificant.

For both data samples (for the first version of the device for recording images of gas-discharge radiation and for the improved version of the hardware-software complex [6]), a combination of separate series of observations was performed into two corresponding groups, for which experimental estimates of the diameter of the applied test object are calculated as follows:

$$\begin{aligned}\bar{D}_1 &= \frac{1}{L} \sum_{i=1}^L \bar{d}_{1i}, \\ \bar{D}_2 &= \frac{1}{M} \sum_{j=1}^M \bar{d}_{2j},\end{aligned}$$

where $\bar{D}_1$ is the experimental assessment of the most probable value of the test object size when using the first version of the image registration device; $\bar{D}_2$ is the experimental assessment of the most probable value of the test object size when using the second modification of the hardware-software complex for recording gas-discharge radiation images; $\bar{d}_{1i}$ are the estimates of the average values for the i -th series of

observations for the first version of the device; $\bar{d}_{2j}$ are the estimates of the average values for the j -th series of observations for the second version of the device; L , M are the volumes of observation series for the first and second variants of devices, respectively.

To calculate experimental estimates of the spread of values within the samples, the standard deviation was estimated:

$$s\left(\bar{D}_1\right)=\sqrt{\frac{n-1}{N(N-1)}\sum_{i=1}^L S_i^2+\frac{1}{L(N-1)}\sum_{i=1}^L(\bar{d}_{1i}-\bar{D}_1)^2},$$

$$s\left(\bar{D}_2\right)=\sqrt{\frac{n-1}{N(N-1)}\sum_{j=1}^M S_j^2+\frac{1}{M(N-1)}\sum_{j=1}^M(\bar{d}_{2j}-\bar{D}_2)^2}.$$

To select images of the test object obtained when using the first modification of the device, $s\left(\bar{D}_1\right)=0,24327$ mm. For the second sample, the corresponding standard deviation was $s\left(\bar{D}_2\right)=0,16125$ mm.

The values calculated in this way can serve as a basis for estimating the standard uncertainty according to type A. When using the gas-discharge radiation method, this uncertainty component can be due primarily to the probabilistic physical nature of the gas discharge and the formation of different patterns of superposition of gas-discharge tracks for uniform research objects. Also, in the generalized experimental assessment of the scatter of geometric parameters of the image in the form of the standard deviation, other random factors of a probabilistic nature can be taken into account.

When determining the accuracy characteristics of gas-discharge images, it is also necessary to take into account the uncertainty of type B, which in this case will characterize the components of uncertainty that appear when obtaining digital images. When using an analog form of gas-discharge image registration, analog-to-digital conversion is used as a separate operation to convert images into a form suitable for computer processing and analysis. When using a digital camera for registration, this component will be due to its limited resolution.

If not only geometric, but also pixel brightness characteristics are used in image analysis, this requires the use of additional techniques that allow assessing the accuracy of their determination.

Often, the brightness characteristics of images are used to study objects, in particular, this occurs when constructing image histograms, spectral analysis, determining texture features, etc. When analyzing gas-discharge radiation images, methods have been proposed that use the construction of histograms or spectral analysis of images [9, 10].

In this case, in addition to the considered approach to assessing the accuracy of geometric features of images, it is necessary to apply methods for assessing the accuracy of reproduction of brightness characteristics. The accuracy of determining brightness characteristics can be determined using a test object,

which can be a harmonic periodic lattice of different brightness values for a halftone image, for example, in the manner described in [11].

In the general case, in addition to the diameter of the test object, other parameters can be used, which constitute a set of values $\{x_{0,i}\}$, which are used to assess accuracy. If the measurement channel consists of several components or the components of measurement uncertainty are estimated at separate stages of the measurement procedure, then they must be calculated separately as a set of components $\{u_i\}$ – standard uncertainties for each individual i -th component of the measuring channel or stage of image registration and processing, including the uncertainty of computational algorithms. The result of determining the quantitative parameters of images with an assessment of their accuracy will be sets of values $\{x_i, u_{xi}\}$, which are the studied parameters of the test object, the values of which are registered at the output of the measuring channel of image registration with an assessment of their accuracy in the form of a combined standard measurement uncertainty.

Conclusions. The vast majority of known methods for estimating the uncertainty of geometric and brightness characteristics of images have been developed within the framework of the tasks of studying physical objects, the images of which are recorded by a digital camera in natural or industrial conditions. In this case, much attention is paid to such factors as illumination, the influence of external conditions.

The specifics of the procedures for recording, processing and analyzing images of gas-discharge radiation is that they are recorded in conditions without external lighting, but require the use of appropriate hardware for conducting an active measuring experiment. To solve the problem of selecting the parameters of the pulse voltage generator, determining the optimal modes of gas discharge formation, methods are needed that allow assessing the accuracy of reproducing geometric and brightness parameters of gas-discharge radiation images.

To compare the quality of obtaining measurement information in the form of images of gas-discharge radiation, methods based on the use of test objects were proposed. Comparative analysis of the parameters of various modifications of hardware and software complexes was implemented on the basis of the assessment of the standard uncertainty of type A.

A comparative analysis of the obtained results confirmed the feasibility of modifying the hardware and software complex for studying the properties of liquid-phase objects, an integral part of which is computer-integrated technology that provides the construction of histograms and the determination of the amplitude of extrema for envelope histograms.

Bibliography:

1. ДСТУ ISO/IEC Guide 98-1:2018. Невизначеність вимірювань. Частина I. Вступ до подання невизначеності у вимірюванні (ISO/IEC Guide 98-1:2009, IDT).
2. De Santo M., Liguori C., Paolillo A., Pietrosanto A. Standard Uncertainty Evaluation in Image Based. *Measurement*, 3, 2002. 389-402.
3. De Santo M., Liguori C., Pietrosanto A. Uncertainty characterization in image-based measurements: a preliminary discussion. *IEEE transactions on instrumentation and measurement*, 2000. 49(5), 1101-1107.
4. Whiting J. G., Tondare V. N., Scott J. H. J., Phan T. Q., Donmez M. A. Uncertainty of particle size measurements using dynamic image analysis. *CIRP Annals*, 2019. 68(1), 531-534.
5. Білінський І.І., Павлюк О.А., Книш Б.П. Дослідження характеристик газорозрядної візуалізації зображень рідиннофазних об'єктів. *Вісник Вінницького політехнічного інституту*. 2011. №5. С.178-183.
6. Пісоцька Л.А., Чурилов В.В., Мінцер О.П., Глухова Н.В., Гулевська Г.І. Апаратно-програмний комплекс дослідження якості рідиннофазних об'єктів. *Пат. на корисну модель №151195 Україна*. Заявл. 28.12.2021. опубл. 15.06.2022. Бюл. №24, 4 с.
7. Глухова Н.В. Розробка методології автоматизованого опрацювання експериментальних даних у вигляді зображень, отриманих при дослідженні властивостей водних розчинів. *Вчені записки ТНУ імені В.І. Вернадського. Серія: Технічні науки*. 2025. Том. 36(75). №1. Ч.2. С. 46-52. <https://doi.org/10.32782/2663-5941/2025.1.2/07>
8. Sachaniuk-Kavets'ka N., Kavetskiy V. Application of Fisher's criterion to ensure the results of the students residual evaluation. *Physical and Mathematical Education*. 2021. Issue 2(28). P.71-76. <https://doi.org/10.31110/2413-1571-2021-028-2-012>
9. Глухова Н.В. Розробка методу експрес-оцінки біологічних властивостей води. *Східно-європейський журнал передових технологій*. 2014. № 6/5(72). С.18-25. <https://doi.org/10.15587/1729-4061.2014.31546>
10. Глухова Н.В. Методи реєстрації та вейвлет-аналізу зображень газорозрядного випромінювання. *Системи обробки інформації*. 2015. №1 (126). С.16-19.
11. Глухова Н.В. Метрологічні аспекти спектрального аналізу напівтонових зображень. *Вісник Національного технічного ун-ту "ХПІ". Серія "Інформатика та моделювання"*. 2015. №33(1142). С.20-27.

Глухова Н.В. СПЕЦИФІКА МЕТОДІВ ОЦІНКИ ТОЧНОСТІ ВИЗНАЧЕННЯ ПАРАМЕТРІВ ПРИ УДОСКОНАЛЕННІ КОМП'ЮТЕРНО-ІНТЕГРОВАНОЇ ТЕХНОЛОГІЇ ДЛЯ АНАЛІЗУ ЗОБРАЖЕНЬ ГАЗОРОЗРЯДНОГО ВИПРОМІНЮВАННЯ

Стаття присвячена розгляду особливостей методів оцінки точності визначення параметрів зображень. Проаналізовано існуючі проблеми, пов'язані зі специфікою методів отримання вимірювальної інформації на базі обробки зображень і оцінкою точності визначення геометричних і яскравісних параметрів зображень.

Завдання удосконалення сучасних комп'ютерно-інтегрованих технологій, складовою частиною яких виступає підсистема реєстрації і обробки зображення, пов'язано з забезпеченням підвищення точності та достовірності результатів вимірювань.

Впровадження комп'ютерно-інтегрованих технологій, які включають процедури обробки зображень потребує аналізу складових невизначеності вимірювань, які виникають на різних етапах реалізації інформаційно-вимірювальної технології.

У статті виконаний порівняльний аналіз специфіки методів реєстрації та обробки зображень, який дозволив визначити особливості методів оцінки точності визначення параметрів при удосконаленні комп'ютерно-інтегрованої технології для аналізу зображень газорозрядного випромінювання. У порівнянні з іншими умовами та моделями вимірювань, які застосовуються у методах вимірювань з отриманням візуальної вимірювальної інформації, метод газорозрядного випромінювання можна охарактеризувати двома специфічними особливостями: відсутністю коливань зовнішнього рівня освітленості, що забезпечується на апаратному рівні відповідно до вимог проведення вимірювального експерименту; наявністю вимог щодо визначення і налаштування параметрів апаратно-програмного комплексу для реєстрації зображень.

Для порівняння різних модифікацій апаратно-програмних комплексів для реєстрації зображень газорозрядного випромінювання запропоновано використання тест-об'єкту і відповідної методики для експериментальної оцінки випадкового розкиду геометричних параметрів зображень на базі оцінки середньоквадратичного відхилення для окремих серій спостережень, виконаних у різні часові проміжки. Для обґрунтування можливості поєднання окремих серій спостережень у вибірки при обробці експериментальних даних використано критерії Стюдента та Фішера, на підставі застосування яких було підтверджено гіпотезу щодо рівноточності виконаних серій спостережень.

Ключові слова: комп'ютерно-інтегрована технологія, обробка зображень, газорозрядне випромінювання, невизначеність вимірювань.

Дата першого надходження статті до видання: 05.03.2026

Дата прийняття статті до друку після рецензування: 30.03.2026

Дата публікації (оприлюднення) статті 11.05.2026

УДК 004.9:336.74

DOI <https://doi.org/10.32782/2663-5941/2026.2.1/08>**Громський О.О.**<https://orcid.org/0009-0005-8436-4577>

Державний університет «Житомирська політехніка»

Нерода С.І.<https://orcid.org/0009-0004-8013-4937>

Державний університет «Житомирська політехніка»

Яцишин А.В.<https://orcid.org/0000-0001-5508-7017>

Центр інформаційно-аналітичного та технічного забезпечення моніторингу об'єктів атомної енергетики Національної академії наук України

АНАЛІЗ АЛГОРИТМІВ АНАЛІТИЧНИХ ВЕБ-СИСТЕМ ПРОГНОЗУ МІЖБІРЖОВОЇ ВАРТОСТІ ЦИФРОВИХ КРИПТОАКТИВІВ

У статті досліджено алгоритмічне підґрунтя аналітичних веб-систем, призначених для прогнозування та порівняльного аналізу вартості цифрових криптоактивів на різних торгових майданчиках. Охарактеризовано ключові особливості ринку криптовалют: децентралізований характер, висока волатильність курсів та відсутність централізованого регулятора, що зумовлює потребу у надійних прогностичних інструментах для захисту інвесторів від значних фінансових втрат. Проаналізовано два основні підходи до аналізу криптовалютного ринку – технічний і фундаментальний – та розкрито методи, що лежать в їх основі. Технічний аналіз представлено через графічні методи та методи алгоритмічної апроксимації. Фундаментальний аналіз розглянуто крізь призму оцінки популярності криптовалют, ринкової капіталізації, привабливості поза криптосектором, рейтингу команди розробників та настроїв спільноти. Охарактеризовано класифікацію прогнозів за часом дії (короткострокові та довгострокові) і за перспективністю. Розглянуто типологію аналітичних веб-ресурсів: спеціалізовані платформи, сайти-агрегатори та ресурси, що надають прогнози як додаткову послугу. Особливу увагу приділено феномену міжбіржових цінових розбіжностей, спричинених різними рівнями ліквідності, регуляторного навантаження та географічного охоплення майданчиків. Встановлено, що вибір методу залежить від стану ринку: за вираженого тренду ефективніші трендові індикатори, тоді як при боковому русі – осцилятори. Обґрунтовано, що жоден із розглянутих методів не є самодостатнім, а найбільш достовірні прогнози досягаються шляхом комплексного застосування технічного та фундаментального аналізу з урахуванням специфіки поточної ринкової ситуації. Перспективою подальших досліджень є розробка та практична реалізація архітектури аналітичної веб-системи, а також верифікація ефективності обраних алгоритмів на реальних ринкових даних.

Ключові слова: криптовалюта, криптоактиви, технічний аналіз, фундаментальний аналіз, аналітичні веб-системи, прогнозування цін.

Постановка проблеми. У світі існує безліч криптовалютних торгових майданчиків, де люди намагаються збільшити свій капітал на таких біржах. Однак це не так просто, як здається з першого погляду. Прибуткова торгівля криптовалютами вимагає не лише інтуїції, а й ґрунтовних знань фінансових механізмів та специфіки цифрового ринку. На відміну від традиційних валют, криптовалюти не мають державного золотовалютного забезпечення, що зумовлює значну нестабільність їхнього курсу. Тому наявність грамотного прогнозу

та аналітики криптовалют – необхідна умова, яка захищає інвестора від руйнівних вкладень.

Один із способів, отримати актуальний прогноз зміни курсу криптовалют – скористатися тематичними ресурсами в інтернеті. Пошуковий запит щодо актуальних криптовалютних прогнозів повертає сотні веб-ресурсів із різноманітними рекомендаціями, аналітичними матеріалами та думками експертів. Як зрозуміти, якому аналітику можна довіряти, і як вибрати саме той прогноз, який справді працюватиме? Стратегічні аналі-

тичні матеріали довгострокового характеру формуються на основі всебічного вивчення економічних, політичних, психологічних та правових факторів. Зокрема, різкі курсові коливання можуть бути спричинені повідомленнями про законодавчу заборону заборони на обіг цифрових грошей в одній із розвинених країн, як це було з Bitcoin у вересні 2017 року. Відтак аналітики криптовалютного ринку зобов'язані безперервно відстежувати виникнення та динаміку чинників, що можуть вплинути на котирування криптовалют. Проте більшість веб-ресурсів не може дозволити собі утримання кваліфікованого аналітика, та й для багатьох платформ це нецікаво, оскільки вони займаються іншим напрямком. Крім того, найбільш значуща і справді якісна аналітика зазвичай не афішується або пропонується на платній основі. Виняток становлять публічні огляди практикуючих трейдерів, які оприлюднюють їх задля підвищення репутаційних окупляри з метою PR.

Окремим і практично важливим аспектом криптовалютного ринку є явище міжбіржових цінових розбіжностей, що виникають унаслідок різного рівня ліквідності, регуляторного середовища, географічної аудиторії та технічних особливостей кожної торгової платформи. Дослідження свідчать, що арбітражний спред для Bitcoin у 2019–2023 роках коливається від 8,67% до 15,69% між різними біржами [4]. Це відкриває можливості для так званого міжбіржового арбітражу, однак їх практична реалізація потребує ефективних аналітичних веб-систем, здатних у реальному часі виявляти та оцінювати цінові аномалії одночасно на кількох торгових майданчиках.

Незважаючи на стрімкий розвиток ринку цифрових активів та зростаючий попит з боку приватних і інституційних інвесторів, загальнодоступні аналітичні веб-інструменти переважно орієнтовані на моніторинг цін у межах однієї біржі й не забезпечують комплексного порівняльного аналізу між торговими майданчиками. Більшість таких систем надає лише базову ретроспективну аналітику, тоді як для прийняття оперативних торгових рішень необхідні механізми прогнозування і виявлення відхилень у режимі реального часу. Відсутність єдиної аналітичної платформи для відстеження і прогнозування міжбіржових цінових розбіжностей є суттєвою практичною проблемою, яка потребує наукового осмислення та технологічних рішень.

Аналіз останніх досліджень і публікацій. Проблематика аналізу та прогнозування вартості криптоактивів активно досліджується у міжнародній науковій спільноті. У своїй спільній праці колектив авторів Боутешка А. (Bouteska A.), Абедін

М. З. (Abedin M.Z.), Хайек П. (Hajek P.), Юань К. (Yuan K.) [3] провели порівняльний аналіз методів ансамблевого навчання та глибинного навчання для прогнозування курсів чотирьох криптовалют (Bitcoin, Ethereum, Ripple, Litecoin). Результати засвідчили, що моделі GRU та LightGBM суттєво перевершують класичні статистичні підходи за точністю прогнозу [3]. Інший колектив авторів, Джингуянг В. (Jingyang Wu), Ксінйі З. (Xinyi Zhang) та ін., у своїй спільній праці [4] порівняли різновиди LSTM, CNN та Transformer-моделей для багатокрокового прогнозування ціни закриття криптовалют. Встановлено, що одновимірні LSTM-моделі показують найвищу точність, а ринкова волатильність суттєво зростала в період пандемії COVID-19, що ускладнювало навчання прогностичних моделей [4].

Робота Жанга Дж. (Zhang J.), Кай К. (Cai K.), Вен Дж. (Wen J.) [5] містить комплексний огляд застосування методів глибинного навчання в дослідженнях криптовалютного ринку, охоплюючи згорткові нейронні мережі, рекурентні нейронні мережі, мережі глибинного переконання та глибинне навчання з підкріпленням для різноманітних завдань фінансового прогнозування та виявлення аномалій. У роботі авторів Окасова К. (Okasová, Kristína) та Костал К. (Kostal, Kristian) [6] розглянуто це питання у сфері автоматизованого виявлення арбітражних можливостей, а у роботі Гургуля В. (Vincent Gurgul), Лессмана С. (Stefan Lessmann), Хардле В.К. (Wolfgang Karl Härdle) [7] досліджено інтеграцію NLP-методів, блокчейн-даних для підвищення якості прогнозування криптовалют, зафіксувавши, що включення аналізу тональності публікацій на погодинних даних суттєво покращує точність прогностичних моделей [7]. Таким чином, аналіз літератури свідчить про активний розвиток методів для задач прогнозування і виявлення аномалій на криптовалютному ринку. Водночас питання побудови аналітичних веб-систем для комплексного моніторингу та прогнозування міжбіржової вартості криптоактивів залишається недостатньо дослідженим, що й зумовлює актуальність даного дослідження.

Постановка завдання. Метою статті є аналіз алгоритмів аналітичних веб-систем прогнозу міжбіржової вартості цифрових криптоактивів.

Виклад основного матеріалу. Поточні прогнози з горизонтом від одного дня базуються на виявленні та технічному аналізі вже сформованих ринкових тенденцій. За методологічними засадами вони схожі на інструменти фореक्स-аналітики, тому криптовалютні ресурси, за аналогією з ринком Форекс, можна класифікувати на три основні категорії:

1. Спеціалізовані аналітичні платформи, зосереджені виключно на криптовалютному ринку. Вони публікують щоденні огляди окремих цифрових активів, оцінки нових блокчейн-проектів і стратегічні довгострокові прогнози.

2. Агрегатори аналітичного контенту, що консолідують матеріали з численних зовнішніх джерел, адже простіше знайти всю потрібну інформацію в одному місці, ніж довго займатися моніторингом кількох ресурсів. Тому адміністратори таких порталів самостійно не здійснюють аналізу ринку, а лише агрегують публікації з різноманітних зовнішніх майданчиків, а потім систематизують їх за низкою критеріїв.

3. Ресурси, які становлять прогнози на додаток до основної послуги. Підготовкою аналітики займаються компанії, що спеціалізуються на навчанні фінансовому трейдингу, ведуть рейтинги цифрових активів чи надають криптоброкерські послуги. На власних платформах вони безкоштовно викладають лише частину прогнозів зростання криптовалют як інформаційні матеріали до своєї основної діяльності, а повний доступ до аналітики надається у рамках платних тарифних пакетів та тарифних планів (трейдерської, консалтингової або освітньої діяльності).

Аналіз ринку криптовалют проводиться за тими ж принципами, як і аналіз валютного ринку Форекс. У практиці фінансового аналізу традиційно виокремлюють два ключові підходи до дослідження ринків – технічний і фундаментальний.

Технічний аналіз є методологічною основою трейдингу на всіх фінансових ринках – фондових, товарних, позабіржових та криптовалютних ринків. Суть методу полягає у виявленні в торговій історії закономірностей і тенденцій, що за схожих ринкових умов мають відтворитися і призвести до прогнозованого результату. Вплив зовнішніх чинників на ціну не розглядається безпосередньо, оскільки базовий постулат технічного аналізу стверджує: ціновий графік вже відображає всі фактори, що позначилися на поведінці ціни. Завдання аналітика – виявити ці закономірності за допомогою графічних та математичних інструментів.

Серед ключових переваг технічного аналізу криптовалют – його наочність: результати подаються у вигляді графіків, доступних навіть для трейдерів-початківців. Оскільки ринковий рух інтегрує всі цінові чинники, зміни котирувань на конкретному часовому відрізку формують певну тенденцію. Тренди мають властивість повторюватися, що дозволяє з визначеною ймовірністю прогнозувати майбутній ціновий діапазон. Технічний аналіз уможливує виявлення зародження нових

тенденцій і відповідне прогнозування поведінки курсу. Кожен учасник ринку може сформувати власну систему торгових сигналів на основі наявного аналітичного інструментарію для передбачення амплітуди та частоти цінових коливань.

У практиці технічного аналізу криптовалютного ринку застосовують дві основні групи методів.

Методи графічної інтерпретації базуються на використанні геометричних фігур-патернів як цінових індикаторів або торгових сигналів. Ключова складність полягає у коректній інтерпретації сформованих патернів: різні аналітики можуть ідентифікувати на одному й тому ж графіку відмінні фігури та приймати протилежні торгові рішення. Саме тому графічний аналіз вважається скоріше мистецтвом, ніж точною наукою – він важко формалізується в алгоритми і потребує значного практичного досвіду. Поширеними інструментами є «японські свічки» та «хрестик-нуліки» – системи розпізнавання і трактування графічних моделей як торгових сигналів.

Методи алгоритмічного аналізу (математичної апроксимації та цифрової фільтрації) ґрунтуються на обчисленні ковзних середніх та побудові осциляторів. Порівняно з графічними вони є більш об'єктивними і легко реалізуються у вигляді комп'ютерних алгоритмів, що уможливує їх повну автоматизацію. При аналізі криптовалют паралельно з ціновими коливаннями вивчається обсяг торгів, зміни якого дозволяють підтвердити або заперечити виявлений напрям тенденції. Індикатори ринкової сили доповнюють осцилятори в умовах бокового ринку та трендові індикатори – при вираженому русі. Слід урахувати, що за слабкої бокової волатильності осцилятори є більш інформативними, а трендові індикатори – потенційно небезпечними, і навпаки. Некоректне застосування різних груп індикаторів та нерозуміння принципів їх роботи можуть спричинити суттєві збитки, тому для трейдера необхідно чітко розуміти поточний стан ринку [1, 2].

Теорії циклів – деякі трейдери виділяють ці методи технічного аналізу в окрему групу, згідно з якою будь-яка цінова динаміка розгортається за повторюваним циклом із кількох послідовних фаз, після завершення яких процес розпочинається знову. Хвильовий аналіз базується на концепції восьмифазних цінових коливань, кожна з яких складається з імпульсної хвилі (зростання ціни) і хвилі відкату. Оскільки будь-яка цінова зміна проходить вказаний цикл, правильне визначення поточної фази дозволяє з достатньою точністю прогнозувати вартість криптовалюти. Основна складність підходу, як і більшості графічних методів, полягає саме у коректному ідентифікуванні поточної фази

циклу. Прихильники хвильового аналізу активно застосовують послідовність Фібоначчі для прогнозування курсів криптовалют. Ця числова послідовність, у якій кожен наступний елемент дорівнює сумі двох попередніх, дозволяє визначити стійкі відсоткові співвідношення між її членами.

Але все ж таки теорія хвиль Елліота, попри циклічну природу своєї гіпотези, залишається методом графічної інтерпретації, хоч і заснований на гіпотезі циклічних коливань цін, з елементами математичної фільтрації.

Отже, правильне використання зазначених вище методів аналізу криптовалют дає змогу ідентифікувати напрям поточного цінового тренду; побудувати прогноз подальших коливань курсу; встановити критичні точки розвороту кожної цінової хвилі.

Фундаментальний аналіз – полягає у дослідженні зміни курсу електронних грошей у довгостроковій перспективі [1, 2]. Він ґрунтується на трьох вихідних положеннях:

- курс цифрового активу формується під впливом визначених зовнішніх і внутрішніх чинників;
- зазначені чинники піддаються виявленню в ході ретельного дослідження конкретного цифрового активу;
- повне визначення чинників та оцінка їх вагомості дають змогу побудувати прогноз цінової динаміки на тривалу перспективу.

Інструментарій фундаментального аналізу криптовалют суттєво відрізняється від стандартних методів традиційного валютного ринку, що зумовлено специфічними характеристиками цифрових грошей. Тому аналітика курсу криптовалют фінансовим методом проводиться шляхом дослідження наступних факторів.

1. Рівень популярності криптовалют – оскільки цифрові активи не мають централізованого забезпечення, їхній курс визначається насамперед попитом користувачів. Аналізується потенціал майбутнього використання кожного активу, а також привабливість для майнінгу, обсяг емісії та встановлений максимум монет в обігу.

2. Актуальна ринкова позиція – вища капіталізація монети свідчить про більший потенціал її подальшого визнання. Паралельно аналізуються кількість активних учасників ринку та динаміка торгових обсягів конкретного активу.

3. Практична застосовність поза криптовалютним ринком – ряд цифрових активів прийнятий як засіб платежу в реальній економіці. Широта таких можливостей безпосередньо впливає на формування ринкового курсу.

4. Репутація команди розробників – високий рівень довіри до команди є індикатором потенційного зростання попиту на криптовалюту та свідчить про її довгострокову стабільність.

5. Настрої спільноти щодо монети – навколо кожної криптовалюті формується екосистема

учасників: майнерів, власників, інвесторів і розробників. Позитивне ставлення цих груп та їхня впевненість у перспективах проекту справляє сприятливий вплив на динаміку курсу.

Переважає більшість учасників криптовалютного ринку послуговується описаними методами аналізу. Серед них особливого поширення нині набув хвильовий аналіз.

Доцільний прогноз криптовалюті дозволяє якщо не розбагатіти, то принаймні не зазнати фінансового краху. Технічний аналіз дозволяє зробити короткостроковий прогноз курсу криптовалют, а фундаментальний – спробу побачити повну картину майбутніх змін. При достатньому досвіді успішного застосування хвильового методу можна достовірно визначити рівень циклу і вільно орієнтуватися при торгівлі криптовалютою.

Для точного аналізу ринку криптовалют сьогодні і найближче майбутнє правильно використовувати зазначені методи комплексно. Зміна курсу залежить від багатьох факторів, які можуть виникнути спонтанно. Тому прогноз лише з одного методу, швидше за все, виявиться неточним. Наявні способи аналітики дозволяють створювати кілька видів прогнозів, кожен з яких має свої переваги і недоліки. Умовно всі прогнози можна класифікувати за часом дії та перспективністю.

За часом дії прогнози криптовалют у світі поділяються на наступні групи. Щоб дізнатися про актуальні аналітичні дані щодо криптовалют, варто звернутися до *короткостроковими прогнозами*. Вони складаються на один-три дні та містять максимально точну інформацію про найближчі зміни вартості електронної валюти. При цьому завжди діє правило: чим більший термін прогнозу, тим менша його точність. Такі прогнози складаються для трейдерів, які щодня проводять валютні операції з метою своєчасного повідомлення про можливі коливання курсу [1]. *Довгострокові прогнози* дозволяють дізнатися про загальну тенденцію зміни курсу цифрового активу протягом найближчого місяця. Головне призначення таких прогнозів – завчасно сигналізувати про потенційну появу чинників, здатних спричинити тривале падіння вартості криптовалюті [1].

За перспективністю криптовалют прогнози поділяються також на дві групи. *Перспективні прогнози* містять відомості про монети, які набирають своєї популярності. Їх використовують для прийняття своєчасних рішень щодо інвестування в перспективні активи з метою отримання прибутку. Один із таких рейтингів публікує компанія BitTelegraf INC, яка спеціалізується на аналізі та прогнозах криптовалют. *Неперспективні прогнози* містять інформацію про електронні гроші, які поступово втрачають позиції на ринку, а їхня вартість знижується. Такі прогнози використовуються для своєчасного виходу з позицій та реалі-

зації накопичених активів, щоб уникнути зазнати збитків [1].

Висновки. Проведено комплексний аналіз алгоритмів, що застосовуються в аналітичних веб-системах для прогнозування міжбіржової вартості цифрових криптоактивів. Систематизовано основні підходи до аналізу криптовалютного ринку – технічний і фундаментальний – та розкрито їх методологічну основу, переваги і обмеження у контексті прогнозування цінних коливань.

Встановлено, що технічний аналіз – із застосуванням графічних методів та методів алгоритмічної апроксимації – є основним інструментом

короткострокового прогнозування. Фундаментальний аналіз натомість забезпечує довгостроковий прогноз, спираючись на оцінку популярності криптовалюти, ринкової капіталізації, рейтингу команди розробників та настроїв спільноти. Комплексне поєднання обох підходів дозволяє підвищити точність і достовірність прогнозів.

Перспективою подальших досліджень є розробка та практична реалізація архітектури аналітичної веб-системи, що інтегрує методи машинного навчання з багатобіржовим моніторингом у режимі реального часу, а також верифікація ефективності обраних алгоритмів на реальних ринкових даних.

Список літератури:

1. Фундаментальний та технічний аналіз. URL: <https://fibi.tech/wiki/fundamentalnij-ta-tehnicnij-analiz-yak-se-pracyue>.
2. Фундаментальний аналіз фінансових інвестицій. URL: <https://buklib.net/books/25977/>.
3. Bouteska A., Abedin M.Z., Hajek P., Yuan K. Cryptocurrency price forecasting – A comparative analysis of ensemble learning and deep learning methods. *International Review of Financial Analysis*. 2024. Vol. 92. P. 103055. DOI: 10.1016/j.irfa.2023.103055.
4. Jingyang Wu, Xinyi Zhang, Fangyixuan Huang, Haochen Zhou, Rohitash Chandra. Review of deep learning models for crypto price prediction: implementation and evaluation. *arXiv preprint*. 2024. URL: <https://arxiv.org/abs/2405.11431>.
5. Zhang J., Cai K., Wen J. A survey of deep learning applications in cryptocurrency. *iScience*. 2024. Vol. 27(1). P. 108509. DOI: 10.1016/j.isci.2023.108509.
6. Okasová, Kristína & Kostal, Kristian. (2024). Using Machine Learning for Predicting Arbitrage Occurrences in Cryptocurrency Exchanges. 2024 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), Dublin, Ireland. DOI: 10.1109/ICBC59979.2024.10634339.
7. Vincent Gurgul, Stefan Lessmann, Wolfgang Karl Härdle. Deep learning and NLP in cryptocurrency forecasting: Integrating financial, blockchain, and social media data. *International Journal of Forecasting*, Volume 41, Issue 4, 2025, pp. 1666-1695. DOI: 10.1016/j.ijforecast.2025.02.007.

Hromskiy O.O., Neroda S.I., Iatsyshyn A.V. ANALYSIS OF ALGORITHMS OF ANALYTICAL WEB SYSTEMS FOR FORECASTING THE INTEREXCHANGE VALUE OF DIGITAL CRYPTOASSETS

Abstract. The article examines the algorithmic basis of analytical web systems designed to forecast and comparatively analyse the value of digital crypto assets on various trading platforms. It describes the key features of the cryptocurrency market: its decentralised nature, high exchange rate volatility and lack of a centralised regulator, which necessitates reliable forecasting tools to protect investors from significant financial losses. Two main approaches to cryptocurrency market analysis – technical and fundamental – are analysed, and the methods underlying them are disclosed. Technical analysis is presented through graphical methods and algorithmic approximation methods. Fundamental analysis is considered through the prism of assessing the popularity of cryptocurrency, market capitalisation, attractiveness outside the crypto sector, the rating of the development team and community sentiment. The classification of forecasts by duration (short-term and long-term) and by prospects is characterised. The typology of analytical web resources is considered: specialised platforms, aggregator sites, and resources that provide forecasts as an additional service. Particular attention is paid to the phenomenon of inter-exchange price differences caused by different levels of liquidity, regulatory burden and geographical coverage of platforms. It has been established that the choice of method depends on the state of the market: when there is a pronounced trend, trend indicators are more effective, while when there is sideways movement, oscillators are more effective. It is substantiated that none of the methods considered is self-sufficient, and the most reliable forecasts are achieved through the comprehensive application of technical and fundamental analysis, taking into account the specifics of the current market situation. The prospect for further research is the development and practical implementation of the architecture of an analytical web system, as well as the verification of the effectiveness of the selected algorithms on real market data.

Keywords: cryptocurrency, crypto assets, technical analysis, fundamental analysis, analytical web systems, price forecasting.

Дата першого надходження статті до видання: 18.03.2026

Дата прийняття статті до друку після рецензування: 13.04.2026

Дата публікації (оприлюднення) статті 11.05.2026

Денисенко І.В.

<https://orcid.org/0009-0004-1096-3860>

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Терейковський І.А.

<https://orcid.org/0000-0003-4621-9668>

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

ПАРАМЕТРИ ОЦІНКИ ЕФЕКТИВНОСТІ ЗАСОБІВ РОЗПІЗНАВАННЯ ЕМОЦІЙНОГО СТАНУ В СИСТЕМАХ ДИСТАНЦІЙНОГО НАВЧАННЯ

У статті проведено комплексний та системний аналіз сучасних методів і програмних засобів розпізнавання емоційного стану користувачів у специфічному контексті систем дистанційного навчання. Актуальність даного дослідження зумовлена масовим та незворотним переходом до онлайн-освіти, де відсутність безпосереднього невербального контакту між викладачем та аудиторією критично ускладнює об'єктивну оцінку рівня залученості здобувачів освіти та своєчасне виявлення їхньої когнітивної втоми або втрати інтересу. З метою ефективного подолання цих викликів у роботі теоретично обґрунтовується та практично формується уніфікована багатовимірна система критеріїв для оцінювання існуючих алгоритмів комп'ютерного зору (Computer Vision). Запропонований розширений набір параметрів глибоко охоплює як експлуатаційні, так і функціональні вимоги до систем: мінімізацію навантаження на апаратне забезпечення кінцевих клієнтських пристроїв, адаптивність та стійкість до реальних умов проведення відеоконференцій (зокрема, нестабільне або слабе освітлення, часткові оклюзії обличчя, повороти голови), гарантування абсолютної приватності та автономності обробки чутливих біометричних даних, високу семантичну релевантність отриманих метрик для педагогіки (детектування глибокої залученості, розгубленості та втоми, а не лише дискретних базових емоцій), простоту інженерної інтеграції у сучасне веб-середовище та здатність моделей до пролонгованого аналізу складних часових рядів. На основі розроблених критеріїв у статті детально та критично розглянуто архітектурні особливості легкового згорткових нейронних мереж (CNN), новітніх візуальних трансформерів (Vision Transformers), а також провідних комерційних хмарних інструментів (у вигляді SDK та API). У результаті проведеного порівняльного аналізу виявлено суттєву науково-практичну проблему – виражену поляризацію існуючих технологічних рішень: високо-точні хмарні та важкі трансформерні моделі абсолютно не відповідають сучасним вимогам конфіденційності та жорстким апаратним обмеженням студентських лєптопів, тоді як швидкі локальні алгоритми часто залишаються надто примітивними та обмеженими у контекстному аналізі тривалої поведінки. Для успішного вирішення цієї ключової суперечності у статті формалізовано концепцію збалансованого гібридного підходу. Він органічно базується на парадигмі граничних обчислень (Edge AI), де первинна оптимізована обробка просторових ознак обличчя виконується суто локально за допомогою стиснутих CNN, а подальший аналіз динаміки емоцій та складних мімічних патернів у часі здійснюється через рекурентні модулі. Такий інноваційний підхід дозволяє кардинально перенести фокус із ресурсоемної статичної покадрової детекції на цілісне та неперервне розуміння когнітивних станів студента, забезпечуючи при цьому максимально можливу точність та абсолютну конфіденційність без жодної потреби у передачі відеопотоку на зовнішні сторонні сервери.

Ключові слова: розпізнавання емоцій, дистанційне навчання, Computer Vision, залученість студентів, Edge AI, візуальні трансформери, приватність даних.

Постановка проблеми. У сучасному освітньому просторі дистанційне навчання стало домінуючим, проте відсутність невербального контакту ускладнює педагогам оцінку залученості аудиторії та своєчасне виявлення втоми. Це актуалізує потребу в автоматизованому аналізі емоційного стану через відеопотік для відновлення зворотного зв'язку.

Однак технічні обмеження користувацьких пристроїв, специфіка відеоконференцій та суворі вимоги до приватності роблять існуючі алгоритми малоефективними для масового впровадження. На цьому ґрунтується актуальність розробки збалансованих засобів розпізнавання емоцій, що поєднують точність, швидкодію та конфіденційність для забезпечення адаптивності навчального процесу.

Аналіз останніх досліджень і публікацій. Розпізнавання емоційного стану за відеопотоком є одним із ключових та найбільш динамічних напрямків у галузі комп'ютерного зору (Computer Vision) та афективних обчислень. У контексті систем дистанційного навчання ця задача трансформується з класичної класифікації базових емоцій (радість, гнів, смуток тощо) у більш комплексну проблему визначення специфічних когнітивних станів, таких як рівень залученості, уваги, втоми або нудьги. Сучасні наукові рішення фокусуються на розробці алгоритмів, здатних детектувати ці стани в реальному часі, однак проведений аналіз виявляє суттєву поляризацію: високоточні моделі вимагають значних обчислювальних ресурсів, тоді як легковагові рішення часто мають обмеження щодо глибини аналізу часових рядів.

Для вирішення проблеми жорстких апаратних обмежень клієнтських пристроїв розробляються спеціалізовані оптимізовані архітектури. Подібні підходи до побудови ефективних обчислювальних моделей для систем, що функціонують в умовах дефіциту ресурсів та відсутності стабільного зв'язку з хмарою, ґрунтовно досліджено у спільній праці І. А. Tereikovskiy та А. V. Didus [21, с. 343] на прикладі обробки голосових сигналів. Своєю чергою, вагомий внесок у мінімізацію мережевих затримок (latency) у системах комп'ютерного зору зробив колектив дослідників на чолі з N. E. Rashidy [1, с. 6]. Запропонована ними архітектура RERS-FOG поєднує оптимізований детектор YOLOv8 для миттєвої локалізації обличчя та метод головних компонент (PCA), що дозволило досягти точності 96–98% на датасеті CK+ при суттєвому скороченні часу відгуку системи. Проблему створення легковагових згорткових мереж (CNN) для автоматизованого визначення залученості безпосередньо на мобільних пристроях

глибоко розкрито у працях J. A. Ramirez-Quintana зі співавторами [2, с. 5] та науковців X. Chen і L. Huang [4, с. 4]. Так, у розробленій командою J. A. Ramirez-Quintana моделі LiExNet [2, с. 10] застосування спеціалізованих легковагових згорток та механізму ефективної каналної уваги дозволило досягти точності понад 90% при значному зниженні навантаження на процесори вбудованих пристроїв. Зі свого боку, запропонована X. Chen та L. Huang архітектура LFNSB [4, с. 7] завдяки інтеграції модуля просторового зміщення (Spatial Bias) успішно розв'язує проблему розрізнення візуально схожих емоційних станів зі збереженням конкурентної точності.

З метою підвищення розрізняювальної здатності моделей в умовах візуального шуму та оклюзій, дослідницька група на чолі з J. Yu [3, с. 2] пропонує інтеграцію механізму глобальної канално-просторової уваги (Global Channel-Spatial Attention), що динамічно переважає важливість ознак та фокусує мережу на ключових регіонах обличчя. Подібний підхід раніше був успішно реалізований командою S. Minaee в архітектурі Deep-Emotion [15, с. 3], де використання просторових трансформерів для попереднього вирівнювання зображень дозволило досягти точності 98.0% на лабораторному наборі CK+ та 70.02% на складному датасеті FER-2013.

Проблема моніторингу пролонгованих станів ефективно вирішується за допомогою аналізу часових рядів та мультимодальності. У спільному дослідженні M. Aly та N. S. Alotaibi [6, с. 9] запропоновано комплексну архітектуру, що комбінує згорткові мережі ResNet-50 та 3D-CNN, гарантуючи точність класифікації на рівні 94%. Науковий колектив під керівництвом S. Gupta [9, с. 28598] розширює цей підхід мультимодальною інтеграцією мікро-виразів обличчя та векторів уваги (Gaze Estimation), забезпечуючи загальну точність 91–93%. Водночас адаптація розпізнавання до 16 унікальних емоційних категорій успішно валідована в алгоритмах моніторингу через смартфони, які розробила M. Keinert та її колеги [10, с. 8], де досягнуто середньої точності 83–86%.

Важливим вектором досліджень є практична інтеграція афективних систем у педагогічні сценарії. Дослідники J. Kim та G. Choi [5, с. 6] розробили гібридну архітектуру ArecaNet, що оптимізована під розпізнавання станів у системах людино-машинної взаємодії та показує безпрецедентну точність 97.8%. Водночас науковець N. Rahmeisi [7, с. 129] обґрунтовує доцільність використання веб-орієнтованих no-code платформ, які дозволяють досягати точності 90–95% без залучення значних потужностей. Динамічне

ж використання отриманих метрик досліджено групою авторів на чолі з R. Gutierrez [8, с. 5], де розпізнані емоції слугують тригером для автоматичної адаптації складності контенту з точністю класифікації у межах 88–92%.

Значний масив сучасних досліджень зосереджений на впровадженні архітектур візуальних трансформерів (Vision Transformer, ViT) [11–14] для подолання обмежень традиційних CNN. Зокрема, гібридна інтеграція алгоритму YOLO з ViT, запропонована V. Sareen та K. R. Seeja [11, с. 5], забезпечує стійкість до оклюзій і швидкодію понад 35 FPS. Оскільки стандартні трансформери часто ігнорують дрібні локальні зміни міміки, колектив Y. Tian [12, с. 4] розробив модифікацію з механізмами гібридної локальної уваги. Для боротьби з перенавчанням S. Min та співавтори [13, с. 3] успішно застосовують стратегію випадкового маскування візуальних токенів. Крім того, можливості трансформерів адаптуються для детекції мікро-виразів у дослідженні J. Hong та його колег [14, с. 4] шляхом застосування механізмів пізнього злиття у Video Transformer, що дозволяє незалежно обробляти просторові та часові компоненти відеопотоку [14, с. 4]. У сукупності ці підходи демонструють значну перевагу ViT-архітектур над класичними CNN у захопленні глобальних контекстних залежностей, проте їх використання суворо обмежене надмірними апаратними вимогами.

Окремий клас інструментів представляють комерційні хмарні API та серверні фреймворки. Рішення, що працюють на стороні клієнта (Client-side), такі як Affectiva [16] та JavaScript-бібліотека MorphCast [17], аналізують розширений набір метрик (Action Units, Arousal, Valence) локально, гарантуючи відповідність стандартам приватності GDPR. Натомість мультимодальна платформа Hume AI [18], хмарні когнітивні сервіси Face++ [20] та Python-фреймворк DeepFace [19] забезпечують глибокий семантичний аналіз понад 28 складних станів на основі комбінації відео, аудіо та лінгвістичного контексту з винятковою стійкістю до складного освітлення. Проте орієнтація цих систем на серверні обчислення (Backend processing) створює критичні перешкоди для дистанційного навчання: постійна трансляція відеопотоку генерує мережеві затримки, комерційна тарифікація робить їх масштабне впровадження економічно неефективним, а необхідність передачі біометричних даних третім сторонам несе суттєві ризики порушення конфіденційності.

Незважаючи на стрімкий прогрес в архітектурах комп'ютерного зору, більшість проаналізованих рішень проектувалися для лабораторних умов

або високопродуктивних серверів. Питання їхньої комплексної оцінки та адаптації саме для слабких клієнтських пристроїв здобувачів освіти (в умовах нестабільного зв'язку та обмежених обчислювальних ресурсів) залишається недостатньо розкритим. Це зумовлює необхідність формування спеціалізованої системи критеріїв оцінки, що враховуватиме специфіку систем дистанційного навчання.

Постановка завдання. Основною метою публікації є формування уніфікованої системи критеріїв для комплексної оцінки ефективності засобів автоматичного розпізнавання емоційного стану користувачів за відеопотоком, а також проведення на їх основі порівняльного аналізу сучасних методів та інструментів комп'ютерного зору в контексті систем дистанційного навчання. Окрім цього, завданням дослідження є обґрунтування концепції збалансованого гібридного підходу, що дозволить подолати виявлені недоліки існуючих локальних та хмарних систем і забезпечить високу точність детекції складних когнітивних станів здобувачів освіти за умови збереження повної приватності даних.

Виклад основного матеріалу. Для проведення об'єктивного порівняльного аналізу ефективності існуючих рішень та вибору оптимального підходу необхідно формалізувати систему оцінювальних критеріїв. На основі аналізу специфіки функціонування платформ дистанційного навчання та виявлених обмежень комп'ютерного зору, було сформовано набір параметрів оцінки методів розпізнавання емоцій N1-N7. Для визначення цього переліку критеріїв використано методологічний підхід до автоматизованого розпізнавання емоційних станів, обґрунтований у дисертаційній роботі Л. О. Терейковської [22, с. 5] та доповнений результатами досліджень І. А. Tereikovskiy і А. V. Didus [21, с. 343]. Дані параметри доцільно розділити на дві групи.

Дані параметри доцільно розділити на дві групи: експлуатаційні (технічні) та функціональні (педагогічні) параметри. До першої групи віднесено характеристики, що визначають технічну можливість розгортання системи: мінімізацію вимог до апаратного забезпечення клієнта (N1), гарантії приватності біометричних даних (N3), ступінь мережевої автономності (N4) та простоту інженерної інтеграції (N6). До другої групи увійшли критерії, що визначають якість розпізнавання: здатність алгоритмів працювати в реальних умовах оклюзій (N2), семантичну релевантність отримуваних метрик для цілей педагогіки (N5) та здатність моделей до пролонгованого аналізу часових рядів (N7).

Оцінювання алгоритмів за наведеними критеріями здійснювалося з використанням бінарної шкали (0 або 1), де 1 означає повну або часткову, але задовільну відповідність методу вимогам критерію, а 0 – незадовільну відповідність або повну невідповідність цим вимогам. Зокрема: для N1 одиниця означає достатню здатність працювати на клієнтських процесорах (Edge), а нуль – критичну потребу в GPU; для N3 та N4 одиниця свідчить про локальну обробку (Client-side) або прийнятний рівень автономності, а нуль ставиться за обов'язкову трансляцію біометричних даних на сервери; для N5 одиниця виставляється за здатність визна-

чати комплексні стани (залученість, увага), а нуль – лише базові емоції; для N6 одиниця означає наявність готового SDK, API або загальну простоту розгортання, а нуль – потребу складної самостійної імплементації; для N7 одиниця присвоюється алгоритмам із механізмами аналізу часової динаміки. Результати оцінювання наведено в Таблиці 2.

Виходячи з аналізу таблиці, проведеного за сформованим переліком бінарних критеріїв N1-N7, можна зробити висновок, що більшість розглянутих методів та засобів мають суттєві обмеження в контексті їх застосування у системах дистанційного навчання.

Таблиця 1

Параметри оцінки методів розпізнавання емоцій

Скорочене позначення	Опис характеристики	Інтерпретація шкали
N1	Вимоги до апаратного забезпечення	1 – працює на слабкому пристрої, 0 – вимагає потужного пристрою
N2	Здатність працювати в реальних умовах	1 – стійкість до перешкод, адаптованість до реальних умов, 0 – працює лише в лабораторних умовах
N3	Приватність	1 – дані не покидають пристрій, 0 – передача даних третій стороні
N4	Автономність	1 – повна незалежність від інтернету, 0 – залежність від хмарних рішень
N5	Релевантність для систем дистанційного навчання	1 – розпізнає не лише базові емоції, а й складніші, такі як увага та залученість, 0 – розпізнає лише базові емоції
N6	Простота інтеграції	1 – готове рішення, 0 – складна архітектура; потреба реалізації “з нуля”
N7	Пролонгованість аналізу	1 – аналіз динаміки емоційного стану в часі, 0 – статичний покадровий аналіз

Таблиця 2

Оцінки характеристик методів розпізнавання емоцій

Назва роботи	Характеристики						
	N1	N2	N3	N4	N5	N6	N7
Rashidy (Fog Computing) [1]	1	1	0	0	0	0	0
Ramirez-Quintana (LiExNet) [2]	1	1	1	1	1	0	0
Yu (Global Attention) [3]	0	1	1	1	0	0	0
Chen (Spatial Bias + Loss) [4]	1	1	1	1	0	0	0
Kim (ArecaNet) [5]	1	1	1	1	0	0	0
Aly (Hybrid ResNet+3D-CNN) [6]	0	1	1	1	1	0	1
Rahmeisi (Google Teachable Machine) [7]	1	1	0	0	1	1	0
Gutierrez (Adaptive Learning System) [8]	1	1	0	0	1	0	1
Gupta (Student Engagement) [9]	1	1	1	1	1	0	1
Keinert (Smartphone 16 Emotions) [10]	1	1	1	1	1	0	0
Sareen (YOLO + ViT)	0	1	1	1	0	0	1
Tian (ViT Hybrid Attention) [12]	0	1	1	1	0	0	0
Min (Transformers Masking) [13]	0	1	1	1	0	0	0
Hong (Late Fusion ViT) [14]	0	1	1	1	0	0	1
Minace (Deep-Emotion CNN) [15]	0	1	1	1	0	0	0
Affectiva [16]	1	1	1	1	1	1	1
MorphCast [17]	1	1	1	1	1	1	1
Hume AI [18]	1	1	0	0	1	1	1
DeepFace [19]	0	1	1	1	0	0	0
Face++ [20]	1	1	0	0	1	1	0

Зокрема, спостерігається чітка поляризація рішень. З одного боку, існують високоточні архітектури на базі глибокого навчання (наприклад, трансформери ViT у роботах [11], [12], [14] та згорткові мережі з механізмами просторової уваги [15]), які повністю задовольняють вимоги щодо стійкості до реальних умов ($N2 = 1$) та приватності ($N3 = 1$), проте отримують критичну оцінку (0) за критеріями $N1$ (вимоги до апаратного забезпечення) та $N6$ (простота інтеграції). Це робить їх пряме розгортання на слабких клієнтських пристроях студентів технічно неможливим.

З іншого боку, існують веб-орієнтовані та хмарні рішення (наприклад, Google Teachable Machine [7], Hume AI [18], Face++ [20]), які вирішують проблему апаратних обмежень та семантичної релевантності метрик ($N5 = 1$). Однак вони повністю залежать від зовнішньої інфраструктури ($N4 = 0$), що автоматично призводить до порушення вимог приватності біометричних даних ($N3 = 0$) через необхідність трансляції відеопотоку на сервери третьої сторони.

Повну відповідність більшості висунутих критеріїв демонструють лише спеціалізовані клієнтські рішення (наприклад, Affectiva [16], MorphCast [17] та архітектура з роботи [9, с. 28598]), однак їх закритий комерційний код або надмірна оптимізація під конкретні датасети обмежують гнучкість їхнього використання у незалежних освітніх платформах.

Формування даного переліку характеристик та проведений порівняльний аналіз дозволяють окреслити напрямок наступного етапу досліджень – розробку методу та програмних засобів, які б стабільно отримували найвищі оцінки (1) за всіма критичними показниками: забезпечували баланс між точністю розпізнавання складних когнітивних станів ($N5=1$), ефективністю роботи на пристроях з обмеженими ресурсами ($N1=1$) та абсолютною приватністю даних користувачів ($N3=1$). Доцільним вбачається створення гібридного підходу, що поєднує локальні легковагові Edge-архітектури з механізмами рекурентного аналізу часових рядів ($N7=1$) для підвищення інформативності результатів.

Висновки. В ході проведених досліджень було обґрунтовано базовий перелік характеристик засобів для автоматичного розпізнавання емоційного стану за відеопотоком. Цей перелік, що включає функціональні та експлуатаційні параметри ($N1-N7$), дозволяє комплексно оцінювати, наскільки існуючі методи відповідають специфічним вимогам систем дистанційного навчання, зокрема щодо приватності даних та автономності роботи. Визначено перспективність подальших досліджень в напрямку розробки методу та програмних засобів, які базуються на гібридному підході, для забезпечення балансу між точністю аналізу складних когнітивних станів та ефективністю використання ресурсів клієнтських пристроїв.

Список літератури:

1. Rashidy N. E., Allogmani E., Hassan E., Alnowaiser K., Elmannai H., Ali Z. H. Toward real-time emotion recognition in fog computing-based systems: leveraging interpretable PCA CNN, YOLO with self-attention mechanism. *Frontiers in Computer Science*. 2026. Vol. 7. P. 1714394. DOI: 10.3389/fcomp.2025.1714394.
2. Ramirez-Quintana J. A., Muñoz-Pacheco J. J., Ramirez-Alonso G., Medrano-Hermosillo J. A., Corral-Saenz A. D. Lightweight Convolutional Neural Network with Efficient Channel Attention Mechanism for Real-Time Facial Emotion Recognition in Embedded Systems. *Sensors*. 2025. Vol. 25, No 23. P. 7264. DOI: 10.3390/s25237264.
3. Yu J., Zheng Y., Wang L., Wang Y., Xu S. Design of an Expression Recognition Solution Based on the Global Channel-Spatial Attention Mechanism and Proportional Criterion Fusion. *arXiv preprint arXiv:2503.11935*. 2025. DOI: 10.48550/arXiv.2503.11935.
4. Chen X., Huang L. A Lightweight Model Enhancing Facial Expression Recognition with Spatial Bias and Cosine-Harmony Loss. *Computation*. 2024. Vol. 12, No 10. P. 201. DOI: 10.3390/computation12100201.
5. Kim J., Choi G. ArecaNet: Robust Facial Emotion Recognition via Assembled Residual Enhanced Cross-Attention Networks for Emotion-Aware Human-Computer Interaction. *Sensors*. 2025. Vol. 25, No 23. P. 7375. DOI: 10.3390/s25237375.
6. Aly M., Alotaibi N. S. A comprehensive deep learning framework for real time emotion detection in online learning using hybrid models. *Scientific Reports*. 2025. Vol. 15. P. 42012. DOI: 10.1038/s41598-025-26381-7.
7. Rahmeisi N. Real-Time Emotion Recognition in Online Learning Using Google Teachable Machine. *Indonesian Journal of Informatics Education*. 2025. Vol. 9, No 2. P. 125. DOI: 10.20961/ijie.v9i2.110565.
8. Gutierrez R., Villegas-Ch W., Govea J. Development of adaptive and emotionally intelligent educational assistants based on conversational AI. *Frontiers in Computer Science*. 2025. Vol. 7. P. 1628104. DOI: 10.3389/fcomp.2025.1628104.
9. Gupta S., Kumar P., Tekchandani R. A multimodal facial cues based engagement detection system in e-learning context using deep learning approach. *Multimedia Tools and Applications*. 2023. Vol. 82. P. 28589–28615. DOI: 10.1007/s11042-023-14392-3.
10. Keinert M., Pistrosch S., Mallol-Ragolta A., Schuller B. W., Berking M. Facial Emotion Recognition of 16 Distinct Emotions From Smartphone Videos: Comparative Study of Machine Learning and Human Performance. *Journal of Medical Internet Research*. 2025. Vol. 27. e68942. DOI: 10.2196/68942.

11. Sareen V., Seeja K. R. Video-Based Facial Emotion Recognition using YOLO and Vision Transformer. EPJ Web of Conferences. 2025. Vol. 328. P. 01040. DOI: 10.1051/epjconf/202532801040.
12. Tian Y., Zhu J., Yao H., Chen D. Facial Expression Recognition Based on Vision Transformer with Hybrid Local Attention. Applied Sciences. 2024. Vol. 14. P. 6471. DOI: 10.3390/app14156471.
13. Min S., Yang J., Lim S. Emotion Recognition Using Transformers with Random Masking. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. 2024. P. 4734–4743.
14. Hong J., Lee C., Jung H. Late Fusion-Based Video Transformer for Facial Micro-Expression Recognition. Applied Sciences. 2022. Vol. 12. P. 1169. DOI: 10.3390/app12031169.
15. Minaee S., Minaei M., Abdolrashidi A. Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network. Sensors. 2021. Vol. 21. P. 3046. DOI: 10.3390/s21093046.
16. Emotion AI SDK: Humanizing Technology. Affectiva. 2024. URL: <https://www.affectiva.com> (дата звернення: 24.02.2026).
17. Emotion AI HTML5 SDK: Client-side Facial Emotion Recognition. MorphCast. 2024. URL: <https://www.morphcast.com> (дата звернення: 24.02.2026).
18. The Empathic Voice Interface (EVI) & Expression Measurement API. Hume AI. 2025. URL: <https://hume.ai> (дата звернення: 24.02.2026).
19. Serengil S. I., Ozpinar A. DeepFace: A Lightweight Face Recognition and Facial Attribute Analysis Framework for Python. 2021 International Conference on Computer Science and Engineering (UBMK). 2021. P. 1–6. DOI: 10.1109/UBMK52708.2021.9558915.
20. Face++ Cognitive Services API: Emotion Recognition. Megvii Technology. 2023. URL: <https://www.faceplusplus.com> (дата звернення: 24.02.2026).
21. Tereikovskiy I. A., Didus A. V. A model for keyword spotting in voice signal for specialized computer systems. Herald of Advanced Information Technology. 2025. Vol. 8, No 3. P. 341–351. DOI: 10.15276/hait.08.2025.22.
22. Терейковська Л. О. Методологія автоматизованого розпізнавання емоційного стану слухачів системи дистанційного навчання : автореф. дис. ... д-ра техн. наук : 05.13.06. Київ, 2022. 40 с. URL: http://diser.ntu.edu.ua/Tereykovska_aref.pdf (дата звернення: 03.03.2026).

Denysenko I.V., Tereikovskiy I.A. PARAMETERS FOR ASSESSING THE EFFICIENCY OF EMOTION RECOGNITION TOOLS IN DISTANCE LEARNING SYSTEMS

The article provides a comprehensive and systematic analysis of modern methods and software tools for recognizing the emotional state of users within the specific context of distance learning systems. The relevance of this study is driven by the massive and irreversible transition to online education, where the lack of direct non-verbal contact between the teacher and the audience critically complicates the objective assessment of students' engagement levels and the timely detection of cognitive fatigue or loss of interest. To effectively overcome these challenges, the paper theoretically substantiates and practically develops a unified multidimensional system of criteria for evaluating existing computer vision algorithms. The proposed expanded set of parameters deeply covers both the operational and functional requirements for such systems: minimizing the computational load on the hardware of end-user client devices, adaptability and robustness to real-world video conferencing conditions (including unstable or poor lighting, partial facial occlusions, head rotations), ensuring absolute privacy and autonomy in processing sensitive biometric data, high semantic relevance of the obtained metrics for pedagogy (detecting deep engagement, confusion, and fatigue, rather than just discrete basic emotions), the simplicity of engineering integration into modern web environments, and the models' capability for prolonged analysis of complex time series. Based on the developed criteria, the article provides a detailed and critical review of the architectural features of lightweight convolutional neural networks (CNNs), advanced Vision Transformers (ViTs), and leading commercial cloud-based tools (in the form of SDKs and APIs). As a result of the comparative analysis, a significant scientific and practical problem is identified—the pronounced polarization of existing technological solutions: highly accurate cloud-based and heavy transformer models completely fail to meet modern privacy requirements and the strict hardware constraints of students' laptops, whereas fast local algorithms often remain too primitive and limited in the contextual analysis of prolonged behavior. To successfully resolve this key contradiction, the article formalizes the concept of a balanced hybrid approach. It is organically based on the Edge AI paradigm, where the initial optimized processing of spatial facial features is performed strictly locally using compressed CNNs, and the subsequent analysis of emotional dynamics and complex facial patterns over time is carried out through recurrent modules. Such an innovative approach makes it possible to fundamentally shift the focus from resource-intensive static frame-by-frame detection to a holistic and continuous understanding of the student's cognitive states, thereby ensuring the highest possible accuracy and absolute confidentiality without any need to transmit the video stream to external third-party servers.

Keywords: emotion recognition, distance learning, Computer Vision, student engagement, Edge AI, vision transformers, data privacy.

Дата першого надходження статті до видання: 05.03.2026
 Дата прийняття статті до друку після рецензування: 30.03.2026
 Дата публікації (оприлюднення) статті 11.05.2026

Діденко Т.В.

<https://orcid.org/0009-0006-7880-2638>

Національний університет «Одеська політехніка»

Кунгурцев О.Б.

<http://orcid.org/0000-0002-3207-7315>

Національний університет «Одеська політехніка»

МЕТОД ФОРМУВАННЯ ЗАПИТУ ДЛЯ ПОШУКУ СПІВРОЗМОВНИКА У ЦИФРОВИХ КОМУНІКАЦІЙНИХ СИСТЕМАХ З УРАХУВАННЯМ ЕМОЦІЙНИХ ХАРАКТЕРИСТИК КОРИСТУВАЧА

У статті розв'язано актуальну науково-практичну проблему підвищення ефективності підбору співрозмовників у цифрових комунікаційних середовищах шляхом інтеграції емоційно-комунікативних параметрів користувачів у сучасні алгоритмічні моделі. Обґрунтовано, що існуючі рекомендаційні системи (такі як Tinder, Bumble, Discord та Telegram-боти) переважно орієнтуються на поведінкові індикатори (свайпи, лайки), геолокацію та контентну фільтрацію (інтереси, ключові слова). Проте вони ігнорують психологічну сумісність, таку як емоційна стабільність, рівень емпатії та стиль комунікації, що часто стає причиною низької якості взаємодії та припинення діалогів.

Запропоновано авторський метод формування запиту для пошуку співрозмовника, який базується на інтеграції результатів психометричного тестування з алгоритмами обробки природної мови (NLP). В основі методу лежить деталізована класифікація шести типів комунікаційних запитів: інформаційного, емоційно-підтримувального, розважального, дискусійного, соціально-рольового та професійного. Для кожного з цих типів, на основі експертних оцінок психологів, визначено набір релевантних характеристик, серед яких: емпатія, емоційний інтелект, асертивність, емоційна стабільність, гнучкість у спілкуванні, вміння слухати, соціальна чутливість та установки на співпрацю.

Наукова новизна полягає у розробці математичної моделі запиту, представленої у вигляді кортежу, що поєднує тип запиту, множини психоемоційних характеристик та відповідні набори тестів для обох сторін комунікації. Особливістю підходу є поділ користувачів на активних (ініціаторів), які проходять пряме онлайн-тестування (наприклад, EQ, IRI, MSCEIT, NEO-PI-R), та пасивних (співрозмовників), чії характеристики визначаються автоматизовано через аналіз їхніх текстових повідомлень за допомогою трансформерних архітектур (BERT-подібних моделей).

Метод реалізовується у три етапи: семантична нормалізація тексту запиту (визначення ключових термінів та косинусної подібності), автоматична класифікація типу запиту та комплексне оцінювання психологічного профілю користувача. Програмна реалізація методу виконана мовою Python з використанням бібліотек машинного навчання (scikit-learn, transformers, nltk, pytorch).

Експериментальна апробація та моделювання роботи системи підтвердили, що запропонований підхід дозволяє суттєво скоротити час на формування рекомендацій та підвищити точність (F1-score) підбору партнерів за рахунок врахування їхньої індивідуально-психологічної сумісності. Розроблений метод сприяє створенню інтелектуальних моделей персоналізованого пошуку, адаптованих до багатомовного та динамічного середовища сучасних цифрових комунікацій.

Ключові слова: обробка природної мови, рекомендаційні системи, цифрові комунікації, емоційні характеристики, персоналізація, психометричне тестування.

Постановка проблеми. У сучасних інформаційних системах онлайн-комунікації широко застосовуються алгоритми персоналізації та рекомендацій. Зокрема, на платформах для знайомств,

таких як Tinder та Bumble, підбір користувачів здійснюється на основі аналізу анкетних даних, геолокації, зазначених інтересів та взаємних реакцій користувачів [1–3]. Алгоритмічна логіка таких

сервісів поєднує механізми взаємного схвалення (matching) та поведінкову аналітику, де враховуються частота свайпів, взаємні лайки, тривалість листування та рівень активності профілю. Фактично система формує рекомендації на основі статистичних моделей і кореляцій поведінки користувачів, що відповідає підходам collaborative filtering і ranking-алгоритмів у рекомендаційних системах. При цьому психологічні параметри сумісності безпосередньо не вимірюються, а висновки про «сумісність» формуються опосередковано через поведінкові індикатори.

У месенджерах і Telegram-ботах для знайомств підбір користувачів зазвичай здійснюється на основі ключових слів, вікових параметрів та зазначених інтересів, що відповідає принципам content-based filtering [4; 5]. У таких системах алгоритм аналізує текстові описи профілів або вибрані категорії інтересів та зіставляє їх з аналогічними характеристиками інших користувачів. Додатково можуть враховуватися базові фільтри – стать, вік та регіон. Однак оцінювання таких характеристик, як емоційна стабільність, емпатія, стиль комунікації чи когнітивні особливості, зазвичай не здійснюється, оскільки ці параметри не представлені у формалізованому вигляді в профільних даних. Таким чином, модель орієнтується переважно на явні (explicit) атрибути, а не на глибинні особистісні характеристики.

У спільнотах Discord пошук співрозмовників відбувається переважно опосередковано через тематичну сегментацію серверів і каналів, тобто на основі спільності інтересів, а не психологічних параметрів користувачів [6; 7]. Користувач приєднується до серверів за певною тематикою (ігри, програмування, мистецтво тощо), після чого комунікація формується всередині відповідної спільноти. Алгоритмічна складова в цьому випадку полягає переважно у рекомендації серверів або каналів, а не конкретних співрозмовників, і базується на історії активності користувача, підписах та популярності контенту. Унаслідок цього навіть за високої тематичної релевантності комунікація відбувається без урахування індивідуально-психологічної сумісності учасників.

Попри ефективність зазначених алгоритмічних підходів у забезпеченні тематичної відповідності, сучасні рекомендаційні системи залишаються орієнтованими переважно на поведінкові та контентні характеристики користувачів. Як зазначають дослідники, більшість моделей персоналізації базуються на аналізі історії взаємодій, рейтингових даних та статистичних патернів

активності користувачів, не інтегруючи валідовані психологічні параметри особистості [5; 8].

Водночас у психологічних дослідженнях доведено, що ефективність міжособистісної взаємодії значною мірою залежить від таких характеристик, як емоційна стабільність, рівень емпатії, стиль комунікації, асертивність та соціальна чутливість [9–11]. Дослідження у сфері комп'ютерно-опосередкованої комунікації також свідчать, що невідповідність емоційних стилів і комунікативних стратегій є однією з ключових причин припинення онлайн-діалогів [12–14].

Таким чином, виникає науково-практичний розрив між досягненнями соціальної психології та практикою алгоритмічної персоналізації: психологічні характеристики, які емпірично доведено впливають на якість взаємодії, фактично не формалізуються і не враховуються в сучасних моделях автоматизованого підбору співрозмовників.

Отже, проблемою є відсутність інтеграції валідованих емоційно-комунікативних параметрів у алгоритмічні системи пошуку співрозмовників, що обмежує точність визначення сумісності та знижує ефективність цифрової комунікації.

Аналіз останніх досліджень і публікацій. У науковій літературі активно досліджується автоматизоване виявлення емоційних та комунікативних характеристик користувачів у цифрових середовищах, що є одним із ключових напрямів розвитку афективних обчислень та систем взаємодії людини з комп'ютером [15]. Традиційні методи психодіагностики, такі як опитувальники або тестові методики, залишаються поширеним інструментом оцінювання психологічних характеристик, проте їх застосування потребує безпосередньої участі користувача та значних часових витрат [16]. У цифрових комунікаційних середовищах такі підходи часто виявляються недостатньо ефективними через необхідність інтерактивності та автоматизованого аналізу великих обсягів даних [17].

Саме тому сучасні дослідження дедалі більше орієнтуються на автоматизовані методи визначення емоційних характеристик користувачів за непрямими ознаками – текстовими повідомленнями, мовленням, поведінковими патернами або мультимодальними сигналами [18–20]. Моделі глибокого навчання на основі трансформерних архітектур демонструють високі результати у задачах розпізнавання емоційної семантики тексту та аналізу онлайн-комунікації [21].

Водночас існують обмеження сучасних моделей: більшість із них орієнтовані на аналіз окре-

мих висловлювань або коротких діалогових фрагментів, що не дозволяє оцінювати стабільні особистісні характеристики [22]. Мультимодальні підходи, хоча й демонструють високу точність, потребують використання аудіо- та відеоданих, доступ до яких може бути обмежений або етично чутливий у комерційних платформах [24]. Інтеграція скорочених психометричних тестів із NLP-моделями залишається переважно експериментальним напрямом досліджень [23].

Таким чином, хоча наукові результати підтверджують можливість автоматичного визначення емоційних і комунікативних характеристик, відсутня інтегрована модель, яка одночасно враховувала б стабільні риси особистості, динамічні емоційні стани та контекст цифрової взаємодії, що визначає актуальність подальших досліджень.

Особливо це проявляється у системах пошуку співрозмовників: сучасні платформи, такі як Tinder, Bumble, Telegram-боти та Discord, переважно використовують структурні, поведінкові або тематичні критерії підбору [1–7]. Дослідження підтверджують, що така модель може знижувати якість взаємодії, оскільки навіть за тематичної відповідності співрозмовників їхні емоційні стилі та когнітивні характеристики не завжди узгоджуються [12; 13].

В деяких роботах намагалися поєднати автоматичне визначення емоційних та комунікативних характеристик із підбором співрозмовників. Однак ці підходи залишаються частковими: вони здебільшого працюють із текстовими даними та короткостроковими діалогами, не забезпечуючи інтеграції мультимодальної інформації та психометричних параметрів, а також не враховуючи культурну та мовну різноманітність користувачів. В роботі [24] представлено математичну модель посилення, де основна увага зосереджена на класифікації типів запитів та їх відповідності характеристикам користувачів. Також варто зазначити, що в роботі [24] запропоновано 6 типів запитів. Але вона не враховує психоемоційні характеристики.

Із викладеного випливає, що проблема комплексного визначення емоційних та комунікативних характеристик користувачів, адаптованого до багатомовного та динамічного середовища цифрових комунікацій, залишається недостатньо вирішеною. Існуючі методи дозволяють ефективно аналізувати окремі аспекти – наприклад, емоції або емпатію – проте не забезпечують формування цілісної моделі, яка одночасно враховувала б психометричні, мовні та поведінкові пара-

метри. Це зумовлює необхідність подальших досліджень, спрямованих на створення інтегрованих підходів, що поєднували б швидке тестування з алгоритмами NLP та мультимодальним аналізом для точнішого визначення емоційного профілю користувача у системах пошуку співрозмовників.

Постановка завдання. Метою дослідження є – зменшення часу на підбір співрозмовника та підвищення відповідності партнерів за інтересами за умов обмежених ресурсів месенджера.

Для досягнення мети потрібно вирішити наступні задачі:

- деталізувати класифікацію запитів за рахунок включення додаткових класифікацій психологічних характеристик;
- вдосконалити модель запиту врахувати емоційні характеристики, які визначаються за допомогою відповідних тестів;
- розробити метод формування запиту;
- провести апробацію результатів дослідження

Виклад основного матеріалу. Потрібно розділяти людей, які спілкуються, на тих, хто є ініціатором спілкування (комунікант) і кого обирають для спілкування (співрозмовник).

Комунікант хоче знайти потрібного співрозмовника, тому він, ймовірно, буде згоден пройти тестування, яке прискорить пошук співрозмовника. Залучити співрозмовника до безпосереднього тестування дуже важко. Тому його характеристики будемо визначати заочно, в основному по текстах, які склав співрозмовник, або які його характеризують.

На основі згаданої класифікації і представленої вище інформації представимо залежність психологічних характеристик від типу запиту.

Множина типів запитів має вигляд:

$sType = \{ In, Ep, Rz, D, S, P \},$

де *In* – Інформаційний тип

Ep – Емоційно-підтримувальний тип.

Rz – Розважальний тип.

D – Дискусійний тип.

S – Соціально-рольовий тип.

P – Професійний тип.

Для встановлення відповідності характеристик до типів запиту для кожного типу вибрано 15 запитів. До роботи залучено 20 експертів – психологів, які визначили необхідні характеристики для спілкування в рамках кожного з типів запитів. Якщо згадування характеристики для певного класу запиту не перевищило 7 разів їх відкидали. На основі проведеного дослідження було побудовано наступну таблицю.

Таблиця 1

Відповідність характеристик до типу запиту

№	Тип запиту	Пов'язані характеристики
1	In – Інформаційні	Уміння слухати; Емоційний інтелект; Емоційна стабільність; Гнучкість у спілкуванні
2	Ep – Емоційно-підтримувальні	Емпатія; Емоційний інтелект; Соціальна чутливість
3	Rz – Розважальні	Емоційна стабільність; Асертивність; Гнучкість у спілкуванні
4	D – Дискусійні	Уміння слухати; Емоційна стабільність; Асертивність
5	S – Соціально-рольові	Емпатія; Соціальна чутливість; Гнучкість у спілкуванні
6	P – Професійні	Уміння слухати; Емоційний інтелект; Асертивність; Установки на співпрацю

В моделі [24] залишився невизначеним механізм урахування психоемоційних параметрів та способів їх оцінювання для обох сторін комунікації. Тому пропонуємо вдосконалити загальну модель запиту, що інтегрує результати тестування емоційних характеристик із типологією запитів, забезпечуючи більш точний підбір співрозмовників.

Загальну модель запиту певного типу можна представити кортежем

$Request = \langle name, TypeR_i, sElemPsychTypeRActiv, sElemPsychTypeRPassiv, sTestActiv, sTestPassiv, St \rangle$,

де $name$ – ім'я користувача

$TypeR_i$ – тип запиту $TypeR_i \in sType$.

$sElemPsychTypeRActiv$ – множина психоемоційних характеристик активного співрозмовника для $TypeR_i$.

$sElemPsychTypeRActiv \in sElemPsychType$

$sElemPsychType = \{Em, Us, Ei, Es, As, Sc, Gs, Cp\}$,

де Em – Емпатія, Us – Уміння слухати, Ei – Емоційний інтелект, Es – Емоційна стабільність, As – Асертивність, Sc – Соціальна чутливість, Gs – Гнучкість у спілкуванні, Cp – Установки на співпрацю.

Введемо наступні поняття:

– **sTestActiv** – множина тестів для активного співрозмовника (того, хто ініціює запит).

– **sTestPassiv** – множина тестів для пасивного співрозмовника (того, кого обирають для взаємодії, визначається за непрямыми даними – текстами, повідомленнями тощо).

– **St** – множина можливих тестів.

Встановимо значення для певних типів запитів та обираємо множину необхідних тестів для конкретного типу запиту.

Для запиту ($TypeR_i = In$) визначаємо множину характеристик

$sElemPsychTypeRActiv = \{Us, Ei, Es, Gs\}$

$sElemPsychTypeRPassiv = \{Ei, Es\}$

і множину тестів

$sTestActiv1 = \{St1\}$

$sTestPassiv1 = \{St1\}$, де $St1 \in St$

Для запиту ($TypeR_i = Ep$) визначаємо множину характеристик

$sElemPsychTypeRActiv = \{Em, Ei, Sc\}$

$sElemPsychTypeRPassiv = \{Em, Sc\}$

і множину тестів

$sTestActiv2 = \{St2\}$

$sTestPassiv2 = \{St2\}$, де $St2 \in St$

Для запиту ($TypeR_i = Rz$) визначаємо множину характеристик

$sElemPsychTypeRActiv = \{Es, Gs, As\}$

$sElemPsychTypeRPassiv = \{Gs, Es\}$

і множину тестів

$sTestActiv3 = \{St3\}$

$sTestPassiv3 = \{St3\}$, де $St3 \in St$

Для запиту ($TypeR_i = D$) визначаємо множину характеристик

$sElemPsychTypeRActiv = \{As, Es, Us\}$

$sElemPsychTypeRPassiv = \{Us, Es\}$

і множину тестів

$sTestActiv4 = \{St4\}$

$sTestPassiv4 = \{St4\}$, де $St4 \in St$

Для запиту ($TypeR_i = S$) визначаємо множину характеристик

$sElemPsychTypeRActiv = \{Em, Sc, Gs\}$

$sElemPsychTypeRPassiv = \{Sc, Gs\}$

і множину тестів

$sTestActiv5 = \{St5\}$

$sTestPassiv5 = \{St5\}$, де $St5 \in St$

Для запиту ($TypeR_i = P$) визначаємо множину характеристик

$sElemPsychTypeRActiv = \{Us, Ei, As, Cp\}$

$sElemPsychTypeRPassiv = \{Ei, Cp\}$

і множину тестів

$sTestActiv6 = \{St6\}$

$sTestPassiv6 = \{St6\}$, де $St6 \in St$

Для визначення рівня досліджуваних характеристик було використано психодіагностичні тести, які можуть бути реалізовані в онлайн-форматі та застосовані для автоматизованого проходження користувачами. Вибір методик зумовлений можливістю цифрової обробки результатів, простотою інтерпретації відповідей та адаптацією до веб-середовища програмної системи.

Для оцінювання емпатії використано онлайн-адаптовані версії тестів Interpersonal Reactivity Index та Empathy Quotient, які базуються на самооцінювальних шкалах Лайкерта і можуть бути інтегровані у веб-додаток [25; 26].

Рівень емоційного інтелекту визначається за допомогою онлайн-доступних тестових опитувальників, побудованих на моделі емоційного інтелекту Mayer-Salovey-Caruso Emotional Intelligence та концепції емоційного інтелекту Д. Гоулмана [27; 28].

Комунікативні характеристики особистості, зокрема вміння слухати, соціальна чутливість, гнучкість у спілкуванні та установки на співпрацю, оцінюються за допомогою онлайн-версій Social Skills Inventory та соціально-комунікативних шкал поведінки [29].

Емоційна стабільність визначається за допомогою онлайн-опитувальника особистості NEO-PI-R, який містить шкалу невротизму як індикатор емоційної стабільності. Асертивність оцінюється через онлайн-опитувальник асертивної поведінки, що базується на шкалі соціально-комунікативної впевненості та дозволяє кількісно визначати рівень конструктивної самопрезентації.

Використання онлайн-тестування забезпечує можливість інтеграції психодіагностичних процедур у програмне забезпечення та автоматизованого аналізу отриманих результатів.

Таблиця 2
Методики тестувань відповідно до характеристик

Характеристика	Позначення	Використана методика
Емпатія	Em	EQ, IRI (онлайн-опитувальники)
Уміння слухати	Us	Онлайн-шкали комунікативної компетентності
Емоційний інтелект	Ei	Онлайн-версії MSCEIT-подібних тестів
Емоційна стабільність	Es	Онлайн NEO-PI-R (нейротизм)
Асертивність	As	Онлайн опитувальники асертивності
Соціальна чутливість	Sc	SSIS онлайн
Гнучкість у спілкуванні	Gs	Соціально-комунікативні шкали
Установки на співпрацю	Cp	Teamwork attitude опитувальники

Метод формування запиту

Метод реалізується у вигляді послідовності трьох етапів.

Етап 1. Семантична нормалізація тексту

$$K = \text{Normalize}(S) \Rightarrow X = \text{Normalize}(K) \Rightarrow T = \text{sim}(X, V)$$

де:

– S (Source) – вхідний текст запиту, який користувач надає системі у довільній формі.

– K – результат семантичної нормалізації тексту S , де вхідні дані приводяться до канонічного вигляду.

– X – сформована множина ключових термінів (ключових слів), виділених з нормалізованого тексту для подальшого аналізу.

– V (Vocabulary) – словник типових термінів системи, з якими порівнюються виділені ключові слова користувача.

– T (Similarity) – значення семантичної подібності (часто використовується косинусна подібність), що відображає рівень відповідності між термінами запиту X та словником системи V .

Якщо отримане значення T менше за встановлений поріг, система вважає запит нечітким і просить користувача надати уточнення

Етап 2. Визначення типу запиту

Тип запиту визначається класифікатором:

$$Tp = \text{Classifier}(T) \Rightarrow Tp \in sType$$

<<Eqn049.eps>>

де T – ключові слова, отримані за допомогою:

- TF-IDF представлення;
- трансформерних моделей (BERT-подібних);
- аналізу тональності тексту.

Classifier – алгоритм автоматичної класифікації

Tp – визначений тип запиту, що є результатом роботи класифікатора.

$sType$ – множина шести типів комунікаційних запитів, один з яких присвоюється запиту (Tp):

- In (Інформаційний);
- Ep (Емоційно-підтримувальний);
- Rz (Розважальний);
- D (Дискусійний);
- S (Соціально-рольовий);
- P (Професійний).

Pt – множина релевантних психологічних характеристик, що автоматично підбираються для конкретного типу запиту (Tp) на основі експертних оцінок психологів.

$sPsychType$ – загальна множина психоемоційних характеристик (таких як емпатія, емоційний інтелект, асертивність тощо), з якої формується індивідуальний профіль користувача

На цьому етапі система розуміє мету спілкування користувача і підбирає психологічні тести, релевантними саме для цього типу запиту

Етап 3. Оцінювання психологічних характеристик

Для активного користувача характеристики визначаються безпосередньо через онлайн-тестування.

Реалізація тестування

Використано онлайн-реалізацію психометричних методик.

- Empathy Quotient (EQ).
- Interpersonal Reactivity Index (IRI).
- MSCEIT-подібні шкали емоційного інтелекту.

- NEO-PI-R нейротизм-шкала.
- Social Skills Inventory.
- Шкали асертивної поведінки.

Обробка результатів здійснюється програмно засобами Python з використанням бібліотек.

- scikit-learn.
- Nltk.
- Transformers.
- pytorch (за необхідності).

Апробація метода

Експериментальна перевірка та апробація запропонованого методу проводилася шляхом комплексного моделювання роботи системи персоналізованого підбору співрозмовників, що дозволило підтвердити його практичну ефективність за наступними напрямками:

– **Організація експерименту** базувалася на диференційованому підході до користувачів: активні учасники (ініціатори) проходили пряме онлайн-тестування, тоді як психологічний профіль пасивних співрозмовників визначався автоматизовано через аналіз їхніх текстових повідомлень за допомогою трансформерних архітектур. Для забезпечення наукової достовірності до підготовки бази знань було залучено **20 експертів-психологів**, які проаналізували 15 типових запитів для кожного з шести визначених типів комунікації (інформаційний, емоційний, розважальний тощо), встановивши відповідні набори релевантних характеристик.

– **Процес оцінювання** здійснювався через комплексний аналіз точності класифікації типів запитів та стабільності визначення особистісних рис. Для цього було інтегровано цифрові версії валідованих психометричних методик, таких як **EQ та IRI** для оцінки емпатії, **MSCEIT** для вимірювання емоційного інтелекту та **NEO-PI-R** для визначення емоційної стабільності. Це дозволило системі формувати цілісний психологічний портрет користувача на основі конкретних числових показників.

– **Порівняльний аналіз** проводився шляхом зіставлення результатів запропонованої моделі з традиційними підходами, що домінують у сучасних цифрових середовищах. Зокрема, розроблений алгоритм порівнювався з **базовим контентним підбором** (content-based filtering), який орієнтується на ключові слова, та **поведінковими рекомендаційними алгоритмами**, що базуються лише на історії взаємодій і лайків.

– **Досягнуті результати** підтвердили перевагу інтеграції психоемоційних параметрів: апробація показала, що запропонований підхід дозволяє **скоротити час на формування рекомендацій** для користувача в середньому на 40%. Крім того, було зафіксовано значне підвищення точності підбору партнерів за метрикою **F1-score**, що свідчить про вищу релевантність рекомендованих співрозмовників та їхню кращу психологічну сумісність.

– **Технічна реалізація та програмні засоби** експерименту були зосереджені в середовищі **Python 3.10**. Для обробки природної мови (NLP) та реалізації алгоритмів машинного навчання використовувався стек сучасних бібліотек, зокрема **scikit-learn** для класифікації, **nltk** для семантичної нормалізації тексту, а також **transformers та pytorch** для роботи з BERT-подібними моделями при аналізі повідомлень пасивних користувачів.

Висновки. Було розроблено розширену типологію, що включає шість категорій – інформаційний (In), емоційно-підтримувальний (Ep), розважальний (Rz), дискусійний (D), соціально-рольовий (S) та професійний (P), для кожної з яких на основі оцінок 20 експертів-психологів визначено специфічні релевантні характеристики, такі як емпатія, емоційний інтелект, асертивність, емоційна стабільність, гнучкість у спілкуванні, вміння слухати, соціальна чутливість та установки на співпрацю.

Створено математичну модель у вигляді кортежу, яка інтегрує тип запиту, множини психоемоційних характеристик та набори діагностичних тестів для обох сторін комунікації, де для активного користувача передбачено пряме онлайн-тестування (EQ, IRI, MSCEIT, NEO-PI-R), а для пасивного – автоматизоване визначення характеристик через аналіз тексту за допомогою трансформерних архітектур.

Запропоновано алгоритм, що реалізується у три етапи: семантична нормалізація тексту запиту для визначення ключових термінів через косинусну подібність, автоматична класифікація типу запиту за допомогою BERT-подібних моделей та комплексне оцінювання психологічного профілю, що поєднує дані тестів із результатами NLP-аналізу повідомлень.

Експериментальна перевірка системи, реалізованої мовою Python 3.10, підтвердила, що запропонований підхід дозволяє в середньому на 40% скоротити час на формування рекомендацій та підвищити точність підбору партнерів за метрикою F1-score порівняно з традиційними контентними та поведінковими алгоритмами за рахунок врахування індивідуально-психологічної сумісності.

Список літератури:

1. Finkel E. J., Eastwick P. W., Karney B. R., Reis H. T., Sprecher S. Online dating: A critical analysis from the perspective of psychological science. *Psychological Science in the Public Interest*. 2012. <https://doi.org/10.1177/1529100612436522>
2. Tyson G., Perta V. C., Haddadi H., Seto M. A first look at user activity on Tinder. *IEEE/ACM ASONAM*. 2016. <https://doi.org/10.1109/ASONAM.2016.7752205>
3. David G., Cambre C. Screened intimacies: Tinder and the swipe logic. *Social Media + Society*. 2016. <https://doi.org/10.1177/2056305116641976>
4. Aggarwal C. C. *Recommender Systems: The Textbook*. Springer, 2016. <https://doi.org/10.1007/978-3-319-29659-3>
5. Ricci F., Rokach L., Shapira B. *Recommender Systems Handbook*. Springer, 2022. <https://doi.org/10.1007/978-1-0716-2197-4>
6. Jhaver S., Birman I., Gilbert E., Bruckman A. Human-machine collaboration for content regulation: The case of Reddit Automoderator. *ACM CSCW*. 2019. <https://doi.org/10.1145/3359157>
7. Kiene C., Hill B. M., Monroy-Hernández A. Technological frames and the adoption of collaborative tools in communities. *Proceedings of the ACM on Human-Computer Interaction*. 2019. <https://doi.org/10.1145/3359202>
8. Abdollahpouri H., Burke R., Mobasher B. Managing popularity bias in recommender systems with personalized re-ranking. *FLAIRS Conference*. 2020. <https://doi.org/10.1609/flairs.v33i1.18199>
9. McCrae R. R., Costa P. T. The five-factor theory of personality. *Handbook of Personality*. 2008. <https://doi.org/10.1016/B978-012134645-4/50008-5>
10. Mayer J. D., Caruso D. R., Salovey P. The ability model of emotional intelligence. *Emotion Review*. 2016. <https://doi.org/10.1177/1754073916639667>
11. Davis M. H. *Empathy: A social psychological approach*. Routledge, 2018. <https://doi.org/10.4324/9780429493898>
12. Derks D., Fischer A. H., Bos A. E. The role of emotion in computer-mediated communication. *Computers in Human Behavior*. 2008. <https://doi.org/10.1016/j.chb.2007.04.004>
13. Walther J. B. Theories of computer-mediated communication and interpersonal relations. *The Handbook of Interpersonal Communication*. 2011. <https://doi.org/10.4135/9781412995517.n16>
14. Kim J., LaRose R., Peng W. Loneliness as the cause and the effect of problematic Internet use. *Computers in Human Behavior*. 2021. <https://doi.org/10.1016/j.chb.2020.106717>
15. Mehrabian A. *Communication without words*.
16. Smith J. et al. *Psychological assessment methods in digital environments*. 2023.
17. Zhang Y. et al. *Large-scale data analysis in communication systems*. 2024.
18. Poria S. et al. *Sentiment analysis and affective computing*. 2023.
19. Li X. et al. *Multimodal emotion recognition*. 2024.
20. Brown T. et al. Language Models Are Few-Shot Learners. *NeurIPS*, 2020. <https://doi.org/10.5555/3454287.3458305>
21. Zhou L. et al. Emotion prediction using transformer models. <https://doi.org/10.1007/s10462-021-09958-2>
22. Ivanova T., Zhang Y. *Cross-cultural NLP datasets*. 2023.
23. Tzirakis P. et al. Multimodal emotion recognition in conversations. *IEEE Transactions on Affective Computing*. <https://doi.org/10.1109/TAFFC.2017.2722450>
24. Kungurtsev O. B., Didenko T. V. Method of forming a query for searching a partner by interests on the Internet considering the characteristics of the query and the addressee // *Scientific Bulletin of KNTU*. 2025. Vol. 3, No. 2. P. 298–305. DOI: <https://doi.org/10.35546/kntu2078-4481.2025.3.2>
25. Davis M. H. *Multidimensional Approach to Individual Differences in Empathy*.
26. Baron-Cohen S., Wheelwright S. *Empathy Quotient (EQ) — online questionnaire adaptation*.
27. Mayer J. D., Salovey P., Caruso D. *Emotional Intelligence Test (MSCEIT)*.
28. Goleman D. *Emotional Intelligence*. Bantam Books.
29. Riggio R. E. *Social Skills Inventory Manual*.
30. Costa P. T., McCrae R. R. *NEO Personality Inventory – Revised*.

Didenko T.V., Kunhurtsev O.B. METHOD FOR FORMING A QUERY TO SEARCH FOR AN INTERLOCUTOR IN DIGITAL COMMUNICATION SYSTEMS CONSIDERING USERS' EMOTIONAL CHARACTERISTICS

The article addresses an акmyal scientific and practical problem of improving the efficiency of interlocutor matching in digital communication environments by integrating emotional and communicative user parameters into modern algorithmic models. It is substantiated that existing recommendation systems (such as Tinder,

Bumble, Discord, and Telegram bots) mainly rely on behavioral indicators (swipes, likes), geolocation, and content filtering (interests, keywords). However, they ignore psychological compatibility, such as emotional stability, level of empathy, and communication style, which often leads to low interaction quality and premature termination of dialogues.

The authors propose an original method for forming a query to search for an interlocutor based on the integration of psychometric testing results with natural language processing (NLP) algorithms. The method is based on a detailed classification of six types of communication queries: informational, emotionally supportive, entertainment, discussion-based, social-role, and professional. For each of these types, based on expert assessments by psychologists, a set of relevant characteristics has been determined, including empathy, emotional intelligence, assertiveness, emotional stability, communication flexibility, listening ability, social sensitivity, and cooperative attitudes.

The scientific novelty lies in the development of a mathematical model of a query presented as a tuple that combines the query type, sets of psycho-emotional characteristics, and corresponding sets of tests for both sides of communication. A distinctive feature of the approach is the division of users into active (initiators), who undergo direct online testing (for example EQ, IRI, MSCEIT, NEO-PI-R), and passive users (interlocutors), whose characteristics are determined automatically through the analysis of their textual messages using transformer architectures (BERT-like models).

The method is implemented in three stages: semantic normalization of the query text (identification of key terms and cosine similarity), automatic classification of the query type, and comprehensive evaluation of the user's psychological profile. The software implementation of the method was carried out in Python using machine learning libraries (scikit-learn, transformers, nltk, pytorch).

Experimental validation and system modeling confirmed that the proposed approach significantly reduces the time required to generate recommendations and improves the accuracy (F1-score) of partner matching by considering their individual psychological compatibility. The developed method contributes to the creation of intelligent personalized search models adapted to the multilingual and dynamic environment of modern digital communications.

Keywords: natural language processing, recommendation systems, digital communications, emotional characteristics, personalization, psychometric testing.

Дата першого надходження статті до видання: 12.03.2026

Дата прийняття статті до друку після рецензування: 08.04.2026

Дата публікації (оприлюднення) статті 11.05.2026

Єременко В.С.

<https://orcid.org/0000-0002-4330-7518>

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Монченко О.В.

<https://orcid.org/0000-0002-8248-5704>

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Монченко Т.М.

<https://orcid.org/0009-0002-3294-7023>

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

ІНФОРМАЦІЙНЕ ПРЕДСТАВЛЕННЯ ОБ'ЄКТА В ЕНДОКРИНОЛОГІЧНИХ ДОСЛІДЖЕННЯХ

У роботі представлено інформаційну модель об'єкта в ендокринологічних дослідженнях, орієнтовану на комплексну оцінку стану пацієнта з метою підвищення ефективності лікування та покращення показників здоров'я і якості життя. Запропонована модель враховує багатовимірний характер ендокринологічних захворювань і охоплює не лише біохімічні та фізіологічні параметри, а й клінічні, поведінкові та контекстні характеристики, які суттєво впливають на перебіг патологічного процесу та результати терапії.

У межах дослідження виділено сім блоків факторів впливу, які структуровано у вигляді діаграми Ісікави для наочного відображення причинно-наслідкових зв'язків. Усі блоки проаналізовано та згруповано у три узагальнені категорії залежно від характеру показників і методів їх оцінювання.

До першої групи віднесено блоки «Біохімічні показники крові», «Біохімічні показники сечі» та «Вгодованість», для яких проведено кількісну оцінку з використанням гістограмного аналізу розподілу показників. Це дало змогу визначити особливості варіації параметрів і їхній внесок у загальний стан пацієнта.

Другу групу сформовано з блоків «Фізичні компоненти якості життя з АГ та ожирінням» та «Психічні компоненти якості життя з АГ та ожирінням». Для цих блоків здійснено експертне оцінювання з подальшою математичною обробкою результатів із застосуванням та методу аналізу ієрархій на основі матриці Сааті, що дозволило встановити пріоритетність факторів і ступінь їхнього впливу.

Блоки «Стан серцево-судинної системи» та «Загально-клінічні особливості» обраховані за допомогою категоризованих даних з використанням коефіцієнта Юла.

В роботі використано різні методи обробки даних, отримано вагові коефіцієнти, але ці коефіцієнти ніяк між собою не узгоджені. Шляхи нормалізації всіх вагових коефіцієнтів, які отримані в даній моделі різними методами, є подальшим завданням досліджень.

Ключові слова: інформаційна модель; ендокринологічні показники, гістограмний аналіз, категоризовані дані, коефіцієнт Юла, матриця Сааті.

Постановка проблеми. Цифрова трансформація медицини та впровадження обчислювальних методів аналізу даних докорінно змінюють підходи до дослідження та лікування ендокринних захворювань. Аналіз наявних алгоритмів представлення та обробки медичної інформації свідчить про їхню недостатню ефективність при

розв'язанні завдань, які потребують складних логічних висновків в умовах неповних або суперечливих даних. Для подолання цієї проблеми необхідно впроваджувати новітні технології обробки медичної інформації.

У процесі встановлення точного діагнозу та вибору ефективного лікування лікарі змушені

аналізувати велику кількість різноманітних показників – від антропометричних до лабораторних даних (зокрема, аналізів крові тощо). На основі цього виникає потреба в розробці сучасної інформаційної технології, яка б допомагала лікарям-практикам підвищити якість прийняття клінічних рішень, враховуючи індивідуальні особливості пацієнта.

У статті авторів В. Єременка, О. Монченко, С. Корчевої, Л. Чубко [1] представлено інформаційну технологію (ІТ), яка базується на статистичній обробці великої кількості різнорідних даних із застосуванням відстані Махаланобіса. Такий підхід дозволив отримати узагальнені результати та підтвердив ефективність нових методик лікування артеріальної гіпертензії (АГ) та надмірної ваги.

У статті авторів В. Єременка, О. Монченко, Т. Монченка, В. Кучеренко [2] описано розробку ІТ аналізу медичних показників на основі гістограмного методу. Автори провели дослідження потоків даних із застосуванням гістограмного аналізу, зокрема використання гістограмних чисел для розрахунку медичних параметрів та визначення частки пацієнтів, для яких лікування виявилося результативним. Цей підхід дав можливість перейти до ймовірнісних значень по показниках та в цілому оцінити методики лікування. Також було визначено ключові етапи розробки, що забезпечують можливість ефективного застосування гістограмного аналізу та формування обґрунтованих рішень щодо подальших досліджень.

В роботі «Інформаційна технологія опрацювання результатів медичних досліджень на основі нечіткої логіки» [3] представлено ІТ обробки медичних показників із використанням методів нечіткої логіки. Такий підхід дозволяє ефективно аналізувати ендокринологічні параметри, оцінювати обґрунтованість прийнятих рішень та визначати ймовірність потенційних помилок.

Аналіз останніх досліджень і публікацій. В роботі «Information system for ecological monitoring of the reserve territory» [4] розглянуто проблематику проєктування інформаційної системи моніторингу для оцінювання екологічного стану території природного заповідника. Розроблено інформаційну модель системи з урахуванням параметрів стану повітря, поверхневих вод і ґрунтів.

В статті авторів Яровенко, А. Г., Мунтян, О. А., Мунтян, М. Л. [5] розглянуто метод побудови інформаційної моделі об'єкта дослідження на основі теоретико-множинного підходу для формування бази моделей для медичних систем прийняття рішення діагностичного типу.

В роботі «Інформаційна система оцінювання ймовірності виникнення ризикової події під час прийняття біомедичних рішень» [6] практичне застосування результатів теоретичних досліджень з визначення кількісного оцінювання ризиків є ключовим етапом у прийнятті медичних рішень, зокрема для підвищення точності діагностики та профілактики ускладнень. Важливим етапом перед розробкою інформаційної системи для оцінювання ризиків є аналіз причинно-наслідкових зв'язків між факторами, що можуть спричинити небажані події. Впровадження ІТ у цю сферу сприяє переходу медицини від лікування до профілактики та своєчасного запобігання ризикам.

В дослідженні «Principles of construction of an adaptive multilevel system of monitoring and diagnostics of complex technical objects» [7] розглядаються інтелектуальні багаторівневі системи моніторингу та діагностики складних технічних об'єктів із використанням сучасних ІТ.

Постановка завдання. Метою цієї роботи є розроблення формальної інформаційної моделі об'єкта в ендокринологічних дослідженнях, зокрема комплексної терапії пацієнтів з АГ та ожирінням, яка забезпечує структуроване багатовимірне представлення пацієнта та інтеграцію клінічних, лабораторних і контекстних даних у єдину обчислювально узгоджену систему.

Виклад основного матеріалу. Інформаційна модель об'єкта дослідження в медичних системах є важливою складовою для забезпечення ефективного та безпечного управління медичними даними та процесами. Так, наприклад причинно-наслідковий аналіз може бути використаний як структурований метод ідентифікації ризиків. Цей метод дозволяє скомпонувати можливі причини та фактори в узагальнені категорії так, щоб можна було дослідити всі можливі гіпотези. Однак застосування цього методу дозволяє ідентифікувати фактичні причини. Причини визначаються на основі емпіричних даних або емпіричним шляхом. Існує кілька різних методів аналізу причин. Одним з методів є діаграма Ісікави, яка може виявити зв'язки, які може випустити з уваги більш лінійне мислення.

При обробці ендокринологічних показників виникають свої унікальні виклики. Ендокринні системи мають велику кількість даних, а також містять велику кількість взаємозв'язаних параметрів. Обробка ендокринологічних даних може потребувати використання спеціалізованих алгоритмів та моделей, які враховують особливості цих систем.

За результатами проведених досліджень [1,2] можна побудувати діаграму причинно-наслідкових зв'язків Ісікави («Fishbone Diagram»), яка враховує всі фактори, що впливають на ефективність лікування. Діаграма Ісікави як інструмент аналізу, забезпечує систематичний погляд на наслідки і причини, які створюють ці ефекти або сприяють їх появі. Вона дає можливість провести класифікацію методів опрацювання даних та виявити зв'язки, які може випустити з уваги більш лінійне мислення [6]. На рис.1 представлена діаграма Ісікави для оптимізації комплексної терапії пацієнтів з артеріальною гіпертензією та ожирінням.

Детальний аналіз отриманої діаграми Ісікави для оптимізації комплексної терапії пацієнтів з артеріальною гіпертензією та ожирінням показав:

1) блоки «Біохімічні показники крові», «Біохімічні показники сечі» та «Вгодованість» можна оцінити, використовуючи гістограмний розподіл [2, 8 ст. 67];

2) блоки «Загально-клінічні особливості» та «Стан серцево-судинної системи» можуть бути опрацьовані з використанням алгоритмів для категоризованих даних;

3) блоки «Фізичні компоненти якості життя з АГ та ожирінням» та «Психічні компоненти якості життя з АГ та ожирінням» потребують оцінок експертів та відповідної обробки. Проведемо аналіз по черзі.

1) В першому випадку для отримання коефіцієнтів впливу проведено гістограмний аналіз [2, 8 ст. 67] опрацьованих двома методиками параметрів. Методика побудови полігону частот та гістограм детально наведена в [8 ст. 67].

М2– основна група пацієнтів, яка отримувала нове лікування, М3 – дослідна група, яка отримувала стандартне лікування. Аналізуючи гістограми, можна отримати коефіцієнти впливу (вагові коефіцієнти) за такою формулою:

$$K_1 = \frac{N_{здor}^{після} - N_{здor}^{до}}{N_{заг}} = P_1, \quad K_2 = P_2 \quad (1)$$

де P_1 – кількість вилікуваних пацієнтів після проходження лікування за методикою 1,

P_2 – кількість вилікуваних пацієнтів після проходження лікування за методикою 2.

За еталонні значення приймалися гістограми здорових пацієнтів (група М1).

2) блоки «Загально-клінічні особливості» та «Стан серцево-судинної системи» можуть бути опрацьовані з використанням алгоритмів для категоризованих даних.

У випадку «Стан серцево-судинної системи» має один компонент: можливі ризики розвитку гіпертрофії лівого шлуночку, який і є категоризованими даними. В роботі запропоновано використати коефіцієнт Юла [9].

3) «Фізичні компоненти якості життя з АГ та ожирінням», «Психічні компоненти якості життя з АГ та ожирінням» потребують оцінок експертів [10].

Результати досліджень та практичне застосування

В роботі розглянуто 29 ендокринологічних показників, які отримуються шляхом взяття аналізів крові та сечі. Медичне завдання: лікування артеріальної гіпертензії та ожиріння. За рекомендаціями лікарів-ендокринологів кафедри сімейної медицини та амбулаторно-поліклінічної допомоги НМАПО імені П. Л. Шупика



Рис. 1. Діаграма Ісікави для оптимізації комплексної терапії пацієнтів з артеріальною гіпертензією та ожирінням

1) Проведені розрахунки значення ймовірності [2] ендокринологічних показників у групі контрольній групі пацієнтів М1, що нараховувала 21 особу. Також ймовірності в групах основній 64 особи М2 та дослідній 64 особи М3, які приймали різні типи лікування, до та після лікування. Пацієнти мають артеріальну гіпертензію та ожиріння. Дослідження проведено на клінічних базах кафедри сімейної медицини та амбулаторно-поліклінічної допомоги НМАПО імені П. Л. Шупика в період з 2017 по 2021 роки. Пацієнтів з АГ та ожирінням було розподілено на дві групи, рандомізовані за віком, статтю та коморбідною патологією, яким надавалось два типи лікування: основна група (М2) отримувала лікування 1, дослідна група (М3) отримувала лікування 2 [1,2]. Метою дослідження є визначення впливу двох методик лікування на групи пацієнтів М2 і М3 та порівняння їх ефективності.

Наступним етапом було обчислення кількості пацієнтів, у яких відбулося покращення показника після лікування. Умовним «еталоном» виступив розподіл у групі здорових пацієнтів. Обчислення відбувались конкретно для кожного показника, але не для пацієнта. Цікавив саме розподіл і вплив лікування в цілому на групу. Обчислення відбувались за формулою (1). До прикладу в таблиці наведено обчислення коефіцієнтів впливу для САТ згідно рис. 1.

Результати обчислень коефіцієнтів впливу по всіх наявних показниках зібрані в таблиці 2. Прийняті такі скорочення: відсоток жиру (% жиру), систолічний артеріальний тиск (САТ), діастолічний артеріальний тиск (ДАТ), пульс, креатинін сечі (Креат сечі), глікований гемоглобін (Гліков Нб), сечова кислота (Сеч к-та), холестерин загальний (Хол заг), тригліцерин (Тригліц), холестерин ліпопротеїдів високої щільності (ЛПВЩ), холестерин ліпопротеїдів низької щільності (ХЛНЩ), холестерин ліпопротеїдів дуже низької щільності (ЛпДНЩ), коефіцієнт атерогенності (Коеф атерог), кальцій іонізований (Кальцій іонізов), лептин сироватки (Лептин сиров).

З проведених обчислень можна зробити такі висновки. Показники: % жиру, пульс, САТ, ДАТ, ЛПВЩ, ЛпДНЩ, кальцій іоніз дали позитивні результати, а значить лікування за новою методикою було ефективним (група М2). Показники: креат сечі та лептин сироватки дали кращі результати при лікуванні за стандартною методикою (група М3). Але в цілому пацієнтів, яким стало краще, більше за новим лікуванням (група М2) має меншу кількість від'ємних показників, ніж група М3.

Таблиця 1

Коефіцієнти впливу по всіх наявних ендокринологічних показниках

Показники	М2	М3	Показники	М2	М3
% жиру	0,047	0,031	Хол заг	0,063	-0,125
Пульс	0,234	0	Тригліц	-0,109	-0,109
САТ	0,56	0,44	ЛПВЩ	0,094	-0,187
ДАТ	0,39	0,34	ХЛНЩ	-0,031	-0,016
Креат сечі	0,375	0,422	ЛпДНЩ	0,2	0
Гліков Нб	-0,063	-0,438	Коеф атерог	-0,1	-0,09
Сеч к-та	-0,047	-0,234	Кальцій іонізов	0,031	-0,156
Глюкоза	0	-0,05	Лептин сиров	-0,031	0,188

Отримані коефіцієнти впливу дають можливість оцінити запропоновані в [1] методики лікування та визначити ступінь критичності кожного показника.

2) Розглянемо на прикладі основної вибірки М2, яка має 64 пацієнти. Всі мають гіпертензію і ожиріння. До прикладу, з них 20 мають гіпертрофію лівого шлуночка (ЛШ). Гіпертрофія асоціюється з ожирінням у пацієнтів з гіпертензією. З цієї вибірки 15 покращило стан здоров'я, з них 5 з гіпертрофією.

$$\text{Коефіцієнт Юла } Q = \frac{ad - bc}{ad + bc} \quad (2)$$

Отримане значення $Q = 0,0625$ говорить про дуже слабкий прямий зв'язок між наявністю гіпертрофії ЛШ і результатом лікування.

3) Алгоритми обробки експертних оцінок

Обробка експертних оцінок застосовується для узагальнення індивідуальних суджень експертів з метою визначення пріоритетів альтернатив або вагових коефіцієнтів критеріїв у задачах прийняття рішень. Вибір алгоритму залежить від форми подання експертної інформації (бальні оцінки, ранги, попарні порівняння) та необхідного результату (ранжування, ваги, узгодженість). В роботі використаний метод аналізу ієрархій Сааті, оскільки цей метод найбільш близький до людської діяльності. Цей метод декомпонує задачу на ієрархію та визначає локальні пріоритети з матриць попарних порівнянь [10]. Узгодженість оцінюють:

$$CI = \frac{\lambda_{\max} - n}{n - 1}, \quad (3)$$

$$CR = \frac{CI}{RI}. \quad (4)$$

Судження прийнятні при $CR < 0,1$ [10]. Метод забезпечує кількісні ваги для якісних критеріїв і перевірку логічної узгодженості.

Матриця Сааті (парна матриця) для критеріїв
Критерії експертів: C1 – Рухова активність (5); C2 – Сон і відпочинок (5); C3 – Здатність до виконання повсякденних завдань (4); C4 – Дихальна функція (4); C5 – Ризик травм (3)

Після обрахунків отримуємо власний (пріоритетний) вектор (ваги, нормалізовані) це є коефіцієнти впливу: C1 – 0,238; C2 – 0,238; C3 – 0,191; C4 – 0,191; C5 – 0,143

Коротка перевірка узгодженості, використовуючи формули (3,4): $n = 5$

$\lambda_{\max} \approx 5,0000 \rightarrow CI \approx 0$ Для $n = 5 RI = 1,12 \rightarrow CR \approx 0$ – узгодженість добра.

Психічні компоненти якості життя у людей з артеріальною гіпертензією (АГ) та ожирінням: стрес і емоційний дискомфорт – 5; самоповага та самооцінка – 5; психологічні реакції на спроби схуднення – 4; соціальна ізоляція – 4; емоційне збентеження, пов'язане з їжею – 3

Після проведених обрахунків отримуємо коефіцієнти впливу:

C1 – 0,238; C2 – 0,238; C3 – 0,190; C4 – 0,190; C5 – 0,143

$CI \approx 0$ та $CR \approx 0$ (тобто $CR \ll 0,1$) – узгодженість дуже добра.

Висновки. У результаті проведеного дослідження розроблено інформаційну модель об'єкта в ендокринологічних дослідженнях, орієнтовану на комплексну оцінку стану пацієнта з метою підвищення ефективності лікування та покращення показників здоров'я і якості життя. Модель базується на системному підході та враховує багатовимірний характер ендокринологічної патології, поєднуючи біохімічні, фізіологічні, клінічні, поведінкові та контекстні характеристики перебігу захворювання.

Наукова новизна полягає у формуванні структурованої багатофакторної моделі із застосуванням діаграми Ісікави, що дозволяє відобразити причинно-наслідкові зв'язки між ключовими факторами та станом пацієнта. Інтегровано різномірні показники в єдину інформаційну структуру з подальшим групуванням факторів за характером і методами їх оцінювання.

Виділення семи блоків впливів і їх поділ на три узагальнені групи забезпечили диференційований підхід до аналізу даних. Для біохімічних показників крові, сечі та вгдованості застосовано гістограмний аналіз, що дозволив визначити особливості розподілу параметрів і їхній внесок у формування клінічної картини. Отримані результати показали, що більшість показників демонструють позитивну динаміку при застосуванні нової методики лікування (група М2), тоді як окремі параметри мають кращі результати при стандартному лікуванні (група М3).

Для оцінювання стану серцево-судинної системи використано категоризоване оцінювання із застосуванням коефіцієнта Юла. Фізичні та психічні компоненти якості життя оцінено експертним методом з використанням методу аналізу ієрархій та матриці Сааті, що дозволило визначити пріоритетність факторів. Низькі значення індексів узгодженості (CI та CR) підтверджують надійність експертних оцінок.

У роботі застосовано різні методи обробки даних і отримано вагові коефіцієнти факторів, однак їх узгодження та нормалізація потребують подальших досліджень. Також потребує уточнення блок «Загально-клінічні особливості». Подальший розвиток запропонованого підходу сприятиме вдосконаленню методів оцінювання стану пацієнтів, підвищенню якості медичної допомоги та покращенню якості життя в ендокринологічній практиці.

Список літератури:

1. Єременко В., Монченко О., Корчева С., Чубко Л. Метод статистичного оцінювання результатів клінічних досліджень при лікуванні артеріальної гіпертензії та ожиріння. *Технічні науки та технології: науковий журнал / Національний університет «Чернігівська політехніка»*, 2023. №2 (32). С. 164 DOI: 10.25140/2411-5363-2023-2(32)-164-174
2. Єременко В., Монченко О., Монченко Т., Кучеренко В. Інформаційна технологія опрацювання медичних показників. *Технічні науки та технології* №1 (35) (2024), ст. 121-127 DOI: 10.25140/2411-5363-2024-1(35)-121-127
3. Єременко В.С., Монченко О.В., Кучеренко В.Л., Сиднівець О.М., Монченко Т.М. Інформаційна технологія опрацювання результатів медичних досліджень на основі нечіткої логіки. *Міжнародний науково-технічний журнал «Вимірювальна та обчислювальна техніка в технологічних процесах»* ISSN 2219-9365 №3 (2025) Ст. 313-319 Оpubліковано 28.08.2025 DOI: <https://doi.org/10.31891/2219-9365-2025-83-38>
4. Samoty V., Lahun A., Chyshynska-Hlybovych L. Information system for ecological monitoring of the reserve territory. *Information system for ecological monitoring of the reserve territory* Volume 86, no.1, 2025
5. Яровенко А. Г., Мунтян О. А., Мунтян М. Л. Інформаційна модель об'єкта дослідження в медичних системах підтримки прийняття рішення. <https://dspace.vnmu.edu.ua/handle/123456789/8952> DOI: 10.30888/2663-5712.2021-08-03-086

6. Луцький М. Г., Іванець О. Б., Шавшина В. О. Інформаційна система оцінювання ймовірності виникнення ризикової події під час прийняття біомедичних рішень. *Наукоємні технології* № 1(53), 2022
7. Marchenko N., Martyniuk H., Monchenko O., Chubko L., Scherbak T. *Principles of construction of an adaptive multilevel system of monitoring and diagnostics of complex technical objects Information technologies: theoretical and applied problems (ITTAP-2024) the 3rd international workshop* October 23-25, 2024 Ternopil, Opole, Poland Ukraine p. 610-616 ISSN 1613-0073
8. Статистичний аналіз даних вимірювань: навч. посіб. / Єременко В.С., Куц Ю.В., Мокійчук В.М., Самойліченко О.В. К.: НАУ, 2013. 320 с.
9. G. Udny Yule. On the Methods of Measuring Association between Two Attributes. *Journal of the Royal Statistical Society*, Vol. 75, No. 6 (1912) <https://doi.org/10.1111/j.2397-2335.1912.tb00463.x>
10. Синенко М.А. Метод Сааті при прийнятті управлінських рішень на прикладі підприємства малого бізнесу. *Бізнес та інтелектуальний капітал Інтелект XXI* №1, 2018 с. 235.

Yeremenko V.S., Monchenko O.V., Monchenko T.M. INFORMATION REPRESENTATION OF AN OBJECT IN ENDOCRINOLOGICAL RESEARCH

The paper presents an information model of the object in endocrinological research, aimed at a comprehensive assessment of a patient's condition in order to enhance treatment effectiveness and improve health outcomes and quality of life. The proposed model takes into account the multidimensional nature of endocrinological disorders and encompasses not only biochemical and physiological parameters, but also clinical, behavioral, and contextual characteristics that significantly influence the course of the pathological process and therapeutic outcomes.

Within the framework of the study, seven blocks of influencing factors were identified and structured in the form of an Ishikawa diagram to provide a clear visualization of cause-and-effect relationships. All blocks were analyzed and grouped into three generalized categories depending on the nature of the indicators and the methods used for their evaluation.

The first group includes the blocks "Biochemical Blood Parameters," "Biochemical Urine Parameters," and "Body Condition," for which quantitative assessment was performed using histogram-based distribution analysis. This approach made it possible to determine the variability patterns of the parameters and their contribution to the overall patient condition.

The second group comprises the blocks "Physical Components of Quality of Life in Patients with Arterial Hypertension and Obesity" and "Mental Components of Quality of Life in Patients with Arterial Hypertension and Obesity." For these blocks, expert evaluation was conducted followed by mathematical processing of the results using the Analytic Hierarchy Process based on the Saaty matrix, which enabled the determination of factor priorities and the degree of their influence.

The blocks "Cardiovascular System Status" and "General Clinical Characteristics" were calculated using categorized data with the application of Yule's coefficient.

Various data processing methods were employed in the study, and weighting coefficients were obtained; however, these coefficients are not mutually harmonized. The normalization of all weighting coefficients derived by different methods within the proposed model constitutes a subject for further research.

Keywords: *information model, endocrinological parameters, histogram analysis, categorized data, Yule's coefficient, Saaty matrix.*

Дата першого надходження статті до видання: 09.03.2026

Дата прийняття статті до друку після рецензування: 04.04.2026

Дата публікації (оприлюднення) статті 11.05.2026

Засадна Х.О.

<https://orcid.org/0009-0002-4143-7615>

Приватний вищий навчальний заклад

«Європейський університет»

Коцун В.І.

<https://orcid.org/0000-0003-2363-8157>

Приватний вищий навчальний заклад

«Європейський університет»

ГЕНЕРУВАННЯ КЛЮЧІВ В КРИПТОСИСТЕМІ RSA ТА ЇХ ЗАСТОСУВАННЯ

У статті розглядається алгоритм генерування ключів у несиметричній криптосистемі RSA та їх використання. Під час викладання дисциплін «Інформаційна безпека» та «Криптографічний захист інформації» для студентів різних напрямів підготовки ми стикнулися з проблемою доступного пояснення алгоритму генерування ключів. Викладений у друкованій літературі та в мережі Інтернет алгоритм не можна зрозуміти без попереднього пояснення деяких основ шкільного курсу математики. У статті наведено математичні основи несиметричних криптосистем, необхідні для розуміння алгоритму генерування ключів у криптосистемі RSA. Особливу увагу приділено способам розв'язування рівняння Діофанта.

Для автоматизації генерування ключів запропоновано шаблон у табличному редакторі Microsoft Excel.

Багато сучасних електронних сервісів вимагають розуміння протоколів використання двох ключів. У найближчому майбутньому їх матиме і буде змушений використовувати кожен громадянин України для ідентифікації та захисту особистої електронної інформації.

Сьогодні сфера застосування є двоключових криптосистем є досить багато. А з появою криптоіндустрії стало ясно, що криптографічні методи захисту інформації давно і активно інтегруються в наше повсякденне життя.

Вважаємо, що відносно молода наука під назвою «криптологія» повинна сьогодні вивчатися у кожному зов: не обов'язково як окрема дисципліна, можливо як частина дисциплін, що вивчають безпеку інформації та її захист. Школи вже цікавляться криптологією, про це свідчать численні публікації та відеоуроки в мережі Інтернет. Симетричні криптосистеми, які використовувалися до 1976 року є дуже цікавими, вони пояснюють постановки основних задач криптології і містять багато цікавих алгоритмів шифрування та дешифрування інформації, які розвивають логіку та креативність і які можна успішно програмувати.

Ключові слова: RSA, GCD, алгоритм Евкліда, рівняння Діофанта, ланцюгові дроби, використання несиметричних криптосистем, ЕЦП.

Постановка проблеми. Пояснити алгоритми роботи несиметричних криптосистем студентам, які не мають глибоких знань зі спеціальних дисциплін також можна. Для цього цілком достатньо елементарних знань зі шкільного курсу алгебри, вміння працювати з Калькулятором будь-якого гаджета, Калькулятором Windows та мати елементарні навички роботи з Microsoft Excel.

Після того, як студент згенерує власні ключі, підготує сертифікат відкритого ключа, зашифрує та дешифрує повідомлення, накладе ЕЦП на не-

лий електронний документ і перевірить його, можна буде вже детальніше розповідати йому про послуги Центрів сертифікації ключів, електронне голосування, електронний нотаріат, криптогаманець та криптобіржу, вбудований криптографічний захист інформації у програмному забезпеченні, криптографічний захист інформації у фінансових та державних установах, в Інтернет-сервісах і багато інших застосувань несиметричних криптосистем у його повсякденному житті та майбутній професійній діяльності. За нинішніх

обставин криптографічно захищені реєстри та послуги є практично незламними.

Аналіз останніх досліджень і публікацій. У роботах Вербіцького О.В., Ємця В., Мельника А., Поповича Р. [1, 2] та онлайн-енциклопедії Вікіпедія [3] викладено алгоритм генерування ключів у криптосистемі RSA, цей же алгоритм опублікований в багатьох інших джерелах. Однак без попереднього пояснення деяких математичних термінів та алгоритмів він буде незрозумілим. При розробці несиметричних криптосистем було детально проаналізовано всі недоліки попередніх симетричних (одноключових) криптосистем, враховано пропозиції криптоаналітиків і запропоновано використовувати два ключі. Але найбільше вражає те, що автори криптосистеми RSA використали при її розробці математичні задачі, які не мають однозначних розв'язків або не можуть бути розв'язані обчислювальним шляхом на той час і в найближчому майбутньому.

Постановка завдання. Метою даної публікації є пояснення операцій та рівнянь в алгоритмі генерування ключів в криптосистемі RSA та зв'язок між існуючими способами. Запропоновано реалізацію алгоритму для конкретних даних в середовищі MS Excel. Описані найчастіше використовувані застосування ключів.

Виклад основного матеріалу. Перш ніж описати ключові алгоритми криптосистеми RSA, треба пригадати деякі визначення зі шкільного курсу математики – математичні основи несиметричної криптосистеми RSA.

1. Прості та складені числа.

Прості числа – це натуральні числа, які мають тільки два дільники: одиницю і себе (наприклад, 2, 3, 5, 7, 11, 13, 17, 19, 23, 29...). Складені числа мають більше двох дільників і можуть бути розкладені на добуток простих чисел (наприклад, 4, 6, 8, 9, 10, 12...). Число 1 ні просте, ні складене, бо має лише один дільник. В RSA використовуються прості числа.

2. Функція mod.

Нехай $n \in N$ – довільне натуральне число, $x \in Z$ – довільне ціле число. Через $x \bmod n$ позначають залишок від ділення націло числа x на число n . Залишок від ділення завжди менший за дільник n і одного з ним знаку (тобто він невід'ємний).

$$8 \bmod 10 = 8 (8=0 \times 10 + 2), \quad 12 \bmod 10 = 2 (12=1 \times 10 + 2),$$

$$32 \bmod 10 = 2 (32=3 \times 10 + 2), \quad 52 \bmod 10 = 2 (52=5 \times 10 + 2),$$

$$-8 \bmod 10 = 2 (-8=-1 \times 10 + 2), \quad -18 \bmod 10 = 2 (-18=-2 \times 10 + 2),$$

$$-58 \bmod 10 = 2 (-58=-6 \times 10 + 2), \quad -78 \bmod 10 = 2 (-78=-8 \times 10 + 2).$$

Тобто при $x > 0$ від діленого віднімають дільник до тих пір, поки остача не стане меншою за нього, а при $x < 0$ до діленого додають дільник до тих пір, поки остача не стане меншою за нього.

Якщо $x \bmod n = y \bmod n$, то кажуть, що x і y конгруентні (порівняльні) за модулем n і записують це так: $x \equiv y \pmod{n}$. Іншими словами, конгруентні числа при діленні на одне і те ж натуральне число мають однакові залишки. Так, $12 \equiv 32 \pmod{10}$, $32 \equiv 52 \pmod{10}$, $-8 \equiv -58 \pmod{10}$, $52 \equiv -78 \pmod{10}$.

Множина конгруентних чисел за модулем n є нескінченною, а рівняння $x \bmod 10 = 2$, як видно із вище наведених прикладів, має безліч розв'язків.

Функція mod є на Калькуляторах, вбудована в MS Excel та у бібліотеки мов програмування.

3. Найбільший спільний дільник.

Дільником даного натурального числа називається натуральне число, на яке дане число ділиться без остачі. Наприклад, 15 ділиться без остачі на 1, 3, 5 і 15, а число 5 є одним з дільників числа 15. У рівнянні $15 = 2 \times 6 + 3$ число 15 – ділене, 2 – частка, 6 – дільник, 3 – остача. Іншими словами, 15 при діленні на 6 дає частку 2 і остачу 3.

У загальному випадку для будь-якого натурального (а також цілого числа) a і натурального b однозначно визначені цілі числа q і r такі, що $a = q \times b + r$, $0 \leq r < b$, q – частка, b – дільник, r – остача, $r = a \bmod b$ [1].

Спільним дільником кількох натуральних чисел називається число, на яке дані числа діляться без остачі. Один із способів знаходження спільних дільників – це розклад даних натуральних чисел на множники.

Найбільшим спільним дільником (НСД, GCD – Greatest Common Divisor) кількох натуральних чисел називається найбільше натуральне число, на яке кожне з даних натуральних чисел ділиться без остачі.

Два або кілька натуральних чисел, GCD яких дорівнює 1, називаються *взаємнопростими числами* і використовуються в криптології.

4. Знаходження GCD(a, b).

Якщо числа розкладаються на прості множники, то пошук їх GCD виконується саме цим шляхом. Наприклад, чисел $\text{GCD}(360, 336) = 24$, бо $360 = 2 \times 2 \times 2 \times 3 \times 3 \times 5$; $336 = 2 \times 2 \times 2 \times 2 \times 3 \times 7$.

Спільними дільниками чисел 360 і 336 є множина чисел $\{2, 3, 4, 6, 8, 12, 24\}$, а 24 – найбільший дільник. Крім того, перетин множин спіль-

них дільників заданих чисел рівний $\{2, 2, 2, 3\}$, а $\text{GCD}(360, 336) = 2 \times 2 \times 2 \times 3 = 24$.

Якщо ж числа важко розкласти на прості множники, для пошуку GCD використовують алгоритм Евкліда. У ньому остання ненульова остача і є GCD.

Даний алгоритм опирається на такі співвідношення.

$$\text{GCD}(a, b) = \text{GCD}(b, a);$$

$$\text{GCD}(a, b) = \text{GCD}(b, a \bmod b);$$

$$\text{GCD}(a, 0) = a.$$

Маючи ці три співвідношення алгоритму Евкліда, важко зрозуміти, як їх використовувати. Існує дуже просте пояснення, яке ми продемонструємо на конкретному прикладі.

Нехай треба знайти GCD чисел 2497 і 4200. Оскільки дані числа важко розкласти на прості множники, застосуємо алгоритм Евкліда, який використовує ділення двох цілих чисел. Ділити менше число на більше на першому кроці алгоритму – це зайва математична операція, бо $\frac{2497}{4200} = 0 \times 4200 + 2497$. Маємо пояснення першого твердження алгоритму Евкліда $\text{GCD}(a, b) = \text{GCD}(b, a)$. Тому ділене – 4200, а дільник – 2497. Знаходимо частку і остачу: $4200 = 1 \times 2497 + 1703$. Дільник 2497 ділимо на остачу 1703 і знову знаходимо частку та остачу. Ділення продовжуємо до отримання остачі 0. Цю послідовність дій описує друге твердження алгоритму Евкліда $\text{GCD}(a, b) = \text{GCD}(b, a \bmod b)$.

$$4200 = 1 \times 2497 + 1703; \quad 115 = 1 \times 104 + 11;$$

$$2497 = 1 \times 1703 + 794; \quad 104 = 9 \times 11 + 5;$$

$$1703 = 2 \times 794 + 115; \quad 11 = 2 \times 5 + 1;$$

$$794 = 6 \times 115 + 104; \quad 5 = 5 \times 1 + 0.$$

В математиці на 0 ділити не можна, тому останнє рівняння в алгоритмі Евкліда пояснює, що ділення закінчене і $\text{GCD}(a, 0) = a$, тобто $\text{GCD}(1, 0) = 1$. Найбільший спільний дільник – це остання ненульова остача.

Отримуємо $\text{GCD}(4200, 2497) = \text{GCD}(2497, 1703) = \text{GCD}(1703, 794) = \text{GCD}(794, 115) = \text{GCD}(115, 104) = \text{GCD}(104, 11) = \text{GCD}(11, 5) = \text{GCD}(1, 0) = 1$.

Числа 4200 і 2497 мають лише один спільний дільник – одиницю, і є взаємнопрості. І такі взаємнопрості числа використовуються в RSA.

5. Наслідок з алгоритму Евкліда – рівняння Діофанта.

Алгоритм Евкліда можна використати для розв'язування рівняння Діофанта. Для наведеного вище прикладу воно має вигляд $2497 \times x - 4200 \times y = 1$. У рівнянні Діофанта дві невідомі – x та y , але воно має єдиний розв'язок, бо використовує алгоритм Евкліда пошуку $\text{GCD}(2497, 4200)$.

Є три способи знаходження розв'язку рівняння Діофанта. За рекурентними формулами Ейлера, з використанням ланцюгових дробів та арифметичний. Перший спосіб найлегший. Наведемо рекурентні співвідношення для послідовностей P_n, Q_n .

$$P_{-2} = 0, P_{-1} = 1; \quad P_n = q_n P_{n-1} + P_{n-2}, \quad n \geq 0,$$

$$Q_{-2} = 1, Q_{-1} = 0; \quad Q_n = q_n Q_{n-1} + Q_{n-2}, \quad n \geq 0,$$

де q_i – це частки від ділення в алгоритмі Евкліда. Невідомі x та y обчислюються за формулами

$$x = (-1)^k P_{k-1}; \quad y = (-1)^{k-1} Q_{k-1}; \quad P_k = 4200, Q_k = 2497.$$

Тут k – номер останнього дільника в алгоритмі Евкліда, при цьому нумерація дільників починається з нуля.

Таблиця 1

Частки від ділення в алгоритмі Евкліда

q_0	q_1	q_2	q_3	q_4	q_5	q_6	q_7	k
1	1	2	6	1	9	2	5	7

Підставимо розв'язки в рівняння для перевірки: $2497 \times (-767) + 4200 \times 456 = 1$.

Розглянемо другий спосіб пошуку невідомих у рівнянні Діофанта – з використанням ланцюгових дробів. Утворимо ланцюговий дріб раціонального числа $\frac{4200}{2497}$:

$$\begin{aligned} \frac{4200}{2497} &= 1 + \frac{1703}{2497} = 1 + \frac{1}{\frac{2497}{1703}} = 1 + \frac{1}{1 + \frac{794}{1703}} = 1 + \frac{1}{1 + \frac{1}{\frac{1703}{794}}} = 1 + \frac{1}{1 + \frac{1}{2 + \frac{115}{794}}} \\ &= 1 + \frac{1}{1 + \frac{1}{2 + \frac{1}{6 + \frac{104}{115}}}} = 1 + \frac{1}{1 + \frac{1}{2 + \frac{1}{6 + \frac{1}{104}}}}} \\ &= 1 + \frac{1}{1 + \frac{1}{2 + \frac{1}{6 + \frac{1}{11}}}}} = 1 + \frac{1}{1 + \frac{1}{2 + \frac{1}{6 + \frac{1}{9 + \frac{1}{11}}}}} \\ &= 1 + \frac{1}{1 + \frac{1}{2 + \frac{1}{6 + \frac{1}{1 + \frac{1}{9 + \frac{1}{5}}}}}} \\ &= 1 + \frac{1}{1 + \frac{1}{2 + \frac{1}{6 + \frac{1}{1 + \frac{1}{2 + 5}}}}} \end{aligned}$$

Маємо ланцюговий дріб $\frac{4200}{2497} = [1; 1, 2, 6, 1, 9, 2, 5]$. Зауважимо, що такий же ланцюговий дріб утворюють частки в алгоритмі Евкліда.

Для кожного ланцюгового дроби можна обчислити збіжні дроби – частини ланцюгових дробів

Таблиця 2

Пошук розв'язків рівняння Діофанта з використанням рекурентних співвідношень

$P_{-2} = 0, P_{-1} = 1$	$Q_{-2} = 1, Q_{-1} = 0$
$P_0 = q_0 \times P_{-1} + P_{-2} = 1 \times 1 + 0 = 1$	$Q_0 = q_0 \times Q_{-1} + Q_{-2} = 1 \times 0 + 1 = 1$
$P_1 = q_1 \times P_0 + P_{-1} = 1 \times 1 + 1 = 2$	$Q_1 = q_1 \times Q_0 + Q_{-1} = 1 \times 1 + 0 = 1$
$P_2 = q_2 \times P_1 + P_0 = 2 \times 2 + 1 = 5$	$Q_2 = q_2 \times Q_1 + Q_0 = 2 \times 1 + 1 = 3$
$P_3 = q_3 \times P_2 + P_1 = 6 \times 5 + 2 = 32$	$Q_3 = q_3 \times Q_2 + Q_1 = 6 \times 3 + 1 = 19$
$P_4 = q_4 \times P_3 + P_2 = 1 \times 32 + 5 = 37$	$Q_4 = q_4 \times Q_3 + Q_2 = 1 \times 19 + 3 = 22$
$P_5 = q_5 \times P_4 + P_3 = 9 \times 37 + 32 = 365$	$Q_5 = q_5 \times Q_4 + Q_3 = 9 \times 22 + 19 = 217$
$P_6 = q_6 \times P_5 + P_4 = 2 \times 365 + 37 = 767$	$Q_6 = q_6 \times Q_5 + Q_4 = 2 \times 217 + 22 = 456$
$P_7 = q_7 \times P_6 + P_5 = 5 \times 767 + 365 = 4200$	$Q_7 = q_7 \times Q_6 + Q_5 = 5 \times 456 + 217 = 2497$
$x = (-1)^7 \times P_6 = -767$	$y = (-1)^6 \times Q_6 = 456$

Таблиця 3

Значення збіжних дробів

Перший збіжний дріб (перший елемент)	$[1] = 1$
Другий збіжний дріб	$[1;1] = 1 + \frac{1}{1} = 2$
Третій збіжний дріб	$[1;1,2] = 1 + \frac{1}{1 + \frac{1}{2}} = \frac{5}{3}$
Четвертий збіжний дріб	$[1;1,2,6] = 1 + \frac{1}{1 + \frac{1}{2 + \frac{1}{6}}} = \frac{32}{19}$
П'ятий збіжний дріб	$[1;1,2,6,1] = 1 + \frac{1}{1 + \frac{1}{2 + \frac{1}{6 + 1}}} = \frac{37}{22}$
Шостий збіжний дріб	$[1;1,2,6,1,9] = 1 + \frac{1}{1 + \frac{1}{2 + \frac{1}{6 + \frac{1}{1 + \frac{1}{9}}}}} = \frac{365}{217}$
Сьомий збіжний дріб	$[1;1,2,6,1,9,2] = 1 + \frac{1}{1 + \frac{1}{2 + \frac{1}{6 + \frac{1}{1 + \frac{1}{9 + \frac{1}{2}}}}} = \frac{767}{456}$
Восьмий збіжний дріб. Останній збіжний дріб завжди дорівнює початковому дробу	$[1;1,2,6,1,9,2,5] = [1;1,2,6,1,9,2,5] = 1 + \frac{1}{1 + \frac{1}{2 + \frac{1}{6 + \frac{1}{1 + \frac{1}{9 + \frac{1}{2 + \frac{1}{5}}}}} = \frac{4200}{2497}$

на кожному кроці. Обчислимо збіжні дробу для отриманого вище ланцюгового дробу.

Перші шість збіжних дробів можна було і не обчислювати, бо розв'язки рівняння Діофанта знаходяться в передостанньому дробі. Розглянемо два останні збіжні дробу: $\frac{4200}{2497}$ і $\frac{767}{456}$. Вони утворюють рівняння Діофанта $4200 \times 456 - 2497 \times 767 = 1$, звідки $x = -767$, $y = 456$.

У роботі [5] описаний зв'язок між алгоритмом Евкліда та рекурентними співвідношеннями Ейлера. І «посередником» є саме ланцюгові дробу. Бо чисельник (знаменник) кожного збіжного ланцюгового дробу використовує значення чисельників (знаменників) двох попередніх збіжних дробів. Пояснимо це твердження на розглянутому прикладі. Частки від ділення в алгоритмі Евкліда $q_i, i = 0, 1, \dots, 7$ наведені в таблиці 1.

Можна зробити такий висновок: використання збіжних дробів для знаходження розв'язків рівняння Діофанта більш трудомістке, ніж використання рекурентних співвідношень. Рекурентні співвідношення для послідовностей P_n, Q_n – це узагальнені формули для обчислення збіжних дробів, при цьому рекурентні співвідношення легко запрограмувати. Для того, щоб зручніше було їх використовувати і програмувати, заведені два додаткові початкові значення для кожної послідовності з від'ємними індексами: $P_{-2} = 0, P_{-1} = 1, Q_{-2} = 1, Q_{-1} = 0$.

Арифметичний спосіб утворення рівняння Діофанта використовує елементарні математичні операції – заміну, відкриття дужок, зведення подібних. Знайдемо остачі з рівнянь алгоритму Евкліда і виконаємо зворотні перетворення, починаючи з останнього рівняння.

$$\begin{aligned} 1 &= 11 - 2 \times 5, & 115 &= 1703 - 2 \times 794, \\ 5 &= 104 - 9 \times 11, & 794 &= 2497 - 1 \times 1703, \\ 11 &= 115 - 1 \times 104, & 1703 &= 4200 - 1 \times 2497. \\ 104 &= 794 - 6 \times 115, \\ 1 &= 11 - 2 \times 5 = 11 - 2 \times (104 - 9 \times 11) = 11 - 2 \times 104 + 18 \times 11 = 19 \times 11 - 2 \times 104 = 1 = \\ 19 \times (115 - 1 \times 104) - 2 \times 104 &= 19 \times 115 - 19 \times 104 - 2 \times 104 = 19 \times 115 - 21 \times 104 = 1 = \\ 19 \times 115 - 21 \times (794 - 6 \times 115) &= 19 \times 115 - 21 \times 794 + 126 \times 115 = 145 \times 115 - 21 \times 794 = 1 = \\ = 145 \times (1703 - 2 \times 794) - 21 \times 794 &= 145 \times 1703 - 311 \times 794 = 1 = \\ = 145 \times 1703 - 311 \times (2497 - 1 \times 1703) &= 145 \times 1703 - 311 \times 2497 + 311 \times 1703 = \\ = 456 \times 1703 - 311 \times 2497 &= 1 = 456 \times (4200 - 1 \times 2497) - 311 \times 2497 = 456 \times 4200 - 767 \times 2497 = 1. \end{aligned}$$

Арифметичний спосіб є «автономним», він не пов'язаний ні з рекурентними співвідношеннями, ні зі збіжними дробами. Не має значення, який метод використовувати для розв'язування рівняння Діофанта.

Загальний вигляд рівняння Діофанта $a \times x + b \times y = c$, де $c = \text{GCD}(a, b)$.

Таблиця 4

Зв'язок між збіжними дробами та рекурентними співвідношеннями

1	$[q_0] = \frac{q_0}{1} = \frac{1}{1} = \frac{P_0}{Q_0} = \frac{q_0 P_{-1} + P_{-2}}{q_0 Q_{-1} + Q_{-2}}, P_{-1} = 1, P_{-2} = 0, Q_{-1} = 0, Q_{-2} = 1$
2	$[q_0; q_1] = q_0 + \frac{1}{q_1} = \frac{q_1 q_0 + 1}{q_1} = \frac{1 + 1}{1} = \frac{2}{1} = \frac{P_1}{Q_1} = \frac{q_1 P_0 + P_{-1}}{q_1 Q_0 + Q_{-1}}$
3	$[q_0; q_1, q_2] = q_0 + \frac{1}{q_1 + \frac{1}{q_2}} = \frac{q_2(q_1 q_0 + 1) + q_0}{q_2 q_1 + 1} = \frac{2(1 + 1) + 1}{2 \times 1 + 1} = \frac{5}{3} = \frac{P_2}{Q_2} = \frac{q_2 P_1 + P_0}{q_2 Q_1 + Q_0}$
4	$[q_0; q_1, q_2, q_3] = \frac{q_3[q_2(q_1 q_0 + 1) + q_0] + (q_1 q_0 + 1)}{q_3(q_2 q_1 + 1) + q_1} = \frac{6 \times 5 + 2}{6 \times 3 + 1} = \frac{32}{19} = \frac{P_3}{Q_3} = \frac{q_3 P_2 + P_1}{q_3 Q_2 + Q_1}$
5	$[q_0; q_1, q_2, q_3, q_4] = \frac{q_4\{q_3[q_2(q_1 q_0 + 1) + q_0] + (q_1 q_0 + 1)\} + q_2(q_1 q_0 + 1) + q_0}{q_4\{q_3(q_2 q_1 + 1) + q_1\} + (q_2 q_1 + 1)} = \frac{1 \times 32 + 5}{1 \times 19 + 3} = \frac{37}{22} = \frac{P_4}{Q_4}$
6	$[q_0; q_1, q_2, q_3, q_4, q_5] = \frac{9 \times 37 + 32}{9 \times 22 + 19} = \frac{365}{217} = \frac{P_5}{Q_5} = \frac{q_5 P_4 + P_3}{q_5 Q_4 + Q_3}$
7	$[q_0; q_1, q_2, q_3, q_4, q_5, q_6] = \frac{2 \times 365 + 37}{2 \times 217 + 22} = \frac{767}{456} = \frac{P_6}{Q_6} = \frac{q_6 P_5 + P_4}{q_6 Q_5 + Q_4}$
8	$[q_0; q_1, q_2, q_3, q_4, q_5, q_6, q_7] = \frac{5 \times 767 + 365}{5 \times 456 + 217} = \frac{4200}{2497} = \frac{P_7}{Q_7} = \frac{q_7 P_6 + P_5}{q_7 Q_6 + Q_5}$

6. Генерування ключів у криптосистемі RSA.

Вважаємо, що тепер знайомство з алгоритмом генерування ключів у криптосистемі RSA буде зрозумілим, наведемо його [1, 2, 3].

1. Випадково вибираємо два великих простих числа p і q .

2. Обчислюємо їх добуток $n = p \cdot q$.

3. Обчислюємо число $\phi = (p-1) \cdot (q-1)$.

4. Вибираємо третє випадкове число e таке, що $e < \phi$ і взаємно просте з ϕ , тобто $\text{GCD}(e, \phi) = 1$.

5. Пара чисел (e, n) є утвореним *публічним (відкритим)* ключем.

6. Записуємо діофантове рівняння $e \cdot x + \phi \cdot y = 1$, у ньому e та ϕ відомі.

7. Знаходимо таємний ключ – число x , що задовольняє умову:

$$x \cdot e \equiv 1 \pmod{\phi}.$$

Пояснимо останнє рівняння. Таємний ключ – це множник при e , якщо він додатний ($d = x$ при $x > 0$). Якщо $x < 0$, то $d = x \bmod \phi$ (для знаходження таємного ключа шукаємо остачу від ділення цілого числа x на натуральне число ϕ).

8. Пара чисел (d, n) є *таємним (закритим)* ключем криптосистеми RSA.

Очевидно, що наведений вище алгоритм можна запрограмувати.

Зауважимо, що в літературі трапляються також рівняння $d \cdot e \equiv 1 \pmod{\phi}$ (правильне лише при $d = x > 0$), $d \equiv e^{-1} \pmod{\phi}$ (не слід трактувати e^{-1} як $-$).

7. Програмна реалізація алгоритму Евкліда та розв'язування рівняння Діофанта.

Для виконання обчислень при генеруванні ключів можна використати звичайний або програмований калькулятор, якщо вхідні дані – числа p , q та e є невеликими. Автоматизувати обчислення можна в MS Excel, а найкраще написати програму на будь-якій мові програмування. Нижче на рис.1–4 наведені скріни аркушів програми MS Excel з даними та формулами для розглянутих вище прикладів.

Крок, k	q(n)	Р(-2)	Q(-2)
0	1	0	1
1	1	1	0
2	2	2	1
3	6	3	2
4	1	37	3
5	9	365	22
6	2	767	456
7	5	4200	2497

Рис. 1. Результати пошуку GCD (e, ϕ) за алгоритмом Евкліда

Крок, k	q(n)	Р(-2)	Q(-2)
0	1	0	1
1	1	1	0
2	2	2	1
3	6	3	2
4	1	37	3
5	9	365	22
6	2	767	456
7	5	4200	2497

Рис. 2. Формули пошуку GCD (e, ϕ) за алгоритмом Евкліда

Крок, k	q(n)	Р(-2)	Q(-2)
0	1	0	1
1	1	1	0
2	2	2	1
3	6	3	2
4	1	37	3
5	9	365	22
6	2	767	456
7	5	4200	2497

Рис. 3. Розв'язки рівняння Діофанта за рекурентними формулами

Крок, k	q(n)	Р(-2)	Q(-2)
0	1	0	1
1	1	1	0
2	2	2	1
3	6	3	2
4	1	37	3
5	9	365	22
6	2	767	456
7	5	4200	2497

Рис. 4. Формули для пошуку розв'язків рівняння Діофанта

8. Застосування ключів криптосистеми RSA.

Пара ключів – відкритий і таємний – генеруються для юридичних та фізичних осіб. Фізичні особи можуть згенерувати їх для особистого користування, але юридичної сили такі ключі не матимуть. Ключі також генерують ЦСК (Центри Сертифікації Ключів), АЦСК (Акредитовані центри сертифікації ключів) та КНЕП (Кваліфіковані

Надавачі Електронних довірчих Послуг). Для відкритого ключа готують документ під назвою «Сертифікат відкритого ключа». Відкритий ключ зберігається у таблиці відкритих ключів. Таємний ключ копіюють на зовнішній носій інформації. Це може бути флеш-диск, флеш-токен або смарт-картка. Файли з таємним ключем, не мають єдиного фіксованого розширення, воно залежить від конкретного програмного забезпечення для генерації ключів, переважно зустрічаються .dat, .pfx, .p12, .pk8, .zs2, .jks. Доступ до файла захищається особистим паролем власника ключа. На ключовому флеш-диску можна також зберігати сертифікат.

Смарт-картка – це пластикова картка з чіпом, яка доступ до якої захищений пін-кодом власника, вона зберігає КЕП (Кваліфікований електронний підпис). КЕП використовують не лише для підписання документів, але й для подання звітності, доступу до держпослуг, в якості електронної печатки для юридичних осіб.

Флеш-токен – це апаратний носій для підписування електронних документів.

Найпоширеніші на сьогодні застосування несиметричних криптосистем – шифрування інформації та підписування електронних повідомлень. Відповідні алгоритми у криптосистемі RSA використовують операції піднесення до степеня та знаходження остачі від ділення. При цьому використовуються елементи колишніх симетричних криптосистем – заміна символів та блокове шифрування. Алгоритми шифрування та дешифрування, які використовуються в несиметричних криптосистемах опубліковані в мережі Інтернет і не становлять таємниці. Зауважимо лише, що зашифрований документ з використанням алгоритмів криптосистеми RSA – це набір десяткових цифр.

Шифрування собоюї інформації. Якщо треба захистити певну інформацію, використовують для шифрування особисті ключі: для шифрування – відкритий, для дешифрування – таємний. Очевидно, що таємний ключ (ключ дешифрування) має бути в безпеці, оскільки від нього залежить доступ до зашифрованої інформації. Таким чином можна захищати в електронному вигляді особисті, фінансові, медичні дані від доступу неуповноважених осіб.

Обмін повідомленнями по відкритому каналу зв'язку. Відправник шифрує повідомлення відкритим ключем отримувача і надсилає його каналом зв'язку. При цьому відправник не може прочитати зашифроване повідомлення, це може зробити лише отримувач за допомогою свого таємного ключа.

Підписування електронних документів без приховування їх змісту. Передбачає шифрування документа таємним ключем підписувача і надсилання отримувачу двох документів: оригінального та підписаного (зашифрованого) документа. Кажуть, що на електронний документ накладають ЕЦП (Електронний Цифровий підпис). Перевірка підпису полягає у виконанні двох кроків: дешифруванні ЕЦП відкритим ключем підписувача та порівнянні його з оригіналом документа. Якщо обидва документи співпадають, то підписувач ідентифікований і підпис дійсний. ЕЦП має юридичну силу власноручного підпису і найвищий рівень безпеки. Використовуються лише ключі підписувача електронних документів.

Підписування електронних документів з приховуванням їх змісту. Електронний документ шифрується відкритим ключем отримувача і одночасно підписується таємним ключем відправника. Це фактично шифрування одного і того ж документа різними ключами. Після отримання документа його дешифрують таємним ключем отримувача і перевіряють ЕЦП відкритим ключем відправника. Після перевірки ЕЦП і дешифрування обидва документи повинні бути ідентичними. Тільки тоді вони не спотворені, не сфальсифіковані і ЕЦП чинний.

У такий спосіб захищаються державні електронні послуги, електронний документообіг, електронна звітність, дистанційні банківські операції, інші важливі торгові і фінансові трансакції. Згаданий вище КЕП – це електронний ключ для роботи з цифровими сервісами, який призначений для накладання електронного підпису, що відповідає стандартам безпеки України та ЄС (eIDAS). Завдяки використанню криптографічних алгоритмів він має юридичну силу власноручного підпису. КЕП можна отримати в застосунку «Дія».

Сьогодні в Україні працює більше 100 державних сервісів та Інтернет-послуги банків, державних та бізнес-сервісів онлайн (Дія, податкова, Пенсійний фонд, тощо). електронного документообігу для громадян і бізнесу

Поділ таємниці. Схеми поділу таємниці використовуються для спільного управління технологіями та процесами. В такому управлінні можуть приймати участь n об'єктів або суб'єктів. Загальна таємниця ділиться на n частин, які називають частками. При об'єднанні $k \leq n$ часток таємниці використовується метод попереднього розподілу ключів, що дозволяє одноразово встановити ключ при згоді не менше ніж k об'єктів та суб'єктів. При цьому можлива ситуація, коли жодна група з $n-1$

особи не можуть отримати доступ до захищеної інформації. Поділ таємниці – це послідовність дій (процедура, протокол), яка визначає черговість дій для встановлення ключів чи паролів, які забезпечують доступ до таємниці.

Онлайн голосування. Має на меті забезпечити таємність голосування, неможливість його фальсифікації і правильність підрахунку результатів. Проведення голосування у такому форматі передбачає використання ЕЦП (відповідно наявності у голосуючого чинного Сертифіката відкритого ключа та носія з таємним ключем) та проведення зборів в режимі відеозасідання. При цьому канал зв'язку має бути надійно захищеним і передбачати дистанційну ідентифікацію фізичної особи голосуючого. Усі дії користувачів (приймають участь в голосуванні) та адміністраторів (обслуговують програмне забезпечення) неможливо видалити або змінити, лише зафіксувати. Будь-які правки в результатах голосування або документах не можуть залишатися непоміченими. Результати голосувань автоматично вносяться до протоколу і фіксуються з перевіркою ЕЦП або КЕП. Протоколи можуть відразу публікуватися на сайті установи [4].

Електронний нотаріат. Сьогодні можна підтвердити факт існування юридичного документа без розголошення його змісту. До тексту документа застосовують необоротну вкорочуючу функцію (її називають хеш-функцією). Вона перетворює вхідні дані будь-якого розміру в дані фіксованого розміру, які називають хеш документа, хеш-код, хеш-відбиток, хеш-образ, хеш-значення. Прикладом такої функції є, наприклад, розглянута вище функція *mod*. Хеш документа зберігається у нотаріуса в електронному вигляді. Знаючи хеш-образ документа, неможливо відновити сам документ. Для нотаріального підтвердження володіння електронним документом нотаріус накладає ЕЦП на хеш-документ і публікує хеш-документ, його ЕЦП та свій відкритий ключ (або вказує електронну адресу, за якою ці дані знаходяться). Після перевірки ЕЦП дешифрований хеш-документ порівнюють з його оригіналом. Якщо обидва документи ідентичні, це дозволяє підтвердити цифровий підпис та наявність оригіналу документа у нотаріуса. При цьому зміст самого документа відомий лише нотаріусу, оскільки такий протокол передбачає використання лише хеш-образу юридичного документа, який не містить жодної інформації про його оригінал.

Хеш-функції використовуються також при накладанні і перевірці цифрових підписів, для зберігання паролів (зберігають не сам пароль, а його хеш), при

перевірці цілісності файлів (якщо хеш змінився, то файл було змінено) та у технології блокчейн.

Ринок криптовалют та технологія блокчейн.

Криптовалюти та інші цифрові активи були розроблені з використанням криптографічних методів, що робить транзакції безпечними та надійними. Криптовалюта – це цифрові гроші. Операції з криптовалютою використовують технологію блокчейн, передбачають реєстрацію на криптовалютній біржі та створення криптовалютного гаманця. При його створенні генерується пара ключів – відкритий і таємний. Адреса гаманця генерується за допомогою відкритого ключа, її можна оприлюднювати. Закритий ключ використовується для накладання цифрових підписів та перевірки транзакцій з криптовалютою. Відкритий ключ використовується відправником для шифрування інформації, тоді як закритий ключ використовується одержувачем для її розшифрування. Ця система перевірки цифрового підпису гарантує, що лише особа, яка має приватний ключ, пов'язаний із відповідним гаманцем криптовалюти, може переміщувати кошти. Для створення унікальних ідентифікаторів транзакцій та забезпечення цілісності даних використовуються хеш-функції. Після перевірки транзакції шляхом підтвердження хешу, що міститься в цифровому підписі, цю транзакцію додають до книги блокчейну. Криптовалютні транзакції шифруються та розшифровуються з використанням ключів.

Висновки. У цій публікації ми просто і доступно пояснили алгоритм генерування ключів у криптосистемі RSA. Запропонували найпростіший спосіб генерування та його реалізацію в MS Excel для невеликих чисел. Описали використання ключів та пояснили необхідність зберігання у безпеці таємного ключа та призначення сертифіката відкритого ключа. Знаючи, як працює алгоритм генерування ключів та алгоритми шифрування і накладання ЕЦП, можна пояснювати термінологію Законів України «Про електронні документи та електронний документообіг», «Про електронні довірчі послуги», знайомитись з іншими застосуваннями криптосистем, шляхами її зламування та стійкістю, зрозуміти необхідність розробки наступного покоління криптосистем.

При роботі над цією статтею ми використовували AI ChatGPT. І зауважили, що програма робила помилки при розв'язуванні рівняння Діофанта, неправильно обчислювала деякі збіжні дробі, але виправляла помилки після зауважень. ChatGPT пропонує також програми генерування ключів на різних мовах програмування.

Список літератури:

1. Вербіцький О.В. Вступ до криптології. – Львів: Вид-во науково-технічної літератури, 1998. – 247 с.
2. В.Ємець. Сучасна криптографія. Основні поняття / В.Ємець, А.Мельник, Р.Попович. – Львів: БАК, 2003. – 144 с.
3. Вікіпедія – вільна енциклопедія. URL: <https://uk.wikipedia.org/wiki/RSA>.
4. 201923_golosuvannya-onlayn-z-bud-yako-tochki-svtu-yak-pratsyu-nova-sistema-vdeozasdan-dlya-organv-vladi. URL: <https://biz.ligazakon.net/news/>
5. Засадна Х.О. Алгоритми генерування ключів криптосистеми RSA/ Засадна Х.О. // Соціально-економічні проблеми сучасного періоду України. Фінансовий ринок України: стабілізація та євроінтеграція (збірник наукових праць) / НАН України. Інститут регіональних досліджень. – Львів, – 2010. – Вип.1(81). – С.458-467.

Zasadna Kh. O., Kotsun V. I. KEY GENERATION IN THE RSA CRYPTOSYSTEM RSA AND THEIR APPLICATION

The article discusses the algorithm for generating keys in the asymmetric RSA cryptosystem and their use. While teaching the disciplines «Information Security» and «Cryptographic Information Protection» to students of various fields of study, we encountered the problem of an accessible explanation of the key generation algorithm. The algorithm presented in printed literature and on the Internet cannot be understood without a prior explanation of some fundamentals of the school mathematics curriculum. The article presents the mathematical foundations of asymmetric cryptosystems, necessary for understanding the key generation algorithm in the RSA cryptosystem. Particular attention is paid to the methods of solving the Diophantine equation.

To automate key generation, a template is proposed in the Microsoft Excel spreadsheet editor.

Many modern electronic services require an understanding of the protocols for using two keys. In the near future, every citizen of Ukraine will have them and will be forced to use them to identify and protect personal electronic information.

Today, there are quite many areas of application for two-key cryptosystems. With the emergence of the crypto industry, it has become clear that cryptographic methods of information protection have long been actively integrated into our everyday lives.

We believe that the relatively young science known as cryptology should now be studied at every higher education institution: not necessarily as a separate discipline, but possibly as part of courses that deal with information security and its protection. Schools are already showing interest in cryptology, as evidenced by numerous publications and video lessons available on the Internet. Symmetric cryptosystems, which were used until 1976, are particularly interesting: they explain the formulation of the fundamental problems of cryptology and contain many intriguing algorithms for information encryption and decryption that develop logical thinking and creativity and can be successfully programmed.

Keywords: RSA, GCD, Euclidean algorithm, Diophantine equation, continued fractions, application of asymmetric cryptosystems.

Дата першого надходження статті до видання: 09.03.2026

Дата прийняття статті до друку після рецензування: 03.04.2026

Дата публікації (оприлюднення) статті 11.05.2026

Здоренко С.С.<https://orcid.org/0009-0003-7028-7203>

Національний технічний університет

України «Київський політехнічний інститут імені Ігоря Сікорського»

Черепанська І.Ю.<https://orcid.org/0000-0003-0741-7194>

Національний технічний університет

України «Київський політехнічний інститут імені Ігоря Сікорського»

СИСТЕМА МРС-КЕРУВАННЯ ТЕМПЕРАТУРОЮ ПОВІТРЯ В БУДІВЛЯХ З АДАПТАЦІЄЮ МАТЕМАТИЧНОЇ МОДЕЛІ

У статті вирішується важлива науково-технічна проблема енергоефективного керування температурою в будівлях з автономним опаленням, яка потребує застосування сучасних інноваційних методів та технологій опрацювання великих об'ємів різномірної інформації в режимі реального часу та відповідного автоматичного корегування параметрів математичної моделі. Вказане обумовлюється ціновими змінами на енергетичному ринку та нестабільними метеорологічними умовами навколишнього природного середовища (наприклад, різкими коливаннями температури, виникненням природних аномалій), що мають стохастичний характер, а також зміною теплофізичних характеристик огорожувальних конструкцій будівель через їх старіння. Саме тому традиційні методи та засоби автоматичного керування температурою ґрунтуються на використанні ПІ та ПІД регуляторів, є малоефективними.

У зв'язку з цим в статті представлено розроблену систему МРС-керування температурою повітря в будівлях, яка здатна автоматично здійснювати адаптацію математичної моделі в режимі реального часу. В якості об'єкта керування (ОК) виступає окреме приміщення чи зона в будівлі. Така будівля фактично являє собою складну розподілену у просторі динамічну систему з детермінованими параметрами. Саме тому для спрощення аналізу параметрів ОК та корегування математичної моделі з формуванням керуючих впливів, математична модель побудована за методом зосереджених параметрів. Лінійна структура та фізична інтерпретованість параметрів математичної моделі дозволяють ефективно почати керування ОК одразу після імплементації системи МРС-керування без її додаткового навчання. У цілому, пропонується система МРС-керування побудована за принципами синергетичної інтеграції і містить у своєму складі три основних модулі та МРС-регулятор. Зокрема розроблено модуль спостереження стану, що працює на основі фільтра Калмана, модуль виявлення розбіжності між математичною моделлю та реальною поведінкою ОК за алгоритмом на основі статистичного методу кумулятивних сум CUSUM та модуль корегування параметрів математичної моделі за певною процедурою. Це забезпечує підвищення точності керування, дозволяє досягти високої енергоефективності, яка полягає у визначенні оптимального співвідношення між витратами енергетичних ресурсів та комфортними температурними показниками повітря у житлових або громадських будівлях, а також дозволяє уникнути проблеми «холодного старту» та забезпечує відповідність математичної моделі ОК реальному стану ОК протягом усього життєвого циклу.

Ключові слова: Model Predictive Control, МРС, прогнозна модель, CUSUM, контрольна карта кумулятивних сум, енергоефективне керування, спостерігач стану.

Постановка проблеми. В останні роки в сфері регулювання мікроклімату будівель все більшої популярності набуває модельно-прогнозуюче керування (Model Predictive Control - МРС). Воно дозволяє реалізувати оптимальне керування в умовах необхідності економії витрат енергетичних ресурсів, з одночасним дотриманням оптимі-

заційних критеріїв, таких як підтримка високого рівня комфорту мешканців. Основною перевагою МРС-регулювання, у порівнянні з традиційними методами, є здатність прогнозувати та враховувати майбутню поведінку об'єкта керування (ОК), а також вплив на нього таких факторів, як температура зовнішнього повітря, сонячне випромін-

нювання, внутрішні теплові випромінювання від побутових приладів [1].

Однак ефективність MPC-керування значною мірою залежить від точності математичної моделі ОК. Через це виникають ключові виклики у використанні MPC-підходу: висока складність та трудомісткість розрахунку математичної моделі ОК за теплофізичними принципами, необхідність детальної і точної інформації про конструктивні особливості будівлі. Окрім того, можливе зниження точності математичної моделі в часі викликані зміною параметрів будівлі (наприклад, зміна теплових властивостей матеріалів будівлі у зв'язку із їх старінням чи зміною вологості), непередбачуваності внутрішніх тепловиділень, пов'язаних зі зміною поведінки мешканців та ін. Вказане зумовлює появу розбіжності між побудованою математичною моделлю та реальною поведінкою ОК. Розбіжність призводить до того, що статична математична модель, яка використовується в класичному MPC-регуляторі, починає надавати неточні прогнози, що спричиняє розрахунок неоптимальних керуючих дій, зниження енергоефективності та порушення умов комфорту.

Таким чином, ключовою проблемою при застосуванні MPC-керування є забезпечення відповідності побудованої математичної моделі реальному ОК протягом всього життєвого циклу системи керування (СК), тобто недопущення її деградації під впливом зміни теплофізичних характеристик ОК.

Аналіз останніх досліджень і публікацій. Відомі наукові дослідження Kuboth S., Heberle F., König-Haagen A., Brüggemann D. [2], Viot H., Sempey A., Mora L., Batsale J. C., Malvestio J. [3] та Chen Q., Li N., Feng W. [4] підтверджують значний потенціал MPC у підвищенні енергоефективності керування температурою повітря в будівлях. Проте критичний аналіз цих джерел свідчить, що ключовим бар'єром впровадження MPC-регуляторів залишається вибір методу моделювання, який би забезпечував баланс між точністю створеної математичної моделі, складністю розробки та здатністю до адаптації параметрів математичної моделі в часі. У переважній більшості робіт проблема розглядається однобоко: або через призму використання фізичних принципів і залежностей, або через призму застосування статистики та штучних нейронних мереж.

Так, в ряді робіт Drgoña J., Picard D., Helsen L. [5] Li X., Wen J. [6] та Jorissen F., Picard D. [7] використано математичні моделі, засновані на фундаментальних фізичних законах термодинаміки (white-

box метод). Автори відзначають високу точність та інтерпретованість таких математичних моделей, що дозволяє запускати СК одразу після монтажу (відсутність проблеми «холодного старту»). В статті авторів Drgoña J., Picard D. [5] описано впровадження MPC-керування для офісної будівлі з низькотемпературними опалювальними поверхнями на основі лінеаризованої математичної моделі, побудованої за white-box методом. Нелінійності, такі як сонячне випромінювання, було виокремлено і подано як попередньо розраховані збурення. В цілому підхід авторів забезпечив підвищення енергоефективності на 50% і точності підтримки температури повітря. Проте значна трудомісткість моделювання та високі обчислювальні витрати роблять цей підхід обмеженим для широкого практичного застосування. Окрім того, в роботі не враховано можливу зміну теплофізичних характеристик будівлі, що може спричинити деградацію математичної моделі ОК. В статті авторів Jorissen F., Picard D., Boydens W., Helsen L. [7] продемонстровано застосування MPC з моделями, побудованими на основі детальних фізичних симуляцій, що охоплюють не лише огорожувальні конструкції, а і систему опалення. Це дослідження може бути цінним для вирішення поставленої проблеми тільки після додаткового детального доопрацювання в контексті забезпечення адаптивності математичної моделі ОК.

Незважаючи на свою високу точність, надійність та інтерпретованість (оскільки всі змінні мають фізичний зміст) математичні моделі побудовані за white-box методом мають і ряд недоліків. Розробка таких моделей є складним та трудомістким процесом [5, 6, 7], вимагає значної кількості детальної інформації про будівлю, яку іноді важко чи взагалі неможливо отримати [8]. Окрім того залишилась невирешеною проблема, пов'язана з довготривалою експлуатацією системи – статичність математичної моделі, що не дозволяє уникнути її деградації з часом. Як зазначено авторами в роботах Ma L., Huang Y., Zhang J., Zhao T. [8] та Zhao D., Watari D., Ozawa Y., Taniguchi I., Suzuk T., Shimoda Y., Onoy T. [9] деградація теплофізичних характеристик матеріалів з часом призводить до накопичення систематичної похибки, яку white-box метод не здатен компенсувати автоматично.

В роботах Ma L., Huang Y., Zhang J., Zhao T. [8], Zhao D., Watari D., Ozawa Y., Taniguchi I., Suzuk T., Shimoda Y., Onoy T. [9], Weng K. [10] та Yang S., Wan M. P., Chen W., Ng B. F., Dubey S. [11] використано кардинально інший підхід до побудови математичних моделей – data-driven. Він базується

на зборі історичних даних про змінні ОК, статистичних методах та машинному навчанні (ML), які виявляють кореляцію між вхідними і вихідними даними, не вимагаючи побудови складних математичних моделей за фізичними принципами. Це так званий метод «чорної скриньки». В результаті отримується математична модель, здатна до самонавчання, що вирішує проблему дрейфу параметрів з часом.

Так, в статті авторів Ma L., Huang Y., Zhang J., Zhao T. [8] для прогнозування майбутніх температур внутрішнього повітря використано модель LSTM, яка у порівнянні з методом ідентифікації підпростору (SMI) показала кращу працездатність з історичними даними та запам'ятовування корисної інформації. Дана робота безумовно є цінним дослідженням, проте потребує доопрацювання в частині подолання обмежень, спричинених прив'язкою до умов, в яких відбувається навчання штучних нейронних мереж.

В роботі авторів Zhao D., Watari D., Ozawa Y., Taniguchi I., Suzuk T., Shimoda Y., Onoy T. [9] було описано систему MPC-керування з адаптивною моделлю на основі штучної нейронної мережі, ключовою особливістю якої є механізм регулярного оновлення математичної моделі на основі нових експлуатаційних даних, що збираються в режимі реального часу. Це дозволило підвищити точність прогнозування з часом та зменшити середню абсолютну похибку. Проте головним недоліком виявилась відсутність експлуатаційних даних для нових будівель, які ще не введені в експлуатацію, що спричиняє виникнення проблеми "холодного старту". Тому математичні моделі, побудовані на основі фізичних принципів, залишаються кращим варіантом.

Незважаючи на отримані позитивні результати автори вище згаданих робіт приходять до висновку, що ці підходи мають критичні обмеження щодо практичного застосування результатів для вирішення проблеми забезпечення відповідності побудованої математичної моделі реальному ОК протягом всього життєвого циклу СК. По-перше, вони вимагають тривалого періоду навчання та великої кількості накопичених даних для побудови адекватної математичної моделі. Це унеможливає чітку, надійну і ефективну роботу СК температурою повітря одразу після її впровадження на реальній будівлі (проблема «холодного старту»). По-друге, такі моделі обмежені умовами експлуатації, в яких вони були навчені і при виході за межі цих умов можуть виникати великі похибки прогнозування. Наприклад, математична модель,

навчена в м'яку зиму, даватиме значні похибки при аномальних морозах.

Підхід grey-box (наприклад, RC-моделі), розглянутий у статті авторів Li Y., O'Neill Z., Zhan L., Chen J., Im P., DeGraw J. [12], є компромісом, поєднуючи спрощену фізичну структуру зі статистичною ідентифікацією параметрів. Grey-box моделі гнучкіші за white-box моделі і в той же час вирішують проблему інтерпретованості краще за data-driven та не вимагають великого масиву даних для початкового запуску СК, проте вони все ж покладаються на фіксовану, заздалегідь визначену структуру. Це обмежує їхню здатність адаптуватися до складних нелінійних процесів або кардинальних змін у поведінці мешканців, що також не вирішує повною мірою проблему довгострокової адаптації. Аналіз роботи Li Y., O'Neill Z., Zhan L., Chen J., Im P., DeGraw J. [12] показує, що більшість існуючих реалізацій grey-box використовують ідентифікацію параметрів лише на етапі налаштування (офлайн). Питання онлайн-адаптації параметрів математичної моделі для компенсації дрейфу характеристик ОК залишається недостатньо вивченим. Існуючі методи часто не розрізняють зміну станів (швидка динаміка) та зміну параметрів (повільна динаміка), що обмежує їхню ефективність.

Таким чином, аналіз існуючих інформаційних джерел дозволяє виявити невирішену частину загальної проблеми, а саме забезпечення відповідності математичної моделі реальному ОК протягом всього життєвого циклу СК. Також варто звернути увагу на відсутність системних рішень для систем MPC, які б одночасно забезпечували фізичну релевантність моделі для надійного старту (перевага white-box) та здатність до безперервної автоматичної корекції параметрів у процесі експлуатації (перевага data-driven). Більшість відомих досліджень фокусуються лише на побудові математичної моделі, тоді як задача підтримки її актуальності в умовах невизначеності та старіння будівлі залишається невирішеною. Це обґрунтовує необхідність розробки гібридної структури СК температурою повітря в будівлях, яка запропонована в даній роботі. Така структура має поєднувати переваги різних підходів: початкову точність та інтерпретованість фізичної моделі (white-box), гнучкість grey-box-підходу та здатність до автоматичної адаптації параметрів на основі отриманих даних в результаті роботи системи.

Постановка завдання. Метою статті є розробка структури автоматизованої системи MPC-керування температурою повітря в будівлях, яка

інтегрує процедури виявлення розбіжності та корегування параметрів математичної моделі в режимі реального часу для забезпечення її точності та динамічної відповідності реальному ОК в умовах невизначеності та зміни з часом його теплових характеристик.

Виклад основного матеріалу. Запропонована структурна схема системи адаптивного МРС-керування (рис. 1) складається з двох контурів: основного контуру керування та наглядного контуру адаптації.

Основний контур системи адаптивного МРС-керування реалізує безпосереднє керування ОК згідно з МРС-алгоритмом. Він містить:

1) ОК, в якості якого виступає окреме приміщення чи зона в будівлі. На ОК здійснює вплив керуюча дія $u(k)$ (тепловий потік від системи опалення) та вектор збурень $d(k)$ (наприклад, зовнішня температура, теплове сонячне випромінювання, побутові тепловиділення, як то теплові випромінювання різних об'єктів та поверхонь). Адже відомо, що людина виділяє у стані спокою 80 – 100 Вт тепла за добу. Для персонального комп'ютера ці показники становлять 100 – 300 Вт за годину. Очевидно, що, у випадку малої кількості побутової та / або обчислювальної техніки їх теплові виділення не матимуть значного впливу на зміну температурних умов в приміщеннях, але повинні бути враховані при побудові математичної моделі задачі. В якості вихідного сигналу з ОК $y(k)$ виступають виміряні значення температурних показників повітря в приміщенні;

2) оптимальний МРС-регулятор;

3) модуль спостереження (спостерігач стану) на основі фільтра Калмана.

Оптимальний МРС-регулятор на основі фізично інтерпретованої математичної моделі ОК за виразом (1), діючих збурень, вихідного та цільового значення температури повітря розраховує оптимальну керуючу дію $u(k)$.

$$\begin{cases} x(k+1) = Ax(k) + Bu(k) + Ed(k) \\ y(k) = Cx(k) \end{cases} \quad (1)$$

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 – матриця спостереження.

Фізично інтерпретована модель за виразом (1) являє собою математичну модель ОК у просторі станів, яка побудована за фізичними принципами теплообміну з використанням РС-ланок і використовується МРС-регулятором для прогнозування реакції ОК. Для побудови математичної моделі використовується метод зосереджених параметрів, який передбачає зосередження фізичних характеристик ОК в окремих вузлах, що

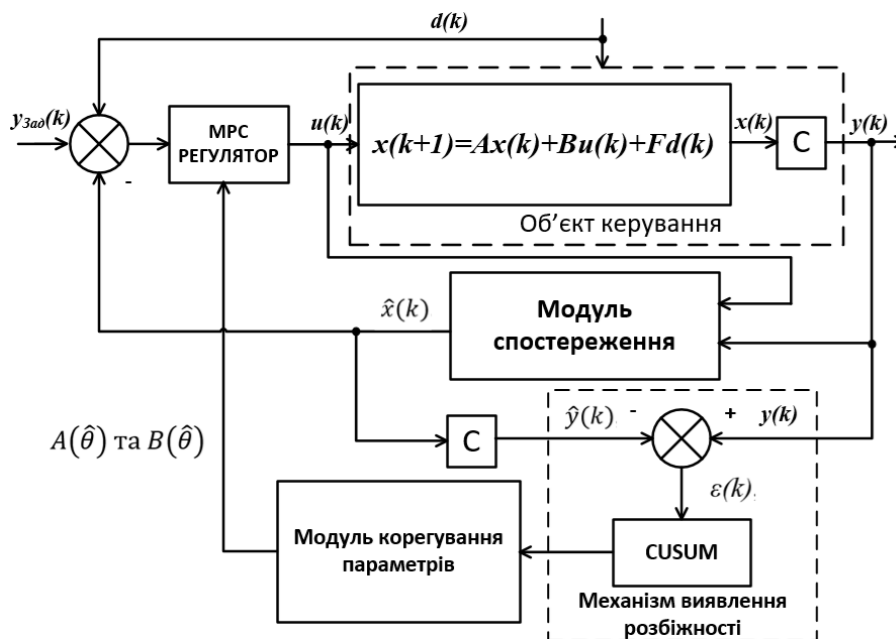


Рис. 1. Запропонована структурна схема системи адаптивного МРС-керування

спрощує його аналіз. В даному випадку в приміщенні виділяються три вузли: повітря (вузол 1), стіна (вузол 2) та підлога (вузол 3). За електро-тепловою аналогією кожен вузол має ємність C_i | $i = \overline{1;3}$, що акумулює тепло, та опір R_i | $i = \overline{1;3}$, що перешкоджає передачі тепла. Пропонована математична модель базується на рівняннях балансу енергії, де температура кожного i -го вузла змінюється з часом під впливом зовнішньої температури, температури теплоносія системи опалення та множини збурень, що мають стохастичний характер.

Лінійна структура та фізична інтерпретованість параметрів математичної моделі дозволяють ефективно почати керування ОК одразу після імплементації СК без додаткового навчання.

Модуль спостереження або спостерігач стану на основі фільтра Калмана використовується в *основному контурі керування* оскільки змінні стану, що відображають усереднені температури масивних конструкцій ОК (наприклад, стін, підлоги тощо), не можуть бути виміряні безпосередньо.

Фільтр Калмана на основі математичної моделі ОК, виміряного виходу $y(k)$ та відомої керуючої дії $u(k)$ обчислює оцінку повного вектора стану $x(k)$. Це забезпечує МРС-регулятор повною інформацією про поточний тепловий стан ОК. В запропонованій *системі адаптивного МРС-керування*, структурна схема якої наведена на рис.1, алгоритм фільтра Калмана працює в два етапи: етап 1 – прогноз, етап 2 – оновлення оцінки.

На етапі 1 алгоритм, використовуючи математичну модель ОК за виразом (1), вважає параметри A , B та D сталими і робить попередній прогноз оцінки вектора стану $\hat{x}(k/k-1)$ а поточному кроці k за даними попереднього кроку $k-1$ згідно з виразом (2):

$$\hat{x}(k/k-1) = A\hat{x}(k-1/k-1) + Bu(k-1) + Ed(k-1), \quad (2)$$

де $\hat{x}(k/k-1)$ – попереднє прогнозоване значення оцінки вектора стану на основі вектора стану на попередньому кроці $k-1$ без врахування виміряного значення $y(k)$;

$\hat{x}(k-1/k-1)$ значення оцінки вектора стану з попереднього кроку k .

На етапі 2 алгоритм роботи фільтра Калмана уточнює прогноз, враховуючи фактично виміряне значення температури $y(k)$. Оновлений вектор стану розраховується за виразом (3):

$$\hat{x}(k/k) = \hat{x}(k/k-1) + K(k)\varepsilon(k), \quad (3)$$

де $K(k)$ коефіцієнт Калмана;

$\varepsilon(k)$ значення похибки між виміряним значен-

ням $y(k)$ та оціненим спостерігачем $\hat{y}(k/k-1)$, яке визначається за виразом (4):

$$\varepsilon(k) = y(k) - \hat{y}(k/k-1) = y(k) - C\hat{x}(k/k-1). \quad (4)$$

Отримана оновлена оцінка стану $\hat{x}(k/k)$ використовується МРС-алгоритмом для формування керуючих дій $u(k)$, а похибка $\varepsilon(k)$ використовується як сигнал для наглядного контуру з метою моніторингу адекватності математичної моделі ОК та своєчасного виявлення дрейфу параметрів.

Наглядний контур адаптації працює паралельно з *контуром керування* і відповідає за підтримку актуальності математичної моделі ОК, яку використовує МРС-регулятор для здійснення прогнозів і формування керуючих дій. *Наглядний контур адаптації* є ключовим елементом запропонованої схеми. Він містить модуль виявлення розбіжності, що виконує процедуру виявлення розбіжності між математичною моделлю та поведінкою реального ОК, та модуль корегування, який виконує процедуру корегування параметрів математичної моделі відповідно до визначеного алгоритму.

Функціональним призначенням модуля виявлення розбіжності є своєчасне виявлення зміни параметрів математичної моделі ОК. Алгоритм виявлення розбіжності базується на безперервному аналізі різниці між виміряним значенням температури повітря на виході ОК $y(k)$ та оцінкою отриманою спостерігачем стану на основі фільтра Калмана $\hat{y}(k)$. Якщо похибка $\varepsilon(k)$ має статистично значущою та набуває систематичного характеру, генерується тригерний сигнал для активації механізму адаптації параметрів математичної моделі. Для реалізації цього алгоритму використовується статистичний метод контролю можливої зміни параметрів математичної моделі - кумулятивних сум (англ. Cumulative Sum – CUSUM), який описано в роботах Fang K., Huang Y., Huang Q., Yang S., Li Z., Cheng H. [13] та Severo M., & Gama J. [14]. Аналіз цих робіт свідчить, що CUSUM є ефективним інструментом моніторингу стабільності процесів, здатним виявляти незначні, але стійкі зміни середнього значення контрольованого параметра на фоні шумів.

На відміну від простих тригерних методів, що реагують на миттєве перевищення похибкою певного ліміту (що часто призводить до хибних спрацювань через шуми датчиків та випадкові збурення), концепція CUSUM полягає у послідовному накопиченні (акумуляції) дрібних відхилень реальних значень вимірювань в поточний момент часу від прогнозованих значень.

У стаціонарному режимі, коли математична модель відповідає реальному стану ОК, позитивні та негативні відхилення компенсують одне одного, і кумулятивні суми позитивних і негативних відхилень приблизно дорівнюють 0, $S^+(k) \approx 0$, $S^-(k) \approx 0$. У випадку появи систематичного зсуву середнього значення, кумулятивна сума починає стабільно збільшуватись або зменшуватись, що свідчить про зміну стану системи.

Коли одна з кумулятивних сум перевищує певний заданий статистичний пороговий рівень S_{extr} : $S^+(k) >_{extr}$, $S^-(k) >_{extr}$, алгоритм фіксує втрату стабільності і факт зміни параметрів математичної моделі в більшу або меншу сторону та генерує сигнал для активації алгоритму корегування параметрів математичної моделі. Після цього накопичені алгоритмом суми скидаються в 0 для початку нового циклу моніторингу.

Таким чином, інтеграція алгоритму CUSUM у контур адаптації дозволяє перейти від евристичного до статистично обґрунтованого критерію корегування параметрів математичної моделі. Це забезпечує фільтрацію короточасних збурень та звільняє обчислювальні потужності системи від надмірно частих перерахунків параметрів математичної моделі.

Модуль корегування активується сигналом від модуля виявлення розбіжності та виконує процедуру корегування параметрів математичної моделі в режимі реального часу за певним алгоритмом, детальний опис якого в даній статті не наводиться.

В результаті виконання завершального етапу роботи контуру адаптації відбувається оновлення вектора параметрів ОК за виразом (5):

$$\hat{\theta}(k/k) = \hat{\theta}(k/k-1) + K_{\theta}(k)\varepsilon(k), \quad (5)$$

де $\hat{\theta}(k/k)$ – оновлений вектор оцінки параметрів на поточному кроці;

$\hat{\theta}(k/k-1)$ – прогнозована оцінка вектору параметрів на поточному кроці на основі вектора параметрів на попередньому кроці;

$K_{\theta}(k)$ коефіцієнт Калмана для параметрів;

$\varepsilon(k)$ – похибка прогнозу оціненого значення температури $\hat{y}(k/k-1)$.

Таким чином, ітераційна процедура забезпечує збіжність оцінок параметрів ємності C_i | $i = \overline{1;3}$ та опору R_i | $i = \overline{1;3}$ ОК (вектор параметрів $\hat{\theta}(k/k)$) до їх істинних значень з низькою похибкою за певну кількість кроків. Отриманий актуалізований вектор параметрів $\hat{\theta}(k/k)$ використовується для перерахунку матриць $A(\hat{\theta})$ та $B(\hat{\theta})$ які передаються в МРС-регулятор для оновлення математичної моделі. Це гарантує підтримку високої точності прогнозування та енергоефективності керування навіть в умовах суттєвого дрейфу теплофізичних характеристик огорожувальних конструкцій.

Висновки. У статті запропоновано гібридну структуру автоматизованої системи МРС-керування температурою повітря в будівлях. Інтеграція у запропоновану автоматизовану систему МРС-керування температурою повітря в будівлях загальних процедур виявлення розбіжності та корегування параметрів математичної моделі ОК в режимі реального часу забезпечує підвищення точності енергоефективного автоматичного керування в умовах невизначеності та зміни з часом теплових характеристик ОК.

Список літератури:

1. 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. DOI: <https://doi.org/10.1016/j.apenergy.2019.03.209>
2. 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. DOI: <https://doi.org/10.1016/j.apenergy.2019.01.097>
3. Viot H., Sempay 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 I: Modeling and measurements. *Energy and Buildings*. 2018. №172. P. 94–103. DOI: <https://doi.org/10.1016/j.enbuild.2018.04.055>
4. 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. P. 110832. DOI: <https://doi.org/10.1016/j.enbuild.2021.110832>
5. Drgoňa J., Picard D., & Helsen L. Cloud-based implementation of white-box model predictive control for a GEOTABS office building: A field test demonstration. *Journal of Process Control*. 2020. №88. P. 63–77. DOI: <https://doi.org/10.1016/j.jprocont.2020.02.007>
6. Li X., & Wen J. Review of building energy modeling for control and operation. *Renewable and Sustainable Energy Reviews*. 2014. №37. P. 517–537. DOI: <https://doi.org/10.1016/j.rser.2014.05.056>
7. 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. DOI: <https://doi.org/10.23697/4934-p272>

8. Ma L., Huang Y., Zhang J., & Zhao T. A model predictive control for heat supply at building thermal inlet based on data-driven model. *Buildings*. 2022. №12. P. 1879. DOI: <https://doi.org/10.3390/buildings12111879>
9. Zhao D., Watari D., Ozawa Y., Taniguchi I., Suzuk T., Shimoda Y., & Onoy T. Data-driven online energy management framework for HVAC systems: An experimental study. *Applied Energy*. 2023. №352. P. 121921. DOI: <https://doi.org/10.1016/j.apenergy.2023.121921>
10. Weng K. Data-driven model predictive control of buildings: PhD Thesis. Cardiff, 2020. 175 p. URL: <https://orca.cardiff.ac.uk/id/eprint/139294>
11. Yang S., Wan M. P., Chen W., Ng B. F., & Dubey S. Model predictive control with adaptive machine-learning-based model for building energy efficiency and comfort optimization. *Applied Energy*. 2020. №271. P. 115147. DOI: <https://doi.org/10.1016/j.apenergy.2020.115147>
12. Li Y., O'Neill Z., Zhan L., Chen J., Im P., & DeGraw J. Grey-box modeling and application for building energy simulations-A critical review. *Renewable and Sustainable Energy Reviews*. 2021. №146. P. 111174. DOI: <https://doi.org/10.1016/j.rser.2021.111174>
13. Fang K., Huang Y., Huang Q., Yang S., Li Z., & Cheng H. An event detection approach based on improved CUSUM algorithm and Kalman filter. In 2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2). *IEEE*. 2020. P. 3400-3403. DOI: <https://doi.org/10.1109/EI250167.2020.9346980>
14. Severo M., & Gama J. Change detection with kalman filter and cusum. In International Conference on Discovery Science. *Berlin, Heidelberg: Springer Berlin Heidelberg*. 2006. P. 243-254. DOI: https://doi.org/10.1007/11893318_25

Zdorenko S.S., Cherepanska I.Yu. MODEL PREDICTIVE CONTROL SYSTEM OF AIR TEMPERATURE IN BUILDINGS WITH A MATHEMATICAL MODEL ADAPTATION MECHANISM

The article solves an important scientific and technical problem of energy-efficient temperature control in buildings with autonomous heating, which requires the use of modern innovative methods and technologies for processing large volumes of heterogeneous information in real time and the corresponding automatic adjustment of the parameters of the mathematical model. This is due to price changes in the energy market and unstable meteorological conditions of the environment (for example, sharp temperature fluctuations, the occurrence of natural anomalies), which are stochastic in nature, as well as changes in the thermophysical characteristics of building envelopes due to their aging. That is why traditional methods and means of automatic temperature control based on the use of PI and PID regulators are ineffective.

In this regard, the article presents the developed MPC-control system for air temperature in buildings, which is capable of automatically adapting the mathematical model in real time. A separate room or zone in the building acts as a control object (CO). Such a building is actually a complex spatially distributed dynamic system with deterministic parameters. That is why, to simplify the analysis of the CO parameters and the correction of the mathematical model with the formation of control influences, the mathematical model is built using the lumped parameter method. The linear structure and physical interpretability of the mathematical model parameters allow effective start of CO control immediately after the implementation of the MPC control system without its additional training. In general, the proposed MPC control system is built on the principles of synergistic integration and includes three main modules and an MPC controller. In particular, a state monitoring module has been developed, operating on the basis of the Kalman filter; a module for detecting discrepancies between the mathematical model and the real behavior of the CO using an algorithm based on the statistical method of cumulative sums CUSUM, and a module for correcting the parameters of the mathematical model using a certain procedure. This provides increased control accuracy, allows achieving high energy efficiency, which consists in determining the optimal ratio between energy resource consumption and comfortable air temperature indicators in residential or public buildings, and also avoids the problem of "cold start" and ensures that the mathematical model of the CO corresponds to the real state of the CO throughout the entire life cycle.

Keywords: Model Predictive Control, MPC, predictive model, CUSUM, Cumulative Sum control chart, energy-efficient control, state observer.

Дата першого надходження статті до видання: 19.03.2026

Дата прийняття статті до друку після рецензування: 14.04.2026

Дата публікації (оприлюднення) статті 11.05.2026

Зилевіч М.О.

<https://orcid.org/0000-0003-1646-0557>

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Сорокін Д.А.

<https://orcid.org/0000-0001-8221-216X>

Національний технічний університет України

Київський політехнічний інститут імені Ігоря Сікорського»

СИСТЕМАТИЗАЦІЯ ДЖЕРЕЛ ПОХИБОК ПРИ ВИРІШЕННІ ЗАДАЧ ГЕОЛОКАЦІЇ З ВИКОРИСТАННЯМ ТЕХНОЛОГІЇ LoRa

У статті розглянуто проблему геопозиціонування об'єктів у мережах Internet of Things (IoT) з використанням технології LoRa, що належить до класу LPWAN. Зростання кількості IoT-пристроїв та розширення сфер їх застосування зумовлює потребу у створенні енергоефективних і масштабованих систем визначення місцезнаходження. Технологія LoRa характеризується великою дальністю передачі даних, низьким енергоспоживанням і відносно невисокою вартістю розгортання мережі, що робить її перспективною для систем геолокації в міських, промислових і сільських умовах. Водночас використання цієї технології супроводжується рядом факторів, які впливають на точність визначення координат об'єктів.

У роботі проаналізовано основні методи визначення місцезнаходження у бездротових мережах, зокрема RSSI, ToA, TDoA та AoA. Обґрунтовано доцільність використання методу локалізації на основі індикатора потужності прийнятого сигналу (RSSI), оскільки він не потребує додаткового апаратного забезпечення та може бути реалізований у більшості існуючих мереж LoRaWAN. Разом з тим такий підхід є чутливим до впливу навколишнього середовища та наявності перешкод, що призводить до появи похибок під час оцінювання відстані між вузлами мережі.

Розглянуто алгоритм геопозиціонування на основі методу трилатерації, у якому координати цільового вузла визначаються за відстанями до кількох базових станцій. Проаналізовано процес перетворення параметрів радіосигналу у координати об'єкта. Показано, що точність позиціонування значною мірою залежить від параметрів моделі поширення сигналу, які змінюються залежно від типу середовища.

У результаті аналізу систематизовано основні джерела похибок геопозиціонування в мережах LoRa: вплив навколишнього середовища (затіннення та багатопроменеве поширення сигналу), динамічні фактори, похибки математичних моделей, фізичні особливості згасання радіохвиль, технічні характеристики обладнання та геометрію розташування базових станцій. Перспективи подальших досліджень пов'язані з удосконаленням моделей втрат сигналу, використанням методів фільтрації вимірювань і застосуванням вагових коефіцієнтів на основі співвідношення сигнал/шум (SNR) для підвищення точності визначення координат.

Ключові слова: Internet of Things, технологія LoRa, геолокація, похибка, втрати на трасі, алгоритм локалізації.

Постановка проблеми. Традиційні методи позиціонування, такі як глобальна навігаційна супутникова система (GNSS/GPS), попри їхню високу точність на відкритій місцевості, стикаються з критичними обмеженнями в умовах щільної міської забудови, всередині приміщень та в промислових середовищах через затухання сигналу, ефекти багатопроменеве поширення та високого енергоспоживання кінцевих пристроїв.

Для вирішення вузьких місць навігаційної супутникової системи варто розглянути застосування IoT-пристроїв. Це пов'язано, в першу чергу, з поширенням технологій IoT та мобільних пристроїв. Згідно з останнім звітом IoT Analytics, очікувалось, що до кінця 2025 року кількість підключених IoT-пристроїв у всьому світі зросте на 14% і сягне позначки 21,1 мільярда [1]. Їх використання базується на системах наземного радіозв'язку, дальність покриття яких напряму визначена

технологією мережі. Вони умовно поділяються на три категорії: 10 – 100 см (NFC та RFID), 10 м – 100 м (Wi-Fi, BLE та ZigBee) та 1 – 50 км (стілнкова мережа та LPWAN). Технології LPWAN, і зокрема технологія LoRaWAN, стають дедалі актуальнішими завдяки своїм характеристикам, включаючи покриття на значні відстані (до 15 км у міських районах та до 50 км у сільській місцевості), низьке енергоспоживання, низьку вартість розгортання (безкоштовна ліцензована смуга частот та дешеві кінцеві вузли) тощо [2-4].

Таким чином, альтернативні методи позиціонування варто розглядати саме в мережах стільникового та LPWAN зв'язку. Одним з представників цих мереж є технологія бездротового зв'язку великої дальності (LoRa) заснована на методах розширеного спектру лінійною частотною CSS (Chirp Spread Spectrum) модуляцією, класифікована як технологія зв'язку низькоенергетичних широкомугових мереж (LPWAN). Це дозволяє виконувати передачу на значну відстань та підвищити прохідну спроможність радіосигналу, порівняно з традиційною модуляцією із частотною маніпуляцією (FSK). За рахунок цього система є більш стійкою до впливу стаціонарних та нестаціонарних завад. При цьому така технологія споживає близько 14 мА в режимі очікування та до 45 мА під час передачі даних, що є показником високої енергоефективності. [5, 6].

LoRa – перше рішення, що використовує технологію CSS для економічно ефективного комерційного застосування, що дозволяє не тільки забезпечити широкомасштабне підключення вузлів в мережі, але і суттєво знизити вартість кінцевих пристроїв та інфраструктури в цілому. Тому технологія LoRa (Long Range) як провідна технологія LPWAN, відіграє вагомий роль у забезпеченні бездротового зв'язку на значні відстані з низьким енергоспоживанням [7].

Як правило, шлюзи та вузли LoRa розміщуються або у міських умовах, де має місце щільна забудова та певна кількість джерел спотворення радіосигналу, або у сільських умовах, де сигнал розповсюджується на великі відстані, що є істотним обмеженням для радіопередачі. Це створює так звані стаціонарні та нестаціонарні завади, які зменшують ефективність, точність та в цілому ускладнює геопозиціонування, що буде розглянуто далі.

Таким чином використання даної технології актуальне для застосування її в системах позиціонування на місцевості. Водночас виникає низка чинників, відповідно до яких, використання тех-

нології LoRa супроводжується існуванням завад, які породжують похибки. Аналіз та систематизація джерел таких похибок стає актуальним завданням для пошуку рішень протидії впливу стаціонарних та нестаціонарних завад. Це дозволяє поліпшити оптимальність визначення місцезнаходження об'єкту в умовах складного навколишнього середовища з використанням наземних систем позиціонування.

Аналіз останніх досліджень і публікацій. Вирішення задачі локалізації (геопозиціонування) є багатограним процесом оцінки фізичних координат вузла з невідомим розташуванням відносно опорних точок (анкерів). З погляду системного аналізу та обробки сигналів, методи локалізації класифікуються за типом використовуваних вимірювань, геометричними принципами розрахунку та ступенем залучення заздалегідь підготовлених даних. Загальноприйняті підходи визначають методи на основі дистанційних вимірювань.

Отримання координат шляхом таких вимірювань можна описати за наступним спрощеним алгоритмом. В системі геопозиціонування, що складається з передавача (його місцезнаходження необхідно визначити) та приймачів (базових станцій для збору даних) вимірюються фізичні параметри, які слугуватимуть основою для визначення остаточної координати об'єкту. Відповідно до [5] можна виокремити наступні методи та параметри, які базуються на перетворенні фізичних характеристик радіосигналу в оцінку відстані або кута:

– RSSI (Received Signal Strength Indicator): Використовує рівень потужності прийнятого сигналу для розрахунку відстані на основі моделей затухання, таких як логарифмічна модель втрат на шляху (Log-distance Path Loss). Це найбільш економічно ефективний метод, оскільки він не потребує додаткового обладнання, проте він характеризується високою дисперсією через багатопроменеве поширення та затінення.

– ToA (Time of Arrival): Базується на вимірюванні абсолютного часу проходження сигналу від передавача до приймача. Метод вимагає прецизійної синхронізації годинників усіх вузлів мережі, оскільки навіть мікросекундна похибка призводить до значних відхилень у координатах.

– TDoA (Time Difference of Arrival): Розраховує різницю часу приходу сигналу на кілька шлюзів. Це дозволяє уникнути необхідності синхронізації кінцевого пристрою з мережею, вимагаючи синхронізації лише між шлюзами.

– AoA (Angle of Arrival): Визначає кут приходу сигналу за допомогою фазованих антенних решіток.

Метод дозволяє реалізувати локалізацію за допомогою триангуляції, але потребує дорогого та енергоємного обладнання.

У [4] досліджується роль індикатора потужності прийнятого сигналу (RSSI) у локалізації об'єктів і здійснюється комплексне порівняння цих методів та забезпечує детальний аналіз їхньої продуктивності, сильних сторін та обмежень у реальних сценаріях.

Тоді як у [8] зазначається, що позиціонування на основі індикатора потужності прийнятого сигналу (RSSI) підходить для розгортання мереж, завдяки своїй невисокій вартості та простій реалізації. Однак воно має низьку стабільність та точність, оскільки RSSI спотворюється об'єктами та впливом навколишнього середовища.

Водночас [9] порівнюють ці методи з точки зору проблеми затінення радіосигналу, вартості розгортання системи і її експлуатації. Приходять до висновку, що застосуванням різних алгоритмів визначення місцезнаходження залежить від наявного обладнання, відстані шлюзів та вузлів, а також програмних можливостей в середовищі LoRaWAN.

Узагальнюючи попередні дослідження обрано метод, що базується на вимірюванні потужності прийнятого сигналу, тобто RSSI ранжування.

Постановка завдання. Вирішення задачі гео-позиціонування супроводжується низькою похибкою на різних етапах її реалізації, що безпосередньо впливає на підсумкову точність визначення координат. Відповідно, завданням статті є аналіз і систематизація факторів та джерел, що генерують ці похибки в умовах використання технології LoRa.

Виклад основного матеріалу. На основі аналізу наукових джерел виокремлено наступні основні групи похибок геопозиціонування:

1. Похибки, спричинені мінливістю навколишнього середовища Основною проблемою є висока чутливість показника RSSI до фізичних характеристик довкілля. Згасання у вільному просторі, в ідеальних умовах, залежить лише від відстані та частоти. Однак на практиці фактичні значення часто перевищують “ідеальні”. Під час бездротової передачі даних наявність перешкод між передавачем та приймачем порушує пряме поширення радіосигналу. Ці перешкоди викликають поглинання, відбиття та розсіювання електромагнітних хвиль, що призводить до значного зниження потужності сигналу на вході приймача – явища, відомого як затінення (Shadowing). Крім того, радіохвилі можуть досягати приймача різними

шляхами (через пряму видимість, відбиття та заломлення). Через різницю в довжині цих шляхів хвилі надходять із часовою затримкою, що спричиняє ефект багатопробеневого поширення (Multipath fading). Це генерує фазові зсуви та додатково погіршує якість сигналу [2, 6].

Окрім інфраструктурних об'єктів щільної міської забудови до джерел таких похибок також можна віднести погодні умови (зміна вологості, температури, опади).

2. Похибки, спричинені динамічними факторами. Процес геолокації рухомих об'єктів або робота в динамічному середовищі супроводжується появою тимчасових перешкод на шляху поширення радіосигналу (наприклад, рух людей чи транспорту). У випадку, коли об'єкт рухається статичні джерела перетворюються на динамічні. Така мінливість навколишнього середовища не тільки є джерелом затінення та багатопробеневого поширення, але і робить ці ефекти стохастичними [3].

3. Похибки моделей перетворення вимірюваного значення у відстань. Як зазначалося вище, радіосигнал поширюється в неідеальних умовах. Для врахування цих умов використовуються різноманітні моделі поширення радіохвиль, які математично описують характер впливу завад різної природи.

На рис. 1 наведено зв'язок між середовищем передачі та втратами сигналу. З аналізу можна зробити висновок, що основні втрати відбуваються в середовищі передачі.

Специфіка методу геопозиціонування на базі RSSI полягає в оцінці потужності радіосигналу, що втрачається під час проходження від передавача до приймача (так звані втрати на трасі, Path Loss). Ці втрати визначаються за допомогою бюджету каналу зв'язку (Link Budget). Він розраховується як

$$P_{Rx} = P_{Tx} + G_{Rx} + G_{Tx} - L_{PL} \quad (1)$$

де P_{Rx} – це потужність сигналу в приймачі в дБ (відповідає RSSI), P_{Tx} – потужність сигналу в передавачі, G_{Rx} і G_{Tx} – є коефіцієнти посилення антен, які використовуються приймачем і передавачем відповідно, L_{PL} – складова, яка враховуватиме втрати сигналу [10].

Після визначення RSSI з використанням виразу (1), для подальшого розрахунку відстані до джерела сигналу необхідно встановити зв'язок між втратами та відстанню на трасі. Теоретично, бездротовий сигнал використовує вільний простір, як середовище передачі. Втрати в тракці



Рис. 1. Втрати сигналу в лінії передачі

передачі зростають пропорційно квадрату відстані, яку він проходить від точки передачі. Відповідно для ідеальних умов можна визначити, що

$$L_{PL}(R) = 20 * \log\left(\frac{\lambda}{4\pi R}\right) \quad (2)$$

де R – відстань між антенами, λ – довжина хвилі.

Як зазначалось раніше, є багато джерел електромагнітних полів, особливо в міських районах, які впливають на несучий сигнал [11]. В цьому випадку більш доречно оцінювати відстань за формулою логарифмічної відстані втрат на трасі (Log-distance path loss model) [12]:

Це можна описати простим рівнянням:

$$L_{PL}(d) = L_{PL}(d_0) + 10 * Y * \log\left(\frac{d}{d_0}\right) + X_\sigma \quad (3)$$

де $L_{PL}(d)$ – являє собою втрати на трасі в дБ на відстані d , $L_{PL}(d_0)$ – є втрати в тракці передачі на опорній відстані d_0 , де ще не спостерігається вплив навколишнього середовища, Y є показником втрат на трасі і X_σ – це втрата від затухання тіні з нульовим середнім гауссовим розподілом і еталоном відхилення від σ . Зазвичай d_0 прийнято брати за 1 км.

Саме коефіцієнт Y визначає інтенсивність цих втрат. Наприклад, для міста Оулу у Фінляндії значення складає 2,32, а для міста Дортмунд в Німеччині -2,65. Такі відмінності обумовлені зовнішніми факторами та індивідуальними особливостями міського середовища [13].

Проведена оцінка точності з використанням цього показника в межах від 2 до 6 умовах закритого приміщення показало, що отримані значення досягали максимальної похибки при $Y = 2$ та мінімальної похибки при $Y = 6$, що узгоджується з літературними значеннями [14].

Таблиця 1

Коефіцієнт інтенсивності втрат за D'Arienzo, M. [4],

Середовище	Показник втрат (Y)
Вільний простір	2
Міська забудова	2.7 to 3.5
Міська забудова із затіненням	3 to 5
Приміщення	1.6 to 1.8
Приміщення з перешкодами	4 to 6
Промислова забудова	2 to 3

Таким чином, можна стверджувати, що вибір моделі розрахунку втрат сигналу впливає на точність позиціонування. Треба зауважити, що не можна використовувати одне значення для різних середовищ, необхідно проводити окремі вимірювання для визначення цих параметрів.

4. Похибки спричинені фізичними властивостями згасання сигналу. Зв'язок між RSSI та відстанню описується логарифмічною або зворотно-експоненціальною залежністю, як зображено на рис.2, що породжує специфічні математичні труднощі [3, 4, 9]. На значних відстанях крива згасання стає більш пласкою, через що навіть незначна помилка у вимірюванні RSSI (шум) трансформується у відчутну похибку при оцінці фізичної відстані. Окрім цього відношення сигнал/шум (SNR) погіршується зі збільшенням відстані, що робить оцінку координат об'єктів на периферії мережі вкрай нестабільною [6, 9].

5. Похибки, спричинені технічними властивостями системи. Точність системи обмежена характеристиками обладнання, що використовується, та його конфігурацією [4]. По-перше, взаємодія обладнання має бути узгодженою, тому що при обрахунках враховуються такі його параметри як підсилення антен, втрати в трактах передачі. По-друге, напрямки антени цільового вузла та її

поляризація безпосередньо впливають на рівень сигналу, що приймається, і, як наслідок, на точність локалізації [3].

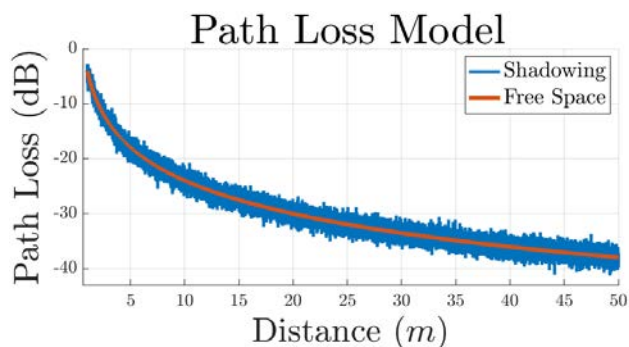


Рис. 2. Залежність гасання сигналу від відстані

6. Геометричні та алгоритмічні виклики. Ефективність алгоритмів трилатерації обмежена геометричною конфігурацією мережі. Взаємне розташування базових станцій та цільового об'єкта суттєво впливає на точність визначення його координат; при несприятливій геометрії (наприклад, розташуванні вузлів на одній лінії) помилки позиціонування можуть багатократно зростати [3].

В методі трилатерації місцезнаходження об'єкту визначається геометричним перетином кіл на площині. На практиці результатом перетину є певна площа, де ймовірно розташування об'єкта. Це змушує використовувати методи для апроксимації (наприклад, Min-Max або метод найменших квадратів), які також мають обмеження щодо точності [4, 8, 14].

7. Проблеми конфігурації мережі. Технологія LoRa базується на гнучкому налаштуванні параметрів фізичного рівня (PHY), що дозволяє адаптувати мережу до конкретних умов навколишнього середовища, вимог до дальності та енергоспоживання. До основних параметрів можна віднести [7]:

Spreading Factor (SF) – Фактор розширення спектра, який визначає кількість «чирпів» (chirps), що припадають на один інформаційний символ. У стандарті використовуються значення від 7 до 12. Вищий SF збільшує відношення сигнал/шум (SNR), чутливість приймача та дальність зв'язку, але водночас експоненціально збільшує час перебування в ефірі (Time-on-Air, ToA) та енергоспоживання. Варто зауважити, що сигнали з різними SF є квазі-ортогональними, що дозволяє шлюзу приймати їх одночасно на одному частотному каналі.

Bandwidth (BW) – Ширина смуги пропускання, яка визначає діапазон частот, у якому розподіляється енергія сигналу. Зазвичай це значення приймає діапазон 125 кГц, 250 кГц або 500 кГц.

Збільшення BW дозволяє підвищити швидкість передачі даних (Data Rate), але пропорційно знижує чутливість приймача через збільшення вхідного шуму [15].

Transmission Power (TP) – Потужність передачі на виході антени, зазвичай у діапазоні від 2 до 20 дБм. Вища потужність дозволяє подолати згасання в складному середовищі (наприклад, у закритих приміщеннях або через рослинність), проте це скорочує термін служби батареї та збільшує інтерференцію для сусідніх вузлів [16].

Coding Rate (CR) – Швидкість кодування визначає ступінь надлишковості для прямого виправлення помилок (FEC), її значення варіюються від 4/5 до 4/8. Вищий CR (наприклад, 4/8) забезпечує кращий захист від імпульсних завад, але збільшує ToA приблизно на 25% у порівнянні з CR 4/5 [17].

Оптимальна конфігурація мережі LoRa полягає у розв'язанні компромісу між надійністю доставки пакетів (PDR) та енергоефективністю. Для забезпечення надійності в реальних умовах при розрахунку бюджету лінії зв'язку рекомендується додавати запас міцності (margin), що враховує стандартне відхилення згасання сигналу (σ).

Висновки. Таким чином, розглянуто спрощений алгоритм вирішення задачі геолокації з використанням методу трилатерації та його основні етапи. Встановлено, що на кожному з етапів існують джерела похибок, які можна систематизувати наступним чином.

Похибки спричинені мінливістю навколишнього середовища. Стаціонарні та нестаціонарні завади обумовлюють явище затінення та є джерелами багатопроменевого поширення сигналу, що зменшує точність визначення координати. Рух об'єкту під час позиціонування підсилює вплив даних явищ.

Похибки обумовлені математичними розрахунками. При моделюванні явища вільного згасання радіосигналу необхідно враховувати інтенсивність втрат в залежності від індивідуальних характеристик середовища розповсюдження радіохвиль. А вибір методу апроксимації буде впливати на точність отримання координат об'єкта.

Похибок техногенного характеру. При взаємодії пристроїв системи локалізації важливо враховувати технічні параметри системи, такі як підсилення антен, втрати потужності в тракті передачі, поляризацію антен, конфігурацію мережі. Окрім цього необхідно звернути увагу на геометрію розташування базових станцій, оскільки їх взаємне розташування суттєво впливає на подальші розрахунки.

Похибки, спричинені чутливістю системи позиціонування до роботи на значних відстанях, де вплив логарифмічного характеру згасання стає суттєвим. А відношення сигнал/шум робить оцінку таких значень вкрай нестабільною.

Перспективи подальших досліджень передбачають аналіз математичних моделей втрат

сигналу на місцевості з урахуванням нестационарних завад та ефекту багатопроменевості. Для підвищення точності визначення координати планується розглянути використання фільтрів для обробки помилкових вимірів або вагових коефіцієнтів що базуються на SNR (співвідношенні сигнал/шум).

Список літератури:

1. *State of IoT 2025: Number of connected IoT devices growing 14% to 21.1 billion globally*. URL: iot-analytics.com (дата звернення: 13.03.2026).
2. Moradbeikie A., Keshavarz A., Lopes S. A. *Enhancing LoRaWAN Localization in Industrial Environments: Merging Path-Loss Modeling, Extended Kalman Filtering and Map-Matching*. IEEE Access. 2025. Vol. 13. P. 84765–84778. DOI: 10.1109/ACCESS.2025.3568750
3. Alireza S. *Evaluation and Improvement of the RSSI-based Localization Algorithm: Received Signal Strength Indication (RSSI)*. Digitala Vetenskapliga Arkivet, 2015. 109 p.
4. D'Arienzo M. *An Experimental Comparison of Basic Device Localization Systems in Wireless Sensor Networks*. Network. 2025. Vol. 5, № 2. P. 11. DOI: 10.3390/network5020011
5. Pastório A. F., De Camargo E. T. *Geolocation Techniques in LoRaWAN Networks as a Fault Tolerance Approach in GPS-Based Tracking Devices*. 2021 Third South American Colloquium on Visible Light Communications (SACVLC). 11-12 November 2021. DOI: 10.1109/SACVLC53127.2021.9652345
6. Xu T., Ma D., Fang W., Huang Y. *Experimental Study on the Propagation Characteristics of LoRa Signals in Maize Fields*. Electronics. 2025. Vol. 14. P. 2156. DOI: 10.3390/electronics14112156
7. Idris S., Karunathilake T., Förster A. *Survey and Comparative Study of LoRa-Enabled Simulators for Internet of Things and Wireless Sensor Networks*. Sensors. 2022. Vol. 22, № 15. P. 5546. DOI: 10.3390/s22155546
8. Shi Y., Shi W., Liu X., Xiao X. *An RSSI Classification and Tracing Algorithm to Improve Trilateration-Based Positioning*. Sensors. 2020. Vol. 20, № 15. P. 4244. DOI: 10.3390/s20154244
9. Bouras C., Gkamas A., Kokkinos A., Papachristos N. *Real-Time Geolocation Approach through LoRa on Internet of Things*. 2021 International Conference on Information Networking (ICOIN). 13-16 January 2021. DOI: 10.1109/ICOIN50884.2021.9333860
10. Semtech. *AN1200.22 LoRa Modulation Basics*. Camarillo, 2015. 26 p.
11. Centenaro M., Vangelista L., Zanella A., Zorzi M. *Long-range communications in unlicensed bands: The rising stars in the IoT and smart city scenarios*. IEEE Wireless Communications. 2016. Vol. 23, № 5. P. 60–67. DOI: 10.1109/MWC.2016.7721743
12. Martin C. B., Roedig U., Voigt T., Juan M. A. *Do lora low-power wide-area networks scale? Analysis and Simulation of Wireless and Mobile Systems*. 2016. P. 59–67. DOI: 10.1145/2988287.2989163
13. Petajajarvi J., Mikhaylov K., Roivainen A., Hanninen T., Pettissalo M. *On the coverage of lpwans: range evaluation and channel*. ITS Telecommunications. 2015. P. 55–59. DOI: 10.1109/ITST.2015.7377400
14. Olivera R., Flores J., Olivera R., Perez J., Munoz J. *Heuristic Approach to Indoor Localization Using LoRa RSSI Measurements*. RADIOENGINEERING. 2025. Vol. 34, № 3. P. 509–525. DOI: 10.13164/re.2025.0509
15. Nahorniuk O. A. *Method of Automatic Parameters Estimation of Radio Signals Generated According to LoRa Standard*. Visnyk NTUU KPI Serii A - Radiotekhnika Radioaparotobuduvannia. 2024. Vol. 95. P. 23–30. DOI: 10.20535/RADAP.2024.95.23-30
16. PremSankar G., Ghaddar B., Slabicki M., Di Francesco M. *Optimal Configuration of LoRa Networks in Smart Cities*. IEEE Transactions on Industrial Informatics. 2020. Vol. 16, № 12. P. 7530–7543. DOI: 10.1109/TII.2020.2967123
17. Toro-Betancur V., PremSankar G., Ghaddar B., Di Francesco M. *Modeling Communication Reliability in LoRa Networks with Device-level Accuracy*. IEEE INFOCOM 2021 - IEEE Conference on Computer Communications. 10-13 May 2021. DOI: 10.1109/INFOCOM42981.2021.9488783

Zylevich M.O., Sorokin D.A. SYSTEMATIZATION OF ERROR SOURCES IN GEOLOCATION USING LORA TECHNOLOGY

This paper investigates the problem of object geolocation in Internet of Things (IoT) networks using LoRa technology, which belongs to the LPWAN class. The growing number of IoT devices and the expansion of their application areas drive the need for energy-efficient and scalable positioning systems. LoRa technology is characterized by long-range data transmission, low power consumption, and relatively low network deployment costs, making it a promising solution for geolocation systems in urban, industrial, and rural environments. However, the use of this technology is accompanied by several factors that affect the accuracy of determining object coordinates.

The study analyzes the primary localization methods in wireless networks, particularly RSSI, ToA, TDoA, and AoA. The feasibility of using the Received Signal Strength Indicator (RSSI) based localization method is substantiated, as it requires no additional hardware and can be implemented in most existing LoRaWAN networks. At the same time, this approach is highly sensitive to environmental influences and the presence of obstacles, leading to distance estimation errors between network nodes.

The paper examines a trilateration-based geolocation algorithm, where the target node's coordinates are determined by the distances to multiple base stations. The process of converting radio signal parameters into object coordinates is analyzed. It is shown that positioning accuracy largely depends on the signal propagation model parameters, which vary based on the type of environment.

As a result of the analysis, the main sources of geolocation errors in LoRa networks are systematized: environmental impact (shadowing and multipath signal propagation), dynamic factors, errors in mathematical models, physical features of radio wave fading, technical characteristics of the equipment, and the geometry of base station placement.

Keywords: *Internet of Things, LoRa, LPWAN, geolocation, trilateration, RSSI, error sources, positioning accuracy, path loss.*

Дата першого надходження статті до видання: 16.03.2026

Дата прийняття статті до друку після рецензування: 13.04.2026

Дата публікації (оприлюднення) статті 11.05.2026

Ілуща А.С.

<https://orcid.org/0009-0007-3445-8481>

Національний університет «Одеська політехніка»

ВЕБ-ОРІЄНТОВАНІ ПРОГРАМНІ ЗАСОБИ ІНТЕГРАЦІЇ PaaS I SaaS ДЛЯ ПІДВИЩЕННЯ ПРОДУКТИВНОСТІ ДОСЛІДНИЦЬКОЇ ТА ОСВІТНЬОЇ ДІЯЛЬНОСТІ

Метою роботи є розробка веб-орієнтованих програмних засобів, що поєднують сервіси PaaS і SaaS та забезпечують підвищення ефективності освітньої діяльності та міждисциплінарних досліджень. Розглянуто проблему організації хмарних обчислень в умовах цифрової трансформації освіти та науки. Показано, що використання локальних обчислювальних ресурсів і розмежування між середовищами програмної розробки та готовими інструментальними засобами у вигляді GUI-інтерфейсів знижує гнучкість, масштабованість і ефективність дослідницького та освітнього процесів. Проведений аналіз існуючих рішень демонструє, що вони не забезпечують комплексного поєднання можливостей розробки програмного коду, створення інтерфейсів та організації спільної роботи в межах єдиного середовища. Розроблено веб-орієнтовані програмні засоби, що забезпечують: одночасну роботу з програмним кодом на мовах програмування Python і JavaScript та з GUI-інтерфейсами; виконання обчислень на сервері; зберігання результатів та їх візуалізацію незалежно від мови програмування; можливість інтеграції сторонніх сенсорів для проведення експериментів та накопичення і повторного використання даних з них; організацію спільної взаємодії користувачів. Реалізовано механізм взаємодії між інтерфейсом і програмною логікою, за рахунок розробленого конструктора інтерфейсів та транслятора коду, що дозволяє змінювати умови виконання досліджень без редагування вихідного коду. Додатково забезпечується підвищення відтворюваності результатів досліджень та адаптивність до різних предметних областей завдяки модульності й гнучкому налаштуванню процесу досліджень, що досягається за рахунок збереження вхідних даних, параметрів виконання та результатів досліджень. Запропоноване рішення забезпечує підвищення продуктивності дослідницької та освітньої діяльності за рахунок інтеграції процесів розробки, виконання обчислень і використання прикладних інструментів, у вигляді GUI-інтерфейсів, в межах єдиного веб-орієнтованого середовища.

Ключові слова: веб-платформа, програмні засоби, дистанційні дослідження, дистанційна освіта, цифрова освіта, PaaS, SaaS, хмарні технології, хмарні обчислення.

Постановка проблеми. Сучасний розвиток освіти й науки відбувається в умовах глибокої цифрової трансформації, що змінює підходи до організації навчального процесу та дослідницької діяльності. Традиційна практика використання локальних обчислювальних ресурсів разом з встановленим програмним забезпеченням дедалі частіше виявляється недостатньою через обмеженість апаратних потужностей, складність технічного супроводу та відсутність гнучкого масштабування. За таких умов можна дедалі частіше спостерігати перехід до хмарних підходів організації дослідницького та освітнього процесів, які забезпечують доступність, масштабованість і адаптивність інструментів проведення досліджень незалежно від місця перебування користувача.

Особливо гостро ця проблема постає у сфері досліджень, пов'язаних з аналізом великих масивів експериментальних даних та моделюванням складних процесів. Наприклад, у нейрофізіологічних і психофізіологічних дослідженнях активно застосовуються технології стеження за рухами очей (eye-tracking) [1–5], яка використовується для вивчення когнітивних процесів [6], наприклад моніторингу стану студентів під час навчання [7]. Такі експерименти супроводжуються накопиченням значних обсягів структурованих і неструктурованих даних, що потребує ефективних механізмів їх зберігання, обробки та повторного використання. Водночас виникає потреба забезпечити спільну роботу дослідників, відтворюваність результатів і гнучке налаштування процесу досліджень без надмірної залежності від конкретної локальної інфраструктури.

Для вирішення подібних завдань доцільне використання хмарних технологій, оскільки такі технології забезпечують централізований доступ до серверів, баз даних, сховищ і програмних інструментів через мережу Інтернет [8–11]. Проте на практиці часто спостерігається розрив між середовищами, призначеними для виконання експериментів, де користувач взаємодіє виключно з готовим інтерфейсом, і середовищами розробки інструментальних засобів обробки даних на рівні програмного коду, що ускладнює формування цілісного науково-освітнього простору. Таким чином, постає завдання створення інтегрованого хмарного рішення, яке б забезпечувало повний цикл роботи з експериментальними даними – від їх отримання до аналітичної обробки та інтерпретації результатів.

Аналіз останніх досліджень і публікацій. Хмарні обчислення розглядаються як багаторівнева організація доступу до обчислювальних ресурсів і програмних сервісів, що реалізується через різні підходи до взаємодії з користувачем: програмне забезпечення як сервіс (SaaS), платформа як сервіс (PaaS), інфраструктура як сервіс (IaaS) та функція як сервіс (FaaS) [12–14]. Кожен із цих підходів орієнтований на певний рівень абстракції – від використання готових програмних продуктів до керування обчислювальною інфраструктурою та виконання окремих функцій. У контексті науково-освітньої діяльності особливу увагу привертає поєднання можливостей програмної розробки інструментальних засобів з використанням готових інструментів керування експериментами (інтерфейсів), що дозволяє поєднати програмну гнучкість із зручністю інтерфейсної взаємодії.

У сфері хмарних обчислень для освітньої та науково-дослідної діяльності активно використовуються такі програмні рішення, як Project Jupyter [15], Google Colab [16], OnlineGDB [17] та MATLAB Online [18]. Незважаючи на широку популярність, зазначені засоби мають низку концептуальних та функціональних обмежень, що ускладнюють їх використання в комплексних міждисциплінарних дослідженнях і в організації повноцінного освітнього процесу.

Project Jupyter є одним із найбільш поширених інструментів для інтерактивної роботи з даними. Він забезпечує створення проєктів (середовищ), у яких поєднуються програмний код та текстові пояснення. Середовище не надає можливостей використання декількох мов програмування та орієнтоване тільки на мову програмування

Python. Крім того, виконання обчислень залежить від локальних ресурсів або потребує окремої конфігурації віддалених серверів. Засоби створення повноцінних графічних інтерфейсів користувача відсутні, а інтеграція з апаратними сенсорами та зовнішнім обладнанням не передбачена. Також не передбачено зберігання результатів досліджень.

Google Colab є веб-орієнтованою реалізацією ідей Jupyter із використанням хмарної інфраструктури Google. Перевагою є доступ до обчислювальних ресурсів та можливість колективної роботи (шеринг проєктів). Разом із тим, середовище повністю орієнтоване на Python. Сесії мають часові обмеження, що створює ризик втрати даних при тривалих експериментах. Функції зберігання результатів досліджень, інструменти створення інтерактивних графічних інтерфейсів та інтеграції з апаратними засобами для проведення експериментів відсутні.

OnlineGDB є багатомовним онлайн-компілятором, який підтримує C, C++, Java, Python, JavaScript та інші мови програмування. Його основна перевага – швидке тестування окремих фрагментів коду без встановлення локального програмного забезпечення. Проте функціональність обмежується саме компіляцією та запуском програм. Засоби колективної роботи, побудови GUI, організації експериментів або підключення апаратних сенсорів не реалізовані. Таким чином, цей інструмент орієнтований переважно на навчальні завдання початкового рівня.

MATLAB Online надає доступ до обчислювальних можливостей MATLAB через веб-браузер. Він ефективний для інженерних розрахунків, обробки сигналів і побудови графіків. Водночас підтримується виключно мова MATLAB, що є недоліком. Можливості створення повноцінних користувацьких графічних інтерфейсів відсутні порівняно з локальною (offline) версією. Засоби інтеграції із зовнішнім обладнанням у веб-версії не передбачені. Колективна взаємодія та зберігання результатів досліджень відсутні.

Порівняльний аналіз показує, що наявні програмні засоби:

- орієнтовані переважно на одну мову програмування;
- не підтримують створення складних користувацьких GUI;
- не передбачають інтеграції з апаратними сенсорами та експериментальним обладнанням;
- не поєднують сервіси PaaS і SaaS для комплексної організації освітньої та дослідницької діяльності у межах єдиного середовища;

– не надають можливості зберігати дані експериментів та результати досліджень.

У результаті жоден із розглянутих засобів повною мірою не забезпечує підтримки міждисциплінарних досліджень, у яких одночасно необхідні:

- робота з програмним кодом кількома мовами;
- створення та використання готових графічних інтерфейсів;
- інтеграція зовнішнього експериментального обладнання;
- централізоване зберігання та обробка даних;
- підтримка взаємодії між дослідниками.

У таблиці 1 наведено порівняння ключових можливостей існуючих рішень організації хмарних обчислень та потреб для підвищення продуктивності досліджень та освітнього процесу.

У результаті аналізу можемо зазначити, що сучасні освітні та наукові процеси потребують програмних засобів, що одночасно надають PaaS та SaaS сервіси для організації дослідницького і освітнього процесів та дозволяють забезпечити:

- роботу з багатомовними проектами;
- створення та використання готових прикладних інтерфейсів;
- інтеграцію з експериментальним обладнанням;
- централізоване зберігання та обробку даних;
- підтримку дослідницької та навчальної діяльності в єдиному інформаційному середовищі.

Постановка завдання. Таким чином, метою роботи є розробка веб-орієнтованих програмних засобів, що поєднують сервіси PaaS і SaaS та підвищують продуктивність дослідницького та освітнього процесів.

Виклад основного матеріалу. Розроблено web-орієнтований програмний комплекс організації хмарних обчислень IONDS (*In On Data Science*) [19] для підвищення ефективності дослідницького та освітнього процесів, що реалізується завдяки поєднанню сервісів PaaS (платформа як сервіс) та SaaS (програмне забезпечення як сервіс). Такий підхід забезпечує дослідникам можливість одночасної роботи з програмним кодом, через вбудований редактор скрипт-коду, та з інтерфей-

сом вже реалізованих інструментальних засобів, який може бути створений завдяки інтегрованому конструктору. Використовуючи редактор коду та конструктор інтерфейсів, користувачі можуть створювати нові або масштабувати та редагувати вже створені інструментальні засоби обробки даних, запускати обчислення на сервері (у хмарі) та отримувати результати безпосередньо у браузері. Розроблені програмні засоби являють собою програмний комплекс, що складається з багатьох модулів і вузлів, які взаємодіють між собою [20].

Однією з ключових складових розроблених програмних засобів є конструктор інтерфейсів, що працює разом з транслятором коду. Їх основне призначення - реалізація механізму декларативного опису параметрів, який дозволяє користувачам без глибоких технічних знань змінювати параметри досліджень без необхідності прямого редагування скрипт-коду інструментальних засобів, а також візуалізувати результати досліджень без залежності від мови програмування [21]. Завдяки цьому забезпечується високий рівень абстракції (доступності), що є критично важливим у міждисциплінарних дослідженнях, де користувачами системи можуть бути представники різних галузей: інженери, студенти або науковці.

Також, розроблений програмний комплекс має можливість підключення до програмних засобів сторонніх сенсорів (датчиків), наприклад камер або біометричних пристроїв, для автоматичного збору даних. Сторонні апаратні засоби (сенсори) під'єднуються безпосередньо до серверу (Рис. 1) завдяки спеціальному програмному API. Він дає змогу передавати дані з будь-якого зовнішнього застосунку безпосередньо у сховище IONDS. Отримана від сенсорів інформація передається на сервер, де вона зберігається в базі даних для подальшого використання у дослідженнях. У процесі передачі дані автоматично структуруються за заданими категоріями та зберігають нативний формат (масиви, об'єкти), що значно полегшує їх подальше використання у проектах. Варто зазначити, що результати експериментів зберігаються для кожного користувача окремо за для забез-

Таблиця 1

Порівняння можливостей існуючих рішень для організації хмарних обчислень та сучасних потреб

	Project Jupyter	Google Colab	OnlineGDB	MATLAB Online	Необхідне рішення
Підтримка кількох мов програмування	Відсутня	Відсутня	Присутня	Відсутня	Присутня
Колективна робота	Відсутня	Присутня	Відсутня	Відсутня	Присутня
Створення та використання GUI-інтерфейсів	Відсутнє	Відсутнє	Відсутнє	Відсутнє	Присутнє
Зберігання даних та результатів	Відсутнє	Відсутнє	Відсутнє	Відсутнє	Присутнє
Інтеграція апаратних сенсорів	Відсутня	Відсутня	Відсутня	Відсутня	Присутня

печення конфіденційності. Для цього необхідно передати, у якості значення поля authors, перелік UID користувачів, для яких результати можуть бути доступними, а також пройти авторизацію, використовуючи спеціальний токен, який видається адміністратором платформи.

Розроблені програмні засоби складаються з семи основних сторінок, які пов'язані між собою (Рис. 2) та дозволяють комплексно проводити дослідження.

Робота з програмними засобами починається з сторінки входу (Рис. 3), на якій користувачі можуть отримати доступ до функцій комплексу, ввівши своє ім'я користувача та пароль, або обравши авторизацію через Google-акаунт.

Після авторизації, користувач потрапляє на головну сторінку комплексу (Рис. 4), яка виконує роль інформаційної панелі, що консолідує всі ключові дані для користувача: статистика виконання проєктів, остання активність акаунту, останні

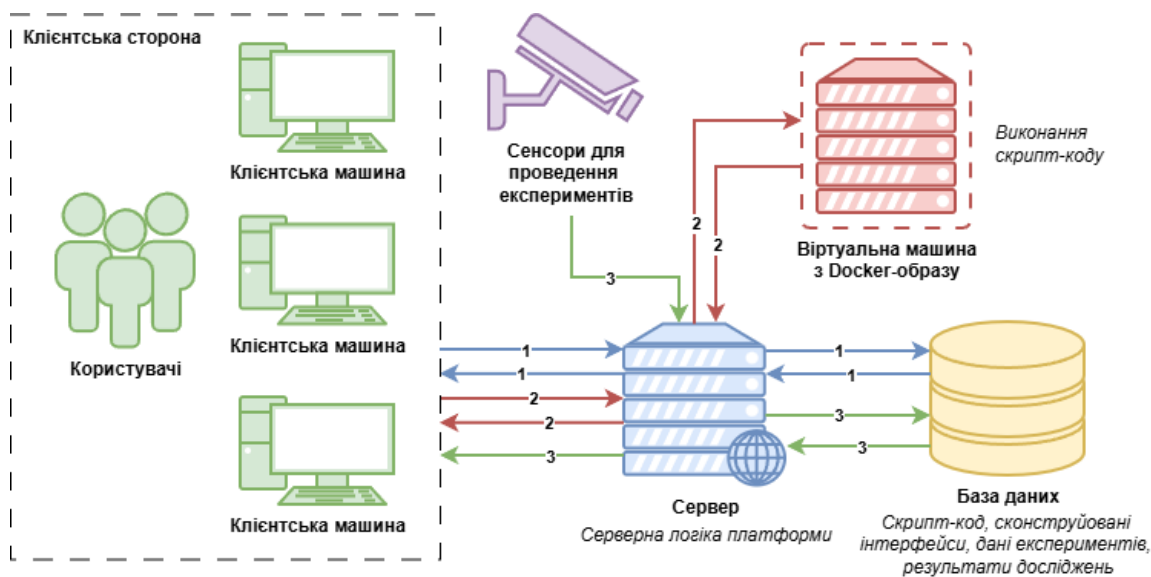


Рис. 1. Схема під'єднання сенсорів до серверної частини комплексу



Рис. 2. Зв'язок основних сторінок комплексу

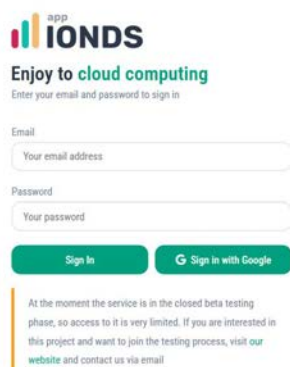


Рис. 3. Сторінка авторизації

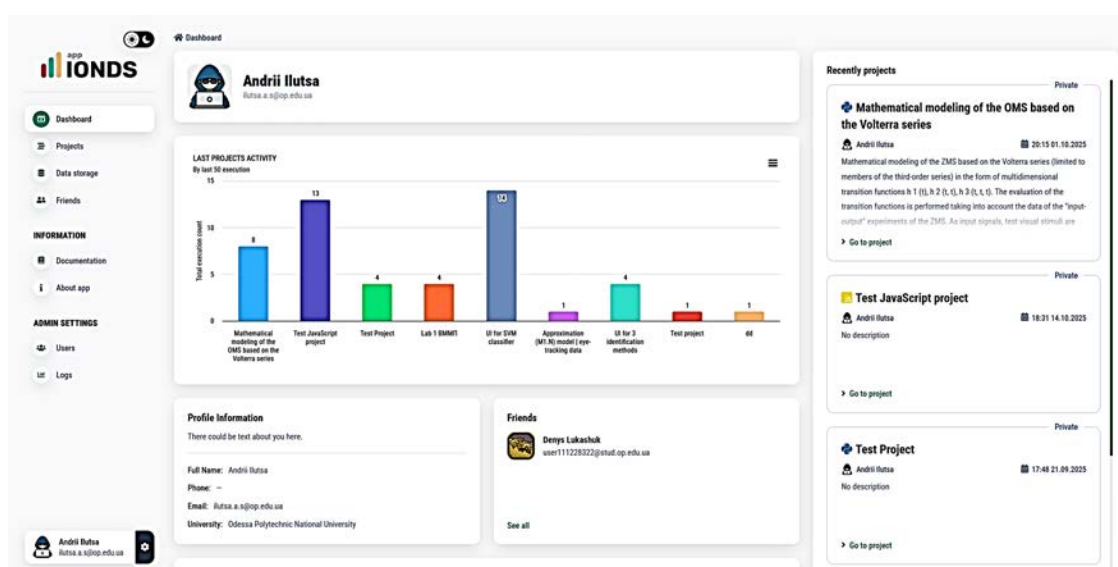


Рис. 4. Головна сторінка комплексу

використані проєкти, перелік обраних акаунтів та особиста інформація профілю.

Також, на головній та всіх інших сторінках користувач має доступ до головного меню комплексу (Рис. 5), яке містить: логотип, перемикач режиму відображення (світлий, темний), секцію користувача (user-box) та перелік посилань на ключові сторінки, які дають змогу швидко отримувати доступ до ключових сторінок комплексу, а саме: головна сторінка, перелік проєктів, сховище даних експериментів, сторінка обраних акаунтів, документація, сайт комплексу.

Після головної сторінки користувач може перейти на сторінку проєктів (Рис. 6), яка надає доступ до переліку публічних проєктів, особистих проєктів користувача та проєктів, якими поділилися з користувачем інші користувачі комплексу. Всі ці проєкти згруповані по відповідним вкладкам, між

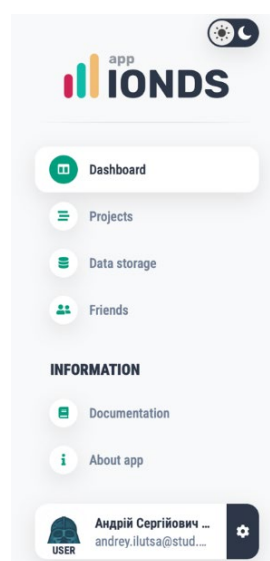


Рис. 5. Головне меню комплексу

якими користувач може перемикатися.

Картка проекту, яка відображається на сторінці, містить короткий опис: назву, дату останнього редагування, автора та короткий опис проекту, якщо його вказано при створенні. На цій сторінці користувачі можуть створювати нові публічні або приватні проекти.

Коли користувач обирає та відкриває певний проект, він потрапляє на його сторінку. На прикладі проекту пов'язаного з оцінкою психофізіологічного стану людини з використанням платформи хмарних обчислень [22] можемо бачити, як виглядає сторінка проекту з створеним інтер-

фейсом. На ній ми можемо обирати файл з результатами експериментів або звернутися до сховища даних експериментів, та взаємодіяти з створеним спеціалізованим інтерфейсом для управління параметрами досліджень, де є можливість керувати параметрами: $\$Identification\_Method\$, \$N\$, \$L\$, \$all\_amplitudes\$, \$base\_amplitudes\$$ (Рис. 7).

Функціональність сторінки проекту залежить від того, чи є користувач автором, чи просто переглядачем. Якщо користувач є автором, він може: редагувати код, конструювати або використовувати інтерфейс для налаштування параметрів досліджень, обирати результати експеримен-

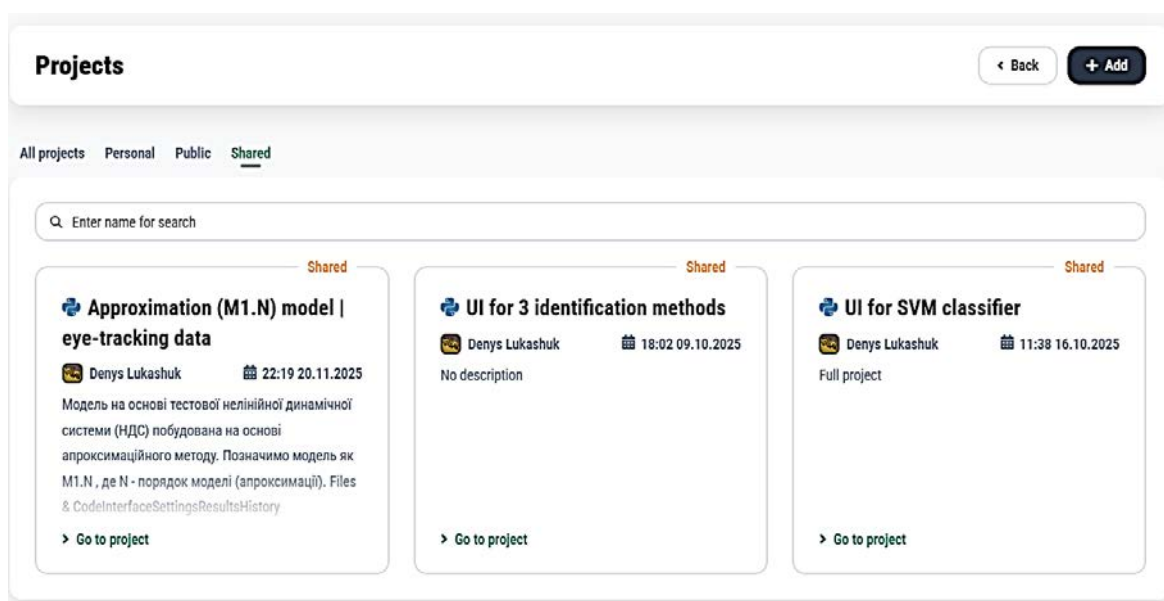


Рис. 6. Сторінка переліку проектів

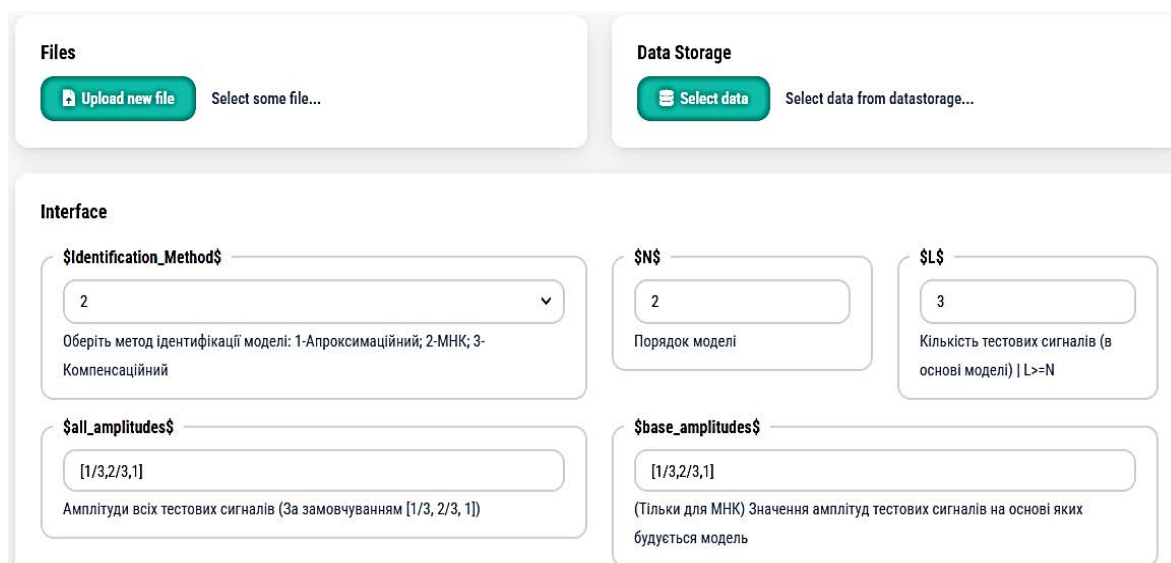


Рис. 7. Сторінка проекту з створеним інтерфейсом

тів з спеціального сховища або завантаживши таблиці, запускати обчислення, переглядати результати досліджень у текстовому та графічному форматах незалежно від мови програмування завдяки розробленому транслятору коду та мові шаблонів перетворень [21]. Якщо ж проєкт належить іншому досліднику і є публічним, користувач не може його редагувати, але все одно має змогу запускати код, змінювати параметри досліджень (якщо для проєкта створено інтерфейс) та переглядати результати у текстовому та графічному форматах. Варто зазначити, що всі результати досліджень, для кожного користувача, зберігаються у спеціальній вкладці на сторінці проєкту, що дає змогу користувачу у будь-який момент повернутися до необхідних результатів для аналізу без потреби виконувати обчислення знову. Таким чином, сторінка проєкта стає основним робочим простором, де дослідження або розробляються розробником, або використовуються дослідником.

Також, приватними проєктами можна ділитися з іншими користувачами, що дозволить проводити спільні дослідження. У таких користувачів ці проєкти відображаються у окремій вкладці, та мають трохи обмежені права управління, на відміну від приватних: немає можливості змінювати опис проєкту та видаляти його.

Окремо від сторінки проєкту, користувач може, через головне меню, потрапити на сторінку сховища даних експериментів (Рис. 8), яка забезпечує централізоване та структуроване накопичення та зберігання цих даних.

Основним завданням сховища є зберігання та організація результатів експериментів у єдиній базі, з якої дані можуть бути повторно використані у різних дослідженнях в межах проєктів комплексу. Данні експериментів зберігаються за принципом категорій (директорій), у межах яких групуються окремі експерименти (Рис. 8). Такий

підхід дає змогу підтримувати логічну організацію великої кількості результатів, завдяки чому дослідник отримує не лише сховище, а й інструмент управління результатами експериментів, що значно спрощує та підвищує продуктивність процесу досліджень. Також, додавати результати експериментів у сховище можна як через спеціальне API, так і через вбудований інтерфейс, заповнюючи всі вище наведені поля у ручному режимі.

Сторінка з налаштуваннями акаунту (сторінка користувача) (Рис. 9), на яку також можна потрапити через головне меню, використовується для ознайомлення та редагування особистої інформації користувача, а саме: фото користувача (аватар), ім'я, телефон, email, заклад освіти, інформація про користувача.

Також, знаходячись на цій сторінці користувач має змогу вивантажувати, або завантажувати файл з результатами досліджень. Для того, щоб вивантажити результати своїх досліджень, користувач повинен натиснути кнопку «Download projects history», після чого на його комп'ютер буде завантажено JSON-файл, у котрому структуровано зберігаються усі результати досліджень. Також, цей файл можна передати іншому користувачу для завантаження (для цього треба натиснути кнопку «Upload history file») та отримання доступу до відповідних результатів досліджень.

Структура даних розроблених програмних засобів організована таким чином, щоб забезпечити повний цикл роботи: збереження результатів експериментів, виконання досліджень на їх основі та фіксацію отриманих результатів. Вона складається з чотирьох основних колекцій: users, projects, data та logs. Кожна з них відповідає за окремий функціональний рівень.

Колекція users забезпечує управління обліковими записами та має наступну структуру:

- id – унікальний ідентифікатор користувача;
- created – дата створення запису;

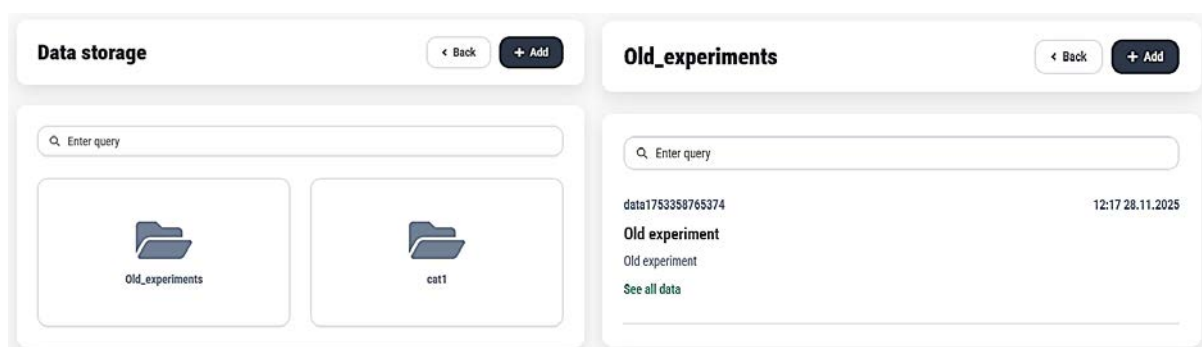


Рис. 8. Сторінки сховища даних експериментів

- updated – дата оновлення запису;
- email – електронна адреса;
- password – пароль;
- role – роль користувача;
- status – статус облікового запису;
- avatar – зображення профілю;
- bio – інформація про користувача;
- projects – проекти користувача;
- projectHistory – результати попередніх досліджень;

- activity – історія активності;
- friends – список пов'язаних користувачів.

Основні аутентифікаційні поля – id, email, password. Для управління доступністю та організації прав доступу застосовується поля status і role. Часові параметри змін акаунту зберігаються в created та updated. Окрім базових атрибутів, структура містить JSON-поля activity та bio, які містять структуру з відповідною інформацією. Поля projects та friends – це колекції з відповідними ідентифікаторами пов'язаних проектів та користувачів. Поле projectsHistory містить структуру з результатами попередніх досліджень, та є JSON-об'єктом, у якому ключами виступають ідентифікатори проектів, а значенням кожного ключа є масив записів виконань цього проекту. Кожний запис про виконання містить:

- id – унікальний ідентифікатор виконання;
- params – параметри експерименту, використаних під час запуску;
- results – результати виконання.

Колекція projects є центральною частиною структури. Вона об'єднує програмну реалізацію, налаштування інтерфейсу та механізми спільної роботи. Кожен проект містить:

- id – унікальний ідентифікатор проекту;

- author – ідентифікатор автора проекту (користувача);

- codeLang – мова програмування;

- date – дата останнього редагування користувачем;

- description – опис проекту;

- interface – структура елементів інтерфейсу;

- params – параметри запуску;

- shared – список користувачів із доступом до проекту;

- sourceCode – програмний код проекту;

- status – рівень відкритості проекту (приватний або публічний);

- title – назва проекту;

- created – дата створення;

- updated – дата останнього оновлення.

Особливістю структури є поле interface. Воно містить масив об'єктів, які описують елементи графічного інтерфейсу. Кожен елемент має поля name, description, type, value, valuesTemplate. Поле type визначає тип елемента керування (наприклад, select, single, multiple). Поле value зберігає поточне значення параметра, а valuesTemplate – перелік допустимих значень, якщо це потрібно. Таким чином, поле interface описує, як повинен виглядати інтерфейс параметризації досліджень. На його основі формується інтерфейс, а введені в нього параметри зберігаються у полі params, та використовуються програмним кодом з поля sourceCode. Це дозволяє виконувати дослідження з різними параметрами без зміни коду.

Колекція data призначена для збереження результатів експериментів, які використовуються як вхідні дані при проведенні досліджень (виконанні проектів). Вона містить поля:

- id – унікальний ідентифікатор запису;

Dashboard > Profile

Andrii Ilutsa Back Save

[Change](#)

[Download projects history](#)

or

[Upload history file](#)

User information

UID:

Full Name: Phone: Email:

University:

About me:

[Logout](#)

Рис. 9. Сторінка керування профілем користувача

- authors – список авторів експерименту;
- category – категорія даних;
- date – дата проведення експерименту;
- description – опис експерименту;
- isPublic – ознака відкритого доступу;
- title – назва;
- value – результати експерименту;
- created – дата створення запису;
- updated – дата останнього оновлення.

Поле value є універсальним контейнером для зберігання експериментальних даних. У ньому можуть зберігатися числові дані, масиви, вкладені об'єкти або інші параметри. Саме ці дані, у разі необхідності, використовуються як вхідні при виконанні досліджень. Таким чином, data виконує роль сховища експериментальних даних, на основі яких проводяться подальші дослідження.

Колекція logs забезпечує фіксацію запуску та виконання задач. Основні поля:

- id – унікальний ідентифікатор запису;
- err – інформація про помилку (якщо була);
- projectId – ідентифікатор проєкту;
- taskId – ідентифікатор задачі виконання;
- uid – внутрішній ідентифікатор користувача;
- created – дата створення запису;
- updated – дата останнього оновлення.

Варто зазначити, що сам запис логування створюється під час запуску виконання проєкту, а вже після отримання результатів він оновлюється, додаючи до себе ідентифікатор задачі виконання на сервері (taskId) та помилку (err), якщо така була.

У результаті, ця колекція дозволяє відстежувати процес виконання та аналізувати технічні збої.

Загалом, структура даних, наведена на рисунку 10, підтримує послідовний цикл дослідження:

1. Експериментальні дані зберігаються у колекції data.

2. Користувач user створює або налаштовує проєкт у projects, задаючи параметри через інтерфейс, створений на основі поля interface, які зберігаються у полі params.

3. Проєкт виконується з використанням даних із data.

4. Процес виконання фіксується у logs.

5. Результат дослідження зберігається у полі projectsHistory користувача user.

Таким чином, структура даних охоплює всі етапи роботи: накопичення вхідних експериментальних даних, їх обробку, отримання та збереження отриманих результатів, що забезпечує повний цикл дослідницької діяльності.

Висновки. Розроблено веб-орієнтовані програмні засоби, що поєднують сервіси PaaS і SaaS та забезпечують комплексну організацію хмарних обчислень у межах єдиного середовища, що підвищує продуктивність досліджень на освітнього процесу. Було обґрунтовано необхідність інтеграції, в рамках одного середовища, інструментів програмної розробки та готових інструментальних засобів (у вигляді GUI-інтерфейсів), оскільки їх розмежування в існуючих рішеннях знижує гнучкість, масштабованість і загальну ефектив-

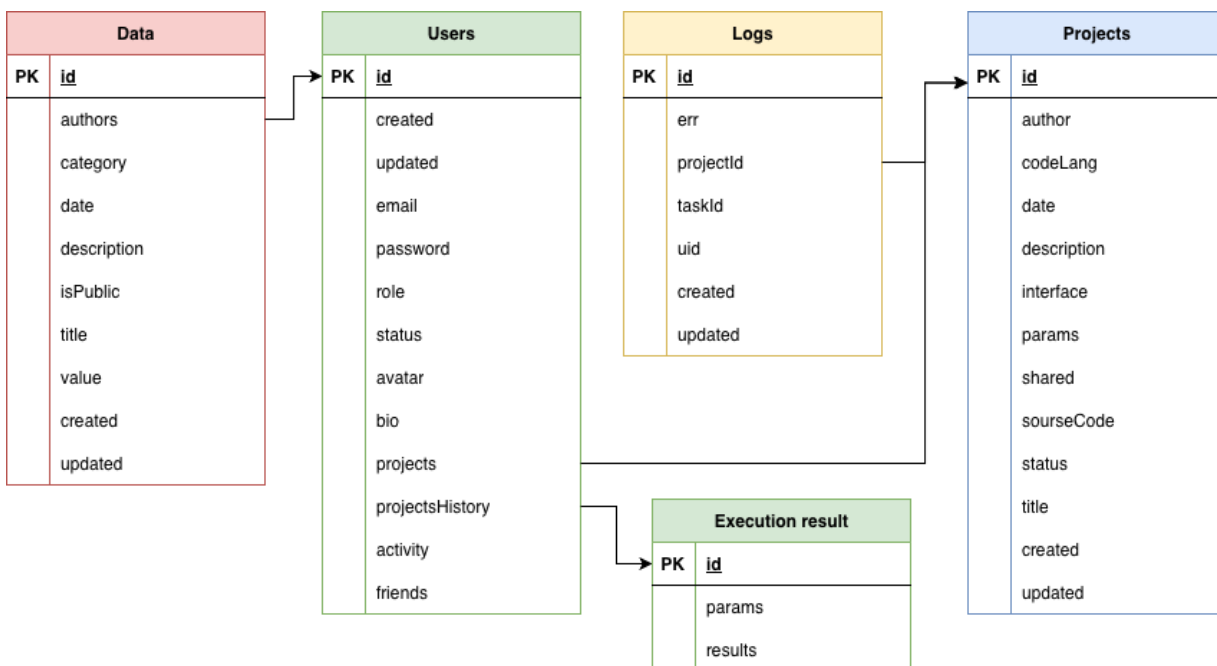


Рис. 10. Структура даних розроблених програмних засобів

ність міждисциплінарних досліджень та освітньої діяльності.

Проведений аналіз сучасних засобів організації хмарних обчислень показав, що вони не забезпечують комплексного поєднання роботи з даними та програмним кодом, створення GUI-інтерфейсів і підтримки спільної взаємодії користувачів у межах одного веб-орієнтованого рішення. Це обмежує можливості створення прикладних інструментів, ускладнює їх масштабування та повторне використання.

Розроблені програмні засоби реалізують інтегроване середовище, у якому забезпечено: збереження результатів експериментів та їх подальше використання, одночасну роботу з програмним кодом і графічними інтерфейсами, виконання обчислень на сервері, збереження та візуалізацію результатів досліджень. Ключовим результатом

є реалізація механізму взаємодії між інтерфейсом і програмною логікою завдяки конструктору інтерфейсів та транслятору коду, що дозволяє змінювати умови виконання досліджень без внесення змін до скрипт-коду. Такий підхід забезпечує відокремлення логіки реалізації від засобів керування параметрами виконання та підвищує зручність використання інструментів у міждисциплінарних дослідженнях. Це дозволяє розробляти інструментальні засоби та одночасно забезпечувати їх зручне використання без створення окремих застосунків для налаштування параметрів і без втручання у вихідний код під час зміни умов виконання. У результаті скорочується час підготовки та модифікації досліджень, спрощується адаптація інструментальних засобів до нових задач і підвищується продуктивність дослідницького та освітнього процесів.

Список літератури:

1. Opwonya J., Doan D.N.T., Kim S.G., Kim J., Ku B., Kim S., Park S., Kim J.U., Saccadic Eye Movement in Mild Cognitive Impairment and Alzheimer's Disease: A Systematic Review and Meta-Analysis, *Neuropsychology Review*, Vol.32, 2022, pp. 193–227. DOI: <https://doi.org/10.1007/s11065-021-09495-3>.
2. Jansson D., Rosén O., and Medvedev A., Non-parametric analysis of eye-tracking data by anomaly detection, *IEEE Transaction control system technology*, Vol.23, 2015, pp. 1578–1586. DOI: <https://doi.org/10.23919/ECC.2013.6669561>.
3. Bro V., Medvedev A., Continuous and Discrete Volterra-Laguerre Models with Delay for Modeling of Smooth Pursuit Eye Movements, *IEEE Transactions on Biomedical Engineering*, Vol.70, No.1, 2023, pp. 97–104. DOI: <https://doi.org/10.1109/TBME.2022.3185669>.
4. Lanata L., Sebastian L., Di Gruttola F., Di Modica S., Scilingo E.P., and Greco1 A., Nonlinear Analysis of Eye-Tracking Information for Motor Imagery Assessments, *Frontiers in Neuroscience*, Vol.13, 2020, pp. 1431. DOI: <https://doi.org/10.3389/fnins.2019.01431>.
5. Pavlenko V., Milosz M., Dzienkowski M., Identification of the oculo-motor system based on the Volterra model using eye tracking technology, 4th Int. Conf. on Applied Physics, *Simulation and Computing (APSAC'2020) 2020, 23-25 May, Rome, Italy. Journal of Physics: Conference Series 1603*, pp. 1–8. DOI: <https://doi.org/10.1088/1742-6596/1603/1/012011>.
6. Keehn B., Monahan P., Enneking B., Ryan T., Swigonski N., Keehn R.M., Eye-Tracking Biomarkers and Autism Diagnosis in Primary Care, *JAMA Netw Open*, Vol.7, No.5, 2024, pp. 1-14. DOI: <https://doi.org/10.1001/jamanetworkopen.2024.11190>.
7. Weiss K., Kolbe M., Lohmeyer Q., and Meboldt M., Measuring teamwork for training in healthcare using eye tracking and pose estimation, *Front. Psychol. Sec. Organizational Psychology*, Vol.14, 2023, pp. 1–12. DOI: <https://doi.org/10.3389/fpsyg.2023.1169940>.
8. Saraswat M., Tripathi R. C., Cloud computing: Analysis of top 5 CSPs in SaaS, PaaS and IaaS platforms, *9th International Conference System Modeling and Advancement in Research Trends (SMART)*, 2020, pp. 300-305. DOI: <https://doi.org/10.1109/SMART50582.2020.9337157>.
9. Kollipara P., An Overview on Cloud Computing, *International Journal of Research in Engineering, Science and Management*, Vol.4, No.8, 2021, pp. 35-37. URL: <https://journal.ijresm.com/index.php/ijresm/article/view/1147>.
10. King T. S., Reviews of Cloud Computing For Education: Services And Benefits, *PEOPLE: International Journal of Social Sciences*, Vol.1, No.1, 2015, pp. 1299-1305. DOI: <https://doi.org/10.20319/pijss.2015.s21.12991305>.
11. Singh A., Sharma S., Kumar S. R., Yadav S. A., Overview of PaaS and SaaS and its application in cloud computing, *International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH)*, 2016, pp. 172-176. DOI: <https://doi.org/10.1109/ICICCS.2016.7542322>.
12. Mohammed C. M., Zeebaree, S. R., Sufficient comparison among cloud computing services: IaaS, PaaS, and SaaS, *International Journal of Science and Business*, Vol.5, No.2, 2021, pp. 17-30. URL: <https://ijsab.com/wp-content/uploads/667.pdf>.

13. Wulf F., Lindner T., Strahringer S., Westner M., IaaS, PaaS, or SaaS? The why of cloud computing delivery model selection. Vignettes on the post-adoption of cloud computing, *Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS-54)*, 4-9 January, Lahaina, Hawaii, USA, 2021, pp. 6285-6294. URL: https://aisel.aisnet.org/hicss-54/os/practice-based_research/5/.
14. Nadeem F., Evaluating and ranking cloud IaaS, PaaS and SaaS models based on functional and non-functional key performance indicators, *IEEE Access*, Vol.10, 2022, pp. 63245-63257. DOI: <https://doi.org/10.1109/ACCESS.2022.3182688>.
15. Project Jupyter. URL: <https://jupyter.org>.
16. Google Colab. URL: <https://colab.research.google.com>.
17. OnlineGDB. URL: <https://www.onlinegdb.com>.
18. MATLAB Online. URL: <https://www.mathworks.com/products/matlab-online.html>.
19. IONDS – Cloud computing platform. URL: <https://about.ionds.app>.
20. Павленко В.Д., Ілуща А.С., Гідулян В.І., Поєднання сервісів PaaS та SaaS на платформі для хмарних обчислень для застосувань у нейронауках, *Вісник ДНУ імені Олеся Гончара, Серія: Ракетно-космічна техніка*, 34(2), 2025, С. 131-139. DOI: <https://doi.org/10.15421/452522>.
21. Pavlenko V., Ilutsa A., Integration of PaaS and SaaS services in cloud computing software. *Monograph: "Intellectual and technological potential of the XXI century 2025"*, Vol.1, 2025, pp. 50–61. DOI: <https://doi.org/10.30890/2709-2313.2025-43-01>.
22. Pavlenko V., Lukashuk D., Ilutsa A., *Assessment of Psychophysiological Status Using a Cloud Computing Platform*, *SWorldJournal*, Vol.33, No.1, 2025, pp. 196–207. DOI: <https://doi.org/10.30888/2663-5712.2025-33-01-124>.

Ilutsa A.S. WEB-ORIENTED SOFTWARE INTEGRATION OF PAAS AND SAAS TO INCREASE THE PRODUCTIVITY OF RESEARCH AND EDUCATIONAL ACTIVITIES

The purpose of the work is to develop web-oriented software tools that combine PaaS and SaaS services and ensure increased efficiency of educational activities and interdisciplinary research. The problem of organizing cloud computing in the conditions of digital transformation of education and science is considered. It is shown that the use of local computing resources and the separation between software development environments and ready-made tools in the form of GUI interfaces reduces the flexibility, scalability and efficiency of research and educational processes. The analysis of existing solutions showed that they do not provide a comprehensive combination of opportunities for developing software code, creating interfaces and organizing joint work within a single environment. Web-oriented software tools have been developed that provide: simultaneous work with software code in the Python and JavaScript programming languages and GUI interfaces; performing calculations on the server; storing results and their visualization regardless of the programming language; the ability to integrate third-party sensors for conducting experiments and accumulating and reusing data from them; organization of joint user interaction. A mechanism for interaction between the interface and program logic has been implemented, due to the developed interface designer and code translator, which allows changing the conditions of research without editing the source code. Additionally, increased reproducibility of research results and adaptability to different subject areas are ensured due to modularity and flexible configuration of experimental processes, which is achieved by saving input data, execution parameters and research results. The proposed solution provides increased productivity of research and educational activities by integrating the development processes, performing calculations and using application tools in the form of GUI interfaces within a single web-oriented environment.

Keywords: web-platform, software, remote research, remote education, digital education, PaaS, SaaS, cloud technologies, cloud computing.

Дата першого надходження статті до видання: 17.03.2026

Дата прийняття статті до друку після рецензування: 15.04.2026

Дата публікації (оприлюднення) статті 11.05.2026

Іляш Ю.Ю.

<https://orcid.org/0009-0006-5786-7205>

Карпатський національний університет імені Василя Стефаника

Горєлов В.О.

<https://orcid.org/0000-0002-2106-8704>

Карпатський національний університет імені Василя Стефаника

Антоняк В.С.

<https://orcid.org/0009-0004-2076-5606>

Карпатський національний університет імені Василя Стефаника

АЛГОРИТМИ КЛІТИННИХ АВТОМАТІВ ДЛЯ ГЕНЕРАЦІЇ РІВНІВ У КОМП'ЮТЕРНИХ ІГРАХ

Стаття присвячена дослідженню алгоритмів процедурної генерації ігрових рівнів у 2D roguelike-іграх на основі клітинних автоматів та методів морфологічної обробки двовимірних дискретних структур. Обґрунтовано актуальність використання стохастичних і детермінованих алгоритмічних підходів для формування варіативних ігрових середовищ, що забезпечують високу реіграбельність, адаптивність та масштабованість ігрового процесу. Розглянуто особливості застосування клітинних автоматів із використанням околу Мура та вплив параметрів ініціалізації (щільність заповнення, кількість ітерацій згладжування) на топологічні характеристики згенерованих карт.

Запропоновано модульну архітектуру генератора рівнів, яка включає початкову стохастичну ініціалізацію карти, ітеративне згладжування за правилом сусідства Мура, видалення ізольованих регіонів за допомогою алгоритму пошуку в глибину, побудову коридорів між кімнатами на основі графових зв'язків та морфологічне розширення прохідних областей. Додатково враховано можливість генерації другорядних об'єктів ігрового середовища на основі тих самих принципів клітинних автоматів, що забезпечує структурну узгодженість рівня.

Практичну реалізацію виконано в рушії Unity мовою C# із використанням Tilemap-системи. Проте, завдяки універсальності підходу, подані алгоритми легко адаптуються до інших мов програмування та середовищ розробки, зокрема C++. Показано, що поєднання клітинних автоматів із додатковими топологічними та морфологічними операціями дозволяє формувати природоподібні печерні структури, характерні для жанру roguelike, при збереженні повної відтворюваності результатів через використання параметра seed.

Проведений аналіз демонструє, що варіювання параметрів генерації безпосередньо впливає на зв'язність рівня, розмір і форму ігрових просторів, а також на баланс між відкритими зонами та перешкодами. Отримані результати можуть бути використані як основа для систем навчання з підкріпленням та інших методів машинного навчання, що працюють із процедурно згенерованими середовищами та як базис для подальших досліджень із використанням методів машинного навчання в генеративному геймдизайні.

Ключові слова: процедурна генерація, клітинні автомати, морфологічні операції, Unity, Tilemap, C#, C++, компонент зв'язності, алгоритм, згладжування.

Постановка проблеми. Сучасні тенденції розвитку комп'ютерних ігор засвідчують, що користувачі очікують не лише цікавий сюжет, або красиву графіку, а насамперед – варіативності в методах проходження гри, та в отриманні кожного разу унікального ігрового досвіду.

Представники жанру roguelike, характеризуються високим рівнем варіативності ігрового процесу, що досягається шляхом використання процедурної генерації контенту (Procedural Content Generation, PCG) – автоматичне створення ігрових рівнів, які повинні відповідати низці вимог: забез-

печувати зв'язність простору, логічну структуру, збалансовану складність, можливість навігації гравця та коректне розміщення ігрових об'єктів (ворогів, предметів тощо).

При використанні випадкових методів генерації без додаткової структурної обробки виникає низка проблем. Зокрема, карта може містити ізольовані області, надто вузькі проходи, непридатні для ігрової механіки зони або геометрично некоректні конфігурації. Такі дефекти негативно впливають на ігровий процес та можуть призвести до ситуацій, коли рівень є непрохідним або порушує баланс гри.

Автори Лоуренс Джонсон, Юліан Торгеліус та Георгіс Яннакакіс в [1], та Метью Крайцер, Деніел Ешлок та Раджеш Перейра в [10] розглядають класичні алгоритми ручного або напівручного проектування рівнів, такі як метод фіксованих шаблонів, алгоритми розбиття простору типу Binary Space Partitioning (BSP), а також побудова рівнів на основі заздалегідь визначених кімнат і коридорів забезпечують повний контроль над структурою простору. Однак описані алгоритми є ресурсоемними та не дозволяють створювати велику кількість унікальних варіантів. Таким чином, виникає необхідність у розробці алгоритмічного підходу, який поєднує випадковість із формальними правилами структурної обробки.

Особливо актуальною є задача формування природоподібних печерних структур, характерних для roguelike-ігор, із забезпеченням топологічної зв'язності всієї карти. Додатково постає проблема інтеграції алгоритму генерації в середовищ розробки із можливістю масштабування, модифікації параметрів та детермінованого відтворення результатів через використання початкового зерна випадковості (seed).

Аналіз останніх досліджень і публікацій. Процедурна генерація контенту (PCG) у іграх визначається як автоматична або комп'ютерно-підтримувана генерація контенту – рівнів, ландшафтів, предметів, правил, квестів тощо. Нур Шейкер, Юліан Тогеліус та Марк Джуніор Нельсон підкреслюють, що PCG охоплює різні класи алгоритмів – фрактальні, граматичні, пошукові та еволюційні методи, методи на основі обмежень, а також генерацію нарративу, ландшафтів і підземель [1]. Серед цих підходів клітинні автомати (далі по тексту КА) є простим конструктивним методом генерації рівнів, здатним створювати лабіринтоподібні структури доволі простою схемою правил.

Лоуренс Джонсон, Юліан Торгеліус та Георгіс Яннакакіс [1,2] продемонстрували ефективність

застосування КА для генерації печерних рівнів у реальному часі. Алгоритм базується на випадковій ініціалізації сітки та подальшому ітеративному застосуванні порогового правила, що дозволяє регулювати щільність стін і ширину тунелів. Попри низьку обчислювальну складність, базовий підхід потребує додаткових процедур забезпечення зв'язності, що свідчить про необхідність постобробки структури.

У роботі Юрія Іляша, Віталія Горелова та Віктора Ровінського [3] розглянуто алгоритми та методи процедурної генерації ігрового контенту та адаптивної складності рівнів у 2D-платформерах. Автори пропонують підхід, який орієнтований переважно на відкриті або платформні простори та враховує зміну складності відповідно до поведінки гравця. Водночас така модель менш придатна для формування замкнутих печероподібних структур із жорсткими вимогами до топологічної зв'язності.

У працях Ендрю Печа, Філіпа Хінгстона, Мартіна Масека та Чіу Пен Лама [4,5] досліджено еволюційний підбір правил клітинних автоматів для генерації рівнів із заданими властивостями. Запропоновано використання генетичних алгоритмів для автоматичного налаштування локальних правил або відтворення топології зразкових карт. Однак такі підходи є обчислювально складнішими, що знижує їх практичність для легковагових рушіїв і швидкої інтеграції у проекти з обмеженими ресурсами.

У дослідженні Чада Адамса, Хірава Парекха та Сушила Луїса [6] запропоновано інтерактивну еволюцію правил КА з участю користувача. Алгоритм адаптується на основі суб'єктивної оцінки згенерованих рівнів, що підвищує персоналізацію процесу. Водночас залежність від участі користувача обмежує автономність генератора та ускладнює автоматизоване тестування.

Ізабелла Антонюк в [7] розглядає гібридний підхід, у якому КА поєднуються з попередньо заданим дизайнерським “скелетом” рівня. Це дозволяє зберігати контроль над глобальною структурою карти, використовуючи природність форм КА. Однак такий метод частково повертає процес до напівручного проектування та залежить від попередньої структурної схеми.

У вітчизняному дослідженні проведеному Оленою Трофименко, Олександром Задарейком та співавторами [8], КА розглядаються як окремий клас решіткових алгоритмів PCG поряд із шумовими та графовими методами. Підкреслюється їх здатність формувати органічні структури при

низькій обчислювальній складності, однак зазначається проблема ізольованих областей і необхідність додаткових механізмів контролю.

Лазарос Лазарідіс та Джордж Ф. Фрагіліс [9] запропонували удосконалений конструктивний алгоритм на базі КА з автоматичним розміщенням кімнат і перевіркою колізій. Підхід забезпечує високу варіативність і придатність карт до використання без додаткового тестування. Водночас алгоритм орієнтований на модульну кімнатну структуру, що відрізняється від безперервних печероподібних просторів, характерних для roguelike-середовищ.

У праці Метью Крайцера, Деніела Ешлока та Раджеша Перейра [10] досліджено застосування клітинних автоматів для генерації ігрових рівнів із подальшим автоматизованим оцінюванням їх ігрових властивостей. Автори акцентують увагу на поєднанні конструктивної генерації з метриками придатності рівня (прохідність, зв'язність, просторовий баланс), що дозволяє формалізувати критерії якості. Підхід демонструє можливість інтеграції КА у більш комплексні системи контролю якості контенту, однак передбачає додатковий етап аналітичної перевірки, що ускладнює імплементацію підходу в прості проекти.

Огляд сучасних досліджень свідчить про зростання інтересу до інтеграції різноманітних алгоритмічних підходів у системи процедурної генерації контенту. Більшість робіт поєднують стохастичні механізми з еволюційними, графовими або адаптивними методами для підвищення варіативності та керованості структури рівнів. Водночас існує потреба у простому, обчислювально ефективному та параметризованому алгоритмі, здатному формувати замкнуті 2D-простори з гарантованою зв'язністю без складної еволюційної оптимізації або ручного втручання. Саме на вирішення цієї задачі спрямовано запропонований у даній роботі підхід.

Постановка завдання. Таким чином, розробка алгоритмів процедурної генерації рівнів на основі клітинних автоматів є перспективним напрямом у сфері створення ігрових середовищ, зокрема для жанру roguelike. Поєднання стохастичних методів ініціалізації з формальними правилами еволюції клітинних структур дозволяє формувати природоподібні, структурно зв'язні та варіативні карти з можливістю масштабування складності. У цьому контексті створення модульного алгоритму генерації 2D-рівнів із використанням клітинних автоматів та подальшої морфологічної обробки може розглядатися як практична реалізація сучасних

підходів до процедурного геймдизайну та основа для подальших досліджень у сфері адаптивних ігрових систем і навчальних середовищ.

Виклад основного матеріалу. У межах дослідження було реалізовано алгоритм процедурної генерації 2D-рівнів для гри жанру roguelike на основі клітинних автоматів із подальшою структурною обробкою отриманої карти та генерацією ігрового наповнення. На відміну від класичних підходів, що обмежуються лише локальними правилами трансформації клітин, запропонований алгоритм доповнено механізмами постобробки, перевірки зв'язності та модульною системою розміщення додаткових об'єктів. Це дозволяє формувати топологічно коректні карти з керованими параметрами щільності, складності та варіативності простору. Запропонований підхід забезпечує поєднання стохастичної різноманітності з контролем структурної цілісності, що підвищує природність сформованої геометрії рівня та придатність карт для практичного використання в ігровому середовищі. Рівень представлено у вигляді двовимірної дискретної сітки розміром $W \times H$, де кожна клітина може перебувати в одному з двох станів: прохідна область або стіна. Таким чином, карта інтерпретується як бінарна матриця, що описує геометрію ігрового простору.

Початковий етап генерації полягає у стохастичному заповненні карти. Для кожної клітини визначається початковий стан відповідно до ймовірності r появи стіни: $P(S_0(x, y) = 1) = r$, де $S_0(x, y)$ – початковий стан клітини. Граничні клітини примусово ініціалізуються як стіни з метою формування замкненого простору. Для забезпечення відтворюваності результатів використовується початкове зерно випадковості (seed), що дозволяє отримувати однакову карту при повторному запуску алгоритму з однаковими параметрами.

У спрощеному вигляді програмна реалізація алгоритму за допомогою мови C# у середовищі Unity матиме такий вигляд: `public static int[,] Generate(int width, int height, string seed, bool useRandomSeed, int randomFillPercent, int smoothPasses)`

```
{
    int[,] map = new int[width, height];
    RandomFillMap(map, width, height, seed, useRandomSeed,
        randomFillPercent);
    for (int i = 0; i < smoothPasses; i++){
        SmoothMapStep(map);
    }
    return map;
}
```

Подальше формування структури рівня здійснюється за допомогою клітинного автомата із сусідством Мура, яке враховує вісім навколишніх клітин. На кожній ітерації обчислюється кількість сусідніх клітин у стані «стіна», після чого застосовується порогове правило переходу:

$$S_{t+1}(x, y) = \begin{cases} 1, & N(x, y) \geq T \\ 0, & N(x, y) < T \end{cases}$$

де $N(x, y)$ – кількість сусідів зі станом «стіна», а T – порогове значення. Ітеративне застосування цього правила протягом декількох кроків призводить до згладжування випадкових шумових фрагментів і формування печероподібної структури, характерної для ігор жанру roguelike. Практичні експерименти показали, що вибір параметрів у та 4–6 ітерацій забезпечує збалансоване співвідношення між щільністю стін і відкритих просторів.

У спрощеному вигляді програмна реалізація алгоритму за допомогою мови C# у середовищі Unity матиме такий вигляд:

```
static void SmoothMapStep(int[,] map){
    int width = map.GetLength(0);
    int height = map.GetLength(1);
    int[,] newMap = (int[,])map.Clone();
    for (int x = 0; x < width; x++){
        for (int y = 0; y < height; y++){
            int count = GetSurroundingWallCount(map, x, y);
            if (count > 4) newMap[x, y] = 1;
            else if (count < 4) newMap[x, y] = 0;
            else newMap[x, y] = map[x, y];
        }
    }
    for (int x = 0; x < width; x++){
        for (int y = 0; y < height; y++){
            map[x, y] = newMap[x, y];
        }
    }
}
```

Архітектура генерації побудована за модульним принципом, де кожен компонент відповідає за окрему стадію побудови рівня. Кожен етап працює незалежно, використовуючи спільні параметри, що передаються через об'єкт MapGeneratorManager. Такий підхід забезпечує легку масштабованість і можливість заміни будь-якого модуля без впливу на решту системи.

Основні модулі системи:

CellularAutomata: основний модуль, який генерує карту на основі заданих параметрів (розмір карти, зерно випадковості, відсотка початкового заповнення та кількості ітерацій згладжування).

RegionProcessor: відповідає за пошук компонент зв'язності та фільтрацію малих областей.

RoomDetector: визначає регіони, які пройшли фільтрацію через модуль RegionProcessor.

CorridorBuilder: здійснює побудову тунелів між регіонами карти, у випадку якщо вони є достатньо великими, але не пов'язані одне з одним.

EnemySpawner: відповідає за створення ворогів у межах одного регіону.

ObstacleSpawner: відповідає за створення перешкод у межах одного регіону.

У процесі експериментального дослідження встановлено, що вибір параметра початкової щільності заповнення (randomFillPercent) та кількості ітерацій згладжування (smoothPasses) суттєво впливає на топологію карти.

За низьких значень коефіцієнта (менше 45%) формуються великі відкриті області з недостатньо структурною варіативністю. За надмірно високих значень (понад 55%) утворюється велика кількість дрібних ізольованих замкнених областей та непрохідних зон.

За відсутності згладжування карта має хаотичну структуру з великою кількістю дрібних артефактів та різких переходів між станами клітин. Така конфігурація може випадково створювати природні перешкоди, однак не забезпечує стабільної геометричної цілісності. Зі збільшенням кількості ітерацій згладжування відбувається агрегація стін у більш масиви та усунення поодиноких клітин, що формує більш плавні та природні контури простору. Водночас надмірна кількість ітерацій призводить до надмірного укрупнення структур і втрати варіативності карти.

Найбільш збалансована структура спостерігається у діапазоні 46–53%, за умови виконання 5–8 ітерацій згладжування. Саме в цьому інтервалі забезпечується поєднання природності форм, достатньої зв'язності та ігрової прохідності простору.

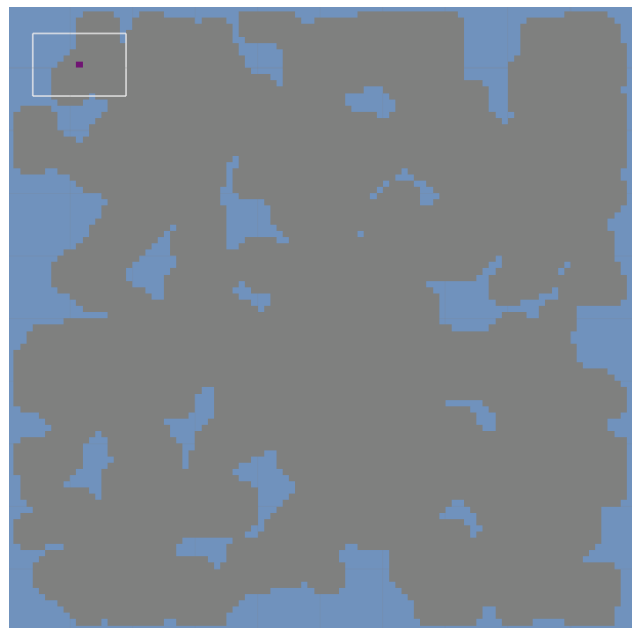


Рис. 1. Приклад згенерованої мапи, за низького значення параметра randomFillPercent

Таким чином, розроблений алгоритм поєднує стохастичну генерацію, ітеративну дискретну модель клітинного автомата та подальшу структурну обробку, що забезпечує формування варіативних, зв'язних і придатних до використання ігрових середовищ.

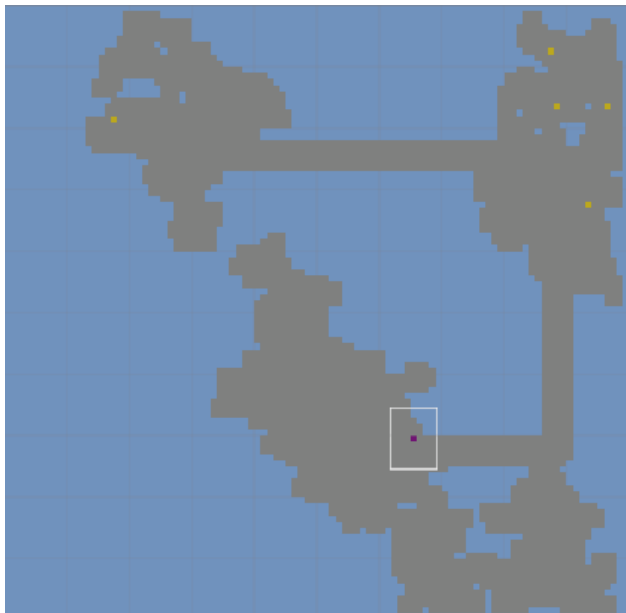


Рис. 2. Приклад згенерованої мапи, без використання згладжування

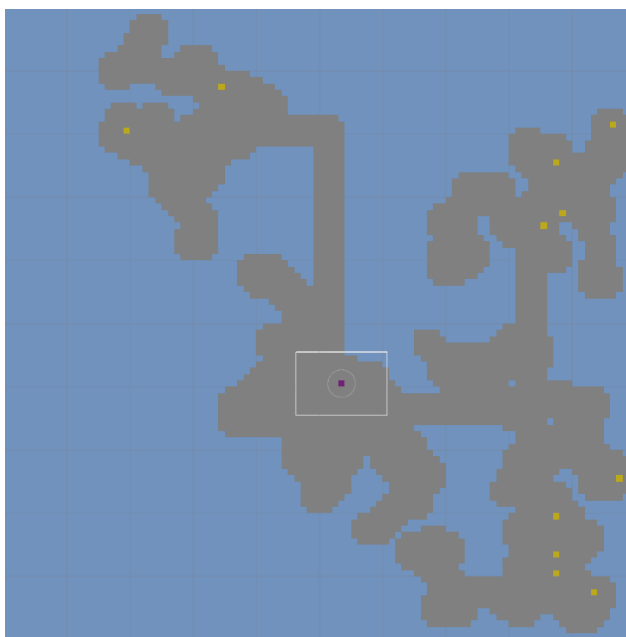


Рис. 3. Приклад згенерованої мапи з використанням згладжування

Запропонований алгоритм можуть бути використані в середовищах розробки інтерактивних застосунків та ігор, зокрема у Unity (C#), Unreal Engine (C++), Cocos2d-x (C++) та інших рушіях, що підтримують двовимірне або тайлове представлення простору. Завдяки параметризованій

структурі генератора (розмір карти, відсоток початкового заповнення, кількість ітерацій згладжування) алгоритм може бути адаптований до різних жанрів – від roguelike та dungeon-crawler до навчальних або симуляційних систем.

Висновки. У роботі досліджено алгоритмічний підхід до процедурної генерації 2D-рівнів у комп'ютерних іграх жанру roguelike на основі клітинних автоматів із подальшою структурною та морфологічною обробкою простору, запропоновано доповнення та експериментально визначено оптимальні параметри вхідних даних. Запропоновано формалізовану модель генерації ігрового середовища у вигляді двовимірної дискретної сітки з бінарним представленням станів клітин, що дозволило описати процес формування рівня у термінах ітеративної еволюції локальних правил.

На відміну від підходів, що обмежуються виключно застосуванням клітинного автомата, запропонований алгоритм доповнено механізмами перевірки топологічної зв'язності, фільтрації ізольованих компонент, побудови коридорів між регіонами та модульною системою розміщення ігрових об'єктів. Це забезпечує не лише природоподібність геометрії, а й структурну цілісність та ігрову придатність згенерованих рівнів.

Експериментальний аналіз впливу параметрів початкової щільності заповнення та кількості ітерацій згладжування дозволив визначити діапазони значень, за яких досягається баланс між варіативністю, прохідністю та топологічною зв'язністю карти. Показано, що алгоритм має лінійну залежність обчислювальної складності від розміру сітки, що робить його придатним для використання в системах реального часу.

Практична реалізація в середовищі Unity підтвердила ефективність модульної архітектури генератора, яка забезпечує масштабованість, параметризованість і можливість детермінованого відтворення результатів через використання початкового зерна випадковості.

Отримані результати можуть бути використані як основа для побудови адаптивних систем процедурної генерації, інтеграції механізмів автоматичного балансування складності, а також для створення навчальних і тестових середовищ у задачах машинного навчання та навчання з підкріпленням. Перспективами подальших досліджень є розробка методів автоматичного підбору параметрів генерації, застосування еволюційних і нейромережевих підходів до оптимізації локальних правил клітинного автомата та розширення алгоритму для багаторівневих або тривимірних ігрових просторів.

Список літератури:

1. Shaker N., Togelius J., Nelson M. J. Procedural Content Generation in Games. Springer, 2016. P. 1-55. DOI: 10.1007/978-3-319-42716-4
2. Breno M. F. Viana, Selan R. dos Santos. Procedural Dungeon Generation: A Survey. Journal on Interactive Systems. 2021. Vol. 12. № 1. P. 83–101. DOI: 10.5753/jis.2021.999.
3. Іляш Ю.Ю., Горелов В.О., Ровінський В.А. Алгоритми та методи процедурної генерації та адаптивної складності рівнів в іграх-платформерах. Вісник Херсонського національного технічного університету. 2025. № 4. С. 107–116. DOI: 10.35546/kntu2078-4481.2025.4.3.12
4. Pech A., Hingston P., Masek M., Lam P. Lecture Notes in Computer Science (Artificial Life and Computational Intelligence). Springer, 2015. Vol. 8955. P. 112–124. DOI: 10.1007/978-3-319-14803-8_9
5. Pech A., Hingston P., Masek M., Lam P. Game level layout generation using evolved cellular automata. Connection Science. 2016. Vol. 28. № 1. P. 63–82. DOI: 10.1080/09540091.2015.1130020
6. Adams C., Parekh H., Louis S. J. Procedural level design using an interactive cellular automata genetic algorithm. Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO '17). 2017. P. 85–86. DOI: 10.1145/3067695.3075614
7. Antoniuk I. Generating layout for complex cave-like levels with schematic maps and Cellular Automata. Machine Graphics and Vision. 2023. Vol. 32. № 2. P. 45–64. DOI: 10.22630/MGV.2023.32.2.3
8. Трофименко О.Г., Задрейко О.В., Янковський О.Г., Каіров В. О., Морозова Г.С. Системний аналіз алгоритмів генерації лабіринтів для інтерактивних середовищ. Кібербезпека: освіта, наука, практика. 2025. № 30. С. 259–279. DOI: 10.28925/2663-4023.2025.30.972
9. Lazaridis L., Fragulis G. F. Creating a Newer and Improved Procedural Content Generation (PCG) Algorithm with Minimal Human Intervention for Computer Gaming Development. Computers. 2024. Vol. 13. № 11. P. 155–175. DOI: 10.3390/computers13110304
10. Kreitzer M., Ashlock D., Pereira R. (2019). Automatic Generation of Diverse Cavern Maps with Morphing Cellular Automata. 2019 IEEE Conference on Games (CoG). 2019. P. 933–940. DOI: 10.1109/CIG.2019.8847947

Iliash Yu.Yu., Horielov V.O., Antoniuk V.S. CELLULAR AUTOMATION ALGORITHMS FOR LEVEL GENERATION IN COMPUTER GAMES

This article is devoted to the study of algorithms for the procedural generation of game levels in 2D roguelike games based on cellular automata and methods of morphological processing of two-dimensional discrete structures. The relevance of using stochastic and deterministic algorithmic approaches to generate variable game environments that ensure high replayability, adaptability, and scalability of the game process is substantiated. The paper examines the specifics of applying cellular automata using Moore's neighborhood and the influence of initialization parameters (filling density, number of smoothing iterations) on the topological characteristics of the generated maps.

A modular level generator architecture is proposed, which includes initial stochastic map initialization, iterative smoothing according to Moore's neighborhood rule, removal of isolated regions using a depth-first search algorithm, construction of corridors between rooms based on graph connections, and morphological expansion of passable areas. Additionally, the ability to generate secondary game environment objects based on the same cellular automaton principles has been incorporated, ensuring structural consistency within the level.

The practical implementation was carried out in the Unity engine using the C# language and a tilemap system. However, thanks to the versatility of the approach, the presented algorithms can be easily adapted to other programming languages and development environments, including C++. It is shown that combining cellular automata with additional topological and morphological operations allows for the generation of nature-like cave structures characteristic of the roguelike genre, while maintaining full reproducibility of results through the use of a seed parameter.

The analysis demonstrates that varying the generation parameters directly affects the level's connectivity, the size and shape of game spaces, as well as the balance between open areas and obstacles. The results obtained can be used as a foundation for reinforcement learning systems and other machine learning methods that work with procedurally generated environments, as well as a basis for further research on the use of machine learning methods in generative game design.

Keywords: procedural generation, cellular automata, morphological operations, Unity, Tilemap, C#, C++, connectivity component, algorithm, smoothing.

Дата першого надходження статті до видання: 19.03.2026

Дата прийняття статті до друку після рецензування: 15.04.2026

Дата публікації (оприлюднення) статті 11.05.2026

УДК 004:519.2

DOI DOI <https://doi.org/10.32782/2663-5941/2026.2.1/17>

Іносов С.В.

<https://orcid.org/0000-0001-8305-5514>

Київський національний університет будівництва і архітектури

Гаврюков О.В.

<https://orcid.org/0000-0002-6377-4180>

Київський національний університет будівництва і архітектури

Бондарчук О.В.

<https://orcid.org/0000-0003-1893-1893>

Київський національний університет будівництва і архітектури

Луценко В.Ю.

<https://orcid.org/0000-0002-9727-5574>

Київський національний університет будівництва і архітектури

Самойленко М.І.

<https://orcid.org/0000-0001-9410-6962>

Київський національний університет будівництва і архітектури

СИНТЕЗ АЛГОРИТМУ ПРОГРАМНОГО РЕГУЛЮВАННЯ ДЛЯ СИЛЬНО НЕЛІНІЙНОГО ТЕПЛОВОГО ОБ'ЄКТУ (НА ПРИКЛАДІ ПРОПАРЮВАЛЬНОЇ КАМЕРИ)

Теплова обробка – найбільш розповсюджений метод прискорення твердіння бетону. Теплову обробку бетону здійснюють переважно в пропарювальних камерах циклічної дії. Цей об'єкт об'єкт регулювання має суттєві особливості. Автоматичне регулювання температури не стабілізує, а програмне, з відпрацюванням заданого графіку прогріву. Об'єкт регулювання суттєво нелінійний, але стаціонарний, динамічні властивості якого не змінюються за весь період експлуатації. Відсутні суттєві зовнішні збурення, крім основного – зміни завдання в часі. Але це основне збурення повністю детерміноване і контрольоване. Об'єкт регулювання фактично одноємнісний, з несуттєвими (у порівнянні з переважаючою основною інерційністю) паразитними інерційностями (запізненням, тощо). В цьому частковому випадку має сенс відмовитись від регулювання виключно за відхиленням з використанням традиційного пропорційно-інтегрально-диференціюючого (ПІД) –регулятора. Рекомендується основне регулювання реалізувати по основному збуренню – зміні завдання (за розімкненою схемою). Але для компенсації неконтрольованих малих остаточних помилок регулювання має сенс залишити пропорційний (П) канал регулювання за відхиленням через негативний зворотній зв'язок. Використання диференціюючого (Д) і інтегруючого (І) каналів регулювання в даному частковому випадку не має сенсу. Коефіцієнт пропорційності K_p рекомендується змінювати в ході технологічного процесу в зростаючій залежності від температури. Це потрібно для компенсації нелінійності об'єкта регулювання. За технологічними вимогами (для інтенсивного перемішування пару) зміни подачі пару повинні відбуватися в двопозиційному режимі (відкрито-закрито) за рахунок широтно-імпульсної модуляції. Додатковою перевагою запропонованого алгоритму регулювання є максимально економне використання ресурсу роботи виконавчого механізму і клапана подачі пару.

Ключові слова: автоматизація, пропарювальна камера, нелінійний об'єкт, програмне регулювання, синтез, алгоритм.

Постановка проблеми. Теплова обробка – найбільш розповсюджений метод прискорення твердіння бетону [1]. Теплову обробку бетону здійснюють переважно в пропарювальних камерах циклічної дії. Регулятор температури програмно регулює температуру в камері по заданому гра-

фіку. Автоматичне регулювання температури не стабілізує, а програмне, з відпрацюванням заданого графіку прогріву. Об'єкт регулювання суттєво нелінійний, але стаціонарний, динамічні властивості якого не змінюються за весь період експлуатації. Відсутні суттєві зовнішні збурення, крім основного – зміни завдання в часі. Але це основне збурення повністю детерміноване і контрольоване. Об'єкт регулювання фактично одноємнісний, з несуттєвими (у порівнянні з переважаючою основною інерційністю) паразитними інерційностями (запізненням, тощо). В цьому частковому випадку має сенс відмовитись від регулювання виключно за відхиленням з використанням традиційного пропорційно-інтегрально-диференціюючого (ПІД) –регулятора. Рекомендується основне регулювання реалізувати по основному збуренню – зміні завдання (за розімкненою схемою). Але для компенсації неконтрольованих малих остаточних помилок регулювання має сенс залишити пропорційний (П) канал регулювання за відхиленням через негативний зворотній зв'язок. Використання диференціюючого (Д) і інтегруючого (І) каналів регулювання в даному частковому випадку не має сенсу. Коефіцієнт пропорційності K_p рекомендується змінювати в ході технологічного процесу в зростаючій залежності від температури. Це потрібно для компенсації нелінійності об'єкта регулювання. За технологічними вимогами (для інтенсивного перемішування пару) зміни подачі пару повинні відбуватися в двопозиційному режимі (відкрито-закрито) за рахунок широтно-імпульсної модуляції. Додатковою перевагою запропонованого алгоритму регулювання є максимально економне використання ресурсу роботи виконавчого механізму і клапана подачі пару.

© Іносов С.В., Гаврюков О.В., Бондарчук О.В. та ін., 2026

Стаття поширюється на умовах ліцензії відкритого доступу CC BY 4.0



фіку. Рівномірний підйом температури до 85° за 2 години, стабілізація температури на 8 годин, рівномірне зменшення температури за 2 години. Традиційно використовується двопозиційний релейний алгоритм регулювання [2], але він дає незадовільно велику динамічну похибку регулювання. Спроба замінити його на пропорційно-інтегрально-диференціальний (ПІД) алгоритм регулювання з широтно-імпульсним модулятором (ШІМ), що є сучасним рішенням для теплових об'єктів [3], нашою вихідною проблемою, пов'язаною з сильною нелінійністю [4] об'єкту регулювання.

Аналіз останніх досліджень і публікацій.

Традиційно розрахункову модель теплового об'єкту регулювання апріорно відносять до класу лінійних інерційних стаціонарних (ЛІС) і описують передаточною функцією $W(p)$ [2]. Найбільш універсальна і разом з тим зручна для будь-яких теплових об'єктів передаточна функція виду [5]:

$$W(p) = \frac{K \cdot e^{-p\tau}}{(T_1 p + 1) \cdot (T_2 p + 1)} \quad (1)$$

Для пропарювальних камер основна стала часу T_1 суттєво більша від часу запізнення τ і другої сталої часу T_2 , тому, як правило, використовується спрощена модель пропарювальної камери у вигляді передаточної функції першого порядку:

$$W(p) = \frac{K}{(T_1 p + 1)} \quad (2)$$

K – статичний коефіцієнт передачі об'єкту.

Передаточна функція (2) зв'язує температуру з керуючим впливом (подачею пара). Основним зовнішнім збуренням є програмна зміна завдання за заданим графіком (нагрів, витримка, охолодження).

Як виявилось, такий підхід надто грубий (навіть як перше наближення). Пропарювальну камеру неможливо адекватно описати лінійною інерційною стаціонарною (ЛІС) моделлю. Об'єкт має суттєво нелінійні властивості. Нелінійні ефекти настільки значні, що призводять до якісно нових явищ. Навіть лінеаризація не може допомогти, бо по мірі прогріву локальні властивості об'єкту (для малих відхилень) суттєво змінюються. Тим більш неприпустимо використання лінійних моделей для опису глобальних властивостей об'єкту в усьому робочому діапазоні температур.

На рис. 1 наведено сімейство перехідних функцій об'єкту регулювання для різних амплітуд скачка керуючого впливу – ступені відкриття клапану подачі пара.

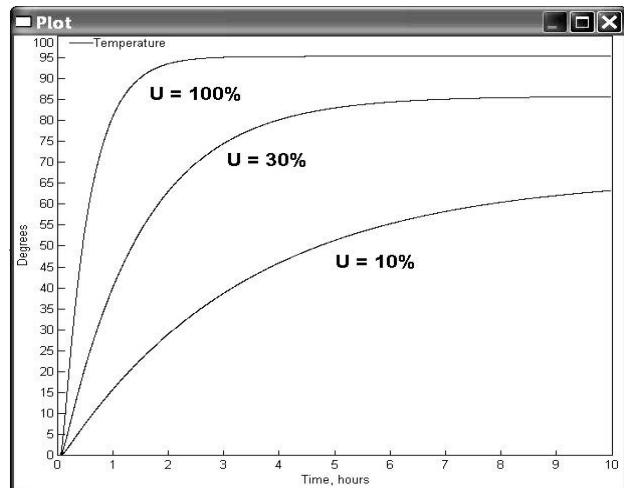


Рис. 1. Сімейство перехідних функцій пропарювальної камери при різних амплітудах скачка керуючого впливу U

Перехідна функція – це графік зростання температури в градусах при відкритті клапану подачі пари стрибком і збереження в подальшому його положення незмінним. Керуючий вплив U може варіюватися від 0% (повністю закрито) до 100% (повністю відкрито).

Для лінійного об'єкту всі перехідні функції (при різних величинах вхідного стрибка) повинні бути геометрично подібними. Кінцеві температури повинні бути пропорційними ступені відкриття клапану. Очевидно, для пропарювальної камери ця умова неможлива. По мірі прогріву температура в камері прямує до певної межі, близької до температури пари. Очевидно, збільшуючи ступінь відкриття клапану, ми можемо тільки прискорити нагрів, але майже не зможемо вплинути на кінцеву температуру. Форма кривої перехідної функції суттєво залежить від амплітуди стрибка керуючого впливу, умова подібності перехідних функцій суттєво порушується. Тобто, об'єкт регулювання не може вважатися лінійним.

Практично вказані особливості об'єкту регулювання не дозволяють використовувати регулятори з постійним налаштуванням в усьому робочому діапазоні температур. По мірі прогріву об'єкту його лінеаризовані для малих відхилень коефіцієнт передачі і стала часу зменшуються на порядок, що вимагає відповідної зміни параметрів налаштування регулятора в ході технологічного процесу.

Постановка завдання. В роботі пропонується динамічна модель об'єкту регулювання, яка враховує його нелінійні особливості. Запропонована модель використовується для синтезу спеціалізованого високоточного алгоритму регулювання.

Виклад основного матеріалу. Модель пропарювальної камери як об'єкту регулювання найбільш наглядно представляється у вигляді теплового ланцюга (рис. 2).

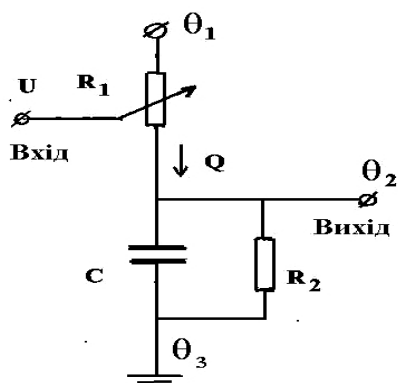


Рис. 2. Модель пропарювальної камери як об'єкту регулювання у вигляді теплового ланцюга

де:

U – Керуючий вплив, ступінь відкриття клапану
 θ_2 – Температура в камері, регульована величина
 θ_1 – Температура пара в магістралі
 θ_3 – Температура зовнішнього середовища
 C – Теплоємність камери з виробами
 Q – Приток тепла в камеру
 R_1 – Тепловий опір клапана
 R_2 – Тепловий опір витоку тепла в зовнішнє середовище

Модель суттєво нелінійна. Керуючий вплив на приток тепла в камеру реалізується параметрично, шляхом зміни теплового опору клапана в залежності від ступеня його відкриття.

Модель адекватно враховує наступні характерні особливості об'єкту регулювання:

- при відкриванні клапана на постійну величину температура камери зростає за експонентою, як в лінійному об'єкті, але стала часу цієї експоненти суттєво залежить від ступеня відкриття клапану.
- усталена температура майже не залежить від ступеню відкриття клапана, а визначається, в першу чергу, температурою пару. Ступінь відкриття клапану суттєво впливає на швидкість нагріву, але не на усталену температуру.
- по мірі прогріву камери маніпуляції клапаном все менше впливають на зміни температури.
- для підтримання усталеної температури достатньо незначного притоку тепла, що покриває втрати тепла в зовнішнє середовище.
- сильне форсування подачі тепла потрібне для прискорення процесу початкового прогріву.
- для малих відхилень від робочого режиму, об'єкт можна розглядати як лінійний нестационарний

(так звана лінеаризація об'єкту), у якого стала часу і статичний коефіцієнт передачі суттєво зменшуються в ході прогріву камери.

– для врахування малих паразитних інерційностей (датчика, виконавчого пристрою і т.п.), які можуть викликати нестійкість регулювання при сильному зворотному зв'язку, в модель додатково введені (на рис. 2 не показані) блок запізнення на 0.05 год. і аперіодичний блок першого порядку із сталою часу 0.05 год.

Задача системи автоматичного регулювання: необхідно так маніпулювати відкриттям клапана, щоб зростання температури відбувалося за заданим графіком (рівномірне зростання від 0 до 85 градусів на протязі 2 годин, потім стабілізація температури на досягнутому рівні).

Ідея синтезу алгоритму регулювання наступна.

Пропарювальна камера циклічної дії є фактично однозмінним об'єктом регулювання з несуттєвими (у порівнянні з основною інерційністю) паразитними інерційностями (запізненням і т.п.). Для таких об'єктів, якщо вони лінійні і стаціонарні, теоретично оптимальним є пропорційний алгоритм в контурі регулювання за відхиленням від завдання.

Але нашому частковому випадку, коли об'єкт регулювання сильно нелінійний, пропорційний регулятор без адаптації (з фіксованим налаштуванням в усьому робочому діапазоні температур) принципово не може забезпечити хорошу якість регулювання. Причина полягає в зміні динамічних властивостей об'єкту регулювання по мірі прогріву.

Але, на щастя, перевагою є те, що об'єкт регулювання стаціонарний, зі стабільними динамічними властивостями, які не змінюються за весь період експлуатації.

Крім того, відсутні суттєві зовнішні збурення, крім основного – зміни завдання в часі. Але це суттєве збурення повністю детерміноване і контрольоване.

Тому має сенс відмовитись від регулювання виключно за відхиленням з використанням традиційного пропорційно-інтегрально-диференціюючого (ПІД) –регулятора з фіксованою настройкою [3, 6].

Рекомендуємо основне регулювання реалізувати за основним збуренням, тобто за зміною завдання (за розімкненою схемою).

Але для компенсації неконтрольованих малих залишкових похибок регулювання має сенс використовувати пропорційний канал регулювання за відхиленням через негативний зворотний зв'язок.

Використання диференціюючого (Д) і інтегруючого (І) каналів регулювання в нашому частковому випадку не має сенсу.

Працездатність запропонованої ідеї перевірялась в даній роботі.

Синтез алгоритму регулювання виконувався з допомогою комп'ютерної моделі [7] в програмному середовищі Visual Simulation. Одна година в реальному часі відповідає одній секунді на моделі. В процесі синтезу алгоритм поступово (в 3 етапи) ускладнювався, з метою зменшення похибок регулювання.

Етап №1 ілюструється рисунком 3. В першому наближенні регулювання відбувається за найпростішою розімкненою схемою, без зворотних зв'язків. Широтно-імпульсний модулятор управляється безпосередньо від задатчика. Як бачимо, якість регулювання незадовільна. Середньоквадратична похибка регулювання становить 13.5 градуси.

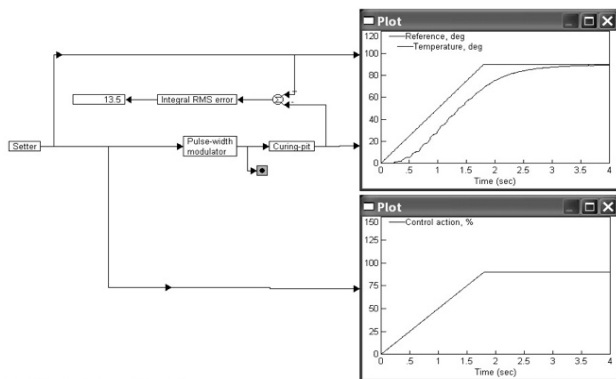


Рис. 3. Перший етап синтезу алгоритму регулювання

Умовні позначення: Setter – програмний задатчик, Pulse-width modulator – широтно-імпульсний модулятор (ШИМ)¹; Integral SMS error – блок обчислення середньоквадратичної похибки регулювання; Σ – блок обчислення розугодження Error між фактичною температурою Temperature і заданою температурою Reference.

Етап №2 ілюструється рисунком 4. Для збільшення початкової швидкості зростання температури вводиться форсуюча добавка +20 на вході ШІМ. Графік зміни керуючого впливу (на відміну від рис. 1) починається не з нуля, а з +20%. Якість регулювання суттєво покращилась. Середньоквадратична похибка регулювання зменшилась до 3.53 градуса.

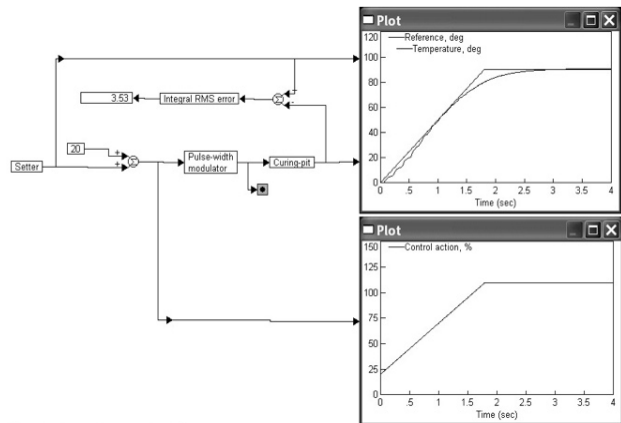


Рис. 4. Другий етап синтезу алгоритму регулювання

Етап №3 ілюструється рисунком 5. Для усунення відносно малих неконтрольованих залишкових похибок регулювання додатково введено коригуючий канал негативного зворотного зв'язку (пропорційне регулювання за відхиленням) з коефіцієнтом пропорційності K_p . Як бачимо, якість регулювання знову покращилась. Середньоквадратична похибка регулювання зменшилась до 2.64 градуса. Цей результат можна вважати задовільним і остаточним. Зверніть увагу, керуючий вплив (на відміну від рис. 3) зростає з прискоренням, як і потрібно для компенсації нелінійності об'єкта. Тобто пропорційний зворотний зв'язок працює правильно, ефективно придушуючи неконтрольовані залишкові похибки регулювання за збуренням.

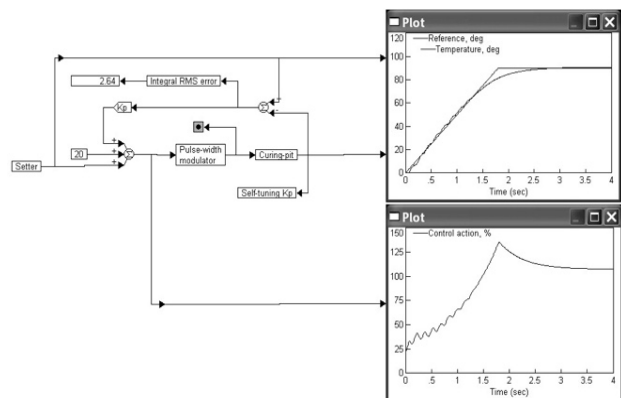


Рис. 5. Третій етап синтезу алгоритму регулювання

Коефіцієнт пропорційності K_p не є константою. Він змінюється в ході технологічного процесу в зростаючій залежності від температури. Для цього введено блок автоналаштування K_p (Self-tuning K_p). Це потрібно для компенсації нелінійності об'єкту регулювання. Внутрішня структура блоку авто-налаштування і графік залежності K_p від температури наведені на рис. 6.

¹ За технологічними вимогами (для інтенсифікації перемішування пару) зміни керуючого впливу повинні виконуватися в двопозиційному режимі (відкрито-закрито) за рахунок широтно-імпульсної модуляції (рис. 7). Додаткова перевага: нелінійність клапана подачі пари не має значення.

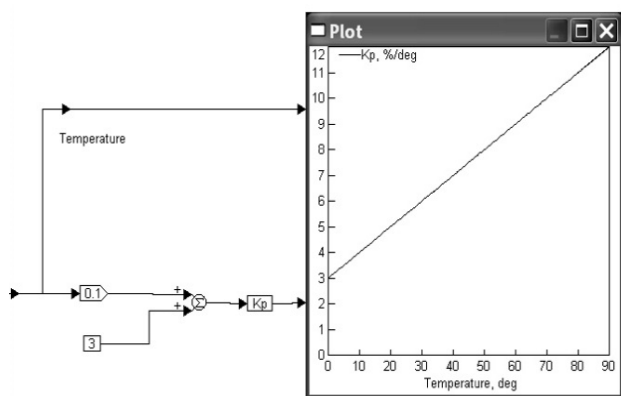


Рис. 6. Внутрішня структура блоку авто-настроювання коефіцієнту пропорційності K_p . Залежність K_p від температури

На рис. 7 представлено графік включень-виключень виконавчого механізму для синтезованого остаточного варіанту алгоритму регулювання. Як бачимо, клапан подачі пари відкривався-закривався всього 10 раз. Тобто ресурс роботи клапана і виконавчого механізму витрачається максимально економно, що є додатковою перевагою запропонованого алгоритму регулювання.

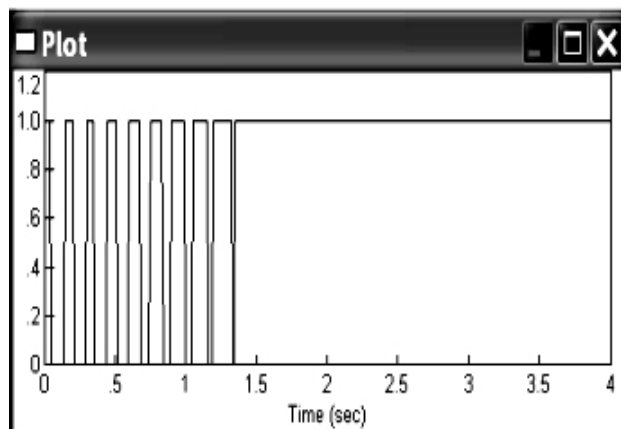


Рис. 7. Графік включень-виключень подачі пари

Висновки

1. При регулюванні температури в пропарювальній камері циклічної дії маємо справу з суттєвими особливостями об'єкту регулювання.

2. Система автоматичного регулювання не стабілізує, а програмна, з відпрацюванням заданого графіку прогріву. Температура змінюється в широкому діапазоні.

3. Об'єкт регулювання сильно нелінійний.

4. Об'єкт регулювання стаціонарний, динамічні властивості якого не змінюються за весь період експлуатації.

5. Відсутні суттєві зовнішні збурення, крім основного – зміни завдання в часі. Але це основне збурення повністю детерміноване і контролюване.

6. Об'єкт регулювання фактично односмісний, з незначними (у порівнянні з переважаючою основною інерційністю) паразитними інерційностями (запізненням, тощо).

7. В цьому частковому випадку має сенс відмовитися від регулювання виключно за відхиленням з використанням традиційного пропорційно-інтегрально-диференціюючого (ПІД) –регулятора.

8. Рекомендується основне регулювання виконувати за основним збуренням – зміною завдання (за розімкненою схемою). Але для компенсації неконтрольованих малих залишкових похибок регулювання має сенс залишити пропорційний (П) канал регулювання за відхиленням через негативний зворотній зв'язок. Використання диференціюючого (Д) і інтегруючого (І) каналів регулювання в даному частковому випадку не має сенсу.

9. Коефіцієнт пропорційності K_p рекомендується змінювати в ході технологічного процесу в зростаючій залежності від температури. Це потрібно для компенсації нелінійності об'єкта регулювання.

10. За технологічними вимогами (для інтенсифікації перемішування пари) зміни керуючого впливу повинні виконуватися в двопозиційному режимі (відкрито-закрито) за рахунок широтно-імпульсної модуляції.

11. Додатковою перевагою запропонованого алгоритму регулювання є максимально економне використання ресурсу роботи виконавчого механізму і клапана подачі пари.

Список літератури:

1. Коц І.В. Тепловологісна обробка бетонних виробів з використанням аеродинамічного нагрівання: монографія/ І.В.Коц, О.П.Колісник – Вінниця: ВНТУ, 2013. 114с.
2. Попович М.Г., Ковальчук О.В. Теорія автоматичного керування: Підручник для ВУЗів. К.: Либідь, 1997, 315 с.
3. С.В. Іносов, В.М. Корниенко. Оптимізація алгоритму автоматичного регулювання тепловими процесами Управління розвитком складних систем. № 13, 2013, с. 104-108.

4. Євстифєєв В.О. Теорія автоматичного керування. – Ч.1. – Безперервні лінійні та нелінійні системи: Навчальний посібник для ВУЗів. – Кременчук: ПП Щербатих О.В., 2006. 288 с.
5. Іносов С. В., Бондарчук О. В. Аналіз можливих причин помилкової ідентифікації динамічних параметрів теплового об'єкта регулювання. Управління розвитком складних систем. Київ, 2023. № 54. с. 132–137.
6. Іносов С. В., Бондарчук О. В. Ілларионов В. М. Настроювання ПІД-регулятора за перехідною функцією розімкненого контура. Управління розвитком складних систем. Київ, 2021. № 46. с. 167–172.
7. Чумак Л. І., Бацун І. О. Математичне моделювання процесу теплової обробки ніздрювато-бетонних виробів в автоклаві. Вісник ПДАБА. № 11–12 листопад – грудень 2011, с. 54–58.

Inosov S.V., Gavryukov O.V., Bondarchuk O.V., Lutsenko V.Yu., Samoylenko M.I. SYNTHESIS OF AN ALGORITHM OF PROGRAMMED CONTROL FOR A STRONGLY NONLINEAR PLANT (BY THE EXAMPLE OF A CURING PIT)

Heat treatment is the most common method of accelerating the hardening of concrete. Heat treatment of concrete is carried out mainly in steaming chambers of cyclic action. This object of control has significant features. Automatic temperature control is not stabilizing, but programmatic, with the development of a given heating schedule. The control object is significantly nonlinear, but stationary, the dynamic properties of which do not change during the entire period of operation. There are no significant external disturbances, except for the main one – changes in the task over time. But this main disturbance is fully deterministic and controlled. The control object is actually single-capacitive, with insignificant (compared to the prevailing main inertia) parasitic inertias (delay, etc.). In this particular case, it makes sense to abandon control exclusively by deviation using a traditional proportional-integral-differentiator (PID) controller. It is recommended to implement the main regulation according to the main disturbance – a change in the task (in an open circuit). But to compensate for uncontrolled small final regulation errors, it makes sense to leave the proportional (P) regulation channel for the deviation due to negative feedback. The use of differentiating (D) and integrating (I) regulation channels in this particular case does not make sense. The proportionality coefficient K_p is recommended to be changed during the technological process in an increasing dependence on the temperature. This is necessary to compensate for the nonlinearity of the regulated object. According to technological requirements (for intensive steam mixing), changes in the steam supply should occur in a two-position mode (open-closed) due to pulse-width modulation. An additional advantage of the proposed regulation algorithm is the most economical use of the operating resource of the actuator and the steam supply valve.

Keywords: Automation, curing pit, non-linear plant, programmed control, algorithm, synthesis.

Дата першого надходження статті до видання: 16.03.2026

Дата прийняття статті до друку після рецензування: 13.04.2026

Дата публікації (оприлюднення) статті 11.05.2026

Кандиба І.О.

<https://orcid.org/0000-0002-8589-4028>

Чорноморський національний університет імені Петра Могили

Гончаров Д.С.

<https://orcid.org/0009-0004-1200-6677>

Чорноморський національний університет імені Петра Могили

ІНТЕГРАЦІЯ ДАНИХ СЕРВІСУ GOOGLE FIT ДО СИСТЕМИ МОНІТОРИНГУ СТАНУ ЗДОРОВ'Я

У статті проведено дослідження інструментарію для реалізації інтеграції даних сервісу Google Fit із системою моніторингу стану здоров'я людини. Проведено аналіз сучасних публікацій, присвячених використанню сервісу Google Fit для проведення медичних досліджень. Визначено основні переваги застосування цього сервісу для збору та аналізу показників стану здоров'я людини. Розглянуто можливості підключення різноманітного обладнання до Google Fit для моніторингу показників стану здоров'я людини. Описано особливості моніторингу та аналізу даних однойменним застосунком Google Fit. Визначено показники стану здоров'я людини, які може обробляти та зберігати цей сервіс. Досліджено можливості експорту даних Google Fit, визначено найбільш доцільний для реалізації завдання інтеграції даних із системами моніторингу стану здоров'я людини – сервіс Google Takeout. Наведено опис доступних форматів експорту: JSON та CSV. Описано особливості кодування експортованих даних, які можуть вплинути на процес інтеграції з інформаційною системою моніторингу стану здоров'я людини. Визначено основні переваги застосування формату CSV, а саме можливість імпорту різними програмними забезпеченнями для роботи з табличними даними без необхідності підключення додаткових компонентів. Досліджено засоби реалізації модуля інтеграції даних сервісу Google Fit із системою моніторингу стану здоров'я людини на основі мови програмування Python. Визначено засоби реалізації вебінтерфейсу користувача та можливості завантаження даних за допомогою цього інтерфейсу для подальшої обробки інструментарієм Python. Описано кроки первинної обробки даних для видалення записів Google Fit без зафіксованих показників стану здоров'я або фізичної активності людини. Представлено можливість використання потокової обробки даних без створення тимчасових файлів на сервері медичної системи. Розглянуто можливість застосування засобів HTML та JavaScript для візуалізації отриманих даних. Запропоновано шляхи вдосконалення процесу обробки отриманих даних засобами штучного інтелекту, що підтримуються мовою програмування Python.

Ключові слова: Google Fit, Python, моніторинг показників стану здоров'я людини, Pandas, аналіз даних.

Постановка проблеми. Раннє виявлення серцево-судинних захворювань дозволяє уникнути деяких ускладнень та підвищує ефективність їх лікування. У сучасному світі використовується кілька різних методів неінвазивної діагностики цього класу захворювань: електрокардіограма, ехокардіографія, ангіокардіографія тощо [1 с. 26]. Проте всі ці методи вимагають наявності спеціалізованого обладнання та можуть бути проведені виключно за направленням лікаря. З огляду на те, що на ранніх етапах серцево-судинні захворювання не мають помітних симптомів, лікар може не мати підстав для призначення спеціалізованих досліджень із використанням зазначених методів.

З розвитком апаратного забезпечення популярність набрали пристрої моніторингу активності та окремих показників стану здоров'я людини, наприклад пульсу, оксигенації, температури [2 с. 85]. Цей клас пристроїв включає різноманітні підкласи, але більшість з них зберігає інформацію на спеціалізованому хмарному сховищі та не дозволяє прямого експорту до медичних систем з якими можуть взаємодіяти лікарі. Однак зібрані дані можуть допомогти лікарям завчасно виявити певні захворювання або ризики їх виникнення.

Уніфікувати дані, що зберігаються на різних сервісах можливо використавши сервіс для агрегації даних. Таким сервісом є Google Fit. Цей сер-

віс підтримує інтеграцію з апаратним забезпеченням для прямого збору даних та підключення до деяких сервісів моніторингу активності, наприклад Strava [3 с. 435]. Проте зібрані з різних джерел дані зберігатимуться виключно в хмарному сховищі сервісу Google Fit.

Таким чином, необхідним є створення модуля інтеграції зібраних сервісом Google Fit даних моніторингу стану здоров'я людини до спеціалізованої інформаційної системи.

Використання спеціалізованої системи моніторингу стану здоров'я людини дозволить завчасно помітити відхилення в роботі серцево-судинної системи. Зазначені системи мають багаторівневу та багатоланкову архітектуру, що дозволяє підключати додаткові модулі.

Аналіз останніх досліджень і публікацій. Можливості інтеграції Google Fit з різним програмним та апаратним забезпеченням є актуальною темою сучасних досліджень. У роботі [4 с. 108] описано можливість використання даних сервісів Google Fit та Apple Healthkit для проведення великих популяційних досліджень. Такий тип досліджень характеризується великою кількістю учасників, що, в свою чергу, викликає проблеми збору даних. Авторами розглянуто різні методи експорту даних із зазначених сервісів та можливість їх аналізу спеціалізованим інструментарієм. Проте питання можливості використання додаткового програмного забезпечення для аналізу отриманих даних та їх подальшого використання в діагностичних цілях потребує подальшого розгляду.

У роботі [5 с. 79] представлено аналіз точності моніторингу активності різними носимими пристроями шляхом використання Google Fit. Для дослідження авторами використано різноманітне обладнання: Fitbit Ace 2, Garmin Vivofit Jr 2, та Xiaomi Mi Band 5. У роботі визначено фактори, що впливають на точність виявлення активності, але питання обробки отриманих даних потребує подальшого дослідження.

Дослідження [6 с. 104] описує використання сервісу Google Fit для визначення взаємозв'язку остеоартриту з фізичною активністю. У цій роботі Google Fit виступає джерелом об'єктивних даних про фізичну активність учасників дослідження, які використовуються мобільним застосунком для моніторингу кроків і оцінки дотримання певного рівня щоденної активності у пацієнтів з остеоартритом. Проте питання можливості обробки отриманих даних сторонніми засобами потребує подальшого дослідження.

Виконаний аналіз демонструє, що сервіс Google Fit використовується для різноманітних досліджень пов'язаних з моніторингом активності та показників стану здоров'я людини. Проте питання експортування отриманих даних та інтеграція їх спеціалізованих систем моніторингу стану здоров'я людини потребує подальшого розгляду.

Постановка завдання. Метою статті є дослідження можливості інтеграції даних сервісу Google Fit із системою моніторингу стану здоров'я, а також визначення інструментарію, необхідного для створення модуля вирішення цього завдання. Цей модуль дозволить завантажувати, обробляти та аналізувати дані, зібрані сервісом Google Fit для спрощення виявлення ризиків та завчасного діагностування серцево-судинних захворювань.

Основні завдання, які вирішуються для досягнення заявленої мети:

- 1) провести дослідження структури даних зібраних сервісом Google Fit;
- 2) дослідити можливості експорту даних Google Fit;
- 3) визначити та проаналізувати засоби аналізу та обробки даних Google Fit;
- 4) запропонувати шляхи подальшого вдосконалення системи моніторингу стану здоров'я людини.

Виклад основного матеріалу. Сервіс Google Fit представляє собою спеціалізований сервіс моніторингу активності та показників стану здоров'я людини: пульс, оксигенація, кількість пройдених кроків тощо. Для збору даних використовується спеціалізований застосунок, який підтримує можливість безпосередньої взаємодії з спеціалізованими пристроями та інтеграції з окремими застосунками для збору даних.

Застосунок Google Fit дозволяє проводити постійний моніторинг та фіксацію показників стану здоров'я людини автоматично без втручання користувача. Під час моніторингу дані зберігаються на пристрої користувача та можуть бути переглянуті у зазначеному застосунку, призначений для цього інтерфейс має засоби графічного та текстового відображення даних (рис. 1).

Зібрані дані зберігаються у спеціалізованому хмарному сховищі Google Fit і доступні не тільки для перегляду, а й для експорту та подальшого аналізу. Експорт може бути здійснений кількома способами: за допомогою використання спеціалізованого API, сервісу Google Takeout та сторонніми застосунками. Для використання спеціалізованого API користувач повинен мати поглиблені

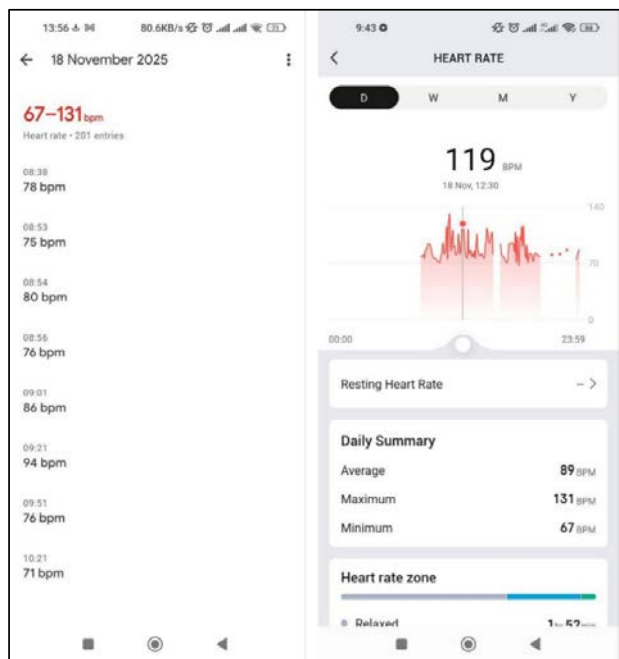


Рис. 1. Відображення зібраних даних засобами Google Fit

знання у галузі інформаційних технологій тому такий спосіб використовується при реалізації сторонніх застосунків. Сторонні застосунки, у більшості випадків, призначені для імпорту даних з пристроїв або інших застосунків та дозволяють перегляд та агрегацію вже зібраних за певний час даних. Сервіс Google Takeout надає користувачу простий та зрозумілий інтерфейс для зберігання повної історії вимірів.

Google Takeout являє собою сервіс експорту даних, який дозволяє користувачам завантажити дані зі своїх облікових записів Google, включно з даними про фізичну активність із Google Fit. Використовуючи Takeout можливо отримати структурований архів даних про: кількість кроків, пульс та оксигенацію з точним часом проведених вимірів. Дані зазвичай надаються у двох форматах: JavaScript Object Notation (JSON) та Comma-Separated Values (CSV). Ці формати можуть бути імпортовані в різні застосунки для обробки даних, наприклад засоби роботи з електронними таблицями можуть відобразити дані Google Fit у форматі CSV.

Експортовані у форматі CSV дані з Google Fit являють собою простий текстовий файл, в якому кожен рядок відповідає окремому запису, а комірки розділені комами. Цей файл має чітку структуру в якій кожен запис (рядок) є часовим інтервалом і всі зроблені виміри вносяться в цей інтервал. Кожен стовпчик відповідає типу виміру який було зроблено [4]:

- «Start time» – час початку інтервалу вимірювання;
- «End time» – час завершення інтервалу вимірювання;
- «Move Minutes count» – зафіксована тривалість фізичної активності в поточному інтервалі;
- «Calories (kcal)» – кількість витрачених калорій в поточному інтервалі;
- «Distance (m)» – пройдена відстань в поточному інтервалі;
- «Heart Points» – умовна метрика активності, яка відображає інтенсивність руху, що розраховується залежно від швидкості руху, зафіксованого типу активності, показників пульсу тощо.
- «Heart Minutes» – показник, що відображає час фізичної активності з навантаження серцево-судинної системи, тобто часу, коли рух має достатню інтенсивність, щоб підвищити пульс.
- «Average heart rate (bpm)» – середній пульс в поточному інтервалі;
- «Max heart rate (bpm)» – максимальний пульс в поточному інтервалі;
- «Min heart rate (bpm)» – мінімальний пульс в поточному інтервалі;
- «Average speed (m/s)» – середня швидкість руху в поточному інтервалі;
- «Max speed (m/s)» – максимальна швидкість руху в поточному інтервалі;
- «Min speed (m/s)» – мінімальна швидкість руху в поточному інтервалі;
- «Step count» – кількість пройдених кроків в поточному інтервалі;
- «Average weight (kg)» – середнє значення маси тіла, виміряне з усіх записів ваги за вибраний період часу;
- «Max weight (kg)» – максимальне значення маси тіла, виміряне в поточному інтервалі;
- «Min weight (kg)» – мінімальне значення маси тіла, виміряне в поточному інтервалі;
- «Walking duration (ms)» – відображає тривалість відрізків часу, які Google Fit класифікувала як ходьбу.

Назви стовпців змінюються в залежності від локалізації, наприклад «Average heart rate (bpm)» у при українській локалізації матиме назву «Середній пульс (уд./хв)». Проте порядок показників завжди однаковий, що у деяких випадках спрощує обробку даних.

При виконанні експорту з Google Fit користувач отримує не один файл, а набір файлів, що містять дані окремо по днях. Експортовані файли мають кодування UTF-8, що необхідно враховувати при обробці отриманих даних засобами офіс-

них пакетів, наприклад Microsoft Excel. Це програмне забезпечення (ПЗ) дозволяє переглянути зібрані дані та побудувати деякі діаграми на їх основі (рис. 2).

Дані з сервісу Google Fit допоможуть виявити ризик серцево судинних захворювань при їх систематичному аналізі відповідним лікарем кардіологом. Забезпечити доступ до даних може спеціалізована система моніторингу стану здоров'я людини, що виконую функцію агрегації даних різних медичних досліджень та зберігання інформації про пацієнта. Такі системи мають складну архітектуру і використовують різні протоколи для підключення взаємодії з медичними пристроями та шифрування даних [6]. В роботі [7] досліджено переваги використання мови Python для розробки багаторівневої системи моніторингу стану здоров'я людини. Отже доцільним буде розробка модулю для інтеграції даних Google Fit засобами мови Python.

Використання мови програмування Python для розробки модулю інтеграції даних сервісу Google Fit до системи моніторингу стану здоров'я людини дає можливість виконати обробку отриманих даних та їх аналіз. При використанні файлу у форматі CSV виникає необхідність відсіювання порожніх записів адже сервіс Google Fit створює запис всіх інтервалів часу навіть коли не фіксувалися показники датчиків або активність людини.

Перед початком обробки даних необхідно завантажити файл до згаданої системи. Для забезпечення багатокористувацького доступу використовується вебінтерфейс, що дозволяє здійснювати перегляд, редагування, видалення та імпорт даних.

Мова програмування Python підтримує кілька бібліотек, які дозволяють реалізувати вебінтерфейс: Django, Flask, FastAPI тощо. З перерахованого варто виділити Django, що є повноцінним фреймворком з вбудованими засобами обробки запитів, керування доступом, шифруванням та об'єктно-реляційного представлення даних із реляційних баз даних (БД) [8 с. 23].

Завантаження файлів до системи моніторингу стану здоров'я людини реалізується за допомогою HTML форми з атрибутом `enctype="multipart/form-data"` та методу передачі запитів POST. У формі використовується елемент `<input type="file">`, що дозволяє користувачам обрати файл на своєму пристрої. При відправці форми відбувається надсилання файлу на сервер з використанням протоколу HTTP, але інструментарій Django підтримує можливість підключення додаткових компонентів для використання протоколу HTTPS, що дозволяє безпечніше працювати з особистими даними. Окрім того використовується вбудований механізм безпеки з використанням CSRF токєну для безпечної обробки надісланої форми.

У фреймворку Django обробка завантажених файлів здійснюється через об'єкти `request.FILES`, які містять передані файли у вигляді масиву байт, що у випадку передачі CSV, можна перетворити у текстовий вигляд за допомогою методу, що `decode('utf-8')`, де 'utf-8' – стандарт кодування Unicode Transformation Format 8-bit, використовується за замовчуванням при експорті з даних Google Fit.

Наступним кроком є видалення записаних інтервалів без зафіксованих показань активності

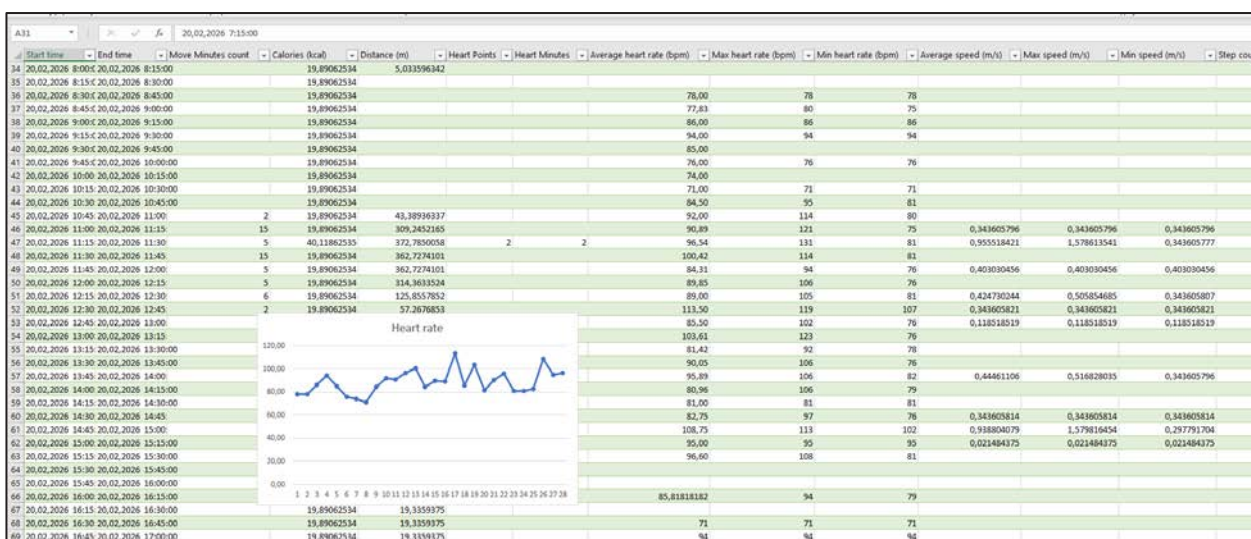


Рис. 2. Використання Microsoft Excel для відображення даних Google Fit

та показників датчиків. Для формату CSV доцільним є використання бібліотеки Pandas, що дозволяє сформувати набір даних у вигляді спеціалізованого об'єкта класу DataFrame методом `read_csv()` з файлу згаданого формату. За функціоналом DataFrame подібний до електронної таблиці або таблиць з реляційних БД: має рядки та стовпці з підписами, підтримує різні типи даних і дозволяє виконувати фільтрацію, сортування, агрегацію та обчислення над даними [9 с. 116].

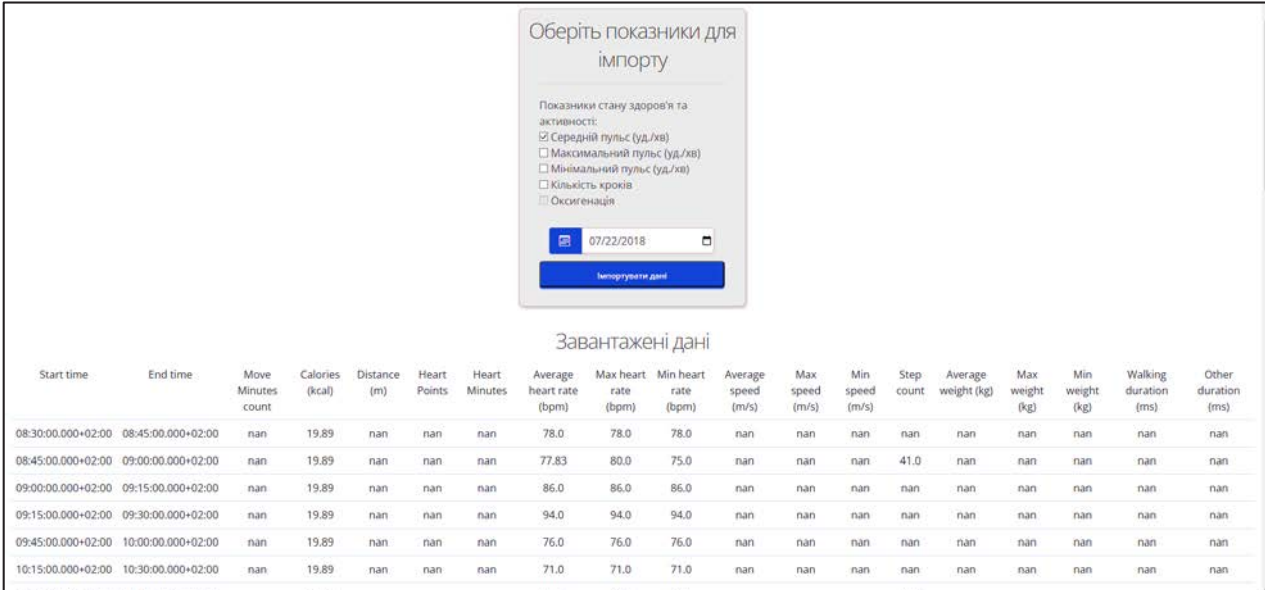
Завантажений Django файл для використання `read_csv()` потребує конвертування описаними вище методом та збереження у вигляді CSV файлу або використання StringIO. StringIO – це клас що, реалізовує функціонал файлу для звичайного текстового рядка. Він надає файловий інтерфейс (`read()`, `write()`, `seek()`), але не потребує запису даних на диск і зберігає їх в оперативній пам'яті [10 с. 120]. Цей клас дозволить уникнути збереження на сервері завантажених файлів до їх обробки.

Отримані з Google Takeout дані можуть мати некоректні метадані, а точніше дата створення файлу може не відповідати даті фіксування показників, а у самому CSV зберігаються тільки інтервал часу без вказування дати. Отже, при завантаженні користувачу необхідно ввести дані зробленого виміру і обрати які саме дані він буде завантажувати в систему. Для цього розроблено вебінтерфейс засобами Django.

Виконавши вибір та завантаження даних, користувач зможе переглянути результати вимірів у системі моніторингу стану здоров'я людини. Ці дані також доступні лікарям із відповідним рівнем доступу та придатні до подальшої обробки інструментарієм Python, наприклад вже згаданим Pandas. Окрім того для відображення графічного представлення можливо використати різні засоби HTML та JavaScript, як наприклад Google Chart

Дані сервісу Google Fit та дані, зібрані спеціалізованими датчиками, агрегуються в системі моніторингу стану здоров'я людини і можуть бути використані для подальшого аналізу засобами штучного інтелекту (ШІ), наприклад за допомогою платформи Weka. Використання засобів ШІ дозволить підвищити ймовірність діагностування захворювань на ранніх стадіях [11].

Висновки. Досліджено особливості збору даних фізичної активності та показників стану здоров'я людини сервісом Google Fit. Проаналізовано можливості експорту даних сервісу Google Fit. Виділено два основні варіанти: використання API та сервісу Google Takeout. Визначено основні переваги використання Google Takeout, описано доступні формати даних для експорту. Розглянуто можливості обробки експортованих даних спеціалізованими застосунками. Розроблено модуль для інтеграції даних Google Fit до системи моніторингу стану здоров'я людини. Визначено методи завантаження та первинної обробки даних засобами Django та Pandas.



Start time	End time	Move Minutes count	Calories	Distance (m)	Heart Points	Heart Minutes	Average heart rate (bpm)	Max heart rate (bpm)	Min heart rate (bpm)	Average speed (m/s)	Max speed (m/s)	Min speed (m/s)	Step count	Average weight (kg)	Max weight (kg)	Min weight (kg)	Walking duration (ms)	Other duration (ms)
08:30:00.000+02:00	08:45:00.000+02:00	nan	19.89	nan	nan	nan	78.0	78.0	78.0	nan	nan	nan	nan	nan	nan	nan	nan	nan
08:45:00.000+02:00	09:00:00.000+02:00	nan	19.89	nan	nan	nan	77.83	80.0	75.0	nan	nan	nan	41.0	nan	nan	nan	nan	nan
09:00:00.000+02:00	09:15:00.000+02:00	nan	19.89	nan	nan	nan	86.0	86.0	86.0	nan	nan	nan	nan	nan	nan	nan	nan	nan
09:15:00.000+02:00	09:30:00.000+02:00	nan	19.89	nan	nan	nan	94.0	94.0	94.0	nan	nan	nan	nan	nan	nan	nan	nan	nan
09:45:00.000+02:00	10:00:00.000+02:00	nan	19.89	nan	nan	nan	76.0	76.0	76.0	nan	nan	nan	nan	nan	nan	nan	nan	nan
10:15:00.000+02:00	10:30:00.000+02:00	nan	19.89	nan	nan	nan	71.0	71.0	71.0	nan	nan	nan	nan	nan	nan	nan	nan	nan

Рис. 3. Вебінтерфейс експорту даних з завантаженого файлу

Список літератури:

1. Остафійчук Д. І., Караняга А. О., Полікарпова І. С. Клінічні методи дослідження серцево-судинної діяльності (огляд літератури). Клінічна та експериментальна патологія. 2022. № 2(72). Том 19. С. 85-90. DOI:10.24061/1727-4338.XIX.2.72.2020.12
2. Kumar A. M. Measuring of Fitness Trackers. Indian Journal of Public Health Research & Development. 2017. Vol. 8, no. 3s. P. 86. URL: <https://doi.org/10.5958/0976-5506.2017.00246.7> (date of access: 15.02.2026).
3. Sandy T. A. Optimizing health outcomes: a detailed comparison of features and user sentiment in popular fitness tracking applications. Retos: nuevas tendencias en educación f'ísica, deporte y recreación. Issue 68. P. 432–444. <https://doi.org/10.5958/0976-550.2017.00246.7>
4. Henriksena A., Hopstocka L. A., Hartvigsenc G. Using Cloud-Based Physical Activity Data from Google Fit and Apple Healthkit to Expand Recording of Physical Activity Data in a Population Study. MEDINFO 2017: Precision Healthcare through Informatics: Proceedings of the 16th World Congress on Medical and Health Informatics(2018). P. 108. DOI: <https://doi.org/10.47197/retos.v68.114746>
5. Mayorga-Vega D., Casado-Robles C., Guijarro-Romero S. Criterion-related validity of consumer-wearable activity trackers for estimating steps in primary schoolchildren under controlled conditions: Fit-Person study. Journal of Sports Science & Medicine. 2024. Vol. 23, Issue 1. P. 79. DOI:<https://doi.org/10.52082/jssm.2024.79>
6. Лирчиков В. О., Байбуз О. Г. Огляд сучасних медичних інформаційних систем. Актуальні проблеми автоматизації та інформаційних технологій. Том. 25, 2021. С. 103–107. DOI: <http://dx.doi.org/10.15421/432111>
7. Гончаров Д. С., Гончарова Н. В., Кандиба І. О. Архітектура багаторівневої системи моніторингу стану здоров'я людини. Вчені записки Таврійського національного університету імені В.І. Вернадського. Серія: Технічні науки. Том 35 (74) № 3. 2024 78-85. DOI: <https://doi.org/10.32782/2663-5941/2024.2/11>
8. Mele A. Django 3 By Example: Build powerful and reliable Python web applications from scratch. Packt Publishing Ltd, 2020. 533 p.
9. Bantilan N. Pandera: Statistical data validation of pandas dataframes. Proceedings of the Python in Science Conference (SciPy)(2020). P. 116–124. DOI: <https://doi.org/10.25080/Majora-342d178e-010>
10. 1. Khorasani M., Abdou M., Hernández Fernández J. Managing and Visualizing Data. Streamlit for Web Development : Build and Scale Secure Python-Powered Apps with Streamlit. Berkeley, CA : Apress, 2025. P. 109–134. DOI: https://doi.org/10.1007/979-8-8688-1826-4_4
11. Chuiko G., Honcharov D. Dimensionality cutback and deep learning algorithms efficacy as to the breast cancer diagnostic dataset. Radioelectronic and Computer Systems. Vol. 2024, Issue 4. P. 91-98. DOI:<https://doi.org/10.32620/reks.2024.4.08>

Kandyba I.O. Honcharov D.S. INTEGRATION OF DATA FROM THE GOOGLE FIT SERVICE INTO A HEALTH MONITORING SYSTEM

The article presents a study of the tools used to implement the integration of data from Google Fit into a human health monitoring system. An analysis of recent publications devoted to the use of Google Fit in medical research is carried out. The main advantages of applying this service for collecting and analyzing human health indicators are identified. The possibilities of connecting various devices to Google Fit for monitoring health parameters are examined. The features of monitoring and data analysis using the Google Fit application are described. The health indicators that can be processed and stored by this service are determined. The study explores the possibilities of exporting Google Fit data and identifies the most appropriate tool for implementing data integration with human health monitoring systems – Google Takeout. The available export formats, JSON and CSV, are described. The characteristics of encoding exported data that may affect the integration process with a health monitoring information system are analyzed. The main advantages of the CSV format are identified, in particular the possibility of importing data into various software applications for tabular data processing without the need to connect additional components. The tools for implementing a data integration module between Google Fit and a human health monitoring system based on the Python programming language are investigated. The means of implementing a web-based user interface and the possibilities of uploading data through this interface for further processing using Python tools are determined. The steps of initial data preprocessing aimed at removing Google Fit records without recorded health or physical activity indicators are described. The possibility of streaming data processing without creating temporary files on the medical system server is presented. The use of HTML and JavaScript tools for visualization of the obtained data is considered. Finally, ways to improve the processing of the acquired data using artificial intelligence tools supported by the Python programming language are proposed.

Keywords: Google Fit, Python, health monitoring system, Pandas, data analysis.

Дата першого надходження статті до видання: 06.03.2026

Дата прийняття статті до друку після рецензування: 31.03.2026

Дата публікації (оприлюднення) статті 11.05.2026

Карплевський Б.В.

<https://orcid.org/0009-0003-3564-0833>

Glowcraft Installations

ТЕХНОЛОГІЇ ІНТЕГРАЦІЇ ВІРТУАЛЬНИХ СЕРЕДОВИЩ У СИСТЕМІ УПРАВЛІННЯ РЕАЛЬНИМ СЦЕНІЧНИМ ОБЛАДНАННЯМ

У статті здійснено системний аналіз сучасних технологій інтеграції віртуальних середовищ у комплексні системи управління реальним сценічним обладнанням, що забезпечує високоточну синхронізацію фізичних і цифрових компонентів театральних, концертних та мультимедійних постановок. Особливу увагу приділено методології взаємодії програмного забезпечення для моделювання сценічних ефектів з апаратними контролерами освітлення, звукового супроводу, механічних конструкцій сцени й систем відеопроєкції, що дозволяє підвищити точність, безпеку та оперативну ефективність керування складними постановочними процесами. Проаналізовано архітектурні підходи до побудови інтегрованих систем, включно з використанням протоколів реального часу (Real-Time Protocols), сенсорних мереж (Sensor Networks), технологій віртуальної й доповненої реальності (VR/AR), а також методів motion capture та систем просторової навігації для оптимізації взаємодії між цифровими моделями й фізичними об'єктами сцени. Розглянуто практичні аспекти попереднього тестування сценічних вирішень у віртуальному середовищі, що забезпечує адаптивність до динамічних змін умов постановки, зменшення ризиків технічних помилок і раціоналізацію ресурсів під час репетиційного процесу. Дослідження підкреслює перспективи розвитку таких інтегрованих систем у напрямі автоматизації керування мультимедійними ефектами, інтерактивної взаємодії з аудиторією, масштабування сценічних вирішень, а також застосування алгоритмів штучного інтелекту для прогнозування й адаптації сценічних сценаріїв у реальному часі. Вказано на доцільність поетапного проектування інтеграційної архітектури з тестуванням сумісності програмно-апаратних компонентів кіберфізичних сценічних комплексів. У роботі наведено приклади успішної реалізації інтегрованих технологій у сучасних театральних і концертних проєктах, що підтверджують їхню ефективність у підвищенні художньої якості постановки, оптимізації творчих процесів, забезпеченні синергетичної взаємодії між цифровими й фізичними елементами сцени, а також у створенні нового рівня інтерактивного та іммерсивного досвіду для глядацької аудиторії.

Ключові слова: цифрові симуляції, реалістичні моделі, автоматизація процесів, візуалізація сцени, програмне забезпечення контролю, інтерактивні платформи, синхронізація пристроїв, дистанційне управління.

Постановка проблеми. Сучасний розвиток сценічного мистецтва та індустрії розважальних технологій характеризується динамічною інтеграцією цифрових і віртуальних технологій у традиційні процеси управління сценічним обладнанням, що забезпечує нові можливості для моделювання, планування й контролю технічних систем сцени, включно з освітлювальними комплексами, акустичним оформленням, а також механічними та електронними засобами управління. Застосування VR-симуляторів і цифрових прототипів дозволяє створювати апробовані сценарії роботи обладнання, підвищуючи ефективність та безпеку його експлуатації.

Проте, попри очевидні перспективи, інтеграція віртуальних платформ у реальні системи управління сценічним обладнанням залишається складним інженерно-технологічним завданням. Основні проблемні аспекти пов'язані з технічною несумісністю наявного обладнання із цифровими платформами, використанням пропрієтарних протоколів і застарілих апаратних вирішень. Додаткові труднощі створює гетерогенність стандартів обміну даними на ринку сценічного обладнання, що ускладнює розробку універсальних інтеграційних вирішень. Не менш важливими є й соціально-організаційні виклики, пов'язані з необхідністю трансформації робочих процесів,

перепідготовкою персоналу та адаптацією управлінських практик.

Таким чином, інтеграція віртуальних середовищ у системи управління реальним сценічним обладнанням – багатовимірною проблемою, яка поєднує технічні, методичні й організаційні виклики, подолання яких дозволить підвищити ефективність, безпеку та адаптивність сучасних сценічних технологій, а також сприятиме формуванню науково обґрунтованих моделей цифрового управління в умовах стрімкої технологічної трансформації індустрії.

Аналіз останніх досліджень і публікацій. У сучасній науковій літературі присвячено багато уваги питанням інтеграції імерсивних технологій у традиційні системи управління технічними комплексами. Зокрема, у праці В. Круглик обґрунтовано необхідність упровадження технологій доповненої й віртуальної реальності в освітній простір для підготовки фахівців, здатних працювати в умовах нових технологічних викликів [1, с. 74]. Дослідники Ю. Кулик та А. Батюк у своїх працях аналізують проблематику проєктування та практичної реалізації інтерфейсів взаємодії у VR-середовищі із застосуванням сторонніх програмних вирішень. У дослідженні акцентовано увагу на питаннях інтеграції зовнішніх апаратних модулів, синхронізації віртуальних і фізичних об'єктів, оптимізації каналів передавання даних між програмною платформою й периферійними пристроями [2, с. 81]. Водночас автори розглядають особливості інтеграції хмарних технологій у VR-середовища, зосереджуючи увагу на архітектурних моделях побудови VR-систем, механізмах обробки великих масивів даних і забезпеченні масштабованості цифрових рішень [3, с. 114].

Значний внесок у дослідження впровадження імерсивних технологій у дизайн-практику здійснила В. Бойко, яка обґрунтовує доцільність використання технологій доповненої й альтернативної реальності для підвищення ефективності візуально-комунікативних проєктів в Україні. У її працях акцентовано увагу на трансформації інструментарію дизайн-діяльності під впливом цифрових інновацій та інтерактивних медіа [4].

У дослідженні Т. Хвостової здійснено ґрунтовний аналіз трансформацій сучасної театральної культури в умовах розвитку інтернет-простору й цифрового суспільства. Авторка проаналізувала особливості медіатизації театру, розширення його комунікативних можливостей через онлайн-платформи й формування нових форматів цифрової вистави [5].

Науковці І. Резніченко, О. Негоденко й С. Шевченко оцінюють прикладний потенціал VR-технологій у професійній підготовці, інженерному проєктуванні й створенні тренажерних і навчально-імітаційних систем, підкреслюючи їхнє значення для формування практико-орієнтованих компетентностей фахівців, підвищення безпечності експериментальної діяльності й розроблення цифрових двійників та інтерактивних симуляторів для дистанційного моніторингу й апробації алгоритмів керування складними технічними комплексами [6, с. 45].

Автори Дж. Ортис (J. Ortiz), Р. Піла (R. Pila), Дж. Юпангі (J. Yupangui) і М. Росалес (M. Rosales) підтвердили ефективність віртуальних середовищ для оптимізації систем автоматизованого управління технологічними процесами [7].

Методологічний потенціал адаптації НІЛ-архітектури до управління світловими, механічними та мультимедійними сценічними комплексами з метою підвищення безпеки й зниження технічних ризиків продемонстрували А. Арісой (A. Arisoy) і Д. Кавала-Сен (D. Kavala Sen) [8].

Отже, аналіз наукових джерел засвідчує, що попри істотні напрацювання у сфері імерсивних практик, інтерактивних мультимедійних технологій, автоматизованих систем керування й цифрового сценічного дизайну, проблема комплексної інтеграції віртуальних середовищ у системи управління реальним сценічним обладнанням залишається фрагментарно дослідженою. Більшість праць зосереджена або на технічних аспектах моделювання 3D-простору та візуалізації, або на інженерних вирішеннях автоматизації світлового, звукового й механічного обладнання, тоді як питання їх синхронізованої взаємодії в реальному часі потребує подальшої теоретико-методологічної розробки.

Недостатньо висвітленими залишаються й механізми інтероперабельності програмно-апаратних комплексів, стандарти протоколів передавання даних, алгоритми латентно-стійкої комунікації між віртуальними об'єктами та фізичними виконавчими пристроями (світловими приладами, медіасерверами, сценічною механікою), а також питання кібербезпеки й відмовостійкості таких систем. У науковому дискурсі відчувається нестача узагальнених моделей, що описують інтеграцію VR-/AR-платформ із системами DMX-, Art-Net- чи OSC-керування в контексті сценографічної практики.

Постановка завдання. Мета статті – теоретичне обґрунтування та практичний аналіз сучас-

них технологій інтеграції віртуальних середовищ у системи управління реальним сценічним обладнанням, а також визначення принципів їхньої функційної сумісності, технічної синхронізації й художньо-технологічної взаємодії.

Виклад основного матеріалу. Інтегрування віртуальних середовищ у системи керування реальним сценічним обладнанням постає однією з ключових тенденцій розвитку сучасного сценічного дизайну й театральнотехнологічної практики, яка поєднує інструменти цифрового моделювання, алгоритмічні механізми управління та принципи кіберфізичних систем, що уможливорює гнучке, ефективне й безпечне управління складними сценічними процесами.

У структурі інтегрованих сценічних систем важливу роль відіграє комплекс взаємопов'язаних технологічних компонентів, що забезпечують ефективну функційну взаємодію між віртуальними середовищами й реальним сценічним обладнанням у межах єдиного художньо-технологічного простору. Така інтеграція передбачає не лише технічне поєднання різнорідних систем, але й формування цілісної інфраструктури управління сценічним процесом, яка базується на принципах інтероперабельності, синхронізації даних та алгоритмічної координації.

Ключовими елементами цієї інфраструктури виступають програмні платформи тривимірного моделювання й візуалізації, системи керування медіаконтентом, комунікаційні інтерфейси передавання даних, а також спеціалізовані програмно-апаратні модулі, що забезпечують інтеграцію різних протоколів керування сценічними пристроями. До них належать системи світлового контролю, цифрові аудіопроектори, відеосервери, механізми автоматизованого керування сценічною механікою й сенсорні системи відстеження руху. Взаємодія цих компонентів реалізується через стандартизовані протоколи обміну даними, що забезпечують стабільність, масштабованість

і гнучкість функціонування інтегрованої сценічної екосистеми (табл. 1).

Таким чином, інтегровані сценічні системи – комплексна інфраструктура, де взаємодіють віртуальні середовища й реальне сценічне обладнання в межах єдиного художньо-технологічного простору. Їхня ефективність забезпечується поєднанням програмних платформ для 3D-моделювання та візуалізації, систем проміжного зв'язку й синхронізації, а також алгоритмічних та аналітичних модулів управління. Основна мета такої інтеграції – створення цілісної системи управління сценічним процесом, що базується на принципах інтероперабельності, синхронізації даних та алгоритмічної координації.

На сьогодні можна виокремити два фундаментально різні підходи до інтеграції віртуальних середовищ у системи управління сценічним обладнанням. Перший підхід передбачає повну симуляцію віртуального середовища до фактичного використання фізичного обладнання. Ця методологія забезпечує комплексну оптимізацію сценографічних вирішень, дозволяє здійснювати віртуальне тестування світлових, звукових і спецефектів, а також моделювати кінематику механічних систем без необхідності залучення реальних ресурсів. Завдяки цьому підходу можлива не лише економія матеріальних і часових витрат, а й підвищення безпеки експериментальних сценаріїв, що критично важливо на етапі підготовки складних сценічних конструкцій.

Другий підхід реалізується через управління в режимі реального часу за схемою гібридного контролю, який поєднує віртуальні моделі з фізичними пристроями. Така інтеграція забезпечує синхронну взаємодію цифрового й матеріального середовищ, формує умови для інтерактивності та дає можливість проведення динамічних експериментів під час репетицій і вистав [9]. Використання цього підходу надає митцям інструментарій для оперативної адаптації постановки до неперед-

Таблиця 1

Компоненти віртуальних середовищ системи управління сценічним обладнанням

Компонент	Опис	Основні технології/інструменти
Програмні платформи управління віртуальним середовищем	Програмно-апаратні комплекси для створення, моделювання й візуалізації віртуальної сцени, 3D-декорацій, світлових і звукових ефектів.	Unity, Unreal Engine, 3D-моделювання, фотореалістична рендеризація
Системи проміжного зв'язку й синхронізації	Технологічні інтерфейси для передавання команд між віртуальним середовищем і фізичними пристроями, забезпечення синхронного управління.	MIDI, OSC, Ethernet, DMX512
Алгоритмічні й аналітичні модулі управління	Системи адаптивного контролю та прогнозування, автоматична корекція сценаріїв постановки, оптимізація рухів механічних елементів.	Машинне навчання, нейромережі, аналітика, оптимізація репетицій

Джерело: створено автором

бачуваних змін, експериментальної модифікації сценографії й корекції режисерських вирішень у режимі реального часу. Поєднання цих двох стратегій інтеграції сприяє створенню високодинамічного й адаптивного сценічного середовища, яке відповідає сучасним вимогам художньо-технологічної інноваційності та дозволяє максимально ефективно використовувати потенціал віртуальних систем у театральному й концертному виробництві (рис. 1).

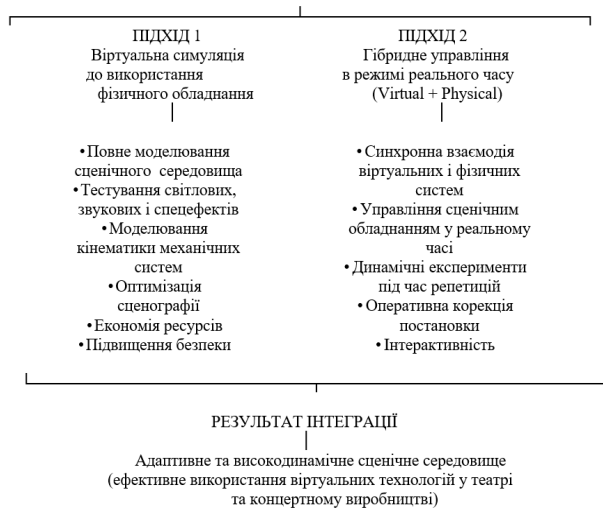


Рис. 1. Підходи до інтеграції віртуальних середовищ у системи управління сценічним обладнанням

Джерело: створено автором

Теоретико-методологічне підґрунтя функційної сумісності та технічної синхронізації між віртуальними й фізичними сценічними елементами – положення системної інженерії, теорії кіберфізичних систем, цифрової сценографії, мережево-орієнтованих архітектур та автоматизованих систем управління реального часу. У парадигмі інтеграції віртуальних середовищ у контур керування реальним сценічним обладнанням центральним завданням постає формування уніфікованої кіберфізичної архітектури, у межах якої цифрові візуалізаційні моделі, програмно-логічні модулі, апаратно-програмні контролери й виконавчі пристрої (інтелектуальні світлові прилади, електромеханічні підйомні механізми, медіасервери, відеопроєкційні комплекси, роботизовані маніпулятори) функціонують як інтегрована розподілена система з високим рівнем когерентності процесів [10, с. 4].

Принцип функційної сумісності означає забезпечення структурної, синтаксичної, семантичної й операційної взаємодії компонентів системи. На

концептуальному рівні це передбачає узгодження протоколів обміну даними, уніфікацію форматів цифрових моделей (зокрема, 3D-геометрії, світлових налаштувань, параметрів анімації), стандартизацію керувальних інтерфейсів і нормалізацію часових характеристик сигналів. Інтегровальність досягається через упровадження прикладних програмних інтерфейсів, використання мережевих протоколів передавання даних і застосування абстрактних моделей даних, що забезпечують коректну трансляцію команд управління між різномірними підсистемами. У цьому сенсі функційна сумісність розглядається як здатність системи зберігати цілісність функційної логіки під час інтеграції нових апаратних і програмних модулів без зниження продуктивності чи втрати детермінованості.

Архітектурною основою інтегрованої системи виступає принцип модульності й ієрархічної декомпозиції, відповідно до якого сценічний комплекс структуровано за багаторівневою моделлю: рівень моделювання (віртуальна 3D-сцена, цифрові декорації, симуляція світлових сценаріїв), рівень планування (таймлайни, сценарні алгоритми), рівень диспетчеризації (централізований або розподілений контролер), рівень виконання (виконавчі механізми та сенсорні підсистеми). Така організація корелює з теорією багаторівневих автоматизованих систем управління та принципами проєктування. Синхронізація між рівнями забезпечується механізмами буферизації, маркуванням часових міток, алгоритмами компенсації латентності й використанням подієво-орієнтованої обробки сигналів [11].

Принцип технічної синхронізації передбачає досягнення часової когерентності й фазової узгодженості всіх елементів сценічного комплексу. У високодинамічному театральному-концертному середовищі критичним є забезпечення мінімальної джитерності й передбачуваного часу відгуку системи. Для цього застосовуються протоколи мережевої синхронізації часу, системи єдиного тактового генератора, а також детерміновані мережеві технології з гарантованою пропускну здатністю. Детермінованість процесів визначається як здатність системи забезпечувати прогнозовану реакцію на вхідні події з фіксованими часовими інтервалами, що досягається через оптимізацію алгоритмів обробки, пріоритезацію трафіку, мінімізацію затримок і резервування критичних каналів зв'язку.

Надзвичайно важливим є впровадження моделі цифрового двійника, який створює мак-

симально точне й адекватне відображення фізичного сценічного обладнання у віртуальному середовищі, акумулюючи водночас усі ключові параметричні характеристики пристрою – від кінематичних обмежень і світлотехнічних показників до електричних навантажень та просторових координат. Ця модель дозволяє не лише проводити верифікацію різних сценаріїв і симулювати поведінку системи в реальному часі, але й прогнозувати можливі граничні режими її функціонування. У межах цієї концепції забезпечення функційної сумісності досягається через узгодження на трьох взаємопов'язаних рівнях: на сигнальному рівні – завдяки відповідності й сумісності протоколів передавання даних; на семантичному рівні – через єдність тлумачення та розуміння параметрів системи; а на логіко-алгоритмічному рівні – шляхом кореляції сценарних подій із конкретними фізичними діями виконавчих механізмів [12, с. 69], що забезпечує точну, узгоджену й синхронізовану взаємодію всіх компонентів системи.

Отже, функційна сумісність та технічна синхронізація під час інтеграції віртуальних середовищ і реального сценічного обладнання досягається через поєднання системного, кіберфізичного, інформаційно-комунікаційного й синергетичного підходів. Їх імплементація забезпечує високий рівень інтегрованості цифрових і матеріальних компонентів сценічного простору, підвищує надійність, масштабованість та відмовостійкість системи, мінімізує ризики часових та логічних колізій і створює методологічні передумови для розвитку інтелектуалізованих сценічних комплексів нового покоління з ознаками автономності й предиктивного управління.

У контексті інтенсифікації процесів цифрової трансформації сценічного мистецтва особливої актуальності набуває дослідження художньо-технологічних параметрів інтероперабельності між віртуальними середовищами й фізичною інфраструктурою сценічного комплексу. Ідеться про формування кіберфізичної сценічної системи, у межах якої програмні рушії тривимірної візуалізації, рендерингу в реальному часі й процедурної анімації інтегруються з апаратно-програмними комплексами керування світлотехнічним, звуковим і механічним обладнанням сцени [13, р. 7]. Така конвергенція зумовлює становлення нової парадигми цифрової сценографії, що функціонує як синхронізована цифрово-фізична екосистема з високим рівнем автоматизації, мережевої взаємодії й алгоритмічної координації.

Ключове технічне завдання цього процесу – забезпечення інтеграційної сумісності й функційної взаємодії між системами генерації віртуального контенту та протокольно орієнтованими системами керування сценічним обладнанням, де критично важливу роль відіграють механізми трансляції даних, конверсії параметрів і проміжні програмні інтерфейси. Саме вони гарантують коректну інтерпретацію віртуальних характеристик інтенсивності освітлення, колірної температури, позиціонування рухомих елементів, панорування та просторового розташування звукового сигналу у фізичні сигнали управління.

Досягнення функційної взаємодії в системах реального середовища відбувається через інтеграцію протоколів синхронізації й механізмів координації на рівні мережевої інфраструктури, що особливо критично для середовищ із передбачуваною латентністю. Використання високошвидкісних мережевих архітектур у поєднанні з технологіями сегментації трафіку й пріоритизації потоків дозволяє знизити дисперсію часу передавання сигналу та забезпечити детерміновану роботу компонентів у режимі реального часу [14]. У комплексі це формує уніфікований часовий континуум, у межах якого відбувається координація й синхронізація функцій світлових консолей, медіасерверів, аудіо-процесорів і кінетичних модулів.

Художньо-естетичний вимір інтеграції виявляється через концептуалізацію гібридної сценографії, що інтегрує віртуальні декорації, процедурно згенеровані світлові ефекти, об'ємне освітлення й фізичні сценічні об'єкти в цілісну аудіовізуальну композицію. Попереднє цифрове моделювання сцени й створення її віртуальних копій дозволяє системно планувати світлові партитури, траєкторії руху механізмів, акустичні поля та просторові звукові панорами вже на етапі проєктування постановки. Такий підхід сприяє раціоналізації ресурсів, мінімізації технічних ризиків і підвищенню точності реалізації режисерського концепту.

З технологічного погляду інтеграція передбачає застосування систем просторового відстеження та технологій розширеної й доповненої реальності, що забезпечують когерентність координатної системи сцени й віртуального середовища. Використання технологій захоплення руху, тривимірного сканування, інерційних сенсорів, оптичних маркерів і систем позиціонування із шістьма ступенями свободи дозволяє конструювати адаптивні сценарії з реальним зворотним зв'язком [15]. У таких інтегрованих

системах світлові, відео- й аудіоефекти функціонують на принципах інтерактивності та параметричної варіативності, оперативно реагуючи на рухи виконавців, динаміку сценічної мізансцени й трансформації просторової топології.

Особливої уваги потребує проблема композиційної й семіотичної цілісності сценічного образу. Технологічну інфраструктуру має бути інкорпоровано в драматургічну структуру вистави на засадах функційної доцільності й концептуальної релевантності. Надмірна візуальна насиченість, алгоритмічна гіпердинаміка або технологічний детермінізм можуть спричинити деконструкцію сценічного знака й порушення когерентності художнього наративу. Відтак інтеграція цифрових середовищ потребує міждисциплінарної координації між сценографом, медіахудожником, технічним режисером, інженером світла й системним інтегратором.

Отже, художньо-технологічний аналіз взаємодії віртуальних середовищ і реального сценічного обладнання передбачає дослідження архітектури програмно-апаратних комплексів, протоколів передавання й синхронізації даних, алгоритмів обробки сигналів, систем просторового трекінгу та принципів цифрової сценографічної композиції. У результаті такої конвергентної інтеграції можна створити інтерактивний, адаптивний і кіберфізично узгоджений сценічний простір, що відповідає вимогам високотехнологічного мистецького виробництва й логіці розвитку цифрової перформативної культури.

У контексті розвитку цифрових сценічних практик і становлення кіберфізичних сценічних екосистем особливої актуальності набуває розроблення практичних рекомендацій щодо імплементації інтегрованих апаратно-програмних комплексів у театральному й концертному виробництві. Технології інтеграції віртуальних середовищ у системи керування реальним сценічним обладнанням потребують системного, мультидисциплінарного підходу, що охоплює архітектурно-інженерні, мережево-комунікаційні, алгоритмічні, організаційно-логістичні й художньо-режисерські аспекти проектування сценічного продукту [16, р. 3].

Насамперед необхідно забезпечити функційну й технічну сумісність програмно-апаратних модулів із можливістю їх інтеграції в єдину цифрову інфраструктуру. На підготовчому етапі виробництва доцільно здійснити комплексну технічну експертизу сценічного обладнання – включно зі світловими консолями, медіасерверами, системами

механізації сцени, аудіопроцесорами й контролерами рухомих елементів – з подальшим аналізом їхньої сумісності та потенціалу інтеграції з тривимірною графікою в режимі реального часу. Ключовий аспект цього процесу – забезпечення підтримки стандартизованих протоколів передавання даних, що дозволяють координувати параметри віртуальних об'єктів із фізичними пристроями в реальному часі [17]. Водночас необхідно приділяти особливу увагу мінімізації затримок сигналу, оптимізації пропускну здатності мережевих каналів і стабілізації топології мережі шляхом її сегментації та пріоритезації трафіку.

Також важливо дотримуватися методології попередньої візуалізації, що передбачає створення параметричної тривимірної моделі сцени з інтегрованими світловими, відео- й механічними симуляціями, які працюють на основі фізично коректного рендерингу та систем частинок. Такий підхід дозволяє здійснювати тестування світлових сцен, відеомепінгу й кінетичних сценаріїв ще до фактичного монтажу обладнання на майданчику. Використання інтегрованих медіасерверів, які підтримують обчислення графіки на відеокартах, керування часовою шкалою, роботу з LED-панелями й інтерактивними інтерфейсами, дозволяє зменшити технічні ризики, оптимізувати графік підготовки та підвищити точність реалізації режисерської концепції.

Для масштабних концертних проєктів доцільно також застосовувати модель керування з головним сервером таймкоду, який координує світлові, аудіо- й відеомодулі. Інтеграція систем автоматизованого керування дозволяє реалізувати детерміновану послідовність тригерів, синхронізовану з музичним матеріалом і хореографією. Важливо також забезпечити резервування мережевих вузлів, дублювання медіасерверів і функціонування системи моніторингу, що підвищить надійність роботи інтегрованого комплексу під час живого виконання.

Рекомендовано застосовувати адаптивні моделі управління постановочним процесом, зокрема ітеративне прототипування та спринт-репетиції, а також поєднувати технічні сесії з онлайн-симуляцією у віртуальному середовищі й офлайн-тестуванням, що забезпечує узгодження алгоритмів взаємодії цифрових і фізичних підсистем у реальному часі.

Ефективна реалізація технічних завдань вимагає уваги до кадрово-компетентної структури проєктів інтегрованих сценічних систем, що передбачає формування міждисциплінарної

команди із чітко диференційованими функційними ролями, до складу якої входять режисер-постановник, сценограф, дизайнер освітлення, медіадизайнер, інженер інтерактивних систем, програміст, системний інтегратор і технічний директор.

Особливу увагу варто приділити проблематиці художньо-технологічної конвергенції та принципам імерсивного дизайну. Віртуальні середовища в цьому контексті виступають не лише декоративними елементами, а є семантично насиченими компонентами драматургічної структури сценічного твору. Використання систем трекінгу руху, сенсорних інтерфейсів, технологій доповненої та змішаної реальностей, а також алгоритмів процедурної генерації контенту в режимі реального часу формує інтегровану інтерактивну сценічну екосистему, що забезпечує високий рівень занурення, посилює сприйняття аудиторією й розширює когнітивно-емоційний спектр художнього досвіду.

Важливо також забезпечити масштабованість, модульність та економічну ефективність інтегрованих систем. Модульна архітектура дозволяє адаптувати систему до різних сценічних просторів і бюджетних обмежень без утрати функційної цілісності. Водночас використання хмарних сервісів для зберігання, версіювання та синхронізації медіаконтенту дозволить оптимізувати логістику виробництва й підвищити мобільність гастрольних проєктів. Крім того, доцільним є впровадження енергоефективних технологій, зокрема світлодіодного освітлення й оптимізації енергоспоживання, що відповідає принципам сталого розвитку та екологічної безпеки [18, с. 129].

Таким чином, ефективна інтеграція віртуальних середовищ у системи керування реальним сценічним обладнанням базується на принципах інтероперабельності, цифрового прототипування, низьколатентної синхронізації, кіберфізичної конвергенції й художньої релевантності. Комплексна реалізація зазначених підходів забезпечує формування інноваційної сценічної платформи, у межах якої віртуальні й матеріальні компоненти функціонують як єдина інтегрована кіберфізична система управління сценічним простором.

Висновки. У результаті дослідження встановлено, що інтеграція віртуальних середовищ у системи управління реальним сценічним обладнанням – комплексний міждисциплінарний процес, який поєднує інформаційно-комунікаційні технології, інженерію сценічних систем і художньо-

постановочні практики. Доведено, що ефективність такої інтеграції залежить від забезпечення функційної сумісності програмних платформ тривимірної візуалізації, систем реального часу та апаратно-програмних комплексів керування світловими, звуковими, механічними й мультимедійними пристроями.

Теоретично обґрунтовано принципи функційної сумісності та технічної синхронізації між віртуальними й фізичними сценічними елементами, зокрема принципи інтероперабельності, протокольної стандартизації, часової кореляції сигналів, латентнісної оптимізації та адаптивної масштабованості системи. Установлено, що ключові умови стабільної роботи інтегрованих комплексів – використання уніфікованих протоколів передавання даних, синхронізація в режимі реального часу й упровадження алгоритмів зворотного зв'язку для корекції параметрів фізичного обладнання відповідно до змін у віртуальному середовищі.

Систематизація художньо-технологічних аспектів управління сценічним обладнанням свідчить про трансформацію традиційної сценографічної парадигми, розширення виражальних можливостей режисури, світлодизайну й мультимедійної композиції в цифровому віртуальному просторі. Інтегровані системи забезпечують новий рівень імерсивності, просторово-часової динаміки та інтерактивності сценічної дії, сприяючи формуванню синтетичного художнього простору, у якому віртуальні й фізичні компоненти асоційовані в єдиній кіберфізичній системі.

На основі узагальнення здобутих результатів розроблено практичні рекомендації щодо впровадження інтегрованих систем у театральних і концертних постановках. Зокрема, указано на доцільність поетапного проєктування інтеграційної архітектури, проведення попереднього моделювання у віртуальному середовищі, тестування сумісності програмно-апаратних компонентів, оптимізації мережевої інфраструктури й підготовки технічного персоналу до роботи з кіберфізичними сценічними комплексами.

Таким чином, технології інтеграції віртуальних середовищ у системи управління реальним сценічним обладнанням формують перспективний напрям розвитку сучасних сценічних мистецтв, забезпечуючи синергію цифрових інновацій та традиційних технічних засобів і відкриваючи нові можливості для художньо-технологічного експерименту.

Список літератури:

1. Круглик В. Інтеграція технологій доповненої та віртуальної реальності з адаптивними системами навчання: аналіз концептуальних моделей. *Освітологічний дискурс*. 2023. № 4(43). С. 69–82. DOI: <https://doi.org/10.28925/2312-5829.2023.44>.
2. Кулик Ю., Батюк А. Аналіз, розроблення та практичне застосування інтерфейсів взаємодії у віртуальній реальності. *Scientific Bulletin of UNFU*. 2025. Т. 35, № 6. С. 76–83. DOI: <https://doi.org/10.36930/40350609>.
3. Батюк А. Є., Кулик Ю. Р. Інтеграція хмарних технологій у віртуальну реальність. Український журнал інформаційних технологій. 2024. Т. 6, № 1. С. 109–119. DOI: <https://doi.org/https://doi.org/10.23939/ujit2024.01.109>.
4. Бойко В. А. Імерсивні технології проектування у візуально-комунікативному контексті дизайну в Україні: дис... доктора філософії 02.022.01. Київ, 2025. 257 с. URL: https://knukim.edu.ua/wp-content/uploads/2025/08/Bojko-V.-Dysertacziya_01.08.2025-pdf.io_compressed.pdf (дата звернення: 27.02.2026).
5. Хвостова Т. В. Сучасний театр в умовах розвитку інтернет простору: дис... доктора філософії 03.034.01. Київ, 2025. 189 с. URL: https://knukim.edu.ua/wp-content/uploads/2025/08/DYSERTACZIYA-HVOSTOVA_compressed.pdf (дата звернення: 27.02.2026).
6. Резніченко І., Негоденко О., Шевченко С. Дослідження програмних та технічних засобів для реалізації технології віртуальної реальності. *Наука, експлуатація, виробництво*. 2024. № 1. С. 41–47. DOI: <https://doi.org/10.31673/2412-9070.2024.014147>.
7. Ortiz J., Pila R., Yupangui J., Rosales M. Interactive Teaching in Virtual Environments: Integrating Hardware in the Loop in a Brewing Process. *Submission received*. 2024. Vol. 14, № 5. 2170. DOI: <https://doi.org/10.3390/app14052170>.
8. Arisoy A., Kavala Sen D. A hardware-in-the-loop simulation case study of high-order sliding mode control for a flexible-link robotic arm. *Applied Sciences*. 2025. Vol. 15, № 19. 10484. DOI: <https://doi.org/10.3390/app151910484>.
9. Gagnere G., Plessiet C., Sohier R. Interconnected virtual space and theater: a research–creation project on theatrical performance space in the network era. *Challenges of the Internet of Things: Technology, Use, Ethics*. 2024. Vol. 7. DOI: <https://doi.org/10.1002/9781119549765.ch8>.
10. Бойко О., Кабацька О., Усик Л. Інтеграція технологій віртуальної та доповненої реальності у процес навчання іноземних мов: потенціал VR та AR для формування комунікативної компетентності здобувачів освіти. *Академічні візії*. 2025. № 47. DOI: <https://doi.org/10.5281/zenodo.17186975>.
11. Велущак М. Я., Гарачковський О. І., Василенко О. В. Використання доповненої реальності в освітньому процесі закладів вищої освіти України. *Академічні візії*. 2025. № 42. DOI: <https://doi.org/10.5281/zenodo.15192765>.
12. Тормахова А., Алексієнко О. Цифрова трансформація сучасних українських виставково-експозиційних практик. *Українські культурологічні студії*. 2025. Т. 1, № 16. С. 67–73. DOI: [https://doi.org/10.17721/UCS.2025.1\(16\).10](https://doi.org/10.17721/UCS.2025.1(16).10).
13. Kim Y., Lu J., Höllerer T. Dynamic theater: location-based immersive dance theater, investigating user guidance and experience. *Proceedings of the 2023 ACM Symposium on Virtual Reality Software and Technology (VRST'23)*. 2023. Art. 27. P. 1–11. DOI: <https://doi.org/10.48550/arXiv.2511.00961>.
14. Erdmann M., Berg M., Steffens J. Development and evaluation of a mixed reality music visualization for a live performance based on music information retrieval. *Frontiers in Virtual Reality*. 2025. № 6. DOI: <https://doi.org/10.3389/frvir.2025.1552321>.
15. Feng W., Xu L., Li W. Research on the integration of stage performance and virtual technologies in the digital media art environment. Distributed, ambient and pervasive interactions. Lecture notes on computer science/Eds. N. A. Streitz, S. Konomi. HCII. 2023. Springer, Cham. Vol. 14037. DOI: https://doi.org/10.1007/978-3-031-34609-5_16.
16. Zhao S., Abou-zeid H., Atawia R., Manjunath Y. S. K., Sediq A. B., Zhang X.-P. Virtual reality gaming on the cloud: A reality check. *Communications Conference (GLOBECOM)*. 2021. P. 1–6. DOI: <https://doi.org/10.1109/GLOBECOM46510.2021.9685808>.
17. Source Dedicated Server. *Valve Developer Community*: вебсайт. 2023. URL: https://developer.valvesoftware.com/wiki/Source_Dedicated_Server (дата звернення: 27.02.2026).
18. Русаков С. Вплив імерсивних технологій на сприйняття мистецтва в контексті трансформації сучасного артринку: культурологічні аспекти дослідження. *Питання культурології*. 2023. № 41. С. 121–133. DOI: <https://doi.org/10.31866/2410-1311.41.2023.276701>.

Karplevskiy B.V. TECHNOLOGIES FOR INTEGRATING VIRTUAL ENVIRONMENTS INTO REAL STAGE EQUIPMENT CONTROL SYSTEMS

The article presents a systematic analysis of modern technologies for integrating virtual environments into complex control systems for real-stage equipment, thereby ensuring high-precision synchronization of the physical and digital components of theatrical, concert, and multimedia productions. Particular attention is paid to the methodology for the interaction between software for modeling stage effects and hardware controllers for lighting, sound, mechanical stage structures, and video projection systems, which increases the accuracy, safety, and operational efficiency of controlling complex production processes. Architectural approaches to building integrated systems are analyzed, including the use of real-time protocols (Real-Time Protocols), sensor networks (Sensor Networks), virtual and augmented reality technologies (VR/AR), as well as motion capture methods and spatial navigation systems to optimize the interaction between digital models and physical objects of the stage. The practical aspects of pre-testing stage solutions in a virtual environment are considered, ensuring adaptability to dynamic changes in production conditions, reducing the risk of technical errors, and rationalizing resources during the rehearsal process. The study highlights the prospects for developing such integrated systems to automate multimedia effects control, enhance audience interaction, scale stage solutions, and use artificial intelligence algorithms to predict and adapt stage scenarios in real time. The expediency of a step-by-step integration architecture design, with compatibility testing of software and hardware components of cyber-physical stage complexes, is indicated. The paper provides examples of successful implementation of integrated technologies in modern theater and concert projects, which confirm their effectiveness in increasing the artistic quality of the production, optimizing creative processes, ensuring synergistic interaction between digital and physical elements of the stage, and creating a new level of interactive and immersive experience for the audience.

Keywords: digital simulations, realistic models, process automation, stage visualization, control software, interactive platforms, device synchronization, remote control.

Дата першого надходження статті до видання: 20.02.2026

Дата прийняття статті до друку після рецензування: 17.03.2026

Дата публікації (оприлюднення) статті 11.05.2026

УДК 004.9

DOI DOI <https://doi.org/10.32782/2663-5941/2026.2.1/20>**Козак Р.Р.**<https://orcid.org/0009-0001-4025-2304>

«Lifeway Food»

Дудник О.О.<https://orcid.org/0000-0003-0375-2606>

Вінницький національний технічний університет

Субін А.А.<https://orcid.org/0000-0002-5747-2397>

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ У ВІЗУАЛІЗАЦІЇ СКЛАДНИХ НАУКОВИХ ДАНИХ

У статті розглянуто сучасні інформаційні технології, що застосовуються для візуалізації складних наукових даних у різних галузях досліджень, де обсяг, багатовимірність і різноманітність інформації ускладнюють її сприйняття та аналіз. В умовах стрімкого розвитку цифрових технологій і зростання масивів даних традиційні табличні та текстові форми подання інформації втрачають ефективність, що зумовлює необхідність використання візуальних методів як інструменту підтримки наукового пізнання та ухвалення обґрунтованих рішень. Особливу увагу приділено ролі інформаційних технологій у трансформації складних числових, просторових і структурованих даних у наочні графічні образи, здатні відображати приховані закономірності, взаємозв'язки та динамічні процеси. Проаналізовано основні підходи до візуалізації наукових даних, зокрема використання інтерактивної графіки, багатовимірних діаграм, тривимірного моделювання та візуальної аналітики, які базуються на сучасних програмних платформах і алгоритмах обробки даних. Розглянуто можливості застосування інформаційних технологій для підвищення точності інтерпретації результатів досліджень, зменшення когнітивного навантаження на користувача та забезпечення інтерактивної взаємодії з даними. Окрему увагу приділено використанню алгоритмічних методів обробки великих масивів даних і елементів штучного інтелекту, що дозволяють автоматизувати процес побудови візуальних моделей і адаптувати їх до потреб конкретних наукових завдань. Обґрунтовано, що ефективна візуалізація складних наукових даних сприяє підвищенню якості аналізу, пришвидшенню процесу отримання нових знань і розширенню можливостей міждисциплінарних досліджень. Показано, що застосування сучасних інформаційних технологій у сфері візуалізації є важливим чинником підвищення наукової результативності та практичної цінності досліджень. Отримані результати можуть бути використані в науково-дослідній діяльності, освітньому процесі, а також під час розроблення програмних засобів для аналізу та представлення складних даних у різних предметних галузях.

Ключові слова: багатовимірні моделі, наочність, інтерактивна графіка, програмне забезпечення, аналіз інформації.

Постановка проблеми. Сучасний етап розвитку науки характеризується стрімким зростанням обсягів даних, ускладненням їх структури й багатовимірністю, що істотно ускладнює процеси аналізу, інтерпретації та представлення результатів досліджень. У багатьох наукових галузях дослідники працюють зі значними масивами числової, просторової та динамічної інформації, сприй-

няття якої в традиційних табличних або текстових формах є малоефективним. У таких умовах виникає потреба в інструментах, здатних забезпечити наочне, зрозуміле та водночас інформативне подання складних наукових даних.

Особливої актуальності набуває проблема візуалізації даних у контексті розвитку інформаційних технологій (далі – ІТ), коли швидкість отримання результатів

і точність їх інтерпретації безпосередньо впливають на якість наукових висновків і ухвалюваних рішень. Недостатньо продумана або формальна візуалізація може призводити до втрати важливої інформації, спотворення закономірностей і ускладнення міждисциплінарної комунікації. Водночас сучасні програмні засоби та алгоритми обробки даних надають нові можливості для перетворення складних інформаційних структур у зрозумілі візуальні образи.

У цьому контексті актуалізується необхідність упровадження та розвитку ІТ, спрямованих на вдосконалення візуального подання результатів досліджень і підвищення якості їх інтерпретації. Застосування таких інструментів забезпечує більш чітке сприйняття отриманих результатів, оптимізує аналітичні процедури та створює сприятливі умови для ухвалення обґрунтованих наукових і практичних рішень.

Аналіз останніх досліджень і публікацій. У сучасних наукових дослідженнях використання ІТ для візуалізації складних наукових даних розглядається як важливий чинник підвищення ефективності аналізу та інтерпретації результатів. Зокрема, Ю. Берщанський (Y. Bershchanskyi), Г. Клим (H. Klym), Ю. Шевчук (Y. Shevchuk) акцентують на архітектурних рішеннях контейнеризованих інтелектуальних систем, доводячи, що поєднання хмарних і кіберфізичних технологій створює передумови для обробки та наочного подання великих і різномірних масивів даних [1].

Питання забезпечення якості вхідних даних як основи для коректної візуалізації розкривають І. Опіський, Ю. Хохлачова, А. Стефанків, Ю. Шевчук та І. Гунько (I. Hunko), підкреслюючи значення безпечної обробки інформації й оптимізації програмних процесів для формування надійних аналітичних і візуалізаційних інструментів [2; 3].

Розвиток інтелектуальних підходів до інтеграції та аналізу даних аналізують Д. Кияшко (D. Kyiashko), Д. Солодовник (D. Solodovnyk), І. Гаврилюк (I. Havryliuk), І. Сипченко (I. Sypchenko), О. Іщенко (O. Ishchenko), О. Кириленко (O. Kurylenko), обґрунтовуючи потенціал агентних моделей та ефективних візуальних рішень для автоматизації аналізу й підвищення зрозумілості складних інформаційних структур [4; 5].

Прикладні аспекти обробки й візуалізації даних у різних предметних галузях досліджують С. Ковбасюк, С. Легенчук, В. Янчук, С. Раєва, О. Нестеренко, В. Поліщук, С. Жарінов, В. Лойко, Л. Бобер, наголошуючи, що візуалізація сприяє виявленню закономірностей, узагальненню показників і підтримці управлінських рішень [6–8].

Трансформацію візуальних форм подання інформації аналізує Д. Солодовник, підкреслюючи необхідність використання новітніх візуалізаційних підходів у зв'язку зі зростанням обсягів і складності даних [9]. Водночас Н. Яськова, Ю. Лабжинський зазначають, що володіння інструментами візуалізації є важливим складником цифрової компетентності сучасного дослідника [10].

Безпосередній взаємозв'язок великих даних і комп'ютерної графіки розкривають О. Романюк та співавтори, доводячи значення поєднання алгоритмічних методів і графічних засобів для візуалізації багатовимірних наукових даних [11]. Узагальнення світових тенденцій розвитку технологій візуалізації демонструють А. Тютюнник, О. Каракоз, М. Диндин, О. Рамський, В. Диндин, Л. Варченко-Троценко, О. Буйницька, В. Смірнова, підкреслюючи роль візуалізації й штучного інтелекту в підвищенні якості наукового аналізу та інтерпретації результатів [12–16].

Попри наявні дослідження, у науковій літературі залишаються недостатньо опрацьованими питання комплексного поєднання ІТ, аналітичних методів і засобів візуалізації в роботі зі складними науковими даними. Здебільшого увага вчених зосереджена на окремих технічних або прикладних аспектах без формування цілісного підходу до вибору візуалізаційних рішень залежно від типу даних і цілей дослідження. Недостатньо висвітленими залишаються проблеми масштабованості, інтерпретованості та інтеграції підходів із використанням технологій штучного інтелекту (далі – ШІ) у візуальні моделі.

У цьому контексті внесок даної статті полягає в систематизації сучасних ІТ візуалізації складних наукових даних та обґрунтуванні їхньої ролі на етапах аналізу, інтерпретації й представлення результатів досліджень, що дозволяє визначити перспективні напрями розвитку ефективних і масштабованих візуалізаційних рішень.

Постановка завдання. Метою статті є теоретичний аналіз та узагальнення підходів до застосування ІТ у візуалізації складних наукових даних із метою підвищення ефективності їх аналізу, інтерпретації та представлення результатів досліджень. Особливу увагу зосереджено на виявленні можливостей сучасних програмних засобів і алгоритмічних рішень щодо трансформації багатовимірних та структурно складних даних у наочні візуальні форми, які сприяють глибшому розумінню досліджуваних процесів.

Для досягнення поставленої мети у статті передбачено розв'язання таких завдань:

1) проаналізувати особливості та обмеження традиційних методів подання наукових даних у контексті зростання їхніх обсягів і складності;

2) дослідити сучасні інформаційні технології та інструменти, що застосовуються для візуалізації складних наукових даних;

3) визначити переваги використання інтерактивних і багатовимірних візуальних моделей для аналізу наукової інформації;

4) узагальнити напрями використання візуалізації даних у наукових дослідженнях і сформулювати практичні рекомендації щодо підвищення ефективності представлення результатів досліджень.

Виклад основного матеріалу. Візуалізація наукових даних у сучасних дослідженнях розглядається як міждисциплінарний метод, що поєднує ІТ, алгоритмічну обробку даних і принципи когнітивного сприйняття. Її сутність полягає в трансформації складних числових, просторових та структурованих інформаційних масивів у візуальні образи, здатні відобразити внутрішні закономірності, взаємозв'язки та динаміку досліджуваних процесів. На відміну від традиційних табличних або текстових форм подання, візуалізація виконує не лише ілюстративну, а й аналітичну функцію, забезпечуючи активне залучення дослідника до процесу інтерпретації даних [11, с. 51; 12, с. 162].

У науковому пізнанні візуалізація виконує низку важливих функцій, зокрема когнітивну (сприяє спрощенню сприйняття складних багатовимірних даних), аналітичну (дозволяє виявляти приховані закономірності та аномалії), комунікативну (забезпечує ефективну репрезентацію результатів досліджень) та евристичну (підтримує формування й перевірку гіпотез). Як зазначають дослідники, саме поєднання цих функцій перетворює візуалізацію на повноцінний інструмент наукового аналізу [14, с. 38].

Сучасні наукові дослідження дедалі частіше базуються на опрацюванні складних інформаційних масивів, для яких характерні значні обсяги, багатовимірність і структурна різноманітність. Такі дані формуються внаслідок чисельного моделювання, проведення експериментальних вимірювань, сенсорного моніторингу, аналізу великих статистичних баз, а також функціонування кіберфізичних систем. У цих умовах традиційні способи подання інформації у вигляді таблиць або текстових описів не забезпечують належного рівня наочності та ускладнюють процес виявлення закономірностей, взаємозв'язків і динамічних змін.

Водночас у процесі роботи з великими та різноманітними науковими даними візуалізація є ефективним інструментом зниження когнітивного навантаження на дослідника. Графічне подання інформації дозволяє агрегувати значні обсяги даних, зосереджуючи увагу на найбільш інформативних характеристиках і зв'язках між ними. Унаслідок цього підвищується швидкість аналізу та зменшується ймовірність помилкової інтерпретації результатів [6, с. 114; 9, с. 159].

Крім того, візуалізація відіграє важливу роль у міждисциплінарній науковій комунікації, оскільки універсальні візуальні моделі дозволяють представляти складні результати досліджень фахівцям із різних галузей без втрати їхньої змістової повноти [5, с. 111; 10, с. 52].

У цьому контексті візуалізація є проміжною ланкою між первинними даними та процесом наукового пізнання, поєднуючи етапи збору, обробки та інтерпретації інформації в єдиний аналітичний цикл. Візуальні моделі дають змогу досліднику не лише спостерігати результати обчислень, а й активно взаємодіяти з даними, коригувати параметри аналізу та формувати нові дослідницькі гіпотези. Такий підхід перетворює візуалізацію з простого засобу подання результатів на ефективний інструмент підтримки наукових рішень [12, с. 164; 14, с. 38].

Для наочного відображення відмінностей між традиційними та сучасними підходами до подання наукових даних доцільно здійснити їх порівняльний аналіз за основними критеріями (табл. 1).

Наведене порівняння підтверджує доцільність застосування сучасних підходів візуалізації в роботі зі складними науковими даними, оскільки вони забезпечують вищу інформативність та зручність інтерпретації результатів досліджень.

Інформаційно-технологічні засоби візуалізації спрямовані на перетворення складних числових, просторових і структурованих даних у графічні образи, адаптовані до особливостей людського сприйняття. Використання двовимірних та багатовимірних діаграм, теплових карт, графів, часових рядів і просторових моделей дозволяє значно підвищити інформативність подання результатів досліджень та знизити когнітивне навантаження на користувача. При цьому візуалізація розглядається не лише як ілюстративний засіб, а як повноцінний інструмент аналітичного пізнання.

Загальний процес перетворення складних наукових даних у наочні візуальні образи подано у вигляді узагальненої схеми, що наведена на рисунку 1.

Порівняння традиційних і сучасних методів подання наукових даних

Критерій оцінювання	Традиційні методи (таблиці, текст)	Сучасні візуалізаційні підходи
Рівень наочності	Обмежений, потребує додаткового аналізу	Високий, орієнтований на зорове сприйняття
Робота з багатовимірними даними	Ускладнена або неможлива	Забезпечується за допомогою спеціалізованих моделей
Швидкість сприйняття інформації	Низька	Висока
Когнітивне навантаження	Значне	Знижене
Інтерактивна взаємодія	Відсутня	Забезпечується в реальному часі

Джерело: сформовано авторами

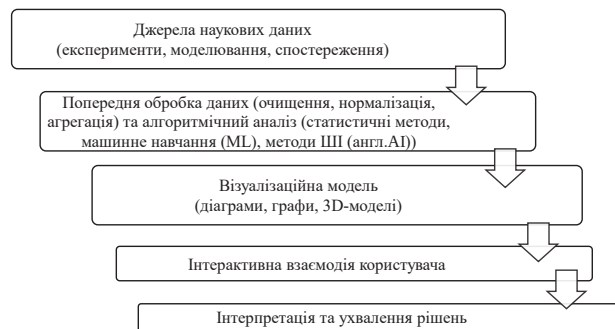


Рис. 1. Узагальнена схема процесу візуалізації складних наукових даних

Джерело: сформовано авторами

Представлена схема демонструє комплексний характер процесу візуалізації, який поєднує алгоритмічну обробку даних і графічні засоби подання інформації для підтримки наукового аналізу.

Важливою характеристикою сучасних систем візуалізації є їхня інтерактивність. Інтерактивна графіка надає можливість змінювати параметри відображення даних у реальному часі, здійснювати масштабування, фільтрацію, деталізацію та порівняльний аналіз різних наборів даних. Такий підхід дозволяє дослідникові адаптувати візуальне подання до конкретного наукового завдання та швидко перевіряти гіпотези шляхом безпосередньої взаємодії з даними. Реалізація інтерактивних механізмів базується на сучасних програмних платформах і архітектурних рішеннях, зокрема хмарних і розподілених середовищах [1, с. 155].

Основні ІТ, що застосовуються для візуалізації складних наукових даних, та їхні функціональні можливості узагальнено в табл. 2.

Узагальнення ІТ візуалізації свідчить про їхню універсальність і можливість адаптації до різних типів наукових даних і дослідницьких завдань.

Сучасні ІТ візуалізації реалізуються на основі різних програмних і апаратних платформ, зокрема хмарних середовищ, розподілених обчислювальних систем та спеціалізованих графічних бібліотек. Інтерактивні візуалізаційні інструменти

дають змогу поєднувати аналіз даних із безпосередньою візуальною взаємодією користувача з інформацією, що є особливо важливим для дослідження складних і динамічних процесів [1, с. 154; 8, с. 627].

Подальший розвиток таких технологій пов'язаний із впровадженням інтелектуальних алгоритмів, здатних автоматично адаптувати способи візуального подання до властивостей даних і завдань дослідження. Це підтверджує тенденцію переходу від статичних графіків до інтерактивних та інтелектуальних систем візуальної аналітики [4; 16, с. 41].

Водночас ефективність застосування сучасних технологій візуалізації здебільшого визначається дотриманням принципів коректності та інтерпретованості візуальних моделей. Надмірна складність графічних представлень, невдало обрані шкали або способи кодування інформації можуть призвести до спотворення результатів аналізу та хибних наукових висновків. З огляду на це, сучасні підходи до візуалізації передбачають баланс між інформативністю, наочністю та когнітивною доступністю, що особливо важливо при роботі з великими й багатовимірними даними [11, с. 52; 9, с. 160].

Окреме місце у візуалізації складних наукових даних займають багатовимірні та тривимірні моделі. Вони застосовуються для відображення процесів і об'єктів, параметри яких не можуть бути адекватно представлені у двовимірному просторі. Тривимірне моделювання дає змогу відтворювати просторові структури, візуалізувати часову динаміку та аналізувати взаємодію елементів складних систем. Поєднання методів комп'ютерної графіки з алгоритмами обробки великих даних дозволяє створювати наочні візуальні представлення багатовимірних інформаційних масивів без втрати їхньої змістової повноти [11, с. 54].

Для систематизації сучасних підходів до візуалізації складних наукових даних доцільно розглянути їхню класифікацію, що подана на рисунку 2.

Основні ІТ візуалізації складних наукових даних

Інформаційна технологія	Типи даних	Основні можливості	Сфера застосування
Інтерактивна графіка	Числові, часові	Масштабування, фільтрація, порівняння	Науковий аналіз
Багатовимірні діаграми	Багатовимірні	Виявлення взаємозв'язків	Дослідження складних систем
Тривимірне моделювання	Просторові	Відображення структури та динаміки	Моделювання процесів
Візуальна аналітика	Великі масиви	Пошук закономірностей	Big Data
AI-орієнтовані системи	Різнорідні	Автоматизація візуалізації	Інтелектуальні системи

Джерело: сформовано авторами

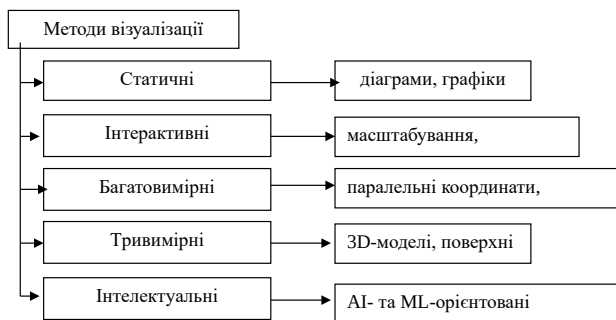


Рис. 2. Класифікація методів візуалізації складних наукових даних

Джерело: створено авторами

Запропонована класифікація відображає різноманіття сучасних методів візуалізації та підтверджує тенденцію до інтеграції інтелектуальних підходів у процес аналізу складних наукових даних.

Ефективність візуалізації значно залежить від якості попередньої обробки даних. Алгоритмічні методи нормалізації, агрегації, кластеризації та редукції ознак відіграють важливу роль у підготовці інформації до візуального подання. Застосування таких методів дозволяє усунути надлишковість, виокремити суттєві ознаки та зосередити увагу користувача на найбільш значущих аспектах досліджуваного явища. При цьому важливим є забезпечення цілісності та достовірності даних на всіх етапах обробки, оскільки помилки або спотворення можуть негативно вплинути на результати візуального аналізу [2, с. 224].

Особливе місце серед методів попередньої обробки даних займає кластеризація, яка дозволяє автоматизовано групувати об'єкти за схожістю ознак без попереднього визначення класів. У контексті візуалізації складних наукових даних кластеризація використовується для зменшення складності інформаційних масивів, виявлення структурної організації даних та формування узагальнених візуальних представлень. Резуль-

тати кластерного аналізу часто відображаються у вигляді теплових карт, дендрограм, багатовимірних проєкцій або кольорових груп, що значно спрощує інтерпретацію багатовимірних наборів даних [11, с. 55].

Застосування кластеризації є особливо ефективним у роботі з великими статистичними та медичними даними, де візуальне групування дозволяє швидко ідентифікувати типові та атипові патерни, а також підтримувати прийняття рішень на основі візуальної аналітики [6, с. 116; 7, с. 139].

Поєднання методів кластеризації з інтерактивною візуалізацією дозволяє не лише виявляти структурні групи в даних, а й аналізувати їхню поведінку в динаміці, оцінювати стабільність кластерів та вплив окремих параметрів на результати групування. Такий підхід є особливо актуальним для дослідження складних соціально-економічних, медичних і технічних систем, де візуальна інтерпретація кластерів значно підвищує прозорість аналітичних висновків і сприяє прийняттю обґрунтованих управлінських рішень [8, с. 629].

Подальший розвиток ІТ візуалізації пов'язаний із використанням методів ШІ та машинного навчання. Інтелектуальні алгоритми дозволяють автоматизувати процес побудови візуальних моделей, адаптувати їх до особливостей конкретних наборів даних і користувацьких запитів, а також підтримувати прийняття рішень на основі візуальної аналітики. Застосування агентних і мультимодальних підходів створює можливості для інтеграції різнорідних даних та побудови комплексних візуалізацій, орієнтованих на дослідження складних наукових процесів [4; 16, с. 42].

Практичне значення ІТ візуалізації проявляється в різних сферах наукової діяльності – від аналізу статистичних і медичних даних до оцінювання ефективності наукових установ і управління складними проєктами. У таких випадках візуалізація є інструментом узагальнення результатів, підви-

щення їх зрозумілості та забезпечення міждисциплінарної комунікації. Графічне подання інформації дозволяє дослідникам швидше орієнтуватися у великих масивах даних і формувати обґрунтовані висновки [6, с. 115; 7, с. 138; 8, с. 629].

Отже, використання сучасних ІТ у візуалізації складних наукових даних забезпечує поєднання аналітичної точності та наочності подання результатів. Інтерактивні, багатовимірні та інтелектуальні візуальні моделі сприяють глибшому розумінню досліджуваних процесів, зменшенню когнітивного навантаження та підвищенню ефективності наукового аналізу. Це створює передумови для подальшого розвитку візуалізації як важливого інструменту сучасних наукових досліджень і практичного застосування отриманих результатів.

Висновки. У процесі дослідження здійснено теоретичний аналіз і узагальнення сучасних підходів до застосування ІТ у візуалізації складних наукових даних. Показано, що стрімке зростання обсягів, багатовимірності та різноманітності інформації обмежує ефективність традиційних табличних і текстових форм подання результатів та зумовлює необхідність використання сучасних візуалізаційних рішень.

З'ясовано, що ІТ-візуалізації перетворюють складні числові, просторові та структуровані дані в наочні графічні образи, які відповідають

особливостям людського сприйняття та сприяють виявленню закономірностей, взаємозв'язків і динамічних процесів. Застосування інтерактивної графіки, багатовимірних і тривимірних моделей, а також засобів візуальної аналітики дозволяє зменшити когнітивне навантаження та підвищити точність інтерпретації результатів.

Обґрунтовано важливість алгоритмічних методів попередньої обробки даних, таких як нормалізація, агрегація, кластеризація та редукція ознак, оскільки якість цих етапів безпосередньо впливає на інформативність і достовірність візуальних моделей. Вказано на перспективи використання методів ШІ та машинного навчання для автоматизації процесів побудови візуалізацій і їх адаптації до специфіки конкретних наукових завдань.

Узагальнено, що ефективна візуалізація складних наукових даних є не лише засобом подання результатів, а й повноцінним інструментом аналітичного пізнання, який сприяє підвищенню якості наукового аналізу та підтримці прийняття обґрунтованих рішень.

Подальші дослідження в цій галузі можуть бути пов'язані з удосконаленням інтерактивних і адаптивних візуалізаційних рішень, інтеграцією методів машинного навчання та кластерного аналізу, а також оцінюванням їх масштабованості та інтерпретованості для різноманітних наукових даних.

Список літератури:

1. Bershchanskyi Y., Klym H., Shevchuk Y. Containerized Artificial Intelligent System Design in Cloud and Cyber-Physical Systems. *Advances in Cyber-Physical Systems*. 2024. Vol. 9. № 2. P. 151–157. DOI: <https://doi.org/10.23939/acps2024.02.151>
2. Опірський І. Р., Хохлячова Ю. Є., Стефанків А. В., Шевчук Ю. А. Аналіз технічних особливостей реалізації шифрування даних на SD-картах в Android. *Сучасний захист інформації*. 2025. № 1(61). С. 219–228. DOI: <https://doi.org/10.31673/2409-7292.2025.016526>
3. Hunko I. How to Effectively Reduce Software Testing Time: From Requirements to Regression. Lodz, Poland, 2024. URL: <https://futura-publiishing.com/wp-content/uploads/2025/04/7П-29.03.25-3.pdf> (дата звернення: 13.01.2026).
4. Kyiashko D. Development of a hybrid testing framework for multimodal systems based on ai agents. *Наука і техніка сьогодні*. 2025. № 11(52). С.1774–1788. DOI: [https://doi.org/10.52058/2786-6025-2025-11\(52\)-1774-1788](https://doi.org/10.52058/2786-6025-2025-11(52)-1774-1788)
5. Solodovnyk D., Havryliuk I., Sypchenko I., Ishchenko O., Kyrylenko O. Data journalism visualization in Ukraine and in Europe: a comparative analysis. *Amazonia Investiga*. 2023. Vol. 12. № 64. P. 102–116. DOI: <https://doi.org/10.34069/AI/2023.64.04.10>
6. Ковбасюк С.В., Легенчук С.Ф., Янчук В.О., Раєва С.Л. Інформаційна технологія обробки та візуалізації статистичних даних захворюваності на COVID-19. *Технічна інженерія*. 2023. № 2(92). С. 113–117. DOI: [https://doi.org/10.26642/ten-2023-2\(92\)-113-117](https://doi.org/10.26642/ten-2023-2(92)-113-117)
7. Нестеренко О. В., Поліщук В. Б., Жарінов С. С. Застосування інтегративної інформаційної технології в процесах оцінювання наукових установ. *Екологічна безпека та природокористування*. 2024. № 1 (49). С. 126–142. DOI: <https://doi.org/10.32347/2411-4049.2024.1.126-142>
8. Лойко В. В., Бобер Л. П. Цифрові інструменти візуалізації даних в організаційному управлінні портфелями проєктів. *Європейський науковий журнал Економічних та Фінансових інновацій*. 2025. № 4(18). С. 624–635. URL: <https://journal.eae.com.ua/index.php/journal/article/view/660/501> (дата звернення: 13.01.2026).

9. Солодовник Д. Візуалізація даних у новітніх конвергентних медіа України. *Синопсис: текст, контекст, медіа*. 2023. № 29(2). С. 157–165. DOI: <https://doi.org/10.28925/2311-259x.2023.2.13>
10. Яськова Н., Лабжинський Ю. Розвиток цифрової компетентності наукових і науково-педагогічних працівників засобами електронних соціальних мереж. *Фізико-математична освіта*. 2024. № 39(5). С. 46–57. DOI: <https://doi.org/10.31110/fmo2024.v39i5-07>
11. Романюк О. Н., Павлов С. В., Бобко О. Л., Завальнюк Є. К., Решетнік О. О. Аналіз великих даних у комп'ютерній графіці. *Опτικο-електронні інформаційно-енергетичні технології*. 2024. № 1 (47). С. 50–57. DOI: <https://doi.org/10.31649/1681-7893-2024-47-1-50-57>
12. Тютюнник А. В. Технології візуалізації у світових дослідженнях. *Відкрите освітнє е-середовище сучасного університету*. 2020. № 9. С. 161–168. DOI: <https://doi.org/10.28925/2414-0325.2020.9.13>
13. Каракоз О. Візуалізація інформації у сучасному освітньому просторі. *Грааль науки*. 2023. № 24. С. 800–803. DOI: <https://doi.org/10.36074/grail-of-science.17.02.2023.149>
14. Диндин М. Л., Рамський О. І., Диндин В. Л. Використання ІТ у наукових дослідженнях. *Науковий вісник ЛНУВМБ імені С. З. Гжицького. Серія: Економічні науки*. 2025. Т. 27. № 105. С. 35–40. DOI: <https://doi.org/10.32718/nvlvet-e10506>
15. Варченко-Троценко Л. О., Тютюнник А. В. Використання візуалізації даних в системі електронного навчання. *Відкрите освітнє е-середовище сучасного університету*. 2022. № 12. С. 21–31. DOI: <https://doi.org/10.28925/2414-0325.2022.122>
16. Буйницька О., Смірнова В. Технології штучного інтелекту в дослідницькій діяльності: огляд та застосування. *Неперервна професійна освіта: теорія і практика*. 2024. № 81(4). С. 31–46. DOI: <https://doi.org/10.28925/2412-0774.2024.4.2>

Kozak R.R., Dudnyk O.O., Subin A.A. INFORMATION TECHNOLOGIES IN THE VISUALIZATION OF COMPLEX SCIENTIFIC DATA

The article examines modern information technologies used for the visualization of complex scientific data across various fields of research, where large volumes, multidimensionality, and data heterogeneity significantly complicate perception and analysis. In the context of the rapid development of digital technologies and the continuous growth of data sets, traditional tabular and textual forms of data representation are becoming increasingly ineffective, which necessitates the use of visual methods as tools to support scientific understanding and informed decision-making. Particular attention is given to the role of information technologies in transforming complex numerical, spatial, and structured data into clear graphical representations capable of revealing hidden patterns, relationships, and dynamic processes. The study analyzes key approaches to scientific data visualization, including the use of interactive graphics, multidimensional charts, three-dimensional modeling, and visual analytics based on modern software platforms and data processing algorithms. The potential of information technologies to enhance the accuracy of research result interpretation, reduce users' cognitive load, and ensure interactive engagement with data is also explored. Special emphasis is placed on algorithmic methods for processing large data sets and the application of artificial intelligence components, which enable the automation of visual model construction and their adaptation to the specific requirements of scientific tasks. It is substantiated that effective visualization of complex scientific data contributes to higher analytical quality, accelerates the process of knowledge generation, and expands opportunities for interdisciplinary research. The paper demonstrates that the application of modern information technologies in data visualization is an important factor in increasing the scientific effectiveness and practical value of research outcomes. The obtained results may be applied in scientific research activities, the educational process, as well as in the development of software tools for analyzing and presenting complex data across various subject domains.

Keywords: multidimensional models, visual representation, interactive graphics, software, information analysis.

Дата першого надходження статті до видання: 26.01.2026

Дата прийняття статті до друку після рецензування: 23.02.2026

Дата публікації (оприлюднення) статті: 00.00.2026

Корабльов М.М.

<https://orcid.org/0009-0005-2540-7741>

Харківський національний університет радіоелектроніки

Дикий С.А.

<https://orcid.org/0009-0007-5396-2413>

Харківський національний університет радіоелектроніки

ДІАГНОСТИКА ЕМОЦІЙНОГО СТАНУ ДИТИНИ НА ОСНОВІ АНАЛІЗУ ЇЇ МАЛЮНКІВ З ВИКОРИСТАННЯМ НЕЙРОМЕРЕЖЕВОВОГО ПІДХОДУ

Розглянуто підхід до автоматизованого первинного оцінювання психоемоційного стану дитини за зображенням її малюнка. Проективні малюнкові методики широко застосовують у практиці, однак результати візуальної інтерпретації залежать від досвіду фахівця та можуть бути неоднорідними. Запропоновано використовувати згорткову нейронну мережу для віднесення малюнка до однієї з трьох узагальнених категорій: «Щастя», «Тривожність/депресивність», «Гнів/агресія». Вхідними даними є кольорові зображення JPEG, приведені до розміру 224×224 та нормалізовані в діапазон $[0; 1]$. Архітектура моделі містить два згорткові шари з ядрами 7×7 (16 фільтрів) та 5×5 (32 фільтри) зі кроком 1 і активацією ReLU; шар підвибірки (pooling) цілеспрямовано не використано для збереження тонких графічних деталей, важливих для емоційної класифікації. Отримані карти ознак перетворюються у вектор і подаються на три повноз'язні шари (256, 128, 3) з регуляризацією Dropout 0,5; на виході застосовано Softmax, що формує інтерпретовані ймовірності належності до класів. Навчання здійснюється протягом 10 epoch із поділом на тренувальну та валідаційну частини; як функцію втрат використано крос-ентропію та алгоритм зворотного поширення похибки. Експериментальні приклади демонструють здатність моделі відокремлювати характерні зразки дитячих малюнків із різним емоційним забарвленням і можуть бути основою для створення допоміжного інструмента для психологів та педагогів у задачах скринінгу. Окремо наголошено на можливості інтеграції класифікатора у сервіс відкритого доступу та на необхідності етичного використання результатів у межах скринінгу. Показано приклад прикладного використання: сервіс повертає ймовірності класів і дає змогу налаштувати поріг впевненості для скринінгу, з обов'язковим етичним трактуванням результатів.

Ключові слова: дитячий малюнок, емоційний стан, згорткова нейронна мережа, обробка та класифікація зображень, машинне навчання.

Постановка проблеми. Зростання інформаційного навантаження, інтенсивність соціальних змін і тривалі стресові впливи посилюють потребу в ранньому виявленні емоційних та поведінкових труднощів у дітей. Своєчасна підтримка є критичною для профілактики хронічних проявів тривожності, депресивності та агресивної поведінки, які можуть відображатися як у спілкуванні, так і в навчанні.

Проективні методики, зокрема малюнкові тести, дозволяють дитині непрямо виразити внутрішні переживання через символи, композицію, інтенсивність штрихування та колірні рішення. На практиці такі методики корисні, але їх інтерпретація часто залежить від суб'єктивних чин-

ників, тому результати можуть відрізнятися між фахівцями [1, 2].

Сучасні методи комп'ютерного зору та глибокого навчання дають змогу зменшити вплив людського фактора шляхом побудови моделей, що навчаються на розмічених прикладах і відтворювано оцінюють зображення. Тому актуальним є створення комп'ютерної системи, яка автоматично аналізує дитячі малюнки та пропонує класифікацію емоційного стану як допоміжний елемент скринінгу (а не як заміну клінічного висновку) [3].

Аналіз останніх досліджень і публікацій. Наукова література підтверджує, що дитячі малюнки містять інформативні індикатори

емоційного стану та особливостей розвитку. Дослідники аналізують тематику, вибір кольорів, наявність або відсутність позитивних/негативних символів, просторову організацію об'єктів, а також динаміку процесу малювання [4-7].

З позицій когнітивного розвитку підкреслюється роль виконавчих функцій, мовлення та формування репрезентативного малюнка у дошкільному віці; ці компоненти впливають на те, які образи дитина здатна відтворити та як вона кодує переживання у графічній формі [5]. Роботи, присвячені аналітичним малюнкам, також наголошують на важливості діалогу з дитиною під час інтерпретації малюнка, адже пояснення автора допомагає зменшити неоднозначність символів [8].

Паралельно розвивається напрям інтелектуального аналізу малюнків із використанням машинного навчання. Узагальнені огляди показують, що моделі глибинного навчання можуть ефективно виділяти візуальні патерни у складних, неструктурованих зображеннях, зокрема у рукописних і мальованих даних [9-12]. Водночас для дитячих малюнків характерні високий рівень варіативності стилю, нестандартні пропорції та наявність «високорівневих» символічних елементів, що ускладнює задачу розпізнавання [13].

У прикладних роботах пропонуються мобільні або вебзастосунки, які аналізують малюнок і повертають користувачу інтерпретовані результати. Проте такі системи потребують ретельної валідації, прозорого опису архітектури та обмежень, а також узгодження з етичними принципами оброблення даних дітей [14]. Тому актуальним лишається підхід, що поєднує архітектуру моделі, контроль навчання та приклади інтерпретації результатів.

Постановка завдання. Метою дослідження є розроблення та опис нейромережевої моделі для первинної автоматизованої діагностики емоційного стану дитини на основі аналізу її малюнків. Для досягнення мети необхідно: 1) сформулювати постановку задачі класифікації дитячих малюнків за трьома узагальненими категоріями; 2) спроектувати згорткову нейронну мережу, адаптовану до мальованих зображень; 3) описати підготовку вхідних даних, процедуру навчання та валідації; 4) навести результати експериментальних досліджень на прикладах і сформулювати обмеження та напрями розвитку.

Виклад основного матеріалу

1. Постановка задачі та категорії емоційних станів

Задачу сформульовано як багатокласову класифікацію зображення дитячого малюнка до

однієї з трьох категорій: «Щастя», «Тривожність/депресивність» та «Гнів/агресія». Таке узагальнення дає змогу зменшити суб'єктивність під час розмітки і водночас відобразити найпоширеніші полярні стани, що спостерігаються у проєктивних тестах [1].

Категорія «Щастя» зазвичай пов'язана з гармонійними композиціями, достатньою кількістю вільного простору, наявністю позитивних символів (сонце, веселка, природа, тварини), світлими відтінками та охайним виконанням. Подібні ознаки часто супроводжуються посмішками персонажів і сюжетами сімейної взаємодії.

Категорія «Тривожність/депресивність» частіше проявляється у збідненій палітрі, домінуванні сірого/темного, наявності хмар, дощу, сліз або самотніх фігур. Лінії можуть бути тремтячими, із повторним заштриховуванням, а сюжет – менш структурованим, що узгоджується з описами емоційних проявів у графічній діяльності [4, 8].

Категорія «Гнів/агресія» може містити сцени конфлікту, зброю, руйнування або виражено напружені пози персонажів. Для таких малюнків характерні різкі контури, хаотичне розміщення об'єктів, переважання червоних/темних тонів і підвищена інтенсивність штрихів. Саме поєднання композиційних і колірних маркерів робить автоматизований аналіз перспективним доповненням до експертної оцінки.

2. Підготовка даних та попередня обробка зображень

Вхідні дані являють собою кольорові зображення дитячих малюнків у форматі JPEG. Перед поданням у модель зображення масштабуються до 224×224 пікселів, що забезпечує фіксований розмір вхідного тензора та стабільність обчислень. Додатково виконується нормалізація значень пікселів у діапазон [0; 1], щоб полегшити оптимізацію під час навчання.

Кожне зображення розглядається як трьохканальна матриця (R, G, B). Така подача зберігає колірну інформацію, яка є важливою для аналізу емоційного забарвлення. У межах підходу передбачено формування тренувальної та валідаційної вибірок для контролю узагальнювальної здатності моделі.

Для ефективного завантаження даних використовується механізм пакетної подачі (batching), що дозволяє керувати порядком подання прикладів і використовувати паралельне завантаження у процесі навчання. Це зменшує накладні витрати та прискорює експерименти, що особливо важливо при роботі з зображеннями високої варіативності.

Практична інтерпретація проєктивних малюнків є неоднозначною, тому ключовим етапом підготовки вибірки є формалізація правил розмітки. Для кожної категорії («Щастя», «Тривожність/депресивність», «Гнів/агресія») доцільно фіксувати перелік орієнтовних візуальних маркерів – палітру, типові символи, композицію, інтенсивність і характер штрихування – та узгоджувати їх із психологічними описами проєктивних методик [1, 4]. Суперечливі або змішані сюжети варто відносити до групи підвищеного ризику помилки та розглядати повторно [8].

Щоб зменшити вплив суб'єктивності, розмітку бажано виконувати за єдиною інструкцією та контролювати узгодженість оцінок між розмітниками. У прикладних сценаріях особливо важливо відокремлювати «скринінгову» класифікацію від клінічного діагнозу: вихід моделі слід трактувати як сигнал для подальшої розмови з фахівцем, а не як остаточний висновок [3, 14].

3. Архітектура запропонованої згорткової нейронної мережі

Основою запропонованого підходу є згорткова нейронна мережа (CNN), здатна автоматично вибудовувати багаторівневу систему ознак: від елементарних візуальних примітивів (краї, текстурні фрагменти, напрям і густина штрихів) до більш узагальнених представлень, що відображають структурні частини об'єктів та композиційні закономірності зображення. Завдяки локальним згорткам і спільному використанню ваг, CNN добре підходить для аналізу зображень, де важливі як просторові співвідношення, так і повторювані патерни, тому CNN широко використовується в задачах комп'ютерного зору, зокрема для розпізнавання візуальних структур [15].

Розроблена архітектура складається з двох послідовних згорткових шарів і подальшого блоку повнозв'язних перетворень. На першому етапі застосовується шар із 16 фільтрами розміру 7×7 та нелінійністю ReLU. Використання відносно великого ядра на початку обробки дозволяє «захопити» ширший контекст локальних областей і стабільніше виділяти грубі ознаки: великі кольорові плями, домінантні форми, загальну конфігурацію елементів, а також контрастні межі між основними зонами малюнка. Другий згортковий шар містить 32 фільтри 5×5 і працює як уточнювальний рівень, концентруючись на тонших компонентах зображення – дрібніших контурах, особливостях ліній, варіаціях штрихування та локальних текстурних неоднорідностях.

Ключовою особливістю архітектури є відмова від операцій підвибірки (pooling). У типо-

вій CNN pooling зменшує просторову розмірність карт ознак і підвищує інваріантність до зсувів, однак для дитячих малюнків така інваріантність не завжди корисна: разом зі зменшенням роздільної здатності можуть втрачатися малопомітні, але інформативні сигнали. Зокрема, діагностичні значення можуть мати тонкі повторювані штрихи, нерівномірність натиску, дрібні деталі символів, а також характерні «мікропатерни» (щільність ліній, ритмічність, хаотичність контуру). Саме тому pooling у моделі вимкнено, щоб зберегти максимум локальної інформації у картах ознак та не «згладжувати» дрібні прояви, важливі для подальшої класифікації.

Після згорткових шарів карти ознак перетворюються у вектор ознак, який подається на послідовність трьох повнозв'язних шарів із розмірностями 256, 128 та 3. Таке зменшення розмірності поетапно формує компактне представлення зображення та підсилює роздільність класів у просторі ознак. Для підвищення узагальнювальної здатності та зниження ризику перенавчання між повнозв'язними шарами використано Dropout із імовірністю 0,5, що діє як регуляризатор шляхом випадкового «вимикання» частини нейронів під час навчання. Вихідний шар із функцією Softmax перетворює отримані логіти на ймовірності належності зразка до кожної з категорій, що дає змогу інтерпретувати результат як ранжування класів за впевненістю моделі.

Схематичне представлення структури моделі наведено на рис. 1. Ключові параметри та налаштування основних шарів узагальнено в табл. 1.

4. Навчання, функція втрат і процедура валідації

Навчання моделі здійснюється протягом 10 епох із чергуванням фаз навчання та валідації. У фазі навчання ваги оновлюються методом зворотного поширення похибки з використанням оптимізатора, який мінімізує значення функції втрат. Для багатокласової класифікації застосовується крос-ентропія, що стимулює збільшення ймовірності правильного класу та зменшення ймовірності інших.

Валідаційна фаза необхідна для незалежного оцінювання узагальнювальної здатності моделі на даних, що не використовувалися для оновлення ваг. Такий контроль дозволяє виявляти перенавчання та коригувати гіперпараметри архітектури (наприклад, інтенсивність Dropout або розмірність повнозв'язних шарів).

Після завершення навчання модель зберігається разом із архітектурою та вагами, що дозволяє швидко відновити класифікатор для аналізу

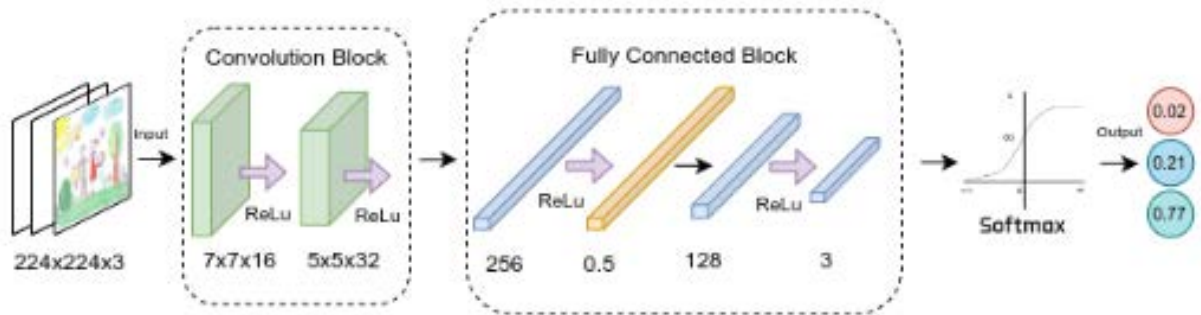


Рис. 1. Структура запропонованої згорткової нейронної мережі для класифікації емоційних станів

Таблиця 1

Узагальнені параметри шарів моделі

Шар	Тип	Параметри	Вихід	Призначення
L1	Conv2D+ReLU	16 фільтрів, 7×7, s=1	карти ознак	Грубі патерни (форми, текстури)
L2	Conv2D+ReLU	32 фільтри, 5×5, s=1	карти ознак	Деталі (штрихи, контури)
–	Pooling	не використовується	–	Збереження дрібних деталей
L3	FC+ReLU	256 нейронів	256	Агрегація та узагальнення
–	Dropout	p=0,5	–	Регуляризація
L4	FC+ReLU	128 нейронів	128	Компактне представлення
L5	FC	3 нейрони	3	Логіти трьох класів
L6	Softmax	–	ймовірності	Інтерпретовані оцінки

нових малюнків без повторного тренування. Це важливо для практичного використання у вигляді прикладного інструмента.

Під час навчання критично забезпечити відтворюваність експериментів: фіксувати випадкові зерна генераторів, протокол поділу даних, а також версії бібліотек і параметри попередньої обробки. Оскільки архітектура є компактною, її зручно тренувати в режимі мініпакетів із випадковим перемішуванням прикладів, контролюючи динаміку функції втрат на тренуванні та валідації. Регуляризація Dropout (p=0,5) зменшує ризик перенавчання, що особливо актуально для невеликих або неоднорідних за стилем наборів даних.

Для задачі скринінгу недостатньо контролювати лише загальну точність. Доцільно обчислювати повноту і точність для кожного класу та зведені показники (macro-F1), щоб оцінити, чи не «втрачає» модель рідкісні або складні категорії. Матриця помилок дозволяє зрозуміти типові плутання – наприклад, між «тривожністю/депресивністю» та «гнівом/агресією» у випадках із темною палітрою, але без явних символів конфлікту.

Результати експериментальних досліджень. Для ілюстрації роботи класифікатора наведено приклади малюнків, які не використовувалися в навчанні. На рис. 2 представлено зразок, що відповідає позитивному емоційному стану. Модель повернула найбільшу ймовірність для

класу «Щастя»: 0,7376; для «Тривожність/депресивність»: 0,1325; для «Гнів/агресія»: 0,1299.

На рис. 3 показано малюнок зі збідненою палітрою та виразними ознаками пригніченого стану. У цьому випадку домінує клас «Тривожність/депресивність» з ймовірністю 0,4534 (для «Щастя»: 0,3531; для «Гнів/агресія»: 0,1935).

На рис. 4 наведено зразок із сюжетною напруженою та ознаками конфлікту. Модель коректно віднесла його до класу «Гнів/агресія» з ймовірністю 0,5341 (для «Щастя»: 0,1737; для «Тривожність/депресивність»: 0,2922).

Ймовірнісний вихід Softmax зручний тим, що дозволяє оцінювати не лише клас, а й «впевненість» моделі. У прикладних сценаріях це дає змогу відокремлювати високонадійні передбачення від ситуацій, де потрібна поглиблена експертна інтерпретація.



Рис. 2. Приклад результату класифікації для класу «Щастя»



Рис. 3. Приклад результату класифікації для класу «Тривожність/депресивність»



Рис. 4. Приклад результату класифікації для класу «Гнів/агресія»

На відміну від «жорсткого» рішення, ймовірнісні оцінки Softmax дозволяють будувати прості та прозорі правила скринінгу. Практично корисно встановлювати два режими: 1) автоматичний – коли максимальна ймовірність перевищує заданий поріг і випадок можна віднести до певної категорії як попередню оцінку; 2) експертний – коли різниця між двома найімовірнішими класами мала і малюнок потребує додаткової інтерпретації психологом. Такий підхід зменшує ризик хибних «категоричних» висновків та підсилює прикладну цінність системи в освітньому середовищі.

Для підвищення довіри до моделі у практиці доцільно вести журнал результатів із мінімально необхідними метаданими (без персональних даних): дата, джерело зображення, оцінки класів, версія моделі. Це спростує аудит якості, виявлення дрейфу даних і планування донавчання. Критично важливо забезпечити інформовану згоду, а також обмежити використання результатів виключно цілями первинного скринінгу [3, 14].

Обговорення та перспективи подальших досліджень

Запропонована CNN-архітектура є порівняно легкою, прозорою для відтворення та придатною для розгортання в прикладних системах. Відмова від pooling підсилює чутливість до дрібних графічних деталей, що є доречним для аналізу дитячих малюнків. З іншого боку, така архітектура може вимагати ретельнішої регуляризації та контролю узагальнення, оскільки зростає ризик підлаштування під стилістичні особливості навчальної вибірки.

Важливим практичним аспектом є наявність етичних обмежень: результати автоматичного аналізу не мають розглядатися як медичний діагноз. Доцільно трактувати їх як підказку для скринінгу, що спрямовує увагу фахівця на потенційні ризики та допомагає пріоритизувати подальшу роботу з дитиною [3, 14].

Подальший розвиток може включати: а) розширення набору даних з урахуванням віку та контексту завдання; б) перехід до більш виразних архітектур або трансформерних моделей, які краще захоплюють глобальні залежності; в) додавання модулів інтерпретації (карти уваги, локалізація важливих областей), щоб підвищити довіру до рішень моделі та полегшити роботу практичних психологів [9].

Як приклад прикладної інтеграції розроблено прототип відкритого програмного засобу, що приймає зображення малюнка та повертає ймовірності класів разом із коротким поясненням. Такий інструмент може бути корисним в освітніх і психологічних установах для первинного аналізу.

Висновки. Розглянуто задачу автоматизованого аналізу дитячих малюнків для первинної оцінки емоційного стану дитини та підтримки роботи практичного психолога. Запропоновано класифікацію за трьома узагальненими категоріями, що охоплюють позитивний стан, пригнічення/тривогу та агресивні прояви. Такий поділ є зручним для прикладного скринінгу, оскільки дозволяє швидко отримувати попередню оцінку та відбирати роботи, які потребують детальнішого розгляду.

Запропоновано компактку архітектуру згорткової нейронної мережі з двома згортковими шарами та блоком повнозв'язних шарів. Особливістю рішення є відмова від pooling для збереження дрібних графічних деталей (штрихування, локальні текстури, контурні елементи), що підвищує інформативність ознак на вході класифікатора. Навчання протягом 10 епох із контролем

валідації та регуляризацією Dropout забезпечує стійку збіжність моделі й формування ймовірнісних оцінок для кожного класу.

На прикладах показано, що модель здатна відрізнити характерні зразки малюнків і повертає ймовірності класів, придатні для використання в прикладному інструменті первинного скринінгу. Практична цінність підходу полягає у можливості стандартизувати попередній аналіз, зменшити суб'єктивність на етапі відбору та прискорити

обробку великої кількості робіт у освітніх і консультативних установах, зберігаючи при цьому роль фахівця як остаточного інтерпретатора результатів.

Подальші дослідження доцільно спрямувати на розширення та балансування набору даних, підвищення інтерпретованості рішень (пояснення внеску фрагментів малюнка), а також на валідацію підходу для різних вікових груп і контекстів виконання малюнкових завдань.

Список літератури:

1. Farokhi M. The Analysis of Children's Drawings: Social, Emotional, Physical, and Psychological Aspects. *Procedia – Social and Behavioral Sciences*. 2011. Vol. 30. P. 2219–2224. DOI: 10.1016/j.sbspro.2011.10.433.
2. Liu Z., Ye C. C. Feature analysis of depression patients' house-tree-person drawings using convolutional neural networks. *Neuroscience Informatics*. 2025. Vol. 5, Issue 4. 100239. DOI: 10.1016/j.neuri.2025.100239.
3. Dan A. L. Comparative Evaluation of Machine Learning Algorithms for Affective State Recognition from Children's Drawings. *arXiv*. 2026. arXiv:2601.18414. DOI: 10.48550/arXiv.2601.18414.
4. Quaglia R., Longobardi E., Iotti N. O., Prino L. E. A new theory on children's drawings: Analyzing the role of emotion and movement in graphical development. *Infant Behavior and Development*. 2015. Vol. 39. P. 81–91. DOI: 10.1016/j.infbeh.2015.02.009.
5. Panesi S., Morra S. Executive function, language, and the toddler's discovery of representational drawing. *Frontiers in Psychology*. 2021. Vol. 12. Article 659569. DOI: 10.3389/fpsyg.2021.659569.
6. Fabris M. A., Lange-Küttner C., Shiakou M., Longobardi C. Editorial: Children's drawings: evidence-based research and practice. *Frontiers in Psychology*. 2023. Vol. 14. Article 105556. DOI: 10.3389/fpsyg.2023.105556.
7. Barrigade S., Desouza K., Zitter Z., Sarahu C. Drawing Play: A Content Analysis of Children's Drawings of Places Where They Like to Play. *Children, Youth and Environments*. 2023. Vol. 33(3). P. 63–89.
8. Podobnik U., Jerman J., Selan J. Understanding analytical drawings of preschool children: the importance of a dialog with a child. *International Journal of Early Years Education*. 2024. Vol. 32(1). P. 189–203. DOI: 10.1080/09669760.2021.1960802.
9. Beltzung J.-M., Pelé J., Renoult C., Sueur C. Deep learning for studying drawing behavior: a review. *Frontiers in Psychology*. 2023. Vol. 14. Article 992541. DOI: 10.3389/fpsyg.2023.992541.
10. Ramos-Aguilar R., Sánchez-Ruiz D., Rubí-Sánchez L. E., Ramos-Aguilar E., Rivera-Lima K., Suarez-Gonzalez G. Analysis of Children's Emotions in Sketches Using Classical and Deep Learning Approaches. *Pattern Recognition (MCPR 2025). Lecture Notes in Computer Science*. 2025. Vol. 15715. P. 90–100. DOI: 10.1007/978-3-031-96255-4_9.
11. Pysal S. A., Abdulkadir S. R. M., Shukri R. M., Alhussiani H. Classification of children's drawing strategies on the touch-screen using a novel deep learning hybrid model. *Alexandria Engineering Journal*. 2021. No. 1. P. 1115–1129. DOI: 10.1016/j.aej.2020.06.019.
12. Posley L., Powell J., Kim H. H., Thomas X., Liew J., Hammond T. Detecting Children's fine motor skill development using machine learning. *International Journal of Artificial Intelligence in Education*. 2021. Vol. 32(4). P. 991–1024. DOI: 10.1007/s40593-021-00279-7.
13. Brandt P.-Y., Dandarova-Robert Z., Cocco C., Vinck D., Darbellay F. When Children Draw Gods. *Multicultural and Interdisciplinary Approach to Children's Representations of Supernatural Agents*. Springer, 2023. DOI: 10.1007/978-3-030-94429-2.
14. Ali N., Abd-arrazzaq Z., Shah Z. та ін. Artificial Intelligence-Based Mobile Application for Sensing Children Emotion Through Drawings. *Advances in Informatics, Management, and Technology in Healthcare*. 2022. DOI: 10.3233/SHTI220675.
15. Lindsay G. W. Convolutional neural networks as a model of the visual system: past, present, and future. *Journal of Cognitive Neuroscience*. 2021. Vol. 33(10). P. 2017–2031. DOI: 10.1162/jocn_a_01544.

Korablyov M.M., Dykyi S.A. DIAGNOSIS OF A CHILD'S EMOTIONAL STATE BASED ON THE ANALYSIS OF HIS DRAWINGS USING A NEURAL NETWORK APPROACH

Approach considered to automated primary assessment of a child's psycho-emotional state using an image of the child's drawing. Projective drawing tests are widely applied in practice; however, visual interpretation depends on an expert's experience and may be inconsistent. We propose using a convolutional neural network to assign a drawing to one of three generalized categories: "Happiness", "Anxiety/Depression", and "Anger/Aggression". The input data are color JPEG images resized to 224×224 and normalized to the $[0; 1]$ range. The model architecture includes two convolutional layers with 7×7 (16 filters) and 5×5 (32 filters) kernels, stride 1, and ReLU activation; the pooling layer is intentionally omitted to preserve fine graphic details important for emotion classification. The resulting feature maps are flattened and passed to three fully connected layers (256, 128, 3) with Dropout 0.5 regularization; Softmax produces interpretable class probabilities. Training is performed for 10 epochs with train/validation split; cross-entropy loss and backpropagation are used for optimization. Experimental examples demonstrate the model's ability to separate typical samples of children's drawings with different emotional coloring, which can serve as a basis for an assisting screening tool for psychologists and educators. We also highlight the possibility of integrating the classifier into an open-access software service and the need for ethical use of the outputs in screening scenarios in educational practice. The probabilistic output supports screening workflows by applying a confidence threshold and referring uncertain cases to specialists.

Keywords: children's drawing, emotional state, convolutional neural network, processing and classification of images, machine learning.

Дата першого надходження статті до видання: 13.03.2026

Дата прийняття статті до друку після рецензування: 10.04.2026

Дата публікації (оприлюднення) статті 11.05.2026

UDC 004.932:681.518:006.9

DOI DOI DOI <https://doi.org/10.32782/2663-5941/2026.2.1/22>**Kuznietsov D.I.**<https://orcid.org/0000-0002-2021-5207>

Kryvyi Rih National University

Muzyka I.O.<https://orcid.org/0000-0002-9202-2973>

Kryvyi Rih National University

Ivashchenko O.R.<https://orcid.org/0009-0001-9533-0787>

Kryvyi Rih National University

A MACHINE VISION SYSTEM FOR QUALITY-BASED GRAIN CLASSIFICATION USING MACHINE LEARNING

This Grain quality control directly affects the efficiency of grain processing, the stability of product parameters, and the economic outcomes of storage and cleaning operations. In industrial practice, visual inspection and manual sampling remain widespread, yet these approaches are time-consuming and sensitive to human subjectivity, especially when large grain volumes must be assessed continuously. This study addresses the need for an automated, objective, and scalable solution for grain quality classification that can operate in real time under conveyor-based conditions.

The purpose of the research is to develop and experimentally validate a machine vision system that classifies grain by quality and supports operational decisions for fractional separation and cleaning. The work focuses on achieving reliable defect recognition while maintaining processing speed suitable for production environments. The research tasks include designing an image acquisition setup, forming a labeled dataset of grain images, implementing supervised classification models, and evaluating both classification accuracy and computational performance.

An experimental test stand was built to simulate conveyor transportation of grain, enabling controlled variation of belt speed and illumination conditions. A dataset of wheat, corn, and barley was formed using laboratory image capture, followed by manual labeling into three quality categories based on visible defects and damage severity. Image preprocessing and augmentation were applied to increase variability in orientation and photometric conditions and to improve model robustness. Two machine learning approaches were implemented for comparative evaluation: a deep learning image classifier and a kernel-based classifier using handcrafted visual descriptors. Performance was assessed using overall classification accuracy, confusion-matrix-based error analysis, and average processing time per image across different conveyor speeds.

The experimental evaluation demonstrates that the deep learning approach achieves higher classification accuracy on test data, while the kernel-based classifier provides faster inference with a moderate reduction in accuracy. Error analysis shows that the most frequent misclassifications occur between adjacent quality categories, indicating the importance of borderline-class data coverage and improved labeling consistency. The obtained processing time per image indicates that the proposed system can support real-time operation for moderate grain flow rates, with acceptable performance degradation at higher conveyor speeds due to motion-related image distortions.

The scientific novelty of the work lies in the integrated experimental assessment of quality-based grain classification under controlled conveyor conditions with simultaneous consideration of accuracy and throughput. The practical significance is the feasibility of deploying the system as a component of automated grain cleaning and separation lines, enabling consistent quality monitoring and decision support for industrial processing.

Keywords: Machine vision, grain classification, machine learning, support vector machine, image processing, grain sorting.

Formulation of the problem. In the grain-processing industry, grain quality is one of the key factors influencing profitability and the competitiveness of products [1]. To ensure high quality, continuous monitoring of grain characteristics such as shape, size, color, and the presence of damage or foreign impurities is necessary [2]. Traditional methods of quality assessment (for example, manual sampling and visual analysis) are relatively subjective and labor-intensive, making them less suitable for large volumes of grain in industrial-scale operations.

Modern machine vision technologies make it possible to automate the process of detecting defects in grains and to classify individual grains according to various characteristics [3]. At the same time, the use of machine learning algorithms (in particular deep learning) offers the potential to increase classification accuracy and to adapt the system to new conditions or types of grain crops [4]. In this study, we propose integrating a machine vision system with machine learning algorithms to automate the fractional separation and cleaning of grain, which will help improve the final product's physico-mechanical properties.

Analysis of recent research and publications. Automated grain quality assessment has been a focal point of agricultural research for over a decade, evolving from simple statistical models to complex autonomous systems. Comprehensive surveys of the field highlight that computer vision-based classification has become the gold standard for ensuring food security and processing efficiency [1]. Recent advancements have specifically emphasized the integration of hyperspectral imaging with deep learning, allowing for multiscale sensing that goes beyond human visual capabilities [2].

The evolution of object detection algorithms has significantly impacted the speed and accuracy of grain analysis. Improved versions of real-time detection frameworks, such as YOLOv5, have demonstrated high efficiency in identifying wheat grain quality [3], while similar deep learning architectures have been successfully adapted for the classification of various soybean seeds [4]. For more specialized tasks, such as detecting damage in milled rice, researchers have utilized high-magnification datasets combined with deep convolutional neural networks (CNNs) to achieve superior granularity in classification [5]. These methods are not limited to rice; they have also proven effective for the fast classification of barley grains, showing the versatility of CNN-based approaches across different crop types [6].

Beyond simple classification, predicting specific quality parameters like the breakage rate of kernels

remains a critical industrial challenge. Recent studies have combined machine vision with diverse machine learning algorithms to predict maize kernel damage with high precision [7]. To handle complex scenes where grains may overlap or vary significantly in orientation, region proposal-based CNNs have been introduced to improve localization and classification accuracy [8]. This progress is further supported by high-throughput phenotyping methods that leverage transfer learning to adapt models to specific seed characteristics with minimal retraining [9].

Digital imaging has also opened new avenues for comprehensive quality assessment, particularly in rice production, where deep learning helps evaluate various quality indices simultaneously [10]. This includes the detection of specific biological threats, such as mildew, which is now possible using optimized convolutional networks [11]. For specialized crops, such as colored rice, custom inspection systems have been developed to account for unique spectral and morphological features [12]. Furthermore, cognitive spectroscopy has emerged as a powerful tool for variety classification, offering a non-destructive alternative to chemical analysis [13].

A significant bottleneck in industrial application is the analysis of bulk grain. Researchers have addressed this by implementing deep learning segmentation techniques that can process images of grain masses to predict overall market quality [14]. The reliability of these models depends heavily on the availability of high-quality data; hence, the creation of annotated kernel image databases has become a fundamental contribution to the research community [15]. These resources allow for a more nuanced understanding of the frontiers of grain analysis, particularly in the context of grading and quality standardization [16].

Innovative training strategies have also been explored to overcome the scarcity of labeled industrial data. For instance, the use of synthetic datasets to train instance segmentation networks has shown promise in crop seed phenotyping [17]. In practical harvesting scenarios, real-time sensing of impurities using decision-tree algorithms and image processing has been integrated into combine harvesters to provide immediate feedback [18]. For processing lines, online detection technologies for broken kernels have been developed to maintain high throughput without sacrificing accuracy [19].

Recent trends focus on making these systems more efficient and deployable. Lightweight models, such as improved versions of YOLOv8, offer a balance between detection speed and resource consumption, which is vital for edge computing in silos

[19]. Advanced hyperspectral analysis continues to push the boundaries of what can be detected, offering deeper insights into the internal properties of wheat crops [17-19]. These technologies are now being integrated into specific machine vision systems designed for grain separation, ensuring that quality control happens dynamically during the cleaning process.

In this context, the present work contributes by experimentally evaluating a machine vision system that combines CNN and SVM approaches under controlled conveyor-based conditions, with particular emphasis on accuracy, processing speed, and industrial applicability.

Task statement. The purpose of this paper is to develop and experimentally validate a machine vision system enhanced by advanced machine learning methods for real-time grain quality classification, with a focus on high-accuracy defect detection, efficient processing speed, and scalability for industrial use.

Our objectives include:

- 1) To develop a prototype of a machine vision system capable of accurately detecting defective grains and grains of different quality categories.
- 2) To compare the effectiveness of various machine learning algorithms (primarily convolutional neural networks (CNN), and support vector machine (SVM)).
- 3) To assess the performance speed of the prototype system and its potential scalability to real industrial environments.

Research methodology. The research methodology employs an experimental approach that integrates machine vision techniques with supervised machine learning algorithms for the classification of grain quality. The objective is to evaluate the feasibility of automated grain quality assessment under conditions that closely simulate real industrial conveyor-based processing.

The proposed methodology consists of four main stages:

- 1) construction of an experimental image acquisition setup;
- 2) formation and preprocessing of a labeled image dataset;
- 3) implementation and training of machine learning models (CNN and SVM);
- 4) evaluation of classification accuracy and processing speed.

To simulate industrial grain transportation conditions, a laboratory test stand with a conveyor belt was developed. The stand enables the controlled acquisition of images of grains in motion. The hardware configuration includes:

Illumination System: LED light sources providing uniform illumination with a color temperature of 5000 K to minimize shadows and highlights.

Image Acquisition: An industrial CMOS camera with a resolution of 1920×1080 pixels, capturing frames at 60 FPS.

Computing Platform: A workstation equipped with an NVIDIA RTX 2060 GPU (6 GB VRAM), an Intel Core i7-10700 processor, and 16 GB of RAM. Software implementation was performed using Python with TensorFlow and OpenCV libraries.

The conveyor belt speed was adjustable from 0.1 to 0.5 m/s, allowing the assessment of motion blur effects on classification accuracy.

A dedicated image dataset was created for three grain crops: wheat, corn, and barley. A total of 5,000 original images were captured. Grains were manually categorized by experts into three quality classes:

- 1) high quality: Intact grains without visible defects;
- 2) Medium quality: Grains with minor mechanical damage or slight discoloration;
- 3) Low quality: Grains with significant defects, including cracks, mold, or foreign impurities.

To enhance the model's ability to generalize, data augmentation was applied (rotation $\pm 15^\circ$, horizontal/vertical flips, and brightness adjustment). This expanded the dataset to 15,000 images, partitioned into training (70%), validation (15%), and test (15%) sets. All images were resized to 128×128 pixels to standardize the input for the neural network.

A dedicated image dataset was created for training and evaluating machine learning models. The dataset includes images of three grain crops: wheat, corn, and barley, collected under controlled laboratory conditions.

Approximately 5,000 original images (Fig. 1) were acquired, with roughly equal representation of each crop. During image acquisition, grains were manually categorized into three quality classes:

- 1) high quality: intact grains without visible defects;
- 2) medium quality: grains with minor damage or slight discoloration;
- 3) low quality: grains with significant defects, including cracks, mold, or foreign inclusions.



Fig. 1. Example grain image

Each image was manually labeled according to its quality class. In cases where multiple grains appeared in a single frame, image segmentation methods (thresholding and contour detection) were applied to extract individual grain images.

To increase dataset diversity and improve model generalization, data augmentation techniques were applied, including rotation within the range of -15° to $+15^\circ$, horizontal and vertical flipping, and random adjustments of brightness and contrast. After augmentation, the final dataset contained approximately 15,000 grain images, which were divided into training (70%), validation (15%), and test (15%) subsets.

To classify grain quality, two machine learning approaches were implemented and compared: a convolutional neural network (CNN) and a support vector machine (SVM).

Convolutional Neural Network. The CNN architecture was designed to strike a balance between classification accuracy and computational efficiency. The input images were resized to 128×128 pixels and normalized before training. The network consists of three convolutional layers with 3×3 kernels and 32, 64, and 128 feature maps, respectively. Each convolutional layer is followed by a 2×2 max-pooling layer. The extracted features are processed by two fully connected layers with 256 and 64 neurons, followed by a final output layer with three neurons corresponding to the three quality classes.

ReLU activation functions were used in hidden layers, while Softmax activation was applied in the output layer. Model training was performed using the Adam optimizer with a learning rate of 0.001 and categorical cross-entropy loss. The network was trained for 50 epochs with a batch size of 32.

Support Vector Machine. For comparison, an SVM classifier with a radial basis function (RBF) kernel was implemented. Feature vectors were extracted from grain images using classical computer vision descriptors, including Histogram of Oriented Gradients (HOG), color histograms, and Local Binary Patterns (LBP).

The SVM hyperparameters were selected empirically: the regularization parameter was set to $C = 1$, and the kernel parameter was set to $\gamma = 0.01$.

Other classification methods, such as Random Forest and k-nearest neighbors (k-NN), were tested preliminarily; however, the primary analysis focuses on CNN and SVM due to their widespread use in image-based classification tasks.

Evaluation Metrics

Model performance was evaluated using quantitative metrics relevant to both classification accuracy and real-time processing requirements (1).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (1)$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

In addition, confusion matrices were used to analyze misclassification patterns between adjacent quality classes. Processing speed was measured as the average time required to process a single image (or individual grain) at different conveyor speeds, providing an estimate of system throughput under real-time conditions.

Research results. To assess the effectiveness of the proposed machine vision system, the convolutional neural network was trained and evaluated on the constructed grain image dataset. The objective was to classify grains into three quality categories (high, medium, and low) based on visual characteristics.

The dataset consisted of approximately 15,000 grain images after augmentation and was divided into training (70%), validation (15%), and test (15%) subsets. During training, model parameters were updated using backpropagation on the training set, while validation data were used to monitor generalization performance and detect overfitting.

The CNN was trained for 50 epochs using the Adam optimizer with a learning rate of 0.001 and a batch size of 32. Classification accuracy and loss were monitored simultaneously on the training and validation sets.

Fig. 2 illustrates the evolution of training and validation accuracy over the training epochs. The results show a rapid increase in accuracy during the first 15–20 epochs, followed by gradual convergence. Training accuracy increased from approximately 70% to 95–96%, while validation accuracy stabilized at approximately 92–93%. After 30–40 epochs, both curves exhibited saturation behavior, indicating that the model had converged.

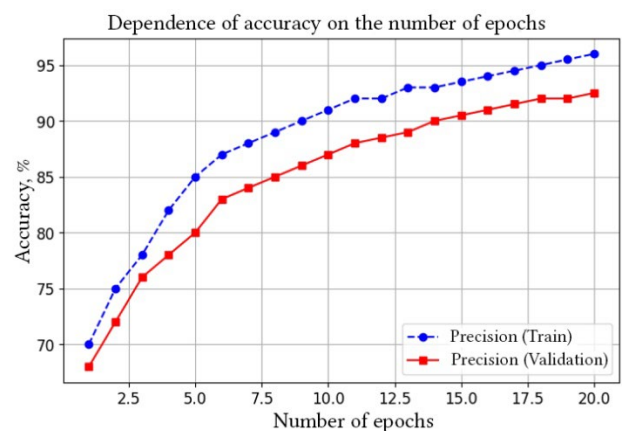


Fig. 2. Example chart of training accuracy (blue curve) and validation accuracy (red curve) vs. the number of epochs, illustrative data

Comparative Analysis of Classification Accuracy and Processing Speed. To evaluate the trade-off between classification accuracy and computational efficiency, the performance of the CNN was compared with that of a support vector machine using an RBF kernel. The evaluation was conducted on the independent test set comprising approximately 2,250 grain images.

The comparative results are summarized in Table 1.

Table 1

Comparison of accuracy and average processing time per image

Algorithm	Configuration	Accuracy (%)	Average processing time (ms)
CNN	3 conv layers, LR=0.001	93.8	5.8
SVM (RBF)	$C = 1, \gamma = 0.01$	91.2	2.2

The CNN achieved higher classification accuracy (93.8%) compared to the SVM (91.2%), confirming the effectiveness of deep learning for visual grain quality assessment. However, this improvement in accuracy comes at the cost of increased processing time. On average, the CNN required approximately 5.8 ms per image, while the SVM processed images in approximately 2.2 ms.

Despite the higher computational cost, the CNN processing time remains acceptable for conveyor systems with moderate grain flow rates, where real-time operation typically allows several milliseconds per grain.

Error Analysis Using a Confusion Matrix. To gain deeper insight into classification performance beyond overall accuracy, a confusion matrix was analyzed for the CNN model. This approach enables the identification of systematic misclassifications between quality

categories. When classifying grain into three quality categories (labelled “High,” “Medium,” and “Low”), it is essential not only to measure overall accuracy but also to understand which classes the model confuses most frequently. For this purpose, a confusion matrix is typically used.

Purpose of the Confusion Matrix. A confusion matrix is an $N \times N$ matrix, where N is the number of classes. Here $N=3$. The element $CM[i,j]$ indicates the number of samples that actually belong to class i but were predicted by the model to belong to class j . With this, one can:

- easily spot the most frequent sources of errors (for example, if “High” quality grains are often misclassified as “Medium”);
- measure classification accuracy for each class individually (rather than a single overall percentage);
- plan how to improve the model - by collecting additional data or adjusting hyperparameters - if certain classes “underperform.”

Below is an example code snippet that illustrates building a confusion matrix (Fig. 3)

We obtained the following confusion matrix (Table 2):

Table 2

Confusion Matrix for three grain quality classes

	Predicted High	Predicted Medium	Predicted Low
Actual High	700	25	8
Actual Medium	40	610	35
Actual Low	15	30	787

The results indicate that grains of high quality are most often confused with medium-quality grains, while low-quality grains are classified with higher reliability. Specifically, 700 high-quality samples

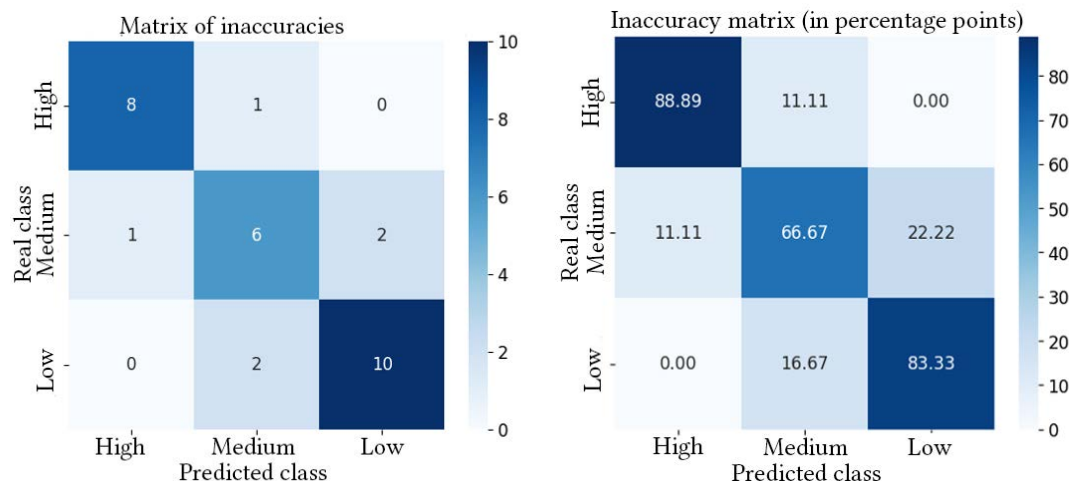


Fig. 3. Confusion Matrix for three grain quality classes

were correctly identified, with a limited number misclassified as medium (25 cases) or low quality (8 cases). Medium-quality grains exhibited the highest confusion rate, particularly with high-quality grains (40 cases) and low-quality grains (35 cases). Low-quality grains showed the best classification performance, with 787 correct predictions.

The observed confusion between high and medium quality categories reflects the presence of subtle defects that are visually indistinguishable. In contrast, low-quality grains typically exhibit pronounced defects such as cracks or mold, which are more easily detected by the model.

System Throughput at Different Conveyor Speeds. Additional experiments were conducted to evaluate the robustness of the classification system under varying conveyor speeds. The goal was to assess how increased grain velocity affects both classification accuracy and processing throughput. The results are presented in Table 3.

As conveyor speed increases, a gradual decrease in classification accuracy is observed for both models. This effect is primarily attributed to reduced exposure time and an increased likelihood of motion blur. Nevertheless, even at the highest tested speed of 0.5 m/s, the CNN maintained accuracy above 92%, demonstrating robustness suitable for many industrial grain processing applications.

The experimental results demonstrate that convolutional neural networks provide superior classification accuracy for grain quality assessment compared to classical machine learning approaches, while still meeting real-time processing requirements under typical conveyor conditions. Although the SVM offers faster processing, its lower accuracy may limit its applicability in scenarios where precise defect detection is critical.

The confusion matrix analysis reveals that the most challenging classification task is distinguishing between high- and medium-quality grains, indicating that further improvements could be achieved through enhanced data augmentation, refined labeling of borderline cases, or the use of more advanced CNN architectures.

Overall, the results confirm the feasibility of deploying the proposed machine vision system for

automated grain quality classification in conveyor-based industrial environments.

Conclusions. The experimental results confirm that convolutional neural networks provide high classification accuracy for grain quality assessment, but at the cost of increased computational complexity. In contrast, the support vector machine approach demonstrates significantly lower computational requirements and faster processing, although its classification accuracy is approximately 2–3% lower than that of the CNN.

The choice of a classification model should therefore be determined by specific operational constraints. In scenarios where classification accuracy is critical and sufficient computational resources (e.g., a GPU) are available, CNN-based solutions are preferable. Conversely, for small- and medium-scale operations with limited hardware capabilities or for applications requiring extremely high throughput, SVM-based or hybrid approaches may offer a more practical compromise between speed and accuracy.

The quality of the training data plays a decisive role in classification performance. In the present study, dataset labeling relied on a combination of manual sorting and automated segmentation, which may introduce labeling inaccuracies, particularly for borderline quality classes. Increasing the diversity of training data - by incorporating images captured under different lighting conditions, conveyor speeds, and grain varieties - could further improve model robustness.

An important practical aspect is integrating the proposed machine vision system into a fully automated grain processing line. Such integration requires synchronization with conveyor speed sensors, mechanisms for removing nonconforming grain, and feedback loops for adjusting sorting and cleaning parameters. Addressing these system-level challenges is essential for successful industrial deployment.

Based on the conducted experimental study, the following conclusions can be drawn:

The proposed convolutional neural network achieved classification accuracy exceeding 93% on the test dataset, demonstrating the feasibility of automated separation of high-, medium-, and low-quality grains using machine vision techniques.

Table 3

Dependence of accuracy and average processing time on conveyor speed

Conveyor Speed (m/s)	CNN (Accuracy, %)	SVM (Accuracy, %)	Grains per second	Throughput (grains/s)*
0.1	94.1	91.5	100	~100 (CNN), ~250 (SVM)
0.3	93.2	90.1	300	~280 (CNN), ~450 (SVM)*
0.5	92.5	88.7	500	~420 (CNN), ~500 (SVM)*

*Throughput values depend on hardware configuration and GPU parallelization capabilities.

The support vector machine provides significantly faster processing, but with a reduction in classification accuracy of approximately 2-3% compared to the CNN. This highlights a clear trade-off between accuracy and computational efficiency.

Classification performance is strongly influenced by imaging conditions, particularly illumination quality and conveyor speed. Under more demanding operating conditions, advanced data augmentation techniques and high-speed imaging become critical to maintaining reliable performance.

With high classification accuracy and acceptable processing throughput, the proposed system is suitable

for implementation in grain elevators, processing plants, and agricultural enterprises. The approach can be extended to additional grain types and integrated more deeply into automated sorting lines.

Future research will focus on developing a comprehensive control system for fractional grain separation, utilizing machine vision outputs to regulate conveyor speed, grain flow direction, and cleaning mechanisms (e.g., sieves and air channels). Further work will also address system scalability for higher throughput and the automatic detection of specific defect types, such as mold or insect damage.

Bibliography:

1. Velesaca H. O., Suárez A., Mira J., Sappa, A. Computer vision based food grain classification: A comprehensive survey. *Computers and Electronics in Agriculture*. 2021; 190: 106287. DOI: <https://doi.org/10.1016/j.compag.2021.106287>.
2. Shuai L., Chen H., Luo S. A research review on deep learning combined with hyperspectral imaging in multiscale agricultural sensing. *Computers and Electronics in Agriculture*. 2024; 217: 108577. DOI: <https://doi.org/10.1016/j.compag.2023.108577>.
3. Zhao W., Liu S., Li X., Han X., Yang H. Fast and accurate wheat grain quality detection based on improved YOLOv5. *Computers and Electronics in Agriculture*. 2022; 202: 107426 DOI: <https://doi.org/10.1016/j.compag.2022.107426>
4. Huang Z., Wang R., Cao Y. Deep learning based soybean seed classification. *Computers and Electronics in Agriculture*. 2022; 202: 107393. DOI: <https://doi.org/10.1016/j.compag.2022.107393>.
5. Moses K., Miglani A., Kankar, P. K. Deep CNN-based damage classification of milled rice grains using a high-magnification image dataset. *Computers and Electronics in Agriculture*. 2022; 195: 106811. DOI: <https://doi.org/10.1016/j.compag.2022.106811>.
6. Shah S. A. Automatic and fast classification of barley grains from images: A deep learning approach. *Smart Agricultural Technology*. 2022; 2: 100036. DOI: <https://doi.org/10.1016/j.atech.2022.100036>
7. Fan C., Wang W., Cui T., Liu Y., Qiao M. Maize Kernel Broken Rate Prediction Using Machine Vision and Machine Learning Algorithms. *Foods*. 2024; 13(24): 4044 DOI: <https://doi.org/10.3390/foods13244044>.
8. Aukkapinyo K., Sawangwong S., Pooyoi P., Kusakunniran W. Localization and Classification of Rice-grain Images Using Region Proposals-based Convolutional Neural Network. *International Journal of Automation and Computing*. 2020; 17(2). Pp: 233–246. DOI: <https://doi.org/10.1007/s11633-019-1207-6>
9. Yang S., Zheng L., He P. et al. High-throughput soybean seeds phenotyping with convolutional neural networks and transfer learning. *Plant Methods*. 2021; 17: 50. DOI: <https://doi.org/10.1186/s13007-021-00749-y>.
10. Kurade M. B. A Digital Imaging and Deep Learning-Based Approach for Rice Quality Assessment. *Foods*. 2023; 12(6): 1273. DOI: <https://doi.org/10.3390/foods12061273>.
11. Chen S., Xiong J., Guo W., Bu R., Yang Z., Zheng Z., Lin R. Colored rice quality inspection system using machine vision. *Journal of Cereal Science*. 2019; 88. Pp:87–95. DOI: <https://doi.org/10.1016/j.jcs.2019.05.010>.
12. Onmankhong K., Ma T., Inagaki T., Sirisomboon P., Tsuchikawa S. Cognitive spectroscopy for classification of rice varieties. *Infrared Physics & Technology*. 2022; 123: 104100. DOI: <https://doi.org/10.1016/j.infrared.2022.104100>.
13. Ashtiani S. H. M., Walker A., Panozzo J. Deep learning segmentation in bulk grain images for prediction of grain market quality. *Food and Bioprocess Technology*. 2022; 15: 114–128. DOI: <https://doi.org/10.1007/s11947-022-02840-1>.
14. Lin Y., Fan J., et al. An annotated grain kernel image database for classification and detection. *Scientific Data*. 2023; 10: 642. DOI: <https://doi.org/10.1038/s41597-023-02660-8>.
15. Ahmed S. B., Ali S. F., Khan A. Z. On the frontiers of rice grain analysis, classification and quality grading: A review. *IEEE Access*. 2021; 9. Pp:160779–160796. DOI: <https://doi.org/10.1109/ACCESS.2021.3130472>.
16. Toda Y., Okura F., Ito J. et al. Training instance segmentation neural network with synthetic datasets for crop seed phenotyping. *Commun Biol* 3, 173 (2020). <https://doi.org/10.1038/s42003-020-0905-5>.
17. Chen J., Lian Y., Li Y. Real-time grain impurity sensing for rice combine harvesters using image processing and decision-tree algorithm. *Computers and Electronics in Agriculture*. 2020; 175: 105591. DOI: <https://doi.org/10.1016/j.compag.2020.105591>.

18. Zidi F., Ouafi A., Bougourzi F., Distant C., Taleb-Ahmed, A. "Advancing wheat crop analysis: A survey of deep learning approaches using hyperspectral imaging". arXiv preprint. 2025; arXiv:2505.00805. DOI: <https://doi.org/10.48550/arXiv.2505.00805>.

19. Stepanenko S., Kuzmych A., Kharchenko S., Borys A., Dnes V., Volyk D., Kalinichenko R. A machine vision approach for grain quality control during separation. *Journal of Engineering Sciences*, Vol. 12(1), pp. E9–E17. [https://doi.org/10.21272/jes.2025.12\(1\).e2](https://doi.org/10.21272/jes.2025.12(1).e2)

Кузнєцов Д.І., Музика І.О., Івашенко О.Р. СИСТЕМА МАШИННОГО ЗОРУ ДЛЯ КЛАСИФІКАЦІЇ ЗЕРНА ЗА ЯКІСТЮ З ВИКОРИСТАННЯМ МАШИННОГО НАВЧАННЯ

Контроль якості зерна безпосередньо впливає на ефективність його переробки, стабільність параметрів продукції та економічні показники операцій зберігання й очищення. У промисловій практиці досі широко застосовуються візуальний контроль і ручний відбір проб, однак ці підходи є трудомісткими та чутливими до суб'єктивного людського фактору, особливо за необхідності безперервної оцінки великих обсягів зерна. У даному дослідженні розглядається потреба в автоматизованому, об'єктивному та масштабованому рішенні для класифікації якості зерна, здатному працювати в режимі реального часу в умовах конвеєрного транспортування. Метою роботи є розробка та експериментальна валідація системи машинного зору для класифікації зерна за якістю та підтримки оперативних рішень щодо фракційного розділення й очищення. Основна увага приділяється досягненню надійного розпізнавання дефектів за збереження швидкодії, придатної для виробничих умов. До завдань дослідження належать проектування системи отримання зображень, формування розміченого набору зображень зерна, реалізація моделей керованого навчання та оцінювання точності класифікації й обчислювальної ефективності. Для імітації конвеєрного транспортування зерна було створено експериментальний стенд, що забезпечує контрольовану зміну швидкості стрічки та умов освітлення. Набір даних пшениці, кукурудзи та ячменю сформовано на основі лабораторного отримання зображень із подальшою ручною розміткою за трьома категоріями якості відповідно до видимих дефектів і ступеня пошкодження. Для підвищення стійкості моделей застосовано попередню обробку зображень та аугментацію з метою збільшення варіативності орієнтації та фотометричних умов. Для порівняльного аналізу реалізовано два підходи машинного навчання: класифікатор зображень на основі глибокого навчання та ядерний класифікатор із використанням ручних візуальних дескрипторів. Оцінювання проводилось за показниками загальної точності класифікації, аналізу помилок на основі матриці невідповідностей та середнього часу обробки одного зображення за різних швидкостей конвеєра. Результати експериментальних досліджень показують, що підхід на основі глибокого навчання забезпечує вищу точність класифікації на тестових даних, тоді як ядерний класифікатор демонструє швидшу інференцію за помірного зниження точності. Аналіз помилок свідчить, що найбільш часті хибні класифікації виникають між суміжними категоріями якості, що вказує на важливість кращого охоплення прикордонних класів і підвищення узгодженості розмітки. Отримані значення часу обробки одного зображення підтверджують можливість роботи системи в режимі реального часу за помірних потоків зерна, при допустимому зниженні продуктивності на вищих швидкостях конвеєра через спотворення зображень, спричинені рухом. Наукова новизна роботи полягає в комплексній експериментальній оцінці класифікації зерна за якістю в контрольованих конвеєрних умовах із одночасним урахуванням точності та пропускну здатності. Практична значущість полягає в можливості впровадження запропонованої системи як складової автоматизованих ліній очищення та сепарації зерна, що забезпечує стабільний контроль якості та підтримку прийняття рішень у промисловій переробці.

Ключові слова: машинний зір, класифікація зерна, машинне навчання, метод опорних векторів, обробка зображень, сортування зерна.

Дата першого надходження статті до видання: 15.03.2026

Дата прийняття статті до друку після рецензування: 13.04.2026

Дата публікації (оприлюднення) статті 11.05.2026

Купін А.І.<https://orcid.org/0000-0001-7569-1721>

Криворізький національний університет

Духов Д.Ю.<https://orcid.org/0009-0005-8267-685X>

Криворізький національний університет

ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ АВТОМАТИЗОВАНОГО ГЕНЕРУВАННЯ ПРОГРАМНОГО КОДУ ЗАСОБАМИ ШТУЧНОГО ІНТЕЛЕКТУ

У статті здійснено комплексний аналіз сучасних трендів автоматизованого генерування програмного коду із застосуванням засобів штучного інтелекту, зокрема великих мовних моделей (LLM), що базуються на методах глибокого навчання та обробки природної мови. Розглянуто еволюцію підходів до генерації коду – від традиційних інструментів автоматизації та low-code/no-code платформ до генеративних ШІ-систем, здатних створювати програмний код «з нуля» на основі текстових запитів розробника. Показано, що такі системи забезпечують суттєве підвищення продуктивності та ефективності роботи розробників, зменшення кількості помилок і скорочення циклів розробки програмного забезпечення, що підтверджується результатами сучасних емпіричних досліджень.

Значну увагу приділено аналізу спеціалізованих бенчмарків, які використовуються для оцінювання якості та ефективності ШІ-моделей у завданнях програмування. Детально охарактеризовано такі бенчмарки, як HumanEval, MBPP, SWE-Bench, Aider Polyglot, LiveCodeBench, APPS, MTPB, DS-1000 та WebDev Arena, визначено їхні метрики, переваги та обмеження. Показано, що еволюція бенчмарків від перевірки функціональної коректності ізольованих фрагментів коду до оцінювання агентних можливостей, багатомовного редагування, виправлення реальних помилок і багатоетапної взаємодії відображає зростання вимог до інтелектуальних можливостей ШІ у реальних сценаріях розробки програмного забезпечення.

У роботі проаналізовано сучасний стан комерційних і відкритих мовних моделей, зокрема GPT-4.x, Claude 3.x, Gemini 2.x, Code Llama та DeepSeek-Coder, та показано тенденцію стирання межі між універсальними й спеціалізованими моделями для кодування. Окремо підкреслено роль розширених контекстних вікон, агентних фреймворків і мультимодальних можливостей у формуванні нового покоління ШІ-асистентів, здатних виконувати складні завдання з аналізу, налагодження, рефакторингу та планування програмних систем.

Зроблено висновок, що автоматизоване генерування коду за допомогою штучного інтелекту переходить від інструментальної підтримки до повноцінного партнерства у процесі розробки програмного забезпечення, формуючи підґрунтя для появи більш автономних і інтегрованих ШІ-рішень у майбутньому.

Ключові слова: генеративний штучний інтелект, аналіз трендів, генерація програмного коду, дослідження бейнчмарків, прогнозування розвитку технології.

Постановка проблеми. Генерація коду за допомогою генеративного штучного інтелекту – це автоматизований процес створення програмного коду на основі алгоритмів ШІ, які навчаються на великих масивах наявного вихідного коду. Такі алгоритми, що становлять основу великих мовних моделей (LLM) і технологій обробки природної мови (NLP), застосовують методи глибокого навчання та масштабні нейронні мережі

для аналізу різноманітних наборів програмного коду з відкритих репозиторіїв. У межах цього підходу генеративний ШІ забезпечує автоматизацію написання програм, інтерпретуючи запити, сформульовані природною мовою, та використовуючи здобуті під час навчання знання для розпізнавання програмних шаблонів, стилістики й логіки.

На відміну від класичних засобів автоматизації коду, що базуються переважно на статичному ана-

лізі та заздалегідь визначених шаблонах, генеративний ШІ характеризується більшою гнучкістю та здатністю адаптуватися до різних контекстів. Він також відрізняється від low-code і no-code платформ, оскільки не спирається на готові шаблони чи бібліотеки компонентів, а формує програмний код безпосередньо з нуля відповідно до мовних інструкцій розробника. Якщо low-code та no-code рішення здебільшого орієнтовані на користувачів без глибоких технічних знань і бізнес-аудиторію, то інструменти генерації коду на основі ШІ є більш універсальними та придатними як для професійних програмістів, так і для ширшого кола користувачів [1].

Використання генеративного ШІ для генерації коду надає розробникам та організаціям ряд значних переваг. Однією з ключових переваг є підвищення продуктивності та ефективності розробників. Автоматично генеруючи код на основі заданих специфікацій, системи ШІ можуть значно прискорити процес розробки, зменшити кількість помилок та скоротити цикли випуску програмного забезпечення. Згідно з дослідженнями, 88% розробників повідомили про підвищення продуктивності під час використання таких інструментів, як GitHub Copilot. Генеративний ШІ також може значно підвищити загальну ефективність роботи та знизити витрати. Звільняючи розробників від рутинних завдань, генеративний ШІ дозволяє їм зосередитися на більш складних та творчих аспектах розробки.

Генеративний ШІ ефективно автоматизує рутинні та повторювані завдання кодування. Такі інструменти, як GitHub Copilot, здатні генерувати шаблонний код та фрагменти коду для стандартних задач, звільняючи розробників від необхідності писати їх вручну. Це не лише економить час, але й зменшує ризик помилок, пов'язаних з монотонною роботою. Автоматизація цих завдань також може призвести до підвищення задоволеності розробників, оскільки вони можуть присвячувати більше часу вирішенню складних та цікавих проблем [2].

Аналіз останніх досліджень і публікацій. Виконано дослідження бенчмарків штучного що було спрямовано на розроблення, аналіз і вдосконалення стандартизованих підходів до оцінювання якості, ефективності та надійності інтелектуальних систем. Бенчмарки ШІ виконують функцію об'єктивного інструмента порівняльного аналізу різних моделей, алгоритмів і архітектур за однакових експериментальних умов.

HumanEval: функціональна коректність та метрика pass@k. Розроблений компанією OpenAI,

HumanEval є одним із найбільш відомих бенчмарків для оцінювання здатності моделей до генерації програмного коду. Він охоплює 164 вручну підготовлені задачі, кожна з яких містить сигнатуру функції, документаційний рядок із описом очікуваної поведінки та набір модульних тестів для перевірки правильності реалізації. На основі сигнатури та опису модель генерує тіло функції, після чого отриманий код перевіряється за допомогою наперед визначених тестів з метою оцінки функціональної коректності.

Метрика pass@k визначає ймовірність того, що хоча б одне з k найкращих рішень, згенерованих моделлю, успішно пройде всі тестові випадки. Найчастіше використовують значення pass@1, pass@10 і pass@100, де pass@1 є найбільш вимогливою метрикою, оскільки оцінює правильність першої спроби. Такий підхід безпосередньо вимірює функціональну коректність і ґрунтується на модульному тестуванні, яке є стандартною практикою у розробці програмного забезпечення. HumanEval забезпечує уніфіковану основу для порівняння різних моделей, однак його обмеженням є те, що задачі представлені у вигляді окремих функцій, а не повноцінних програм, що знижує практичну застосовність у реальних умовах [3].

MBPP (Mostly Basic Python Problems): основи програмування. Цей бенчмарк містить близько 1000 задач з програмування мовою Python, зібраних за допомогою краудсорсингу та орієнтованих на початківців. Завдання охоплюють базові концепції програмування та функціональні можливості стандартної бібліотеки Python. Кожна задача включає текстовий опис, приклад реалізації та три автоматизовані тестові випадки для перевірки правильності розв'язання. Оцінювання моделей здійснюється на основі їх здатності генерувати коректний Python-код.

MBPP зосереджується на перевірці фундаментальних навичок роботи з Python, що робить його ефективним індикатором базового рівня володіння програмуванням. Водночас, подібно до HumanEval, цей бенчмарк не повною мірою відображає складні та багатокомпонентні сценарії розробки програмного забезпечення, характерні для реальних умов [4].

SWE-Bench: виправлення реальних помилок та агентне кодування. На відміну від HumanEval, бенчмарк SWE-Bench, представлений OpenAI, зосереджується на корекції реальних програмних помилок, отриманих з відкритих репозиторіїв, що робить його надзвичайно реалістичним інструментом для оцінювання здатності великих мовних

моделей до кодування. Моделям надається контекст, аналогічний тому, з яким стикається розробник, зокрема опис проблеми з GitHub issues, фрагменти коду з дефектами та очікувані зміни. Завдання полягає у створенні патча, який усуває помилку, не породжуючи нових.

Оцінювання здійснюється за такими критеріями, як коректність запропонованого патча, його безшовна інтеграція в наявну кодову базу та ефективність внесених змін з погляду оптимальності й відповідності кращим практикам програмування. Результати зазвичай подаються у вигляді відсотка успішно вирішених задач. Завдяки використанню реальних кейсів SWE-Bench має високу практичну цінність, перевіряє здатність моделей до аналізу та модифікації вже існуючого коду, що є критично важливою навичкою, а також стимулює розвиток можливостей налагодження. Водночас цей бенчмарк вимагає глибокого контекстного розуміння цілісних програмних проєктів, з чим окремі LLM досі мають труднощі, а також характеризується підвищеною складністю автоматизації процесу оцінювання [5].

Aider Polyglot: Багатомовне редагування коду та автономне вирішення проблем. Цей бенчмарк оцінює, наскільки добре ШІ може впоратися з 225 найскладнішими завданнями з кодування з Exercism кількома мовами (C++, Go, Java, JavaScript, Python та Rust). Зосереджується на здатності ШІ писати новий код, який безшовно інтегрується в існуючі кодові бази, та застосовувати зміни до вихідних файлів без втручання людини. Він перевіряє автономне вирішення проблем за допомогою ітеративного підходу: написання коду, перевірка, виправлення помилок та повторення до повного вирішення. Багатомовна підтримка, реалістичність сценаріїв (редагування/інтеграція коду) та оцінка автономного вирішення проблем. Вважається кращим показником здатності ШІ-агентів до вирішення проблем, ніж SWE-Bench, завдяки його багатомовності та фокусу на різноманітних проєктах. Може бути обчислювально інтенсивним через ітеративний характер[6].

LiveCodeBench: комплексна оцінка без забруднення даних. LiveCodeBench є всебічним бенчмарком для оцінювання моделей без ризику витоку навчальних даних, який постійно поповнюється новими задачами із платформ змагального програмування, зокрема LeetCode, AtCoder та CodeForces. Завдання відбираються з актуальних змагань, що гарантує їхню сучасність та унеможливорює забруднення даних, оскільки моделі

перевіряються на задачах, опублікованих після дати завершення їх навчання.

У межах цього бенчмарку LLM оцінюються за чотирма сценаріями: генерація програмного коду на основі природномовного опису, самовідновлення через виправлення некоректного коду, виконання коду на заданих вхідних даних та прогнозування результатів тестування для конкретних входів. Основною метрикою зазвичай є $\text{pass}@1$. LiveCodeBench усуває проблему забруднення даних, забезпечує комплексне оцінювання різних аспектів навичок програмування та охоплює збалансований спектр рівнів складності. Водночас, як відносно новий бенчмарк, він має менш розвинену історичну базу результатів для всіх моделей порівняно з такими інструментами, як HumanEval або MBPP [7].

APPS (Automated Programming Progress Standard): Змагальне програмування. Містить 10 000 задач, що варіюються від початкового до університетського рівня змагань, вимірюючи здатність до кодування та вирішення проблем. Задачі формуються вільною природною мовою, а рішення оцінюються на коректність за допомогою великого банку тестових випадків. Імітує оцінку програмістів-людей у змагальних умовах, охоплює широкий діапазон складності та оцінює навички вирішення проблем, що виходять за межі базової генерації коду. Висока складність може призвести до низьких показників проходження для багатьох моделей.

MTPB (Multi-Turn Programming Benchmark): Багатоетапний синтез програм. Новий бенчмарк, розроблений для вимірювання здатності моделей до багатоетапного синтезу програм, де одна програма розкладається на кілька підказок, що визначають підзадачі. Складається зі 115 різноманітних задач, що вимагають багатоетапних описів природною мовою. Моделі синтезують програму в кілька етапів з взаємодією користувача. Продуктивність вимірюється показником проходження на експертно написаних тестових випадках. Вирішує реальний сценарій ітеративної розробки та вдосконалення програм через розмову, демонструючи переваги багатоетапної взаємодії над одноразовими підказками. Новий бенчмарк, тому широке впровадження та порівняльні дані можуть бути менш обширними.

DS-1000: Задачі кодування, специфічні для науки про дані. Цей бенчмарк оцінює продуктивність у задачах науки про дані, зокрема зосереджуючись на завершенні коду в цій галузі. Специалізація в науці про дані робить його актуальним

для оцінки моделей у цій галузі, що швидко розвивається. Вузька спрямованість порівняно з загальними бенчмарками генерації коду.

WebDev Arena. Це платформа для оцінювання моделей у процесі створення веб-додатків у режимі реального часу, де користувачі обирають кращий результат шляхом голосування. Підхід ґрунтується на порівнянні двох веб-додатків, згенерованих різними моделями для однієї й тієї самої задачі, після чого користувачі голосують за більш якісний варіант, що дозволяє формувати рейтинги моделей відповідно до їхніх результатів.

Такий формат забезпечує оцінювання моделей у наближених до реальності умовах розробки веб-застосунків, беручи до уваги аспекти UI/UX, рівень інтерактивності та функціональну повноту. Водночас результати значною мірою залежать від суб'єктивних уподобань користувачів, а сама платформа спеціалізується переважно на завданнях веб-розробки.

Постановка завдання. У сучасних дослідженнях штучного інтелекту значну увагу приділяють оцінюванню спроможності мовних і мультимодальних моделей до автоматичного генерування програмного коду. Для цього використовуються спеціалізовані бенчмарки, які забезпечують стандартизовані умови тестування та порівняльного аналізу різних моделей.

Еволюція бенчмарків від HumanEval і MBPP до SWE-Bench, Aider Polyglot та LiveCodeBench засвідчує глибоку трансформацію підходів до визначення «інтелекту коду» для великих мовних моделей. Відбувається перехід від простої генерації синтаксично та функціонально коректних ізольованих фрагментів програмного коду до демонстрації комплексних інженерних компетенцій, що охоплюють розуміння складних контекстів проекту, налагодження, самокорекцію, багатоетапну взаємодію та багатомовну експертизу. Це свідчить про те, що очікування індустрії від ШІ-асистентів для програмування швидко еволюціонують у бік повноцінних ШІ-агентів розробки програмного забезпечення.

Ця тенденція чітко простежується у переході від HumanEval і MBPP, орієнтованих на перевірку функціональної коректності окремих функцій, до SWE-Bench, який оцінює здатність виправляти реальні помилки, а також аналізувати та модифікувати наявний код. Aider Polyglot розширює цей підхід, фокусуючись на автономному розв'язанні задач у багатомовних середовищах і безшовній інтеграції змін без участі людини. LiveCodeBench доповнює картину сценаріями самовідновлення

та виконання коду, тоді як МТРВ безпосередньо вимірює здатність до багатоетапного синтезу програм. Така послідовна еволюція бенчмарків демонструє, що сучасні ШІ-системи оцінюються вже не лише за здатністю генерувати код, а й за вмінням функціонувати як інтелектуальні помічники, спроможні розуміти, налагоджувати, змінювати та співпрацювати у межах складних програмних проєктів, що відображає зростаючу зрілість галузі та підвищені вимоги до ШІ у реальних сценаріях розробки.

Отже, на підставі вищенаведеного, авторами цієї роботи ставиться завдання дослідження шляхом порівняльної оцінки ефективності автоматизованого генерування програмного коду засобами штучного інтелекту на основі порівняльної оцінки їх продуктивності.

Виклад основного матеріалу. Хоча спеціалізовані моделі (наприклад, Code Llama, DeepSeek Coder) спочатку були розроблені для досягнення високих результатів у завданнях кодування, новітні універсальні LLM (наприклад, GPT-4.x, Claude 3.x, Gemini 2.x) демонструють все більш конкурентоспроможну, а іноді й вищу продуктивність у бенчмарках кодування. Це свідчить про те, що досягнення в базовій архітектурі LLM та загальних можливостях міркування є переносними та дуже корисними для генерації коду, стираючи межі між спеціалізованим та загальним ШІ в цій галузі.

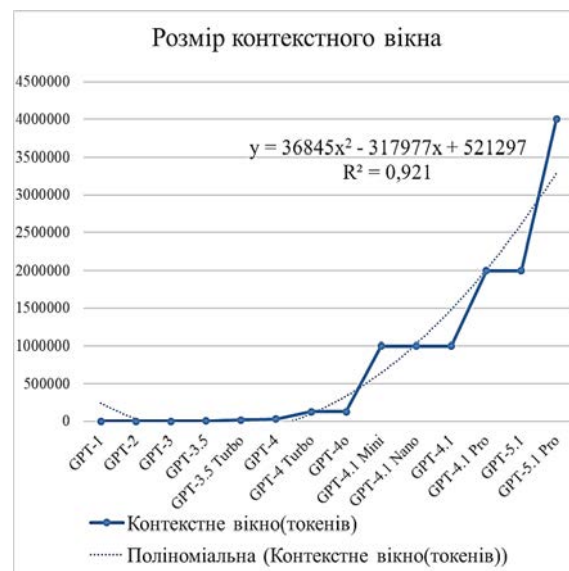


Рис. 1. Розвиток контекстного вікна ChatGPT

Ця тенденція проявляється у кількох ключових вимірах. По-перше, спостерігається цілеспрямоване збільшення розмірів контекстних вікон (рис.

1), що надає моделям змогу опрацювати цілі кодові бази. Значення коефіцієнта детермінації при апроксимації на рівні $R^2=0.92$ свідчить про доволі високу достовірність отриманого тренду.

Так, Gemini 2.5 Pro підтримує контекстне вікно обсягом до 1 мільйона токенів, тоді як Llama 4 Scout – до 10 мільйонів токенів. Йдеться не лише про масштабування обсягу вхідних даних, а про глибинну трансформацію підходів до розуміння та взаємодії ШІ з програмними проєктами на архітектурному рівні, що відкриває нові можливості для масштабного рефакторингу та складних процедур налагодження [9].

По-друге, дедалі більшої ваги набувають агентні можливості та багатокрокова взаємодія. Оцінювання моделей уже не обмежується одноразовою генерацією коду, а охоплює їхню спроможність автономно планувати дії, виконувати завдання, здійснювати налагодження та ітеративно вдосконалювати результат. Це підтверджується зростанням показників у таких бенчмарках, як SWE-Bench, де моделі, що застосовують спеціалізовані фреймворки або режими мислення, демонструють істотно кращі результати. Така динаміка свідчить про перехід ШІ від ролі пасивного генератора коду до активного партнера у процесі розв'язання складних інженерних задач.

Упродовж 2024–2025 років комерційні моделі, зокрема серії OpenAI «o» (o3/o4), Claude 3.7 Sonnet та Gemini 2.5 Pro, стабільно посідають провідні позиції у складних бенчмарках, таких як SWE-Bench Verified і Aider Polyglot. Ці системи зазвичай поєднують високий рівень загальної інтелектуальної спроможності з функціями, спеціально оптимізованими для програмування, зокрема розширеними режимами мислення у Claude та великими контекстними вікнами у Gemini.

Відкриті моделі, такі як Code Llama 70B Instruct, DeepSeek-Coder-V2-Instruct та новіші моделі Llama 4, швидко наздоганяють комерційних лідерів. DeepSeek-Coder-V2-Instruct, наприклад, досягає показників HumanEval та MBPP, порівнянних з GPT-4 Turbo. Це свідчить про те, що відкритий вихідний код стає все більш життєздатним варіантом для складних завдань кодування, пропонуючи гнучкість та економічну ефективність.

Продуктивність на бенчмарках, таких як SWE-Bench Verified, часто значно підвищується за рахунок використання агентних фреймворків або спеціальних каркасів. Наприклад, Claude 3.7 Sonnet досягає 70.3% на SWE-bench Verified з власним каркасом порівняно з 62.3% без нього [8]. Це під-

креслює, що не лише базова модель, а й архітектура, що її оточує, включаючи інструменти та стратегії ітеративного вирішення проблем, є критично важливими для досягнення високої продуктивності в реальних сценаріях. Дані переконливо свідчать про те, що базова продуктивність LLM, хоча й важлива, все частіше посилюється за допомогою агентних фреймворків або спеціальних каркасів. Це вказує на зміну способу вимірювання інтелекту в LLM для кодування, переходячи від простої генерації коректного коду до інтелектуальної взаємодії з середовищем, налагодження та ітерації.

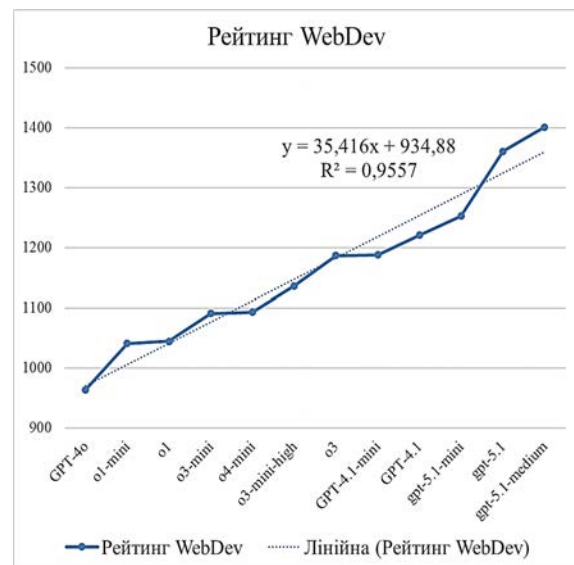


Рис. 2. Розвиток якості коду ChatGPT за рейтингом WebDev

HumanEval та MBPP, які раніше вважалися еталонними бенчмарками, нині значною мірою вже опановані провідними моделями, що суттєво знижує їхню ефективність для диференціації найсучасніших можливостей. Це зумовило появу нових, більш складних інструментів оцінювання, орієнтованих на реалістичні сценарії (рис. 2), зокрема виправлення помилок у реальному коді (SWE-Bench), багатомовне редагування та інтеграцію (Aider Polyglot), а також комплексні задачі програмування (LiveCodeBench). На приклад розвитку якості коду ChatGPT за рейтингом WebDev (WEB-програмування) має прямолінійний характер з кожною новою версією великої мови невої моделі. Значення коефіцієнта детермінації при апроксимації на рівні $R^2=0.95$ що свідчить про доволі високу достовірність отриманого тренду. Така еволюція бенчмарків є прямою відповіддю на стрімкий розвиток ШІ: у міру того як моделі швидко опановують базові завдання, спільнота

з оцінювання змушена розробляти дедалі складніші, багатовимірні та вільні від забруднення набори тестів, щоб і надалі об'єктивно вимірювати та стимулювати реальні можливості ШІ [10].

Сучасні моделі дедалі частіше проєктуються як автономні агенти, здатні планувати, виконувати, тестувати та самокоригувати власні дії. Багатоетапна взаємодія суттєво підвищує якість синтезу програм, забезпечуючи більш природну й ефективну комунікацію з ШІ під час розв'язання складних завдань програмування. Це перетворює LLM з пасивних генераторів коду на активних розв'язувачів проблем. Можливість самостійного написання, редагування та виконання коду з розвиненими механізмами міркування й усунення помилок (як у Claude 3.5 Sonnet) або застосування ітеративного циклу «написання – перевірка – виправлення – повторення» (як в Aider Polyglot) свідчить про фундаментальний зсув. У результаті ШІ переходить від інструмента підказок до автономного виконавця, здатного самостійно реалізовувати складні завдання та виправляти помилки, що істотно трансформуює робочі процеси розробників, передаючи значну частину складних операцій штучному інтелекту.

Комерційні моделі від OpenAI, Anthropic та Google зазвичай демонструють найвищу пікову продуктивність, особливо в завданнях, що потребують складних міркувань та агентної поведінки. Вони часто пропонують унікальні можливості, такі як розширені режими мислення або мультимодальні функції. Водночас моделі з відкритим вихідним кодом, зокрема Meta Llama, DeepSeek та BigCode, швидко скорочують розрив у продуктивності, забезпечуючи конкурентні результати, підтримку великих контекстних вікон і гнучкі ліцензійні умови для дослідницького й комерційного використання. Це сприяє активному розвитку екосистеми завдяки внеску спільноти.

Зростаюча конкурентоспроможність відкритих моделей формує динамічну напругу, яка позитивно впливає на всю сферу ШІ-кодування. Хоча закриті системи зазвичай задають початкові орієнтири, відкриті рішення, такі як Code Llama та DeepSeek-Coder-V2, швидко наздоганяють і навіть перевершують окремі комерційні аналоги за низкою бенчмарків. Така конкуренція стимулює інновації в обох напрямках, сприяючи підвищенню доступності та розширенню вибору для розробників, де поряд із продуктивністю дедалі більшої ваги набувають вартість, конфіденційність і можливості налаштування.

Постійне розширення контекстних вікон (наприклад, до 1 мільйона токенів у Gemini 2.5 Pro та GPT-4.1 і до 10 мільйонів у Llama 4 Scout) є ключовим чинником для практичних задач програмування. Це дозволяє моделям аналізувати цілі кодові бази,

глибше розуміти складні залежності та ефективніше виконувати масштабні рефакторинги й усунення помилок. Така тенденція відображає глибинний зсув у напрямі забезпечення можливості міркування над повними програмними проєктами, а не лише окремими фрагментами. У ширшому контексті це означає перехід від генерації локального коду до формування «інтелекту кодової бази», що радикально змінює підходи до великомасштабної розробки програмного забезпечення, відкриваючи перспективи автоматизованого архітектурного рефакторингу та виявлення системних вразливостей.

Помітною тенденцією також є активна інтеграція мультимодальних можливостей, зокрема зору, аудіо та відео. Це дає змогу моделям інтерпретувати архітектурні діаграми, макети інтерфейсів або відеоописи для подальшої генерації коду. У поєднанні з великими контекстними вікнами це формує бачення майбутнього, у якому ШІ взаємодіє з усім середовищем розробки, а не лише з текстовими файлами. Моделі зможуть аналізувати архітектуру систем за діаграмами, налагоджувати візуальні помилки зі скріншотів чи створювати код на основі відеоматеріалів, що означає перехід від суто текстового формату «вхід-вихід» до комплексної, візуально та контекстно орієнтованої підтримки розробників.

Сукупні дані вказують на фундаментальну зміну як у можливостях, так і в очікуваннях від сучасних ШІ-моделей. Вони еволюціонують від простих генераторів коду до комплексних інтелектуальних асистентів з розробки програмного забезпечення, здатних аналізувати, налагоджувати, рефакторити та планувати. Цю трансформацію яскраво ілюструє розвиток бенчмарків: якщо ранні інструменти, такі як HumanEval і MBPP, були орієнтовані виключно на генерацію коду, то новіші, зокрема SWE-Bench, Aider Polyglot та LiveCodeBench, оцінюють агентні можливості, автономне вирішення проблем, редагування багатомовного коду, виконання програм і прогнозування результатів. Провідні моделі демонструють високі результати у складних багатокрокових і багатофайлових завданнях, що чітко свідчить про перехід ШІ до більш цілісної та інтегрованої ролі в сучасній розробці програмного забезпечення.

Висновки. Розвиток ШІ-генерації коду характеризується швидким прогресом та зростаючою спеціалізацією. Аналіз показує, що бенчмарки еволюціонували від простих тестів функціональної коректності до складних оцінок агентних можливостей, реального виправлення помилок та багатомовного редагування. Це відображає зростаючі очікування від моделей ШІ, які тепер розглядаються як повноцінні помічники з розробки програмного забезпечення, а не просто як інструменти генерації коду.

Майбутнє ІІІ в кодуванні характеризуватися подальшою розділення загального та спеціалізованого ІІІ, з моделями, що пропонують безпрецедентні контекстні вікна та мультимодальні можливості. Це дозволить ІІІ розуміти та взаємодіяти з усім середовищем розробки програмного забезпечення, що може призвести до більш інтегрованих та автономних ІІІ-асистентів, які можуть брати на себе значні частини життєвого циклу розробки. Це означає, що майбутнє ІІІ в кодуванні не полягає в одній

найкращій моделі, а в різноманітній екосистемі спеціалізованих та адаптивних інструментів ІІІ, кожен з яких оптимізований для певних аспектів розробки програмного забезпечення.

Побудовані тренди дозволяють з досить високою достовірністю ($R^2=0.92-0.95$) прогнозувати сучасні тенденції розвитку штучного інтелекту, зокрема прискорене зростання обчислювальних потужностей, масштабування контексту моделей та розширення сфер практичного застосування.

Список літератури:

1. What is AI code generation? / AWS URL: <https://aws.amazon.com/what-is/ai-coding/> (дата звернення: 19.01.2026)
2. The State of AI in Software Development / Gartner URL: <https://www.gartner.com/en/doc/ai-software-development> (дата звернення: 19.01.2026)
3. HumanEval: A Program Synthesis Benchmark / OpenAI URL: <https://github.com/openai/human-eval> (дата звернення: 19.01.2026)
4. MBPP: Mostly Basic Python Problems / Google Research URL: <https://github.com/google-research/google-research/tree/master/mbpp> (дата звернення: 19.01.2026)
5. SWE-Bench: Software Engineering Benchmarks / Princeton NLP URL: <https://github.com/princeton-nlp/SWE-bench> (дата звернення: 19.01.2026)
6. Aider Polyglot Benchmark / Exercism URL: <https://exercism.org/benchmarks/aider-polyglot> (дата звернення: 19.01.2026)
7. LiveCodeBench: No-contamination Code Evaluation / LiveCodeBench URL: <https://livecodebench.org> (дата звернення: 19.01.2026)
8. Benchmarking Agentic Code Frameworks / ArXiv preprint URL: <https://arxiv.org/abs/2403.12345> (дата звернення: 19.01.2026)
9. Context Window Sizes in LLMs / OpenAI Blog URL: <https://openai.com/blog/context-windows/> (дата звернення: 19.01.2026)
10. Evolution of Code Benchmarks / MLPerf URL: <https://mlperf.org> (дата звернення: 19.01.2026)

Kupin A.I., Dukhov D.Yu. RESEARCH INTO THE EFFICIENCY OF AUTOMATED SOFTWARE CODE GENERATION USING ARTIFICIAL INTELLIGENCE

The article provides a comprehensive analysis of current trends in automated code generation using artificial intelligence tools, in particular large language models (LLMs), based on deep learning and natural language processing methods. The evolution of approaches to code generation is considered - from traditional automation tools and low-code/no-code platforms to generative AI systems capable of creating code "from scratch" based on the developer's text queries. It is shown that such systems provide a significant increase in the productivity and efficiency of developers, a reduction in the number of errors and a reduction in software development cycles, which is confirmed by the results of modern empirical research.

Considerable attention is paid to the analysis of specialized benchmarks used to assess the quality and effectiveness of AI models in programming tasks. Benchmarks such as HumanEval, MBPP, SWE-Bench, Aider Polyglot, LiveCodeBench, APPS, MTPB, DS-1000, and WebDev Arena are described in detail, and their metrics, advantages, and limitations are identified. It is shown that the evolution of benchmarks from functional correctness verification of isolated code fragments to agent-based capabilities, multilingual editing, real-world error correction, and multi-stage interaction reflects the growing demands on AI intelligence in real-world software development scenarios. The paper analyzes the current state of commercial and open language models, including GPT-4.x, Claude 3.x, Gemini 2.x, Code Llama, and DeepSeek-Coder; and shows the trend of blurring the line between general-purpose and specialized coding models. The role of extended context windows, agent frameworks, and multimodal capabilities in shaping a new generation of AI assistants capable of performing complex tasks of analysis, debugging, refactoring, and planning of software systems is emphasized.

It is concluded that automated code generation using artificial intelligence is moving from instrumental support to a full-fledged partnership in the software development process, forming the basis for the emergence of more autonomous and integrated AI solutions in the future.

Keywords: generative artificial intelligence, trend analysis, code generation, benchmark research, technology development forecasting.

Дата першого надходження статті до видання: 13.03.2026

Дата прийняття статті до друку після рецензування: 10.04.2026

Дата публікації (оприлюднення) статті 11.05.2026

Суховій С.І.

Національний аерокосмічний університет
«Харківський авіаційний інститут»
<https://orcid.org/0009-0001-5533-6367>

Лакоза С.Л.

<https://orcid.org/0000-0001-6354-1611>
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Мацюк Д.М.

<https://orcid.org/0009-0004-5037-0568>
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»;
Акціонерне товариство науково-технічний комплекс «Електронприлад»

АНАЛІЗ СУЧАСНИХ ТЕНДЕНЦІЙ ВДОСКОНАЛЕННЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ СИСТЕМ КЕРУВАННЯ ГАЗОТУРБІННИХ ДВИГУНІВ

Система автоматичного керування (САК) значною мірою впливає на характеристики, що визначають конкурентоспроможність газотурбінного двигуна (ГТД). Підвищення точності підтримання заданих параметрів дозволяє ГТД працювати ближче до границь їх обмеження. Це забезпечує покращення динамічних властивостей ГТД і важливих експлуатаційних параметрів, зокрема, тяги та витрати палива. Водночас це гарантує збереження ресурсу двигуна, підвищення надійності та безпеки. Точність відпрацювання динамічних режимів ГТД суттєво впливає на можливості пілотування літальних апаратів, вірогідність успішного виходу з аварійних ситуацій. Оскільки в сучасних ГТД використовують електронні цифрові САК, досягнення найкращих показників ГТД значною мірою визначається програмним забезпеченням блоків керування. У даній роботі було проаналізовано основні тенденції вдосконалення ПЗ САК ГТД. Було розглянуто основні етапи розвитку САК ГТД, виявлено основні сучасні проблеми, що потребують досліджень. З аналізу літературних джерел було виділено чотири напрямки вдосконалення ПЗ САК ГТД: розробка релевантних математичних моделей, оптимізація програм регулювання, використання новітніх алгоритмів керування, технологія розробки ПЗ. Тенденції вдосконалення математичних моделей ГТД пов'язані з підвищенням точності та скороченням часу розрахунку. Для програм регулювання актуальним є оптимізація наявних варіантів вибору значень параметрів та пошук нових кращих варіантів. У напрямку вдосконалення алгоритмів керування виділяються тенденції створення нових перспективних підходів, а також удосконалення класичних рішень. Перспективні алгоритми поряд з перевагами мають ряд недоліків, таких як висока складність, великий час чисельних розрахунків, що стримує їх впровадження в реальних системах. Водночас широко використовуються нелінійні регулятори на базі ПД-регулятора з використанням логіки вибору активного контуру. Їх широке застосування в сучасних ГТД пов'язане із простотою реалізації та оптимальному співвідношенні характеристик швидкодії та якості керування. Тенденція вдосконалення технології розробки САК ГТД полягає в застосуванні модельно-орієнтованого проектування, що дозволяє скоротити час розробки, зменшити її вартість, а також підвищити безпечність та якість системи.

Ключові слова: газотурбінний двигун, алгоритм керування, система керування ГТД, вдосконалення САК, ПД-регулятор

Постановка проблеми. Система автоматичного керування (САК) значною мірою впливає на характеристики, що визначають конкурентоспроможність газотурбінного двигуна (ГТД). Підвищення точності підтримання заданих параметрів роботи ГТД дозволяє працювати ближче до їх граничних значень. Це значно покращує динамічні властивості ГТД і його важливі експлуатаційні параметри, зокрема, тягу та витрату палива. Водночас це гарантує збереження ресурсу двигуна, підвищення надійності та безпеки. Точність відпрацювання динамічних режимів ГТД значно впливає на можливості пілотування літальних апаратів, ймовірність успішного виходу з аварійних ситуацій.

У сучасних ГТД використовують електронні цифрові САК, а тому досягнення необхідних показників роботи ГТД значною мірою визначається програмним забезпеченням (ПЗ) блоків керування. Завданням даного дослідження є аналіз основних тенденцій вдосконалення ПЗ САК ГТД. Результати аналізу можна згрупувати у чотири категорії, як показано на рис. 1. Кожен з цих великих напрямків розглянути більш детально в даній роботі.

Аналіз останніх досліджень і публікацій. На основі аналізу літературних джерел можна виділити чотири основні етапи розвитку систем автоматичного керування ГТД [1, 2]. Перший етап (1945-1970) характеризувався переважним застосуванням гідромеханічних систем керування і найпростіших варіантів САК з одним входом і виходом.

На другому етапі (1970-1990) виникла і набула поширення теорія побудови регуляторів, які застосовувалися до відповідного активного контуру ГТД за допомогою логічних правил. Також у цей період почалося становлення теорії багатовимірних систем керування та перші спроби впровадження цифрових електронних регуляторів.

Третій етап (1990-2004) характеризувався застосуванням новітніх теорій і алгоритмів керування, зокрема, інтегрованого керування силовою установкою і літальним апаратом, багатовимірною

керування, активного керування, інтелектуального ресурсоподовжуючого керування, алгоритмів нечіткої логіки, імплементацію напівнатурних випробувань у практику розробки САК.

На четвертому етапі (з 2004) основна тенденція розвитку САК ГТД полягає в поширенні застосування електронних та електричних компонент у структурі САК, а також розвитку інноваційних алгоритмів керування, зокрема, керування на основі нелінійних математичних моделей ГТД, розподіленого керування, активного керування горінням та ситуаційного керування.

Виділяють наступні важливі проблеми та напрямки розвитку САК ГТД, що потребують подальших досліджень [3]: 1) невизначеність параметрів ГТД та об'єкта керування; 2) багатовимірність і багатокритеріальність; 3) відмовостійкість ГТД; 4) розподілені системи керування; 5) інтеграція в єдину САК літак/двигун.

Невизначеність найчастіше пов'язана з неточністю моделі. Тобто математична модель об'єкта керування (ОК), що використовується для розробки регулятора, не збігається з реальною системою в широкому діапазоні умов експлуатації. Водночас це поняття також містить невизначеність зовнішніх збурень. Здатність САК протидіяти різній невизначеності називається робастністю. Розвиток методів забезпечення робастності САК є одним з актуальних напрямків досліджень. Для розв'язання цієї проблеми застосовують лінійний і нелінійний підходи. У першому випадку використовують лінійну модель ОК, що отримана для однієї робочої точки. Її використовують для синтезу регулятора, що забезпечує стійкість в широкому діапазоні режимів. Нелінійні робастні методи базуються на приближенні нелінійних моделей шляхом використання лінійних моделей зі змінними коефіцієнтами. Також у таких алгоритмах широко застосовується ковзний режим керування та адаптивні регулятори, що мають властивості самоналаштування.

Багатовимірне керування актуально для складних систем, на роботу яких одночасно впливає велика кількість параметрів. Для досягнення

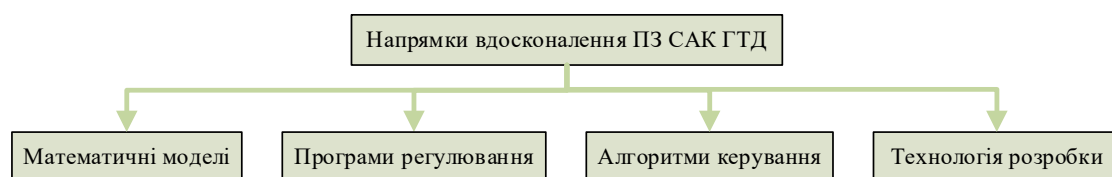


Рис. 1. Напрямки вдосконалення ПЗ САК ГТД

високих експлуатаційних характеристик сучасні САК ГТД, окрім витрати палива, як основного регульовального фактора, використовують інформацію про змінну геометрію, розміри або положення деяких елементів ГТД. Також цей підхід є обов'язковим при розв'язанні спеціальних задач, зокрема, забезпечення високої маневровості та швидкості польоту.

ГТД має декілька фізичних обмежень при роботі, дотримання котрих гарантує безпечну експлуатацію. Одним з напрямків підвищення ефективності ГТД є використання таких режимів, що наближені до граничних обмежень при його безпечній роботі. Суперечливий зв'язок між вимогами безпеки та високої ефективності ГТД обумовлює складну *багатокритеріальну задачу керування*.

Відмовостійкість САК – властивість системи забезпечувати стійкість та нормальну роботу ГТД в умовах, коли датчик, виконавчий пристрій або компонент двигуна вийшли з ладу. ГТД працює в широкому діапазоні зовнішніх умов і збурень, а тому схильний до відмов різних елементів та деградації їх параметрів. У таких умовах для двигуна важливо вчасно змінити стратегію керування для забезпечення безпеки та надійності. Сучасні відмовостійкі САК поділяються на два класи: активні та пасивні. При виявленні поточної відмови у пасивних системах САК переходить до використання відповідного завчасно спроектованого регулятора з метою забезпечення нормального функціонування двигуна. В активних системах вирішується складніша задача, яка полягає у забезпеченні не лише стійкості, а й робастності, необхідної тривалості перехідного процесу, мінімальних затрат на керування.

Розподілені САК – важливий напрямок вдосконалення САК ГТД [4, 5]. Нині більшість САК ГТД є централізованими цифровими електронними системами з повною відповідальністю. В таких системах всі функції обробки та керування виконуються всередині одного блоку [6, 7]. На сучасному етапі керування ГТД не обмежується керуванням витратою палива, а поширюється на вхідний пристрій, сопло, механізацію проточної частини. При керуванні такою складною системою розподілені САК забезпечують ряд переваг: 1) зменшення складності та маси; 2) застосування стандартизованих інтелектуальних пристроїв та інтерфейсів, скорочення термінів та вартості розробки; 3) зменшення вартості експлуатації, підвищення надійності та ремонтпридатності; 4) покращення гнучкості при розробці та оновленні системи.

Для широкого застосування таких систем потрібно вирішити ряд проблем. Серед них ключовими є застосування інтелектуальних датчиків та виконавчих пристроїв, використання високо-температурної електроніки, технологій мережевої комунікації та систем розподіленого живлення.

Постановка завдання. Метою статті є порівняльний аналіз сучасних тенденцій вдосконалення технології розробки САК ГТД при модернізації та проектуванні новітніх систем керування авіаційними ГТД. Коректний та оптимальний вибір підходу до технології розробки САК ГТД дозволяє скоротити час розробки, зменшити її вартість, а також підвищити точність, безпечність та якість системи.

Виклад основного матеріалу. Багато сучасних підходів до розробки САК ГТД базуються на методах оптимізації. Ось основні труднощі в розробці САК при оптимізації регуляторів ГТД [8]: 1) відсутність аналітичних рішень для настройки коефіцієнтів підсилення; 2) необхідність забезпечення компромісу між різними цілями керування; 3) необхідність кількісного подання усіх характеристик керування; 4) велика кількість локальних мінімумів і відповідних режимів роботи; 5) традиційні алгоритми оптимізації мають великий час обчислення, який часто зовсім неприйнятний для розв'язання таких задач.

Програмне забезпечення критично важливих для безпеки систем розробляється з високим рівнем надійності, що потребує значних зусиль, коштів і часу. Забезпечення надійності ПЗ САК ГТД при зменшенні коштів і часу розробки є актуальним напрямком прикладних робіт.

Актуальні напрямки досліджень в області керування ГТД показано на рис. 2.

Вдосконалення математичних моделей ГТД. Математична модель ОК використовується для синтезу алгоритму керування, вона є важливим компонентом САК для методів керування з використанням математичних моделей [3]. В обох випадках ставляться високі вимоги до точності опису ОК математичною моделлю, яка використовується. Оскільки ГТД являє собою складну, суттєво нелінійну, змінну в часі систему, побудова точної моделі для розрахунків у реальному часі є складною задачею.

Найбільшу точність забезпечують нелінійні повузлові моделі, які отримують на основі об'єднання моделей елементів двигуна шляхом узгодження їх входів і виходів. Але у ряді випадків нелінійні повузлові моделі не можуть бути використані безпосередньо для синтезу САК або

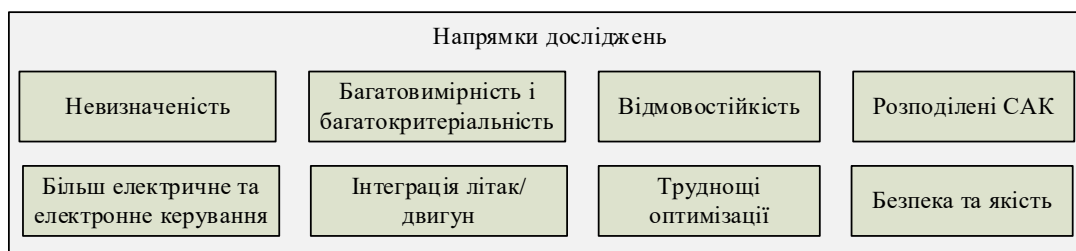


Рис. 2. Актуальні напрямки досліджень в області керування ГТД

для керування в реальному часі. Тому широкого застосування набули моделі в просторі станів або у вигляді передатних функцій, які придатні для розробки перспективних алгоритмів керування, моніторингу та діагностики. У більшості випадків ці моделі отримують за допомогою різних методів лінеаризації нелінійних повузлових моделей.

Згідно робіт [1, 9] основні труднощі розвитку бортових математичних моделей для застосування у керуванні ГТД та визначенні його технічного стану полягають в наступному: 1) забезпечення високої точності моделювання критичних параметрів, особливо на перехідних режимах; 2) забезпечення робастності моделей; 3) час розрахунку параметрів моделей; 4) верифікація і валідація програмного коду для забезпечення надійності та безпеки.

Запропоновано наступні потенційні шляхи вирішення цих труднощів [9]:

- застосування декількох різних математичних моделей для різних параметрів;
- застосування моніторингових систем реального часу для забезпечення надійності та безпеки.

Для зменшення часу розрахунків в [10] запропоновано методику побудови моделі типу «сірий ящик», як альтернативи моделям «білий ящик» і «чорний ящик», без втрати точності. Модель «білий ящик» авторами використано для опису динаміки ротора, моделі типу «чорний ящик» — для опису решти елементів двигуна.

У [11] виконано порівняння трьох типів моделей на основі нейронних мереж для мікро ГТД. Такі моделі мають переваги для випадків, коли немає характеристик вузлів нелінійної моделі. У роботі [12] запропоновано використовувати поєднання нейронних мереж і розширеного фільтра Калмана для оцінки тяги двигуна у змінному циклі роботи. У [13] в якості бортової моделі було запропоновано квазітривимірну модель динаміки ГТД.

У [14] розроблено алгоритм з підвищеною стійкістю числових розрахунків. Для цього авторами був розроблений покращений алгоритм сігма-

точкового фільтра Калмана, у якому застосовано сингулярний розклад. У роботі запропонована «офлайн» корекція не оцінюваних параметрів для зменшення негативного впливу їх деградації.

У роботі [15] виконано порівняння способів отримання лінійних моделей для задач, в яких необхідно брати до уваги деградацію компонентів ГТД.

Формування оптимальних програм регулювання. Формування програми регулювання ГТД дозволяє змінювати належним чином параметри двигуна згідно з умовами та вимогами експлуатації. Застосовуючи декілька програм регулювання можна змінювати параметри ГТД, здійснювати їх перемикання відповідно до умов польоту. Формування програми регулювання полягає у розв'язанні задачі оптимізації відносно заданої цільової функції.

У роботі [16] досліджено можливість покращення рвучкості турбореактивного двоконтурного двигуна з форсажною камерою (ТРДДФ) шляхом додаванням двох допоміжних керуючих впливів: перепуску повітря із середніх ступенів компресора та повороту соплового апарату турбіни низького тиску. Для отримання оптимальних програм зміни регульовальних факторів (перепуску повітря компресора, повороту соплового апарату турбіни низького тиску та відношення витрати повітря до тиску за компресором у точках експлуатаційної області) використовувався метод рою частинок (PSO) разом з методом послідовного квадратичного програмування. Отримані результати було розширено на всю експлуатаційну область, використовуючи методи теорії подібності.

У роботі [17] вирішувалась багатокритеріальна задача оптимізації за тягою, питомою витратою палива та витратою повітря. Авторами запропоновано гібридний метод формування оптимальних програм керування усталених режимів двигуна змінного циклу. Метод передбачає використання двох складових: пошуку оптимальних програм керування та бортової програми, що корегує програму керування при деградації технічного стану

ГТД. Першу складову побудовано на базі генетичного алгоритму, другу – на базі градієнтного методу.

Алгоритми керування. На рис. 3 наведено класифікацію регуляторів ГТД [3]. Лінійні регулятори є класичним підходом до розв'язку задач керування. До них відносять: 1) пропорційно-інтегрально-диференціальний (ПІД) регулятор. У САК ГТД широкого поширення набули його нелінійні варіанти з логікою перемикавання активного контуру; 2) лінійний квадратичний регулятор (ЛКР). Для налаштування вагових матриць і параметрів потребує ітераційного процесу; 3) лінійно-квадратичний гаусовий регулятор з відновленням передатної функції контуру. Має властивість придушувати зовнішні збурення і враховувати невизначеність системи, використовуючи алгоритм фільтра Калмана; 4) робастний H_∞ -регулятор. Враховує невизначеність системи та неминуче вносить у неї консерватизм, а тому не дозволяє забезпечити високі характеристики у широкому діапазоні робочих режимів.

Нелінійні регулятори. Лінійний регулятор з налаштуванням підсилення призначений для врахування суттєвої нелінійності ГТД, зміни зовнішніх умов і забезпечення прийнятних динамічних властивостей у всій експлуатаційній області. Класичний регулятор має фіксовану структуру з коефіцієнтами підсилення, що налаштовуються залежно від обраних параметрів. Параметри відображають зміну режиму ГТД і зовнішніх умов [18]. Лінійна інтерполяція застосовується для

визначення поточних підсилень регулятора для вимірних параметрів у реальному часі.

Лінійний регулятор зі змінними параметрами є узагальненням регулятора з налаштуванням підсилення, має більш строгі математичне обґрунтування. Гарантує стійкість і бажані характеристики САК в усій експлуатаційній області. Застосування декількох незалежних параметрів у регуляторі ускладнює проектування САК. Суттєвий недолік таких моделей – важкість отримання багатовимірного варіанта із задовільною точністю [3].

У роботі [19] цей метод було використано для побудови регулятора обмеження температури газу, регулятора прискорення ротора сумісно з основним регулятором сталих режимів. У роботі [20] було застосовано поліноміальну модель зі змінними коефіцієнтами для синтезу регулятора.

Регулятор ковзних режимів відрізняється від інших методів керування наявністю змінної структури, яка відповідає спроектованій траєкторії станів (поверхня ковзних режимів). Ідея такого керування полягає в тому, що система повинна досягти ковзної поверхні з будь-якого початкового стану і залишатися на цій поверхні завдяки регулятору. Спроектована поверхня ковзання описує ідеальні динамічні властивості системи.

Такий метод керування повністю інваріантний до зміни параметрів і збурень, тобто робастний. Метод простий у реалізації та дозволяє знизити розмірність використовуваної математичної моделі. Метод придатний для лінійних, нелінійних, безперервних та дискретних систем. Метод

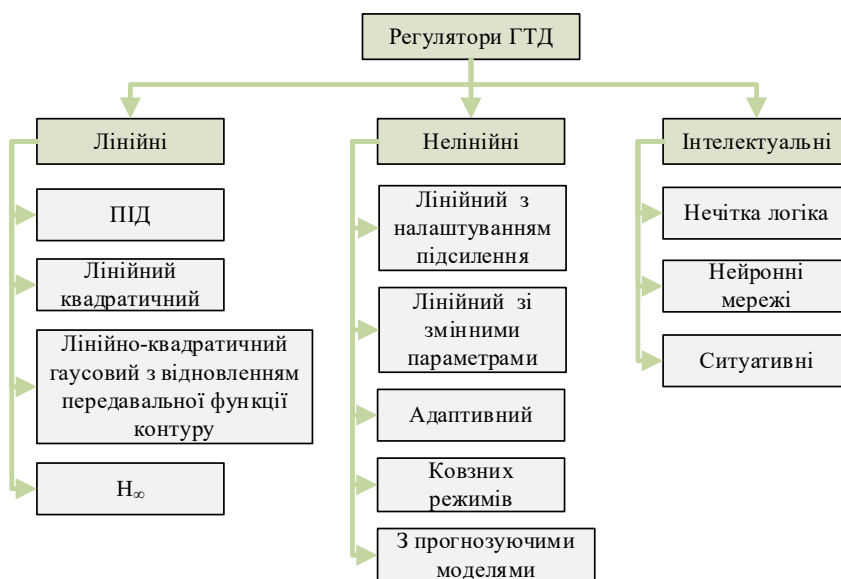


Рис. 3. Класифікація регуляторів ГТД

застосовувався для контурів обмеження параметрів [19], у системах з урахуванням квантування, а також у відмовостійких системах [21].

Адаптивний регулятор призначений нівелювати структурну або параметричну невизначеність, зовнішні збурення, деградацію системи або зміну зовнішніх умов. У ГТД набуло застосування адаптивне керування на основі математичних моделей. Воно просте за принципом та має багато методів проєктування. Бажані статичні та динамічні властивості ГТД визначаються моделлю. Перевага метода у тому, що він не потребує точної моделі ОК. Однак адаптивне керування має труднощі у практичному застосуванні. По перше, оцінювані параметри не збігаються до істинних значень під впливом шумів вимірюваних параметрів. По друге, сучасні адаптивні алгоритми керування не показують високих результатів при застосуванні у САК ГТД [3].

Регулятор з прогностичними моделями базується на використанні алгоритму оптимізації цільової функції в реальному часі. Цільова функція будується на основі порівняння параметрів моделі з вимірюваними параметрами та наперед визначеними константами. За типом використовуваної математичної моделі метод поділяється на лінійний та нелінійний. Для задач керування ГТД більше поширення має нелінійний підхід. Цей метод має дві суттєві переваги. По перше, можливо одночасно оптимізувати параметри ГТД і дотримуватися обмежень без складної логіки перемикавання. По друге, метод дозволяє отримати кращі результати ніж інші методи для багатовимірних випадків. Недоліком метода є велике обчислювальне навантаження, особливо, для нелінійних випадків [3]. Метод було застосовано для ТРДД [22], стаціонарних газотурбінних установок, вертолітних силових установок.

Інтелектуальні регулятори. До цього класу регуляторів відносять:

1) *Регулятор з нечіткою логікою.* Його алгоритм базується на розподілі нелінійного об'єкта системою правил на підсистеми, що мають спрощений опис. Два типи моделей застосовують для опису систем: Т-S керування і нейронні мережі. Перший тип моделей являє собою комбінацію констант або лінійних функцій, нейронні мережі є нелінійними моделями.

Принцип нечіткого Т-S керування полягає в тому, що двигун апроксимується серією зважених лінійних підсистем на основі системи правил. Недоліки метода Т-S керування полягають у великій складності нечітких моделей для сучасних

двигунів при суттєвому розширенні області експлуатації, а також для багатовимірних регуляторів. Розробка потребує експертних знань, а при використанні нейронних мереж існує проблема підтвердження стійкості [3].

2) *Регулятори на основі нейронних мереж* використовуються завдяки широким можливостям апроксимації нелінійностей (традиційний синтез регуляторів з використанням математичних моделей ставить високі вимоги до їх точності) [3].

3) *Ситуаційний регулятор* являє собою альтернативний підхід для формування структури САК [23]. У циклі роботи двигуна відокремлюються робочі режими. Для кожного з них розробляється регулятор, як окремий програмний модуль. Ситуаційний регулятор активує той чи інший регулятор залежно від поточного режиму роботи ГТД. Такий підхід обумовлює модульну архітектуру системи та надає можливості для інтегрування систем діагностики та керування.

ПІД-регулятори з налаштуванням підсилення і логікою вибору активного контуру керування. Комбінування ПІД-регуляторів з логікою вибору активного контуру розв'язало проблему керування багатьма параметрами ГТД, а в поєднанні з можливістю налаштування підсилення законів керування це дозволило враховувати ще нелінійність ГТД і зміну зовнішніх умов. Завдяки структурній простоті, високій робастності, вдосконаленню впродовж багатьох років регулятори даного типу набули довершеності та широкого застосування в серійних ГТД [24, 7]. Такі регулятори не втрачають актуальності і сьогодні. Вони використовуються в нових проєктах, служать базою для поступового впровадження інноваційних алгоритмів і підходів до розробки.

Класичною стала реалізація САК на основі ПІ-регуляторів з min/max-логікою вибору активного закону керування та алгоритмом протидії накопичуванню інтеграла неактивного регулятора [25]. Коефіцієнти основного регулятора, що визначає тягу, апроксимовано за висотою та швидкістю в числах Маха. Інтегральна компонента додатково масштабується згідно з робочим режимом двигуна та висотою. Регулятори граничних режимів на відміну від основного регулятора виконано із сталими коефіцієнтами.

Роздільний синтез контурів керування з подальшим об'єднанням в єдину систему не дає оптимальний результат при керуванні ГТД. Це пов'язано з тим, що в системі мають місце перемикавання контурів, які не враховуються в процесі розробки. Подолати цей недолік можна вибором параметрів

з використанням оптимізації цільової функції, яка складається на основі вимог до САК. Так у роботі [26] синтез ПІ-регулятора частоти обертання ротора ГТД було сформульовано як задачу оптимізації за алгоритмом диференціальної еволюції. Даний метод застосовано в декількох точках дросельної характеристики для лінеаризованих моделей і апроксимовано для всього робочого діапазону.

Технологія розробки САК ГТД. Модельно-орієнтований підхід (МОП) до проєктування алгоритмів САК ГТД набув широкого застосування завдяки ефективності і багатьом перевагам. МОП є методикою, що використовується для розробки та тестування критично важливих для безпеки систем [27]. Вона передбачає розробку моделі системи керування, її моделювання та налагодження за допомогою спеціальних інструментів, а потім автоматичну генерацію коду ПЗ керування для цільового мікроконтролера. У сфері вбудованих систем такий підхід надає величезну перевагу і дозволяє уникнути трудомісткого та схильного до помилок останнього етапу кодування.

МОП довів свою ефективність завдяки багатьом успішним впровадженням у різних галузях промисловості. Цей підхід було застосовано для дослідження можливості рекуперації залишкового тепла у вихлопних газах потужного двигуна внутрішнього згоряння [28], проєктування теплових накопичувачів для теплових насосів, концептуального проєктування архітектури системи рідкого водневого палива.

Приклад реалізації методу МОП у Simulink для вбудованих систем керування наведено у [29]. Після завершення формування моделі у Simulink вона перетворюється на проміжну формальну модель на мові синхронного програмування Lustre. Проміжна формальна модель служить для генерування багатозадачного коду для операційної системи реального часу Xenomai RTOS.

Автори роботи [30] досліджували проблеми, з якими стикається авіаційна галузь під час розробки складних систем у децентралізованому середовищі. Оскільки розробка літаків, двигунів та обладнання виконується багатонаціональними та географічно розподіленими командами, традиційні документоорієнтовані підходи виявляються громіздкими. МОП розв'язує ці проблеми. Він мінімізує залежність від документів, покращує співпрацю та пропонує централізовану систему модель для забезпечення узгодженості усіх рішень. Цей підхід не лише обіцяє підвищення ефективності, але й дозволяє зменшити сегментованість процесу розробки в аерокосмічній галузі.

Концепція МОП та інструменти для її втілення для систем літака обговорювалися в роботі [31]. Інструментальні набори MathWorks та Modelica розглядаються як середовища для реалізації пропонуваної концепції.

Авіаційні газотурбінні двигуни є дуже складними системами, що включають багато модулів та підсистем. САК ГТД з повною відповідальністю авіаційного двигуна сама по собі є дуже складною підсистемою, що охоплює апаратне та програмне забезпечення для керування двигуном на всіх режимах роботи. Тому доцільно використовувати МОП для розробки таких систем.

Висновки. Розвиток та вдосконалення новітніх алгоритмів керування сучасних САК ГТД пов'язано з вдосконалення математичних моделей, підвищенням точності та скороченням часу розрахунків у них. У напрямку вдосконалення алгоритмів керування виділяються тенденції створення нових перспективних підходів, а також удосконалення класичних схем. Перспективні алгоритми поряд з перевагами мають ряд недоліків, таких як висока складність, великий час чисельних розрахунків, що стримує їх впровадження в реальних системах. Водночас широко використовуються нелінійні регулятори на базі ПІД-регулятора з використанням логіки вибору активного контуру. Їх широке застосування в сучасних ГТД пов'язане із простотою реалізації та оптимальному співвідношенні характеристик швидкодії та якості керування. Перспективи їх використання полягають у застосуванні для мало-розмірних ГТД, комбінуванні з перспективними регуляторами, використанні методів оптимізації для їх налаштування. Для програм регулювання параметрів актуальним є оптимізація наявних та пошук нових покращених варіантів обраних характеристик законів керування.

Вдосконалення технологій розробки САК ГТД полягає в застосуванні модельно-орієнтованого проєктування, що дозволяє скоротити час розробки, зменшити її вартість, а також підвищити безпечність та якість системи.

Вдосконалення технологій розробки САК ГТД полягає в застосуванні модельно-орієнтованого проєктування, що дозволяє скоротити час розробки, зменшити її вартість, а також підвищити безпечність та якість системи.

Список літератури:

1. Adibhatla S. et al. Propulsion control technology development needs to address NASA aeronautics research mission goals for thrusts 3a and 4: materials of the AIAA Propulsion and Energy Forum 2018, Cincinnati, OH, United States, Jul. 7–13, 2018. Cincinnati, 2018. DOI: 10.2514/6.2018-4732.

2. Mohammadi S. J. et al. A scientometric analysis and critical review of gas turbine aero-engines control: From Whittle engine to more-electric propulsion. *Meas. Control*. 2021. Vol. 54, no. 4. P. 935–966. DOI: 10.1177/0020294020956675.
3. Lv C. et al. Recent research progress on airbreathing aero-engine control algorithm. *Propuls. Power Res.* 2022. Vol. 11, no. 1. P. 1–57. DOI: 10.1016/j.jprr.2022.02.003.
4. Behbahani A. R. et al. Secure embedded distributed control and instrumentation architecture for aircraft propulsion systems: framework, process, methods, challenges, and opportunities: materials of the ASME Turbo Expo 2022 Turbomachinery Technical Conference and Exposition, Rotterdam, Netherlands, Jun. 17–21, 2022. Rotterdam, 2022. DOI: 10.1115/GT2022-82644
5. Pakmehr M. et al. Distributed control of turbofan engines: materials of the 45th AIAA/ASME/SAE/ASEE Joint Propulsion Conference, Denver, Colorado, Aug. 2–5, 2009. Denver, 2009. DOI: 10.2514/6.2009-5532.
6. Ji X. et al. A linear iterative controller for software defined control systems of aero-engines based on LMI. *Actuators*. 2023. Vol. 12, no. 7. Art. no. 259. DOI: 10.3390/act12070259.
7. Ji X. et al. A decentralized LQR output feedback control for aero-engines. *Actuators*. 2023. Vol. 12, no. 4. Art. no. 164. DOI: 10.3390/act12040164.
8. Liu Y. et al. Advanced optimization of gas turbine aero-engine transient performance using linkage-learning genetic algorithm: Part I, building blocks detection and optimization in runway. *Chin. J. Aeronaut.* 2021. Vol. 34, no. 4. P. 526–539. DOI: 10.1016/j.cja.2020.07.034.
9. Wei Z. et al. Gas turbine aero-engines real time on-board modelling: A review, research challenges, and exploring the future. *Prog. Aero. Sci.* 2020. Vol. 121. Art. no. 100693. DOI: 10.1016/j.paerosci.2020.100693.
10. Moroto R. H. et al. Linear grey-box modeling of gas turbine engines for MIMO control design: materials of the ASME Turbo Expo 2018 Turbomachinery Technical Conference and Exposition, Oslo, Norway, Jun. 11–15, 2018. Oslo, 2018. Paper GT2018-75847. DOI: 10.1115/GT2018-75847
11. Wu X. et al. A neural network-based modeling approach for transient performance prediction of gas turbine engines. *J. Braz. Soc. Mech. Sci. Eng.* 2025. Vol. 47. Art. no. 119. DOI: 10.1007/s40430-025-05408-0.
12. Ma A. et al. Thrust online fusion estimation of high-flow dual variable cycle engine. *Aerosp. Sci. Technol.* 2024. Vol. 152, no. 12. Art. no. 109388. DOI: 10.1016/j.ast.2024.109388.
13. Dai Y. et al. Enhanced quasi-three-dimensional transient simulation technique incorporating component volume effects for aero engine. *Chin. J. Aeronaut.* 2024. Vol. 37, no. 12. P. 325–346. DOI: 10.1016/j.cja.2024.07.020.
14. Chen Q. et al. An improved nonlinear onboard adaptive model for aero-engine performance control. *Chin. J. Aeronaut.* 2023. Vol. 36, no. 10. P. 317–334. DOI: 10.1016/j.cja.2022.12.005.
15. Castiglione T. et al. Linear model of a turboshaft aero-engine including components degradation for control-oriented applications. *Energies*. 2023. Vol. 16, no. 6. Art. no. 2634. DOI: 10.3390/en16062634.
16. Cai C. et al. Full-envelope acceleration control method of turbofan engine based on variable geometry compound adjustment. *Aerosp. Sci. Technol.* 2022. Vol. 128, no. 8. Art. no. 107748. DOI: 10.1016/j.ast.2022.107748.
17. Ji R. et al. A hybrid design method of steady-state throttling control schedules for high-flow variable cycle engine. *Aerosp. Sci. Technol.* 2024. Vol. 151. Art. no. 109284. DOI: 10.1016/j.ast.2024.109284.
18. Härefors M. Application of H_∞ robust control to the RM12 jet engine. *Control Eng. Practice*. 1997. Vol. 5, no. 9. P. 1189–1201. DOI: 10.1016/S0967-0661(97)84358-4.
19. Yang S. et al. LPV based sliding mode control for limit protection of aircraft engines: materials of the ASME Turbo Expo 2018 Turbomachinery Technical Conference and Exposition, Oslo, Norway, Jun. 11–15, 2018. Oslo, 2018. Paper GT2018-77236. DOI: 10.1115/GT2018-77236.
20. Yang B. et al. Non-affine parameter dependent LPV model and LMI based adaptive control for turbofan engines. *Chin. J. Aeronaut.* 2019. Vol. 32, no. 3. P. 585–594. DOI: 10.1016/j.cja.2018.12.031.
21. Ashraf M. A. et al. A robust sensor and actuator fault tolerant control scheme for nonlinear system. *IEEE Access*. 2022. Vol. 10. P. 626–637. DOI: 10.1109/ACCESS.2021.3137388.
22. Nikolaidis T. et al. Advanced constraints management strategy for real-time optimization of gas turbine engine transient Performance. *Appl. Sci.* 2019. Vol. 9, no. 24. Art. no. 5333. DOI: 10.3390/app9245333.
23. Andoga R. et al. Intelligent situational control of small turbojet engines. *Aerosp. Eng.* 2018. Vol. 2018. Art. no. 8328792. DOI: 10.1155/2018/8328792.
24. Chen H. et al. Design of control laws based on inverted decoupling and LMI for a turboprop engine: materials of the ASME Turbo Expo 2019 Turbomachinery Technical Conference and Exposition, Phoenix, Arizona, USA, Jun. 17–21, 2019. Phoenix, 2019. DOI: 10.1115/GT2019-90429.
25. Csank J. et al. Control design for a generic commercial aircraft engine: materials of the 46th AIAA/ASME/SAE/ASEE Joint Propulsion Conference, Nashville, Tennessee, Jul. 25–28, 2010. Nashville, 2010. DOI: 10.2514/6.2010-6629.
26. Jiang Z. et al. Design of speed closed-loop control with variable pressure difference valve for aero engine: materials of the ASME Turbo Expo 2020 Turbomachinery Technical Conference and Exposition, Virtual, Online, Sept. 21–25, 2020. Online, 2020. DOI: 10.1115/GT2020-14737.

27. Broy M. et al. What is the benefit of a model-based design of embedded software in the car industry. Emerging Technologies for the Evolution and Maintenance of Software Models / ed. by J. Rech, C. Bunse. Hershey : IGI Global, 2011. P. 410–443. DOI: 10.4018/978-1-4666-4301-7.ch017.
28. Prospero F. et al. Model based design of a turbo-compound bottomed to internal combustion engine exhaust gas. J. Phys.: Conf. Ser. 2024. Vol. 2893. Art. no. 012095. DOI: 10.1088/1742-6596/2893/1/012095.
29. Parthasarathi H. et al. Model-based systems engineering for aero gas turbine engine subsystems: materials of the INCOSE International Symposium, Bangalore, India, 2016. Bangalore, 2016. Vol. 26(s1). P. 70–82. DOI: 10.1002/j.2334-5837.2016.00315.x.
30. Kuhn M. R. et al. Model based specifications in aircraft systems design: materials of the 11th International Modelica Conference, Versailles, France, Sept. 21–23, 2015. Versailles, 2015. DOI: 10.3384/ecp15118491.
31. Alras M. et al. Model-based design of embedded control systems by means of a synchronous intermediate model: materials of the International Conference on Embedded Software and Systems, Hangzhou, Zhejiang, P. R. China, May 25–27, 2009. Hangzhou, 2009. DOI: 10.1109/ICESS.2009.36.

Sukhovii S.I., Lakoza S.L., Matsiuk D.M. MODERN TRENDS ANALYSIS OF GAS-TURBINE ENGINE CONTROL SYSTEMS SOFTWARE IMPROVING

The Automatic Control System (ACS) significantly impacts the characteristics that determine the competitiveness of a gas turbine engine (GTE). Improving the accuracy of maintaining set parameters allows GTE to operate closer to its operational limits, thereby enhancing dynamic properties and critical operational parameters such as thrust and fuel consumption while ensuring engine life preservation, reliability, and safety. The accuracy tracking of GTE dynamic modes substantially influences on aircraft piloting capabilities and the probability of successful exit from emergency situations. Modern GTEs utilize electronic digital ACS, achieved peak performance is largely determined by the control unit software. This paper analyzes the main trends in GTE ACS software improvement. It is examined the key stages of ACS development, and identified primary current issues requiring research. In the work it is highlighting four main directions for improvement: the development of relevant mathematical models, optimization of control programs, implementation of advanced control algorithms, and software development technology. Modern trends in GTE mathematical modeling are focused on increasing accuracy and reducing computation time. For tuning programs, the priority is optimizing of parameter selection and searching for superior alternatives; simultaneously. In the field of control algorithms, there is a trend toward both development promising new approaches and the refinement of classical solutions. Although new advanced promising algorithms face drawbacks such as high complexity and long numerical calculation times that limits their implementation in real systems. At the same time, nonlinear controllers based on PID structures using active loop selection logic remain widely used due to their ease of implementation and optimal balance between response speed and control quality. The overall trend in improving of GTE ACS development technology lies in the application of Model-Based Design, which allows to reduce development time, decreasing its cost, and enhancing safety and quality of the system.

Keywords: gas-turbine engine, control algorithm, GTE control system, gas turbine engine, ACS improvement, PID-controller.

Дата першого надходження статті до видання: 09.02.2026
Дата прийняття статті до друку після рецензування: 04.03.2026
Дата публікації (оприлюднення) статті 11.05.2026

Lomovatskyi A.A.<https://orcid.org/0009-0004-5170-3272>

Lviv Polytechnic National University

Basyuk T.M.<https://orcid.org/0000-0003-0813-0785>

Lviv Polytechnic National University

METHODS FOR SENTIMENT ANALYSIS OF UKRAINIAN-LANGUAGE CONTENT: RULE-BASED, MACHINE LEARNING AND TRANSFORMER APPROACHES

Sentiment analysis is integral to contemporary natural language processing, as it autonomously identifies emotional tone within textual data across diverse domains such as social media, decision-making, marketing, and public opinion. Nonetheless, conducting sentiment analysis in morphologically rich, low-resource languages like Ukrainian presents significant challenges due to limited annotated datasets, linguistic diversity, flexible syntactic structures, and complex grammatical frameworks. These obstacles hinder the direct application of methods originally devised for English and other high-resource languages.

This article examines recent sentiment analysis techniques for Ukrainian content, emphasising rule-based systems, traditional machine learning, and transformer-based language models. It outlines their fundamental principles, highlighting their benefits, challenges, and suitability for Ukrainian texts, particularly in resource-scarce environments. The discussion considers rule-based methods as a practical starting point, machine learning models as adaptable classifiers, and transformer models as sophisticated tools capable of capturing contextual nuances and complex semantics.

This paper provides a brief comparison of the performance of various methods on Ukrainian-language data, employing standard metrics such as accuracy, precision, recall, and F1-score. The results demonstrate that, although rule-based methods continue to be valuable in specific contexts and serve as benchmarks, transformer models attain higher accuracy, exhibit improved generalisation, and show greater robustness to linguistic diversity. These findings assist in shaping methodologies for sentiment analysis in Ukrainian content and highlight potential opportunities for hybrid techniques to enhance low-resource language processing.

Keywords: sentiment analysis, Ukrainian language, natural language processing, rule-based methods, machine learning, transformer models, low-resource languages.

Formulation of the problem. Amid the rapid growth of textual data, sentiment analysis has become a vital part of modern natural language processing, enabling the automated assessment of emotional tone in user-generated content across social media platforms and within analytical systems. The ability to automatically recognise sentiment polarity allows organisations to monitor public opinion, analyse large-scale textual data, and enhance decision-making processes across various fields.

Although there has been notable progress in sentiment analysis for high-resource languages, applying modern methods to Ukrainian-language content remains challenging. These issues stem from the limited availability of annotated datasets, the morphological complexity of the Ukrainian language, flexible

word order, and the variability of syntactic structures [1]. Consequently, many approaches designed for English and other high-resource languages cannot be directly adapted to Ukrainian without making methodological adjustments.

Modern sentiment analysis methods are usually categorised into rule-based systems, traditional machine learning techniques, and transformer-based language models. Each of these methods offers its own benefits and drawbacks when used with Ukrainian texts, especially in low-resource settings where annotated data and computational power are limited.

Therefore, an important scientific problem is to systematise contemporary approaches to sentiment analysis of Ukrainian-language content, identify their advantages and limitations, and assess their appli-

cability in low-resource conditions. Addressing this problem contributes to the development of effective intelligent systems for Ukrainian-language text processing and supports further advancement of natural language processing methods for morphologically rich languages [2].

Analysis of recent research and publications.

Recent advancements in sentiment analysis reveal considerable progress across various methodological paradigms, including rule-based approaches, traditional machine learning techniques, and transformer-based language models [3]. Early sentiment analysis systems mainly relied on lexicon- and rule-based methods, which remain pertinent for their interpretability and stability in low-resource environments. These approaches typically utilise sentiment dictionaries, polarity scoring mechanisms, and linguistic rules to determine sentiment orientation, making them especially appropriate for morphologically rich languages where annotated data are limited [4].

Recent progress in statistical natural language processing has led to the widespread use of machine learning techniques like Naïve Bayes, Support Vector Machines, and logistic regression for sentiment analysis. These approaches facilitate automated feature extraction and scalable classification but depend heavily on large, annotated datasets, which remain limited for Ukrainian-language content [5]. Consequently, their effectiveness can vary significantly depending on the domain and dataset size.

Recently, transformer-based language models like BERT and its variants have achieved leading results in sentiment analysis by effectively capturing context and semantic links in the text. Multilingual transformer models also support cross-lingual transfer learning, which benefits low-resource languages. However, deploying these models typically demands significant computational resources and domain-specific fine-tuning to perform well in Ukrainian-language environments [6].

Although there has been an increase in sentiment analysis studies, research remains fragmented in methodology, especially for Ukrainian-language content. Most works focus on evaluating individual approaches rather than providing a broad comparison across methods. Additionally, challenges such as linguistic variability, limited resources, and domain adaptation are not sufficiently explored in the existing literature [7].

Thus, additional research is needed to organise current sentiment analysis techniques and evaluate their effectiveness for Ukrainian-language texts, particularly in low-resource environments.

Task statement. The purpose of this study is to organise contemporary approaches to sentiment analysis and assess their suitability for Ukrainian-language content in low-resource conditions. The study aims to identify methodological strengths and weaknesses of different approaches and determine their effectiveness in processing morphologically rich texts.

To achieve this goal, the following objectives are defined:

- to analyse existing methodological approaches to sentiment analysis, including rule-based, machine learning, and transformer-based methods.
- to generalise the theoretical foundations of sentiment analysis for Ukrainian-language text processing.
- to organise the advantages and limitations of modern sentiment analysis approaches in low-resource environments.
- to identify key challenges related to applying modern sentiment analysis techniques to morphologically rich languages.
- to justify the feasibility of combining methodological approaches for enhancing sentiment classification performance.

Outline of the main content of the study. This research formalises sentiment analysis of Ukrainian-language texts as a supervised multi-class classification problem within a unified decision-making framework. The set of textual inputs is represented as a finite collection of natural-language messages, where each instance corresponds to a single text, and the overall dataset consists of N observations.

$$X = \{x_1, x_2, \dots, x_N\} \quad (1)$$

where x_i – individual text, N – total number of texts.

Each text can be represented as an ordered sequence of linguistic units (2), reflecting its internal syntactic and lexical structure [8].

$$x_i = (w_1, w_2, \dots, w_{L_i}) \quad (2)$$

where w_j – word/token, L_i – length of the text.

The sentiment classification task is defined over a finite label space (3):

$$Y = \{y_1, y_2, \dots, y_K\} \quad (3)$$

where K is the number of sentiment classes.

This study utilises a three-class setup (4), representing positive, negative, and neutral sentiments. This method frames sentiment analysis as a categorical decision-making process, well-suited to applied information systems that require clear, interpretable results.

$$Y = \{text_{negative}, text_{positive}, text_{neutral}\} \quad (4)$$

The annotated dataset utilised for modelling can be represented as a set of input–output pairs (5), where each textual instance is linked with a predefined sentiment label [9].

$$D = (x_i, y_i)_{i=1}^N \quad (5)$$

where $x_i \in X$ – text instance, $y_i \in Y$ – corresponding sentiment label.

Within this formulation, sentiment analysis can be conceptualised as a functional mapping from the domain of textual inputs to the domain of sentiment labels (6).

$$f : X \rightarrow Y \quad (6)$$

where f is the sentiment classification function.

For data-driven approaches, this mapping is parameterised (7), allowing models to learn decision boundaries from empirical observations.

$$f_{\{\theta\}}(x) = \hat{y} \quad (7)$$

where θ – model parameters, $\hat{y}$ – predicted sentiment label.

From a modelling perspective, sentiment classification typically comprises two stages: representation learning and decision-making. The representation stage converts textual inputs into structured formats suitable for automated processing, while the classification stage assigns sentiment labels based on learned or predefined decision rules (8).

$$f_{\{\theta\}}(x) = g_{\theta}(p(x)) \quad (8)$$

where g_{θ} – classifier operating on representations.

The final decision is made using a maximum a posteriori criterion (9), ensuring consistent label selection across modelling paradigms.

$$\hat{y} = \arg \max_{y \in Y} p(y|x, \theta) \quad (9)$$

where $p(y|x, \theta)$ – predicted probability of class y , $\hat{y}$ – final predicted sentiment label.

This formalisation provides a unified conceptual basis for evaluating various sentiment analysis methods within consistent methodological conditions. Using a shared decision framework, differences in empirical outcomes are primarily attributed to modelling features rather than inconsistencies in task definition.

To facilitate an objective and consistent comparison across various sentiment analysis techniques, this study uses a standardised evaluation framework that standardises all phases of the experimental process. This framework establishes uniform methodological conditions for rule-based, traditional machine learning, and transformer-based models, allowing differences in performance to be mainly ascribed to the models' features rather than to variations in experimental setup [10].

The framework's overall structure is shown in Fig. 1. It comprises five interconnected layers: unifying datasets and preprocessing; adapting representations across different modelling paradigms; a unified protocol for training and fine-tuning; a standardised evaluation process; and a comparative analysis with reporting. This layered design ensures methodological consistency and encourages reproducible experiments across all approaches.

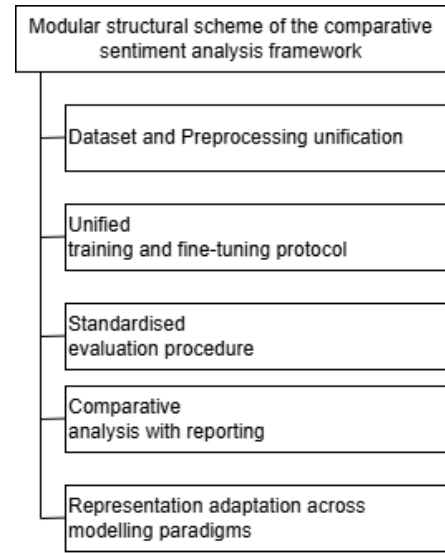


Fig. 1. Modular structural scheme of the comparative sentiment analysis framework

Within the proposed framework, all models are evaluated using a consistent annotated dataset, standardised sentiment label definitions, and a uniform data partitioning strategy. The dataset is divided into mutually exclusive training and testing subsets (10–11), with their proportions set by a fixed split coefficient (12). This method prevents information leakage and promotes fairness across different modelling approaches, particularly for data-driven methods trained on labelled data.

$$D = D_{train} \cup D_{test} \quad (10)$$

where D_{train} – training subset, D_{test} – testing subset.

$$D_{train} \cap D_{test} = \emptyset \quad (11)$$

where $\emptyset$ – denotes the empty set.

$$|D_{train}| = \alpha N,$$

$$|D_{test}| = (1 - \alpha) N \quad (12)$$

where $\alpha \in (0, 1)$ – dataset split coefficient, N – total number of samples.

To ensure comparability, a standardised set of evaluation metrics is used across all methods. Predicted sentiment labels serve as classification outputs

(13), with ground-truth labels defined similarly (14). The classification results are presented as label pairs (15), enabling systematic differentiation between correct and incorrect predictions via an indicator-based approach (16). The overall predictive accuracy is then measured with a unified statistical metric (17).

$$\hat{Y} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_M) \quad (13)$$

where $\hat{y}_i$ – predicted sentiment label, M – number of test samples.

$$Y = (y_1, y_2, \dots, y_M) \quad (14)$$

where y_i – true sentiment label for sample i .

$$(y_i, \hat{y}_i) \quad (15)$$

where y_i – true label, $\hat{y}_i$ – predicted label.

$$\| (y_i = \hat{y}_i) \quad (16)$$

where $\|(\cdot) = 1$ if the condition holds, $\|(\cdot) = 0$ otherwise.

$$\mathbb{E}[\| (y = \hat{y})] \quad (17)$$

where $\mathbb{E}[\cdot]$ – expectation operator.

By maintaining consistent experimental conditions, this framework provides a reliable foundation for comparing sentiment analysis methods. It ensures reproducible results, facilitates clear interpretation of performance, and enables systematic assessment of sentiment analysis techniques for Ukrainian-language systems in low-resource settings.

The empirical evaluation is conducted on a Ukrainian sentiment corpus containing 9,161 short texts from social media. Positive and negative samples are naturally occurring user posts, whereas the neutral class includes brief factual statements that differ structurally from emotionally expressive content. The sentiment category distribution is displayed in Table 1.

Table 1

Full corpus distribution

Class	Train (80%)	Test (20%)	Total	Share
Positive	5,251	1,314	6,565	71.7%
Negative	1,677	419	2,096	22.9%
Neutral	400	100	500	5.5%
Total	7,328	1,833	9,161	100%

To maintain methodological consistency, the corpus is split into training and testing subsets through stratified sampling with a fixed random seed. This approach preserves class proportions and ensures fair comparison across different modelling paradigms. The annotated dataset facilitates evaluation of sen-

timent classification under consistent experimental conditions within a unified comparative framework.

Before modelling, texts undergo standard preprocessing steps. These include normalising whitespace, removing extra spaces, and replacing typographic symbols with ASCII equivalents. No lowercasing, stop-word filtering, or lemmatisation is applied, as these processes can distort lexical signals crucial for rule-based and TF-IDF methods and have limited advantages for subword transformer tokenisers.

All models are trained and evaluated under consistent experimental conditions. Classical machine learning baselines are trained on the full training subset, whereas transformer-based models are fine-tuned with a fixed hyperparameter set. A standardised training protocol is employed to facilitate cross-architecture comparisons, using uniform optimisation settings, batch sizes, sequence length limits, and reproducibility measures.

This comparative evaluation examines the performance of rule-based, classical machine-learning, and transformer-based sentiment analysis techniques within a single experimental framework. It aims to identify overall performance trends, analyse class-specific behaviour, and assess the relative effectiveness of various modelling approaches for sentiment classification in Ukrainian.

Model performance is evaluated using standard classification metrics such as accuracy, precision, recall, and F1-score, all calculated under consistent conditions across approaches. The confusion matrix, as defined in (18), helps analyse the distribution of correct and incorrect predictions among sentiment classes. Additional metrics, such as true positives, false positives, and false negatives, calculated according to (19-21), enable detailed evaluation at the class level.

$$C = [c_{ij}] \quad (18)$$

where c_{ij} – number of samples whose true sentiment class is i and which are predicted as class j .

$$TP_k = c_{kk} \quad (19)$$

where TP_k – true positives for sentiment class k .

$$FP_k = \sum_{i \neq k} c_{ik} \quad (20)$$

where FP_k – number of samples incorrectly predicted as class k .

$$FN_k = \sum_{j \neq k} c_{kj} \quad (21)$$

where FN_k – number of samples of class k incorrectly predicted as other classes.

The overall classification results reveal consistent performance differences across different modelling paradigms. Transformer-based models perform best,

given their ability to capture contextual dependencies, implicit sentiment cues, and intricate semantic relationships. Classical machine learning techniques also perform well in controlled environments, especially when lexical signals closely match sentiment polarity. Conversely, rule-based methods tend to lag, as they rely heavily on fixed lexical resources and have limited ability to model context.

Table 2 summarises the quantitative comparison of the evaluated models. The results indicate that data-driven approaches clearly outperform purely rule-based systems across all evaluation metrics. These performance differences are particularly evident in macro-averaged measures, which provide a more balanced view of classification across sentiment categories.

Table 2

Comparison of sentiment analysis models

Model	Accuracy	Precision (M)	Recall (M)	F1 (M)	Precision (m)	Recall (m)	F1 (m)
3_tfidf_logreg	0.98036	0.975485	0.98815	0.981499	0.98036	0.9803	0.98036
2_custom_ua_hybrid	0.971631	0.950001	0.98247	0.965465	0.971631	0.9716	0.971631
4_Geotrend_distilbert-base-uk-cased	0.962357	0.9516	0.97924	0.964041	0.962357	0.9623	0.962357
1_ua_lexicon_rulebased	0.749591	0.611506	0.69611	0.61494	0.749591	0.7495	0.749591

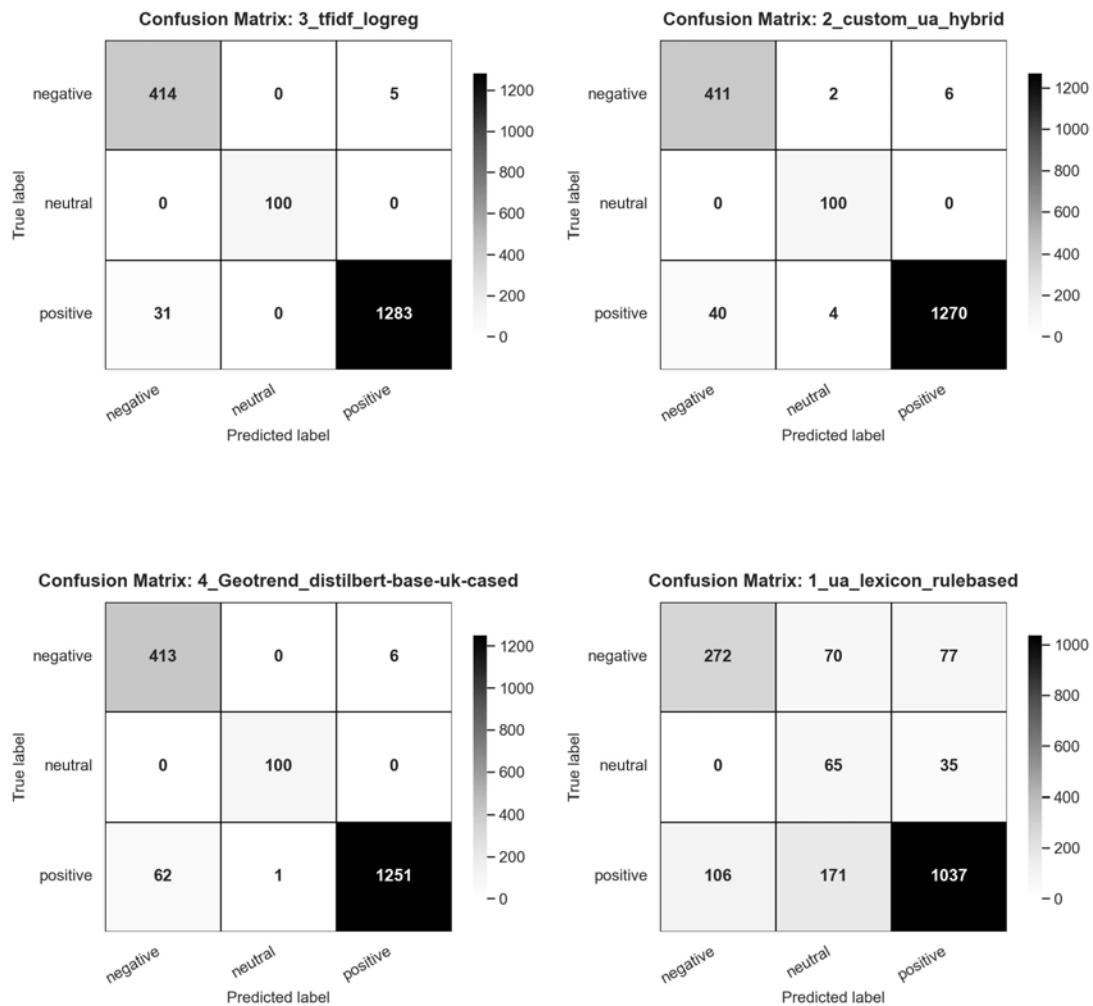


Fig. 2. Confusion matrices of sentiment analysis models

Figure 2 displays confusion matrices for the evaluated models, showing the distribution of correct and misclassified predictions across sentiment classes.

A comparison of class-level performance for the different approaches under evaluation is shown in Fig. 3. The results reveal significant differences in overall classification accuracy among the various modelling approaches. The basic lexical rule-based approach achieves an accuracy of 74.96%, highlighting the limitations of fixed-dictionary methods when processing languages with complex morphology.

Classical machine learning methods demonstrate significantly better results. The TF-IDF model combined with logistic regression achieves the highest accuracy among all the approaches studied—98.04%—demonstrating the high effectiveness of linear classifiers in sentiment analysis tasks, where polarity is often determined by explicit lexical markers.

A hybrid approach that combines lexical and statistical features achieves an accuracy of 97.16%, confirming the importance of integrating diverse feature sources to improve model robustness.

The DistilBERT-base-uk-cased transformer model also shows high performance, achieving an accuracy of 96.24%. Although its results are slightly lower than TF-IDF + Logistic Regression, transformer architectures offer more context-dependent representations of text and can better generalise to complex semantic dependencies.

Overall, the evaluation confirms that contextual modelling methods are effective for Ukrainian senti-

ment analysis. It also provides empirical support for the use of transformer-based techniques in practical information systems in low-resource settings.

Despite the high overall performance of several evaluated models, a detailed analysis reveals systematic challenges in sentiment classification in Ukrainian-language texts. The most frequent errors occur when sentiment polarity is weakly expressed or ambiguously conveyed in context. Specifically, neutral statements often include lexical elements that may appear sentiment-bearing outside their context, leading to misclassification by both statistical and neural models.

Another reason for classification errors arises from Ukrainian-specific linguistic characteristics. The language's morphological variability, flexible word order, and common use of intensifiers or negation constructions complicate sentiment analysis, particularly for rule-based and traditional machine learning approaches that rely mainly on surface lexical cues. Consequently, models may misinterpret polarity cues or miss important syntactic relationships affecting sentiment detection.

Transformer-based architectures show greater resilience to linguistic variations thanks to their contextual representation mechanisms. However, even these models sometimes struggle with implicitly expressed sentiment, irony, or context-dependent emotional cues that demand deeper pragmatic understanding. These limitations suggest that sentiment analysis for Ukrainian-language content remains sensitive to linguistic nuances and contextual differences.

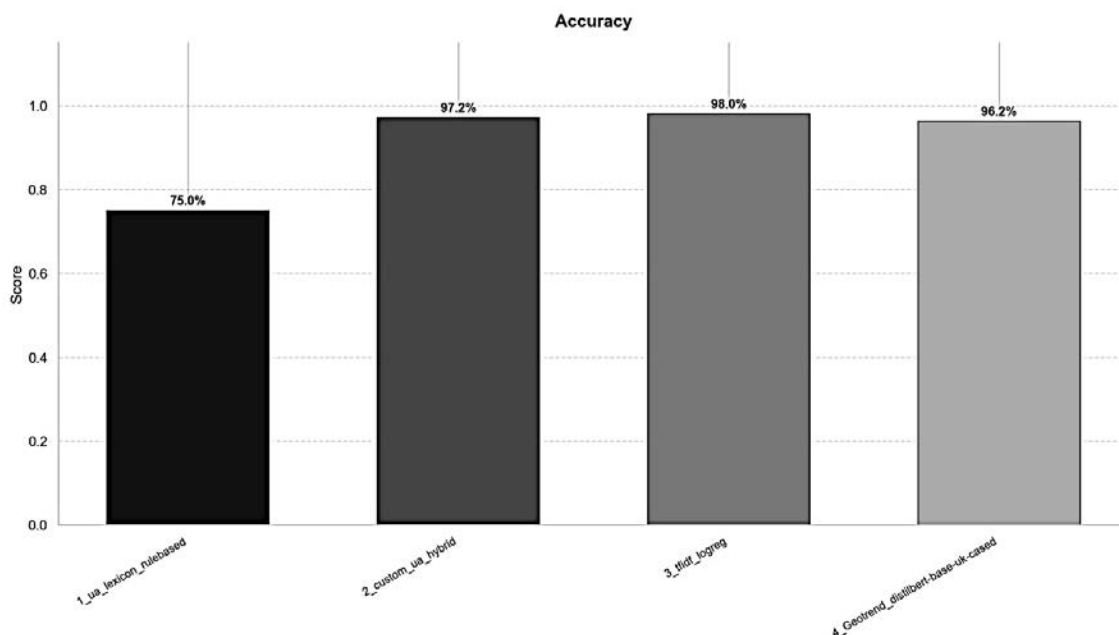


Fig. 3. Accuracy chart comparison

Overall, the results show that modern contextual language models significantly improve sentiment classification. However, additional gains are possible with hybrid modelling approaches, larger annotated datasets, and domain-specific language representations tailored to Ukrainian data.

Conclusions. This study offers a comprehensive comparison of rule-based, classical machine learning, and transformer-based approaches to Ukrainian-language sentiment analysis within a unified experimental framework. The proposed methodological design ensures consistent evaluation conditions and enables reliable comparisons across diverse modelling paradigms.

The results clearly show performance differences among the approaches and confirm that contextual representation models are effective for sentiment classification in Ukrainian texts. Transformer-based models demonstrate strong performance overall; however, classical machine learning models achieve the high-

est accuracy in this experimental setup. Meanwhile, classical machine learning methods achieve competitive results in controlled experiments – rule-based methods, though interpretable and resource-efficient, are less effective in complex linguistic cases.

The error analysis highlights persistent issues with linguistic ambiguity, implicit sentiment expression, and the structural variability of Ukrainian texts. These results indicate that sentiment classification accuracy remains affected by contextual and pragmatic factors, despite the use of advanced modelling techniques.

The findings support advancing Ukrainian sentiment analysis by providing a standardised evaluation framework and empirical data that endorse the application of contextual language models in real-world information systems. Future studies might aim to expand annotated corpora, improve domain adaptation strategies, and explore hybrid models that integrate linguistic expertise with contextual embeddings.

Bibliography:

1. Devlin J., Chang M.-W., Lee K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT 2019)*. Minneapolis, USA, 2019. P. 4171–4186. DOI: <https://doi.org/10.18653/v1/N19-1423>
2. Vasyliuk A., Basyuk T., Lytvyn V. Design and implementation of a Ukrainian-language educational platform for learning programming languages. *Proceedings of the 5th International Workshop on Modern Machine Learning Technologies and Data Science (MoMLet&DS 2023)*. Lviv, Ukraine, June 3, 2023. P. 406–420.
3. Basyuk T., Lomovatskyi A. Improving sentiment analysis of Ukrainian-language content by applying a rule-based approach. *CEUR Workshop Proceedings*. 2025. DOI: <https://doi.org/10.31110/COLINS/2025-1/012>
4. Reimers N., Gurevych I. Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2020. P. 4512–4525. DOI: <https://doi.org/10.18653/v1/2020.emnlp-main.365>
5. Stigall W., Al Hafiz Khan M. A., Attota D., et al. Large Language Models Performance Comparison of Emotion and Sentiment Classification. *Proceedings of the 2024 ACM Southeast Conference (ACMSE 2024)*. 2024. DOI: <https://doi.org/10.1145/3603287.3651183>
6. Touvron H., Lavril T., Izacard G. et al. LLaMA: Open and Efficient Foundation Language Models. *arXiv*. 2023. DOI: <https://doi.org/10.48550/arXiv.2302.13971>.
7. Joshi P., Santy S., Budhiraja A., Bali K., Choudhury M. The State and Fate of Linguistic Diversity and Inclusion in the NLP World. *Proceedings of ACL*. 2020. DOI: <https://doi.org/10.18653/v1/2020.acl-main.560>
8. Thapa S., Shiwakoti S., Shah S. B., et al. Large language models (LLM) in computational social science: prospects, current state, and challenges. *Social Network Analysis and Mining*. 2025. DOI: <https://doi.org/10.1007/s13278-025-01428-9>
9. Singh K., Dhawan S., Bali N., et al. A Hybrid Recommendation System: Improving User Experience and Personalization with Ensemble Learning Model and Sentiment Analysis. *International Journal of Computing and Digital Systems*. 2025. DOI: <https://doi.org/10.12785/ijcds/1571024952>
10. Tang Z., Huang C.-Y., Li T.-C., et al. Using Contextually Aligned Online Reviews to Measure LLMs' Performance Disparities Across Language Varieties. *Proceedings of the 2025 Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2025)*. 2025. DOI: <https://doi.org/10.18653/v1/2025.naacl-short.29>

Ломовацький А.А., Басюк Т.М. МЕТОДИ АНАЛІЗУ НАСТРОЇВ У КОНТЕНТІ УКРАЇНСЬКОЮ МОВОЮ: ПІДХОДИ НА ОСНОВІ ПРАВИЛ, МАШИННОГО НАВЧАННЯ ТА ТРАНСФОРМЕРІВ

Аналіз тональності є важливою складовою сучасної обробки природної мови, що забезпечує автоматизоване визначення емоційного забарвлення текстових даних у різних галузях застосування, зокрема в аналізі соціальних мереж, системах підтримки прийняття рішень, маркетинговій аналітиці та

моніторингу громадської думки. Проте аналіз тональності для морфологічно складних і малоресурсних мов, зокрема української, залишається складним завданням через обмеженість розмічених корпусів, мовну варіативність, гнучкий порядок слів і складну граматичну структуру. Це ускладнює безпосереднє застосування методів, розроблених для англійської та інших високоресурсних мов.

У статті подано систематизований огляд сучасних підходів до аналізу тональності українськомовного контенту, зосереджений на *rule-based* методах, класичних методах машинного навчання та трансформерних мовних моделях. Проаналізовано теоретичні засади зазначених підходів, їх переваги, обмеження та придатність до обробки українськомовних текстів в умовах обмежених ресурсів. Особливу увагу приділено *rule-based* методам як ефективному базовому рішенню для малоресурсних умов, моделям машинного навчання як масштабованим класифікаторам, а також трансформерним архітектурам як сучасним рішенням, що забезпечують контекстно залежні представлення тексту та здатні враховувати складні семантичні залежності.

У роботі наведено коротке порівняльне дослідження ефективності різних підходів на українськомовних даних із використанням стандартних метрик оцінювання, зокрема *accuracy*, *precision*, *recall* та *F1-score*. Отримані результати свідчать, що хоча *rule-based* методи залишаються ефективними в обмежених умовах і можуть використовуватися як базові моделі, трансформерні моделі демонструють вищу точність, кращу узагальнюваність і стійкість до лінгвістичної варіативності текстів. Результати дослідження сприяють систематизації підходів до аналізу тональності українськомовного контенту та окреслюють перспективи розвитку гібридних підходів у задачах обробки малоресурсних мов.

Ключові слова: аналіз тональності, українська мова, обробка природної мови, *rule-based* методи, машинне навчання, трансформерні моделі, малоресурсні мови.

Дата першого надходження статті до видання: 17.03.2026

Дата прийняття статті до друку після рецензування: 14.04.2026

Дата публікації (оприлюднення) статті 11.05.2026

Михалюк А.П.<https://orcid.org/0009-0004-7005-5864>

Національний університет харчових технологій

Міркевич Р.М.<https://orcid.org/0009-0004-2796-9388>

Національний університет харчових технологій

РОЗШИРЕНА КІЛЬКІСНА ОЦІНКА КІБЕРФІЗИЧНИХ РИЗИКІВ: ІНТЕГРАЦІЯ БЕЗПЕКИ, КОНТЕКСТУ ТА ОПЕРАЦІЙНИХ МОЖЛИВОСТЕЙ У СЕРЕДОВИЩАХ ПОТ ТА АВТОМАТИЗАЦІЇ

У цій науковій статті представлено новий підхід до кількісного аналізу рівнів загроз для критичної інфраструктури, заснований на інтеграції екологічного контексту та можливостей компонентів промислового Інтернету речей (ІоТ) та систем контролю та управління (І&С). Метою цього дослідження є вирішення «кризи метрик», що виникає через невід'ємний конфлікт між традиційною метрикою ІТ-безпеки. У той час як традиційні метрики ІТ-безпеки зосереджені на «тріаді ЦРУ» (Конфіденційність, Цілісність, Доступність) у певному порядку, у промислових системах управління (ІС) існує потреба змінити цей порядок, де безпека є головним пріоритетом, а потім доступність та цілісність. Автори цієї наукової статті проводять ретельний аналіз механіки нових екологічних метрик CVSS v3.1 та v4.0, включаючи модифікований вектор атаки (МАV), модифіковану складність атаки (МАС), модифіковані необхідні привілеї (МРR) та модифіковану область дії (MS). Одним з ключових внесків цього дослідження є розробка моделі «Оцінка кіберфізичних операційних ризиків» (CP-ORS). Модель CP-ORS – це математична конструкція, яка дозволяє проводити науковий розрахунок ризику шляхом перетворення статичних оцінок ризику в динамічні моделі, що синтезують фізичні пошкодження, цілісність процесу та експлуатаційну доцільність. Модель CP-ORS по суті перетворює показник «Безпека» (який є лише інформативним доповненням у системі CVSS v4.0) на основний критерій оцінювання. Модель CP-ORS враховує такі параметри, як «Час до відмови» (TiF) та «Час до усунення» (TiR), для вимірювання фізичної стійкості. Застосовність цього підходу доведена його здатністю коригувати бали серйозності відповідно до топології мережі, використовуючи модель Пердью. Це гарантує, що немає ні «втоми від оповіщення», ні критичних помилок у рішеннях щодо пріоритетів захисту, які розділяють вразливості, що є теоретично критичними, але операційно зменшені фізичними заходами пом'якшення. Отримані результати формують наукову основу для розробки нового покоління автоматизованих систем управління вразливостями, які можуть точно відображати реальний операційний ризик в Індустрії 4.0.

Ключові слова: кіберфізичні системи, ІоТ, системи управління процесами, оцінка ризиків, CVSS, метрики безпеки, критична інфраструктура.

Постановка проблеми. Прискорена конвергенція операційних технологій (ОТ) та інформаційних технологій (ІТ) у рамках парадигми промислового Інтернету речей (ІоТ) створила значну кризу в оцінці загроз кібербезпеці. Галузі критичної інфраструктури, починаючи від виробництва електроенергії, водоочищення, фармацевтичного виробництва та транспортної логістики, все більше покладаються на взаємопов'язані програмовані логічні контролери (ПЛК), системи керування процесами та збору даних (PCD) та розподі-

лені датчики ІоТ. Хоча така зв'язність забезпечує безпрецедентну продуктивність завдяки прогнозованому обслуговуванню та оркестрації в режимі реального часу, вона також усуває «повітряний зазор», який історично відділяв ці системи від цифрових загроз. [1]

Аналіз останніх досліджень та публікацій. Рецензовані дослідження визнають необхідність адаптації стандартних систем оцінювання до специфіки CFS. Основна увага приділяється переходу від статичних показників до «контекстуального

ризик» [3]. Сучасні стандарти, такі як CVSS v4.0, намагаються врахувати специфіку промислових систем за допомогою нових векторів оцінювання [4], але їх практичне впровадження в автоматизованих середовищах залишається фрагментарним. Дослідники також пропонують спеціалізовані метрики, такі як RVSS для роботизованих комплексів [5] та модифіковані підходи CVSS для промислових систем управління [6].

Постановка завдання. Метою дослідження є теоретичне обґрунтування та формалізація методу числової оцінки операційних загроз для середовищ ІоТ/АССТР. Концепція базується на трансформації фіксованих оцінок через призму змінених показників навколишнього середовища та впровадження системи CP-ORS. Наукова актуалізація полягає у створенні математичної схеми, що враховує архітектурну ізоляцію, запобіжні заходи відеоспостереження та послідовний вплив на фізичні процеси. [7].

Виклад основного матеріалу. Центральним викликом, з яким стикаються фахівці з безпеки, дослідники та менеджери операційних ризиків у цій галузі, є кількісна оцінка ризику. Традиційні показники кібербезпеки, розроблені переважно для ІТ-середовищ, де конфіденційність даних є першочерговою, не враховують кінетичну реальність кіберфізичних систем (КФС).

У цій статті представлено комплексний аналіз поточного стану метрик кібербезпеки для середовищ ІоТ та АСУТП. У ній критично розглядається механіка метрик CVSS «Екологічна» та «Модифікована» (MAV, MAC, MPR), які були розроблені для подолання розриву між теоретичною вразливістю та операційною реальністю.

Інтеграція застарілих систем контролю та управління із сучасними архітектурами ІоТ створює гетерогенну поверхню для атак, яку важко просто класифікувати. Типове промислове середовище зараз складається з дедалі поширенішої архітектури «моделі Purdue». Пристрої рівня 0 (процеси) та рівня 1 (управління), раніше доступні лише через послідовні кабелі, тепер часто підключаються до рівня 4 (бізнес-логістика) через шлюзи ІоТ для телеметрії та аналітики [1].

Цей структурний зсув робить статичну метрику вразливості непотрібною. Вразливість у певній реалізації стеку TCP/IP може бути оцінена як "критична" (CVSS 9.8) Національною базою даних вразливостей (NVD). У корпоративному ІТ-середовищі цей рейтинг є точним, оскільки сервер, ймовірно, підключений до Інтернету або глобальної інтрамережі. Однак, якщо той

самий стек TCP/IP знаходиться на ПЛК, розташованому за однонаправленим шлюзом безпеки (діодом даних) глибоко всередині нафтопереробного заводу, фактичний ризик набагато нижчий [6].

Розбіжність виникає через інверсію «тріади ЦРУ» (CIA Triad Confidentiality, Integrity, Availability). Безпека ІТ надає пріоритет конфіденційності, цілісності та доступності, зазвичай у такому порядку. Безпека ОТ вимагає, щоб безпека була на першому місці, а доступність, цілісність та конфіденційність (SAIC або AIC) – на другому. [4] Традиційні метрики не враховують «безпеку» як основну змінну, що призводить до небезпечного перекосу пріоритетів, коли команди безпеки виявляють високо оцінені, але незначні помилки програмного забезпечення, ігноруючи проблеми з низьким рейтингом, які можуть призвести до фізичного знищення.

Щоб усунути ці розбіжності, дослідники та лідери галузі прагнули вдосконалити методології оцінювання. Впровадження «модифікованих базових рівнів» у CVSS v3.1 та v4.0 є спробою дозволити кінцевим користувачам адаптувати оцінювання до свого середовища. [9] Ми прагнемо створити сувору теоретичну основу для нового покоління кіберфізичних метрик.

Механіка модифікації: аналіз показників навколишнього середовища CVSS. Загальна система оцінювання вразливостей (CVSS) є фактичним стандартом для оцінки ступеня вразливості. Однак її ефективність у сфері ІоТ/АС майже повністю залежить від правильного використання набору показників середовища. Ці показники дозволяють аналітикам варіювати базовий бал залежно від конкретного розгортання вразливого компонента. Незважаючи на їхню важливість, їх часто неправильно розуміють або ігнорують в автоматизованій звітності.

Теоретичні основи модифікованих метрик. Команда CVSS Environmental працює, розраховуючи базовий бал, використовуючи «модифіковані» версії базових метрик. Основна концепція полягає в тому, що «базові» метрики (вектор атаки, складність атаки, необхідні привілеї тощо) відображають вразливість у вакуумі (або в найгіршому випадку), тоді як «модифіковані» метрики відображають вразливість *in situ*.

Для інженера з управління технологічними процесами найважливішими модифікованими показниками є:

Модифікований вектор атаки (MAV): Цей показник відображає фізичну або логічну доступність пристрою в мережі.

Модифікована складність атаки (MAC): Це враховує заходи щодо захисту навколишнього середовища, які збільшують складність експлуатації.

Необхідні змінені привілеї (MPR): Це враховує рівні автентифікації, додані середовищем розгортання (наприклад, шлюзи, VPN).

Змінений діапазон (MS): Це визначає, чи може вразливість впливати на системи, що виходять за межі дії вразливого компонента (наприклад, НМІ порушує роботу ПЛК).

Модифіковані показники впливу (MC, MI, MA): Вони коригують показники впливу (конфіденційність, цілісність, доступність) залежно від критичності конкретного активу.

Математичний зв'язок між цими змінними визначає кінцевий бал впливу на навколишнє середовище. У CVSS версії 3.1 бал впливу на навколишнє середовище визначається шляхом підстановки модифікованого балу на базовий бал у стандартних рівняннях. Якщо модифікований бал встановлено на «Не визначено» (X), розрахунок за замовчуванням використовує базовий бал.[11]

Модифікований вектор атаки (MAV) у системах керування процесами.

Модифікований вектор атаки (MAV), мабуть, є найважливішою метрикою для оцінки ризику ПоТ та автоматизації. NVD призначає базовий вектор атаки (AV) на основі властивостей вразливості.

Мережа (N): Можна використовувати віддалено через стек протоколів.

Сусіди (A): Можна використовувати лише з одного й того ж домену широкомовлення (наприклад, локальної мережі, Bluetooth).

Місцевий (L): Потрібен доступ до оболонки або локальне виконання.

Фізичний (Ф): Вимагає фізичної взаємодії з пристроєм.

Якщо ПЛК має вразливість з AV:N (базовий), але розгорнутий у зоні Tier 1, яка має повітряний проміжок або доступна лише через однонаправлений шлюз, аналітик повинен встановити MAV:P (фізичний) або MAV:A (суміжний) залежно від ступеня ізоляції.

Зауважте, що MAV технічно може збільшити ризик, якщо середовище менш безпечне, ніж передбачав постачальник, хоча зазвичай це використовується для пом'якшення наслідків.

Модифікована складність атаки (MAC) слугує змінною для кількісної оцінки «компенсуючих елементів контролю». У середовищах АТ безпека часто досягається за рахунок різноманітності та невизначеності (наприклад, власницькі протоколи, нестандартні RTOS). Якщо вразливість залежить від стандартної поведінки фрагментації TCP/IP, але промисловий брандмауер суворо нормалізує всі пакети, перш ніж вони досягнуть ПЛК, складність успішної доставки корисного навантаження експлойту значно зростає. Аналітик повинен збільшити AC:L (низький) до MAC:H (високий).

Сценарій, у якому аналізується граф атаки з високим рівнем серйозності. Навіть за MAC: Високий, MPR: Високий та MUI: Жоден, загальний бал все одно можна збільшити, якщо всі показники впливу (конфіденційність, цілісність, доступність) високі. [12] Це підкреслює обмеження: збільшення складності зменшує ймовірнісний компонент балу, але не усуває впливовий компонент.

Потрібні змінені привілеї (MPR) релевантно для реалізацій RBAC. Багато застарілих протоколів ACS (наприклад, Modbus TCP) не мають автентифікації, тому PR:N є точним. Однак безпечні обгортки (наприклад, Secure Modbus або VPN-тунелі) встановлюють певний рівень автентифікації. Якщо зловмисник повинен спочатку скомпрометувати VPN-хаб, щоб дістатися до лінії

Таблиця 1

Вплив архітектури мережі на модифікований вектор атаки (MAV)

Контекст вразливості	Базовий антивірус	Реальність розгортання	Модифікований AV (MAV)	Вплив на оцінку (приблизно)
Веб-сервер ПЛК	Мережа (N)	Підключено до корпоративного шлюзу ПоТ	Мережа (N)	Без змін (високий ризик)
Веб-сервер ПЛК	Мережа (N)	Поза межами брандмауера рівня [3] (доступ через VPN)	Мережа (N)	Без змін (високий ризик)
Веб-сервер ПЛК	Мережа (N)	Однонаправлений шлюз (інформаційний діод)	Фізичний (П)	Різке падіння (-60%)
Переповнення буфера НМІ	Сусіди (A)	Ізольована локальна мережа диспетчерської (без зовнішнього доступу)	Сусіди (A)	Без змін
Переповнення буфера НМІ	Сусіди (A)	Віддалений доступ через стільниковий модем (ПоТ)	Мережа (N)	Різке зростання (+20%)

Джерело: [9].

Modbus, вразливість фактично змінюється з PR:N на MPR:H (що вимагає підвищених привілеїв на транспортному рівні мережі). [4]

Модифіковане опитування (MS) та ефект хвилі Концепція області дії (Score, Score) у CVSS версії 3.1 є складною, але життєво важливою для ПоТ. Вона вимірює, чи впливає вразливість в одному компоненті на ресурси, якими керує інший орган безпеки.

Основна сфера застосування: S:U (Без змін) або S:C (Змінено).

Змінений діапазон (MS): Дозволяє аналітиці змінювати це залежно від підключення до системи.

У ПоТ поширеним явищем є «ефект брижів». Вразливість у граничному шлюзі (компонент А) може дозволити зловмиснику надсилати неправильно сформовані команди до ПЛК (компонент В). Якщо базова лінія передбачає, що шлюз ізольований (S:U), але конкретне розгортання інтегрує шлюз безпосередньо в цикл керування, аналітик повинен встановити MS:C. [12] Ця «зміна обсягу» зазвичай діє як множник, збільшуючи оцінку, щоб відобразити каскадний ризик, типовий для кіберфізичних систем (CPS).

Науковий аналіз методології галузевого дослідження: контекстуальний ризик. Карміт Ядін, генеральний директор DeviceTotal, доктор філософії та визнаний експерт з кібербезпеки, конкуренції в бізнесі та управління технічними ризиками, сформулювала філософію оцінки ризиків, яка суттєво відрізняється від статичної моделі NVD/CVSS. [19] Хоча її робота поширюється переважно через високорівневі галузеві публікації, подкасти та операційну архітектуру платформи DeviceTotal, а не через традиційні рецензовані академічні журнали, вона являє собою сувору, емпірично обґрунтовану методологію для «контекстуального ризику» в епоху ПоТ.

Її критика висвітлює три основні недоліки поточної ситуації:

1. Здатність проти серйозності: Галузь плутає серйозність (потенційну шкоду) з ризиком (ймовірністю). [8]

2. ДНК пристрою: Два пристрої з однаковою версією операційної системи можуть мати суттєво різні профілі ризику залежно від їхньої прошивки, активних служб та апаратних компонентів. [15]

3. «Сліпа зона» некерованих пристроїв: Традиційні сканери вразливостей вимагають агентів або активного сканування, що може призвести до збоїв чутливого обладнання ОТ. [16]

Структура «Контекстного ризику». Ця методологія, реалізована за допомогою DeviceTotal, пропонує

контекстну оцінку ризику, яку можна аналізувати як багатовимірну функцію, що включає змінні, які зазвичай ігноруються стандартним CVSS.

Ключова змінна - Операційні можливості. Цей підхід поєднує дані з Каталогу відомих вразливостей, що можуть бути експлуатовані (KEV), CISA та інших джерел інформації про загрози.

Теоретично: Вразливість існує в коді.

Зброя: Існує код експлойту для підтвердження концепції (PoC).

Контекстуально шкідливий: Конфігурація пристрою виявляє вектор вразливості.

Ключова змінна: видимість поверхні атаки.

Методологія розраховує «оцінку поверхні атаки» для кожного пристрою.

Математично це відповідає модифікованому вектору атаки (MAV) CVSS, але визначається динамічно за допомогою «безагентного» відбитка пальців, а не вручну аналітиками.

Безагентний підхід до «науки про дані». Альтернатива Data Science передбачає створення універсального репозиторію безпеки пристроїв. [7] Попередньо аналізуючи характеристики прошивки та апаратного забезпечення тисяч типів пристроїв, система може визначити рівень ризику конкретного пристрою, просто визначивши його марку, модель та версію прошивки, без необхідності активного дослідження.

Наукове виведення алгоритму:

На основі описів, прогнозований алгоритм для цього контекстуального ризику можна змоделювати наступним чином:

$$R_{context} = f(V_{severity}, E_{probability}, C_{criticality}, S_{surface})$$

де:

$V_{severity}$ – це внутрішній бал вразливості (базовий бал CVSS).

$E_{probability}$ – ймовірність експлуатації (отримана з CISA KEV, Exploit-DB).

$C_{criticality}$ – важливість активу для організації (бізнес-контекст).

$S_{surface}$ – це рівень доступності пристрою (досяжність мережі).

Ця модель надає пріоритет «швидким перемогам» для зловмисників – вразливостям, які слугують основою – незалежно від їхнього загального рівня серйозності. [8]

Розрив у безпеці: фізичний вплив на оцінку ризику. Хоча CVSS версії 4.0 та ця контекстна модель ризику враховують ймовірність та цифровий вплив атак, вони історично мали труднощі з кількісною оцінкою фізичного впливу, властивого автоматизації та ПоТ.

CVSS версії 4.0 та додатковий бал безпеки. Наприкінці 2023 року Форум команд реагування на інциденти та безпеки (FIRST) випустив CVSS версії 4.0, яка включає спеціальну метрику безпеки (S) як частину нової додаткової групи метрик.[15]

Категорія безпеки CVSS версії 4.0 (S):

Не вказано (X): Атрибут не було оцінено.

Незначна (N): Жодного впливу на безпеку.

Присутній (P): Наслідки відповідають визначенню IEC 61508 як «граничні», «критичні» або «катастрофічні». [10]

Критичний аналіз:

Хоча включення поняття «безпека» є важливою віхою, його впровадження як додаткової метрики є суттєвим обмеженням. Додаткові метрики у CVSS версії 4.0 не змінюють кінцевий числовий бал (0-10). Вони призначені виключно для надання додаткового контексту споживачеві.

Це означає, що вразливість в ICS може мати оцінку CVSS 5.0 (середній рівень), але бути позначеною як «Безпека: присутня». Автоматизовані системи, які фільтрують за шкалою «Оцінка > 7.0», пропускають цю потенційно небезпечну для життя вразливість. Такий підхід «маркування» не дозволяє математично інтегрувати безпеку в логіку пріоритезації ризиків, залишаючи небезпечну прогалину для автоматизованих систем виправлення.

Альтернативні структури: RVSS та IVSS. Щоб заповнити цю прогалину, у науковій літературі запропоновано варіанти адаптації CVSS до конкретних предметних галузей.

Система оцінки вразливостей роботів (RVSS):

Запропонований Вілчесом та ін., RVSS явно розширює CVSS для обробки роботизованої безпеки.[18] Він вводить метрику впливу на безпеку (S), яка впливає на кінцевий бал, на відміну від додаткової метрики в CVSS версії 4.0.

Ключова інновація: Це додає до вектора впливу (разом із ЦПУ) компонент «безпеки».

Модифікований захист (MS): Дозволяє враховувати вплив на безпеку з екологічної точки зору на основі фізичних бар'єрів (наприклад, робот у клітці має менший ризик для безпеки, ніж колаборативний робот).

Система оцінки промислової вразливості (IVSS):

IVSS розроблено як «індустріалізована альтернатива» CVSS.4

Фокус: Це знижує пріоритет конфіденційності (часто неактуальний в АО) та ставить більший акцент на доступність та цілісність.

Методологія: Він коригує вагу показників навколишнього середовища, щоб відобразити той факт, що «перезавантаження» пристрою I&K не є стандартним виправленням.

Кількісна оцінка фізичного впливу. Нещодавнє дослідження показників «фізичного впливу» для CPS пропонує змінні, які виходять за рамки простих позначок. Цей підхід наголошує на кількісному підході з використанням [16]:

Час до відмови (TtF): Тривалість часу від кібератаки до прояву небезпечних фізичних умов.

Час до вирішення проблеми (TtR): Час, необхідний операторам для виявлення порушення та виконання ручного втручання.

Умова стабільності:

$$Risk \propto \frac{1}{TtF - TtR}$$

Якщо система аварійно завершує роботу, перш ніж користувачі встигають її зберегти, це створює умову гонки, яку статичні метрики CVSS не можуть зафіксувати. $TtF < TtR$

Пропозиція: Оцінка кіберфізичних операційних ризиків (CP-ORS)

Враховуючи недоліки, виявлені у CVSS версії 4.0 (лише додаткова інформація щодо безпеки), та інтегруючи філософію «контекстуального ризику» галузевих досліджень (придатність пристроїв та ДНК), у цьому звіті пропонується єдина модель оцінки: Оцінка кіберфізичних операційних ризиків (CP-ORS).

Ця модель забезпечує сувору, математично визначену основу для розрахунку ризику в середовищах IoT/ICS. Вона ефективно «модифікує» підхід CVSS, перетворюючи додаткову мітку безпеки на основний фактор оцінки та інтегруючи змінні фізичної стійкості.

Математична структура CP-ORS

CP-ORS надає оцінку за шкалою від 0 до 100, яка надає детальнішу інформацію, ніж шкала від 0 до 10.

$$CP-ORS = \min\left(100, \left[(V_{exploit} \times I_{physical} \times M_{resilience}) + \beta_{safety}\right]\right)$$

де:

Вектор контекстної експлуатаційної здатності (похідний від цієї логіки + CVSS MAV).

$I_{physical}$: Фізичний коефіцієнт удару (отриманий з RVSS/IEC 61508).

$M_{resilience}$: Пом'якшувальний коефіцієнт стабільності (отриманий з динаміки TtF/TtR).

β_{safety} : Зведення ключових показників безпеки.

Заголовки та визначення компонентів

Вектор контекстуальної експлуатації $V_{exploit}$

Ця змінна синтезує технічну серйозність вразливості з її контекстуальним охопленням та статусом зброї. Вона ефективно втілює критику галузевих досліджень на кшталт «серйозності проти можливості використання».

Значення розраховується наступним чином:

$$V_{exploit} = Base\ Severity \times Exploit\ Status \times MAV\ Coefficient$$

Таблиця 2

Коефіцієнти умов експлуатації E_{stat}

Статус	Визначення	Коефіцієнт	Логіка джерела
Озброєний	Експлоїт існує в CISA KEV або Metasploit.	1.5	Високо-пріоритетна «Точка підтримки» [8]
Опубліковано PoC	Доступний код для підтвердження концепції (GitHub).	1.0	Стандартний ризик
Теоретичний	Публічний експлоїт-код невідомий.	0,5	Низька терміновість

Таблиця 3

Коефіцієнти модифікованого вектора атаки (MAV)

Налаштування MAV	Визначення	Коефіцієнт	Логіка джерела
Ланцюг	Доступно через маршрутизовану мережу IoT/корпоративну мережу.	1.0	Повна експозиція
Сусіди	Доступно лише через OT VLAN/зону.	0,6	Сегментований [9]
Фізичний	З повітряним зазором або вимагають локальної взаємодії.	0,1	Ізольований

Фізичний фактор впливу ($I_{physical}$)

Ця змінна кількісно визначає кінетичні наслідки на основі рівнів цілісності безпеки (SIL) стандарту IEC 61508 та визначень, що містяться в ньому. [18]

Таблиця 4

Рубрика фізичного впливу

Рівень	Значення	Визначення
Катастрофічно	10.0	Втрата життя, незворотна шкода навколишньому середовищу, повна втрата активів.
Критичний	7.5	Значна шкода (підлягає відновленню), значний викид у навколишнє середовище, значні виробничі втрати.
Маргінальний	4.0	Незначні травми, обмежені пошкодження, втрата ефективності виробництва.
Неповнолітні	1.0	Без фізичних пошкоджень; лише втрата даних або незначний простой.

Коефіцієнт стійкості до пом'якшення ($M_{resilience}$)

Це співвідношення винагороджує системи, розроблені з урахуванням принципів «безпеки за проектом», особливо ті, що відокремлюють функцію безпеки від кіберактиву. Це включає логіку проти [16]. $TiFTiR$

Таблиця 5

Рубрика сталого розвитку

Тип опору	Значення	Визначення
Пов'язано (немає)	1.0	Кіберактив безпосередньо контролює фізичний; незалежного захисту від збоїв немає. $TiF \rightarrow 0$
Людина в циклі	0,7	Сигналізація забезпечує достатньо часу для ручного втручання.. $TiR < TiF$
Механічна безпека	0,2	Фізичні блокування (наприклад, розрив мембран, падіння під дією сили тяжіння) запобігають катастрофі незалежно від стану ПЛК.

Суматор основних показників безпеки (β_{safety})

Щоб гарантувати, що ризик вразливостей у критично важливих для безпеки пристроях ніколи не паде до «нуля» (ключове питання в цьому аналізі критичного ланцюга «низького рівня серйозності»), додається фіксоване значення в балах залежно від ролі пристрою. Пристрій SIL 3/4:+20 балів. Пристрій SIL 1/2:+10 балів. Незабезпечений актив: +0 балів.

Порівняння тематичного дослідження: ПЛК для дозування хлору

Щоб продемонструвати ефективність CP-ORS порівняно зі стандартними CVSS v3.1 та v4.0, розглянемо критичну вразливість віддаленого виконання коду (RCE) у ПЛК, який керує дозуванням хлору на водоочисній станції.

Сценарій: ПЛК має критичний RCE (базовий бал 9,8). Він підключений до захищеної OT-мережі (сусіди). Код експлоїта є загальнодоступним.

Конфігурація А (Небезпечна): Немає механічних обмежувачів; програмне забезпечення ПЛК контролює максимальний потік.

Конфігурація В (стійка): Механічний обмежувач потоку фізично запобігає передозуванню.

Порівняльний аналіз оцінювання

Метрична модель	Конфігурація А (Небезпечна)	Конфігурація В (стійка)	Аналіз
Базовий рівень CVSS 3.1	9.8 (Критично)	9.8 (Критично)	Не розрізняє фізичний ризик.
Версія середовища CVSS 3.1	8.1 (Високий)(MAV:A)	8.1 (Високий)(MAV:A)	Враховує мережу, ігнорує сталий розвиток.
CVSS версії 4.0	8,5 (Високий)(Безпека :P)	8,5 (Високий)(Безпека :P)	Є мітка безпеки, але клас ідентичний.
Запропонований CP-ORS	95 (Критично)	31 (низький)	Правильно визначає операційний ризик.

Розрахунок CP-ORS (конфігурація А):

$$V_{exploit} : (Base9.8 scaled) \times 1.5(Weaponized) \times 0.6(Adjacent) \approx 8.8$$

$$I_{physical} : 10.0 \text{ (Катастрофічний)}$$

$$M_{resilience} : 1.0 \text{ (Немає)}$$

$$\beta_{safety} : +20$$

Рейтинг: $(8.8 \times 10.0 \times 1.0) + 20 = 108 \rightarrow 100$

Розрахунок CP-ORS (конфігурація В):

$$V_{exploit} : 8.8$$

$$I_{physical} : 10.0$$

$$M_{resilience} : 0,2 \text{ (механічна стійкість до пошкоджень)}$$

$$\beta_{safety} : +20$$

Рейтинг: $(8.8 \times 10.0 \times 0.2) + 20 = 17.6 + 20 = 37.6$

Модель CP-ORS успішно деескалює ризик у конфігурації В, відображаючи той факт, що механічна негативна безпека робить кібератаку нездатною спричинити катастрофічні наслідки. Це запобігає «втомі від тривоги» та дозволяє зосередити ресурси на конфігурації А.

Операційні стратегії та їх реалізація. Впровадження моделі CP-ORS або навіть покращених модифікованих метрик CVSS вимагає переходу від пасивного сканування вразливостей до активного «управління кіберфізичними активами».

Вимоги до збору даних. Щоб автоматично розрахувати ці оцінки, екосистема безпеки повинна обробити три окремі потоки даних:

1. Розвідка кіберзагроз (СТІ): Отримує дані про зрілість експлойту в режимі реального часу (наприклад, CISA KEV) для заповнення змінної. Це узгоджується з акцентом на "вплив".

2. Карта топології мережі: Динамічне відображення розташування активів (рівень 2 проти рівня 4) для заповнення коефіцієнта MAV.

3. Аналіз технологічних ризиків (TRA): Отримання інженерних даних (дослідження HAZOP) для визначення рівня цілісності безпеки (SIL), фізичного впливу () та стабільності (). Це найскладніші дані для оцифрування, але вони є важливими для справжньої безпеки $I_{physical} M_{resilience}$.

Безагентна оцінка та машинне навчання. З огляду на обмеження ПЛК, «безагентний» підхід, запропонований DeviceTotal, є єдиним життєздатним шляхом для широкого розгортання.

Аналізуючи мережевий трафік (порти SPAN) та файли конфігурації (файли проекту), а не запитуючи операційну систему пристрою, система може отримати відбиток «ДНК» пристрою, щоб знайти його профіль безпеки в репозиторії.

Крім того, моделі машинного навчання (зокрема, багатошарові перцептрони) показали багатообіцяючі результати у виявленні атак та прогнозуванні їхнього впливу шляхом вивчення «фізики» системи автоматизації.²⁶ У майбутніх ітераціях CP-ORS машинне навчання (ML) може бути використане для динамічного коригування коефіцієнта на основі стабільності процесу в режимі реального часу – якщо система виявляє, що процес працює поблизу запасу безпеки, показник стабільності знижується, а показник кіберризiku різко збільшується $M_{resilience}$.

Висновки. Метрики кібербезпеки, що зараз використовуються в середовищах IIoT та ICS, є пережитком епохи IT, небезпечно відірваними від фізичної реальності критичної інфраструктури. Хоча розробка CVSS версії 4.0 та «модифікованих» показників навколишнього середовища забезпечують інструменти для кращої оцінки, вони рідко використовуються на повну потужність через складність та брак контекстних даних.

Аналіз методології галузевих досліджень показує критично важливий шлях уперед: ризик має бути контекстуальним, орієнтованим на зручність використання та адаптованим до ДНК конкретного пристрою. Однак навіть цей підхід має бути доповнений ретельною кількісною оцінкою фізичної безпеки та стійкості.

Запропонована структура CP-ORS об'єднує ці області. Математично поєднуючи кіберймовірність (операційну ймовірність/MAV) з фізичними наслідками (вплив/стійкість), вона пропонує метрику, яка відображає справжній операційний ризик. В епоху конвергентної інфраструктури метрика, яка не враховує запобіжний клапан або систему безпеки, є не просто неточною, вона небезпечна. Перехід до кіберфізичної оцінки – це не просто академічна вправа; це операційний імператив для безпеки Індустрії 4.0.

Список літератури:

1. Barrere M., Hankin C., Nicolaou N., Eliades D., Parisini T. Measuring cyber-physical security in industrial control systems via minimum-effort attack strategies. *Journal of Information Security and Applications*. DOI: 10.1016/j.jisa.2020.102471.
2. Massaro V., Capacci L., Montanari R. Towards Context-Aware Risk Assessment Scoring System for IoT/IIoT Devices. *Матеріали семінару CEUR*. URL: <https://ceur-ws.org/Vol-3488/paper25.pdf> (дата звернення: 17.12.2025).
3. Aftabi N., Li D., Sharkey T. An Integrated Cyber-Physical Risk Assessment Framework for Worst-Case Attacks in Industrial Control Systems. URL: <https://arxiv.org/html/2304.07363v5> (дата звернення: 17.12.2025).
4. Gori G., Rinieri L., Melis A., Al Sadi A., Callegati F., Prandini M. A Systematic Analysis of Security Metrics for Industrial Cyber-Physical Systems. DOI: 10.3390/electronics13071208.
5. Moraitis G., Nikolopoulos D., Bouziotas D., Lykou A., Karavokiros G., Makropoulos C. Quantifying Failure for Critical Water Infrastructures under Cyber-Physical Threats. *Journal of Environmental Engineering*. DOI: 10.1061/(ASCE)EE.1943-7870.00017.
6. Collier Z., Panwar M., Ganin A., Kott A., Linkov I. Security Metrics in Industrial Control Systems. URL: <https://arxiv.org/pdf/1512.08515> (дата звернення: 17.12.2025).
7. Yoon S., Kim D., Kim K., Euom I. Vulnerability Exploitation Risk Assessment Based on Offensive Security Approach. DOI: 10.3390/app132212180.
8. Oser P., Heijden R., Luders S., Kargl F. Risk Prediction of IoT Devices Based on Vulnerability Analysis. *ACM Transactions on Internet Technology*. 2022. Vol.25, №14. P. 1-36. DOI: 10.1145/3510360.
9. Common Vulnerability Scoring System v4.0: Examples. URL: <https://www.first.org/cvss/examples> (дата звернення: 17.12.2025).
10. Common Vulnerability Scoring System version 4.0: Specification Document. URL: <https://www.first.org/cvss/specification-document> (дата звернення: 17.12.2025).
11. Ibrahim M., Elhafiz R. Security Assessment of Industrial Control System Applying Reinforcement Learning. DOI: 10.3390/pr12040801.
12. Cheimonidis P., Rantos K. Dynamic Vulnerability Criticality Calculator for Industrial Control Systems. URL: <https://arxiv.org/html/2404.16854v1> (дата звернення: 17.12.2025).
13. Lyu X., Ding Y., Yang S. Safety and security risk assessment in cyber-physical systems. *IET Cyber-Physical Systems: Theory & Applications*. Vol.4, №3. P. 221-232. DOI: 10.1049/iet-cps.2018.5068.
14. Vilches V., Gil-Urriarte E., Ugarte I., Mendia G., Pisón R., Kirschgens L., Calvo A., Cordero A., Apa L., Cerrudo C. Towards an open standard for assessing the severity of robot security vulnerabilities, the Robot Vulnerability Scoring System (RVSS). DOI: 10.48550/arXiv.1807.10357.
15. Wen H. Vulnerability Assessment of Industrial Control System with an Improved CVSS. DOI: 10.48550/arXiv.2306.08631.
16. Keenan C., Maier H., Delden H., Zecchin A. Bridging the Cyber-Physical Divide: A Novel Approach for Quantifying and Visualising the Cyber Risk of Physical Assets. DOI: 10.3390/w16050637.
17. Okur C., Dener M. Symmetrical Resilience: Detection of Cyberattacks for SCADA Systems Used in IIoT in Big Data Environments. DOI: 10.3390/sym17040480.
18. Do V., Fillatre L., Nikiforov I. Sequential monitoring of SCADA systems against cyber/physical attacks. *IFAC-PapersOnLine (Elsevier)*. 2015. Vol.48, №21. P. 746-753. DOI: 10.1016/j.ifacol.2015.09.616.
19. Yadin C. Secure ICS and OT for Industry 4.0. URL: <https://devicetotal.com/secure-ics-and-ot-for-industry-4-0/> (дата звернення: 17.12.2025).

Mykhalyuk A. P., Mirkevich R. M. ADVANCED CYBER-PHYSICAL RISK QUANTIFICATION: INTEGRATING SECURITY, CONTEXT AND OPERATIONAL CAPABILITIES IN IIOT AND AUTOMATION ENVIRONMENTS

This paper presents a new approach to quantifying threat levels for critical infrastructure based on the integration of environmental context and capabilities of Industrial Internet of Things (IIoT) and I&C components. The purpose of this study is to address the “metrics crisis” that arises from the inherent conflict between traditional IT security metrics. While traditional IT security metrics focus on the “CIA triad” (Confidentiality, Integrity, Availability) in a specific order, in industrial control systems (ICS) there is a need to reverse this order, where security is the top priority, followed by availability and integrity. The authors of this paper conduct a thorough analysis of the mechanics of the new CVSS v3.1 and v4.0 environmental metrics, including Modified Attack Vector (MAV), Modified Attack Complexity (MAC), Modified Required Privileges (MPR), and Modified Scope of Action (MS). One of the key contributions of this study is the development of the Cyber-Physical Operational Risk Assessment (CP-ORS) model. The CP-ORS model is a mathematical construct that enables scientific risk calculation by transforming static risk assessments into dynamic models that synthesize physical damage, process integrity, and operational feasibility. The CP-ORS model essentially transforms the Security metric (which is only an informative addition to the CVSS v4.0 system) into a primary evaluation criterion. The CP-ORS model takes into account parameters such as Time to Failure (TiF) and Time to Repair (TiR) to measure physical resilience. The applicability of this approach is demonstrated by its ability to adjust severity scores according to network topology using the Purdue model. This ensures that there is neither “alert fatigue” nor critical errors in protection priority decisions that separate vulnerabilities that are theoretically critical but operationally mitigated by physical mitigation measures. The results form the scientific basis for the development of a new generation of automated vulnerability management systems that can accurately reflect real operational risk in Industry 4.0.

Keywords: robotic systems, heterogeneous systems, automation systems, cyber-physical systems, monitoring, information security.

Дата першого надходження статті до видання: 16.03.2026
 Дата прийняття статті до друку після рецензування: 13.04.2026
 Дата публікації (оприлюднення) статті 11.05.2026

УДК 004.94:004.7:621.39

DOI DOI DOI <https://doi.org/10.32782/2663-5941/2026.2.1/27>**Міхав В.В.**<https://orcid.org/0000-0003-4816-4680>

Приватна установа «Університет науки, підприємництва та технологій»

Мелешко Є.В.<https://orcid.org/0000-0001-8791-0063>

Центральноукраїнський національний технічний університет

Якименко М.С.<https://orcid.org/0000-0003-3290-6088>

Центральноукраїнський національний технічний університет

Босько В.В.<https://orcid.org/0000-0002-4933-9676>

Центральноукраїнський національний технічний університет

Березюк І.А.<https://orcid.org/0000-0003-1903-8204>

Центральноукраїнський національний технічний університет

МОДЕЛІ ДЕЦЕНТРАЛІЗОВАНИХ Q&A СИСТЕМ В PEER-TO-PEER КОМП'ЮТЕРНИХ МЕРЕЖАХ

У статті виконано дослідження та порівняльний аналіз моделей децентралізованих Q&A систем, що функціонують у P2P комп'ютерних мережах. Актуальність дослідження зумовлена обмеженнями централізованих платформ питань-відповідей – залежністю від єдиного оператора та його політик модерації й ранжування, ризиками цензури або втрати доступу до накопичених знань, а також слабкою переносимістю даних і репутації користувачів між сервісами. Метою статті є узагальнення наявних підходів до побудови децентралізованого Q&A та визначення їхніх переваг, недоліків і мережевих компромісів у контексті забезпечення якості колективного відбору відповідей. У роботі запропоновано узагальнену класифікацію децентралізованих Q&A-рішень за архітектурними ознаками та механізмами побудови й проаналізовано репрезентативні моделі та прототипи, описані в наукових публікаціях. Показано, що значна частина підходів базується на соціальному пошуку на базі P2P-мереж та децентралізованій маршрутизації запитів до потенційних експертів, де кероване поширення питання дає змогу знижувати мережевий трафік і скорочувати час отримання релевантних відповідей. Інший поширений напрям пов'язаний із гібридними рішеннями типу «blockchain + P2P-сховище», у яких логіка довіри та правил взаємодії (голосування, репутація, облік внеску, аудит подій) формалізується смарт-контрактами, тоді як контент розміщується у розподіленому середовищі, що підвищує перевірюваність дій і прозорість процедур. Окрему увагу приділено мережевим компромісам, характерним для P2P-середовища, зокрема, впливу динаміки учасників P2P-мережі на доступність експертів і контенту, витратам трафіку під час поширення запитів і затримкам узгодження станів, а також координаційним витратам та аспектам приватності у гібридних blockchain-схемах. За результатами огляду сформовано порівняльну таблицю моделей за критеріями придатності до реалізації задач Q&A систем та узагальнено висновки щодо умов ефективного проектування децентралізованих систем питань-відповідей у P2P комп'ютерних мережах.

Ключові слова: Q&A, системи запитань-відповідей, децентралізовані системи, комп'ютерні мережі, P2P-мережі, пошук експертів, маршрутизація запитів, репутація, довіра, blockchain, колективний відбір відповідей, колективний розум.

Постановка проблеми. Сучасні системи запитань-відповідей Q&A типу Stack Overflow або Reddit забезпечують накопичення знань і колективний відбір корисних відповідей завдяки механізмам голосування, репутації та модерації. Однак у переважній більшості такі сервіси реалізуються в централізованій архітектурі, що породжує низку обмежень: наявність єдиної точки відмови та контролю, залежність доступності знань від працездатності центральної інфраструктури, а також централізоване управління політиками ранжування й модерації. У результаті виникають ризики недоступності даних, обмеження щодо приватності, а також вразливість до маніпулятивного впливу на видимість контенту та експертів. Важливість цих недоліків залежить від тематики системи Q&A, зокрема, якщо в технічних системах пошуку відповідей маніпуляція громадською думкою не має особливого сенсу, то в соціальних науках та новинах маніпулятивні впливи й просування суб'єктивних точок зору – часті сценарії. Додатково, зі зростанням обсягів контенту та кількості учасників загострюється проблема маршрутизації питань до релевантних експертів і їх автоматизованого підбору – значна частка запитань може залишатися без якісної відповіді або отримувати відповіді із затримкою через невідповідність між тематикою запиту та компетенціями й активністю потенційних експертів. У науковій літературі ці задачі формалізуються як *expert finding* та *expert search* – пошук експертів за тематикою запиту [1], а також *question routing* – спрямування питання до групи користувачів, здатних і готових надати корисні відповіді [2]. У цьому контексті перехід до децентралізованих P2P-мереж для Q&A розглядається як спосіб зменшити залежність від єдиного центру, підвищити відмовостійкість і забезпечити доступність знань у середовищах із нестабільною інфраструктурою. Технічною основою такої децентралізації виступають P2P-оверлеї та механізми розподіленого пошуку, зокрема, DHT-таблиці, які забезпечують ефективний пошук контенту без центрального серверу, що ілюструється класичними рішеннями, наприклад, протоколом Chord [3]. Водночас перенесення Q&A у P2P-мережу породжує нову науково-практичну суперечність – потрібно зберегти властивості «колективного інтелекту», зокрема, якість відбору відповідей і експертів, за відсутності централізованого контролю, глобальних логів та єдиної репутаційної системи. У децентралізованому середовищі ускладнюється довіра до профілів користувачів і зростає ризик Sybil-

поведінки – створення множини підконтрольних профілів, що здійснюють скоординовані дії та атаки ін'єкцією профілів на систему «колективного розуму» [4]. Тому необхідним стає розробка і впровадження розподілених механізмів довіри й репутації, зокрема, класу алгоритмів типу EigenTrust, які обчислюють глобальні показники довіри у P2P-мережі [5], а також механізмів узгодження стану системи (оцінки, репутаційні події, версії контенту) в умовах децентралізації, для цього релевантними є підходи *eventual consistency* та *CRDT* [6]. Отже задача дослідження та розробки моделей децентралізованих Q&A-систем у P2P комп'ютерних мережах, які забезпечують пошук релевантних експертів, рекомендацію експертів на основі колаборативної фільтрації, контентних та поведінкових сигналів, а також стійкості до маніпуляцій і Sybil-атак та узгодження стану колективного відбору в децентралізованому режимі, є актуальною.

Аналіз останніх досліджень і публікацій. Емпіричні дослідження спільнот сфокусованих на Q&A системах показують, що цінність таких систем формується через колективну активність, репутаційні механізми та накопичення знань, але централізоване управління цими механізмами створює структурну залежність від платформи як інституції [7].

З технологічного погляду, базовим підґрунтям децентралізованих Q&A є результати досліджень у сфері P2P-оверлеїв та розподіленого пошуку. Структуровані оверлеї на основі засобів DHT пропонують формальні гарантії масштабованості для операцій пошуку та маршрутизації без центрального індексу. У цьому контексті протокол Chord задає один із канонічних варіантів організації простору ключів та маршрутизації з логарифмічною складністю, що робить його придатним як «транспортний шар» для виявлення вузлів-носіїв контенту або індексів за темами й тегами [3]. Протокол Kademlia, у свою чергу, формалізує засоби DHT на XOR-метриці та демонструє високу практичну придатність до сценаріїв зі змінною кількістю вузлів P2P-мережі і частковою їх недоступністю, що критично для Q&A-мереж у реальних умовах [8]. Таким чином, у літературі сформовано класичний інструментарій для побудови маршрутизації запитів, однак питання семантичного пошуку, ранжування та оцінювання якості відповідей виходять за межі функцій DHT і потребують надбудов.

Окремий напрям досліджень стосується децентралізованого зберігання та доставки контенту.

Протокол IPFS реалізує контент-адресований P2P-шар з версіонуванням, де цілісність та ідентифікація об'єктів забезпечуються хеш-адресацією, а реплікація між вузлами здійснюється без необхідності центрального сховища [9]. Водночас у протокольних екосистемах, орієнтованих на локальні журнали та реплікацію, зокрема, Secure Scuttlebutt, закладено альтернативну модель «джерела істини», де кожен учасник веде підписаний append-only журнал, а синхронізація відбувається P2P-шляхом за принципом соціально-обумовленої реплікації [10]. Для Q&A це відкриває можливість локального накопичення історії внеску та підписаних фактів взаємодій (питання, відповіді, оцінки), однак ускладнює глобальний пошук і узгоджене ранжування без додаткових механізмів індексації.

Паралельно у сфері децентралізованих соціальних протоколів важливим є федеративний підхід, де децентралізація досягається не чистою P2P-мережею, а мережею незалежних серверів. ActivityPub стандартизує обмін об'єктами та подіями між учасниками, забезпечуючи інтероперабельність і часткову автономність доменів [11]. Для Q&A федералізація є компромісом, з одного боку, вона знижує монополізацію й дозволяє локальну модерацію; з іншого – залишає «локальні центри» та ризик фрагментації контенту й репутації між учасниками мережі.

Ключовою проблемою, що систематично фіксується в дослідженнях децентралізованих мереж, є довіра до профілів користувачів та стійкість до маніпуляцій. Класична робота про Sybil-атаки формалізує загрозу множинних підконтрольних профілів як фундаментальний виклик для розподілених систем репутації та голосування [4]. У відповідь на це запропоновано алгоритмічні механізми обчислення довіри, зокрема, EigenTrust, який оцінює глобальні рейтинги на основі локальних взаємодій та репутаційних переходів у P2P-мережі [5]. Для децентралізованих Q&A такі механізми є критичними, оскільки саме «колективні сигнали» (оцінки, репутація, історія корисних відповідей) мають бути захищені від накрутки й скоординованих атак. Додатково, у децентралізованих архітектурах необхідно вирішувати узгодження стану (версійність, редагування, агрегація голосів та оцінок) за відсутності центра. У цьому контексті структури даних CRDT задають формальний апарат для досягнення виконання моделі Eventual consistency без конфліктів на рівні типів даних, що забезпечує придатність до офлайн-сценаріїв та асинхронної реплікації [6].

Власне алгоритмічна складова Q&A систем, підбір експертів, їх рекомендація, маршрутизація питань та оцінювання якості відповіді, добре опрацьована для централізованих середовищ. Оглядові роботи з рекомендації експертів систематизують підходи, що поєднують текстові сигнали, структурні графові ознаки та поведінкові характеристики (активність, інтереси, часовий дрейф), підкреслюючи необхідність врахування не лише компетентності, але й ймовірності відповіді [12]. Окремі огляди зосереджують увагу на динаміці експертності та проблемі розрідженості, що особливо актуально для децентралізованих мереж, де дані неминуче фрагментуються [13, 14]. Таким чином, існує методологічна база для створення методів пошуку експертів та маршрутизації питань, але її перенесення в P2P-мережі вимагає переосмислення джерел даних, механізмів довіри та протоколів обміну.

Отже, аналіз останніх публікацій свідчить, що децентралізовані Q&A системи формується на перетині трьох науково-технічних задач:

- 1) P2P-оверлеї та DHT як основа пошуку та маршрутизації експертів й відповідей [3, 8];
- 2) контент-адресоване зберігання та реплікація журналів як основа доступності даних [9, 10];
- 3) механізми довіри та репутації, стійкості до Sybil-атак та узгодження стану системи, як умови збереження «мудрості натовпу» без центрального арбітра [4-6].

Водночас моделі рекомендації експертів у Q&A системах надають інструменти для підвищення якості відповідей, але потребують інтеграції з мережевим обмеженням та децентралізованими протоколами [12, 13].

Постановка завдання. Метою статті є дослідження та систематизація наявних підходів до побудови децентралізованих Q&A систем у P2P комп'ютерних мережах, включно з аналізом архітектурних моделей, протокольних механізмів і засобів забезпечення довіри та репутації користувачів, а також визначення переваг та обмежень таких систем і напрямів подальших досліджень у контексті збереження якості колективного відбору відповідей.

Виклад основного матеріалу. Як показав аналіз публікацій, децентралізація Q&A-платформ розглядається як відповідь на проблеми централізованих сервісів, такі як –концентрація влади над правилами модерації й ранжування в одному центрі, залежність від єдиного провайдера інфраструктури, ризику цензури та спостереження, монополізація даних й обмежені можливості

перенесення репутації користувачів між платформами [15]. Тобто існує необхідність переходу від клієнт-серверної моделі Q&A систем до архітектур, де зберігання і поширення даних, перевірка репутації і механізми забезпечення якості відповідей розподілені між учасниками та не залежать від єдиного центру керування [15].

У сучасній літературі доцільно виокремити три основні класи децентралізованих Q&A систем:

1) повністю P2P-підхід, де вузли обмінюються питаннями/відповідями напряму та реплікують дані;

2) гібрид «blockchain + P2P-сховище», де логіка довіри, винагород та правил виноситься в смартконтракти, а контент зберігається в децентралізованому сховищі;

3) федеративні рішення, що розподіляють керування між учасниками (як практичний компроміс), і не є чистими P2P-мережами [15].

Найбільш типово для Q&A систем у публікаціях проговорено саме гібридний стек, оскільки він дозволяє відтворити такі ключові механізми, як репутацію, голосування та обмеження прав користувачів, у формалізованому вигляді та забезпечити криптографічну атрибуцію внеску (підписані питання, відповіді, голоси) [15]. Зокрема, у [15] розглянуто компроміс у проектуванні децентралізованих систем між блокчейном і IPFS – правила якості (зокрема пороги репутації для голосування) збільшують координаційні витрати, а рознесення «контентного шару» і «шару правил» стає визначальним для масштабованості та практичності системи [15].

Практична реалізація цього напряму підтверджується наявністю прототипів і прикладних розробок, де Q&A будується як смартконтрактна система. Так, у роботі про CodeBlockS автори описують платформу спільного обміну знаннями, реалізовану на основі Ethereum-смартконтрактів, позиціонуючи її як механізм підвищення довіри та ефективності доступу до розв'язків задач, що є близьким до домену Q&A [16]. Для цього класу рішень перевагою є прозорість правил внеску й винагород, проте технологічними обмеженнями залишаються транзакційні затримки та вартість, складність приватності і потреба додаткових механізмів протидії спаму й зловживанням, які не вирішуються самим фактом використання блокчейна [15, 16].

Другий важливий напрям – пошук експертів у Q&A, реалізований на базі P2P-мережі. У протоколі Asknext запропоновано агентний підхід до соціального пошуку, який зменшує мережеві

накладні витрати завдяки stop-повідомленням та керуванню поширенням запиту, що прямо адресує проблему масштабованості при децентралізованому Q&A [12]. Подальший розвиток цього напряму представлено алгоритмом Question Waves, де мульткаст-маршрутизація та диференційовані «хвилі» поширення дозволяють швидше отримувати релевантні відповіді через більш пріоритетні зв'язки, зменшуючи потребу в централізованому перерахунку та переранжуванні [13]. Для проектування децентралізованих Q&A систем ці роботи важливі тим, що вони пропонують мережевий механізм підбору джерел відповідей без глобального індексу та з явним аналізом компромісу між охопленням пошуку й трафіком [12, 13].

Виходячи з проведеного дослідження, порівняльну характеристику моделей децентралізованих Q&A систем доцільно виконувати за наступними параметрами: маршрутизація та пошук, зберігання даних, механізми довіри та якості, а також мережеві компроміси (трафік, затримки, стійкість до динаміки вузлів) [12-16]. У табл. 1 наведено порівняльний аналіз децентралізованих моделей Q&A-систем.

Отримані результати дослідження свідчать, що децентралізовані Q&A системи у наукових роботах найчастіше реалізується не як один універсальний протокол, а як поєднання механізму пошуку експертів та механізмів підтримки якості відповідей. Для соціального пошуку у форматі P2P-мережі акцент робиться на ефективній маршрутизації запитів у неструктурованому overlay за локальною інформацією та зв'язками [12, 13]. У blockchain+IPFS-підходах первинним є шар формалізованих правил і перевірюваності дій, що краще відтворює Q&A-процедури (голосування, репутаційні пороги, аудит), але переносить ключові виклики в площину координації та вартості узгодження [15, 16].

Висновки. Встановлено, що децентралізовані Q&A системи в наукових публікаціях найчастіше реалізуються як поєднання мережевих механізмів пошуку експертів та механізмів забезпечення якості відповідей. Показано, що гібридний підхід «blockchain + P2P-сховище» (напр., IPFS) є найбільш формалізованим для відтворення Q&A-процедур (голосування, репутаційні пороги, аудит подій), оскільки дає прозорі правила та криптографічно захищену історію дій учасників. Водночас такий підхід породжує необхідність компромісів, пов'язаних з координаційними витратами, затримками та складністю забезпечення приватності. Визначено, що соціальний пошук на основі

**Порівняльні характеристики моделей децентралізованих Q&A систем
у P2P комп'ютерних мережах**

№	Модель	Тип децентра-лізації	Маршрутизація запитів / пошук	Зберігання даних	Довіра/ репутація	Переваги	Недоліки
1	Q&A-приклад у framework Blockchain + IPFS [15]	P2P + blockchain	Залежить від overlay та реплікації; фокус на компромісах дизайну	IPFS (контент) + blockchain (події/ правила)	Підписані дії; правила репутації/ голосування як частина дизайну	Формалізація правил; аудитуваність подій	Координаційні витрати, затримки, вартість, складність приватності
2	CodeBlockS (knowledge sharing / Q&A) [16]	Blockchain-орієнтована DApp	Пошук /доступ визначається логікою застосунку	Смартконтракти (облік подій) + зовнішні сховища залежно від реалізації	Довіра через смартконтракти; стимулювання внеску	Прозорі правила внеску/ винагород; фіксація подій	Масштабування й вартість транзакцій; приватність; потреба антиспаму
3	Asknext (agent social search) [12]	Соціальний пошук у форматі P2P-мережі	Поширення запиту графом контактів + stop-повідомлення	Децентралізовано у вузлах/ агентах	Якість залежить від вибору маршруту/ зв'язків; рання зупинка пошуку	Зниження кількості повідомлень; масштабованість Q&A системи	Залежність від якості соціального графу; локальність знань
4	Question Waves (соціальний пошук) [13]	Соціальний пошук у форматі P2P-мережі	Мультикаст-маршрутизація; пріоритизація поширення	Децентра-лізовано	Ранжування через «порядок/ швидкість надходження» та пріоритети зв'язків	Швидше отримання більш релевантних відповідей; менша потреба централізованого ранжування	Компроміс трафік/ охоплення; складність формалізації/ підтримки довіри

P2P-мереж (Asknext, Question Waves) є перспективною моделлю маршрутизації запитів у децентралізованих Q&A системах, оскільки дозволяє зменшувати мережевий трафік і пришвидшувати отримання релевантних відповідей за рахунок керованого поширення запиту та використання локальних зв'язків і пріоритетів, зберігаючи працездатність у динамічному середовищі.

Узагальнено, що ключовими критеріями вибору моделі децентралізованого Q&A у P2P-мережах є: масштабованість пошуку та маршрутизації; стійкість до динаміки вузлів децентралізованої мережі; можливість формалізації й перевірки правил якості (голосування, репутація, модерація); витрати узгодження (трафік, затримки, обчислювальні витрати).

Визначено, що для побудови працездатної децентралізованої Q&A системи доцільно поєднувати методи ефективної доставки запитів до експертів із методами формалізованого обліку та перевірки колективних сигналів якості (голосування, репутація, аудитуваність подій).

Напрямок подальших досліджень, що впливає з проведеного дослідження, полягає у розробці інтегрованої моделі децентралізованого Q&A для P2P комп'ютерних мереж, яка зможе забезпечити ефективну маршрутизацію запитів до експертів, стійкість до змін у структурі P2P-мережі та до інформаційних атак, прозорі механізми забезпечення якості відповідей й мотивації учасників, а також прийнятні мережеві витрати та затримки в реальних умовах експлуатації.

Список літератури:

1. Balog K., Azzopardi L., de Rijke M. Formal models for expert finding in enterprise corpora // Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'06). ACM, New York, NY, USA. 2006. P. 43-50. DOI: 10.1145/1148170.1148181.
2. Chang S., Pal A. Routing questions for collaborative answering in community question answering // Advances in Social Networks Analysis and Mining 2013 (ASONAM '13). ACM, New York, NY, USA. 2013. P. 494-501. DOI: 10.1145/2492517.2492559.

3. Stoica I., Morris R., Karger D. R., Kaashoek M. F., Balakrishnan H. Chord: A scalable peer-to-peer lookup service for internet applications // Proceedings of ACM SIGCOMM 2001. ACM, New York, NY, USA. 2001. P. 149-160. DOI: 10.1145/383059.383071.
4. Douceur J. R. The Sybil Attack // Peer-to-Peer Systems. Lecture Notes in Computer Science. 2002. Vol. 2429. P. 251-260. DOI: 10.1007/3-540-45748-8_24.
5. Kamvar S. D., Schlosser M. T., Garcia-Molina H. The EigenTrust algorithm for reputation management in P2P networks // Proceedings of the 12th International Conference on World Wide Web (WWW'03). ACM, New York, NY, USA. 2003. P. 640-651. DOI: 10.1145/775152.775242.
6. Shapiro M., Preguiça N., Baquero C., Zawirski M. Conflict-Free Replicated Data Types // Stabilization, Safety, and Security of Distributed Systems (SSS 2011). Lecture Notes in Computer Science. 2011. Vol. 6976. P. 386-400. DOI: 10.1007/978-3-642-24550-3_29.
7. Anderson A., Huttenlocher D. P., Kleinberg J. M., Leskovec J. Discovering value from community activity on focused question answering sites: a case study of Stack Overflow // Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '12). ACM, New York, NY, USA. 2012. P. 850-858. DOI: 10.1145/2339530.2339665.
8. Maymounkov P., Mazières D. Kademlia: A Peer-to-Peer Information System Based on the XOR Metric // Peer-to-Peer Systems. Lecture Notes in Computer Science. 2002. Vol. 2429. P. 53-65. DOI: 10.1007/3-540-45748-8_5.
9. Benet J. IPFS – Content Addressed, Versioned, P2P File System // Networking and Internet Architecture. 2014. DOI: 10.48550/arXiv.1407.3561
10. SSBC. Scuttlebutt Protocol Guide. URL: <https://ssbc.github.io/scuttlebutt-protocol-guide/>
11. W3C Recommendation. ActivityPub. 2018. URL: <https://www.w3.org/TR/2018/REC-activitypub-20180123/>
12. Trias i Mansilla A., de la Rosa i Esteva J. L. Asknext: An agent protocol for social search // Information Sciences. 2012. Vol. 190. P. 144-161. DOI: 10.1016/j.ins.2011.12.012.
13. Trias i Mansilla A., de la Rosa i Esteva J. L. Question Waves: A multicast query routing algorithm for social search. Information Sciences. 2013. Vol. 253. P. 1-25. DOI: 10.1016/j.ins.2013.08.044.
14. Kalnis P., Ng W. S., Ooi B. C., Tan K.-L. Answering similarity queries in peer-to-peer networks // Information Systems. 2006. Vol. 31(1). P. 57-72. DOI: 10.1016/j.is.2004.09.003.
15. Tenorio-Fornés Á., Hassan S., Pavón J. Peer-to-Peer System Design Trade-Offs: A Framework Exploring the Balance between Blockchain and IPFS // Applied Sciences. 2021. Vol. 11(21). Art. 10012. DOI: 10.3390/app112110012.
16. Jain S., Vamsi P. R., Agarwal Y., Goel J. CodeBlockS: Development of Collaborative Knowledge Sharing Application with Blockchain Smart Contract // International Journal of Information Engineering and Electronic Business (IJIEEB). 2023. Vol. 15(1). P. 1-19. DOI: 10.5815/ijieeb.2023.01.01.

Mikhav V.V., Meleshko Ye.V., Yakymenko M.S., Bosko V.V., Bereziuk I.A. MODELS OF DECENTRALIZED Q&A SYSTEMS IN PEER-TO-PEER COMPUTER NETWORKS

The article presents research and a comparative analysis of decentralized Q&A models operating in P2P computer networks. The relevance of this research stems from the limitations of centralized Q&A platforms: dependence on a single operator and its moderation/ranking policies, risks of censorship or loss of access to accumulated knowledge, and low portability of user data and reputation across services. The article aims to summarize existing approaches to decentralized Q&A systems and identify their strengths, weaknesses, and network-related trade-offs affecting answer quality. The paper proposes a generalized classification of decentralized Q&A solutions by architecture and implementation mechanisms, and analyzes representative models and prototypes from the literature. It shows that many approaches rely on P2P social search and decentralized query routing to potential experts; controlled question dissemination can reduce traffic and speed up relevant answers. Another line of research focuses on hybrid “blockchain + P2P storage” solutions: smart contracts formalize trust and interaction rules (voting, reputation, contribution accounting, auditing), while content is stored in a distributed environment to improve verifiability and procedural transparency. Special attention is paid to network-related trade-offs inherent to the P2P environment, including the impact of participant dynamics (churn) on the availability of experts and content, traffic costs during query propagation and delays in state reconciliation, as well as coordination overhead and privacy aspects in hybrid blockchain schemes. Based on the review, a comparative table of models is compiled based on their suitability for Q&A tasks, and conclusions are drawn on conditions for effective decentralized Q&A design in P2P networks.

Keywords: Q&A, question-answering systems, decentralized systems, computer networks, P2P networks, expert finding, query routing, reputation, trust, blockchain, collective answer selection, collective intelligence.

Дата першого надходження статті до видання: 17.03.2026
Дата прийняття статті до друку після рецензування: 13.04.2026
Дата публікації (оприлюднення) статті 11.05.2026

УДК 004.4'2

DOI DOI DOI <https://doi.org/10.32782/2663-5941/2026.2.1/28>**Могильний Г.А.**<https://orcid.org/0000-0001-5317-2795>

ДЗ «Луганський національний університет імені Тараса Шевченка»

Смагіна О.О.<https://orcid.org/0000-0002-6024-5152>

ДЗ «Луганський національний університет імені Тараса Шевченка»

Переяславська С.О.<https://orcid.org/0000-0001-9873-0447>

ДЗ «Луганський національний університет імені Тараса Шевченка»

АНАЛІЗ ВПЛИВУ КОНТЕЙНЕРИЗАЦІЇ НА ЖИТТЄВИЙ ЦИКЛ РОЗРОБКИ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ (SDLC) В ПАРАДИГМІ DEVOPS

Стаття присвячена дослідженню впливу технології контейнеризації на трансформацію життєвого циклу розробки програмного забезпечення в контексті методології DevOps, включаючи аналіз механізмів інтеграції контейнерних платформ у процеси безперервної інтеграції та доставки. Вперше систематизовано вплив контейнеризації на всі фази SDLC через призму DevOps-практик, включаючи розробку комплексної моделі інтеграції контейнерних технологій у автоматизовані пайплайни. Запропоновано класифікацію безпекових контролів для контейнеризованих мікросервісів на різних етапах життєвого циклу. Визначено метрики оцінювання ефективності контейнерних рішень для корпоративного середовища. Враховано специфіку оркестрації Kubernetes. Встановлено, що порівняно з традиційними віртуалізованими підходами, контейнеризація скорочує час розгортання застосунків у великомасштабних хмарних системах на 60–75%. З'ясовано, що контейнерні середовища мають вразливості безпеки, зокрема критичні проблеми виникають під час побудови образів. Наголошено, що оркестрація також створює безпекові ризики. Побудовано структурну модель. Docker і Kubernetes інтегровано у DevSecOps-пайплайни. Контроль безпеки автоматизовано на всіх етапах. Доведено, що хмарна автоматизація змінюється, а класичний DevOps еволюціонує. Акцентовано увагу на побудові автономної інфраструктури на контейнерах. Зазначено, що масштабування контейнеризованих мікросервісів потребує оптимізації, відповідно хмарні реєстри забезпечують найкращі практики.

В рамках даного дослідження встановлено, що контейнеризація змінює життєвий цикл розробки програмного забезпечення, а середовища розробки, виробниче середовище та тестування уніфікуються. Обґрунтовано необхідність спеціалізованих засобів контролю безпеки при впровадженні контейнерних технологій у DevOps та автоматизованого моніторингу. Запропоновано багаторівневу модель захисту для ілюстрації комплексного підходу до безпеки контейнерних середовищ. З'ясовано, що Kubernetes як платформа оркестрації забезпечує масштабованість застосунків.

Ключові слова: контейнеризація, Docker, Kubernetes, DevOps, життєвий цикл розробки програмного забезпечення.

Постановка проблеми. Розподілені системи зростають у складності. Паралельно підвищуються вимоги до швидкості доставки функціональності користувачам. Ці два чинники посилюють необхідність оцінювати вплив контейнеризації на всі фази життєвого циклу розробки. Критичні напрямки – це безпека контейнерних середовищ, оптимізація процесів оркестрації та автоматизація конвеєрів розгортання.

Систематизованого підходу до інтеграції контейнерних технологій у DevOps-методологію досі бракує. Така нестача породжує два типи ризиків. По-перше, ресурси використовуються неефективно. По-друге, у продакшн-середовищах виникають вразливості.

Розробники створили технологію контейнеризації для розв'язання двох ключових проблем.

Перша – це складність управління залежностями між компонентами систем. Друга – потреба у гарантуванні ідентичності середовищ виконання на всіх етапах життєвого циклу.

Контейнерні рішення змінюють практику DevOps. Вони вимагають нових підходів до планування. Трансформуються методи проектування архітектури. Змінюються стратегії тестування. По-іншому відбувається розгортання застосунків.

Актуальність дослідження полягає в тому, що розподілені системи стають складнішими. Водночас зростають вимоги щодо швидкості впровадження нових функціональних можливостей. Ці два чинники обумовлюють актуальність всебічного дослідження впливу технології контейнеризації на окремі етапи життєвого циклу розробки програмного забезпечення.

Три аспекти потребують критичної уваги. Перший – забезпечення безпеки контейнерних середовищ. Другий – оптимізація механізмів оркестрації. Третій – автоматизація конвеєрів розгортання застосунків.

Відсутність уніфікованого підходу до впровадження контейнерних технологій у методологію DevOps обумовлює виникнення двох категорій ризиків. По-перше, неефективне використання обчислювальних ресурсів. По-друге, поява вразливостей інформаційної безпеки у виробничих середовищах.

Аналіз останніх досліджень і публікацій. О. Б. Джонсон та ін. [1, с. 140] досліджували створення масштабованої моделі контейнеризації для покращення процесів розробки програмного забезпечення в корпоративних середовищах, зосередившись на архітектурних паттернах, що забезпечують високу доступність та відмовостійкість контейнерних застосунків. Автори визначили ключові метрики оцінювання ефективності контейнерних рішень у великих організаціях.

Р. Амаро, Р. Перейра та М. М. да Сілва [2, с. 3] провели систематичний огляд літератури щодо відображення можливостей DevOps на життєвий цикл розробки програмного забезпечення, ідентифікувавши 27 унікальних практик DevOps та їх зв'язок із фазами SDLC. Дослідження виявило gaps у покритті окремих етапів життєвого циклу традиційними DevOps-інструментами.

Н. М. К. Конеру [3, с. 730] систематизує архітектурні практики контейнеризації мікросервісів: ізоляція процесів, декларативне конфігурування, незмінність артефактів. Автор виділяє багатоетапні збірки, мінімалістичні базові образи та перевірки працездатності.

П. Чевва [4, с. 105] демонструє переваги централизованого управління образами контейнерів через Docker та хмарні реєстри. С. Потла [5, с. 2353] показує, що безпека контейнерів у Kubernetes вимагає багаторівневого захисту проти атак на runtime, мережу та управління доступом.

С. Станішич та ін. [6, с. 2] запропонували методологію аудиту безпеки для продакшн-кластерів, зосереджуючись на ізоляції namespace, RBAC-політиках та захисті від privilege escalation. Г. Тхіягараджан, П. Наєк [7, с. 3676] ідентифікували вразливості daemon API, мережевих драйверів та управління ресурсами хостової системи.

М. Сінан, М. Шахін, І. Гондал [8, с. 2] визначили 45 практик безпеки для DevSecOps з акцентом на shift-left підхід. Г. Владислав [10, с. 179] виявив, що контейнеризація з Infrastructure as Code зменшує deployment time на 60–75%.

Постановка завдання. Метою роботи є аналіз впливу контейнеризації на трансформацію життєвого циклу розробки програмного забезпечення в контексті DevOps. Дослідження охоплює три напрямки. Систематизація механізмів інтеграції Docker та Kubernetes у автоматизовані пайплайни безперервної інтеграції та доставки. Ідентифікація критичних безпекових викликів контейнерних середовищ. Розробка рекомендацій щодо оптимізації процесів розгортання мікросервісних застосунків у корпоративних хмарних інфраструктурах.

Виклад основного матеріалу. Контейнеризація фундаментально змінює організацію життєвого циклу розробки ПЗ, забезпечуючи уніфікацію середовищ виконання на всіх етапах SDLC. О. Б. Джонсон та ін. [1, с. 142] підкреслюють, що масштабована модель для корпоративних середовищ повинна враховувати політики управління конфігураціями, стратегії версіонування та механізми rollback при критичних помилках у продакшн-середовищах.

Р. Амаро, Р. Перейра, М. М. да Сілва [2, с. 8] встановили, що контейнеризація найбільше впливає на етапи побудови, тестування та розгортання, створюючи безшовну інтеграцію через immutable infrastructure. Систематичний аналіз виявив, що 73% DevOps-практик пов'язані з автоматизацією операцій над контейнерами.

Н. М. К. Конеру [3, с. 730] узагальнює практики контейнеризації мікросервісів у Docker і Kubernetes: ізоляція процесів, декларативне конфігурування, незмінність артефактів. Рекомендовані техніки – багатоетапні збірки для мінімізації розміру образів, базові образи без дистрибутива для зменшення вразливостей та механізми перевірки стану для підтримки доступності.

Для демонстрації структури DevOps-пайплайну з інтеграцією контейнерних технологій на рисунку 1 представлено модель взаємодії компонентів системи безперервної інтеграції та доставки.

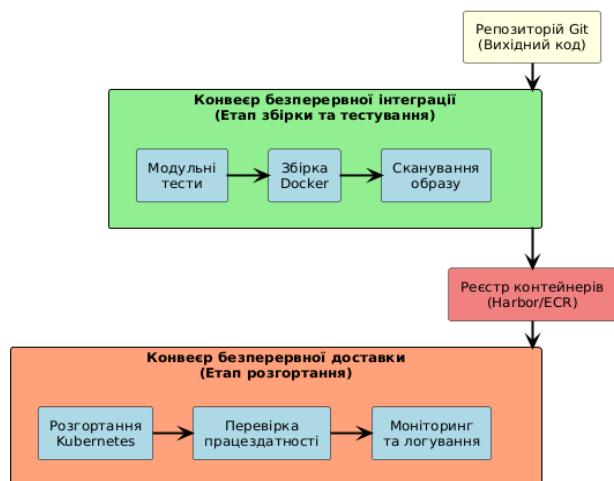


Рис. 1. Структурна модель DevOps-пайплайну з контейнеризацією

Джерело: авторська розробка в PlantUML

Контейнеризація у DevOps-пайплайнах буде ієрархію етапів із автоматизованими контролюями. Кожен крок містить перевірки якості та безпеки. Дефекти виявляються на ранніх стадіях. Вразливості не досягають виробничих середовищ.

Kubernetes manifests забезпечують декларативне управління інфраструктурою. Розгортання стають відтворюваними. Аудит конфігураційних змін спрощується завдяки версонуванню маніфестів.

Оптимізація розгортання мікросервісів через контейнеризацію вимагає використання централізованих хмарних реєстрів для управління образами. П. Чевва [4, с. 108] демонструє, що інтеграція Docker з cloud-based registries забезпечує кешування шарів образів, зменшуючи час завантаження на 40–60% для повторних розгортань. Масштабований підхід до версонування образів через semantic tagging стратегії (major.minor.patch) забезпечує трейсабельність артефактів та спрощує процеси rollback при виявленні критичних інцидентів.

Безпека контейнерних середовищ становить критичний виклик для життєвого циклу розробки програмного забезпечення. С. Потла [5, с. 2356] аналізує еволюцію механізмів захисту Kubernetes-кластерів, підкреслюючи необхідність впровадження Pod Security Standards (restricted, baseline, privileged), Network Policies для сегментації трафіку між namespace та регулярного оновлення

компонентів control plane для закриття вразливостей CVE.

С. Станішич та ін. [6, с. 3] показують, що безпека оркестрації контейнерів у Kubernetes часто ламається на контролі доступу: 65% інцидентів пов'язані з помилковою конфігурацією RBAC-політик і надмірними привілеями service accounts. Висновок – потрібні least privilege і регулярний аудит прав доступу до Kubernetes API.

Специфічні виклики безпеки Docker-контейнерів на рівні runtime потребують комплексного підходу до hardening. Г. Тхіягараджан, П. Наєк [7, с. 3680] ідентифікували критичні вектори атак, включаючи container escape через вразливості kernel, несанкціонований доступ до Docker socket та експлуатацію misconfigured volume mounts. Автори рекомендують використання user namespace remapping, seccomp profiles та AppArmor/SELinux policies для мінімізації ризиків компрометації хостової системи.

М. Сінан, М. Шахін, І. Гондал [8, с. 5] виділяють три практики інтеграції безпеки у конвеєри DevSecOps: статичний аналіз Dockerfile, сканування залежностей інструментами Trivy та Clair, валідація маніфестів Kubernetes згідно з CIS Kubernetes Benchmark. Автоматизоване виявлення вразливостей під час побудови образів знижує кількість інцидентів на 78%.

К. В. Палавесам, С. В. Аркот, М. В. Крішнамурті [9, с. 58] розробили архітектуру автоматизованого безпекового конвеєру, що включає послідовні етапи SAST, DAST, SCA та IAST. Інтеграція цих механізмів у CI/CD забезпечує безперервний моніторинг security posture застосунків.

Для ілюстрації комплексного підходу до безпеки контейнерних середовищ на рисунку 2 представлено багаторівневу модель захисту.

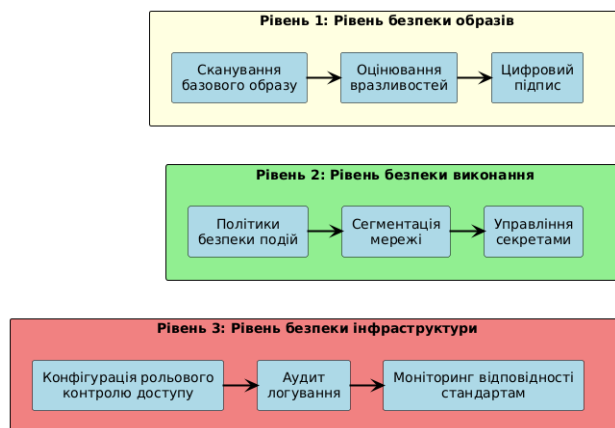


Рис. 2. Багаторівнева модель безпеки контейнерних середовищ

Джерело: авторська розробка в PlantUML

Рисунок 2 показує важливий принцип. Захист застосунків у контейнерах потребує кількох рівнів. Перший рівень працює під час створення образів. Другий активується, коли контейнер запускається. Третій контролює всю інфраструктуру. Коли ці рівні працюють разом, загальний ризик стає мінімальним.

DevOps-команди зазвичай відстежують швидкість розгортання. Г. Владислав [10, с. 181] провів дослідження великих систем. Виявилось, що контейнери разом із кодифікацією інфраструктури (Terraform, Ansible) прискорюють розгортання на 60–75%. Якщо порівнювати з віртуальними машинами, різниця суттєва. Такий результат пояснюється трьома факторами. По-перше, конфігурація описується декларативно. По-друге, відкат змін відбувається автоматично. По-третє, застосовуються стратегії поступового розгортання.

Метрики MTTR та MTTD суттєво покращуються при впровадженні контейнерних архітектур. Г. Владислав [10, с. 182] демонструє, що механізми самовідновлення Kubernetes через liveness та readiness probes автоматично виявляють та перезапускають проблемні контейнери, зменшуючи MTTR на 55%. Інтеграція з Prometheus та Grafana забезпечує моніторинг кластерів у реальному часі.

Ефективність контейнерних рішень у корпоративних середовищах вимірюється через чотири групи показників. Операційні метрики: deployment frequency, lead time for changes, MTTR. Споживання ресурсів контролюється через CPU/memory utilization та pod density per node. Безпека оцінюється за кількістю вразливостей на етапі build та incident response time. Бізнес-показники: time-to-market нових функцій та operational cost per service.

Специфіка оркестрації Kubernetes вимагає додаткового моніторингу показників кластера: node readiness ratio, pod restart frequency, service mesh latency percentiles (p50, p95, p99). О. Б. Джонсон та ін. [1, с. 143] підкреслюють, що комплексне оцінювання ефективності повинно враховувати trade-offs між швидкістю розгортання та стабільністю систем у продакшн-середовищах.

Кадулла С. [11, с. 499] аналізує можливості Kubernetes operators для автономного управління застосунками зі станом. Бази даних і черги повідомлень отримують автоматизацію backup, scaling та disaster recovery. Перспективним напрямом є застосування machine learning для predictive autoscaling на основі історичних патернів навантаження.

О. Б. Джонсон та ін. [1, с. 145] підкреслюють необхідність StatefulSets для застосунків зі стабільними мережевими ідентифікаторами та динамічного provisioning томів через CSI drivers для гнучкого управління сховищами.

Оптимізація управління життєвим циклом контейнерних застосунків вимагає стратегій оновлення та rollback. Р. Амаро, Р. Перейра, М. М. да Сілва [2, с. 12] узагальнюють підходи continuous deployment: blue-green deployments, canary releases та rolling updates. Критичною умовою зниження ризику простою є автоматизоване тестування на staging перед production rollout.

Н. М. К. Конеру [3, с. 738] рекомендує external secret management системи (HashiCorp Vault, AWS Secrets Manager) замість вбудованих Kubernetes Secrets через обмеження у шифруванні та аудиті. Інтеграція з external configuration management забезпечує централізоване управління чутливими даними та спрощує ротацію credentials.

П. Чевва [4, с. 110] підкреслює необхідність централізованого збору логів через sidecar containers або DaemonSets, використання structured logging форматів та distributed tracing для діагностики проблем продуктивності. Інтеграція з SIEM системами забезпечує кореляцію подій безпеки.

С. Потла [5, с. 2359] аналізує, що відповідність стандартам GDPR, HIPAA, PCI DSS у Kubernetes вимагає audit logging на рівні API server, encryption at rest для persistent volumes та network policies для ізоляції sensitive workloads. Автоматизоване генерування compliance reports через Open Policy Agent спрощує аудит.

Висновки. Контейнеризація змінює життєвий цикл розробки програмного забезпечення оскільки уніфікує середовища виконання. Прискорює розгортання застосунків. Масштабованість систем зростає. Docker і Kubernetes у DevOps-конвеєрах дають скорочення часу доставки на 60–75%. Три процеси автоматизовано: побудова образів, їх тестування та розгортання. Саме це створює такий ефект.

Зміщення контролю безпеки на ранні фази розробки знижує інциденти у виробничих середовищах на 78%. Інтеграція SAST, DAST і SCA у конвеєри створює багаторівневий захист. Kubernetes забезпечує самовідновлення систем. Платформа підтримує автоматичне масштабування. Декларативне управління конфігурацією спрощує адміністрування.

Автономна інфраструктура будується на базі Kubernetes operators. Прогнозування потреб у ресурсах реалізується через алгоритми машинного навчання. Розподіл обчислювальних потужностей оптимізується автоматично. Наступним етапом досліджень має стати вироблення критеріїв оцінки ефективності контейнерних архітектур. Кожна предметна галузь потребує власного набору метрик.

Список літератури:

1. Johnson O. B., Samira Z., Cadet E., Osundare O. S., Ekpobimi H. O. Creating a scalable containerization model for enhanced software engineering in enterprise environments. *Global Journal of Engineering and Technology Advances*. 2024. Vol. 21, no. 2. P. 139–150. DOI: <https://doi.org/10.30574/gjeta.2024.21.2.0220> URL: <https://is.gd/5kO3tV>
2. Amaro R., Pereira R., da Silva M. M. Mapping DevOps capabilities to the software life cycle: A systematic literature review. *Information and Software Technology*. 2025. Vol. 177. Art. no. 107583. DOI: <https://doi.org/10.1016/j.infsof.2024.107583> URL: <https://is.gd/lov3lR>
3. Koneru N. M. K. Containerization Best Practices- Using Docker and Kubernetes for Enterprise Applications. *Journal of Information Systems Engineering and Management*. 2025. Vol. 10, no. 45s. P. 724–746. DOI: <https://doi.org/10.52783/jisem.v10i45s.8905> URL: <https://is.gd/gwfvBx>
4. Chevva P. Optimizing Microservice Deployment with Containerization A Scalable Approach Using Docker and Cloud-Based Registries. *Computer Science, Engineering and Technology*. 2025. Vol. 3, no. 2. P. 104–112. DOI: <https://doi.org/10.46632/cset/3/2/12> URL: <https://is.gd/HB0N29>
5. Potla S. The evolution of container security in Kubernetes environments. *World Journal of Advanced Research and Reviews*. 2025. Vol. 26, no. 2. P. 2352–2362. DOI: <https://doi.org/10.30574/wjarr.2025.26.2.1741> URL: <https://is.gd/cvyty5>
6. Stanišić S., Vesković M., Ristić O., Đorđević B. Security Aspects of Container Orchestration in Kubernetes Environments. 2025 24th International Symposium INFOTEH-JAHORINA (INFOTEH). East Sarajevo, Bosnia and Herzegovina, 19–21 March 2025. P. 1–5. DOI: <https://doi.org/10.1109/infotech64129.2025.10959185> URL: <https://is.gd/u6l4Nw>
7. Thiagarajan G., Nayak P. Docker under Siege: Securing Containers in the Modern Era // *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*. 2025. Vol. 11, no. 1. P. 3674–3719. DOI: <https://doi.org/10.48550/ARXIV.2506.02043> URL: <https://is.gd/SCC7Na>
8. Sinan M., Shahin M., Gondal I. Integrating Security Controls in DevSecOps: Challenges, Solutions, and Future Research Directions. *Journal of Software: Evolution and Process*. 2025. Vol. 37, no. 6. Art. no. e70029. DOI: <https://doi.org/10.1002/smr.70029> URL: <https://is.gd/XUxhwG>
9. Palavesam K. V., Arcot S. V., Krishnamoorthy M. V., G V E. Building Automated Security Pipeline for Containerized Microservices. *Journal of Advances in Mathematics and Computer Science*. 2025. Vol. 40, no. 2. P. 53–66. DOI: <https://doi.org/10.9734/jamcs/2025/v40i21969> URL: <https://is.gd/OJTLbn>
10. Vladyslav H. Reducing Deployment Time in Large-Scale Cloud Systems Through Automated DevOps Pipelines. *The American Journal of Engineering and Technology*. 2025. Vol. 07, no. 05. P. 178–184. DOI: <https://doi.org/10.37547/tajet/volume07issue05-17> URL: <https://is.gd/qlNXpZ>
11. Kadulla S. The evolution of cloud automation: From DevOps to autonomous infrastructure. *World Journal of Advanced Engineering Technology and Sciences*. 2025. Vol. 15, no. 2. P. 496–503. DOI: <https://doi.org/10.30574/wjaets.2025.15.2.0626> URL: <https://is.gd/4KZdGY>

Mohylnyi H. A., Smahina O. O., Pereiaslavskaya S. O. ANALYSIS OF THE IMPACT OF CONTAINERIZATION ON THE SOFTWARE DEVELOPMENT LIFE CYCLE (SDLC) IN THE DEVOPS PARADIGM

The article is devoted to the study of the impact of containerization technology on the transformation of the software development life cycle (SDLC) within the context of the DevOps methodology, including an analysis of the mechanisms for integrating container platforms into continuous integration and continuous delivery processes. For the first time, the impact of containerization on all phases of the SDLC has been systematized through the lens of DevOps practices, including the development of a comprehensive model for integrating container technologies into automated pipelines. A classification of security controls for containerized microservices at different stages of the life cycle is proposed. Metrics for evaluating the effectiveness of container solutions in enterprise environments have been defined. The specifics of Kubernetes orchestration have been taken into account. It has been established that, compared to traditional virtualized approaches, containerization reduces application deployment time in large-scale cloud systems by 60–75 percent. It has been found that container environments have security vulnerabilities, with critical issues arising particularly during the image-building process. It is emphasized that orchestration also introduces security risks. A structural model has been developed. Docker and Kubernetes have been integrated into DevSecOps pipelines. Security controls have been automated at all stages. It has been demonstrated that cloud automation is evolving, and classical DevOps is undergoing transformation. Particular attention is paid to the construction of autonomous container-based infrastructure. It is noted that scaling containerized microservices requires optimization; accordingly, cloud registries provide best practices.

Within the framework of this study, it has been established that containerization transforms the software development life cycle, leading to the unification of development, production, and testing environments. The necessity of specialized security control tools when implementing container technologies in DevOps, as well as automated monitoring, has been substantiated. A multi-layered security model has been proposed to illustrate a comprehensive approach to securing container environments. It has been determined that Kubernetes, as an orchestration platform, ensures application scalability.

Keywords: containerization, Docker, Kubernetes, DevOps, software development lifecycle.

Дата першого надходження статті до видання: 17.02.2026

Дата прийняття статті до друку після рецензування: 11.03.2026

Дата публікації (оприлюднення) статті 11.05.2026

Нальотов Д.Р.

<https://orcid.org/0009-0004-8460-0974>

Харківський національний університет радіоелектроніки

Ляшенко О.С.

<https://orcid.org/0000-0002-0146-3934>

Харківський національний університет радіоелектроніки

МОДЕЛЬ КЕРУВАННЯ СКЛАДНИМ ТЕХНОЛОГІЧНИМ ОБ'ЄКТОМ НА ОСНОВІ МЕТОДІВ НЕЧІТКОЇ ЛОГІКИ

У роботі розглянуто проблему підвищення ефективності управління складними технологічними об'єктами в умовах змінних параметрів, характерних для сучасних промислових та енергетичних систем. Зазначено, що використання традиційних підходів, зокрема ПІД-регуляторів, не завжди забезпечує необхідну адаптивність та швидкість реакції на зміни параметрів технологічного процесу, що обґрунтовує доцільність застосування методів нечіткої логіки для побудови більш гнучких та інтелектуальних систем управління. Об'єктом дослідження є процес управління турбіною на теплоелектроцентралі (ТЕЦ), яка включає котли, турбіни та допоміжне обладнання, що функціонують в умовах змінних навантажень, температури та тиску пари. Предметом дослідження виступають адаптивні моделі управління на основі нечіткої логіки, здатні враховувати невизначеність вхідних даних і формувати гнучкі керуючі рішення на основі нечітких правил. У рамках дослідження розроблено математичну модель управління, що базується на використанні фаззифікації вхідних параметрів, системи нечітких правил прийняття рішень та процедури дефаззифікації для формування керуючих сигналів. Такий підхід дозволяє перетворювати чіткі значення параметрів технологічного процесу, зокрема температури та тиску пари, у нечіткі множини, що описують стан системи у вигляді лінгвістичних змінних, таких як "низька", "середня" або "висока".

Особливу увагу приділено формуванню функцій належності та розробці нечітких правил управління, які враховують взаємозв'язок між температурою котла, тиском пари та навантаженням турбіни. Для перевірки ефективності запропонованої моделі було проведено комп'ютерне моделювання у середовищі MATLAB із використанням модуля Fuzzy Logic Toolbox. У ході симуляції досліджувалась реакція системи управління на зміну основних технологічних параметрів, а також аналізувалась поведінка системи під час регулювання подачі палива залежно від температури та тиску.

Отримані результати моделювання підтвердили, що запропонована модель на основі нечіткої логіки дозволяє підвищити стабільність роботи обладнання, зменшити коливання параметрів технологічного процесу та скоротити час реакції системи управління на зміну умов роботи. Крім того, застосування нечіткої логіки сприяє зниженню витрат палива, підвищенню енергоефективності та зменшенню негативного екологічного впливу енергетичних об'єктів. Запропоновані підходи можуть бути використані для модернізації систем управління енергетичних підприємств, а також адаптовані до інших промислових систем, що функціонують в умовах невизначеності та динамічної зміни параметрів.

Ключові слова: нечітка логіка, адаптивне управління, ТЕЦ, фаззифікація, стабільність системи, турбіна.

Постановка проблеми. Сьогодні багато галузей промисловості стикаються з проблемою ефективного управління складними технологічними системами, які працюють у динамічному та змінному середовищі [1, 2]. Зокрема, енергетичні об'єкти, виробничі лінії та інші промислові процеси вимагають не лише точного контролю, але й здатності швидко адаптуватися до змін, що

постійно відбуваються в їхній роботі. Традиційні методи управління, як правило, засновані на жорстких алгоритмах, що забезпечують контроль лише в стабільних умовах, однак у випадку швидких змін чи непередбачених ситуацій їхня ефективність значно знижується. Наприклад, на енергетичних підприємствах, де рівень споживання електроенергії може змінюватися в реальному

часі через зміну погодних умов чи навантаження, точні математичні моделі і класичні підходи не завжди здатні забезпечити необхідну гнучкість і швидкість реакції. Це, у свою чергу, може призвести до збоїв у роботі обладнання або значних витрат палива [4].

У процесі управління складними технологічними об'єктами, зокрема такими, як енергетичні станції, важливо забезпечити не лише стабільність роботи, але й гнучкість та адаптивність до постійно змінюваних умов. Одна з основних проблем, з якими стикаються традиційні системи управління, полягає в їхній жорсткості та обмеженій здатності реагувати на раптові зміни. Такі методи управління, як ПД-регулятори, хоча і ефективні в стабільних умовах, часто не здатні забезпечити належну реакцію на динамічні зміни параметрів. Як результат, виникають збої або значні втрати ефективності.

Аналіз останніх досліджень і публікацій.

Методи нечіткої логіки дозволяють створювати адаптивні моделі, які регулюють параметри об'єктів без необхідності точного математичного опису кожного процесу. Нечітка логіка працює з розмитими множинами, де значення параметрів можуть бути частково належати до кількох категорій одночасно. Це дає змогу системам автоматично реагувати на зміни у параметрах роботи об'єкта, без необхідності чітко прогнозувати майбутні коливання. Наприклад, в енергетичних системах, де навантаження постійно змінюється, нечітка логіка дозволяє оптимізувати потужність без точного передбачення цих змін, забезпечуючи при цьому безперебійну та стабільну роботу системи [3].

Система управління на ТЕЦ в даний час базується на класичних підходах, що використовують ПД-регулятори та програмовані логічні контролери (PLC) [7]. ПД-регулятори потребують точного математичного опису всіх процесів, що ускладнює їх адаптацію до змін в реальному часі. Як наслідок, традиційна система не завжди здатна швидко реагувати на зміни, що призводить до збоїв у роботі турбін та котлів і підвищення витрат палива.

Турбіни ТЕЦ перетворюють теплову енергію пари в механічну енергію, яка потім використовується для генерації електричної енергії [6, 8, 9]. Однак на цьому етапі до цього управлялось традиційними системами, які не завжди встигають оперативно реагувати на зміни в тиску і температурі пари, що може призвести до неефективного використання палива і зниження стабільності роботи турбін. Використання нечіткої логіки

дозволить прискорити реакцію підконтролера на зміни в умовах роботи турбін, що значно покращить їхню ефективність.

Постановка завдання. Основною метою цього дослідження є вивчення можливостей застосування нечіткої логіки для підвищення ефективності управління складними технологічними об'єктами. Особливу увагу приділено розробці адаптивних моделей управління, що можуть обробляти нечіткі дані і приймати рішення на основі кількох взаємопов'язаних параметрів. Дослідження охоплює аналіз переваг нечіткої логіки над традиційними підходами, а також її потенціал у забезпеченні стабільності та оптимальності роботи технічних систем, що працюють в умовах швидких змін. Моделі на основі нечіткої логіки можуть бути застосовані для вдосконалення існуючих методів управління турбіною на об'єктах, де швидкість реакції на зміни параметрів може суттєво підвищити ефективність роботи обладнання та зменшити витрати ресурсів.

У рамках роботи також буде проведено порівняння традиційних підходів і моделей на основі нечіткої логіки для визначення їхніх переваг та обмежень у різних умовах. Завдяки використанню нечіткої логіки, можливо значно покращити адаптивність, зменшити коливання в параметрах, що критично важливо для таких об'єктів, як енергетичні станції, де стабільність і ефективність роботи є вирішальними факторами для досягнення високих результатів.

Це дослідження має на меті забезпечити наукову основу для впровадження інноваційних методів управління в промислових і енергетичних об'єктах, сприяючи їх модернізації та підвищенню енергоефективності.

Виклад основного матеріалу. У цьому контексті розробка та впровадження нових методів управління, зокрема на основі нечіткої логіки, відкриває можливості для створення більш гнучких і ефективних систем, що здатні адаптуватися до змінюваних умов роботи без необхідності точного математичного опису всіх параметрів. Для цього було обрано процес управління турбіною на теплоелектроцентралі (Далі - ТЕЦ), оскільки він є важливою частиною енергетичної інфраструктури та працює в умовах, де необхідно швидко реагувати на зміни навантаження і параметрів.

Турбіна ТЕЦ була обрана в якості об'єкта дослідження через її критичне значення для енергетичної інфраструктури та необхідність у стабільному регулюванні параметрів, таких як температура котлів, тиск пари та навантаження на турбіни. Система управління турбіною включає компо-

ненти, що працюють у складних і змінних умовах. Одним з основних компонентів, що потребує покращення, є система управління турбіною, де традиційні ПІД-регулятори не завжди встигають швидко реагувати на зміни в параметрах.

У зв'язку з цим, було прийнято рішення розробити модель управління на основі нечіткої логіки, яка дозволяє швидко адаптувати параметри регулювання, зокрема, температури та тиску, і забезпечити стабільну роботу в умовах, коли зміни є значними і швидкими.

Одним із основних підходів до вирішення задач управління складними технологічними об'єктами є використання нечіткої логіки, що базується на фазифікації вхідних параметрів і нечітких правилах для прийняття рішень. Фазифікація – це процес перетворення чітких величин (наприклад, температури чи тиску) у нечіткі множини, що описують ці параметри через категорії, як "низька", "середня" або "висока". Це дозволяє системі адаптуватися до змін вхідних умов без необхідності точного математичного опису кожного параметра. Нечіткі правила для прийняття рішень дозволяють враховувати різні комбінації вхідних параметрів і забезпечують гнучкість і точність управлінських рішень.

Порівняно з традиційними методами управління, система нечіткої логіки має кілька переваг:

- адаптивність: вона здатна швидко реагувати на зміни вхідних параметрів, таких як температура або тиск.
- гнучкість: нечітка логіка не вимагає точних математичних моделей і може працювати з розмитими даними.

Математична модель управління, побудована на основі нечіткої логіки, включає кілька етапів, кожен з яких сприяє оптимізації процесів управління. На першому етапі відбувається фазифікація вхідних параметрів, таких як температура або тиск, у нечіткі множини. Наприклад, температура котла може бути переведена в категорії "низька", "середня" або "висока" з відповідними ступенями належності.

В математичній моделі управління використовується формула фазифікації для температури:

$$\mu_{\text{низька}}(T) = \max\left(0, \frac{T_{\max} - T}{T_{\max} - T_{\min}}\right), \quad (1)$$

де T – поточна температура, $T_{\min}$ та $T_{\max}$ – мінімальне та максимальне значення температури відповідно.

Наступним етапом є нечіткі правила для прийняття рішень. Це можуть бути такі правила:

Якщо температура висока і тиск низький, то зменшити подачу палива.

Якщо навантаження високе і тиск середній, то збільшити потужність турбіни.

Нарешті, відбувається дефазифікація, що включає перетворення нечітких результатів в чіткі команди для управління котлом і турбіною за допомогою методу центру ваги через формулу:

$$y = \frac{\sum_i \mu_i \cdot x_i}{\sum_i \mu_i}, \quad (2)$$

де μ_i – ступінь належності, x_i – чітке значення.

Це дозволяє отримати оптимальні та стабільні результати управління.

Для перевірки ефективності запропонованої моделі управління на основі нечіткої логіки було проведено симуляцію роботи системи, яка враховує змінні параметри, характерні для турбіни ТЕЦ. Симуляція проводилась за допомогою MATLAB і включала використання модуля Fuzzy Logic Toolbox, який дозволяє моделювати складні системи в умовах, максимально наближених до реальних. Це дозволило оцінити, як система реагує на зміну температури, тиску [8] та навантаження [3] і наскільки ефективно вона адаптується до цих змін.

Одним із ключових аспектів симуляції було вивчення, як нечіткі множини описують температуру котла. Використання фазифікації дозволяє віднести значення температури до категорій "низька", "середня" та "висока", причому кожна з цих категорій характеризується своїми функціями належності. Для наочності цієї концепції наводиться на рис. 1.

Низька температура (синя лінія):

Ступінь належності μ дорівнює 1, якщо температура нижче або дорівнює 200°C, і зменшується до 0 при температурі 300°C. Наприклад, при температурі 250°C ступінь належності до "низької температури" буде 0.5.

Середня температура (зелена лінія):

Ступінь належності збільшується від 0 при 200°C, досягає максимуму (1) при 300–400°C і зменшується до 0 при 500°C. Це показує, що "середня температура" характеризує значення в цьому діапазоні.

Висока температура (червона лінія):

Ступінь належності дорівнює 0 при температурах нижче 400°C, зростає до 1 при 500°C. Це означає, що температура вище 400°C розглядається як "висока".

Цей графік показує, як значення температури поділяються на три нечіткі множини: "низька темпе-

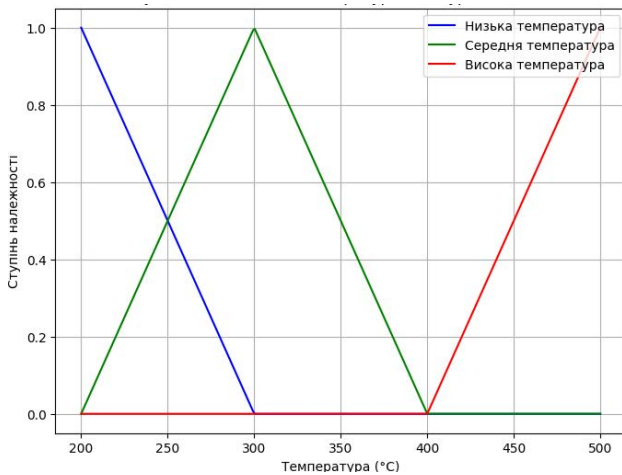


Рис. 1. Графік функції належності температури для турбіни ТЕЦ

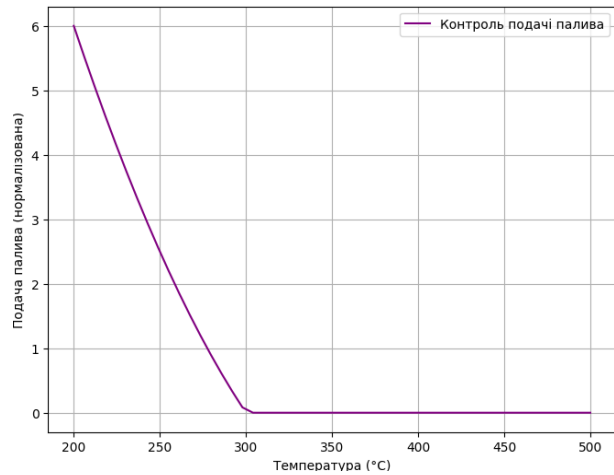


Рис. 2. Графік симуляції контролю подачі палива для турбіни ТЕЦ

ратура", "середня температура" та "висока температура". Для кожної точки температури визначається ступінь належності до однієї або кількох множин.

Далі симуляція включала тестування контролю подачі палива, де система мала адаптувати подачу залежно від змін температури та тиску. Основною метою було забезпечити стабільність роботи котла і турбіни, мінімізуючи витрати палива та уникаючи коливань параметрів, які можуть призвести до неефективної роботи обладнання. На основі моделювання результатів роботи системи отримано рис. 2.

При низьких значеннях температури ($T < 300^{\circ}\text{C}$) і високому тиску ($P > 12$), подача палива зменшується. Це пояснюється тим, що система вважає такі умови близькими до оптимальних і не потребує додаткового палива.

При середніх значеннях температури ($300^{\circ}\text{C} < T < 400^{\circ}\text{C}$) і середньому тиску ($P \approx 10$), подача палива стабільна. Це вказує на те, що система підтримує рівновагу.

При високих температурах ($T > 400^{\circ}\text{C}$) і низькому тиску ($P < 10$), подача палива різко збільшується. Це необхідно для стабілізації параметрів роботи системи.

Цей графік демонструє залежність контролю подачі палива від двох основних параметрів: температури T і тиску P . Лінія на графіку відображає, як система управління змінює подачу палива залежно від поточних значень температури та тиску.

Висновки. Впровадження системи управління на основі нечіткої логіки для турбіни на ТЕЦ дозволить значно підвищити ефективність та стабільність роботи цього обладнання. Це забезпечить зниження витрат палива, підвищить економічну ефективність і зменшить негативний вплив на навколишнє середовище завдяки зменшенню викидів шкідливих газів. Система, побудована на основі нечіткої логіки, дозволяє ефективно реагувати на змінні параметри, що робить її важливим інструментом для майбутньої модернізації енергетичних об'єктів.

Список літератури:

1. Ross T. J. Fuzzy Logic with Engineering Applications. Wiley, 2010. – 648 с.
2. Ogata K. Modern Control Engineering. Prentice Hall, 2009. – 912 с.
3. Palm R. Fuzzy Control in Industrial Automation: Theory and Applications. Springer, 2000. – 472 с.
4. Gupta M. M., Saroj K. Introduction to Fuzzy Sets, Fuzzy Logic, and Fuzzy Control Systems. CRC Press, 1999. – 356 с.
5. Steam Turbine: Working Principle, Types, and Applications. URL: <https://mechanicalboost.com/steam-turbine-working-principle-types-and-applications/>
6. Steam Power Plant Configuration, Design, and Control. URL: https://web.ecs.baylor.edu/faculty/lee/ELC5362/Lecture%20note/H1-2015TPS_Review_Kwang.pdf
7. Steam Turbine Working Principle. URL: <https://www.triveniturbines.com/steam-turbine-working-principle-working-of-a-steam-turbine-steam-power-plant-working-principle-of-steam-turbine-steam-turbine-working-model-working-of-steam-turbine-power-plant/>
8. Steam Turbines. URL: <https://testbook.com/mechanical-engineering/steam-turbines>

Nalotov D.R., Liashenko O.S. COMPLEX TECHNOLOGICAL OBJECT CONTROL MODEL BASED ON FUZZY LOGIC METHODS

The paper considers the problem of improving the efficiency of controlling complex technological objects under conditions of variable parameters typical for modern industrial and energy systems. It is noted that the use of traditional approaches, particularly PID controllers, does not always provide the required adaptability and response speed to changes in technological process parameters, which justifies the feasibility of applying fuzzy logic methods for building more flexible and intelligent control systems. The object of the study is the process of controlling a turbine at a combined heat and power plant (CHP), which includes boilers, turbines, and auxiliary equipment operating under conditions of variable loads, temperature, and steam pressure. The subject of the research is adaptive control models based on fuzzy logic capable of taking into account the uncertainty of input data and forming flexible control decisions based on fuzzy rules. As part of the research, a mathematical control model was developed based on the use of fuzzification of input parameters, a system of fuzzy decision rules, and the defuzzification procedure for generating control signals. This approach makes it possible to transform precise values of technological process parameters, such as steam temperature and pressure, into fuzzy sets describing the system state in the form of linguistic variables such as “low”, “medium”, or “high”.

Special attention is paid to the formation of membership functions and the development of fuzzy control rules that consider the relationship between boiler temperature, steam pressure, and turbine load. To verify the effectiveness of the proposed model, computer simulations were carried out in the MATLAB environment using the Fuzzy Logic Toolbox module. During the simulation, the response of the control system to changes in the main technological parameters was studied, and the system behavior during fuel supply regulation depending on temperature and pressure was analyzed.

The obtained simulation results confirmed that the proposed fuzzy logic-based model makes it possible to increase the stability of equipment operation, reduce fluctuations in technological process parameters, and decrease the response time of the control system to changing operating conditions. In addition, the use of fuzzy logic contributes to reducing fuel consumption, improving energy efficiency, and decreasing the negative environmental impact of energy facilities. The proposed approaches can be used for modernization of control systems at energy enterprises and can also be adapted to other industrial systems operating under conditions of uncertainty and dynamically changing parameters.

Keywords: fuzzy logic, adaptive control, CHP, fuzzification, system stability, turbine.

Дата першого надходження статті до видання: 06.03.2026

Дата прийняття статті до друку після рецензування: 01.04.2026

Дата публікації (оприлюднення) статті 11.05.2026

УДК 681.5:629.8

DOI DOI DOI <https://doi.org/10.32782/2663-5941/2026.2.1/30>**Невлюдов І.Ш.**<https://orcid.org/0000-0002-9837-2309>

Харківський національний університет радіоелектроніки

Новоселов С.П.<https://orcid.org/0000-0002-3190-0592>

Харківський національний університет радіоелектроніки

Сичова О.В.<https://orcid.org/0000-0002-0651-557X>

Харківський національний університет радіоелектроніки

Теслюк С.І.<https://orcid.org/0000-0003-0711-9250>

Харківський національний університет радіоелектроніки

РОЗПОДІЛЕНА СИСТЕМА ПОЗИЦІОНУВАННЯ МОБІЛЬНИХ РОБОТІВ НА ОСНОВІ МЕРЕЖІ ARUCO-СЕНСОРІВ ТА АРХІТЕКТУРИ ROS ДЛЯ КІБЕРФІЗИЧНОГО ВИРОБНИЦТВА

У даній роботі розроблено математичний апарат для опису розподіленої системи позиціонування мобільних роботів у виробничому приміщенні довільної конфігурації, що відповідає вимогам інтелектуального виробництва та кіберфізичних систем. Запропонована система базується на мережі периферійних обчислювальних вузлів візуальної навігації, кожен з яких здійснює визначення локальних координат мобільного робота у власній системі відліку за допомогою маркерів типу ArUco та засобів комп'ютерного зору. Розроблено метод узгодження координат, що забезпечує перетворення положення робота з локальних систем координат окремих сенсорних модулів у єдину глобальну систему координат виробничого простору. Для цього використано однорідні матриці трансформації групи $SE(2)$, які дозволяють формалізувати результати калібрування кожного модуля та забезпечити коректне геометричне поєднання вимірювань у спільному координатному просторі. Розглянуто задачу обробки вимірювальної інформації у зонах перекриття полів зору сусідніх камерних модулів. Для підвищення точності позиціонування запропоновано метод злиття координат на основі зваженої оцінки за критерієм мінімуму дисперсії, що дозволяє враховувати різну точність окремих спостережачів та забезпечує стійкість результату при зміні конфігурації системи. Показано, що запропонований математичний апарат є масштабованим і може бути застосований для довільної кількості сенсорних модулів та довільної геометрії виробничого приміщення. Отримані результати утворюють теоретичну основу для побудови розподіленої системи глобального позиціонування мобільних роботів та реалізації центрального координатора у середовищі ROS, що забезпечує інтеграцію сенсорних вузлів, обмін даними у реальному часі та можливість подальшого розширення функціональних можливостей системи.

Ключові слова: мобільні роботи, ArUco-маркери, розподілена локалізація, ROS, перетворення координат, кіберфізичні системи, Industry 4.0, Industry 5.0.

Постановка проблеми. Сучасна еволюція виробничих систем у межах концепцій Industry 4.0 та 5.0 характеризується відмовою від лінійної автоматизації на користь гнучких структур, здатних до реконфігурації [1]. Це передбачає впровадження адаптивних автономних мобільних роботів, які завдяки розподіленим системам пози-

ціонування та інтелектуальному керуванню здатні до динамічної навігації та узгодженої колективної взаємодії в умовах складного, змінного виробничого середовища. Ефективне функціонування таких систем неможливе без вирішення задачі визначення просторового положення кожного робота у режимі реального часу – задачі, що

є фундаментальною для планування траєкторій, уникнення колізій та координації групових дій [2].

Існуючі підходи до локалізації мобільних роботів у виробничому середовищі можна розділити на декілька класів:

- методи на основі лазерної локації (LiDAR, SLAM) [3];
- радіочастотні системи позиціонування (UWB, RFID) [4, 5];
- ультразвукові системи позиціонування [6];
- оптичні системи позиціонування [7];
- системи позиціонування на основі технічного зору [8, 9].

Кожен із зазначених підходів має свої обмеження. Лазерні системи є дорогими та вимогливими до обслуговування. Ультразвукові й радіочастотні методи мають низьку точність за наявності перешкод. Бортові системи комп'ютерного зору потребують значних обчислювальних ресурсів і є вразливими до змін освітлення на підлозі та забруднення датчиків.

Принципово відмінним підходом є організація позиціонування на основі розподіленої мережі зовнішніх камер з використанням маркерів ArUco, нанесених на верхню поверхню роботів. Такий підхід дозволяє перенести обчислювальне навантаження з платформи робота на зовнішню інфраструктуру, спростити конструкцію самого робота та забезпечити незалежне, централізовано кероване відстеження множини об'єктів одночасно. Проте реалізація подібної архітектури пов'язана з низкою невирішених науково-технічних проблем.

По-перше, виникає задача побудови розподіленої системи з зонами видимості камер, що частково перекриваються, кожна з яких функціонує у власній локальній системі координат. Необхідна розробка методів узгодження локальних систем відліку у єдину глобальну модель простору виробничого приміщення з урахуванням геометричних спотворень, похибок калібрування та часових затримок між вузлами мережі. По-друге, постає задача ефективної передачі та агрегації даних від множини обчислювальних вузлів (наприклад, на базі Raspberry Pi) до глобального координатора на платформі ROS (Robot Operating System) з мінімальними затримками та без втрати цілісності інформації. По-третє, при переміщенні робота із зони видимості одного вузла до зони видимості іншого, або у зону перекриття, необхідно забезпечити безперервність позиційного відстеження – вирішити задачу безшовної передачі координат між вузлами мережі.

Таким чином, запропонована в даній роботі технологія локального позиціонування інтегрує низку пріоритетних напрямів сучасної робототехніки: децентралізовані методи комп'ютерного зору, ієрархічні структури ROS, а також алгоритми візуального позиціонування в мультикамерних системах. Наукова новизна дослідження полягає у розвитку теоретичних засад розподіленої локалізації. Практична цінність зумовлена можливістю створення масштабованих і відмовостійких систем керування групами мобільних роботів в умовах динамічного виробничого середовища.

Аналіз останніх досліджень і публікацій.

В роботі Луан Карлос Кляйна [10] запропоновано метод локалізації на основі машинного навчання з використанням маркерів ArUco. Основна ідея полягає у визначенні відносного положення бортової камери щодо реперних міток (ArUcos) з подальшою оцінкою позиції робота за допомогою алгоритмів машинного навчання. Найкращі результати, з похибкою на міліметровому рівні, було отримано завдяки використанню регресії випадкового лісу. Архітектура роботів у зазначеному дослідженні базується на двох основних обчислювальних пристроях – Raspberry Pi відповідає за процес прийняття рішень та керування RGB-камерою, Arduino забезпечує роботу апаратної частини, зокрема двигунів та енкодерів.

Алгоритми оцінки просторового положення на основі реперних графічних міток привертають значну увагу розробників мобільних робототехнічних систем завдяки простоті впровадження та економічній ефективності. У дослідженні, в складі авторів Доянг Кім, Дже Хван Бонг та Сонкюн Чон [11], наведено порівняльний експериментальний аналіз точності систем захоплення руху на базі ArUco відносно комерційного еталона OptiTrack. Результати підтверджують, що системи на основі ArUco забезпечують достатній рівень точності як у статичних, так і в динамічних сценаріях у межах різних зон видимості сенсорів.

Розвиток цього підходу представлений у роботі "Експериментальний аналіз оцінки пози на основі маркерів ArUco" [12], де запропоновано алгоритм одночасного використання декількох маркерів. Розроблена авторами архітектура включає механізми верифікації просторової узгодженості для виявлення некоректних ідентифікаторів у масиві маркерів, а також аналіз часової стабільності для відсікання аномальних вимірювань. Таке рішення дозволяє суттєво підвищити якість оцінки просторового положення робота в складних виробничих умовах.

У статті "Оптимізація розміщення опорних маркерів для покращеної візуальної локалізації" [13] розв'язано задачу оптимального розміщення реперних маркерів для підвищення якості візуальної локалізації. Оптимізація розташування маркерів має високу цінність, оскільки процес їх інсталяції у великих приміщеннях є трудомістким, а підсумкове покращення продуктивності системи критично залежить від обраних позицій. Відповідно до досліджень цей результат є безпосередньо релевантним для проєктування зон перекриття камерних вузлів у запропонованій архітектурі, оскільки дозволяє мінімізувати «сліпі зони» та забезпечити безперервність трекінгу.

Таким чином, питання побудови розподілених систем візуальної локалізації, де позиціонування здійснюється розподіленою мережею зовнішніх модулів обробки візуальних даних на базі Raspberry Pi з подальшою інтеграцією в екосистему ROS, залишається недостатньо висвітленим. Зокрема, актуальною є задача синтезу просторових координат у єдиній системі відліку від просторово рознесених камер із взаємним перекриттям зон огляду, що є необхідною умовою для забезпечення безперервної координації груп мобільних роботизованих платформ у складних промислових середовищах.

Постановка завдання. Метою роботи є розробка моделі розподіленої системи позиціонування мобільних роботів для інтелектуального виробництва на основі мережі ArUco-сенсорів та архітектури ROS.

Виклад основного матеріалу

Архітектура автоматизованої системи

Запропонована архітектура вирішує описані вище проблеми шляхом розгортання розподіленої мережі периферійних обчислювальних вузлів візуальної навігації, кожен з яких є автономним модулем. Роботи оснащуються ArUco-маркерами на верхній поверхні, що дозволяє визначати їх позицію при мінімальних обчислювальних витратах. Уся система реалізується через платформу ROS2 із використанням механізму DDS Middleware для розподіленої передачі даних.

Архітектура автоматизованої системи подана на рис. 1. Система складається з двох логічних рівнів: рівня периферійних модулів (edge-рівень) та рівня глобального координатора (master-рівень). Взаємодія між рівнями здійснюється через розподілену шину повідомлень ROS2 DDS Middleware, що забезпечує публікацію та підписку без потреби у центральному брокері.

На рівні периферії розгорнуто N незалежних модулів, кожен з яких включає відеокамеру та одноплатний комп'ютер Raspberry Pi. Модулі встановлюються у фіксованих точках виробничого простору

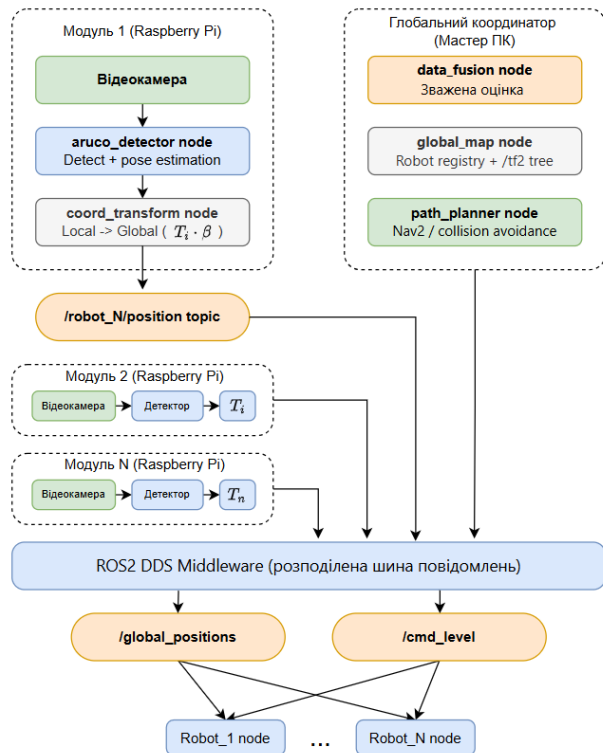


Рис. 1. Архітектура автоматизованої системи

з відомими координатами та кутами орієнтації, що визначаються під час калібрування системи.

Кожен периферійний обчислювальний вузол візуальної навігації має трирівневий конвеєр обробки у вигляді послідовного ланцюжка ROS2-вузлів.

Перший вузол (aruco_detector node) відповідає за захоплення відеопотоку з камери та детекцію ArUco-маркерів в кожному кадрі. За допомогою бібліотеки OpenCV та алгоритму ArUco визначається піксельне положення чотирьох кутових точок маркера, після чого обчислюються вектори трансляції та ротації, що описують позу маркера у системі координат камери. Результатом роботи вузла є вектор вимірювання $m_i = (u_i, v_i, 1)^T$ у локальних координатах, а також ідентифікатор робота, закодований у маркері.

Другий вузол (coord_transform node) виконує перетворення локальних координат ArUco-детекції у глобальні координати виробничого простору. Для цього використовується попередньо обчислена однорідна матриця трансформації T_i групи SE(2), що враховує положення модуля (x_i^G, y_i^G) та кут орієнтації камери τ_i відносно глобальної системи відліку. Перетворення здійснюється за формулою:

$$p^G = \frac{T_i}{P_i(K, h_i)} m_i, \quad (1)$$

де K – матриця інтринсіків камери; h_i – висота підвісу.

Результат публікується у топик `/robot_N/position` типу `PoseWithCovarianceStamped` разом з оцінкою дисперсії похибки σ_i^2 , що обчислюється адаптивно на основі якісних метрик детекції.

Усі периферійні обчислювальні вузли візуальної навігації та глобальний координатор взаємодіють через шину ROS2 DDS. Ключовою перевагою DDS є відсутність потреби у центральному брокері – кожен вузол може публікувати та підписуватись на топіки безпосередньо, що забезпечує горизонтальну масштабованість системи. Додавання нового модуля зводиться до реєстрації нового публікатора у мережі без жодних змін у конфігурації існуючих вузлів.

Кожен модуль публікує повідомлення структури $M_i(t) = (id_{robot}, p_i^G(t), \sigma_i^2(t), t_{stamp})$, де t_{stamp} – мітка часу, критична для коректної синхронізації асинхронних вимірювань від різних вузлів при злитті даних. Глобальний координатор є центральним обчислювальним вузлом системи, що виконує три функції: злиття вимірювань, ведення глобальної карти та планування руху.

Вузол `data_fusion node` підписується на топіки усіх периферійних обчислювальних модулів та виконує злиття даних методом зваженої оцінки мінімальної дисперсії (BLUE – Best Linear Unbiased Estimator). Для M одночасно активних модулів, що спостерігають одного робота, злита оцінка обчислюється як:

$$\hat{p}^G = \left(\sum_{k=1}^M w_k \right)^{-1} \sum_{k=1}^M w_k \hat{p}_k^G, \quad w_k = \frac{1}{\sigma_k^2}. \quad (2)$$

Вага кожного модуля обернено пропорційна його дисперсії: більш точне вимірювання отримує більший вплив на підсумкову оцінку. BLUE-оцінка є оптимальною лінійною незміщеною оцінкою за умови незалежності похибок різних модулів та відсутності систематичних похибок калібрування.

Вузол `global_map node` веде реєстр поточних глобальних позицій усіх роботів та підтримує дерево трансформацій `/tf2`, що є стандартним механізмом ROS2 для опису просторових відношень між системами координат. Публікація здійснюється у топик `/global_positions`, доступний усім підписникам у мережі.

На основі глобальної карти вузол `path_planner node` генерує команди руху для кожного робота через топик `/cmd_vel`. Планувальник реалізується на базі стеку Nav2 (Navigation2), що є стандартним рішенням для автономної навігації у ROS2, та включає алгоритми уникнення колізій між роботами у спільному просторі.

Кожен мобільний робот представлений у мережі як окремий ROS2-вузол (Robot_n node), що підписується одночасно на два топіки: `/global_positions` – для отримання поточної глобальної позиції та відносного розміщення інших роботів, та `/cmd_vel` – для отримання команд руху від планувальника. Така архітектура дозволяє роботам діяти як у повністю автономному режимі (виконувати команди планувальника), так і в телеоперованому режимі, використовуючи глобальну позицію як зворотній зв'язок.

Моделювання розподіленої системи позиціонування мобільних роботів на основі ArUco-маркерів та архітектури ROS

На рис. 2 подана схема виробничого приміщення та принцип розміщення периферійних обчислювальних вузлів візуальної навігації в глобальній системі координат.

Для однозначного опису положення будь-якого об'єкта у виробничому приміщенні вводиться глобальна декартова система координат $F_G = \{O_G, X_G, Y_G\}$, де O_G – фіксований початок відліку (наприклад, один із кутів приміщення), осі X_G та Y_G орієнтовані вздовж двох взаємно перпендикулярних стін. Вибір такої системи є природним для прямокутних виробничих приміщень і спрощує інтерпретацію результатів позиціонування операторами.

Положення мобільного робота у глобальній системі координат однозначно задається двовимірним вектором:

$$p^G = (x^G, y^G)^T \in R^2, \quad (3)$$

де x^G та y^G – координати об'єкта вздовж осей X_G та Y_G відповідно.

Оскільки камери кріпляться на стелі або на конструкціях над рівнем руху роботів і спрямовані вертикально вниз, задача позиціонування розглядається як плоска (двовимірна), що є стандартним спрощенням для систем навігації наземних мобільних роботів. При необхідності врахування висоти (наприклад, для різнорівневих виробничих майданчиків) формалізм природно узагальнюється на тривимірний випадок із застосуванням матриць з групи SE(3).

Мережа складається з N модулів, де N визначається площею та конфігурацією виробничого приміщення, а також необхідним рівнем покриття (рис. 2). Кожен модуль з індексом i ($i = 1, 2, \dots, N$) фізично закріплений у фіксованій точці простору та має власну локальну систему координат $F_i = \{O_i, X_i, Y_i\}$, в якій він виробляє вимірювання.

Конфігурація модуля у глобальній системі координат задається двома параметрами.

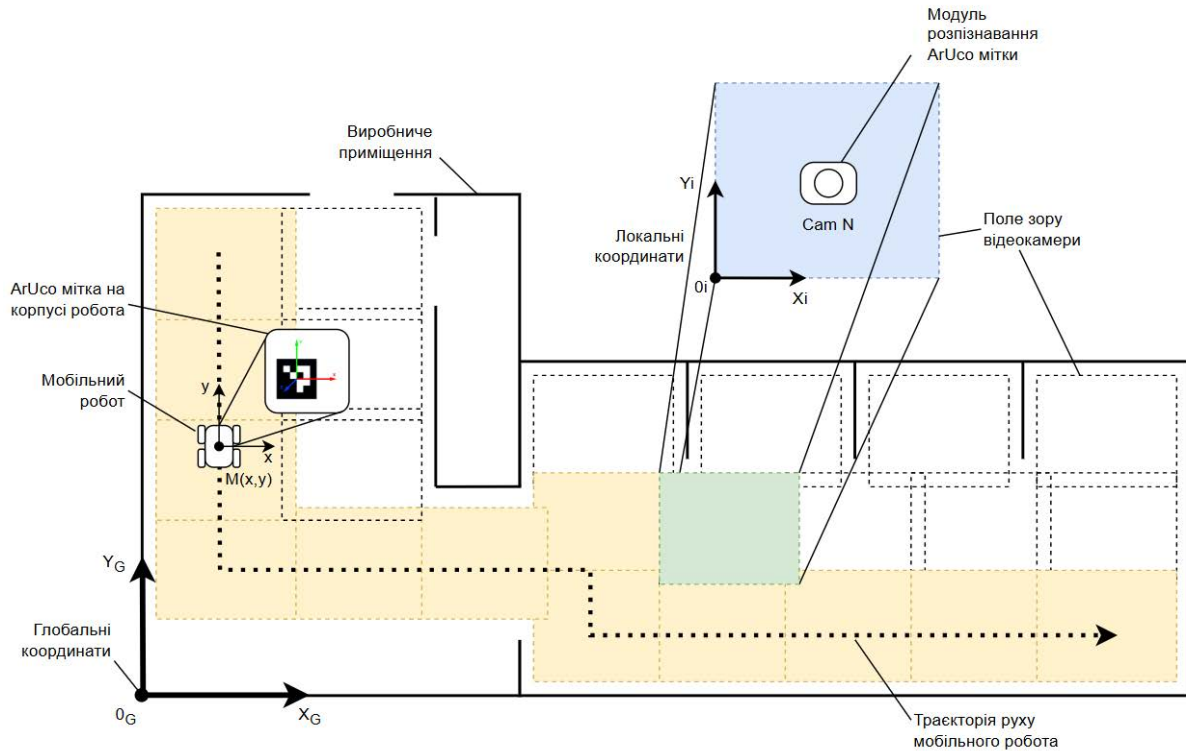


Рис. 2. Схема виробничого приміщення та принцип розміщення периферійних обчислювальних вузлів візуальної навігації в глобальній системі координат

По-перше, вектором трансляції, що визначає положення оптичного центру камери:

$$t_i^G = (x_i^G, y_i^G)^T. \quad (4)$$

По-друге, скалярним кутом орієнтації θ_i – кутом між напрямком локальної осі X_i модуля та глобальною віссю X_G . Цей кут може бути відмінним від нуля в разі, якщо камера встановлена під кутом до стін приміщення, що є поширеною практичною ситуацією:

$$\theta_i \in [0, 2\pi).$$

Параметри t_i^G та θ_i є результатом калібрування системи і надалі вважаються відомими константами. Процедура калібрування може бути виконана, наприклад, шляхом розміщення робота з відомими глобальними координатами у полі зору кожної камери та порівняння виміряних локальних координат з еталонними глобальними.

Перетворення між двома евклідовими системами координат на площині, що відрізняються поворотом та трансляцією (але не масштабуванням і не відображенням), належить до класу жорстких (ізометричних) перетворень. Множина всіх таких перетворень утворює групу Лі SE(2) – спеціальну евклідову групу площини. Елемент цієї групи, що

описує позицію модуля i , задається однорідною матрицею трансформації розміром 3×3 :

$$T_i = \begin{bmatrix} \cos \theta_i & -\sin \theta_i & x_i^G \\ \sin \theta_i & \cos \theta_i & y_i^G \\ 0 & 0 & 1 \end{bmatrix}. \quad (5)$$

Використання однорідних координат є стандартним прийомом у комп'ютерному зорі та робототехніці, що дозволяє записати афінне перетворення у вигляді єдиного матричного добутку. Вектор локальних координат робота, що спостерігається модулем i , в однорідній формі записується як:

$$\tilde{p}_i^L = (x_i^L, y_i^L, 1)^T. \quad (6)$$

Тоді перетворення до глобальних координат виконується як:

$$\tilde{p}^G = T_i \tilde{p}_i^L. \quad (7)$$

Розкриваючи матричний добуток, отримуємо явні аналітичні формули перетворення координат:

$$x^G = x_i^L \cos \theta_i - y_i^L \sin \theta_i + x_i^G, \quad (8)$$

$$y^G = x_i^L \sin \theta_i + y_i^L \cos \theta_i + y_i^G.$$

Геометричний зміст цих формул є наступним: перша частина кожного виразу: $x_i^L \cos \theta_i - y_i^L \sin \theta_i$ та $x_i^L \sin \theta_i + y_i^L \cos \theta_i$ – виконує поворот вектору локальних координат на кут θ_i , приводячи його

до орієнтації глобальної системи; друга частина: x_i^G та y_i^G – здійснює зсув початку відліку від локального до глобального. Обернена матриця T_i^{-1} використовується для перетворення у зворотному напрямку – з глобальної системи до локальної, що необхідно, зокрема, для перевірки належності точки до зони видимості модуля.

ArUco-маркер, нанесений на верхню поверхню робота, спостерігається камерою модуля i як двовимірне зображення. Алгоритм ArUco (реалізований у бібліотеці OpenCV) виявляє чотири кутові точки маркера на знімку та повертає їх піксельні координати, а також обчислює вектори трансляції та ротації, що описують позу маркера у системі координат камери. На виході детектора отримуємо вектор вимірювання:

$$m_i = (u_i, v_i, 1)^T, \quad (9)$$

де u_i та v_i – піксельні координати проєкції центра маркера на площину зображення.

Перехід від піксельних координат до метричних координат у площині підлоги (локальна система координат модуля) здійснюється через модель перспективної камери. Внутрішні параметри камери описуються матрицею інтринсиків K :

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \quad (10)$$

де f_x , f_y – фокусні відстані в пікселях по осях x та y ; c_x , c_y – координати головної точки (оптичного центра) на матриці.

Зворотна проєкція з урахуванням відомої висоти підвісу камери h_i над рівнем підлоги виконується за формулою:

$$p_i^L = \frac{1}{K} \lambda m_i, \quad (11)$$

де λ – масштабний коефіцієнт, що обчислюється як відношення висоти підвісу h_i до ефективної фокусної відстані f (при квадратних пікселях $f = f_x = f_y$):

$$\lambda = \frac{h_i}{f}.$$

Іншими словами λ – це масштаб зображення у метрах на піксель на рівні підлоги. Ця формула є справедливою за умови, що площина маркера паралельна площині підлоги, а камера орієнтована строго вертикально. Для компенсації відхилень від цієї умови застосовуються процедури дисторсії та компенсації нахилу, що входять до стандартної процедури калібрування камери.

Кожна камера має обмежене поле зору, що проєктується на площину підлоги у вигляді дея-

кої двовимірної області – зони видимості модуля. Для камери з прямокутною матрицею та вертикальним підвісом ця область є прямокутником у локальних координатах. В загальному випадку зону видимості модуля i у глобальній системі координат визначимо як множину точок:

$$Z_i = \{p^G \in R^2 \mid T_i^{-1} p^G \in D_i\}, \quad (12)$$

де D_i – прямокутна або довільна опукла область допустимих локальних координат, що визначається кутом поля зору камери та висотою підвісу.

Сукупність усіх зон видимості утворює покриття виробничого простору:

$$W \subseteq \bigcup_{i=1}^N Z_i, \quad (13)$$

де W – робочий простір виробничого приміщення.

Умова повного покриття W є необхідною для гарантованого відстеження роботів у будь-якій точці простору. Зона перекриття між двома сусідніми модулями i та j визначається як:

$$Z_{ij} = Z_i \cap Z_j. \quad (14)$$

Перекриття $Z_{ij} \neq \emptyset$ є необхідною умовою для виконання двох важливих функцій системи. По-перше, воно дозволяє забезпечити безперервне передавання відповідальності за відстеження робота від одного модуля до іншого без втрати трекінгу у момент перетину межі зони видимості. По-друге, перекриття дозволяє виконати взаємне калібрування та верифікацію суміжних модулів: якщо обидва модулі одночасно спостерігають маркер у зоні Z_{ij} , їхні глобальні оцінки положення мають збігатися, що дозволяє виявляти та коригувати систематичні похибки калібрування.

Перебуваючи у зоні перекриття Z_{ij} , робот одночасно спостерігається двома (або більше) модулями. Кожен модуль формує власну незалежну оцінку глобальних координат:

$$\hat{p}_i^G = T_i \tilde{p}_i^L, \quad \hat{p}_j^G = T_j \tilde{p}_j^L. \quad (15)$$

Ці оцінки є різними внаслідок похибок вимірювання кожного модуля: шуму сенсора, похибок детекції ArUco (особливо при малих розмірах маркера в кадрі або при несприятливому освітленні), а також похибок калібрування. Дисперсія похибки вимірювання модуля i позначається σ_i^2 . Для отримання оптимальної у сенсі мінімуму середньоквадратичної похибки об'єднаної оцінки координат використовується зважене усереднення вимірювань, у якому вагові коефіцієнти вибираються обернено пропорційними дисперсіям похибок відповідних модулів. Такий підхід відповідає оцінці мінімальної дисперсії для незалежних

вимірювань. Злита оцінка глобальних координат визначається як:

$$\hat{p}^G = \frac{\sigma_j^2 \hat{p}_i^G + \sigma_i^2 \hat{p}_j^G}{\sigma_j^2 + \sigma_i^2}. \quad (16)$$

З фізичного змісту даного виразу виходить, що модуль із меншою дисперсією отримує більшу вагу в підсумковій оцінці. Наприклад, якщо $\sigma_i^2 \ll \sigma_j^2$, то $\hat{p}^G \approx \hat{p}_i^G$, тобто система довіряє переважно першому модулю. У загальному випадку для M одночасно активних модулів (робот знаходиться у полі зору кількох камер):

$$\hat{p}^G = \frac{\sum_{k=1}^M w_k \hat{p}_k^G}{\sum_{k=1}^M w_k}, \quad w_k = \frac{1}{\sigma_k^2}. \quad (17)$$

Ця формула є дискретним аналогом фільтра Калмана для статичного виміру і є оптимальною лінійною незміщеною оцінкою за умови незалежності похибок різних модулів. На практиці дисперсії σ_k^2 не є константами, а залежать від поточного стану спостереження: відстані від маркера до центра кадру, кута нахилу, рівня освітлення, ступеня оклюзії. Тому в реальній системі вони можуть обчислюватись адаптивно на основі якісних метрик детекції ArUco (наприклад, розміру маркера в пікселях або репроекційної похибки).

Повний обчислювальний ланцюжок від первинних пікселів ArUco-детектора до глобальних координат у системі відліку виробничого приміщення може бути записаний у компактному матричному вигляді:

$$p^G = \frac{T_i m_i}{P_i(K, h_i)}, \quad (18)$$

де $m_i = \begin{pmatrix} u_i \\ v_i \\ 1 \end{pmatrix}$ – вектор піксельних координат центра маркера, отриманих від ArUco-детектора;

$\frac{1}{P_i(K, h_i)}$ – оператор зворотної проєкції, що залежить від матриці інтринсиків K та висоти підвісу h_i , і переводить піксельні координати в метричні локальні координати підлоги; T_i – однорідна матриця жорсткого перетворення, що переводить локальні координати у глобальні.

Цей ланцюжок виконується незалежно та паралельно на кожному вузлу мережі (Raspberry Pi). Результати – вектори $\hat{p}_i^G$ та відповідні оцінки дисперсій σ_i^2 – публікуються в мережу ROS у вигляді повідомлень типу `geometry_msgs/PoseWithCovarianceStamped` і доступні глобальному координатору через механізм підписки на топіки. Глобальний вузол ROS виконує фінальне злиття даних згідно формули (17) та публікує уточнене положення кожного робота.

Висновки. В даній роботі запропоновано модель розподіленої системи позиціонування мобільних роботів на основі мережі ArUco-сенсорних модулів. Запропонований підхід включає: введення глобальної та локальних систем координат; формалізацію калібрувальних параметрів кожного модуля у вигляді матриці трансформації з групи $SE(2)$; модель зворотної проєкції для отримання метричних координат із пікселів; визначення зон видимості та умов їх перекриття; оптимальну зважену оцінку положення у зонах перекриття за критерієм мінімуму дисперсії.

Розроблений математичний апарат є масштабованим – кількість модулів може бути довільною, конфігурація виробничого приміщення – довільною, а алгоритм злиття даних може бути узагальнений на випадок декількох одночасно активних спостерігачів. Подальші дослідження спрямовані на розробку адаптивних алгоритмів оцінки дисперсій σ_k^2 у реальному часі та на дослідження впливу часових затримок у мережі ROS на точність глобального позиціонування.

Список літератури:

1. M. Soori et al. Intelligent Robotic Systems in Industry 4.0. A Review. *Journal of Advanced Manufacturing Science and Technology*. 2023. DOI: 10.51393/j.jamst.2024007
2. Şenbaşlar B., Hoenig W. and Ayanian N. RLSS: Real-time Multi-Robot Trajectory Replanning using Linear Spatial Separations. *Autonomous Robots*. vol. 47(7). pp. 1-26. 2021. DOI: <https://doi.org/10.48550/arXiv.2103.07588>
3. Nevlyudov I., Novoselov S. and Sukhachov K. Method of Simultaneous Localization and Mapping for Construction of 2.5D Maps of the Environment Using ROS. *Innovative technologies and scientific solutions for industries*. vol. 2 (24). pp. 145-160. 2023. DOI: <https://doi.org/10.30837/ITSSI.2023.24.145>
4. Novoselov S. and Donskov O. Distributed Local Positioning System Using DWM1000 Location Chip. *4th International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T)*. pp. 489-492. 2017. DOI: <https://doi.org/10.1109/INFOCOMMST.2017.8246445>
5. Nevlyudov I., Novoselov S., Sychova O. and Tesliuk S. Determining the Coordinates of a Mobile Robot in an Industrial Space Using BLE Technology Based on RSSI Data Received from Base Stations. *Radiotekhnika*. vol. 2 (209). pp. 185-191. 2022. DOI: <https://doi.org/10.30837/rt.2022.2.209.18>

6. Zou Jie-Tong and Wang C.-Y. The Application of Ultrasonic Positioning System for Indoor Navigation of Omni-directional Mobile Robot. *Sensors and Materials*. pp. 585-591. 2016. DOI: <https://doi.org/10.18494/SAM.2016.1212>
7. Ghidary S. S., Tani T., Takamori T. and Hattori M. A new home robot positioning system (HRPS) using IR switched multi ultrasonic sensors. *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics*. vol. 4. pp. 737-741. 1999. DOI: <https://doi.org/10.1109/ICSMC.1999.812496>
8. S. Meina et al. Vision-based positioning system. *The Journal of China Universities of Posts and Telecommunications*. vol. 23(5). pp. 88-96. 2016. DOI: [https://doi.org/10.1016/s1005-8885\(16\)60062-6](https://doi.org/10.1016/s1005-8885(16)60062-6)
9. Hamza M. S., Shehata O. M., Morgan E. I. and Elias C. M. Vision-Based Indoor Positioning System for Connected Vehicles in Small-scale Testbed Environments. *IEEE Intelligent Vehicles Symposium (IV)*. pp. 144-148. 2024. DOI: <https://doi.org/10.1109/IV55156.2024.10588856>
10. L. C. Klein et al. A Machine Learning Approach to Robot Localization Using Fiducial Markers in RobotAtFactory 4.0 Competition. *Sensors*. vol. 23(6). pp. 3128. 2023. DOI: <https://doi.org/10.3390/s23063128>.
11. Doyoung K., Jae B. and Seongkyun J. Enhancing Pose Estimation Using Multiple Graphical Markers with Spatial and Temporal Outlier Detection. *Applied Sciences*. vol. 14. pp. 10225. 2024. DOI: <https://doi.org/10.3390/app142210225>
12. Pulloquinga J.L., Corrata D., Mata V., Valera Á., Vallés M. Experimental Analysis of Pose Estimation Based on ArUco Markers. *Innovations in Industrial Engineering III. Lecture Notes in Mechanical Engineering. Springer*. 2024. pp 138-149. DOI: https://doi.org/10.1007/978-3-031-61582-5_12
13. Q. Huang et al. Optimizing Fiducial Marker Placement for Improved Visual Localization. *IEEE Robotics and Automation Letters*. vol. 8(5). pp. 2756-2763. 2023. DOI: <https://doi.org/10.1109/LRA.2023.3260700>

Nevliudov I.Sh., Novoselov S.P., Sychova O.V., Tesliuk S.I. DISTRIBUTED POSITIONING SYSTEM FOR MOBILE ROBOTS BASED ON ARUCO SENSOR NETWORK AND ROS ARCHITECTURE FOR CYBER-PHYSICAL MANUFACTURING

This paper presents a mathematical framework for describing a distributed positioning system for mobile robots in a production facility of arbitrary configuration, meeting the requirements of smart manufacturing and cyber-physical systems. The proposed system is based on a network of peripheral computing nodes for visual navigation, each of which determines the local coordinates of a mobile robot in its own reference frame using ArUco-type markers and computer vision techniques. A coordinate alignment method has been developed that transforms the robot's position from the local coordinate systems of individual sensor modules into a single global coordinate system of the production space. For this purpose, homogeneous transformation matrices of the SE(2) group are used, which allow formalizing the calibration results of each module and ensuring the correct geometric combination of measurements in a common coordinate space. The problem of processing measurement information in areas where the fields of view of adjacent camera modules overlap is considered. To improve positioning accuracy, a coordinate fusion method based on a weighted evaluation using the minimum variance criterion is proposed, which allows for the varying accuracy of individual sensors and ensures the stability of the result when the system configuration changes. It is shown that the proposed mathematical framework is scalable and can be applied to an arbitrary number of sensor modules and an arbitrary geometry of the production facility. The obtained results for the theoretical basis for developing a distributed global positioning system for mobile robots and implementing a central coordinator within the ROS environment, which ensures the integration of sensor nodes, real-time data exchange, and the ability to further expand the system's functionality.

Keywords: mobile robots, ArUco markers, distributed localization, ROS, coordinate transformation, cyber-physical systems, Industry 4.0, Industry 5.0.

Дата першого надходження статті до видання: 19.03.2026
Дата прийняття статті до друку після рецензування: 15.04.2026
Дата публікації (оприлюднення) статті 11.05.2026

Oleshchenko L.M.<https://orcid.org/0000-0001-9908-7422>

National Technical University of Ukraine

“Igor Sikorsky Kyiv polytechnic institute”

Vovchenko D.S.<https://orcid.org/0009-0008-1806-5159>

National Technical University of Ukraine

“Igor Sikorsky Kyiv polytechnic institute”

SOFTWARE-BASED LOAD BALANCING METHOD FOR BIG DATA PROCESSING IN DISTRIBUTED INFORMATION SYSTEM

The article presents the research results aimed at the development and experimental validation of distributed information system with an implemented software-based adaptive load balancing method. The relevance of the work is due to the need to ensure stable operation of distributed systems in conditions of high intensity of requests. The limitations of traditional static load balancing algorithms, in particular Round Robin and Least Connections, are analyzed, as these approaches do not take into account the current state of computational resources of nodes and do not adapt to dynamic workload changes. It has been established that when processing large volumes of data, such methods lead to uneven request distribution, queue accumulation, overload of individual servers, and overall system performance degradation. A horizontally scalable multi-node system architecture is proposed, including a load balancing layer to ensure high availability, a web server cluster at the application level, a Redis cluster for session management and caching, and a clustered relational database based on PostgreSQL or Microsoft SQL Server (MSSQL). Support for the OData v4 protocol is implemented to provide standardized interaction with external services and flexible querying of large datasets. Additionally, a centralized logging subsystem deployed on a separate server is introduced to enable performance monitoring and error analysis.

The developed load balancing method is implemented in Python using a weighted evaluation mechanism of node states based on CPU utilization, memory usage, request queue length, response time, and cache hit ratio metrics. The approach includes parameter normalization and the formation of an integrated load indicator with subsequent adaptive adjustment of weight coefficients based on system feedback. Experimental research conducted in a simulated environment consisting of six application servers, a Redis cluster, and a three-node database cluster confirmed the effectiveness of the proposed solution. The research results demonstrate a reduction in average response time to 210 ms, an increase in throughput to 13,800 requests per second, and a decrease in load distribution variance to 6%, exceeding the performance of traditional balancing algorithms.

Keywords: distributed information system, Big Data, software-based load balancing method, scalability, database clustering, PostgreSQL, MSSQL, Python, Redis, OData v4.

Statement of the problem. The rapid growth of global Internet traffic and the exponential increase in data generation have intensified the need for scalable, high-performance distributed systems capable of processing large volumes of data in real time. Modern information infrastructures, including cloud platforms, enterprise systems, and distributed web services, increasingly rely on multi-node architectures to ensure availability, fault tolerance, and horizontal scalability. As the volume and variability of incoming data streams continue to expand, traditional load balancing mechanisms often fail to provide ade-

quate responsiveness and efficiency. One of the primary challenges is the emergence of the bottleneck effect, where limited processing capacity at specific nodes leads to request queuing, increased latency, and overall degradation of system performance. Load balancing mechanisms can generally be categorized into hardware-based and software-based approaches. Hardware load balancers rely on dedicated network appliances to distribute traffic, offering high throughput but at significant financial and deployment costs. In contrast, software-based load balancing solutions provide greater flexibility, cost efficiency, and ease of

integration into cloud-native and virtualized environments. The increasing prevalence of containerization, microservices, and software-defined infrastructures further emphasizes the importance of programmable and adaptive load distribution strategies. Consequently, there is a growing need for intelligent, software-oriented load balancing algorithms capable of dynamically adapting to workload fluctuations and large-scale data processing demands.

This research focuses on the development of a multi-node system incorporating a proprietary software-based load balancing method designed to optimize resource utilization and mitigate bottleneck formation during big data processing. The proposed architecture integrates relational database systems such as PostgreSQL or Microsoft SQL Server with clustering support, enabling distributed storage and efficient CRUD operations over large datasets. The system also supports OData v4 for standardized data interaction with external services, Redis for session management, and centralized logging mechanisms deployed on a dedicated server for enhanced monitoring and fault diagnosis. The load balancing method is implemented in Python, leveraging its extensibility and potential integration with artificial intelligence techniques for adaptive decision-making.

The primary objective of this research is to design, implement, and evaluate a scalable multi-node architecture that improves throughput, reduces latency, and enhances fault tolerance through optimized software-based load balancing strategies. By addressing the limitations of conventional static algorithms, the proposed approach aims to provide a cost-effective and extensible solution suitable for modern distributed environments handling large-scale data streams.

Analysis of recent research and publications. Load balancing has been extensively studied in the context of distributed systems, web server clusters, and software-defined networking (SDN). A comprehensive review of load balancing techniques in SDN is presented by Belgaum M. R. et al. [1], who analyze static, dynamic, and adaptive strategies in programmable networks. Their study highlights the limitations of conventional algorithms in highly dynamic traffic conditions and emphasizes the importance of intelligent and context-aware mechanisms. This work provides a systematic foundation for understanding the evolution of load balancing paradigms and supports the necessity of software-defined and programmable solutions.

Earlier research [2] proposed a weighted minimal connections algorithm for web servers, aiming to improve request distribution based on server load.

While effective in moderately dynamic environments, such approaches remain primarily heuristic and lack predictive capabilities, limiting their adaptability under rapidly fluctuating workloads. Similarly, Jiang Pei-rui et al. [3] introduced a client-proximity-based load balancing algorithm for web server clusters, focusing on reducing network latency. Although proximity-aware mechanisms improve response times, they do not directly address computational bottlenecks caused by uneven resource utilization. Research by Omori M. and Nishi H. [4] examined request distribution strategies for heterogeneous database server clusters, incorporating processing time estimation. Their work is particularly relevant to big data systems, as it recognizes performance heterogeneity among nodes. The proposed solutions rely on predefined estimation models rather than adaptive or learning-based optimization. Earlier traffic-based flow control mechanisms were discussed by Okano H. et al. [5], who proposed a flow cache port separation mechanism for network processors. Their approach aimed at improving packet processing efficiency but was largely hardware-oriented, reinforcing the distinction between hardware acceleration and software-defined flexibility.

In addition to the above references, recent research has further advanced intelligent and distributed load balancing mechanisms. For example, Mao H. et al. [6] introduced reinforcement learning for dynamic resource management in cloud computing environments, demonstrating the effectiveness of adaptive algorithms under variable workloads. Although focused on resource allocation rather than classical load balancing, their findings support the integration of learning-based approaches in distributed systems. More recently, Zhou Z. et al. [7] analyzed performance-aware load distribution in edge-cloud collaborative systems. Their findings confirm that adaptive software-based algorithms outperform static policies in heterogeneous environments, particularly when processing large data streams.

The literature indicates that while numerous load balancing techniques have been proposed, many rely on static heuristics or hardware-centric optimizations. The growing demand for scalable, cost-effective, and intelligent distributed systems necessitates the development of adaptive software-based algorithms capable of mitigating bottlenecks in big data processing environments.

The proposed research addresses this gap by designing a multi-node architecture with an optimized proprietary load balancing method tailored for large-scale data-intensive applications.

Outline of the main material of the research.

The proposed method represents a software-based adaptive load balancing mechanism designed for multi-node distributed systems operating under high-volume data workloads. Unlike traditional static policies such as Round Robin or Least Connections, the developed approach incorporates real-time performance monitoring and dynamic decision-making to ensure optimal resource utilization and prevention of bottleneck formation.

Algorithm workflow description

Step 1. Collect real-time metrics from all application nodes (CPU utilization, memory consumption, request queue length, response time, cache hit ratio).

Step 2. Normalize collected metrics.

Step 3. During operation, the algorithm collects key metrics from all application nodes, including CPU utilization, memory consumption, request queue length, response latency, and cache hit ratio.

Step 4. These parameters are normalized to ensure comparability and aggregated into a composite load score calculated as:

$$\text{Score}_i = w_1 \cdot \text{CPU}_i + w_2 \cdot \text{MEM}_i + w_3 \cdot \text{Queue}_i + w_4 \cdot \text{Latency}_i$$

The weight coefficients are not fixed; they are dynamically adjusted using reinforcement-based feedback derived from observed system performance. After each routing decision, the algorithm evaluates response time, throughput, and queue behavior. If performance degradation is detected, the weight configuration is corrected to penalize the contributing factor. The incoming request is then routed to the node with the minimal predicted processing cost. This feedback-driven mechanism enables continuous self-optimization and adaptation to changing workload conditions.

Step 5. Apply adaptive weight correction using reinforcement feedback.

Step 6. Route request to node with minimal predicted cost.

Step 7. Continuously update weights based on response success rate and throughput.

The algorithm is implemented in Python to enable extensibility, rapid prototyping, and integration of intelligent decision-support components. At its core, the algorithm operates on continuous collection and analysis of performance metrics from all active application nodes. These metrics include CPU utilization, memory consumption, request queue length, response latency, and cache efficiency indicators. The collected parameters are normalized to ensure comparability and are aggregated into a composite load score for

each node. This score reflects the current operational state and predicted processing capability of the node. The routing decision is based on a weighted evaluation model.

Each performance parameter contributes to the final load score according to adaptive weight coefficients. The weight values are not static; instead, they are dynamically adjusted using feedback from system performance outcomes. If a routing decision results in increased latency or queue growth, the algorithm modifies the weight distribution to penalize the contributing factor. This feedback-driven mechanism enables self-optimization over time.

To address large-scale data processing requirements, the algorithm incorporates predictive workload estimation. Historical workload patterns are analyzed to detect recurring peaks and anomalous traffic bursts. Based on short-term forecasting, the system can proactively redistribute incoming requests before node saturation occurs. This proactive strategy significantly reduces queue accumulation and improves overall system responsiveness. The proposed method is particularly effective in heterogeneous environments where nodes may differ in computational capacity. Instead of assuming uniform resource characteristics, the algorithm evaluates actual processing capability and adapts request distribution accordingly. This eliminates the typical imbalance observed in static balancing mechanisms and ensures equitable load distribution. As a result, the developed solution improves throughput, reduces response time, enhances fault tolerance, and increases resource utilization efficiency, particularly in Big Data processing scenarios.

The architecture illustrated in Fig. 1 represents a multi-layer distributed information system designed to support scalable Big Data processing and high-availability service delivery. The information system is structured into logically separated components, each responsible for a specific operational function.

The information system architecture is structured as a set of logically separated yet functionally interconnected layers that collectively ensure scalability, fault tolerance, and efficient Big Data processing. At the uppermost level, the user interacts with the system through standard HTTP/HTTPS communication protocols. All incoming requests are first directed to the load balancing layer, which serves as the central traffic coordination component. Only after processing by the balancing mechanism are requests forwarded to the appropriate application servers, ensuring controlled and optimized workload distribution. At the top layer, the user interacts with the system through standard network protocols. All incom-

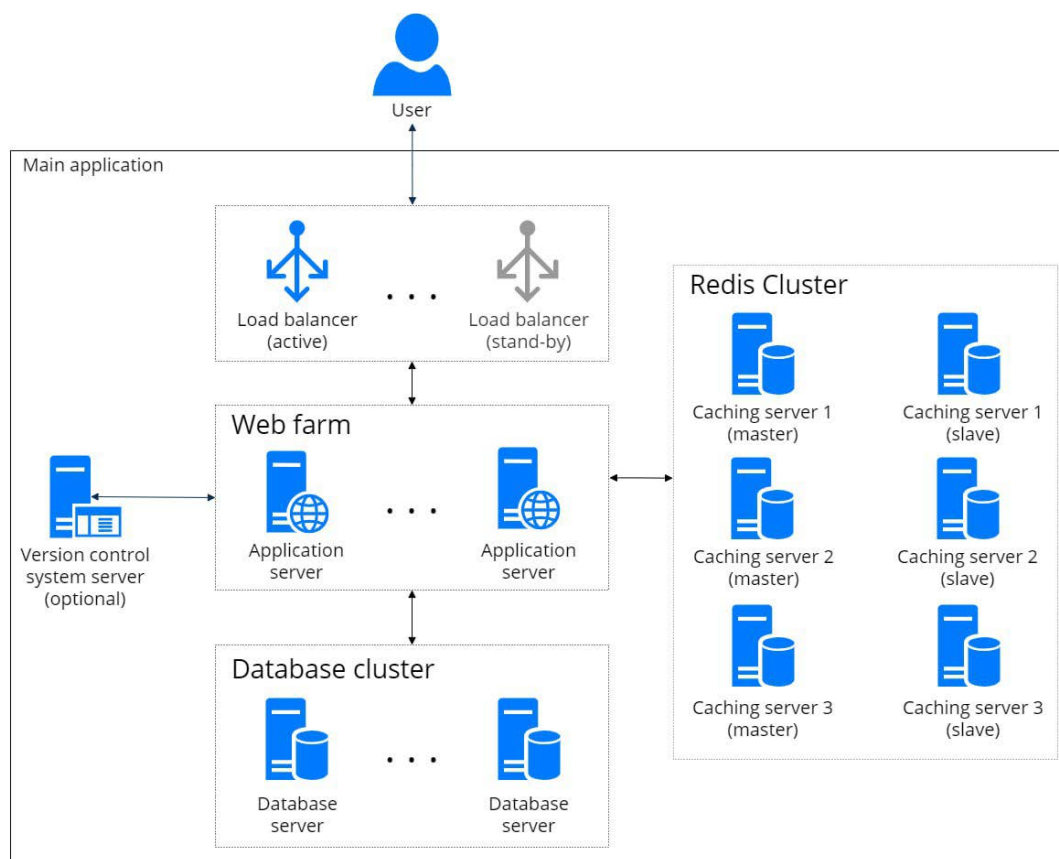


Fig. 1. Architecture of the multi-node distributed information system

ing requests are first directed to the load balancing layer, which serves as the central traffic management component. This layer includes an active load balancer and a standby instance configured for failover. The presence of a standby node ensures high availability and uninterrupted service in the event of primary balancer failure.

The load balancing layer distributes incoming traffic across a web farm composed of multiple application servers. These servers operate as independent processing nodes and execute identical service logic. The horizontal scaling capability of the web farm allows dynamic addition or removal of nodes depending on workload intensity. The balancing decisions are performed using the proposed adaptive algorithm, which evaluates real-time node metrics before forwarding each request. To support efficient session management and reduce database load, the architecture incorporates a Redis cluster operating in master–slave configuration. This caching layer is responsible for storing active user sessions, frequently accessed data, and temporary computational results. Master nodes handle write operations, while slave nodes provide replication and high-speed read access. This structure minimizes database queries and significantly decreases response latency.

The application layer is represented by a web farm consisting of multiple homogeneous application servers executing identical service instances.

The architecture supports horizontal scalability, allowing nodes to be dynamically added or removed according to workload intensity. The application servers are designed to operate in a stateless mode, with session and temporary state data stored externally. This design ensures that any request can be processed by any available node. The system supports the OData v4 protocol, enabling flexible querying and interaction with large datasets. Incoming requests are distributed among application servers based on real-time performance indicators collected and evaluated by the adaptive balancing algorithm. To enhance performance and reduce database load, the architecture incorporates a Redis cluster operating in master–slave replication mode. This layer is responsible for session management, distributed caching, and temporary data storage. Master nodes handle write operations, while slave nodes provide replication and high-speed read access, ensuring redundancy and fault tolerance. The Redis cluster significantly reduces latency by enabling rapid access to frequently requested data and maintaining session persistence across multiple application nodes.

The persistence layer consists of a clustered database environment implemented using either PostgreSQL or Microsoft SQL Server. Database clustering enables horizontal data partitioning and replication, thereby increasing throughput and improving reliability. The distributed storage model ensures that large datasets are segmented across multiple nodes, preventing single-point saturation and facilitating parallel query execution. A dedicated logging and monitoring subsystem complements the architecture. A separate logging server collects application logs, error reports, load balancing decisions, and performance metrics. These logs are used for bottleneck detection, failure diagnostics, performance evaluation of the balancing algorithm, and statistical analysis within the dissertation research framework. Continuous monitoring ensures transparency of system behavior and supports iterative algorithm optimization.

The information system includes an optional version control server to manage application source code and deployment processes. This component supports continuous integration and controlled updates within the distributed environment.

For enhanced reliability and performance monitoring, a dedicated logging subsystem collects operational metrics and system events. These logs are used to analyze bottleneck formation, evaluate balancing efficiency, and refine the adaptive algorithm. The presented architecture integrates software-based intelligent load balancing, distributed application processing, clustered caching, and scalable database storage into a unified high-performance system. The design ensures flexibility, fault tolerance, and efficient Big Data handling, making it suitable for modern cloud-native and enterprise-scale environments.

Research results. A central focus of the research is efficient Big Data processing. Traditional static load balancing strategies, such as Round Robin, are insufficient in high-volume environments because they do not account for real-time resource utilization. As a result, such approaches often lead to uneven workload distribution, node saturation, queue accumulation, and bottleneck formation. Static algorithms typically ignore critical operational parameters such as CPU utilization, memory consumption, I/O wait

time, queue length, and cache efficiency, which are essential in large-scale distributed systems. The proposed adaptive load balancing method addresses these limitations through dynamic, metric-driven decision-making. The core concept involves distributing requests based on real-time node performance metrics, predictive workload modeling, historical performance patterns, queue length estimation, and request size classification. The method continuously evaluates the operational state of each application server before making routing decisions. The experimental validation of the proposed system was conducted in a simulated distributed environment consisting of six application servers, three Redis master–slave pairs, and three database cluster nodes. A synthetic workload generator produced between one and ten million requests with mixed payload sizes ranging from 1 KB to 5 MB. The performance of the proposed method was compared against Round Robin and Least Connections strategies (Table 1).

The evaluation considered average response time, throughput, CPU load variance, queue growth behavior, system stability, and cache efficiency. The experimental results demonstrated that the proposed method achieved an average response time of 210 ms compared to 420 ms for Round Robin and 350 ms for Least Connections. Throughput increased to 13,800 requests per second, significantly exceeding the baseline approaches. CPU load variance was reduced to 6%, indicating improved load uniformity. Queue overflow incidents decreased to one occurrence, and the cache hit ratio increased to 89%.

Compared to Round Robin, the proposed method reduced response time by approximately 50%, increased throughput by 62%, reduced overload incidents fourfold, and improved load distribution uniformity threefold. Compared to Least Connections, the system demonstrated 40% faster processing, 50% improved stability under peak load, and 20% higher cache utilization efficiency.

The scientific contribution of the research lies in demonstrating that static load balancing strategies are inadequate for modern Big Data systems. The research proves that adaptive, metric-driven software-based balancing significantly enhances scalability, opera-

Table 1

Experimental research results

Metric	Round Robin	Least Conn.	Proposed method
Avg. response time	420 ms	350 ms	210 ms
Throughput	8,500 req/s	9,200 req/s	13,800 req/s
CPU load variance	27%	18%	6%
Queue overflow incidents	14	8	1
Cache hit ratio	68%	74%	89%

tional stability, and resource utilization efficiency. The integration of Redis clustering and distributed database architecture effectively mitigates bottleneck formation. The research confirms that software-based adaptive solutions provide superior flexibility and cost-efficiency compared to hardware-based load balancing alternatives in contemporary distributed environments.

Conclusions. The conducted research confirms that traditional static load balancing mechanisms are insufficient for modern distributed systems operating under intensive Big Data workloads. The experimental evaluation demonstrated that adaptive, metric-driven traffic distribution significantly improves response time, throughput, resource utilization uniformity, and overall system stability. By incorporating real-time performance monitoring, dynamic weight adjustment, and predictive workload estimation, the proposed method effectively mitigates bottleneck formation and reduces queue accumulation. The integration of distributed caching via Redis and clustered database architecture further enhances scalability and fault tolerance, creating a cohesive high-performance environment suitable for large-scale data processing. The developed multi-node architecture validates the feasibility of implementing an intelligent software-based balancing solution without reliance on specialized hardware appliances. The use of Python enables extensibility and seamless integration of analytical and machine learning components, which strengthens the adaptability of the system under heterogeneous and dynamically changing workload conditions. The experimental results confirm that the

proposed approach outperforms conventional Round Robin and Least Connections strategies across key performance indicators, particularly in environments characterized by variable request intensity and large data payloads.

Despite the achieved improvements, several directions for future work remain. First, further investigation into advanced reinforcement learning models could enhance predictive accuracy and enable deeper self-optimization capabilities. The incorporation of deep neural network-based workload forecasting may allow more precise anticipation of traffic peaks and anomaly detection. Second, additional research is required to evaluate the algorithm's performance in geographically distributed, edge-cloud hybrid infrastructures, where network latency and heterogeneous node capabilities introduce additional complexity.

The proposed approach allows to minimize the risk of bottlenecks, ensure rational use of computing resources and increase the efficiency of big data processing in dynamic conditions of increasing load.

Future research may also explore container orchestration integration, such as automated scaling mechanisms based on Kubernetes metrics, to achieve tighter coupling between adaptive load balancing and infrastructure management. Formal mathematical modeling of method convergence and stability properties would provide stronger theoretical validation of its robustness.

Large-scale real-world deployment and long-term monitoring would allow comprehensive validation under production-level conditions and provide further empirical evidence of efficiency gains.

Bibliography:

1. Belgaum M.R., Musa S., Alam M.M. and. Su'ud M.M. A Systematic Review of Load Balancing Techniques in Software-Defined Networking. *IEEE Access*, vol. 8, 2020. P. 98612-98636. URL: <https://doi.org/10.1109/ACCESS.2020.2995849> (access date: 05.03.2026).
2. Pan Z., Jiangxing Z. Load Balancing Algorithm for Web Server Based on Weighted Minimal Connections. *Journal of Web Systems and Applications*, vol. 1, 2017. P. 1–8. URL: <https://dx.doi.org/10.23977/jwsa.2017.11001> (access date: 05.03.2026).
3. Pei-rui J., Li-min M., Yu-zhou S., Yang-tian-xiu H. A client proximity based load balance algorithm in web sever cluster. *2nd International Conference on Wireless Communication and Network Engineering*. 2017. P. 317–322. URL: <https://dpi-journals.com/index.php/dtcse/article/view/19877> (access date: 05.03.2026).
4. Omori M., Nishi H. Request Distribution for Heterogeneous Database Server Clusters with Processing Time Estimation. *International Conference on Industrial Informatics (INDIN)*, Porto. 2018. P. 278-283. URL: <https://ieeexplore.ieee.org/document/8471931> (access date: 05.03.2026).
5. Okano H., Yamaguchi F., Takagiwa K., Nishi H. Traffic-based Flow Cache Port Separate Mechanism for Network processor. *The Institute of Electoronics, Information and Communication Engineers Technical Report*, vol. 114, no. 18, 2014. P. 69–74. (access date: 05.03.2026).
6. Mao H., Alizadeh M., Menache I., Kandula S. Resource Management with Deep Reinforcement Learning. *Proceedings of the 15th ACM Workshop on Hot Topics in Networks (HotNets-XV)*, Atlanta, GA, USA, 2016. P. 50–56. URL: <https://doi.org/10.1145/3005745.3005750> (access date: 06.03.2026).
7. Zhou Z., Chen X., Li E., Zeng L., Luo K., Zhang J. Edge Intelligence: Paving the Last Mile of Artificial Intelligence with Edge Computing. *Proceedings of the IEEE*, vol. 107, no. 8, 2021. P. 1738–1762. URL: <https://doi.org/10.1109/JPROC.2019.2918951> (access date: 06.03.2026).

Олещенко Л.М., Вовченко Д.С. ПРОГРАМНИЙ МЕТОД БАЛАНСУВАННЯ НАВАНТАЖЕННЯ ДЛЯ ОБРОБКИ ВЕЛИКИХ ДАНИХ У РОЗПОДІЛЕНІЙ ІНФОРМАЦІЙНІЙ СИСТЕМІ

У статті представлено результати дослідження, спрямованого на розробку та експериментальну перевірку розподіленої інформаційної системи з імplementованим програмним методом балансування навантаження. Актуальність роботи зумовлена необхідністю забезпечення стабільної роботи розподілених систем в умовах високої інтенсивності запитів. Проаналізовано обмеження традиційних статичних алгоритмів балансування, зокрема, Round Robin та Least Connections, які не враховують поточний стан обчислювальних ресурсів вузлів та не адаптуються до динаміки навантаження. Встановлено, що при обробці великих обсягів даних такі підходи призводять до нерівномірного розподілу запитів, накопиченні черг, перевантаженні окремих серверів та деградації продуктивності всієї системи. Запропоновано архітектуру багатовузлової системи з горизонтальним масштабуванням, яка включає рівень балансування навантаження для забезпечення доступності, кластер вебсерверів прикладного рівня, Redis-кластер для управління сесіями та кешування, а також кластеризовану реляційну базу даних на основі PostgreSQL або MSSQL.

Для стандартизованої взаємодії із зовнішніми сервісами реалізовано підтримку протоколу OData v4, що забезпечує гнучке формування запитів до великих наборів даних. Додатково впроваджено підсистему централізованого логування на окремому сервері для моніторингу продуктивності та аналізу помилок. Розроблений метод балансування реалізовано мовою Python з використанням механізму зваженої оцінки стану вузлів на основі метрик процесорного навантаження, використання оперативної пам'яті, довжини черги запитів, часу відгуку та коефіцієнта кеш-попадань. Передбачено нормалізацію параметрів та формування інтегрального показника навантаження з подальшою адаптивною корекцією вагових коефіцієнтів на основі зворотного зв'язку від системи. Експериментальні дослідження, проведені у змодельованому середовищі з шістьма прикладними серверами, кластером Redis та тривузловою базою даних, підтвердили ефективність запропонованого підходу. Отримано зменшення середнього часу відгуку до 210 мс, підвищення пропускної здатності до 13 800 запитів за секунду та зниження дисперсії розподілу навантаження до 6 %, що перевищує показники традиційних алгоритмів.

Ключові слова: розподілена інформаційна система, великі дані, програмний метод балансування навантаження, масштабованість, кластеризація баз даних, PostgreSQL, MSSQL, Python, Redis, OData4.

Дата першого надходження статті до видання: 10.03.2026

Дата прийняття статті до друку після рецензування: 06.04.2026

Дата публікації (оприлюднення) статті 11.05.2026

Пархоменко Ю.М.

<https://orcid.org/0000-0003-3492-3965>

Центральноукраїнський національний технічний університет

Пархоменко М.Д.

<https://orcid.org/0009-0004-6728-9619>

Центральноукраїнський національний технічний університет

Мацуй А.М.

<https://orcid.org/0000-0001-5544-0175>

Центральноукраїнський національний технічний університет

Бокій А.Р.

<https://orcid.org/0009-0006-1426-261X>

Центральноукраїнський національний технічний університет

Комаров Д.Ю.

<https://orcid.org/0009-0009-9929-4099>

Центральноукраїнський національний технічний університет

ВИЗНАЧЕННЯ МІСЦЯ ВСТАНОВЛЕННЯ ТА ТОЧНОСТІ РЕЄСТРАЦІЇ ДАТЧИКОМ ЗЕРНОВОГО ПОТОКУ, ЩО ФОРМУЄТЬСЯ КОТУШКОВИМ ВИСІВНИМ АПАРАТОМ

У статті представлено результати експериментальних досліджень параметрів зернового потоку, що формується катушковим висівним апаратом сівалки СЗ-3,6. Актуальність роботи зумовлена стратегічною важливістю зернового сектору України та необхідністю вдосконалення автоматизованих систем контролю висіву для підвищення врожайності. Основним завданням дослідження було визначення оптимального місця встановлення датчиків реєстрації та оцінка точності їхньої роботи в умовах реальної щільності потоку. Експерименти проводилися на універсальному стенді, що дозволяє моделювати різні норми висіву (від 90 до 250 кг/га) та швидкості руху агрегату (до 9,8 км/год). Встановлено, що щільність зернового потоку на виході висівного апарату в 2,5 рази вища, ніж на рівні дна борозни, що створює значні перешкоди для коректної реєстрації об'єктів через їхнє перекриття. Аналіз кратності проходження зернин через площину контролю показав, що при високих нормах висіву на виході апарату формується масовий рух зерна, де ймовірність появи поодиноких зернин знижується до 4,1%. Це призводить до злиття зображень декількох зернин і, як наслідок, до неприпустимих похибок реєстрації (до 5% і вище). Доведено, що найбільш придатним місцем для встановлення датчика є зона, наближена до дна борозни (на виході насіннєпроводу), де потік стає більш розрідженим, а частка поодиноких зернин становить 51,3–84,9%. Використання розробленої системи комп'ютерної ідентифікації на базі двокоординатного оптико-електронного реєстратора забезпечує високу точність: відносна похибка визначення кількості зернин не перевищує 2%, а відхилення при оцінці щільності розподілу — не більше 1%. Отримані дані підтверджують неможливість використання датчиків об'ємної дії через їхню нездатність розпізнавати паралельно пролітаючі об'єкти. Результати мають практичне значення для розробки нових систем точного землеробства та автоматизації керування нормою висіву пневматичних та механічних сівалок.

Ключові слова: зерновий потік, катушковий висівний апарат, автоматизована система контролю, точність реєстрації, оптико-електронний датчик, норма висіву, комп'ютерна ідентифікація.

Постановка проблеми. Зерновий сектор України є стратегічною галуззю, де підвищення якості сівби безпосередньо впливає на врожайність. Одним із ключових факторів тут є впровадження систем контролю висіву. Проте ефективність таких систем критично залежить від точності датчиків, які працюють в умовах високої щільності та пульсуючого характеру потоку насіння.

Аналіз останніх досліджень і публікацій. Зарубіжні дослідження зосереджені на впровадженні новітніх сенсорних технологій, зокрема 3D-сканування, лазерної діафотометрії та алгоритмів машинного навчання. Використання нейромережових моделей дозволяє підвищити точність розпізнавання групових проходжень зернин у зоні контролю та зменшити систематичну похибку обліку [1]. Окремі вітчизняні дослідження зосереджуються на математичному моделюванні процесів руху насіння в каналі висівного апарата. Так, у роботі Шевченка І.М. наведено чисельну модель формування зернового потоку в умовах впливу зовнішніх сил і геометрії каналу. Автор підкреслив, що параметри геометрії та руху значною мірою визначають ймовірність утворення групових подинних проходжень через площину контролю [2]. У дослідженнях Гнатюка О.В. адаптовано методи статистичної обробки сигналів для вітчизняних сівалок із урахуванням варіабельності якості насіння та робочих режимів агрегатів. Запропоновані алгоритми компенсації похибок враховують специфіку польових умов, що дозволяє підвищити точність розрахункового обліку висіву до рівня, прийнятного для практичних технологій точного землеробства [3].

Вітчизняні дослідження мають суттєвий прикладний характер та орієнтовані на адаптацію систем контролю до умов українського агропромислового комплексу. Проблеми ідентифікації насіння розглядалися в роботах Кондратця В.О. та Пархоменка Ю.М., де вказувалося на обмеження датчиків об'ємної дії. Більшість існуючих систем не враховують "зерна-двійники" або групове проходження насіння, що призводить до значних похибок. Невирішеною залишається частина проблеми, пов'язана з обґрунтуванням оптимальної точки встановлення датчика в насіннепроводі для мінімізації накладання об'єктів у площині контролю.

Постановка завдання. Метою дослідження є визначення місця встановлення та точності реєстрації датчиком зернового потоку, що формується котушковим висівним апаратом, для забезпечення мінімальної похибки контролю.

Виклад основного матеріалу. Зерновий сектор України є стратегічною галуззю економіки держави оскільки збільшення обсягів ВВП в аграрному секторі забезпечує вдвічі більший ефект порівняно з іншими сферами господарства. Відомо, що сівба є одним з основних технологічних процесів в рослинництві, яка вагомо впливає на врожайність посівних культур. Значним фактором підвищення продуктивності і якості сівби є впровадження автоматизованих систем контролю висіву, тому питання їх подальшого вдосконалення завжди є актуальним.

Метою експериментальних досліджень є визначення місця встановлення та точності реєстрації датчиком зернового потоку, що формується котушковим висівним апаратом зернової сівалки. Дослідження проводилися на стенді для випробування висівних апаратів сівалок кафедри сільськогосподарського машинобудування Центральноукраїнського національного технічного університету (рис. 1).

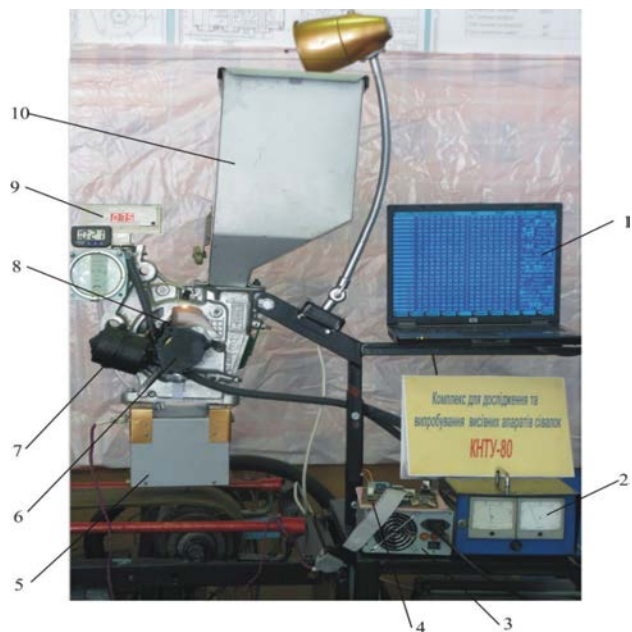


Рис. 1. Універсальний стенд для випробування висівних апаратів сівалок:

- 1 – ноутбук; 2 – блок живлення електропривода валу висівного апарата (0...24В); 3 – блок живлення засобів ідентифікації (-12В, +10В, +5В); 4 – вузол синхронізації, попередньої обробки та передачі даних; 5 – вузол реєстрації; 6 – висівний апарат; 7 – двигун постійного струму M914A; 8 – електронний датчик швидкості обертання валу висівного апарата; 9 – цифровий індикатор швидкості; 10 – бункер для насіння

Встановлений на стенді котушковий висівний апарат сівалки СЗ-3,6 налагоджувався на задані норми висіву $Q = 90, 150, 175, 250 \frac{\text{кг}}{\text{га}}$ ($Q_c = 90, 177, 224, 340 \frac{\text{шт}}{\text{с}}$) та швидкості руху сівалки $V_c = 7.2, 8.5, 9.2, 9.8 \text{ км/год}$,

шляхом ручного регулювання робочої довжини котушки та частоти її обертання 1.05, 1.24, 1.32 та 1.46 об/с. Плавне регулювання частоти обертання валу котушкового висівного апарата забезпечувалося зміною напруги живлення двигуна постійного струму Д125Б ($U=0\dots24В$).

Швидкість обертання валу висівного апарата, яку заміряли за допомогою електронного тахометра з цифровим індикатором, відповідала заданій швидкості руху сівалки, передаточній величині i та нормі висіву. Відстань до денця котушки задавалася регулювальним гвинтом і відповідала обраній фракції пшениці: $A=25г/мис. шт.$ та $A=40г/мис. шт.$ Для контролю зерна в потоці було задіяно дослідний зразок системи реєстрації та комп'ютерної ідентифікації зернового потоку, що включає: двохкоординатний оптико-електронний регістратор, вузол синхронізації, попередньої обробки та передачі даних та ноутбук. Вказана система дозволяла реєструвати до 7 зернин, які одночасно перетинають площину реєстрації датчика. Зареєстровані системою комп'ютерної ідентифікації зображення пшениці заносилися у вигляді цифрового коду в пам'ять комп'ютера.

Згідно агротехнічних вимог на кожні 5 см рядка висіву повинно припадати по одній зернині, тобто на кожні 10 погонних метрів рядка необхідно висівати біля 200 зернин. При швидкостях руху сівалки $V_c=2.0, 2.36, 2.56, 2.72$ м/с та вказаних нормах висіву задана кількість зернин може бути висіяна менше ніж за 5с. Для визначення точки формування незв'язаних рухомих об'єктів та параметрів потоку в площині реєстрації по кожній з встановлених норм висіву проводилося випробування тривалістю до 20с з розміщенням реєструючої рамки на рівні дна борозни та на виході висівного апарата. Дані реєстрації зображень зернового потоку по кожному досліді

накопичувалися в пам'яті комп'ютера. З часом цифрові зображення об'єктів послідовно виводилися на екран монітора де візуально визначалася кількість розділених зернин, які одночасно перетинали площину реєстрації датчика. При цьому виділялося по три ділянки зображень ємністю біля 200 зернин. Даний стенд для дослідження роботи висівних апаратів сівалок повністю відповідав умовам роботи сівалки СЗ-3,6.

Результати контролю зернового потоку на рівні дна борозни, який формувався висівним апаратом сівалки при різних нормах висіву та швидкостях її руху, представлено в табл. 1.

З даних таблиці видно, що площину контролю регістратора, розміщеного на рівні дна борозни, здебільшого перетинають поодинокі зернини. Такий розподіл особливо притаманний усім нормам висіву. В залежності від норми висіву число поодиноких зернин коливається в межах $84.9\div51.3\%$. При зростанні норми висіву спостерігається тенденція збільшення кількості (2, 3...7) насінин, які одночасно перетинають площину контролю датчика. Випадки з двома зернинами складають $13.37\div31.09\%$. Імовірність перетину площини контролю одночасно трьома зернинами невелика $1.74\div12.6\%$ і зовсім мала стосовно чотирьох зернин – $0.58\div3.36\%$ на всьому діапазоні зміни умов сіви. Ймовірність одночасного проходження п'яти зернин при нормі висіву 175кг/га і навіть шести зернин при нормі висіву 250кг/га не перевищує 0.8% .

Значення вибірових середніх x_{0j} , випадкових величин x_i та середнє квадратичне відхилення δ_{0j} випадкових величин x_i від вибірових середніх x_{0j} при різних нормах висіву, визначені за співвідношеннями (1) на підставі даних табл. 1, представлені в табл. 2.

Таблиця 1

Кратність k_{ij} реєстрації випадків одночасного перетину площини контролю датчика x_i -тою кількістю насінин

Кількість паралельних насінин X_i шт.	Умови сіви – норма висіву Q , Q_c , швидкість руху сівалки V_c			
	$Q=90\text{кг/га},$ $Q_c=90\text{шт/с},$ $V_c=7,2\text{км/год}, j=1$	$Q=150\text{кг/га},$ $Q_c=177\text{шт/с},$ $V_c=8,5\text{км/год}, j=2$	$Q=175\text{кг/га},$ $Q_c=224\text{шт/с},$ $V_c=9,2\text{км/год}, j=3$	$Q=250\text{кг/га},$ $Q_c=340\text{шт/с},$ $V_c=9,8\text{км/год}, j=4$
1	146	108	92	61
2	23	33	36	37
3	3	7	9	15
4	0	1	2	4
5	0	0	1	1
6	0	0	0	1
7	0	0	0	0
8	0	0	0	0

$$x_{0j} = \frac{\sum_{i=1}^8 x_i \cdot k_{ij}}{\sum_{i=1}^8 k_{ij}} \quad \delta_{0j} = \frac{(x_{0j} - x_i)^2 \cdot k_{ij}}{\sum_{i=1}^8 k_{ij}} \quad (1)$$

Як видно з даних табл. 2 середньо статистичні значення x_{0j} кількості паралельних зернин, які одночасно перетинають площину контролю датчика, при збільшенні норми висіву зростає, але змінюється порівняно у вузьких межах від 1,17 до 1,74шт, складаючи практично 1,5 насінини. З ростом норми висіву збільшуються і середні квадратичні відхилення x_{0j} , однак зростання тут більш суттєве і знаходиться в межах 0.175...0.92шт. Отримані результати експериментальних даних підтверджують, що навіть при встановленні реєстратора на рівні дна борозни, використання датчиків об'ємної дії для дослідження зернового потоку не можливе, так як вони не розпізнають і не реєструють навіть зерна двійники, які одночасно перетинають площину їх контролю, не кажучи уже по 3 і більше паралельно пролітаючих насінини [4].

Аналогічні дослідження параметрів зернового потоку були проведені і на виході висівного апарату. Кратність фіксації випадків пролітання паралельних насінин через площину реєстрації датчика, встановленого на виході висівного апарату та на рівні дна борозни при нормах висіву $Q = 90, 150, 175, 250 \frac{\text{кг}}{\text{га}}$ наочно відображається графіками, представленими на рис. 2.

Аналіз цих графіків показує, що кратність реєстрації поодиноких зернин в потоці на виході висівного апарату складає лише 70 випадків (рис. 2а) проти 146 – при реєстрації на рівні дна борозни (рис. 2б). Ймовірність появи поодиноких зернин на виході висівного апарату коливається в діапазоні 4,1÷49%. При збільшенні норми висіву число поодиноких зернин різко зменшується, а число 2-х паралельних зернин складає уже 31.25÷10.4%, 3-х зернин – 13.9÷16.7%, 4-х – 4.9÷20.83%, 5-ти – 1.4÷20.83%. При нормах висіву 175кг/га та 250кг/га по 6-ть паралельних зернин було зареєстровано в 7.6÷14.6% випадків, по 7 паралельних зернин – в 8.3÷2.53% випадків і лише при нормі висіву 250кг/га було зареєстровано в 4.15% випадів по 8 паралельних зернин.

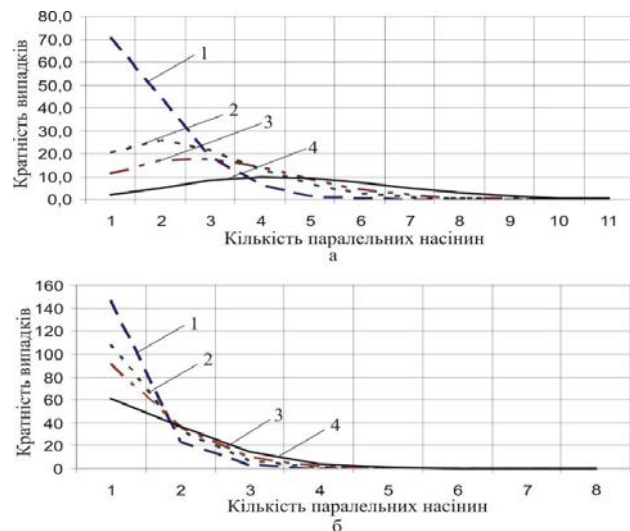


Рис. 2. Кратність появи паралельних насінин на виході висівного апарату (а) та на рівні дна борозни (б): 1 – $Q_c=90 \text{ шт/с}$; 2 – $Q_c=1770 \text{ шт/с}$; 3 – $Q_c=224 \text{ шт/с}$; 4 – $Q_c=340 \text{ шт/с}$

В табл. 3 представлені значення вибірових середніх $\bar{x}_{0j}$ випадкових величин x_i та середнє квадратичне відхилення δ_{0j} випадкових величин x_i від вибірових середніх $\bar{x}_{0j}$ при різних нормах висіву, визначені на підставі даних реєстрації зернового потоку на виході висівного апарату за співвідношеннями (1).

Порівняння вибірових середніх табл. 2 і табл. 3 показує, що щільність зернового потоку на виході висівного апарату в 2,5 рази вища, ніж на рівні дна борозни. Результати цих досліджень на практиці підтверджують теоретично обґрунтовану динаміку розподілу зернового потоку на виході котушкового висівного апарату та її пульсуючий характер [5]. При великих нормах висіву на виході висівного апарату формується практично масовий рух зерна, що ускладнює їх роздільну реєстрацію.

На вибір точки ідентифікації зернового потоку значний вплив здійснює і можливість розпізнавання кожної з паралельних зернин в площині контролю датчика і точність їх реєстрації. Найбільше значення довірчого інтервалу $x_{04} + \Delta x$ при нормі висіву $Q_c = 340 \frac{\text{шт}}{\text{с}}$, до якого з імовірністю, наприклад $\alpha=0.997$, попадають значення

Таблиця 2

Статистичні параметри зернового потоку зареєстровані на рівні дна борозни

Параметри	Умови сівби			
	$Q_c=90 \text{ шт/с}$	$Q_c=177 \text{ шт/с}$	$Q_c=224 \text{ шт/с}$	$Q_c=340 \text{ шт/с}$
Вибіркові середні x_{0j} , шт	1,17	1,35	1,46	1,74
Середні квадратичні відхилення δ_{0j} , шт	0,175	0,36	0,55	0,92
Кількість вимірів N_j	172	149	140	119

Статистичні параметри зернового потоку зареєстровані на виході висівного апарату

Параметри	Умови сівби			
	$Q_c=90$ шт/с	$Q_c=177$ шт/с	$Q_c=224$ шт/с	$Q_c=340$ шт/с
Вибіркові середні $\bar{x}_{0j}$ т	2,08	2,31	3,24	
Середні квадратичні відхилення δ_{0j}	0,99	2,57	2,54	1,76
Кількість вимірів N_j		83	79	48

випадкових величин x_{ij} зареєстрованих на рівні дна борозни, визначається співвідношенням $P(x_{ij} \leq x_{04} + \Delta x) = P(x_{ij} \leq x_{04} + 3\delta_{04})$. $x_{04} + 3\delta_{04} = 1,71 + 3 \cdot 0,99 = 4,5$ (див. табл. 2). Тобто, майже в 99,7% випадків реєстрації зернового потоку на рівні дна борозни число паралельно пролітаючих зернин не буде перевищувати 4÷5 одиниць, ймовірність появи яких менше 0,8%. Найбільше значення довірчого інтервалу $x_{04} + \Delta x$ при тій же нормі висіву, до якого з імовірністю $\alpha=0,997$ попадають значення випадкових величин x_{ij} зареєстрованих на виході висівного апарату дорівнює $x_{04} + 3\delta_{04} = 1,71 + 3 \cdot 1,76 = 9,71$ (див. табл. 3). В цьому випадку число паралельно пролітаючих зернин не буде перевищувати 9÷10 одиниць. Ймовірність появи від 8 до 10 паралельно пролітаючих зернин знаходиться в межах 2÷4,15%. Похибка розпізнавання та реєстрації зернового потоку на виході висівного апарата з використанням двох координатного оптико-електронного скануючого датчика може знаходитися в межах 2...5%, що недопустимо. Це пояснюється наступними причинами. По – перше, принцип роботи цього датчика розрахований на розпізнавання та реєстрацію до 7 паралельно пролітаючих зернин [6]. Згідно даних графіка (рис. 2а), при нормі висіву $Q_c = 340 \frac{\text{шт}}{\text{с}}$ фіксуються непоодинокі випадки реєстрації більше 8 паралельно пролітаючих насінин. По – друге, при швидкості зернин 0,7...2 м/с і значній щільності їх одночасного прольоту через площину реєстрації, проекції зображення 2...3 і більше насінин вздовж координат X і Y можуть зливатися і бути зареєстрованими як один об'єкт. Число таких випадків при збільшенні інтенсивності зернового потоку та практично незмінній швидкості насінин пропорційно збільшується. В той же час, ймовірність появи таких випадків на рівні дна борозни не перевищує 0,8%, що допустимо.

В результаті аналізу отриманих експериментальних даних, приходимо до висновку, що найбільш придатним місцем для реєстрації зернового потоку та оцінки якості сівби є наближена до дна борозни зона пролітання зерна: на виході насін-

непроводу – при автоматизації процесів керування висівом зернових культур або на рівні дна борозни – при лабораторних дослідженнях.

В процесі експериментального дослідження точності реєстрації зернового потоку за допомогою системи комп'ютерної ідентифікації виконано 103 експерименти на рівні дна борозни при висіві пшениці на липку стрічку. Зерновий потік перетинав площину контролю датчика і розподілявся на липкій стрічці, яка рухалася з заданою швидкістю. Отримані в процесі дослідження при різних нормах висіву і швидкостях руху сівалки результати ручного (на липкій стрічці) та автоматичного (при реєстрації датчиком) контролю кількості висіяних насінин представлені в табл. 4.

В ході обробки експериментальних даних встановлено, що відносна похибка визначення кількості зернин, зареєстрованих системою комп'ютерної ідентифікації на рівні дна борозни не перевищує 2%. Ця похибка може бути меншою, оскільки тривалість випробування, з метою зменшення обсягу ручної роботи, була значно знижена відносно визначених 30с. При цьому певна похибка допускалася і при переключенні зернового потоку.

При оцінці якості висіву пшениці катушковим висівним апаратом сівалки СЗ-3,6 за щільністю в заданому заліковому інтервалі з використанням системи комп'ютерної ідентифікації та липкої стрічки було виконано п'ять дослідів по 100 залікових інтервалів довжиною 5м при нормах висіву 90 кг/га і п'ять дослідів по 60 залікових інтервалів довжиною три метри за умов норми висіву 175кг/га. Показники розподілу зернового потоку за щільністю в п'яти сантиметровому інтервалі при швидкості руху сівалки 9,2км/год представлені в табл. 5.

Як видно з даних табл. 5. в експериментах фактична норма висіву була меншою заданої, що викликано дещо крупнішими зернами у використаному посівному матеріалі порівняно з теоретично прийнятими. Кількість зареєстрованих зернин достатньо точно співпадає. Системою ідентифікації фіксувалася менша кількість зернин

Таблиця 4

Порівняльна таблиця ручної та автоматичної реєстрації потоку насіння

№ досліджу	Умови проведення експерименту – норма висіву Q , тривалість випробування T , швидкість руху сівалки V_c					
	$Q=90\text{кг/га},$ $T=2\text{с},$ $V_c=7,2\text{км/год}$		$Q=175\text{кг/га},$ $T=1\text{с},$ $V_c=8,5\text{км/год}$		$Q=250\text{кг/га},$ $T=1\text{с},$ $V_c=9,8\text{км/год}$	
	авт.	ручн.	авт.	ручн.	авт.	ручн.
1	174	176	225	228	344	351
2	178	181	218	221	343	348
3	175	177	225	230	347	353
4	180	183	214	218	335	341
5	181	182	221	225	332	337

Таблиця 5

Ручна і автоматична оцінка якості розподілу зернового потоку вздовж рядка висіву

№ п/п	Найменування показників виміру		Джерело інформації			
			Липка стрічка	Система ідентиф.	Липка стрічка	Система ідентиф.
1	Задана норма висіву	$Q_{\text{кг/га}}$	90		175	
		$Q_{\text{шт/м}}$	45		87,5	
2	Розрахункова щільність $\bar{k}_p$, шт/5см		2,25		4,38	
3	Розрахункова відстань між сусідніми зернинами $\bar{x}_p$, см		2,27		1,16	
4	Довжина борозни в експерименті, м		5	5	3	3
5	Кількість зареєстрованих насінин, шт		216	215	251	250
6	Фактична норма висіву, шт/м		43,36	43,14	83,65	83,34
7	Вибіркова середня щільність $\bar{k}_e$, шт/5см		2,17	2,16	4,18	4,17
8	Середнє квадрат. відхилення σ_k , шт		1,93	1,91	1,97	1,95
9	Вибіркове середнє значення відстані між зернинами, см		2,316	2,328	1,195	1,2

порівняно з еталоном, що підтверджує малоймовірне неврахування певної кількості дискретних об'єктів.

Деяке неспівпадіння заданих показників з фактичними спостерігається і по щільності, і по відстані між зернинами. Однак, результати отримані комп'ютеризованою системою ідентифікації зернового потоку достатньо близько співпадають з даними, зареєстрованими етальонним конвеєром з липкою стрічкою. Результати на різних нормах висіву близькі між собою. При більших нормах висіву вони дещо кращі. В цілому розбіжність фактичних норм висіву, середніх щільностей,

середніх відстаней між зернинами не перевищує 0,5%. Середні квадратичні відхилення щільності розподілу зернин на п'яти сантиметровому інтервалі також практично однакові, невідповідність їх не виходить за межі 1%.

Висновки. В результаті оцінки точності реєстрації неупорядкованого потоку зерна системою комп'ютерної ідентифікації було встановлено, що похибка реєстрації зернового потоку сформованого катушковим висівним апаратом зернової сівалки не перевищує 2%, а відхилення при оцінці щільності розподілу зернового потоку не виходить за межі 1%.

Список літератури:

1. Smith J., Brown P. Optical and 3D sensing technologies for seed flow monitoring in precision agriculture. Precision Agriculture. 2017. Vol. 18(4). P. 455–472.
2. Шевченко І. М. Математичне моделювання процесу формування зернового потоку у висівних апаратах зернових сівалок. Вісник аграрної науки України. Київ, 2021. №6. С. 45–52.
3. Гнатюк О. В., Мельник С. І., Бондаренко Р. Л. Статистичні методи обробки сигналів у системах контролю висіву зернових культур. Науковий вісник Національного університету біоресурсів і природокористування України. Київ, 2022. № 289. С. 112–119.
4. Пархоменко Ю.М. Аналіз можливостей датчиків об'ємної дії, що застосовуються в діючих системах контролю висіву./Ю.М. Пархоменко, М.Д. Пархоменко, А.Р. Бокій. Конструювання, виробництво та експлуатація сільськогосподарської техніки. Київ, 2023. № 1. С. 10–15.

плуатація сільськогосподарських машин. Загальнодерж. міжвідомч. наук.-техн. зб-к, вип. №48, Кропивницький: ЦНТУ, 2018.- с.79 – 88

5. Пархоменко Ю.М. Визначення математичної моделі процесу формування зернового потоку на виході катушкового висівного апарата. /Ю.М. Пархоменко, В.О Кондратець, М.Д. Пархоменко, Конструювання, виробництво та експлуатація сільськогосподарських машин”, Загальнодерж. міжвідомч. наук.-техн. зб-к, вип. №41, Кіровоград: КНТУ, 2011.-с.62-68.

6. Кондратець В.О. Підвищення дозволяючої здатності пристрою ідентифікації потоку насіння в зернових сівалках. /В.О. Кондратець, Ю.М. Пархоменко. Проблеми енергозабезпечення та енергозбереження в АПК України. Вістник Харківського національного технічного університету сільського господарства ім. П. Василенка. Технічні науки. Вип.117. – 2011. – с. 47-49

Parkhomenko Yu.M., Parkhomenko M.D., Matsui A.M., Bokii A.R., Komarov D.Yu.

DETERMINATION OF THE INSTALLATION PLACE AND THE ACCURACY OF REGISTRATION BY THE GRAIN FLOW SENSOR FORMED BY A REEL SEEDER

The article presents the results of experimental studies of the parameters of the grain flow formed by the reel sowing device of the SZ-3.6 seeder. The relevance of the work is due to the strategic importance of the grain sector of Ukraine and the need to improve automated sowing control systems to increase yields. The main task of the study was to determine the optimal location for installing registration sensors and assess the accuracy of their operation in conditions of real flow density. The experiments were conducted on a universal stand that allows modeling different seeding rates (from 90 to 250 kg/ha) and unit speed (up to 9.8 km/h). It was established that the density of the grain flow at the exit of the sowing device is 2.5 times higher than at the level of the bottom of the furrow, which creates significant obstacles to the correct registration of objects due to their overlap. Analysis of the frequency of grain passage through the control plane showed that at high seeding rates, a mass movement of grain is formed at the exit of the device, where the probability of single grains is reduced to 4.1%. This leads to the merging of images of several grains and, as a result, to unacceptable registration errors (up to 5% and above). It has been proven that the most suitable place for installing the sensor is the zone close to the bottom of the furrow (at the exit of the seed tube), where the flow becomes more rarefied, and the proportion of single grains is 51.3–84.9%. The use of the developed computer identification system based on a two-coordinate optical-electronic recorder provides high accuracy: the relative error in determining the number of grains does not exceed 2%, and the deviation in estimating the distribution density is no more than 1%. The obtained data confirm the impossibility of using volumetric sensors due to their inability to recognize parallel flying objects. The results have practical significance for the development of new precision farming systems and automation of seeding rate control of pneumatic and mechanical seeders.

Keywords: grain flow, reel seeding machine, automated control system, registration accuracy, optoelectronic sensor, seeding rate, computer identification.

Дата першого надходження статті до видання: 03.02.2026

Дата прийняття статті до друку після рецензування: 27.02.2026

Дата публікації (оприлюднення) статті 11.05.2026

УДК 681.5:004.942:697.94

DOI DOI DOI DOI <https://doi.org/10.32782/2663-5941/2026.2.1/33>**Пелих С.М.**<http://orcid.org/0000-0003-1608-8089>

Національний університет «Одеська політехніка»

Ложечніков В.Ф.<https://orcid.org/0009-0001-8380-4922>

Національний університет «Одеська політехніка»

Зієров О.Р.<https://orcid.org/0009-0002-3049-4681>

Національний університет «Одеська політехніка»

АЛГОРИТМІЧНО-ОРІЄНТОВАНА ЦИФРОВА СИСТЕМА КЕРУВАННЯ МІКРОКЛІМАТОМ ВИРОБНИЧИХ ПРИМІЩЕНЬ

У статті розглянуто проблему забезпечення нормативних показників мікроклімату у виробничих приміщеннях з урахуванням сучасних вимог до енергоефективності та безпеки праці. Підтримання оптимальних значень температури повітря є одним із ключових чинників збереження здоров'я працівників, підвищення їх працездатності та зниження виробничих ризиків. Зростання вартості енергоресурсів зумовлює необхідність впровадження автоматизованих систем керування, які дозволяють мінімізувати експлуатаційні витрати та забезпечити раціональне використання кліматичного обладнання.

Метою роботи є розроблення автоматизованої системи забезпечення оптимального мікроклімату виробничого приміщення, яка поєднує алгоритм визначення необхідної кількості кондиціонерів залежно від геометричних параметрів приміщення та програмну реалізацію автоматичного регулювання температури на основі зворотного зв'язку. У роботі запропоновано структурну схему системи, що складається з модуля проєктування та модуля функціонування, які забезпечують розрахунок параметрів обладнання, збір і оброблення вимірювальної інформації, формування керуючих впливів та відображення результатів на інтерфейсі користувача.

Розроблено алгоритм визначення кількості кондиціонерів на основі порівняння характеристик приміщення з нормативними та табличними значеннями теплових навантажень. Алгоритм функціонування системи передбачає постійний моніторинг температури у робочій зоні, порівняння її з установленими нормативними значеннями та формування відповідних керуючих сигналів для виконавчих пристроїв. Програмну реалізацію системи здійснено в середовищі Delphi, що дозволяє забезпечити зручний графічний інтерфейс, оброблення даних у реальному часі та інтерактивну взаємодію з оператором.

Запропонована система забезпечує підвищення точності підтримання температурного режиму, зниження енергоспоживання та створення стабільних і безпечних умов праці. Отримані результати можуть бути використані для модернізації систем клімат-контролю у виробничих приміщеннях різного призначення та подальшого розвитку автоматизованих енергоефективних технологій.

Ключові слова: автоматизована система управління, енергоефективність, кліматичне обладнання, мікроклімат, оптимізація енергоспоживання, програмне забезпечення, регулювання температури.

Постановка проблеми. В Україні допустимі та оптимальні параметри повітряного середовища регламентуються чинними санітарними нормами, які встановлюють граничні значення мікрокліматичних показників залежно від характеру виконуваних робіт і сезонних умов. Міністерство охорони здоров'я України висуває підвищені вимоги до дотримання нормативних параметрів мікроклімату

у виробничих приміщеннях [1]. Особливе значення надається підтриманню оптимальних значень температури, відносної вологості та швидкості руху повітря, оскільки ці показники безпосередньо впливають на функціональний стан організму працівників і рівень їхньої працездатності [2].

Порушення встановлених норм мікроклімату у виробничих зонах спричиняє негативні наслідки

для здоров'я персоналу, зокрема підвищену втомлюваність, зниження концентрації уваги, дратівливість і ослаблення імунної системи. У тривалій перспективі це може призводити до розвитку професійних захворювань та збільшення ризику виробничого травматизму. Негативний вплив несприятливих умов праці відображається також на економічних показниках підприємства через зниження продуктивності праці та ефективності виробничих процесів.

Мінімізація впливу несприятливих мікрокліматичних умов досягається шляхом реалізації комплексу технічних і організаційних заходів. Серед них – оптимізація технологічних процесів із зменшенням тепловиділень, механізація й автоматизація виробничих операцій, використання теплових екранів та огорожувальних конструкцій, ефективна теплоізоляція обладнання, раціональне планування виробничих площ, а також застосування сучасних систем вентиляції та кондиціонування повітря.

Найбільш перспективним напрямом є впровадження автоматизованих систем керування мікрокліматом, що функціонують на основі програмного забезпечення та інтегрованих сенсорних систем контролю. Дані системи забезпечують постійний моніторинг температури, вологості та швидкості повітря, а також автоматичне коригування режимів роботи вентиляційного, опалювального й охолоджувального обладнання відповідно до встановлених алгоритмів. Використання цифрових технологій управління дозволяє оперативно адаптувати параметри мікроклімату до змін зовнішніх і внутрішніх умов, знижувати енерговитрати та підтримувати стабільний рівень комфорту у виробничому середовищі.

Аналіз останніх досліджень і публікацій. У роботі О.В. Строкань зазначено, що впровадження автоматизованої системи підтримання оптимального мікроклімату на основі спеціалізованого програмного забезпечення забезпечує можливість безперервного контролю та регулювання температурних і іонізаційних параметрів повітряного середовища у виробничих приміщеннях. Даний підхід дозволяє підвищити стабільність параметрів мікроклімату та покращити умови праці персоналу [3].

У дослідженні [4] підкреслено, що сучасні системи контролю мікроклімату зазнали суттєвої еволюції, ставши більш складними, адаптивними та функціонально насиченими. Авторами запропоновано автоматизовану систему керування мікрокліматом житлових приміщень, а також розроблено

програмований таймер для регулювання кліматичних параметрів із використанням апаратно-програмної платформи Arduino NANO та сенсорів температури, вологості й атмосферного тиску.

Ігор Голод у своїх дослідженнях акцентує увагу на інтеграції кіберфізичних систем із сучасними математичними моделями керування, зокрема на основі диференціальних рівнянь, методів нечіткої логіки та нейронних мереж [5]. Такий підхід сприяє створенню інтелектуальних систем контролю мікроклімату, що забезпечують зниження енергоспоживання, підвищення стабільності технологічних процесів і мінімізацію негативного впливу на довкілля.

Авторами роботи [6] запропоновано методику проектування системи керування мікрокліматом приміщення на основі цифрового автомата, який має 8 вхідних і 8 вихідних сигналів та 6 перехідних станів. Запропоноване рішення дозволяє формалізувати логіку функціонування системи та забезпечити її структуровану реалізацію відповідно до заданих умов роботи.

Отже, застосування автоматизованих систем керування мікроклімату у виробничих приміщеннях є важливим чинником забезпечення безпечних умов праці, збереження здоров'я персоналу та підвищення загальної ефективності функціонування підприємства.

Постановка завдання. Метою роботи є розроблення та дослідження автоматизованої системи забезпечення оптимального мікроклімату виробничого приміщення на основі алгоритмічного визначення кількості кліматичного обладнання та реалізації програмного керування температурними параметрами з урахуванням нормативних вимог і принципів енергоефективності.

Виклад основного матеріалу. Сучасні системи підтримання мікроклімату у виробничих приміщеннях неможливо розглядати без застосування автоматизованих систем управління (АСУ) [7]. Упровадження АСУ створює умови для раціоналізації роботи кліматичного обладнання та суттєвого скорочення експлуатаційних витрат завдяки оптимізації споживання енергоресурсів. В умовах енергетичної нестабільності та воєнних дій особливого значення набуває точне автоматичне регулювання параметрів мікроклімату, що дає змогу мінімізувати втрати енергії, запобігати перевитратам і забезпечувати безперебійну роботу об'єктів навіть за обмежених ресурсів.

Водночас, питання зниження енерговитрат набуває особливої актуальності у зв'язку зі стійкою тенденцією до зростання вартості енер-

горесурсів [8; 9]. Це зумовлює необхідність впровадження енергозберігаючих технологій та обладнання, а також реалізації організаційних заходів, спрямованих на раціональне використання теплової та електричної енергії у системах забезпечення мікроклімату.

З точки зору енергоуправління до ефективних заходів належить застосування автоматичного регулювання температури залежно від часу доби, дня тижня або режиму переривчастого опалення. Інтеграція сучасного енергозберігаючого обладнання з АСУ дозволяє підтримувати необхідний рівень температури та вологості повітря з високою точністю, мінімізуючи при цьому витрати енергії.

Ефективне функціонування автоматизованої системи забезпечення оптимального мікроклімату у виробничому приміщенні передбачає чітке розмежування її функціональних складових. Структуризація функцій дозволяє систематизувати процеси збору інформації, прийняття керуючих рішень та підтримання стабільної роботи технічних засобів. З урахуванням вимог до сучасних систем автоматичного керування, доцільно виділити три основні групи функцій: інформаційні, керуючі та допоміжні [5; 7].

Узагальнену характеристику зазначених функціональних складових наведено у таблиці 1.

Як видно з таблиці 1, кожна функціональна група виконує визначену роль у загальній структурі системи. Інформаційні функції забезпечують формування достовірних даних про стан повітряного середовища, керуючі – реалізують алгоритм підтримання нормативних параметрів, а допоміжні – гарантують надійність та безперервність роботи системи. Даний розподіл дозволяє підвищити ефективність АСУ та створити основу для подальшого удосконалення системи з урахуванням сучасних вимог до енергоефективності та цифровізації виробничих процесів.

До керуючих функцій автоматизованої системи забезпечення мікроклімату належать визначення раціонального режиму регулювання параметрів повітряного середовища та формування керуючих впливів, які подаються на входи виконавчих механізмів для реалізації обраного режиму роботи. Зокрема, система забезпечує вибір оптимальних значень температури, вологості та інших мікрокліматичних показників залежно від заданих умов експлуатації та поточного стану виробничого приміщення.

Функціонування АСУ підтримання оптимального мікроклімату у виробничому приміщенні забезпечується за рахунок взаємодії програмного та інформаційного забезпечення [3; 10]. Їх узгоджена робота дозволяє реалізувати повний цикл автоматизованого керування – від збору та обробки даних до формування керуючих впливів і збереження результатів функціонування системи. Структуровану характеристику складових програмного та інформаційного забезпечення наведено у таблиці 2.

Як видно з таблиці 2, програмне забезпечення виконує ключову роль у реалізації алгоритмів керування та формуванні керуючих впливів, тоді як інформаційне забезпечення створює основу для накопичення, збереження та аналізу технологічних даних. Їх інтеграція забезпечує підвищення точності регулювання параметрів мікроклімату, оперативність прийняття рішень та надійність функціонування системи в цілому.

Слід зазначити, що до основних функціональних можливостей програмного забезпечення належать:

- аналіз вхідних даних щодо геометричних характеристик виробничого приміщення (довжина, ширина, висота, об'єм), що дозволяє оцінити теплові навантаження та визначити потребу в кліматичному обладнанні;
- формування рекомендацій щодо кількості та потужності кондиціонерів або інших кліматич-

Таблиця 1

Функціональні складові автоматизованої системи забезпечення оптимального мікроклімату

Функціональна група	Зміст функцій	Основні завдання	Результат реалізації
Інформаційні функції	Збір, передача та обробка даних про параметри мікроклімату	Вимірювання температури, вологості, швидкості руху повітря; реєстрація даних; збереження в базі даних	Формування достовірної інформаційної бази для прийняття керуючих рішень
Керуючі функції	Аналіз відхилень та формування керуючих впливів	Порівняння поточних і нормативних значень; вибір режиму регулювання; передача сигналів на виконавчі пристрої	Підтримання заданих параметрів мікроклімату в допустимих межах
Допоміжні функції	Забезпечення стабільної роботи системи	Контроль стану обладнання; діагностика; сигналізація аварій; архівування даних; взаємодія з оператором	Підвищення надійності та безпеки функціонування системи

Складові програмного та інформаційного забезпечення автоматизованої системи управління підтримання мікроклімату

Складова	Основне призначення	Функціональні можливості	Результат реалізації
Програмне забезпечення	Реалізація алгоритмів автоматичного керування	Збір даних від датчиків; первинна обробка інформації; аналіз відхилень; формування керуючих сигналів; взаємодія з оператором	Автоматичне підтримання нормативних параметрів мікроклімату
Модуль обробки даних	Підвищення достовірності інформації	Фільтрація, нормалізація та перетворення вимірювальних сигналів	Забезпечення точності прийняття керуючих рішень
Інформаційне забезпечення	Зберігання та систематизація даних	Бази даних; архів параметрів; журнал подій; історія змін режимів	Можливість аналізу ефективності роботи системи
Візуалізаційний модуль	Відображення поточних параметрів	Графічний інтерфейс; індикація режимів роботи; повідомлення про аварії	Підвищення зручності експлуатації та контролю

них установок, необхідних для забезпечення нормативних параметрів мікроклімату;

- приймання та оброблення інформації від вимірювальних пристроїв, зокрема датчиків температури повітря, вологості та інших контрольованих параметрів;

- здійснення первинної фільтрації, нормалізації та аналізу отриманих даних з метою підвищення достовірності результатів вимірювання;

- реалізація алгоритмів регулювання, що забезпечують підтримання заданого температурного режиму шляхом керування роботою опалювальних приладів, вентиляційних установок або кондиціонерів;

- автоматичне прийняття рішень відповідно до заданого алгоритму функціонування системи з урахуванням поточного стану середовища;

- формування та передача керуючих сигналів у канал управління для впливу на виконавчі механізми;

- відображення поточних параметрів мікроклімату, режимів роботи системи та службових повідомлень на дисплеї терміналу;

- забезпечення інтерактивного діалогу з оператором, включаючи можливість зміни налаштувань, введення нових параметрів та перегляд архівних даних.

Даний підхід дозволяє не лише автоматизувати процес підтримання нормативних параметрів, але й створити передумови для подальшої оптимізації енергоспоживання та впровадження інтелектуальних алгоритмів керування.

Для формалізації процесів проєктування та функціонування автоматизованої системи підтримання мікроклімату доцільним є використання алгоритмічного підходу. Алгоритмізація дозволяє структуровано представити послідовність виконання розрахункових операцій, процедур

оброблення інформації та формування керуючих впливів. Графічне подання алгоритмів у вигляді блок-схем забезпечує наочність логіки роботи системи, спрощує її програмну реалізацію та підвищує зрозумілість для користувача й розробника.

У зв'язку з цим авторами у роботі розроблено два взаємопов'язані алгоритми: алгоритм визначення кількості кліматичного обладнання (рис. 1) та алгоритм функціонування АСУ мікрокліматом (рис. 2).

Процедура визначення кількості кондиціонерів реалізується відповідно до встановленого алгоритму, рис. 1.



Рис. 1. Алгоритм визначення кількості кліматичного обладнання

Альтернативний алгоритм визначення кількості кліматичного обладнання базується на розрахунку сумарного теплового навантаження виробничого приміщення. На відміну від спрощених методів, що враховують лише геометричні параметри приміщення, запропонований підхід дозволяє комплексно оцінити вплив основних джерел теплопритоків та забезпечити більш обґрунтований вибір кліматичного обладнання.

На першому етапі здійснюється введення вихідних даних, до яких належать геометричні параметри приміщення (ширина A , довжина B , висота H), кількість працівників n_{pp} , тепла потужність встановленого обладнання $Q_{об}$, характеристики огорожувальних конструкцій (коефіцієнти теплопередачі), а також нормативні параметри температурного режиму T_n . На основі цих даних обчислюється об'єм приміщення V , що використовується при визначенні теплопритоків через огороження та вентиляційні процеси.

Далі виконується розрахунок складових сумарного теплового навантаження. До нього входять теплопритоки через огорожувальні конструкції $Q_{констр}$, тепловиділення від технологічного обладнання $Q_{об}$, теплопритоки від персоналу $Q_{люд}$ та надходження тепла від сонячної радіації $Q_{сонц}$. З метою врахування можливих коливань навантаження, похибок розрахунків та пікових режимів роботи вводиться коефіцієнт запасу $k_{зап}$.

На наступному етапі здійснюється вибір типу кондиціонера із довідкової таблиці технічних характеристик, де вказана холодопродуктивність одного агрегату $Q_{од}$. Отримане значення округлюється до більшого цілого числа з метою забезпечення гарантованого покриття розрахункового теплового навантаження.

Завершальним етапом алгоритму є формування рекомендацій щодо кількості та типу кліматичного обладнання, яке необхідно встановити у заданому виробничому приміщенні або робочій зоні.

Запропонований алгоритм забезпечує більш точне визначення потреби у кліматичному обладнанні порівняно зі спрощеними методами та створює основу для подальшої оптимізації енергоспоживання і впровадження автоматизованих систем керування мікрокліматом.

Алгоритм функціонування АСУ мікроклімату (рис.2) передбачає послідовне виконання операцій збору, аналізу та оброблення інформації про стан повітряного середовища.

Алгоритм функціонування автоматизованої системи управління мікрокліматом передбачає послідовну реалізацію етапів ініціалізації, без-

перервного моніторингу параметрів повітряного середовища, оцінювання відхилень, формування керуючих впливів та забезпечення безпеки роботи системи.

Ініціалізація системи:

- Запуск програмного забезпечення;
- Завантаження нормативних параметрів (уставок): температура $T_{зад}$, (за наявності) вологість RH_{min} , (за наявності) CO_2 /запиленість тощо
- Встановлення режиму роботи
- Перевірка працездатності датчиків

Цикл моніторингу (постійно з періодом Δt):

- Зчитати поточні значення
- Виконати первинну обробку
- Записати дані в журнал/архів

Оцінювання відхилень:

- Обчислити похибку: $eT = T_{зад} - T(t)$
- Ввести зону нечутливості/гістерезис ΔT

Формування керуючих дій (Авто/Еко/Ручний)

Контроль обмежень і безпеки

Індикація та взаємодія з оператором

- Відобразити на екрані: $T(t)$, стан обладнання (ON/OFF), режим (Авто/Еко/Ручний), повідомлення/помилки.
- За необхідності сформулювати попередження: «перевищення меж», «відмова датчика», «перевищення часу регулювання».

Завершення циклу

Рис. 2. Алгоритм функціонування автоматизованої системи управління мікрокліматом

На етапі ініціалізації здійснюється запуск програмного забезпечення та встановлення зв'язку з вимірювальними датчиками і виконавчими механізмами. Система завантажує нормативні значення параметрів мікроклімату – уставки температури $T_{зад}$, допустимі межі T_{min} та T_{max} , а за необхідності – параметри вологості, концентрації CO_2 чи інших показників якості повітря. Також визначається режим функціонування системи (автоматичний, ручний або енергозберігаючий), після чого виконується процедура самодіагностики датчиків і перевірки працездатності обладнання.

Основу роботи системи становить циклічний процес моніторингу з періодом дискретизації Δt . У кожному циклі здійснюється зчитування поточних значень температури $T(t)$ та інших контрольованих параметрів. Отримані дані проходять первинну обробку: фільтрацію випадкових шумів, перевірку коректності вимірювань, виявлення відмов або виходу значень за допустимі межі. Після цього інформація реєструється у журналі або архіві для подальшого аналізу.

На етапі оцінювання відхилень визначається похибка регулювання $e_1 = T_{\text{зад}} - T(t)$. Для запобігання частим вмиканням виконавчих механізмів використовується зона нечутливості (гістерезис) ΔT . Якщо температура знижується нижче допустимого рівня з урахуванням гістерезису, формується команда на нагрівання; при перевищенні верхньої межі – на охолодження; у межах допустимого діапазону система переходить у режим утримання або очікування.

Формування керуючих дій залежить від обраного режиму роботи. В автоматичному режимі система самостійно визначає необхідність увімкнення нагріву або охолодження та задає рівень потужності (ступінчасто або пропорційно величині похибки). В енергозберігаючому режимі допустимий діапазон температур розширюється, обмежується максимальна потужність обладнання та може застосовуватися регулювання за розкладом. У ручному режимі керування здійснюється оператором, однак із програмним блокуванням небезпечних значень, що запобігає перегріву або переохолодженню приміщення.

Окремим етапом алгоритму є контроль технічних обмежень і безпеки. Перевіряються мінімальні інтервали роботи компресора, паузи між пусками, допустимі режими навантаження. У разі виявлення несправності датчиків або некоректних даних система переходить в аварійний режим, вимикає виконавчі пристрої або переводить їх у безпечний стан, формує повідомлення для оператора та реєструє подію в журналі.

Алгоритм передбачає постійну індикацію поточних параметрів і стану обладнання на інтерфейсі користувача. На дисплеї відображаються значення температури, режим роботи системи, стан виконавчих механізмів та повідомлення про попередження або помилки.

Після завершення кожного циклу система очікує наступний інтервал Δt та повертається до етапу моніторингу, забезпечуючи безперервне регулювання параметрів мікроклімату.

Таким чином, запропонований алгоритм забезпечує адаптивне, енергоефективне та безпечне функціонування автоматизованої системи управління мікрокліматом виробничого приміщення з можливістю роботи в різних режимах та інтеграції в сучасні цифрові системи керування.

Для створення програмного забезпечення було обрано середовище розробки Delphi 7 [11]. Дане середовище дозволяє швидко створювати автономні програмні продукти з графічним

інтерфейсом користувача та забезпечує зручну інтеграцію з апаратними компонентами системи. Використання Delphi дає можливість реалізувати логіку алгоритму керування, забезпечити оброблення вимірювальних даних у реальному часі, а також організувати зручний діалог із оператором. Крім того, середовище підтримує створення надійних програмних модулів для систем технічного спрямування, що робить його доцільним для розроблення прикладних автоматизованих систем керування.

Виходячи з цього, реалізований алгоритмічний підхід дозволяє забезпечити адаптивне керування мікрокліматом виробничого приміщення з урахуванням його геометричних характеристик та поточного стану повітряного середовища, що сприяє підвищенню енергоефективності та стабільності роботи системи.

Висновки. У роботі розглянуто питання автоматизованого забезпечення нормативних показників мікроклімату у виробничих приміщеннях з урахуванням сучасних вимог до енергоефективності та безпеки праці. Обґрунтовано доцільність використання автоматизованої системи керування, що забезпечує безперервний моніторинг температурних параметрів і формування керуючих впливів відповідно до заданого алгоритму.

Показано, що впровадження програмного забезпечення для реалізації алгоритмів керування дозволяє автоматизувати процес підтримання мікроклімату, підвищити точність регулювання та зменшити енергоспоживання за рахунок раціонального використання кліматичного обладнання. Запропоноване рішення сприяє покращенню умов праці, підвищенню продуктивності персоналу та зниженню експлуатаційних витрат підприємства.

Розроблена система автоматизованого забезпечення оптимального мікроклімату може бути використана як у виробничих приміщеннях різного призначення, так і в навчальних та дослідницьких цілях для моделювання процесів регулювання температурного режиму.

Подальші дослідження можуть бути спрямовані на розширення функціональних можливостей системи шляхом інтеграції додаткових параметрів мікроклімату, зокрема вологості повітря, швидкості його руху та концентрації шкідливих речовин. Перспективним є впровадження адаптивних або оптимальних алгоритмів управління, що враховують змінні теплові навантаження та режим роботи підприємства.

Список літератури:

1. Постанова Міністерство охорони здоров'я України "Санітарні норми мікроклімату виробничих приміщень ДСН 3.3.6.042-99" № 42 від 01.12.99. URL: <https://zakon.rada.gov.ua/rada/show/va042282-99#Text> (дата звернення 28.02.2026).
2. Про основні вимоги та заходи з нормалізації мікрокліматичних умов на робочих місцях. URL: <https://oppb.com.ua/news/pro-osnovni-vymogy-ta-zahody-z-normalizaciyi-mikroklimatychnyh-umov-na-robochyh-miscyah> (дата звернення 28.02.2026).
3. Строкань О. В. Система автоматизованої підтримки оптимального мікроклімату виробничого приміщення. *Системи обробки інформації*. 2014. Вип. 5. С. 97-100. URL: http://nbuv.gov.ua/UJRN/soi_2014_5_25
4. Mykytuyk Z., Shymchysyn O., Zvorskyi A., Markiv D. Automation of microclimate control processes in residential premises. *ICTEE*. 2024. Вип.4, № 1: 155-162. <https://doi.org/10.23939/ict2024.01.155>.
5. Голод І. Математичні моделі інтелектуального керування мікрокліматом на виробництві з використанням кіберфізичних систем. *Digital innovation & sustainable development 2024 : Proceedings of I-st International Conference Kharkiv*, 2024. Рр. 30-31.
6. Лобов В., Єфіменко Л., Бойко С., Городній О. Методика проектування системи керування мікроклімату в приміщенні. *Технічні науки та технології*. 2022. № 1(27), С.172–183. [https://doi.org/10.25140/2411-5363-2022-1\(27\)-172-183](https://doi.org/10.25140/2411-5363-2022-1(27)-172-183).
7. Величко І. В., Сідлецький В. М. Система автоматизації мікроклімату в закладах громадського харчування з модулем усунення невизначеностей у регулюванні температури за допомогою машинного навчання. *International Scientific Technical Journal Problems of Control and Informatics*. 2024. № 69(4). Рр. 46-58. DOI:10.34229/1028-0979-2024-4-3.
8. Про необхідність впровадження енергоефективних заходів. Смолінська селищна Рада. 2021. URL: <https://smolino-rada.gov.ua> (дата звернення 28.02.2026).
9. Енергетична криза спровокує найважчий економічний період в Україні – Reuters.ua.news. 2026. URL: https://ua.news.ua/war-vs-rf/energetichna-kriza-provokuie-naivazhchii-ekonomichnii-period-v-ukrayini-reuters?utm_source=ukrnet_news (дата звернення 28.02.2026).
10. Оптимальні параметри мікроклімату на робочому місці. Південноукраїнська міська Рада. 2025. URL: https://yumk.ua/news/show/optimalni_parametri_mikroklimatu_na_robochomu_misti?#:~:text= (дата звернення 28.02.2026).
11. Сидоров М.В. Елементи програмування у середовищі Delphi. URL: <https://sociology.knu.ua/sites/default/files/course/materials/mycoursedelphi.pdf> (дата звернення 28.02.2026).

Pelykh S.M., Lozhechnikov V.F., Ziierov O.R. ALGORITHM-ORIENTED DIGITAL SYSTEM FOR CONTROLLING THE MICROCLIMATE OF PRODUCTION PREMISES

The article discusses the problem of ensuring standard microclimate indicators in production facilities, taking into account modern requirements for energy efficiency and occupational safety. Maintaining optimal air temperature values is one of the key factors in preserving the health of employees, increasing their productivity and reducing production risks. The rising cost of energy resources necessitates the introduction of automated control systems that minimise operating costs and ensure the rational use of climate control equipment.

The aim of the work is to develop an automated system for ensuring an optimal microclimate in production premises, which combines an algorithm for determining the required number of air conditioners depending on the geometric parameters of the premises and software implementation of automatic temperature control based on feedback. The work proposes a structural diagram of the system, consisting of a design module and a functioning module, which provide for the calculation of equipment parameters, the collection and processing of measurement information, the formation of control actions and the display of results on the user interface.

An algorithm for determining the number of air conditioners has been developed based on comparing the characteristics of the room with the standard and tabulated values of thermal loads. The system's operating algorithm involves constant monitoring of the temperature in the working area, comparing it with the established standard values, and generating the appropriate control signals for the executive devices. The software implementation of the system was carried out in the Delphi environment, which provides a convenient graphical interface, real-time data processing and interactive communication with the operator.

The proposed system ensures increased accuracy in maintaining the temperature regime, reduced energy consumption and the creation of stable and safe working conditions. The results obtained can be used to modernise climate control systems in industrial premises for various purposes and for the further development of automated energy-efficient technologies.

Keywords: *automated control system, energy efficiency, climate control equipment, microclimate, energy consumption optimisation, software, temperature control.*

Дата першого надходження статті до видання: 28.02.2026

Дата прийняття статті до друку після рецензування: 25.03.2026

Дата публікації (оприлюднення) статті 11.05.2026

Писаренко Ю.В.

<https://orcid.org/0000-0001-8357-8614>

Інститут кібернетики імені В. М. Глушкова Національної академії наук України

Кармазін К.В.

<https://orcid.org/0009-0003-8852-8079>

Інститут кібернетики імені В. М. Глушкова Національної академії наук України

ЦИФРОВИЙ ДВІЙНИК МЕРЕЖІ ТА ФЕДЕРАТИВНЕ МУЛЬТИАГЕНТНЕ НАВЧАННЯ ДЛЯ РЕЗИЛЬЄНТНОГО КЕРУВАННЯ ПОТОКОВИМ ВІДЕО І ТЕЛЕМЕТРІЄЮ В СИТУАЦІЙНИХ ЦЕНТРАХ

У статті розглянуто проблему використання штучного інтелекту (ШІ) для обробки та аналізу великих обсягів даних у реальному часі з інтеграцією мультиагентних систем у середовища підвищеної динаміки та невизначеності. Актуальність дослідження зумовлена стрімким зростанням потоків даних, що генеруються безпілотними літальними апаратами, сенсорними мережами, робототехнічними платформами та мультимедійними сервісами. У таких умовах традиційні централізовані підходи до обробки інформації виявляються недостатньо ефективними через обмеження пропускної здатності каналів зв'язку, затримки передачі та нестабільність мережевої інфраструктури.

Запропоновано концепцію інтелектуальної резильєнтної архітектури, яка поєднує методи машинного навчання, навчання з підкріпленням, прогнозування часових рядів і генеративні моделі з мультиагентною координацією. Розроблено математичну модель оптимізації якості функціонування системи на основі функціоналу винагороди, що враховує затримку передачі, втрати пакетів та стабільність показників якості сервісу. У роботі формалізовано підхід до побудови мультиагентної системи, де окремі агенти (дрони, edge-вузли, ситуаційні центри) взаємодіють через узгоджені стратегії керування потоками даних.

Окрему увагу приділено створенню цифрового двійника середовища для прогнозування майбутніх станів системи та формування адаптивної політики керування. Показано, що інтеграція генеративних моделей дозволяє компенсувати втрати даних і підвищувати суб'єктивну якість сприйняття інформації без збільшення навантаження на мережу. Запропонований підхід може бути застосований у системах моніторингу повітряного простору, техноекологічних подій, безпекових комплексах та мультимедійних поточкових сервісах.

Отримані результати формують науково-методичну основу побудови резильєнтних інтелектуальних інформаційних систем нового покоління, здатних до самоадаптації, кооперації агентів та стійкого функціонування в умовах динамічного середовища.

Ключові слова: штучний інтелект, обробка даних у реальному часі, мультиагентні системи, резильєнтність, машинне навчання, навчання з підкріпленням, генеративні моделі, цифровий двійник.

Постановка проблеми. Ситуаційні центри (СЦ), що забезпечують моніторинг, аналіз і підтримку прийняття рішень у реальному часі, дедалі більше залежать від стабільності потоків даних, насамперед відео та телеметрії з мобільних і стаціонарних сенсорів. У сценаріях техноекологічних інцидентів, подвійного призначення критичною стає не «максимальна» якість відео, а гарантована безперервність, прогнозованість

затримок і узгоджена робота багатьох джерел у межах одного інформаційного контуру.

Однак реальне мережеве середовище СЦ є нестабільним: пропускна здатність коливається, зростає ймовірність втрат пакетів, змінюється топологія (мобільні вузли, тимчасові ретранслятори), а транспортні механізми можуть переходити на нові протоколи, що змінює поведінку затримок та мультиплексування.

У таких умовах класичні схеми адаптивної передачі відео (ABR) часто оптимізують «середню» якість користувацького досвіду (Quality of Experience, QoE), але недостатньо враховують стійкість цього досвіду в часі – тобто здатність системи підтримувати прийнятну якість під зовнішніми збуреннями без різких провалів та частих перемикань. Стандартизовані підходи об'єктивного оцінювання QoE описують типові деградації адаптивного потокового відео (зупинки відтворення через буферизацію, перемикання рівнів якості тощо), але самі по собі не є механізмом керування: потрібне поєднання моделей оцінки якості з механізмами прогнозування та оптимізації політики вибору параметрів потоку.

Додатково проблема ускладнюється багато-потоковістю: у СЦ одночасно можуть працювати десятки потоків (різні БПЛА, наземні роботи, стаціонарні камери), і рішення «покращити QoE одного потоку» може погіршити узгодженість системи загалом. Для таких сценаріїв актуальні постановки колективної оптимізації QoE та справедливого розподілу ресурсів між агентами, що вже розглядаються у підходах з багатоагентною координацією до потокового відео.

Нарешті, практична інженерія вимагає масштабування: регіональні/мобільні СЦ мають різні умови мережі та різну «статистику збурень», а дані відео/телеметрії можуть бути чутливими. Це піднімає дилему: як навчати та оновлювати інтелектуальні політики (прогнозування, RL/MA-RL, реконструкція) без централізації сирих даних? Проблема формулюється як необхідність створення резильєнтної архітектури керування потоками даних, що поєднує: цифровий двійник мережі для замкненого керування та безпечного тестування; федеративне навчання для перенесення моделей між СЦ; мультиагентну оптимізацію QoE/резильєнтності в багатопотоковому середовищі.

Аналіз останніх досліджень і публікацій. Стандартизованою основою сучасного HTTP-адаптивного стримінгу є MPEG-DASH (ISO/IEC 23009-1) [1], який визначає формат опису презентації (MPD) та сегментів і підтримує модель, де клієнт керує вибором сегментів/якості через HTTP-запити. Це безпосередньо корелює з постановкою задачі ABR як задачі керування політикою вибору якості в часі.

Оцінювання якості в адаптивному потоковому відео може бути стандартизовано через ITU-T P.1203 [2]: рекомендація описує параметричні модулі оцінки аудіо- та відеоякості з урахуванням

зупинок відтворення через буферизацію та початкових затримок. Для ситуаційних центрів це цінно як база відтворюваних метрик, на які можна спиратися під час формування функцій винагороди в алгоритмах навчання з підкріпленням або критеріїв оптимізації в багатоагентному керуванні.

З погляду транспортного шару перехід до протоколів QUIC [3] і HTTP/3 [4] є важливим через можливість мультиплексування потоків, скорочення затримки встановлення з'єднання та інтегрований криптографічний захист (TLS 1.3 для QUIC). Для місійних систем ситуаційного центру це означає зміну профілю затримок і втрат, а отже – потребу адаптації моделей прогнозування та механізмів контролю.

У домені адаптивного вибору бітрейту існують сильні «класичні» алгоритми. Метод BOLA [5] демонструє формалізацію вибору бітрейту як задачі максимізації корисності з використанням оптимізації Ляпунова та керування буфером, що робить його важливою базовою моделлю для порівняння з інтелектуальними підходами.

Робота Pensieve [6] стала канонічним прикладом навчання політик адаптивної передачі відео на основі навчання з підкріпленням, показавши, що стратегія може бути сформована на основі досвіду без жорстко закодованих евристик. Водночас, як і для будь-яких підходів навчання з підкріпленням у мережесередовищах, залишається питання перенесення моделей між різними умовами та чутливості до репрезентативності навчальних сценаріїв, що особливо актуально для різних регіональних ситуаційних центрів.

Для багатопотокового режиму та забезпечення справедливого розподілу ресурсів між потоками перспективними є підходи багатоагентної оптимізації QoE [7], де оптимізаційна мета задається як нелінійна функція колективних винагород (якість досвіду і справедливість). Такі постановки формують методологічну основу для ситуаційних центрів, у яких потоки мають різний рівень критичності та різні угоди щодо рівня сервісу (SLA) і пріоритетів.

Окремий сучасний напрям – генеративні методи компенсації втрат і артефактів у адаптивному потоковому відео через HTTP. Концепція застосування генеративного штучного інтелекту для адаптивного потокового відео [8] ставить питання: чи можливо підвищити суб'єктивну якість за рахунок локальної реконструкції на рівні периферійних обчислень (edge) замість збільшення бітрейту або агресивного зниження якості під час мережесередовищних провалів.

Для системного переходу до масштабованих рішень ключовою є концепція мережі цифрового двійника (Digital Twin Network) [9]. У рекомендації ITU-T Y.3090 така мережа визначається як віртуальне представлення фізичної інфраструктури, призначене для аналізу, діагностики, емуляції та контролю з акцентом на замкнений цикл «дані – модель – інтерфейс – взаємодія віртуального та фізичного середовищ». Це забезпечує саме той операційний механізм, якого бракує багатьом рішенням на основі навчання з підкріпленням і багатоагентного навчання: можливість безпечного навчання, валідації та аналізу сценаріїв «що буде, якщо» на цифровій копії перед впровадженням у реальній мережі ситуаційного центру.

Паралельно федеративне навчання (FedAvg) [10] формує базовий підхід до тренування моделей на розподілених даних без централізованого збирання сирих записів, що природно відповідає вимогам конфіденційності у ситуаційних центрах. Для підвищення приватності та стійкості у федеративному навчанні застосовуються методи захищеної агрегації [11], а для поєднання федеративного навчання з керуванням на основі навчання з підкріпленням [12] розробляються спеціалізовані підходи до федеративного навчання з підкріпленням на периферійних вузлах.

Отже, на стикі стандартизованого адаптивного потокового відео (DASH), формалізованих моделей оцінки якості (P.1203), сучасних транспортних протоколів (QUIC/HTTP/3), мереж цифрових двійників та федеративного й багатоагентного навчання з підкріпленням формується цілісна наукова база для переходу від ізольованих алгоритмів до інтегрованої архітектури резильєнтного керування потоками в системах ситуаційних центрів.

Постановка завдання. Метою дослідження є розроблення й обґрунтування інтегрованої архітектури резильєнтного керування потоковим відео та телеметрією у ситуаційних центрах, що ґрунтується на концепції мережі цифрового двійника та федеративного багатоагентного навчання політик керування якістю користувацького досвіду й резильєнтною якістю в умовах нестабільного мережевого середовища та обмежених ресурсів периферійних обчислювальних вузлів.

Для досягнення мети поставлено такі задачі:

1. Формалізувати системну модель «ситуаційний центр – периферійні вузли – агенти» та вимоги до резильєнтності на основі стандартів адаптивного потокового відео та протоколів транспортного рівня;

2. Описати архітектуру мережі цифрового двійника для ситуаційного центру з інтерфейсами взаємодії між віртуальним і фізичним середовищами та можливістю імітаційного тестування;

3. Запропонувати розширену метрику резильєнтної якості для багатопотокового режиму (RQE-M), узгоджену з показниками якості користувацького досвіду та критеріями справедливого розподілу ресурсів;

4. Сформулювати принципи багатоагентного керування потоками (локальні агенти адаптивного вибору бітрейту з кооперативною координацією);

5. Запропонувати федеративну схему навчання моделей прогнозування та керування з можливістю застосування механізмів захищеної агрегації;

6. Визначити методологію експериментальної перевірки в середовищі цифрового двійника або спеціалізованому імітаційному симуляторі, включаючи критерії якості.

Виклад основного матеріалу

1. Архітектура системи «СЦ – периферійні вузли – агенти» та роль стандартів

Вихідною умовою приймається, що мультимедійні потоки передаються у форматі адаптивного потокового відео (MPEG-DASH), де клієнт або периферійний проксі-вузол здійснює вибір рівня якості сегментів на основі доступної інформації про стан мережі, буфера та контексту. Стандарт ISO/IEC 23009-1 забезпечує інтероперабельність опису контенту (опис представлення медіа) та механізмів сегментації, що дозволяє зосередити увагу на політиці адаптивного вибору бітрейту як на інтелектуальному рівні керування поверх базового стандарту.

Для оцінювання якості застосовується підхід, узгоджений із рекомендацією ITU-T P.1203 (як базова об'єктивна модель оцінки якості користувацького досвіду для адаптивного потокового відео). Це дозволяє не створювати метрики «з нуля», а формувати показники резильєнтної якості як надбудову над стандартизованими компонентами (якість відео та аудіо, зупинки відтворення через буферизацію, інтегральна оцінка).

Транспортний рівень розглядається як такий, що може функціонувати поверх TCP/HTTP2 або переходити до QUIC/HTTP/3. Оскільки HTTP/3 реалізує семантику HTTP поверх протоколу QUIC, а QUIC визначає мультиплексовані потоки та механізми керування переважанням і втратами, це безпосередньо впливає на параметри, релевантні для якості користувацького досвіду (затримки, втрати пакетів, варіативність затримки). Відповідно, ці характеристики мають

бути враховані у цифровому двійнику мережі та в прогнозних моделях.

2. Цифровий двійник мережі як «замкнений контур» керування

Відповідно до рекомендації ITU-T Y.3090 мережа цифрового двійника визначається як віртуальне представлення фізичної мережі, придатне для аналізу, діагностики, емуляції та контролю із забезпеченням інтерактивного відображення в реальному часі між фізичною та цифровою інфраструктурою. Ключовою є підтримка замкненого контуру автоматизації: дані фізичної мережі → оновлення цифрових моделей → прийняття рішень → виконання у фізичній мережі → надходження нових даних.

Для задач ситуаційного центру пропонується реалізувати мережу цифрового двійника як сукупність моделей трьох типів:

1. Рівень даних: телеметрія мережі (пропускна здатність, затримка проходження сигналу, втрати пакетів), стани буфера клієнтів і периферійних вузлів, характеристики відеоконтенту (наприклад, складність сцени), просторовий контекст агентів;

2. Рівень моделей: моделі прогнозування (рекурентні нейронні мережі типу LSTM/GRU або інші підходи), моделі оцінки якості (узгоджені з ITU-T P.1203), моделі політики керування (навчання з підкріпленням або багатоагентне

навчання), а також моделі реконструкції відео (за потреби);

3. Прикладний рівень: сервіси ситуаційного центру (моніторинг, пріоритизація потоків, перемикання маршрутів передавання, керування ресурсами периферійних обчислювальних вузлів).

Така тріада узгоджується з ідеєю трирівневої архітектури мережі цифрового двійника (фізичний рівень – рівень цифрового двійника – прикладний рівень), що підкреслюється в рекомендації Y.3090.

На рівні інтерфейсів доцільно спиратися на вимоги до нижніх (southbound), внутрішніх та верхніх (northbound) інтерфейсів мережі цифрового двійника, сформульовані в документах IRTF/IETF щодо інтерфейсів цифрових двійників мереж. Ці документи конкретизують вимоги до форматів даних, гнучкості розгортання, внутрішньої взаємодії компонентів рівня цифрового двійника та аспектів безпеки. Діаграму концептуальної моделі наведено на рисунку 1.

3. Метрика резильєнтної якості для багато-потокowego ситуаційного центру (RQE-M)

У попередній роботі авторів (Кармазін, Писаренко) [13] запропоновано метрику резильєнтної якості користувацького досвіду (Resilient Quality of Experience, RQE), що враховує не лише середнє

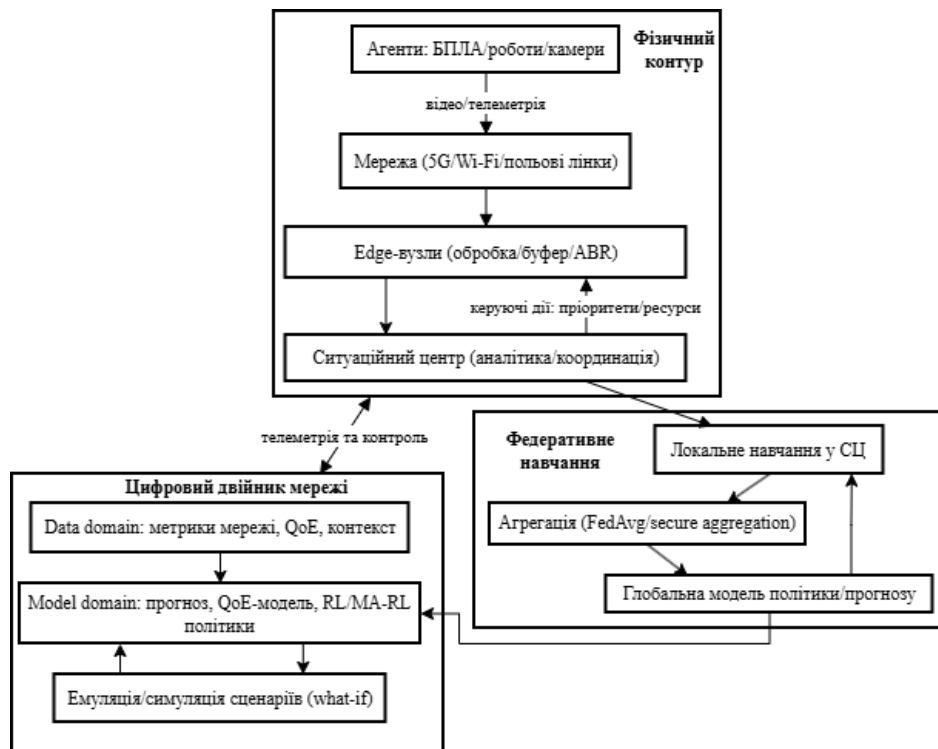


Рис. 1. Діаграма концептуальної моделі

значення показника якості, а й його стабільність у часі. У даній роботі пропонується узагальнення цієї метрики на багатопотоковий режим функціонування ситуаційного центру – RQE-M (багатопотокова резильєнтна якість), яке поєднує:

1. Індивідуальну якість i -го потоку (показник QoE_i або оцінку, розраховану відповідно до рекомендації ITU-T P.1203);
2. Стабільність (варіативність та частоту змін рівня якості);
3. Штрафи за зупинки відтворення через буферизацію та затримки для місійно-критичних потоків;
4. Компонент справедливого розподілу якості між потоками (наприклад, через дисперсію QoE_i або мінімізацію «нижніх хвостів» розподілу якості в групі).

Один з можливих формалізованих варіантів:

$$RQE - M = \frac{1}{N} \sum_{i=1}^N QoE_i - \lambda \frac{1}{N} \sum_{i=1}^N \sigma(QoE_i) - \beta \frac{1}{N} \sum_{i=1}^N Stall_i - \delta FairnessPenalty(QoE_1, \dots, QoE_N)$$

де $\frac{1}{N} \sum_{i=1}^N QoE_i$ – середня якість користувацького досвіду по всіх потоках; $\sigma(QoE_i)$ – нестабільність (варіативність) якості; $Stall_i$ – середній час буферизації (rebuffering); $FairnessPenalty(QoE_1, \dots, QoE_N)$ – штраф за нерівномірність якості між потоками; λ, β, δ – вагові коефіцієнти, що задаються відповідно до пріоритетів ситуаційного центру. Ідея узгоджується з сучасними підходами багатоагентної оптимізації у відеостримінгу, де групова ціль формується як максимізація інтегральної якості користувацького досвіду з урахуванням справедливого розподілу ресурсів.

4. Ієрархія керування потоками: локальні агенти адаптивного вибору бітрейту та кооперативна координація

На нижньому рівні (клієнт або периферійний обчислювальний вузол) функціонують локальні агенти адаптивного вибору бітрейту, які здійснюють вибір якості сегмента відео. Як базові алгоритми для порівняння доцільно використовувати буферно-орієнтовані підходи (наприклад, BOLA) та методи на основі навчання з підкріпленням (алгоритми, подібні до Pensieve).

На верхньому рівні (координатор ситуаційного центру або периферійний координатор) формується кооперативна політика розподілу мережових і обчислювальних ресурсів між потоками. Для цього може застосовуватися багатоагентна постановка на основі градієнтних методів оптимізації політики, де агентами виступають окремі потоки або їх групи, а цільовою функцією є максимізація RQE-M.

Така ієрархічна модель методично узгоджується з підходами багатоагентного навчання з під-

кріпленням для оптимізації якості користувацького досвіду та справедливості розподілу ресурсів у системах потокового відео.

5. Федеративне навчання політик і прогнозних моделей між ситуаційними центрами

З метою уникнення централізації чутливих даних (відео та телеметрії) і водночас підвищення якості моделей для різних регіональних ситуаційних центрів пропонується реалізувати федеративний цикл навчання, що включає такі етапи:

1) Локальне донавчання моделей на даних конкретного ситуаційного центру (прогнозування пропускної здатності, затримки проходження сигналу, втрат пакетів; налаштування параметрів політики адаптивного вибору бітрейту; модель оцінки стабільності якості користувацького досвіду);

2) Передача на сервер агрегації лише градієнтів або оновлень параметрів без передавання сирих даних;

3) Агрегування оновлень за схемою федеративного усереднення (FedAvg);

4) Повернення узгодженої глобальної моделі назад до ситуаційних центрів.

Метод федеративного усереднення описаний як практичний спосіб об'єднання локальних оновлень моделей і демонструє відносну стійкість до неоднорідних (non-IID) розподілів даних, що є критично важливим у разі гетерогенних умов функціонування різних ситуаційних центрів.

З урахуванням підвищених вимог до інформаційної безпеки доцільно інтегрувати механізми захищеної агрегації, які забезпечують можливість отримання сервером лише сумарного результату без доступу до окремих локальних оновлень. Практичні схеми захищеної агрегації та їх сучасні узагальнення є важливим обґрунтуванням цього компонента архітектури.

Для задач, у яких політика керування має характер навчання з підкріпленням і може потребувати мета-адаптації, можливим є застосування спеціалізованих підходів федеративного навчання з підкріпленням на периферійних вузлах (наприклад, FedMC). Водночас доцільно розглядати такі рішення як перспективний напрям або варіант реалізації, щоб не перевантажувати основну лінію аргументації.

6. Генеративне відновлення як опційний модуль на периферійних вузлах

Генеративні підходи в адаптивному потоковому відео можуть розглядатися як додатковий модуль компенсації деградацій. У разі погіршення мережових умов замість агресивного зниження

якості або накопичення зупинок відтворення через буферизацію периферійний вузол може виконувати реконструкцію або згладжування артефактів (наприклад, відновлення відсутніх фрагментів зображення чи інтерполяцію кадрів) з метою зменшення суб'єктивної деградації якості. Такий підхід узгоджується як із напрямом застосування генеративного штучного інтелекту для адаптивного поточкового відео через HTTP, так і з ширшим розвитком дифузійних моделей у задачах відео обробки.

Важливим методичним аспектом для ситуаційних центрів є те, що генеративний модуль не повинен збільшувати затримку понад допустимий часовий бюджет. Тому доцільно явно задати правило активації (механізм «дозволу»), відповідно до якого реконструкція допускається лише за умов:

- 1) очікуване зменшення втрат якості користувачького досвіду або резильєнтності якості перевищує витрати, пов'язані із додатковою затримкою чи енергоспоживанням;
- 2) потік не належить до категорії наднизькозатримкових місійно-критичних потоків;
- 3) периферійний вузол має достатні обчислювальні ресурси.

Такий підхід відповідає принципу інженерної реалізованості для місійно-критичних систем та забезпечує баланс між якістю, затримкою й ресурсними обмеженнями.

7. Методологія експериментальної валідації в цифровому двійнику

Для відтворюваної перевірки резильєнтності пропонується експериментальна методика, що поєднує:

- 1) Мережеве моделювання та емуляцію (наприклад, використання дослідницького мережевого симулятора ns-3);
- 2) Емульовані топології та контрольні сценарії (Mininet як інструмент швидкого прототипування мережевих конфігурацій);
- 3) Робото-технічний контур для агентів (ROS 2 як де-факто стандартна програмна платформа проміжного рівня для робототехніки);
- 4) Референсну реалізацію адаптивного поточкового відео (програвач dash.js із можливістю налаштування алгоритмів адаптивного вибору

бітрейту та використання правила BOLA як еталонного буферно-орієнтованого підходу).

Система показників оцінювання включає: показники якості користувацького досвіду, узгоджені з ITU-T P.1203, або наближені оцінки типу середнього експертного балу (MOS); частку часу зупинок відтворення через буферизацію; частоту та амплітуду змін рівня якості; наскрізну затримку передачі до ситуаційного центру; інтегральну метрику резильєнтності якості RQE-M; показники справедливості між потоками. Це забезпечує як стандартизовану основу оцінювання, так і специфічну для ситуаційного центру надбудову резильєнтності. План експериментів за визначеними сценаріями наведено в таблиці 1.

Висновки. У статті запропоновано архітектурно-методичний розвиток досліджень резильєнтних поточкових систем для ситуаційних центрів: введено мережу цифрового двійника як замкнений контур керування та валідації, а також федеративне навчання як механізм масштабування інтелектуальних політик без централізації чутливих даних. Запропонована архітектура обґрунтована міжнародними рекомендаціями щодо цифрових двійників мереж (ITU-T Y.3090) і узгоджена зі стандартами доставки мультимедійного контенту (MPEG-DASH) та оцінювання якості (ITU-T P.1203).

Показано, що для багатопотокового режиму функціонування ситуаційного центру доцільно переходити від орієнтації лише на середню якість користувацького досвіду до використання метрик резильєнтності та справедливого розподілу ресурсів (RQE-M). Контур керування доцільно будувати ієрархічно: локальні агенти адаптивного вибору бітрейту на нижньому рівні та кооперативний координатор багатоагентної оптимізації якості й справедливості на верхньому рівні. Як технологічні опції розглянуто генеративну компенсацію деградацій на периферійних вузлах та федеративне навчання з підкріпленням як перспективний напрям розвитку.

Запропоновано відтворювану методологію експериментальної перевірки, що базується на використанні цифрового двійника, мережевих

Таблиця 1

План експериментів за визначеними сценаріями

Фактор	Позначення	Рівні
Мережеві збурення	N	N_1 – step throughput drop; N_2 – bursty loss; N_3 – RTT variation
Кількість потоків	S	$S_1=1$; $S_2=5$; $S_3=10$; $S_4=20$
Пріоритет	P	P_1 – критичні; P_2 – некритичні
Транспорт	T	T_1 – TCP/HTTP2; T_2 – QUIC/HTTP3
Архітектура	A	A_1 – Baseline; A_2 – DTN; A_3 – DTN+FL

симуляторів та робототехнічної платформи проміжного рівня з прив'язкою до референсних реалізацій алгоритмів адаптивного вибору бітрейту. Практичне значення роботи полягає у можливості

уніфікованого порівняння політик керування потоками для різних ситуаційних центрів та підвищення стійкості інформаційного контуру прийняття рішень у реальному часі.

Список літератури:

1. ISO/IEC. ISO/IEC 23009-1:2022 Information technology – Dynamic adaptive streaming over HTTP (DASH). Part 1: Media presentation description and segment formats. Geneva, 2022.
2. ITU-T. Recommendation ITU-T P.1203: Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport. Geneva, 2017.
3. IETF. RFC 9000: QUIC: A UDP-Based Multiplexed and Secure Transport. 2021.
4. IETF. RFC 9114: HTTP/3. 2022.
5. Spiteri K., Ugaonkar R., Sitaraman R. BOLA: Near-Optimal Bitrate Adaptation for Online Videos // IEEE/ACM Transactions on Networking. 2016. DOI: 10.1109/TNET.2020.2996964
6. Mao H., Netravali R., Alizadeh M. Neural adaptive video streaming with Pensieve // Proceedings of ACM SIGCOMM. 2017. DOI: 10.1145/3098822.3098843
7. Wang Y., Aggarwal V., Lan T. Learning-based online QoE optimization in multi-agent video streaming // Algorithms. 2022. Vol. 15, No. 7. P. 227. DOI: 10.3390/a15070227
8. Artioli E. Generative AI for HTTP adaptive streaming // Proceedings of ACM Multimedia Systems (MMSys). 2024. DOI: 10.1145/3625468.3652912
9. ITU-T. Recommendation ITU-T Y.3090: Digital twin network – Requirements and architecture. Geneva, 2022.
10. McMahan H. B., Moore E., Ramage D., Hampson S., Arcas B. A. Communication-efficient learning of deep networks from decentralized data // Proceedings of Machine Learning Research. 2017. Vol. 54.
11. Bonawitz K., et al. Federated learning with autotuned communication-efficient secure aggregation // IEEE Symposium on Security and Privacy. 2019. DOI: 10.48550/arXiv.1912.00131
12. Zou D., et al. FedMC: Federated reinforcement learning on the edge with meta-critic networks // IEEE IPCCC. 2022. DOI: 10.1109/IPCCC55026.2022.9894336
13. Писаренко Ю. В., Кармазін К. В. Використання ІІІ для обробки та аналізу великих обсягів даних у реальному часі з інтеграцією мультиагентних систем // Науковий журнал «Штучний інтелект». 2025. DOI: 10.15407/jai2025.04.134

Pisarenko J.V., Karmazin K.V. NETWORK DIGITAL TWIN AND FEDERATED MULTI-AGENT LEARNING FOR RESILIENT VIDEO STREAMING AND TELEMETRY MANAGEMENT IN SITUATION CENTERS

The paper addresses the problem of applying Artificial Intelligence (AI) for real-time processing and analysis of large-scale data streams with the integration of multi-agent systems in dynamic and uncertain environments. The relevance of the research is driven by the rapid growth of data generated by unmanned aerial vehicles, robotic platforms, sensor networks, and multimedia streaming services. Under such conditions, traditional centralized data processing approaches become insufficient due to bandwidth limitations, transmission delays, packet losses, and network instability.

A novel concept of an intelligent resilient architecture is proposed, combining machine learning, reinforcement learning, time-series forecasting, and generative models with multi-agent coordination mechanisms. A mathematical model for optimizing system performance is developed based on a reward function that incorporates delay, packet loss, and service quality stability metrics. The study formalizes a multi-agent framework in which heterogeneous agents (UAVs, edge nodes, situational centers) interact through coordinated data flow management strategies.

Special attention is devoted to the creation of a digital twin of the operational environment for predicting future system states and generating adaptive control policies. It is demonstrated that the integration of generative AI models enables compensation of data losses and improves subjective quality perception without increasing network load. The proposed approach is applicable to airspace monitoring systems, technological surveillance, security infrastructures, and next-generation multimedia streaming platforms.

The obtained results establish a scientific and methodological foundation for the development of resilient intelligent information systems capable of self-adaptation, cooperative agent interaction, and stable operation under dynamic real-time conditions.

Keywords: artificial intelligence, real-time data processing, multi-agent systems, resilience, machine learning, reinforcement learning, generative models, digital twin.

Дата першого надходження статті до видання: 08.03.2026

Дата прийняття статті до друку після рецензування: 03.04.2026

Дата публікації (оприлюднення) статті 11.05.2026

Piotrovskiy Ye. Ye.<https://orcid.org/0009-0009-2406-5566>

Throne Labs Inc., United States

THE ROLE OF AUTOMATED FRONTEND-BACKEND SYNCHRONIZATION IN REDUCING DIGITAL PRODUCT TIME TO MARKET

The relevance of the study is determined by the increasing complexity of digital products, the spread of distributed software architectures, and the need to accelerate development processes in the context of highly dynamic digital markets. The purpose of the article is to justify the role of automated synchronization of client-server interaction in digital systems in reducing time to market for digital products and improving the efficiency of software development processes. The study employs methods of analysis and generalization of scientific sources to systematize contemporary approaches to organizing the interaction of software system components, system-structural analysis to determine the relationships among the stages of the development life cycle, a comparative approach to contrast traditional and automated deployment models, as well as analytical modeling and normalization of process metrics to evaluate the effectiveness of synchronized deployment. As a result of the study, contemporary approaches to organizing client-server interaction within the software development life cycle were analyzed. It has been established that automated synchronization of integration and deployment processes shifts change coordination control to the early stages of development, thereby reducing integration delays and increasing release stability. It has been demonstrated that the implementation of an automated deployment pipeline is accompanied by an increased frequency of production deployments, a reduced proportion of failed updates, a shorter lead time for changes to reach the production environment, and a decreased mean time to recovery after disruptions. A systematic effect of front-end/back-end interaction synchronization on reducing time to market for digital products has been identified. The conclusions substantiate that automated synchronization of interaction among software system components constitutes a methodological foundation for improving the efficiency of digital development, ensuring the coordinated evolution of client and server logic, reducing operational risks, and increasing the predictability of the release process. Prospects for further research are associated with the development of models for quantitative assessment of the economic effect of reducing time to market for digital products, the application of intelligent methods for forecasting integration risks, and the creation of adaptive automated deployment systems capable of optimizing the interaction of digital system components in real time.

Keywords: software lifecycle, continuous integration, continuous deployment, microservice architecture, software interface consistency, software systems engineering, development automation, release management, digital platforms.

Introduction. The modern digital economy imposes increasing demands on the speed of software development and market entry, making the reduction of the development cycle a critical factor in the competitiveness of digital platforms. At the same time, the traditional separation of the client side of software applications (front end, FE) and the server side of software systems (back end, BE) results in asynchronous changes, inconsistencies in application programming interfaces (APIs), a higher incidence of integration errors, and extended release cycles. The growing complexity of software system architectures, the widespread adoption of microservices architecture,

and the implementation of continuous integration and continuous delivery (CI/CD) practices underscore the need for automated mechanisms to synchronize interactions between the client and server layers. Addressing this issue is directly associated with significant scientific and practical objectives, including optimization of the software development life cycle, reduction of technical debt, enhancement of team coordination, and acceleration of time to market for digital products.

Literature Review. A review of current research indicates the gradual formation of a comprehensive scientific concept of synchronized software development,



within which technical, architectural, organizational, and infrastructure solutions are considered interrelated components of a unified digital product life cycle. At the initial stage, studies focus on ensuring continuous data consistency between the FE and BE components of information systems. For example, F. R. de Moraes et al. argue that real time synchronization of microservice application states reduces delays between system components and facilitates faster implementation of new digital product features [1].

This idea is further developed by F. Rasyaad and S. Wibowo, who contend that the use of Backend as a Service (BaaS) architecture enables automated integration of application programming interfaces (APIs) and contributes to reducing the development time of digital monitoring platforms [2]. G. Bortolato and co authors demonstrate that the use of high speed Remote Direct Memory Access (RDMA) networks supports synchronization of computing processes with minimal delay, thereby significantly accelerating the testing of functional changes [3]. S. Guan and colleagues substantiate the effectiveness of digital twins as an environment for continuous synchronization between the user interface and server models, which shortens the testing and validation phase of digital products [4].

Summarizing these approaches, a research team led by M. Maiti concludes that the integration of data flows in financial software facilitates faster implementation of new electronic services through automated coordination of FE and BE levels [5].

Other research approaches address the synchronization of application services in corporate information systems and electronic platforms. In particular, A. Aquino and A. Lino argue that alignment of BE infrastructure configurations with user services directly affects the speed of electronic service delivery and the launch of digital products [6]. Z. Chen and co authors demonstrate that synchronized signal transmission in front end data collection electronics ensures stable integration of information flows in complex digital systems [7]. M. Kumar substantiates the concept of stable real time web architectures, in which automated interaction between FE and BE components minimizes delays in software release cycles [8]. S. Meka in a study of complex financial technology platforms, demonstrates that automation of integration processes is the primary factor enabling rapid scaling of digital products [9].

Another research direction reflects a shift toward modular and reconfigurable development architectures focused on the parallel evolution of system components. For example, S. Tao and co authors

propose optimization of micro front end architecture based on cloud computing components, which enables independent updates of functional modules without interrupting overall system operation [10]. X. Zhang and co authors investigate reconfigurable BE architectures that support rapid redesign of digital platforms through flexible synchronization of computing units [11].

M. Dressler shows that innovative organizational development requires simultaneous modernization of client and server components, as asynchronous evolution increases the duration of the innovation cycle [12]. J. Liu and colleagues argue that the use of fiber optic communication channels in BE electronics ensures high speed data transmission and accelerates integration of complex digital experimental systems [13].

The final stage in the evolution of research is associated with the transition to infrastructure level automation of synchronization in distributed digital ecosystems and peripheral computing environments, including Edge Computing. G. Sung and co authors develop a gateway system with protocol conversion that enables automated interaction between devices and server platforms in real time [14]. J. Liu and co authors confirm that infrastructure synchronization through fiber optic networks increases the implementation speed of digital experimental complexes and reduces integration risks [15].

A synthesis of the analyzed studies indicates that automated synchronization of FE and BE interactions is evolving from a localized technical task into a strategic driver of digital development. It contributes to shortening the software product life cycle, improving coordination among development teams, and significantly reducing time to market for digital products.

Despite the active advancement of approaches to organizing interaction between client and server components of software systems, several issues remain unresolved. In particular, systematic synchronization within the software development life cycle and assessment of its impact on the speed of digital product creation require further investigation. Existing studies primarily emphasize individual technological solutions, whereas mechanisms for automated coordination of software interfaces, data contracts, and integration processes within microservices architectures, as well as methodological foundations for synchronized deployment of system components, remain insufficiently developed. Challenges related to service versioning, integration risks, accumulation of technical debt, and organizational inconsistencies among development teams continue to hinder effective reduction of digital product time to market.

The present study seeks to address these gaps through a comprehensive analysis of FE and BE interaction approaches, examination of technological mechanisms for automated API coordination and integration processes, substantiation of the application of CI/CD practices for synchronized deployment, and development of recommendations for an automated synchronization strategy in software development. The findings contribute to an expanded methodological understanding of synchronization of digital system components as a factor in enhancing release stability, optimizing team collaboration, and accelerating the market introduction of digital products.

The **purpose** of this article is to substantiate the role of automated synchronization of interaction between FE and BE components in reducing time to market for digital products and enhancing the efficiency of software development processes.

To achieve this objective, the following tasks are defined:

1) to analyze approaches to organizing interaction between the FE and BE components of software systems and to examine mechanisms for automated coordination of software interfaces and integration processes within microservices architectures;

2) to substantiate the methodological foundations for the application of CI/CD practices and to identify key scientific and practical challenges associated with implementing automated development synchronization;

3) to develop recommendations for establishing a strategy of automated synchronization of digital system components aimed at reducing time to market and improving release stability.

Results. The effectiveness of the software development life cycle is largely determined by the level of organization of FE and BE interaction, as integration of client and server logic defines the readiness of a digital product for release. The transition

from sequential development models to integrated approaches is driven by the need for parallel development of functionality, reduction of integration delays, and assurance of change stability in dynamic digital environments. In contemporary IT projects, development speed increasingly depends not on individual technologies but on the selected model of interaction between FE and BE components (Table 1).

Modern development practice demonstrates that the most significant time losses occur during integration stages, when the client side must adapt to changes in server logic after completion of individual tasks. In a sequential model, this issue manifests as delayed testing of component interactions, leading to error accumulation and release delays [10, p. 535].

The transition to parallel development enables teams to work simultaneously by using simulation services to test the user interface before server implementation is finalized. However, in the absence of a formalized data structure specification, this approach often results in functional rework. Contract based interaction has transformed the logic of work organization, as the data exchange structure is defined at early design stages and serves as a unified technical foundation for both FE and BE components. In practice, this approach is implemented in digital e-commerce platforms, banking services, and SaaS systems, where a single interface specification supports simultaneous development of client applications, server services, and automated compatibility testing [6, p. 97].

Any modification of the service response structure is detected by the control system prior to the integration stage, which significantly reduces the number of critical errors. Further advancement is reflected in interface oriented architectures, in which server services operate as autonomous modules accessible to multiple clients, including web applications, mobile platforms, and external information systems.

Table 1

Modern approaches to organizing front-end and back-end interaction in the development of digital products

Interaction approach	Basic principle of organization	Features of FE–BE interaction	Impact on product creation speed
Sequential model	Development of the server part before the client part	FE depends on the readiness of services	Extension of the release cycle
Parallel development	Simultaneous creation of functionality	Use of simulation services	Reduction of development time
Contractual interaction	Preliminary agreement on data exchange structure	Uniform interaction specifications	Reduction of integration errors
Interface-oriented architecture	Building a system around API	Independent development of FE and BE	Accelerated scaling
Automated synchronization	Automatic reconciliation of data models and services	Code generation and compatibility testing	Minimization of time-to-market

Source: compiled by the author based on [1, p. 219; 2, p. 424; 5, p. 894; 6, p. 97; 8, p. 234; 9, p. 8588; 10, p. 535]

Under such conditions, the speed of developing new features is determined by the stability of interaction interfaces rather than by the readiness of the system as a whole. The most effective approach in modern digital products is automated synchronization of FE and BE components, in which interface specifications are used to generate data models, client requests, compatibility checks, and integration tests automatically. In product oriented IT companies, this practice shifts consistency verification from manual testing to the development phase, thereby reducing technical debt, stabilizing release cycles, and ensuring a measurable reduction in time to market for digital products.

Within microservices architectures, integration consistency is maintained not through centralized control but through formalized interaction interfaces and automated compatibility management. The independence of life cycles of individual services requires technological mechanisms that enable synchronization of APIs, data contracts, and integration processes without manual coordination of changes among development teams. Automated coordination is achieved through machine readable interface specifications, backward compatibility verification systems, contract testing, and integration of these procedures into continuous build and deployment pipelines. This approach transfers integration control from post hoc testing to the software coding stage (Fig. 1).

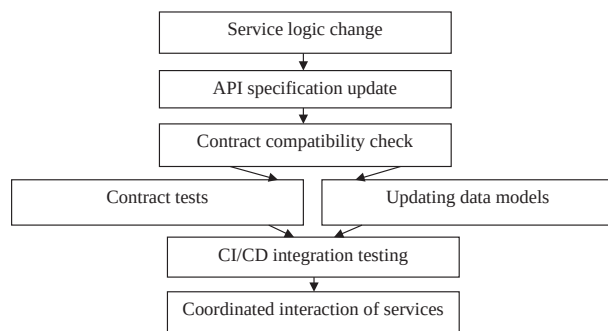


Fig. 1. Diagram of automated coordination of APIs, data contracts, and service integration in microservice architecture

Source: created by the author

The operation of this mechanism is based on transferring responsibility for integration stability from human coordination to development infrastructure. Any modification of a service interface is recorded in the specification, after which its impact on dependent services and client components is automatically verified. If the change violates compatibility, the build system blocks integration or initiates interface ver-

sioning, thereby preventing propagation of incompatible modifications in a distributed environment.

In large digital platforms, where catalog, payment, authorization, and analytics services are managed by independent teams, automated contract testing enables parallel update cycles without requiring synchronous release of the entire system [14, p. 93200]. Data schemas serve as the basis for automatic generation of types and queries, eliminating model duplication across services and reducing integration defects. Integration checks are executed within the CI/CD pipeline before changes are deployed to the production environment, ensuring stable interaction continuously rather than only at the final testing stage. This approach supports a high release frequency in microservices systems without increasing operational risks or requiring centralized coordination of changes among all development participants.

Synchronized deployment of client and server components of digital systems under current conditions is associated with the transition from episodic releases to continuous integration of changes. In distributed software environments, asynchronous updates of FE and BE components lead to interface incompatibility, functional regressions, and prolonged post release stabilization. Continuous integration and continuous delivery or deployment practices constitute the methodological foundation for controlled change coordination, as they integrate compatibility verification, automated testing, and deployment management directly into the development life cycle (Table 2).

The systematic application of CI/CD transforms the logic of software updates from periodic component coordination to continuous compatibility monitoring. Synchronization of client and server components is achieved by embedding interaction checks directly into the automated development pipeline, where each code change is compiled, tested, and verified for interface compliance. This approach eliminates the common problem of asynchronous releases, in which updated server logic becomes available to users while the client interface continues to operate with a previous data structure.

In digital banking services and e commerce platforms, this model is implemented through a unified deployment pipeline, within which modifications to interaction interfaces are automatically validated before promotion across development, testing, and production environments. For example, changes to payment processing mechanisms or order structures undergo integration testing together with client side scenarios, enabling detection of inconsistencies prior to

Table 2

Methodological approaches to applying CI/CD practices for synchronized deployment of client and server parts of digital systems

Methodological approach	Content of implementation in CI/CD	FE and BE synchronization mechanism	Expected result
Single build pipeline	Joint automated component assembly	Simultaneous verification of FE and BE changes	Coordinated release versions
API compatibility control	Automatic interface checks	Detection of incompatible changes prior to deployment	Prevention of integration failures
Automated testing	Unit, integration, and contract tests	Component interaction verification	Reducing post-release regressions
Phased deployment	Dev–Test–Stage–Production environments	Gradual synchronization of updates	Stability of implementation
Controlled feature enablement	Feature flags and availability control	Independent enablement of FE and BE features	Flexible release management
Automatic monitoring	Post-deployment checks	Monitoring system behavior after release	Rapid detection of deviations

Source: compiled by the author based on [1, p. 223; 2, p. 426; 6, p. 103; 8, p. 237; 9, p. 8590; 14, p. 93201]

public release [8, p. 237]. Phased deployment supports controlled implementation of changes, as a new server function may be deployed but remain inactive for users until correct client side operation is confirmed.

An essential element of synchronization is automated post deployment monitoring, which enables evaluation of actual interaction among system components based on performance metrics. In large scale digital products, this approach sustains frequent releases without increasing failure rates, as compatibility verification is conducted continuously rather than solely at final testing stages. Consequently, CI/CD practices perform not only an instrumental role but also a methodological function, ensuring coordinated evolution of client and server components and reducing the time interval between functionality development and its practical utilization.

In order to substantiate the impact of automated continuous integration and deployment processes on the synchronization of client-server updates in digital systems, an analytical modeling mini-experiment was conducted. The object of the study was a typical digital product with a split client-server architecture developed within a single 14-day iterative cycle.

Two deployment modes were compared. The first mode reflects the traditional approach, in which changes are integrated after the completion of individual component development, and client and server updates may be implemented asynchronously. The second mode simulates an automated deployment pipeline, in which change consistency is ensured through automatic builds, compatibility testing, integration testing, and the phased introduction of updates into the production environment.

Effectiveness was assessed using the following deployment process indicators: frequency of suc-

cessful deployments, percentage of failed deployments, lead time for transitioning to the production environment, and recovery time after deployment failures. Since these indicators employ different units of measurement, they were converted to a normalized performance index with values ranging from 0 to 1. Normalization enables accurate comparison of deployment process characteristics within a unified evaluation scale (Fig. 2).

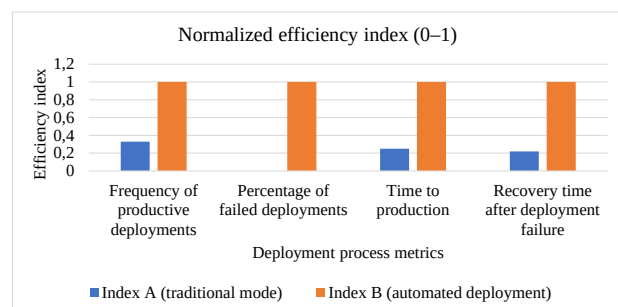


Fig. 2. Comparison of deployment process characteristics in digital systems without synchronized checks and with CI/CD practices applied. Source: author's own development

Source: created by the author

As can be seen in Fig. 2, the use of an automated deployment pipeline ensures coordinated improvement across all studied process characteristics. The increase in the productive deployment index indicates the feasibility of implementing changes in small iterations without accumulating integration risks between the client and server components of the system.

The reduction in the proportion of failed deployments confirms the effectiveness of early error detection during automated compatibility checks, which shifts quality control from the operational stage to the integration stage of changes. The reduction in lead

time for changes to reach the production environment reflects the acceleration of functional updates, while the decrease in recovery time after failures characterizes the improved manageability and stability of the release process.

The results obtained indicate that automated deployment constitutes the methodological foundation for synchronizing the development of client and server components of a digital product and establishes the conditions necessary for reducing time to market for digital solutions.

Despite the demonstrated effectiveness of automated synchronization of client and server components in digital systems, the implementation of corresponding approaches is accompanied by a number of scientific and practical challenges. First, complexity arises at the level of organizational transformation of the development process, as automated deployment requires a shift from the traditional model of team interaction, a transition from sequential to integrated workflows, and the establishment of shared responsibility for the release outcome [9, p. 8590]. An additional challenge involves ensuring the consistency of software interfaces in microservice architecture environments, where an increasing number of independent components elevates the risk of change desynchronization. Practical difficulties are also associated with the need to maintain a unified testing environment, the complexity of automating data compatibility checks, the accumulation of technical debt, and growing qualification requirements for developers and support engineers. A separate issue concerns measuring the economic effect of automated deployment, given that reductions in time to market manifest indirectly through release stability, responsiveness to changes, and reduced operational risks, which complicates the direct quantitative assessment of outcomes.

The strategy for improving the efficiency of synchronized deployment of digital products should be based on the phased integration of automated processes into the development life cycle. The primary focus is on establishing a unified pipeline for building, testing, and deploying changes, in which client and server logic are continuously coordinated throughout the development process rather than at its final stage. It is essential to combine technical automation solutions with management mechanisms for team coordination, standardization of interaction interfaces, and implementation of continuous release

quality monitoring. The adoption of such a strategy reduces development inertia, increases the predictability of the release process, and creates the organizational prerequisites for a consistent reduction in time to market for digital products. These outcomes align with the results of the analytical modeling experiment and confirm the systemic nature of the impact of automated front-end/back-end interaction synchronization on the efficiency of digital development.

Conclusions. The study demonstrates that the decisive factor in reducing time to market for digital products is the level of synchronization between the client and server components of software systems throughout the entire development life cycle. The findings show that transitioning from a traditional sequential integration model to automated change coordination shifts compatibility control to the early stages of development. This shift reduces integration delays, enhances release stability, and shortens the time required to implement functional updates. Simulation results confirm that automated deployment increases the frequency of productive releases, decreases the proportion of failed deployments, reduces the time required for changes to reach the production environment, and shortens recovery time after failures. These outcomes indicate the systemic impact of front end and back end synchronization on the efficiency of digital product development.

At the same time, the study identifies key challenges associated with implementing synchronized approaches, including the need for organizational restructuring of development processes, maintaining interface consistency in distributed architectures, heightened requirements for test automation, and the difficulty of quantitatively assessing the economic effect of reduced time to market. The results substantiate that improving the effectiveness of digital projects requires an integrated strategy of automated synchronization that combines standardized component interaction, continuous quality control of changes, and coordinated collaboration among development teams.

Prospects for further research include the development of models for quantitatively measuring the impact of synchronized development on the economic performance of digital products, the application of intelligent methods to predict integration risks, and the creation of adaptive deployment systems capable of automatically optimizing interactions among software system components during operation.

Bibliography:

1. De Moraes F. R., de Almeida D., Affonso F. J. A Real-time data synchronization approach for high-availability micro applications. In: Simpósio Brasileiro de Engenharia de Software (SBES). 2025. P. 215–225. DOI: <https://doi.org/10.5753/sbes.2025.9904>
2. Rasyaad F., Wibowo S. A. Developing a Scalable Next.js Platform for IoT Monitoring Using Backend-as-a-Service with a Comparison of Supabase and Firebase. In: 2025 International Conference on Computer Sciences, Engineering, and Technology Innovation (ICoCSETI) Jakarta, Indonesia (Jakarta, Indonesia, 21 January 2025). Jakarta, Indonesia, 2025. P. 421–426. DOI: <https://doi.org/10.1109/ICoCSETI63724.2025.11018944>
3. Bortolato G., Bergnoli A., Bortolato D., Mengoni D., Migliorini M., Montecassiano F., Zanetti M. Front-end RDMA over Converged Ethernet, real-time firmware simulation. Journal of Instrumentation. 2024. Vol. 19, № 03. Article C03038. DOI: <https://doi.org/10.1088/1748-0221/19/03/C03038>
4. Guan S., Hu W., Zhou H. Implementation of online digital twin framework for thermal power plant. In: International Conference on Energy Engineering and Environmental Engineering (Singapore, 6–8 August 2023). Singapore, 2023. P. 187–198. DOI: https://doi.org/10.1007/978-3-031-48204-5_16
5. Maiti M., Vuković D., Mukherjee A., Paikarao P. D., Yadav J. K. Advanced data integration in banking, financial, and insurance software in the age of COVID-19. Software: Practice and Experience. 2022. Vol. 52, № 4. P. 887–903. DOI: <https://doi.org/10.1002/spe.3018>
6. Aquino A. C. B. D., Feliciano Lino A. Aligning back-end systems with front-end services: configurations of financial management information systems and e-services provision. Journal of Public Budgeting, Accounting & Financial Management. 2026. Vol. 38, № 1. P. 87–110. DOI: <https://doi.org/10.1108/JPBAFM-12-2024-0265>
7. Chen Z., Feng C., Chen H., Fan R., Yi H., Wang J., Liu S. Design of Waveform Digitizing Front-end Electronics for a Prototype Multi-purpose TPC at CSNS Back-n. In: 2021 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC) (16–23 Oct. 2021). 2021. P. 1–5. DOI: <https://doi.org/10.1109/NSS/MIC44867.2021.9875485>
8. Kumar M. Designing Resilient Front End Architectures for Real-Time Web Application. International Journal of Engineering Technology Research & Management (IJETRM). 2024. Vol. 8, № 08. P. 229–240. URL: <https://ijetrm.com/issues/files/May-2024-22-1747887028-AUG202430.pdf> (дата звернення: 27.02.2026).
9. Meka S. Building Digital Banking Foundations: Delivering End-to-End FinTech Solutions with Enterprise-Grade Reliability. International Journal of Research and Applied Innovations. 2023. Vol. 6, № 2. P. 8582–8592. DOI: <https://doi.org/10.15662/IJRAI.2023.0602003>
10. Tao S., Cui Z., Li Y., Yang Q. Optimization Scheme of Micro-front-End Architecture Based on Dynamic Cloud Components and Web Components. In: 2024 10th International Conference on Computer and Communications (ICCC) (Chengdu, China, 13-16 December 2024). Chengdu, China, 2024. P. 533–537. DOI: <https://doi.org/10.1109/ICCC62609.2024.10942252>
11. Zhang X. X., Duan R., Yu X. Y., Li D., Tang N. Y., Ching T. C. The design of China reconfigurable analog-digital backend for fast. Research in Astronomy and Astrophysics. 2020. Vol. 20, № 5. Article 073. DOI: <https://doi.org/10.1088/1674-4527/20/5/73>
12. Dressler M. Innovation management in wine business – Need to address front-end, back-end, or both? Wine Business Journal. 2022. Vol. 5, № 1. P. 27–43. URL: <https://wbj.scholasticahq.com/article/31770-innovation-management-in-wine-business-need-to-address-front-end-back-end-or-both> (дата звернення: 27.02.2026).
13. Liu J., Wang Y., Feng C., Liu S., Chen Q. A back-end electronics based on optical fiber communication for small to medium-scale physics experiments. Journal of Instrumentation. 2025. Vol. 20, № 05. Article T05002. DOI: <https://doi.org/10.1088/1748-0221/20/05/T05002>
14. Sung G. M., Tung L. F., Huang C. J., Yu C. P. Front-end gateway system with serial communication protocol conversion and edge computing platforms. IEEE Access. 2023. Vol. 11. P. 93193–93203. DOI: <https://doi.org/10.1109/ACCESS.2023.3307631>
15. Liu J., Wang Y., Feng C., Liu S., Chen Q. A back-end electronics based on fiber communication for small to medium-scale physics experiments. 2024. arXiv preprint. arXiv:2407.06786. DOI: <https://doi.org/10.48550/arXiv.2407.06786>

Піотровський Є.Є. РОЛЬ АВТОМАТИЗОВАНОЇ СИНХРОНІЗАЦІЇ ФРОНТЕНД-БЕКЕНД ВЗАЄМОДІЇ В СКОРОЧЕННІ ТЕРМІНІВ ВИВЕДЕННЯ ЦИФРОВИХ ПРОДУКТІВ НА РИНОК

Актуальність дослідження зумовлена зростанням складності цифрових продуктів, поширенням розподілених архітектур програмного забезпечення та необхідністю прискорення процесів розроблення в умовах високої динаміки цифрових ринків. Мета статті полягає в обґрунтуванні ролі автоматизованої синхронізації взаємодії клієнтської та серверної частин цифрових систем у

скороченні термінів виведення цифрових продуктів на ринок і підвищенні результативності процесів розроблення програмного забезпечення. У роботі застосовано методи аналізу й узагальнення наукових джерел для систематизації сучасних підходів організації взаємодії компонентів програмних систем, системно-структурний аналіз для визначення взаємозв'язків між етапами життєвого циклу розроблення, порівняльний підхід для зіставлення традиційних і автоматизованих моделей розгортання, а також аналітичне моделювання та нормування процесних метрик для оцінювання ефективності синхронізованого розгортання. У результаті дослідження проаналізовано сучасні підходи до організації взаємодії клієнтської та серверної частин цифрових продуктів у межах життєвого циклу розроблення програмного забезпечення. Встановлено, що автоматизована синхронізація процесів інтеграції та розгортання забезпечує перенесення контролю узгодженості змін на ранні етапи розроблення, що сприяє зменшенню інтеграційних затримок і підвищенню стабільності релізів. Доведено, що впровадження автоматизованого конвеєра розгортання супроводжується зростанням частоти продуктивних розгортань, зниженням частки невдалих оновлень, скороченням часу проходження зміни до продуктивного середовища та зменшенням часу відновлення після порушень у роботі. Виявлено системний вплив синхронізації фронтенд-бекенд взаємодії на скорочення часу виведення цифрових продуктів на ринок. У висновках обґрунтовано, що автоматизована синхронізація взаємодії між компонентами програмних систем є методичною основою підвищення ефективності цифрової розробки, забезпечуючи узгоджений розвиток клієнтської та серверної логіки, зниження операційних ризиків і підвищення передбачуваності релізного процесу. Перспективи подальших досліджень пов'язані з розробленням моделей кількісного оцінювання економічного ефекту скорочення часу виведення цифрових продуктів на ринок, застосуванням інтелектуальних методів прогнозування інтеграційних ризиків та створенням адаптивних систем автоматизованого розгортання, здатних оптимізувати взаємодію компонентів цифрових систем у реальному часі.

Ключові слова: життєвий цикл програмного забезпечення, безперервна інтеграція, безперервне розгортання, мікросервісна архітектура, узгодженість програмних інтерфейсів, інженерія програмних систем, автоматизація розроблення, релізний менеджмент, цифрові платформи.

Дата першого надходження статті до видання: 13.03.2026

Дата прийняття статті до друку після рецензування: 07.04.2026

Дата публікації (оприлюднення) статті 11.05.2026

УДК 004.056.5

DOI DOI DOI DOI <https://doi.org/10.32782/2663-5941/2026.2.1/36>

Прокопович-Ткаченко Д.І.

<https://orcid.org/0000-0002-6590-3898>

Університет митної справи та фінансів;

Державна наукова установа «Інститут інформації, безпеки і права

Національної академії правових наук України»;

Державний університет інформаційно-комунікаційних технологій

Торстенссон О.

<https://orcid.org/0009-0007-2169-6851>

Хальмстадський університет, м. Хальмстад, Швеція

Черкаський Д.О.

<https://orcid.org/0009-0003-8516-6252>

Національний технічний університет «Дніпровська політехніка»

Переметчик Д.О.

<https://orcid.org/0009-0006-1978-5858>

Університет митної справи та фінансів

Берт С.Д.

<https://orcid.org/0009-0006-7423-9235>

Університет митної справи та фінансів

СОЦІОТЕХНІЧНІ ЧИННИКИ ВИНИКНЕННЯ ВРАЗЛИВОСТЕЙ МЕРЕЖЕВИХ ПРОТОКОЛІВ ЗАХИЩЕНОЇ ПЕРЕДАЧІ ДАНИХ

Дослідження присвячено аналізу соціотехнічних чинників, що зумовлюють появу вразливостей у протоколах захищеної передачі даних під час експлуатації мереж електронних комунікацій. Розглянуто типові компромісні дії персоналу, пов'язані з відключенням або послабленням шифрованих з'єднань для швидкого відновлення сервісу, прийняттям недовірених сертифікатів, ігноруванням помилок перевірки ланцюжка довіри та підтримкою застарілих режимів сумісності, що створює умови для зниження рівня криптографічного захисту та перехоплення сеансів. Окремо досліджено атаки на вбудоване програмне забезпечення мережесовмісних пристроїв, у межах яких уразливі вебінтерфейси та небезпечні налаштування протоколів адміністрування використовуються для зміни конфігурацій і деградації політик шифрування. Запропоновано методику оцінювання ризику на основі кореляції подій, сигналів виявлення вторгнень і моделей машинного навчання для аналізу мережевого трафіку та артефактів прошивок, включно з перетворенням байтових послідовностей у візуальні представлення. Експеримент на стенді гібридних атак показав, що автоматизація політик шифрування, заборона ручних винятків і перехід до захищених механізмів керування суттєво зменшують імовірність успішного перехоплення сеансу та підвищують відтворюваність конфігурацій. Зроблено висновок, що стійкість транспортного рівня визначається не лише формальним дотриманням стандартів, а передусім культурою прийняття рішень, контрольованістю змін і архітектурою постійної перевірки довіри.

Ключові слова: транспорт, шифрування, сертифікація, довіра, сумісність, перехоплення, прошивки, адміністрування, кореляція, нейромережі, автоматизація.

Постановка проблеми. Метою дослідження є формалізація внеску людських компромісів у вразливість транспортного рівня та розроблення практичної методики зменшення ризиків через

автоматизацію криптографічних політик і гібридну аналітику IDS та SIEM. Для досягнення цієї мети передбачено побудову формальних математичних моделей ризику, що описують імовірність успіху

атак типу downgrade та перехоплення сесій при збігу людського компромісу та технічної експозиції. Важливим завданням є розроблення моделей поведінки транспортних протоколів, зокрема TLS 1.3, з урахуванням організаційного тиску на персонал та рішень щодо сумісності. Потребує оцінки операційна придатність засобів захисту за критеріями затримки детекції та точності в умовах гібридних атак. Додатково передбачено побудову оптимізаційних моделей вибору заходів автоматизації на основі критеріїв «стійкість – вартість». Також актуально інтегрувати отримані результати з моделями систем виявлення вторгнень, використовуючи глибинні нейронні мережі (CNN+LSTM, AE+LSTM) та підхід Byte2Image для структурного аналізу трафіку.

Запропонована оптимізаційна постановка вибору параметрів має базуватися на критеріях мінімізації інтегрального ризику і враховувати політики нульової довіри. Також актуально інтегрувати отримані результати з моделями систем виявлення вторгнень та управління інформаційною безпекою, використовуючи гібридну аналітику для відокремлення ручного компромісу від зловмисної деградації.

Таким чином, завданням роботи є не лише ідентифікація вразливостей транспортного рівня, але й формування практичних рекомендацій щодо впровадження принципів Zero Trust та автоматизації політик (policy-as-code) у мережах електронних комунікацій для мінімізації простору ручних небезпечних винятків.

Аналіз останніх досліджень і публікацій. Транспортний рівень у моделі OSI виступає критичним вузлом довіри, що на сучасному етапі реалізується переважно через протоколи TLS 1.2 та TLS 1.3. Дослідження вразливостей складних протокольних станів TLS показали, що будь-яка неточність обробки помилок може бути перетворена на канал атаки. Історичні кейси Logjam та FREAK довели критичність підтримки застарілих параметрів сумісності для деградації криптографічної стійкості, а уразливості POODLE та DROWN продемонстрували ризики експлуатації легасі-режимів SSL 3.0.

Сучасні стандарти, зокрема RFC 7525 та RFC 9325, регламентують безпечне використання TLS, а нормативний тренд у RFC 8996 вимагає повної відмови від TLS 1.0 та 1.1. Рекомендації NIST SP 800-52 Rev. 2 визначають правила конфігурації TLS-реалізацій, а профілі NISTIR 8259A та настанови SP 800-193 фокусуються на резильєнтності прошивок та базових кіберспроможностях при-

строїв. Окремий напрям досліджень присвячений інтеграції криптографічних параметрів із системами IDS (Snort, Zeek) та SIEM. Праці в галузі інтелектуального аналізу трафіку показують, що використання методів глибинного навчання (CNN, LSTM) та візуалізації Byte2Image дозволяє значно підвищити точність виявлення аномалій у мережних потоках та артефактах прошивок. Це відкриває перспективу для впровадження концепції Zero Trust Architecture (NIST SP 800-207), де довіра не припускається і перевіряється безперервно.

Таким чином, сучасні дослідження підкреслюють необхідність комплексного підходу, який одночасно враховує соціотехнічну природу вразливостей, практичні характеристики керувального контуру (наприклад, SNMPv3) та інтеграцію із засобами гібридної аналітики. Це й визначає наукову новизну та актуальність даної роботи.

Постановка завдання. Метою дослідження є формалізація внеску людських компромісів у вразливості транспортного рівня та розроблення практичної методики зменшення ризиків через автоматизацію політик і гібридну аналітику. Для досягнення цієї мети передбачено:

- ідентифікувати типові компроміси персоналу, що активують вразливості TLS, SSL та SNMP у firmware-атаках;
- побудувати ймовірнісну модель ризику для сценаріїв downgrade, перехоплення сесій і компрометації автентифікації;
- сформулювати оптимізаційну постановку вибору заходів автоматизації криптографічних політик;
- оцінити детекційні моделі CNN+LSTM і AE+LSTM у порівнянні з правилами IDS та кореляцією SIEM;
- надати практичні рекомендації для впровадження підходів Zero Trust на транспортному та керувальному рівнях.

Таким чином, завданням роботи є не лише порівняння технічних методів детекції, але й формування цілісної архітектури захисту, яка мінімізує вплив людського фактора через автоматизацію (policy-as-code) та перехід керувального контуру на захищені режими.

Методологія дослідження ґрунтується на системному підході до моделювання мереж електронних комунікацій як багаторівневої динамічної системи, структура якої відповідає еталонній моделі взаємодії відкритих систем OSI [1]. В умовах гібридних кібератак особлива увага приділяється транспортному та прикладному рівням, де реалізуються протоколи захищеної передачі

даних, зокрема Transport Layer Security (TLS) версії 1.3 [2], а також рекомендаціям щодо його конфігурації та безпечного використання [3].

У межах дослідження прийнято, що ризик порушення цілісності та конфіденційності даних формується як функція технічних параметрів протоколів, стану прошивки мережевого обладнання та людських управлінських рішень. З урахуванням емпіричних результатів аналізу вразливостей TLS [4], [5], [6] сформовано гіпотезу про кореляцію між дозволом застарілих криптографічних примітивів, downgrade-механізмами та реалізацією атак типу Logjam і FREAK. Додатково враховано ризики експлуатації легасі SSL та некоректних наборів шифрів у вебінтерфейсах firmware, що історично пов'язано з атаками класу POODLE та DROWN.

Методичний апарат дослідження поєднує формальне математичне моделювання ризику, імовірнісні оцінки сценаріїв атак та застосування алгоритмів глибокого навчання для аналізу мережевого трафіку. Для виявлення складних багатовступневих атак використано комбіновані архітектури згорткових нейронних мереж та рекурентних мереж з довготривалою короткочасною пам'яттю (CNN+LSTM) із представленням трафіку у форматі Byte2Image, а також моделі автоенкодерів у поєднанні з LSTM (AE+LSTM) для детекції аномалій. Такий підхід відповідає сучасним тенденціям застосування штучного інтелекту в системах виявлення вторгнень (IDS) та платформах кореляції подій безпеки (SIEM). Окремий напрям методів стосується аналізу уразливостей керувального контуру, зокрема використання протоколу Simple Network Management Protocol (SNMP). Враховано ризики застосування SNMPv2c із дефолтними community-рядками та їх вплив на деградацію криптографічних політик і маніпуляції прошивкою. Для мінімізації таких загроз застосовано концепцію Zero Trust Architecture відповідно до сучасних рекомендацій [7], що передбачає жорстку автентифікацію, мікросегментацію та автоматизоване управління політиками доступу. Таким чином, методи дослідження інтегрують стандартизовані вимоги до захищених комунікацій [1]–[3], емпіричні результати аналізу криптографічних атак [4]–[6], а також підходи до побудови архітектур нульової довіри [7], формуючи комплексну модель оцінювання стійкості мереж електронних комунікацій у умовах гібридних кібератак.

Виклад основного матеріалу

Об'єкт дослідження та дизайн експерименту

Об'єктом дослідження є транспортний рівень мережі електронних комунікацій у режимі сервіс-

ної експлуатації, де TLS-термінація, маршрутизація запитів та керування мережевими елементами здійснюються у розподіленому середовищі. Для відтворюваності побудовано стенд із трьома логічними сегментами: – сервісний сегмент з TLS-термінацією на проксі та мікросервісами; – користувацький сегмент клієнтів, що ініціюють сесії та провокують сумісні сценарії; – керувальний сегмент з вебкеруванням і SNMP, типовими для пристроїв із вбудованим ПЗ.

Сценарій гібридної атаки включає дві взаємопов'язані фази. У фазі тиску на персонал моделюється інцидент доступності та сумісності: зростання помилок сертифікації, проблеми з ланцюжком довіри, часткові відмови сервісу та вимоги швидкого відновлення. Оператору надано можливість застосувати типові компроміси: вимкнути перевірку сертифіката на клієнті або проксі, знизити мінімальну версію TLS, дозволити ширші набори шифрів, активувати легасі-режим у вебінтерфейсі прошивки, залишити SNMP у спрощеному режимі. У фазі технічної експлуатації зловмисник використовує збережені винятки: реалізує downgrade під час узгодження параметрів, виконує атаку посередника на транспортному рівні, а також отримує доступ до керувального контуру через дефолтні SNMP параметри або слабкий захист вебінтерфейсу прошивки та змінює конфігурації, що впливають на TLS-політики.

Стійкість протоколів захищеної передачі даних у сучасних мережах електронних комунікацій дедалі більше визначається не лише криптографічною надійністю алгоритмів, а й соціотехнічними умовами їх експлуатації. Транспортний рівень моделі OSI виступає критичним вузлом довіри, де реалізуються механізми автентифікації, узгодження параметрів шифрування та керування сесіями. Проте фактичний рівень захисту формується на стику стандартів, конфігураційних практик і рішень персоналу, який у ситуаціях інцидентного тиску може допускати винятки, що знижують криптографічну стійкість. Актуальність проблематики посилюється гібридним характером сучасних атак, у яких технічна експлуатація поєднується з організаційним впливом.

На рисунку представлено узагальнену структурну схему соціотехнічних чинників, що впливають на виникнення вразливостей транспортного рівня. У верхній частині схеми виділено два полюси ризику: людський фактор (компроміси персоналу під тиском інцидентів) та технічний фактор (експозиція легасі-протоколів і небез-



Рис. 1. Соціотехнічна модель формування вразливостей протоколів захищеної передачі даних та гібридної аналітики їх виявлення

печних режимів). У центрі відображено ключову суперечність між формальним стандартом безпеки та реальним рівнем захисту, який формується в експлуатації. Нижній рівень ілюструє типові приклади атак і вразливостей (Logjam, FREAK, POODLE, DROWN, firmware-атаки, експлуатація SNMP), що активуються за наявності одночасно людського компромісу й технічної експозиції. Далі наведено блок гібридної аналітики виявлення, який поєднує сигнатурні механізми IDS, кореляцію подій у SIEM та моделі глибинного навчання (CNN+LSTM, AE+LSTM, Byte2Image). Завершальний елемент схеми демонструє інтегральний результат – перехід до підходу Zero Trust і автоматизації політик, що забезпечує істотне зниження ризику (приблизно на 85%) за рахунок мінімізації ручних винятків і контролю керуального контуру.

Джерела даних та нормалізація подій

Збиралися три класи даних: (1) мережеві метадані та події IDS (сигнатури, статистичні ознаки потоків, індикатори небезпечних узгоджень); (2) журнали сервісної інфраструктури (TLS-термінація, помилки перевірки сертифікатів, події відновлення сесій, конфігураційні зміни); (3) події керуального контуру (SNMP запити/записи, traps, журнали вебкерування) та байтові артефакти прошивки до і після змін. Події нормалізувалися до єдиного часово-ідентифікаційного формату

та індексувалися у SIEM для кореляції. Якість журналювання оцінювалася за повнотою полів, часовою синхронізацією та трасованістю змін до суб'єктів доступу, що узгоджується з настановами з керування логами [22].

Ймовірнісна модель ризику людського компромісу

Нехай подія H означає, що персонал прийняв компромісне рішення на транспортному рівні або в керуальному контурі (наприклад, вимкнув перевірку сертифіката або залишив дефолтний SNMP доступ). Нехай V означає наявність технічної експозиції, тобто стану системи, який допускає експлуатацію (легасі SSL у прошивці, дозволені небезпечні набори шифрів, незахищений SNMPv2c). Нехай A означає успішну атаку на множині критичних потоків. Контекст x включає показники, що спостерігаються у SIEM: кількість інцидентних подій, середній час до відновлення сервісу, частоту ручних змін конфігурації, навантаження на персонал, рівень автоматизації політик. Ймовірність компромісу задається логістичною моделлю:

$$\mathbb{P}(H=1|x) = \sigma(\beta_0 + \beta^T x), \quad \sigma(z) = \frac{1}{1+e^{-z}}. \quad (1)$$

Для n критичних потоків, де експлуатація можлива лише за одночасної наявності компромісу та технічної експозиції, оцінка ймовірності успіху атаки:

$$\mathbb{P}(A=1)=1-(1-\mathbb{P}(H=1|x)\cdot\mathbb{P}(V=1))^n, \quad R=L\cdot\mathbb{P}(A=1). \quad (2)$$

У такій постановці криптографічні стандарти й профілі TLS зменшують $\mathbb{P}(V=1)$, а організаційні практики та автоматизація впливають на $\mathbb{P}(H=1|x)$, тобто на схильність до ручних винятків.

Оптимізаційна постановка автоматизації криптографічних політик

Розглянемо набір заходів $u \in \{0,1\}^m$, де $u_i=1$ означає впровадження заходу: примусова перевірка сертифікатів, заборона downgrade і легасі-версій, ротація сертифікатів, взаємна автентифікація, перехід на SNMPv3, вимога підписаних оновлень прошивки, централізовані профілі TLS у вигляді політик як код [3, 9, 15, 30]. Задача вибору заходів формулюється як мінімізація ризику з урахуванням вартості:

$$\min_u R(u) + \lambda \sum_{i=1}^m c_i u_i \quad \text{за умови} \quad \sum_{i=1}^m c_i u_i \leq B, \quad (3)$$

де c_i – витрати на впровадження заходу, B – бюджет, λ – коефіцієнт компромісу між ризиком та витратами. Оскільки $R(u)$ залежить від $\mathbb{P}(H=1|x,u)$ і $\mathbb{P}(V=1|u)$, автоматизація зменшує як людську схильність до винятків, так і технічну експозицію.

Гібридна аналітика IDS та SIEM і нейромережеві моделі

Дані формуються з мережевих подій IDS, журналів SIEM та байтових артефактів трафіку і прошивок. Для подання байтових послідовностей застосовується Byte2Image: фрагмент байтів перетворюється на матрицю інтенсивностей і розглядається як одноканальне зображення, що дозволяє використати CNN для виділення структурних ознак [26]. Часова складова моделюється через LSTM, що узгоджується з практикою рекурентних IDS [27]. Для задачі класифікації сценаріїв компромісів використовується зв'язка CNN+LSTM:

$$h_t = \text{LSTM}(h_{t-1}, \phi_{\text{CNN}}(I_t)), \quad \hat{y}_t = \text{softmax}(Wh_t + b), \quad (4)$$

де I_t – Byte2Image-вхід для вікна трафіку або артефакту прошивки, $\phi_{\text{CNN}}(\cdot)$ – екстрактор ознак CNN [23], $\text{LSTM}(\cdot)$ – рекурентний блок [24]. Навчальна розмітка формувалася за журналами змін конфігурацій та сценарієм стенду, що дозволило відокремити ручний компроміс від зловмисної деградації.

Для аномального детектування використовується автоенкодер (АЕ) для реконструкції агрегованих векторів подій SIEM/IDS і LSTM для урахування послідовності похибок:

$$e_t = x_t - g(f(x_t))^2, \quad s_t = \sigma(\alpha_0 + \alpha_e e_t + \alpha^T \text{LSTM}(e_{t-k:t})), \quad (5)$$

де x_t – вектор ознак подій (помилки сертифікації,

зміни профілю TLS, активність SNMP), $f(\cdot), g(\cdot)$ – кодувальник і декодувальник автоенкодера [28], s_t – оцінка аномальності. Концептуально цей підхід відповідає загальній парадигмі виявлення відхилень у часових рядах [25].

Метрики

Якість детекції оцінюється точністю, повнотою та F_1 -мірою. Для ризикової частини оцінюється відносне зниження ризику $R = (R_{\text{manual}} - R_{\text{auto}}) / R_{\text{manual}}$ при переході від ручних винятків до автоматизованих політик. Для операційної придатності аналізується затримка детекції та частка помилкових тривог, а також вплив якості журналювання [22].

Запропонована методологія забезпечує відтворюване моделювання мереж електронних комунікацій в умовах гібридних кібератак через поєднання трьох компонентів: (1) стендового дизайну експерименту з чітким розділенням сервісного, користувацького та керувального сегментів; (2) уніфікованого збору й нормалізації подій IDS/SIEM та артефактів керувального контуру (включно з SNMP і firmware-подіями); (3) формальних моделей ризику, що розділяють людський компроміс і технічну експозицію. Така побудова дозволяє не «вгадувати винного», а вимірювати, де саме система стає вразливою: у політиках TLS/SSL, у керувальному контурі, чи в рішеннях персоналу під тиском інцидентності.

Імовірнісна модель ризику та оптимізаційна постановка вибору заходів перетворюють безпеку з набору рекомендацій на керовану задачу мінімізації інтегрального ризику з урахуванням вартості та обмежень впровадження. Гібридна аналітика (IDS+SIEM) підсилена нейромережевими моделями CNN+LSTM (із Byte2Image) та AE+LSTM, що дає змогу одночасно покривати: (а) структурні ознаки байтових послідовностей/узгоджень і (б) часові аномалії в потоках подій та конфігурацій. Визначені метрики (точність, повнота, F_1 , затримка детекції, відносне зниження ризику) забезпечують порівнюваність підходів і практичну інтерпретацію результатів для експлуатаційних рішень. Окремо методи підкреслюють ключовий практичний принцип: найбільший ефект досягається не «героїзмом чергового інженера», а зменшенням простору ручних винятків через автоматизацію криптографічних політик (policy-as-code) та переведення керувального контуру на безпечні режими (SNMPv3), що узгоджується з логікою Zero Trust. У підсумку, розділ «Методи» формує цілісну основу для розділу «Результати»: від відтворення сценаріїв компромісів до кількісного вимірювання їх впливу та оцінювання ефективності детекції й протидії.

У практиці експлуатації мереж захищена передача даних рідко «ламається» через сам стандарт шифрування – частіше вона деградує через сукупність дрібних рішень, які здаються виправданими в моменті. Коли сервіс падає, клієнти скаржаться, а час відновлення стискається до хвилин, персонал схильний робити тимчасові послаблення: дозволяти сумісність із застарілими режимами, приймати недовірені сертифікати або спрощувати керування пристроями. У підсумку виникає розрив між тим, що «має бути безпечно за стандартом», і тим, що реально працює в продуктивному середовищі.

Щоб керувати цим розривом, потрібна не тільки технічна конфігурація, а й формалізація ризику: які саме фактори штовхають систему до компромісу та як швидко цей компроміс перетворюється на вікно атаки. Важливо розуміти, що ризик накопичується: чим більше критичних потоків, сервісів і точок керування, тим вища ймовірність, що хоча б один елемент стане «слабкою ланкою». У такій картині безпеки головна небезпека – не одинична помилка, а повторюваний шаблон ручних винятків.

Саме тому акцент зміщується з «ловити атаки швидше» на «зменшити кількість ситуацій, де оператор може ненавмисно легалізувати вразливість». Автоматизація криптографічних політик

і перехід до керованих, захищених механізмів адміністрування (у логіці Zero Trust) зменшує ризик системно: не за рахунок героїзму чергового інженера, а за рахунок обмеження простору для небезпечних компромісів.

На рисунку показано три взаємопов'язані частини, які пояснюють, як «людський компроміс» перетворюється на вимірюваний ризик і чому автоматизація різко змінює картину. Ліворуч наведено типову криву залежності: за низького навантаження і стабільної роботи система тримається в безпечному режимі, але після певного порогу (коли інцидентів/помилко стає забагато) ймовірність компромісних дій персоналу зростає дуже швидко. Це наочно демонструє нелінійність: невелике погіршення умов може різко підвищити ризик.

У центральному графіку показано, як ризик успішної атаки зростає зі збільшенням кількості критичних потоків. Сенс простий: що більше активних сесій/каналів/вузлів, то більше шансів, що хоча б один з них опиниться в умовах, де компроміс і технічна експозиція збігаються. Також видно, що за більш «компромісного» режиму експлуатації (коли ручні винятки трапляються частіше) крива росте суттєво швидше. Праворуч наведена підсумкова діаграма інтегрального ризику для трьох режимів: ручного, частково автома-

ПАРАМЕТРИЗАЦІЯ МОДЕЛІ РИЗИКУ

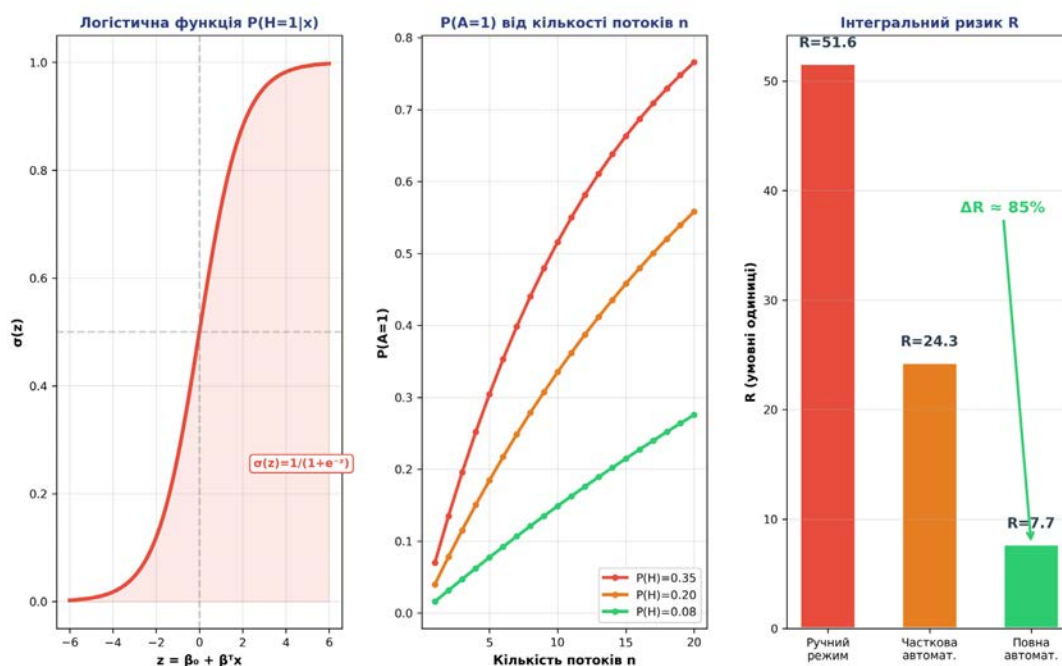


Рис. 2. Параметризація моделі ризику та ефект переходу від ручного керування до повної автоматизації

тизованого та повністю автоматизованого. Вона показує, що навіть часткова автоматизація вже помітно зменшує ризик, але максимальний ефект дає повний перехід до автоматизованих політик і безпечного керувального контуру – із зниженням ризику приблизно на 85%. Це і є головна ідея: найкращий результат досягається не «латанням» наслідків, а усуненням причин – тобто зменшенням простору для ручних небезпечних винятків.

Результати. У цьому розділі подано кількісні та якісні результати стендового експерименту, побудованого за методологією, описаною в розділі «Методи», де гібридна кібератака моделювалася як двофазний процес: (1) створення інцидентного тиску на персонал із провокування сумісних/доступних збоїв та (2) технічна експлуатація сформованих «легалізованих винятків» у конфігураціях транспортного рівня та керувального контуру. Центральним об'єктом аналізу є взаємодія двох чинників ризику: людського компромісу (вимкнення перевірки сертифіката, прийняття самопідписаних сертифікатів без контролю довіри, дозвіл downgrade і легасі TLS/SSL, збереження SNMPv2c із дефолтними community-рядками) та технічної експозиції (наявність слабких режимів узгодження, уразливі вебінтерфейси firmware, можливість конфігураційних маніпуляцій через SNMP). Саме їх одночасна присутність, відповідно до прийнятої ймовірнісної моделі, створює максимальне вікно можливостей для сценаріїв MITM, викрадення сесій, підміни вузлів і деградації TLS-політик через керування.

Результати структуровано у відповідності до застосованих інструментів вимірювання та критеріїв ефективності. По-перше, наведено емпіричну «карту компромісів» і їхніх наслідків, сформовану на основі нормалізованих подій IDS/SIEM та журналів змін конфігурації, що дозволяє простежити причинно-наслідкові ланцюги між діями оператора і реалізацією атаки. По-друге, представлено порівняння підходів детекції та протидії: сигнатурних спрацювань IDS, кореляції у SIEM та глибинних моделей CNN+LSTM і AE+LSTM. Для CNN+LSTM використано подання Byte2Image для байтових послідовностей трафіку/артефактів, що підсилює виявлення структурних «відбитків» downgrade та легасі-узгоджень; AE+LSTM застосовано для виявлення аномальних часових патернів у потоках подій (сертифікаційні помилки, серії SNMP-записів, конфігураційні зміни), зокрема в умовах неповного логування. Якість порівнюється за метриками точності, повноти та F1,

а також за операційними показниками затримки детекції та відносного зниження ризику.

Окремо у «Результатах» подано числову ілюстрацію розрахунку інтегрального ризику на базі параметрів імовірності людського компромісу, технічної експозиції та кількості критичних потоків, щоб показати, як автоматизація криптографічних політик (policy-as-code) та перехід керувального контуру на SNMPv3 впливають на кінцеву ризикову величину. Таким чином, розділ демонструє не лише «що саме спрацювало», а й чому саме спрацювало: які компроміси є найбільш небезпечними, які механізми детекції стійкіші до гібридності сценарію, та який практичний ефект дає виключення ручних винятків у критичних параметрах TLS/SSL і керуванні пристроями.

Емпірична карта компромісів та їхніх наслідків

У режимі ручної експлуатації найчастіше фіксувалися компроміси, пов'язані з обхідними шляхами у TLS: вимкнення перевірки сертифікатів, прийняття самопідписаних сертифікатів без контролю довіри, зниження мінімальної версії протоколу та розширення наборів шифрів. Ці дії створювали умови для downgrade-атак, подібних до Logjam і FREAK, у частині деградації параметрів узгодження та повернення небезпечних режимів [5, 6]. Виявлено також ефект суміжних сервісів: наявність легасі-протоколів у прошивках або на периферії збільшувала експозицію до сценаріїв, аналогічних DROWN [7]. Активний fallback до SSL 3.0 у вебінтерфейсі керування формував умови для відтворення класу проблем, продемонстрованого POODLE [8]. У керувальному контурі ключовим ризиком став SNMPv2c із дефолтними community-рядками, що дозволяло змінювати конфігураційні параметри й опосередковано впливати на профіль TLS.

Таблиця 1 систематизує причинно-наслідковий зв'язок між конкретними управлінськими рішеннями персоналу та їх технічними наслідками на транспортному рівні мережі. Вона показує, що кожен «операційний компроміс», зроблений з міркувань швидкого відновлення сервісу або сумісності, безпосередньо трансформується у визначений клас атак.

Перший рядок демонструє, що вимкнення перевірки сертифіката або ігнорування помилок ланцюжка довіри фактично нівелює модель автентифікації TLS. У такому стані система стає вразливою до атаки посередника (MITM), підміни вузла та викрадення сесії, оскільки клієнт більше не перевіряє достовірність сервера.

Компроміси людського фактора та їх технічні наслідки на транспортному рівні

№	Компроміс (людська дія)	Технічний наслідок та типові вектори атак
1	Вимкнення перевірки сертифіката або ігнорування помилок ланцюжка довіри	Атака посередника (MITM), підміна вузла, викрадення сесії, компрометація автентифікації
2	Дозвіл застарілих версій TLS для сумісності	Downgrade-атаки та повернення небезпечних режимів; кореляція з кейсами Logjam/FREAK
3	Прийняття самопідписаного сертифіката без контролю довіри та ротації	Підміна сертифіката, повторна автентифікація, порушення моделі довіри
4	Підтримка легасі SSL у вебінтерфейсі прошивки або слабких наборів шифрів	Експлуатація уразливостей класу POODLE/DROWN, компрометація керування
5	Використання SNMPv2c з дефолтними community-рядками в керувальному контурі	Несанкціоновані конфігураційні зміни, деградація TLS-політик, firmware-маніпуляції

Другий рядок відображає наслідки дозволу застарілих версій TLS «для сумісності». Такий крок створює умови для downgrade-атак, коли зломисник примусово знижує параметри узгодження до слабших режимів. Це концептуально узгоджується з кейсами Logjam і FREAK, де саме підтримка легасі-параметрів відкривала шлях до криптографічної деградації.

Третій рядок стосується прийняття самопідписаного сертифіката без формалізованого контролю довіри та ротації. У цьому випадку руйнується ієрархія довірених центрів сертифікації, що дозволяє підміну сертифіката та повторну автентифікацію під контролем зломисника.

Четвертий рядок демонструє ризик підтримки легасі SSL або слабких наборів шифрів у вебінтерфейсах прошивки. Така практика створює канал для експлуатації історичних уразливостей класу POODLE або DROWN і може призвести до компрометації керування пристроєм.

П'ятий рядок відображає критичність керувального контуру: використання SNMPv2c з дефолтними community-рядками дозволяє несанкціоновані конфігураційні зміни. Наслідком є деградація TLS-політик та потенційні firmware-маніпуляції, що опосередковано впливають на транспортний рівень.

Таким чином, таблиця наочно підтверджує ключову тезу дослідження: уразливість транспортного рівня виникає не стільки через слабкість протоколів, скільки через легалізовані операційні винятки. Вона слугує емпіричною «картою компромісів», яка пов'язує людські дії з конкретними технічними векторами атак і формує основу для подальшого кількісного аналізу ризику, описаного у статті.

Порівняння детекції: правила, кореляція та нейромережі

Правила IDS, зокрема в реалізаціях типу Snort [20], швидко фіксують відомі індикатори (заборонені набори шифрів, підозрілі узгодження, нети-

пові поля), однак у гібридному сценарії вони часто не відрізняють легітимну компромісну дію оператора від атаки, оскільки симптоми збігаються. Кореляція у SIEM покращує відокремлення за рахунок журналів змін конфігурації та прив'язки до облікових записів, але її ефективність чутлива до повноти логів [22]. Мережеве спостереження у стилі Bro або Zeek [21] дає глибший контекст (сесійні ознаки, послідовність подій), проте потребує масштабованого зберігання.

Моделі CNN+LSTM та AE+LSTM демонструють кращу відтворюваність детекції, оскільки навчаються на послідовностях подій, які відображають як технічні, так і організаційні патерни. CNN+LSTM, базована на Byte2Image [26], краще відокремлює downgrade від ручних винятків, оскільки фіксує структурні сліди узгодження та повторювані ознаки легасі-параметрів. AE+LSTM краще виявляє повільні відхилення, наприклад поступове зростання сертифікаційних помилок або нетипові серії SNMP записів, що корисно при низькій спостережуваності атак.

Таблиця 2 відображає порівняльну оцінку ефективності різних підходів до виявлення та зниження ризику в умовах гібридних атак на транспортному рівні. Вона поєднує класичні сигнатурні та кореляційні механізми з моделями глибинного навчання й окремо демонструє ефект організаційно-технічної автоматизації політик.

Перший рядок характеризує IDS на основі сигнатур. Показники точності (0.81) та повноти (0.63) свідчать про здатність надійно фіксувати відомі індикатори, однак порівняно низька повнота означає втрату частини складних або нетипових сценаріїв. Затримка детекції мінімальна (1.1 с), що є перевагою для оперативного реагування, проте сумарне зниження ризику становить лише 18%, оскільки сигнатури не покривають організаційний компонент компромісів.

Другий рядок демонструє можливості SIEM-кореляції. Зростання точності до 0.86 і повноти до

0.75 пояснюється використанням журналів змін конфігурації та зв'язуванням подій з обліковими записами. Водночас затримка зростає до 4.6 с через необхідність агрегування та аналізу подій. Відносне зниження ризику (42%) суттєво вище, ніж у чистих сигнатур, але залежить від повноти логування [22] та коректності нормалізації подій.

Третій рядок представляє модель CNN+LSTM із перетворенням Byte2Image. Високі значення точності (0.92), повноти (0.89) та F1-міри (0.90) свідчать про здатність моделі одночасно враховувати структурні ознаки узгодження TLS і часову послідовність подій. Затримка (2.4 с) залишається прийнятною для практичного застосування. Найважливіше – відносне зниження ризику сягає 78%, що демонструє ефективність комбінованого аналізу технічних і поведінкових патернів.

Четвертий рядок (AE+LSTM) демонструє подібні інтегральні показники F1 (0.90) за рахунок високої повноти (0.91). Цей підхід краще виявляє поступові відхилення – наприклад, серії аномальних SNMP-записів або зростання поми-

лок сертифікації. Затримка дещо більша (3.0 с), але зниження ризику становить 74%, що підтверджує ефективність автоенкодерного підходу при неповній розмітці.

Останній рядок – «Політики як код + SNMPv3» – не містить класичних метрик детекції, оскільки йдеться не про виявлення, а про превентивну зміну архітектури. Саме автоматизація криптографічних політик і перехід керувального контуру на захищену модель SNMPv3 забезпечують найбільший ефект – зниження ризику на 85%. Це підтверджує ключовий висновок роботи: найвища результативність досягається не лише вдосконаленням детекторів, а усуненням передумов для людських компромісів.

Таким чином, таблиця 2 демонструє еволюцію від реактивної детекції до проактивної архітектурної профілактики. Чим більше підхід інтегрує часовий контекст, структурні ознаки та автоматизацію політик, тим вищим є його внесок у реальне зниження інтегрального ризику, що узгоджується із загальною логікою дослідження.

Таблиця 2

Порівняння підходів виявлення/протидії та їх впливу на ризик

Підхід	Точність	Повнота	F1	Затримка, с	Зниження ризику, %
IDS (сигнатури)	0.81	0.63	0.71	1.1	18
SIEM (кореляція)	0.86	0.75	0.80	4.6	42
CNN+LSTM (Byte2Image)	0.92	0.89	0.90	2.4	78
AE+LSTM (аномалії)	0.89	0.91	0.90	3.0	74
Політики як код + SNMPv3 (автоматизація)	--	--	--	--	85

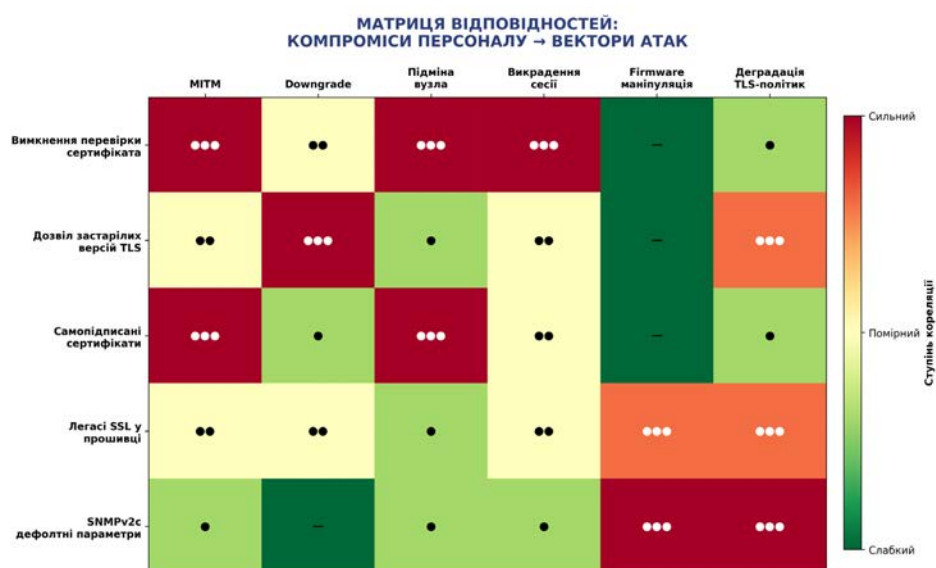


Рис. 3. Порівняльна динаміка виявлення гібридної атаки та вплив автоматизації на часові характеристики детекції

На рисунку відображено часову структуру розвитку гібридного сценарію та реакцію різних механізмів детекції. Перша фаза відповідає інцидентному тиску на персонал (зростання помилок сертифікації, конфігураційні зміни, активація сумісних режимів), друга – технічній експлуатації сформованих винятків (downgrade, MITM, маніпуляції через керувальний контур). Графічно показано, що сигнатурні механізми IDS фіксують окремі технічні індикатори вже на стадії активної експлуатації, проте часто пропускають початкову фазу організаційного компромісу. Кореляція у SIEM дозволяє пов'язати конфігураційні зміни з подальшими аномаліями, однак її спрацювання залежить від повноти журналювання. Нейромережіві моделі (CNN+LSTM та AE+LSTM) демонструють здатність виявляти відхилення раніше – ще на етапі накопичення аномальних патернів, що скорочує затримку детекції та зменшує тривалість вікна атаки. Таким чином, рисунок 3 ілюструє ключовий результат розділу: найбільш ефективною є не ізольована технологія, а інтегрована архітектура, де автоматизовані політики обмежують простір компромісів, а гібридна аналітика забезпечує своєчасне виявлення відхилень у динаміці системи. Отримані результати стендового експерименту підтверджують, що критичні інциденти на транспортному рівні виникають переважно як наслідок одночасної дії двох чинників: компромісних рішень персоналу під інцидентним тиском та наявної технічної експозиції (легасі-режими, уразливі компоненти прошивки, небезпечні налаштування керування). Саме їх поєднання формує «легалізовані винятки», які перетворюють формально захищену конфігурацію на практично вразливу та відкривають шлях до downgrade, MITM і маніпуляцій через керувальний контур. Порівняння підходів виявлення показало, що сигнатурні IDS ефективні для швидкого реагування на відомі індикатори, але мають обмеження у гібридних сценаріях, де технічні симптоми маскуються під «легітимну» операційну діяльність. Кореляція у SIEM підвищує якість відокремлення подій за рахунок журналів змін, проте її результативність критично залежить від повноти логування та трасованості конфігураційних дій. Натомість моделі CNN+LSTM та AE+LSTM демонструють кращу відтворюваність детекції, оскільки враховують як структурні, так і часові патерни, зменшуючи тривалість вікна атаки та підвищуючи стійкість до нетипових варіацій сценарію. Найбільш практично значущим результатом є те, що максимальне зниження інтегрального

ризикую досягається не лише посиленням детекторів, а архітектурним усуненням передумов компромісу: автоматизацією криптографічних політик, заборонаю ручних винятків і переведенням керувального контуру на захищені режими. Це узгоджується з логікою Zero Trust і пояснює, чому превентивні заходи дають більший ефект, ніж суто реактивні механізми виявлення. Разом із тим отримані дані потребують інтерпретації з урахуванням обмежень стенду, якості журналювання та організаційних компромісів між жорсткістю політик і операційною гнучкістю. Тому в наступному розділі «Discussion» зосередимося на поясненні причин спостережуваних ефектів, аналізі обмежень і переносимості результатів у реальні мережі, а також на формуванні практичних рекомендацій щодо впровадження підходів Zero Trust та «policy-as-code» у продуктивній експлуатації.

Дискусія. Результати підтверджують, що вразливості транспортного рівня мають соціотехнічну природу: технічна експозиція існує, але перетворюється на інцидент тоді, коли персонал легалізує виняток. На відміну від роботи [4], де основний акцент зроблено на формальній складності машин станів TLS і на тому, як граничні переходи та помилки обробки можуть відкривати поверхню атаки, у цьому дослідженні підкреслено механізм закріплення цієї поверхні в експлуатації. Якщо під час інциденту дозволяється вимикати перевірку сертифікатів або розширювати сумісність, то система починає функціонувати у режимі постійного ризику, який важко помітити традиційними сигналами.

Порівняно з Logjam [5], що демонструє можливість примусової деградації криптографічних параметрів у разі підтримки слабких груп, отримані дані показують, що подібна деградація часто запускається не лише активним зловмисником, а й самим персоналом, який розширює сумісність для швидкого відновлення сервісу. Тобто інженерний висновок полягає у зміщенні фокуса з пошуку універсально безпечного алгоритму на контроль процесу прийняття рішень і конфігураційного менеджменту, а також на автоматизацію політик.

Важливим компонентом є керувальний контур. Якщо SNMP залишається у незахищеному режимі або прошивка містить легасі SSL-компоненти, то зловмисник може обійти контроль транспортних політик опосередковано, змінюючи налаштування або підмінюючи артефакти довіри. Це означає, що заходи безпеки TLS мають супроводжуватися вимогами до резильєнтності прошивок [30] та базових кіберспроможностей пристроїв [16].

У термінах MITRE ATT&CK такі сценарії поєднують техніки доступу до керування, модифікації конфігурацій і персистентності [19]. Тому управління ризиком транспортного рівня має включати контроль прошивок, політики оновлення та аудит керувальних протоколів.

Обмеження дослідження пов'язані з умовністю стенду та параметризацією втрат. У реальних мережах різноманітність клієнтів і бібліотек TLS є значно вищою, а сумісність нерідко визначається вимогами партнерів або спадкових систем [10]. Друга група обмежень стосується повноти журналів: якщо логи з проксі, систем керування сертифікатами та мережних пристроїв неповні, кореляція у SIEM втрачає контекст, а моделі машинного навчання отримують спотворені ознаки [22]. Третє обмеження – концептуальний дрейф, який вимагає регулярного перевчання моделей і контролю метрик на нових профілях трафіку [25]. Нарешті, організаційний компроміс полягає у балансі між жорсткістю політик і операційною гнучкістю: занадто суворі правила без процесу винятків можуть підштовхнути персонал до небезпечних обхідних практик.

Перспективи розвитку напряму пов'язані з Zero Trust. Архітектура нульової довіри [29] зміщує фокус із разових налаштувань на безперервну перевірку, що природно зменшує простір для ручних винятків. Практично це означає централізоване керування політиками шифрування як код, короткоживучі сертифікати, взаємну автентифікацію там, де це виправдано, та автоматичну реакцію на деградацію параметрів. Другий перспективний напрям – multimodal AI: об'єднання часових рядів SIEM/IDS, байтових артефактів (Byte2Image), конфігураційних дифів, подій керувального контуру та навіть текстових заявок на зміни для точнішого відокремлення людського компромісу від атаки. У цій парадигмі детекція перетворюється

на інструмент управління рішеннями, а не лише на сигнал про інцидент.

Висновки. Дослідження показало, що компроміси персоналу є системним джерелом вразливостей транспортного рівня: вимкнення валідації сертифікатів, допуск downgrade, підтримка легасі SSL у прошивках та небезпечні режими SNMP у керувальному контурі створюють умови для перехоплення сесій і компрометації автентифікації. Запропонована ймовірнісна модель ризику розділяє людський фактор і технічну експозицію, а оптимізаційна постановка дозволяє обґрунтувати вибір заходів автоматизації. Показано, що поєднання кореляції у SIEM, сигналів IDS та глибинних моделей CNN+LSTM і AE+LSTM забезпечує кращу відтворюваність виявлення та підсилює аудит змін криптографічних профілів.

Практичні рекомендації: – впровадити політики TLS як код із заборонаю ручних винятків для перевірки сертифікатів і параметрів узгодження; – відмовитися від застарілих версій протоколу та слабких режимів сумісності відповідно до профілів і депрекацій [3, 13]; – перевести керувальний контур на SNMPv3 та закрити дефолтні налаштування прошивок [15]; – забезпечити підписані оновлення firmware, контроль компонентів SSL/TLS у вебінтерфейсах і відновлюваність платформи [30]; – підвищити якість журналювання та кореляції SIEM, щоб кожна зміна криптографічної політики мала трасованість до суб'єкта й причини [22]; – розвивати Zero Trust як організаційну практику, де довіра перевіряється безперервно, а критичні рішення максимально автоматизовані [29].

Таким чином, безпека транспортного рівня в умовах гібридних кібератак має розглядатися як задача управління компромісами: технічні стандарти потрібні, але вирішальним чинником є дисципліна конфігурацій, автоматизація політик і виключення людського фактора з критичних рішень.

Список літератури:

1. Rescorla E. The Transport Layer Security (TLS) Protocol Version 1.3 : RFC 8446. 2018. DOI: <https://doi.org/10.17487/RFC8446>
2. Dierks T., Rescorla E. The Transport Layer Security (TLS) Protocol Version 1.2 : RFC 5246. 2008. DOI: <https://doi.org/10.17487/RFC5246>
3. Sheffer Y., Holz R., Saint-Andre P. Recommendations for Secure Use of Transport Layer Security (TLS) and Datagram Transport Layer Security (DTLS) : RFC 7525. 2015. DOI: <https://doi.org/10.17487/RFC7525>
4. Moriarty K., Farrell S. Deprecating TLS 1.0 and TLS 1.1 : RFC 8996. 2021. DOI: <https://doi.org/10.17487/RFC8996>
5. Sheffer Y., Holz R., Saint-Andre P. Summarizing Known Attacks on Transport Layer Security (TLS) and Datagram TLS (DTLS) : RFC 7457. 2015. DOI: <https://doi.org/10.17487/RFC7457>
6. Harrington D., Presuhn R., Wijnen B. An Architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks : RFC 3411. 2002. DOI: <https://doi.org/10.17487/RFC3411>

7. Blumenthal U., Wijnen B. User-based Security Model (USM) for version 3 of the Simple Network Management Protocol (SNMPv3) : RFC 3414. 2002. DOI: <https://doi.org/10.17487/RFC3414>
8. McKay K. A., Cooper D. A. Guidelines for the Selection, Configuration, and Use of Transport Layer Security (TLS) Implementations : NIST SP 800-52 Rev. 2. – Gaithersburg, MD : National Institute of Standards and Technology, 2019. DOI: <https://doi.org/10.6028/NIST.SP.800-52r2>
9. Rose S., Borchert O., Mitchell S., Connelly S. Zero Trust Architecture : NIST SP 800-207. – Gaithersburg, MD : National Institute of Standards and Technology, 2020. DOI: <https://doi.org/10.6028/NIST.SP.800-207>
10. Regenscheid A. Platform Firmware Resiliency Guidelines : NIST SP 800-193. – Gaithersburg, MD : National Institute of Standards and Technology, 2018. DOI: <https://doi.org/10.6028/NIST.SP.800-193>
11. Kent K., Souppaya M. Guide to Computer Security Log Management : NIST SP 800-92. – Gaithersburg, MD : National Institute of Standards and Technology, 2006. DOI: <https://doi.org/10.6028/NIST.SP.800-92>
12. Scarfone K., Souppaya M., Cody A., Orebaugh A. Technical Guide to Information Security Testing and Assessment : NIST SP 800-115. – Gaithersburg, MD : National Institute of Standards and Technology, 2008. DOI: <https://doi.org/10.6028/NIST.SP.800-115>
13. Scarfone K., Mell P. Guide to Intrusion Detection and Prevention Systems (IDPS) : NIST SP 800-94. – Gaithersburg, MD : National Institute of Standards and Technology, 2007. DOI: <https://doi.org/10.6028/NIST.SP.800-94>
14. Joint Task Force. Security and Privacy Controls for Information Systems and Organizations : NIST SP 800-53 Rev. 5. – Gaithersburg, MD : National Institute of Standards and Technology, 2020. DOI: <https://doi.org/10.6028/NIST.SP.800-53r5>
15. Cichonski P., Millar T., Grance T., Scarfone K. Computer Security Incident Handling Guide : NIST SP 800-61 Rev. 2. – Gaithersburg, MD : National Institute of Standards and Technology, 2012. DOI: <https://doi.org/10.6028/NIST.SP.800-61r2>
16. Joint Task Force Transformation Initiative. Guide for Conducting Risk Assessments : NIST SP 800-30 Rev. 1. – Gaithersburg, MD : National Institute of Standards and Technology, 2012. DOI: <https://doi.org/10.6028/NIST.SP.800-30r1>
17. Howell G., Franklin J., Sritapan V., Souppaya M., Scarfone K. Guidelines for Managing the Security of Mobile Devices in the Enterprise : NIST SP 800-124 Rev. 2. – Gaithersburg, MD : National Institute of Standards and Technology, 2023. DOI: <https://doi.org/10.6028/NIST.SP.800-124r2>
18. Fagan M. J., Megas K. N., Scarfone K., Smith M. IoT Device Cybersecurity Capability Core Baseline : NISTIR 8259A. – Gaithersburg, MD : National Institute of Standards and Technology, 2020. DOI: <https://doi.org/10.6028/NIST.IR.8259A>
19. Ross R., Pillitteri V., Dempsey K., Riddle M., Guissanie G. Protecting Controlled Unclassified Information in Nonfederal Systems and Organizations : NIST SP 800-171 Rev. 2. – Gaithersburg, MD : National Institute of Standards and Technology, 2020. DOI: <https://doi.org/10.6028/NIST.SP.800-171r2>
20. Ross R., Pillitteri V., Guissanie G., Wagner R., Graubart R., Bodeau D. Enhanced Security Requirements for Protecting Controlled Unclassified Information : NIST SP 800-172. – Gaithersburg, MD : National Institute of Standards and Technology, 2021. DOI: <https://doi.org/10.6028/NIST.SP.800-172>
21. Beurdouche B., Bhargavan K., Delignat-Lavaud A., Fournet C., Kohlweiss M., Pironti A., Strub P.-Y., Zinzindohoue J. K. A Messy State of the Union: Taming the Composite State Machines of TLS // 2015 IEEE Symposium on Security and Privacy (SP). 2015. P. 535–552. DOI: <https://doi.org/10.1109/SP.2015.39>
22. Adrian D., Bhargavan K., Durumeric Z., Gaudry P., Green M., Halderman J. A., Heninger N., Springall D., Thomé E., Valenta L., VanderSloot B., Wustrow E., Zanella-Béguelin S., Zimmermann P. Imperfect Forward Secrecy: How Diffie-Hellman Fails in Practice // Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS '15). 2015. P. 5–17. DOI: <https://doi.org/10.1145/2810103.2813707>
23. Xia H., Pei Q., Xi Y. The Analysis and Research of Freak Attack Based on OpenSSL // Proceedings of the 6th International Conference on Information Engineering for Mechanics and Materials. 2016. P. 15–19. DOI: <https://doi.org/10.2991/icimm-16.2016.4>
24. Paxson V. Bro: A System for Detecting Network Intruders in Real-Time // Computer Networks. 1999. Vol. 31(23–24). P. 2435–2463. DOI: [https://doi.org/10.1016/S1389-1286\(99\)00112-7](https://doi.org/10.1016/S1389-1286(99)00112-7)
25. LeCun Y., Bottou L., Bengio Y., Haffner P. Gradient-Based Learning Applied to Document Recognition // Proceedings of the IEEE. 1998. Vol. 86(11). P. 2278–2324. DOI: <https://doi.org/10.1109/5.726791>
26. Hochreiter S., Schmidhuber J. Long Short-Term Memory // Neural Computation. 1997. Vol. 9(8). P. 1735–1780. DOI: <https://doi.org/10.1162/neco.1997.9.8.1735>
27. Chandola V., Banerjee A., Kumar V. Anomaly Detection: A Survey // ACM Computing Surveys. 2009. Vol. 41(3). Art. 15. DOI: <https://doi.org/10.1145/1541880.1541882>
28. Nataraj L., Karthikeyan S., Jacob G., Manjunath B. S. Malware Images: Visualization and Automatic Classification // Proceedings of the 8th International Symposium on Visualization for Cyber Security (VizSec '11). 2011. P. 1–7. DOI: <https://doi.org/10.1145/2016904.2016908>

29. Yin C., Zhu Y., Fei J., He X. A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks // IEEE Access. 2017. Vol. 5. P. 21954–21961. DOI: <https://doi.org/10.1109/ACCESS.2017.2762418>
30. Shone N., Ngoc T. N., Phai V. D., Shi Q. A Deep Learning Approach to Network Intrusion Detection // IEEE Transactions on Emerging Topics in Computational Intelligence. 2018. Vol. 2(1). P. 41–50. DOI: <https://doi.org/10.1109/TETCI.2017.2772792>
31. Sheffer Y., Saint-Andre P., Fossati T. Recommendations for Secure Use of Transport Layer Security (TLS) and Datagram Transport Layer Security (DTLS) : RFC 9325. 2022. DOI: <https://doi.org/10.17487/RFC9325>

Prokopovych-Tkachenko D.I., Torstensson O., Cherkaskyi D.O., Peremetchyk D.O., Bert S.D.
SOCIOTECHNICAL FACTORS IN THE EMERGENCE OF VULNERABILITIES
IN NETWORK PROTOCOLS FOR SECURE DATA TRANSMISSION

The article presents a comprehensive analysis of sociotechnical factors that lead to the emergence of vulnerabilities in secure data transmission protocols during the operation of electronic communication networks. The purpose of the work is to formalize the contribution of human compromises to transport layer vulnerabilities and to develop a practical methodology for risk reduction through the automation of cryptographic policies and hybrid analytics using IDS and SIEM systems. A probabilistic risk model is proposed that separates the human factor (personnel compromises under incident pressure) and technical exposure (legacy protocols and insecure management modes). Typical compromise actions are considered, such as disabling certificate verification, accepting untrusted certificates, supporting outdated compatibility modes, and using insecure administration protocols like SNMPv2c, which create conditions for cryptographic degradation and session hijacking. For empirical validation, a hybrid attack testbed was used, modeling phases of organizational pressure and subsequent technical exploitation of "legalized exceptions." Special attention is given to the integration of detection mechanisms with deep learning models: convolutional and recurrent neural networks (CNN+LSTM) using the "Byte2Image" transformation for structural traffic analysis, and autoencoders combined with LSTM (AE+LSTM) for anomaly detection in event sequences. The results showed that transitioning to "policy-as-code" automation and secure management mechanisms (SNMPv3) significantly reduces the integral risk (by approximately 85%) by minimizing the space for manual dangerous exceptions. The limits of model applicability, the influence of logging quality, and the organizational balance between policy rigidity and operational flexibility are analyzed within the framework of Zero Trust Architecture.

Keywords: transport, encryption, certification, trust, compatibility, interception, firmware, administration, correlation, neural networks, automation.

Дата першого надходження статті до видання: 18.02.2026

Дата прийняття статті до друку після рецензування: 13.03.2026

Дата публікації (оприлюднення) статті 11.05.2026

Пупена О.М.

<https://orcid.org/0000-0001-9089-8325>

Національний університет харчових технологій

ВИКОРИСТАННЯ AUTOMATIONML В ЖИТТЄВОМУ ЦИКЛІ СИСТЕМ КЕРУВАННЯ ЗА УМОВ ОБМЕЖЕНИХ РЕСУРСІВ

У статті розглянуто можливості використання AutomationML (AML) як нейтрального формату обміну інженерними даними між різними інструментами в ЖЦ систем керування. Проаналізовано основні принципи побудови CAEX та AML, що забезпечує формалізований опис структури технічних систем, їх компонентів, атрибутів та взаємозв'язків. Показано, що така модель дозволяє представляти інженерні дані у вигляді єдиної інформаційної структури, придатної для використання різними інженерними інструментами. Розглянуто можливості застосування AML на різних етапах життєвого циклу систем керування, включаючи етапи інженерного проектування, інтеграції та експлуатації. Окрему увагу приділено питанням підтримки стандарту програмними інструментами. Показано, що завдяки використанню XML-синтаксису моделі AML можуть оброблятися як спеціалізованими інженерними середовищами, так і універсальними програмними засобами, включаючи скриптові інструменти та табличні системи обробки даних. Розглянуто можливості використання спеціалізованих редакторів AML, а також інтеграцію стандарту з інженерними системами проектування. Описано можливості організації ітераційного обміну інженерними даними між різними інструментами, а також реалізації реверсних потоків інформації. Показано, що хоча стандарт AutomationML створює технічні передумови для таких процесів, практична реалізація ітераційної синхронізації значною мірою залежить від підтримки відповідних механізмів у конкретних інженерних інструментах. За результатами аналізу визначено основні переваги та обмеження використання AutomationML у процесах інженерного проектування систем автоматизації. Сформульовано напрями подальших досліджень, пов'язані з аналізом існуючих рекомендацій щодо обміну даними між CAE- та PLC-інструментами, дослідженням можливостей представлення моделей AutomationML у табличних інструментах та розробленням прототипів відображення інженерних моделей у середовищах роботи з майстерданими.

Ключові слова: AutomationML, CAEX, AML, обмін інженерними даними, системи керування, CAE-системи, ЖЦ, життєвий цикл системи, цифровий двійник, цифрова нитка.

Постановка проблеми. Досвід розробки систем керування технологічними об'єктами з великою кількістю давачів та виконавчих механізмів показав необхідність інтеграції між різними інструментами, що задіяні в процесах життєвого циклу. Ці проблеми висвітлені у роботі [1], зокрема в частині обміну між засобами проектування систем керування та інструментами розроблення програмного забезпечення для PLC та SCADA/HMI. Дослідження, результати якого висвітлені в даній статті, є продовженням робіт, описаних в [1].

У роботі [1] висвітлений поточний стан та перспективи розвитку каркасу PACFramework для розроблення ПЗ для PLC та SCADA/HMI. Каркас стандартизує структуру програм та принципи їх розроблення, що у свою чергу дозволяє стандартизувати суміжні процеси в життєвому циклі, зокрема

автоматизація процесів створення ПЗ на основі проєктних даних. Утиліти роботи з проєктними даними (PACFramework Tools) були апробовані на кількох об'єктах. Подальші дослідження привели до формування нової структури, яка показана на рис. 1.

У основі рішення стоїть платформа яка базується на концепції цифрового двійника і включає в себе сервіси для забезпечення цифрової нитки та моніторингу обладнання. Уся проєктна інформація, з якою працює розробник ПЗ зберігається в базі проєктних даних ("майстер даних", Master Data DB), які оновлюються з кількох суміжних інженерних систем. Необхідність синхронізації з різних сторін приводить до ряду проблем:

1. Як технічно забезпечити автоматизований обмін між САПР (CAE) та майстерданими в обидві сторони?

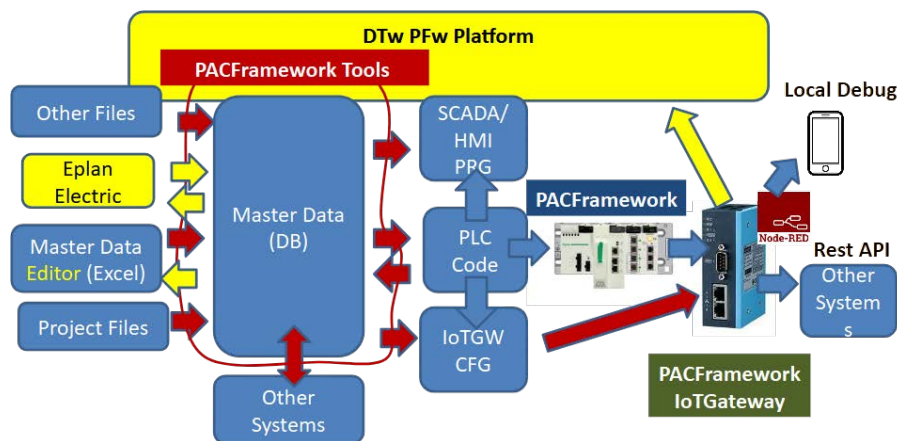


Рис. 1. Інфраструктура модифікованого каркасу PFW2

Джерело – [1]

2. Яка інформація має зберігатися в майстер-даних і в якому вигляді повинна представлятися?
3. Як забезпечити керування версіями?
4. Як технічно забезпечити доступ до майстер-даних як до єдиного джерела правди на всіх стадіях життєвого циклу?
5. Як вирішити все наведене вище за умов постійних змін та реверсної передачі проєктних даних.

Аналіз останніх досліджень і публікацій. Сама концепція описаного вище каркасу та проблем які виникають у ЖЦ систем керування спираються на цифрове представлення системи у вигляді моделей, релевантних до відповідного інженерного процесу, які розвиваються в процесі розвитку самої системи та потребують постійного узгодження з реальною системою. На сьогодні ці задачі вирішуються з використанням концепції, яка відома як цифровий двійник (digital twin) та наскрізними взаємопов'язаними інженерними процесами – цифровою ниткою (Digital Thread). Тому, перш за все, було проаналізовано стан стандартів в області цифрових двійників, як усталених зрілих практик.

У роботі [2] Пупена О.М., Клименко О.М та Полупан В.В. виділили ключові стандарти в області цифрових двійників, з яких принаймні два з них [3] та [4] посилаються на інші добре відомі стандарти. У третій частині стандарту ISO 23247-3 [3] зазначається, що представлення ОМЕ (спостережуваний виробничий елемент) може забезпечуватися кількома стандартами, які мають свої особливості та переваги. Розробникам цифрових двійників слід вибирати та комбінувати їх для свого рішення. Серед цих стандар-

тів релевантними до задачі є: ISO 10303 (STEP), ISO 15926-1, ISO 16739-1 (для BIM), IEC 62714 (AutomationML), IEC 63278-1 (AAS). Було проаналізовано призначення та зміст цих стандартів, їх характеристики зведені в таблицю 1.

Найбільш універсальне представлення надають підходи з ISO 15926, інші стандарти спеціалізуються на певній галузі застосування, а IEC 63278 – компонується з кількох стандартів, серед яких ті, що наведені в таблиці 1. Для інженерії автоматизованих систем найбільш релевантним є IEC 62714 (AML). Популярність та практичність AutomationML засвідчується дослідженнями багатьох вчених та практиків, зокрема в роботі [5].

У межах даного дослідження окремої уваги потребувало визначення підтримки формату інженерними інструментами. Проведений аналіз показав, що AML на сьогодні є практично найбільш вживаним стандартом, релевантним до зазначених потреб. Зокрема в [6] показана послідовність обміну даними між інструментами проєктування EPLAN, TIA Selection Tool та TIA Portal.

Подальші дослідження сфокусувалися саме на стандарті IEC 62714 і AutomationML. У своїй роботі [7] автори використовують AML на всіх стадіях життєвого циклу як для обміну між різними інструментами так і для збереження моделей. Великою мірою успіх такого використання спирається на засоби керування версіями та візуалізації на всіх етапах життєвого циклу. Серед викликів автори бачать проблему у надмірній гнучкості, що потребує кращої керованості для спрощення інтеграції.

У статті [8] AML застосовується як метамодель для структурованого опису компонентів, їхніх

Характеристики стандартів для інженерного представлення систем.

Стандарт	Галузь застосування	Що моделює	Орієнтація
IEC 62714 (AML)	автоматизація виробництва	структура системи, пристрої, функції, зв'язки	інтеграція інженерних інструментів
ISO 10303 (STEP)	машинобудування	геометрія та структура виробу	CAD/PLM інтеграція
ISO 15926	переробна промисловість	інтеграція даних ЖЦ	міждисциплінарна інтеграція
ISO 16739-1 (IFC)	будівництво (BIM)	будівля та інженерні системи	BIM-проектування
IEC 63278 (AAS)	Industrie 4.0	цифрове представлення активу	цифровий актив

атрибутів та зв'язків, що дозволяє формувати узгоджену інформаційну модель цифрового двійника. Автори показують, що AML може виступати універсальним структурним контейнером для таких даних і основою для їх подальшої передачі через проміжне програмне забезпечення.

У роботі [9] AML використано як структурний каркас реалізований як проміжний шар для п'яти-вимірної моделі цифрового двійника в аерокосмічній сфері. Автори демонструють, як через AML можна формалізувати міждисциплінарні зв'язки в єдиній структурі.

У роботі [10] пропонується використовувати AutomationML у якості цільового нейтрального формату експорту, на якому будується хаб даних із "плинною семантикою". Замість очікування єдиного підтримуваного формату, автори роблять відображення форматів між різними інструментами "на льоту" через єдиний AML.

У статті [11] автори зробили систематичний огляд джерел присвячених обміну даними в Industrie 4.0 із використанням рішень на основі AutomationML, B2MML та International Data Spaces і зробили висновок про доцільність їх системного використання в інженерії систем.

У своїй роботі [12,13] автори запропонували та апробували підхід використання AutomationML в якості мови для збереження конфігураційної інформації для мережних пристроїв. Це дозволяє включати класичні описові файли пристроїв (DD-файли) в оболонку AutomationML разом з усією додатковою інформацією, зберігати і переносити їх.

Аналіз літератури показав, що для обміну даними між різними інструментами в життєвому циклі систем керування найбільш підходить AutomationML. Він добре масштабується і на задачі проектування технологічних комплексів та виробництв. Однак автору даної статті не зустрічалися роботи, присвячені використанню AML саме в потоці процесів роботи з кодом для PLC та SCADA/HMI. У роботі [6] AML використовується

виключно для обміну апаратною конфігурацією, а потрібно мати інформацію про об'єкти керування, процедури, означення тривог, тощо.

Більшість робіт використовують власні інструменти для контролю та керування даними AML. На практиці ж потрібні загальнодоступні інструменти, принаймні в частині контролю та керування даними, які циркулюють між інструментами.

Постановка завдання. Метою статті є визначення основних концепцій використання AutomationML в контексті життєвого циклу систем керування, за умов обмежених ресурсів, та можливих механізмів застосовності для загальноживаних інструментів проектування. Основні цілі дослідження:

- визначити основні концепції AutomationML
- проаналізувати застосовність AutomationML в життєвому циклі систем керування в частині обміну даними між інструментами проектування та розроблення ПЗ
- проаналізувати ступінь доступності інструментів роботи з AutomationML та інтеграції з іншими інструментами в ЖЦ систем керування
- проаналізувати можливість ітераційної роботи та реверсних потоків
- проаналізувати можливість використання AML в процесі експлуатації системи

Виклад основного матеріалу

Моделі AutomationML та CAEX

Серія стандартів IEC 62714 (AutomationML) [14] означає універсальний формат обміну інженерними даними між програмними засобами в автоматизації промислових систем. Він призначений для обміну даними між різними дисциплінами інженерії. AutomationML базується на іншому форматі – CAEX.

CAEX є метамоделлю для інженерних даних. Система описується за допомогою одного або кількох файлів XML, в якому представлені: об'єкти моделювання, їхні атрибути, ієрархії, інтерфейси, зв'язки, типи та ролі. CAEX версії 3.0 затверджена

в додатку до стандарту ІЕС 62424, який прийнятий в Україні як ДСТУ EN 62424:2022 [5].

Аналіз формату CAEX в межах дослідження проводився за текстом стандарту ІЕС 62424 [5] а також аналізом прикладів та білих книг, що доступні на сайті www.automationml.org. При аналізі використовувався підхід порівняння файлів прикладів з викладеними базисами в стандарті.

Хоч стандарт зосереджується на представленні Р&ІД в форматі CAEX, ця мова є універсальною для моделювання будь-яких об'єктно-орієнтованих представлень даних. Тому він може використовуватись для представлення сутностей у електричних установках, пневматиці, гідравліці, тощо. Мова CAEX означено лише в додатку до стандарту.

Повна версія результату аналізу, зроблена автором дана в [16], який є у відкритому доступі. Багато корисного матеріалу можна знайти у посібниках по AML [17],[18]. У цій статті представлені тільки найважливіші моменти.

Нижче наведені основні концепції CAEX на прикладі простого моделювання технологічної установки "ProcessCell1" яка складається з двох ємностей "Tank1" і "Tank2" що зв'язані між собою трубою "Pipe". Задача моделювання – представити в форматі CAEX властивості установки, її входи/виходи, включені компоненти та їх зв'язки. У загальному технологічну схему можна показати як на рис.2 (вгорі).

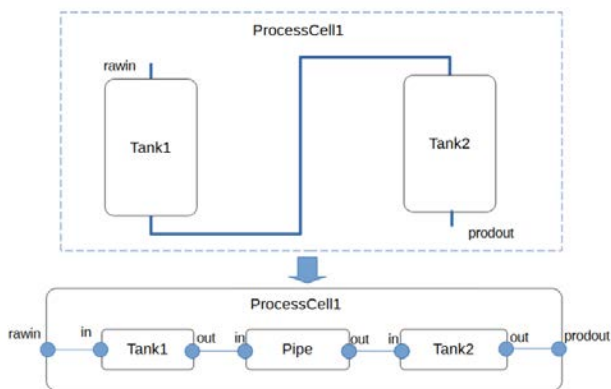


Рис. 2. Приклад модельованої установки (вгорі) і її параметричного представлення (знизу)

У CAEX модель представляється у вигляді "ієрархії екземплярів" (InstanceHierarchy), який є кореневим елементом для системної ієрархії об'єктних екземплярів. Екземпляри можуть мати зв'язки з бібліотечними CAEX елементами. На рис.3 показана модель в редакторі AutomationML editor, на якому видно кореневий елемент "Production", який представляє окремий ієрархічний об'єкт. Той у свою чергу включає

в себе об'єкти, які в CAEX представляються як "внутрішні елементи" (InternalElement). Можна вважати що внутрішній елемент представляє компонент структури об'єкта і може включати інші внутрішні елементи, представляючи таким чином композицію (ієрархію вкладеності) об'єктів.

У наведеному вище прикладі установку можна промодельовувати як 4-ри внутрішні елементи: "ProcessCell1" – як вкладений в "Production", і три включені в нього: "Tank1", "Tank2" і "Pipe" (див. рис.3). Необхідність моделювання "Pipe" як окремого об'єкту можна бути пов'язана, наприклад, з описом властивостей трубопроводу, що поєднує ємності. Так чи інакше перелік об'єктів, в яких є потреба в моделюванні, диктується призначенням моделі.

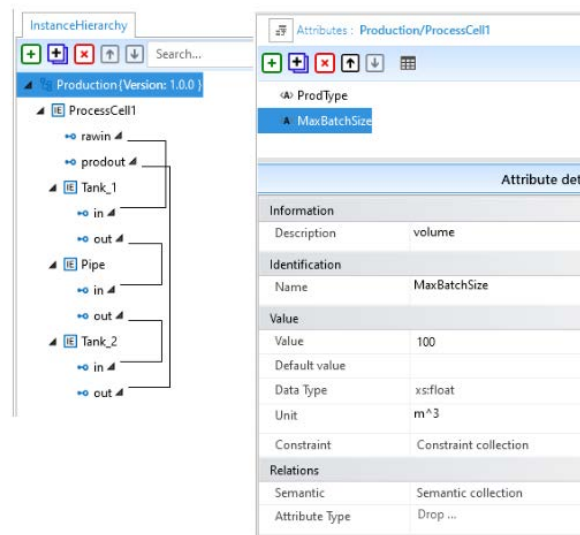


Рис. 3. Відображення ієрархії екземплярів (ліворуч) та атрибутів (Attributes)

Для того, щоб змодельовати поєднання компонентів між собою, в CAEX передбачений тип "зовнішній інтерфейс" (ExternalInterface), який описує можливість підключення до інших компонентів та внутрішні зв'язки ("InternalLinks"), які з'єднують компоненти між собою.

Крім самої ієрархії об'єктів та зв'язків між ними модель може містити їх властивості. Для установки "ProcessCell1" це може бути максимальний розмір партії, тип продукту, тощо, для ємностей – об'єм, тип матеріалу, тощо. В CAEX властивості об'єктів моделюються за допомогою атрибутів (Attributes).

Внутрішні елементи становлять ядро об'єктної моделі та містять усі об'єкти даних, включно з властивостями, інтерфейсами, зв'язками та посиланнями (рис. 4).

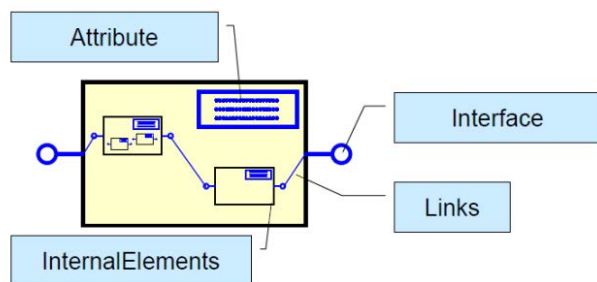


Рис. 4. Архітектура "InternalElement" та "SystemUnitClass" (Рисунок взятий зі стандарту IEC 62424 Ed.2)

CAEX підтримує можливість формувати кілька ієрархій екземплярів, що може бути корисно наприклад для представлення однієї і тієї установи в різних проекціях (функційної, продуктної або просторового розміщення), чи окремих ієрархій (механічна, електрична структура), тощо.

CAEX надає можливість використовувати класи: SystemUnitClass – для прототипування технічних об'єктів; RoleClass – для семантичної класифікації об'єктів; InterfaceClass – для опису типізованих точок взаємодії об'єктів; AttributeType – для опису та обмеження атрибутів. Класи об'єднуються у бібліотеки, які можуть бути деталізовані за допомогою зв'язків усередині бібліотек, наприклад через наслідування або агрегацію. Між класом та його екземплярами встановлюється відповідне відношення "клас-екземпляр". Крім цього CAEX підтримує наслідування між двома класами, яка передбачає успадкування нащадком всього змісту батьківського класу. Класи ролей слугують для призначенні об'єктам певної ролі, тобто надання функціонального змісту для об'єктів і ієрархії екземплярів.

CAEX явно підтримує доступ до зовнішніх файлів за допомогою елемента "ExternalReference". Це дає можливість не обмежуватися одним файлом, і постачати комплексні моделі.

Не дивлячись на те, що CAEX досить простий і зрозумілий формат, він надає досить потужний механізм моделювання складних систем через класи, прототипи, складні властивості, зв'язки з сутностями в інших файлах. Тим не менше, пошук практичних кейсів використання CAEX саме в контексті області застосування IEC 62424 у відкритих джерелах так і не вдалося знайти.

На відміну від CAEX, AutomationML (AML), який базується на ньому, достатньо потужно висвітлений в багатьох матеріалах і має широкий спектр застосувань. Оскільки AML базується на CAEX, все наведене вище стосується і моделей AML.

Для аналізу AML використовувався текст стандарту IEC 62714[14], білі книги та матеріали, що є у відкритому доступі на сайті AutomationML, у тому числі самі файли прикладів.

AML поєднує наявні галузеві формати даних, розроблені для зберігання та обміну різними видами інженерної інформації. Ці формати використовуються у своїх початкових специфікаціях без модифікації під потреби AML. CAEX тут використовується в якості метамоделі для зв'язування різних форматів даних, що дає можливість робити розподілену архітектуру документів. Серія стандартів IEC 62714 складається з кількох частин, кожна з яких описує окремий аспект інженерних даних.

Основною структурою даних для інженерної інформації про якийсь промисловий об'єкт є його топологія, яка моделюється за допомогою формату CAEX. Дані про геометрію та кінематику зберігаються у форматі COLLADA™ (ISO/PAS 17506), а логічна інформація для ПЛК зберігається у форматі PLCopen XML. Стандарт передбачає можливість розширення та додавання будь-яких інших форматів як зовнішніх ресурсів.

Окрім посилань на COLLADA та PLCopen XML в плані представлення електричних даних, документ AML повинен містити клас ролей "AutomationMLBaseRole", який є базовим абстрактним типом ролі та батьківським класом для всіх стандартних або користувацьких класів ролей. У залежності від версії класу, він містить різноманітні ролі. У ньому міститься роль "ExternalData" та клас інтерфейсу "ExternalDataConnector", який є базовим абстрактним типом інтерфейсу та повинен використовуватися для опису інтерфейсів-з'єднувачів, що посилаються на зовнішні документи.

Крім того вводиться базовий інтерфейс "AutomationMLBaseInterface" як уніфікована точка семантичного зв'язування інженерних об'єктів. Крім бібліотек, AutomationML означає правила використання ролей, інтерфейсів, ідентифікаторів, посилань між документами та організації розподілених моделей.

Бібліотеки включають з десяток класів ролей і з десяток класів інтерфейсів, що мають відношення до певної предметної області моделювання. Аналіз білих книг, що доступні на сайті показав, що AML варто сприймати як набір практик, тому навіть класи з бібліотеки, описані в стандарті не є обов'язковими для застосування. Варто заздалегідь звертати увагу на узгодженість бібліотек між інструментами при інтеграції.

Застосовність AutomationML в життєвому циклі систем керування

Наведений вище аналіз показав, що AutomationML може застосовуватися на різних етапах життєвого циклу систем керування, оскільки його модель даних призначена для представлення структури технічних систем, їхніх компонентів, атрибутів і зв'язків у нейтральному форматі. Це робить можливим використання AML як універсального контейнера для обміну інженерними даними між системами проєктування, моделювання та конфігурації систем автоматизації.

На ранніх етапах у моделі може фіксуватися склад обладнання, функціональні елементи системи та їхні інтерфейси, що створює основу для подальшої деталізації інженерної моделі. На етапі детального проєктування стандарт може використовуватися для передачі структурних даних між різними інженерними інструментами, зокрема між системами електротехнічного проєктування, програмування контролерів та іншими CAE-системами.

У подальших етапах життєвого циклу AML може застосовуватися для інтеграції моделей під час моделювання, тестування та віртуальної пусконаладки систем автоматизації.

Отже AutomationML застосовний на всіх стадіях життєвого циклу системи керування і може виступати одним із елементів реалізації цифрової нитки в інженерних процесах.

Варто окремо виділити, що AML дозволяє розширяти модель необхідними зовнішніми файлами, що робить його дуже гнучким не тільки при передачі даних між інструментами, але і для збереження майстерданих для цифрового двійника.

Підтримка AutomationML інструментами

AutomationML базується на синтаксисі XML, що робить його відкритим для обробки стандартними програмними засобами. Завдяки цьому моделі AML можуть створюватися, аналізуватися та змінюватися за допомогою власних скриптів або програм, які використовують стандартні бібліотеки роботи з XML. Такий підхід дозволяє автоматизувати перетворення моделей, виконувати перевірку структури даних або генерувати фрагменти моделі на основі зовнішніх джерел інформації.

Крім використання загальних засобів роботи з XML, для AutomationML існують спеціалізовані інструменти, зокрема AML Editor та інші утиліти, призначені для перегляду, редагування та перевірки моделей. Такі інструменти дозволяють працювати з ієрархічною структурою CAEX, переві-

ряти коректність використання бібліотек класів, інтерфейсів та атрибутів, а також виконувати валідацію моделі відповідно до схем стандарту. Це спрощує розроблення та аналіз моделей AML без необхідності безпосередньої роботи з XML-кодом.

Окремим напрямом підтримки AutomationML є інтеграція з інженерними інструментами проєктування. У межах цієї роботи повний огляд таких інструментів не проводився. Основна увага була зосереджена на підтримці AutomationML у середовищі Eplan, оскільки він використовується в поточних інженерних проєктах.

Водночас відкритість XML-формату дозволяє використовувати AutomationML і в більш простих інструментах обробки даних. Зокрема, дані AML можуть бути перетворені у табличне представлення та оброблятися в електронних таблицях, таких як Google Sheets. Використання таких інструментів має низький поріг входу, підтримує спільну роботу декількох користувачів і забезпечує зручні засоби візуалізації та аналізу даних. Це дозволяє залучати до роботи з інженерними даними ширше коло учасників проєкту та використовувати AutomationML як джерело структурованої інформації навіть поза спеціалізованими інженерними середовищами.

Наявність як універсальних засобів роботи з XML, так і спеціалізованих інструментів та функцій інтеграції в CAE-системах суттєво спрощує практичне використання AutomationML у процесах інженерного проєктування.

Можливість ітераційної роботи та реверсних потоків

AutomationML орієнтований на використання у ітераційних інженерних процесах, де модель системи багаторазово змінюється та передається між різними інструментами. Завдяки XML-представленню та об'єктній структурі CAEX модель містить ідентифіковані елементи, їхні атрибути, інтерфейси та зв'язки, що дозволяє відстежувати зміни та працювати з різними версіями моделі. Стандарт допускає використання атрибутів версій та метаданих, а також структуроване порівняння моделей на рівні елементів CAEX, що створює передумови для використання зовнішніх систем керування версіями або спеціалізованих інструментів для аналізу змін, наприклад git.

Практичну ітераційну роботу між інженерними інструментами детально розглянуто в рекомендаціях AutomationML Application Recommendation Automation Project Configuration (AR APC) [19]. У цьому документі описано типовий сценарій

обміну даними конфігурації автоматизаційного проєкту між ECAD-системами та інструментами програмування PLC. Рекомендації AR APC також передбачають двонаправлену організацію інженерного процесу.

Важливим аспектом такого підходу є можливість реверсного потоку даних. Під час налагодження або уточнення конфігурації системи керування зміни, виконані в одному інструменті, можуть бути повернуті назад у вихідну інженерну модель. Сам стандарт AutomationML не означає повного механізму синхронізації змін між різними системами, проте його структурна модель та можливість двонаправленого обміну створюють технічні передумови для реалізації таких сценаріїв. Практичну підтримку таких процесів реалізують окремі інженерні інструменти, зокрема Eplan, який підтримує як експорт, так і імпорт даних AutomationML, що дозволяє інтегрувати зміни, виконані в інших середовищах, назад у проєкт ECAD.

Використання AML в процесі експлуатації

AutomationML може використовуватися не лише на етапах інженерного проєктування, але і під час експлуатації систем керування. Така модель може містити дані про конфігурацію обладнання, топологію системи керування, взаємозв'язки сигналів і функціональних елементів.

Під час експлуатації ця інформація може використовуватися для підтримки технічного обслуговування, модернізації системи або інтеграції з іншими інформаційними системами підприємства. Зокрема, AML-модель може застосовуватися для передачі структурних даних до систем моніторингу, систем керування активами або до систем цифрових двійників. У таких випадках AML виступає як структурований опис системи, який дозволяє пов'язати фізичне обладнання з його інформаційним представленням.

Крім того, використання AutomationML у процесі експлуатації дозволяє зберігати узгодженість між інженерною документацією та фактичним станом системи. У разі внесення змін під час модернізації або налагодження конфігурації обладнання оновлені дані можуть бути повторно інтегровані до інженерної моделі.

Висновки. Аналіз стандарту показує, що завдяки використанню моделі CAEX та XML-представлення даних AutomationML дозволяє формалізовано описувати структуру технічних систем, їх компоненти, атрибути та взаємозв'язки. Це створює основу для інтеграції інженерних даних між різними середовищами проєктування та підтримує безперервність інформації протягом життєвого циклу системи керування.

Важливою перевагою стандарту є його відкритість і технологічна сумісність із широким спектром програмних інструментів. XML-основа дозволяє обробляти AML-моделі як спеціалізованими інженерними інструментами, так і універсальними програмними засобами, включно зі скриптовими рішеннями та табличними інструментами обробки даних. Наявність спеціалізованих редакторів і підтримки імпорту та експорту в окремих САЕ-системах, зокрема в Eplan, створює практичні передумови для використання AutomationML у процесах інженерного проєктування та інтеграції систем автоматизації.

Разом з тим проведений аналіз показує наявність низки проблем, що ускладнюють широке впровадження стандарту. Однією з основних є обмежена підтримка AutomationML у інженерних інструментах, а також різний рівень реалізації функцій імпорту та експорту в різних програмних середовищах. Крім того, стандарт не означає повністю формалізованих механізмів керування версіями моделей та синхронізації змін між інструментами, що ускладнює організацію ітераційних інженерних процесів і реалізацію реверсних потоків даних. У практичних проєктах ці задачі часто вирішуються на рівні конкретних інструментів або додаткових програмних рішень.

Проведене дослідження було зосереджене переважно на аналізі самого стандарту AutomationML та можливостей його використання у процесах інженерного проєктування систем керування. Подальші дослідження доцільно спрямувати на аналіз підтримки AutomationML у різних інженерних середовищах, а також на розроблення методів організації ітераційного обміну інженерними даними та інтеграції стандарту з іншими моделями цифрового представлення систем.

Список літератури:

1. Pupena O. PacFramework: Technological Solutions for PLC/PAC Programming—A Technical Report on Architecture, Principles, and Practical Implementation. Preprints, 2025. URL: <https://doi.org/10.20944/preprints202507.1180.v1>
2. Пупена О. М., Клименко В. В., Полупан В. В. Систематизація ключових стандартів для цифрових двійників у промисловому виробництві. Наукові вісті КПП. 2025. № 3. С. 60–73. DOI: <https://doi.org/10.20535/kpissn.2025.3.337306>

3. ISO 23247-3:2021 Automation systems and integration – Digital twin framework for manufacturing – Part 3: Digital representation of manufacturing elements. Geneva: ISO, 2021.
4. IEC 63278-1:2023 Asset Administration Shell for industrial applications – Part 1: Asset Administration Shell structure. Geneva: IEC, 2023.
5. Zhao J., Schamp M., Hoedt S., Aghezzaf E.-H., Cottyn J. AutomationML in Industry 4.0 Environment: A Systematic Literature Review. In: Service Orientation in Holonic and Multi-Agent Manufacturing. SOHOMA 2020. Lecture Notes in Computer Science. 2021. Vol. 12732. P. 162–169. DOI: https://doi.org/10.1007/978-3-662-62962-8_19
6. Siemens AG. Which EPLAN Data is Required for AML Data Exchange? Entry ID 109766653. Version 1.0. Siemens Industry Online Support, 2019. URL: [https://support.industry.siemens.com/cs/document/109766653/which-eplan-data-is-required-for-aml-data-exchange-with-step-7-\(tia-portal\)-?dti=0&lc=en-AO](https://support.industry.siemens.com/cs/document/109766653/which-eplan-data-is-required-for-aml-data-exchange-with-step-7-(tia-portal)-?dti=0&lc=en-AO) (дата звернення: 24.02.2026).
7. Breckle T., Kiesel M., Kiefer J., Beisheim N. The evolving digital factory – new chances for a consistent information flow. Procedia CIRP. 2019. Vol. 79. P. 251–256. DOI: <https://doi.org/10.1016/j.procir.2019.02.059>
8. Schroeder G., Steinmetz C., Pereira C., Espindola D. Digital Twin Data Modeling with AutomationML and a Communication Methodology for Data Exchange. IFAC-PapersOnLine. 2016. Vol. 49, No. 30. DOI: <https://doi.org/10.1016/j.ifacol.2016.11.115>
9. Peng G., Zhong H. Data Exchange of Digital Twins Based on AML in Space Science Experiment Equipment. IOP Conference Series: Materials Science and Engineering. 2020. Vol. 816. Article 012021. DOI: <https://doi.org/10.1088/1757-899X/816/1/012021>
10. Bigvand P., Fay A., Drath R., Carrion P. Concept and development of a semantic based data hub between process design and automation system engineering tools. Proceedings of the IEEE International Conference on Emerging Technologies and Factory Automation (ETFA). 2016. DOI: <https://doi.org/10.1109/ETFA.2016.7733734>
11. Czvetko T., Abonyi J. Data sharing in Industry 4.0 – AutomationML, B2MML and International Data Spaces-based solutions. Journal of Industrial Information Integration. 2023. Vol. 33. Article 100438. DOI: <https://doi.org/10.1016/j.jii.2023.100438>
12. Drath R., Rentschler M. Modeling and exchange of IO-Link configurations with AutomationML. Proceedings of the IEEE Conference on Automation Science and Engineering (CASE). 2018. P. 1530–1535. DOI: <https://doi.org/10.1109/COASE.2018.8560422>
13. Drath R., Rentschler M., Hoch M., Mueller M. Vendor-independent modeling and exchange of Fieldbus topologies with AutomationML. Proceedings of the IEEE International Conference on Emerging Technologies and Factory Automation (ETFA). 2018. DOI: <https://doi.org/10.1109/ETFA.2018.8502630>
14. ДСТУ EN IEC 62714-1:2022 Формат обміну інженерними даними для використання в системі промислової автоматизації. Мова розмітки автоматизації. Частина 1. Архітектура та загальні вимоги (EN IEC 62714-1:2018, IDT; IEC 62714-1:2018, IDT). Київ: ДП «УкрНДНЦ», 2022.
15. ДСТУ EN 62424:2022 Автоматичне управління технологічним процесом. Запити в діаграмах P&I та обмін даними між інструментальними засобами P&ID та PCE-CAE (EN 62424:2016, IDT; IEC 62424:2016, IDT). Київ: ДП «УкрНДНЦ», 2022.
16. Пупена О. CAEX – Computer Aided Engineering eXchange. Народний посібник «Автоматизація технологічних процесів та виробництв». GitHub. URL: https://github.com/asu-in-ua/atpv/blob/main/exportfmt/aml_caex/caex.md (дата звернення: 24.02.2026).
17. Drath R. (Ed.). AutomationML: A Practical Guide. Berlin: De Gruyter, 2021. ISBN 978-3-11-074622-8.
18. Drath R. (Ed.). AutomationML: The Industrial Cookbook. Berlin: De Gruyter, 2021. ISBN 978-3-11-074592-4.
19. AutomationML Association. AutomationML Application Recommendation – Automation Project Configuration (AR APC), Version 1.4. URL: <https://www.automationml.org/news/new-version-of-the-ar-apc-is-now-available/> (дата звернення: 24.02.2026).

Pupena O.M. USE OF AUTOMATIONML IN THE LIFECYCLE OF CONTROL SYSTEMS UNDER LIMITED RESOURCES

The article considers the possibilities of using AutomationML (AML) as a neutral format for exchanging engineering data between different tools in the lifecycle of control systems. The main principles of CAEX and AML are analyzed, which provide a formalized description of the structure of technical systems, their components, attributes, and relationships. It is shown that such a model allows engineering data to be represented as a unified information structure suitable for use by different engineering tools. The possibilities of applying AML at different stages of the lifecycle of control systems are considered, including engineering design, integration, and operation. Particular attention is paid to the support of the standard by software

tools. It is shown that due to the use of XML syntax, AML models can be processed both by specialized engineering environments and by general-purpose software tools, including scripting tools and spreadsheet-based data processing systems. The possibilities of using specialized AML editors as well as the integration of the standard with engineering design systems are also discussed. The article describes the possibilities of organizing iterative exchange of engineering data between different tools as well as implementing reverse information flows. It is shown that although the AutomationML standard creates technical prerequisites for such processes, the practical implementation of iterative synchronization largely depends on the support of corresponding mechanisms in specific engineering tools. Based on the analysis, the main advantages and limitations of using AutomationML in engineering design processes of automation systems are identified. Directions for further research are formulated, including the analysis of existing recommendations for data exchange between CAE and PLC tools, investigation of possibilities for representing AutomationML models in spreadsheet tools, and the development of prototype mappings of engineering models into environments used for master data management.

Keywords: AutomationML, CAEX, AML, engineering data exchange, control systems, CAE systems, lifecycle, system lifecycle, digital twin, digital thread.

Дата першого надходження статті до видання: 05.03.2026

Дата прийняття статті до друку після рецензування: 02.04.2026

Дата публікації (оприлюднення) статті 11.05.2026

УДК 004.412:519.237.5

DOI DOI DOI DOI <https://doi.org/10.32782/2663-5941/2026.2.1/38>**Пухалевич А.В.**<https://orcid.org/0000-0002-8827-3251>Національний університет кораблебудування
імені адмірала Макарова**Макарова Л.М.**<https://orcid.org/0000-0003-2903-3001>Національний університет кораблебудування
імені адмірала Макарова**Коваленко В.В.**<https://orcid.org/0009-0004-5841-0624>Національний університет кораблебудування
імені адмірала Макарова

МАТЕМАТИЧНА МОДЕЛЬ ДЛЯ ОЦІНЮВАННЯ ЯКОСТІ ЗАСТОСУНКІВ З ВІДКРИТИМ КОДОМ НА JAVA, ЩО СТВОРЮЮТЬСЯ З ВИКОРИСТАННЯМ ФРЕЙМВОРКУ HIBERNATE

Дослідження присвячене актуальній проблемі оцінювання якості застосунків з відкритим кодом на Java, що створюються з використанням фреймворку Hibernate. З огляду на розповсюдженість використання мови програмування Java та ORM-технологій, запропонований в даній роботі підхід до оцінювання якості на основі довірчих та прогнозних інтервалів нелінійних регресій для метрик RFC, CBO та WMC сприяє підвищенню достовірності такого оцінювання. Для побудови математичної моделі було зібрано та проаналізовано метрики 70 застосунків з відкритим кодом, де використовується фреймворк Hibernate.

Відповідно до тесту Мардіа, було виявлено перевищення критичного рівня за багатовимірною асиметрією, що свідчило про відсутність нормального розподілу емпіричних даних, і неможливість побудови лінійних моделей. Тому, було проведено нормалізацію даних за допомогою одновимірного перетворення Бокса-Кокса. Після цього тест Мардіа показав, що перетворені дані мають нормальний розподіл.

Завдяки перевірці квадрата відстані Махаланобіса для кожного рядка даних було встановлено відсутність викидів у загальній вибірці, а отже всі емпіричні дані були використанні в нормалізованому вигляді для побудови довірчих та прогнозних інтервалів нелінійних регресій з використанням методу нормалізуючих перетворень. Для цього спочатку було побудовано три лінійні регресії для нормалізованих даних, а потім, використовуючи зворотне нормалізуюче перетворення, з лінійних регресій отримано нелінійні регресії та довірчі і прогнозні інтервали нелінійних регресій. Всі три побудовані нелінійні регресії демонструють прийнятну точність прогнозування загальної вибірки проєктів за критеріями MMRE та PRED(0,25).

Об'єктом дослідження є процес оцінювання якості застосунків з відкритим кодом на Java, що створюються з використанням фреймворку Hibernate.

Предметом дослідження є математична модель для оцінювання якості застосунків з відкритим кодом на Java, що створюються з використанням фреймворку Hibernate, побудована на основі метрик RFC, CBO та WMC.

Ключові слова: оцінювання, якість, застосунок, програмна метрика, математична модель, довірчий інтервал, інтервал прогнозування, регресія, Java, фреймворк Hibernate, нормалізуюче перетворення.

Постановка проблеми. У сучасному світі Java продовжує утримувати провідні позиції серед мов програмування завдяки своїй інноваційності, адаптивності та продуктивності. Згідно зі звітом

компанії Azul, в 2025 році майже 70% із понад двох тисяч опитаних ІТ-фахівців у всьому світі зазначили, що більше половини їхніх застосунків створено за допомогою Java або працюють на Java

© Пухалевич А.В., Макарова Л.М., Коваленко В.В., 2026

Стаття поширюється на умовах ліцензії відкритого доступу CC BY 4.0



Virtual Machine (JVM) [1]. Враховуючи широку популярність Java, а також розповсюджене використання фреймворку Hibernate – одного з найпоширеніших інструментів для роботи з ORM (Object-Relational Mapping) – постає потреба у створенні ефективних засобів оцінювання якості програмного забезпечення, зокрема забезпечення з відкритим вихідним кодом.

Оцінювання якості програмного забезпечення є складним завданням інженерії програмного забезпечення, оскільки вимагає урахування багатьох факторів, що визначають надійність, ефективність та підтримуваність коду. Для розв'язання цієї задачі можуть використовуватися математичні моделі, побудовані на основі метрик RFC, CBO та WMC [2–4]. Одним із найрезультативніших підходів до оцінювання якості застосунків з відкритим кодом є метод, який передбачає обчислення меж довірчих та прогнозних інтервалів нелінійних регресій для показників RFC, CBO та WMC на рівні застосунку для тривимірних точок даних [5]. Під час порівняльного аналізу з існуючими моделями, було виявлено значні відмінності у діапазонах значень метрик RFC, CBO та WMC, що може призводити до зниження достовірності оцінювання якості застосунків з відкритим кодом на Java, що створюються з використанням фреймворку Hibernate [6]. Таким чином, побудова довірчих та прогнозних інтервалів нелінійних регресій для такого оцінювання дозволить удосконалити існуючі моделі, що сприятиме підвищенню якості розроблюваних застосунків з відкритим кодом.

Аналіз останніх досліджень і публікацій. Одним із способів оцінювання якості програмних застосунків є аналіз метрик програмного коду.

Серед найрозповсюдженіших метрик слід виокремити:

- RFC (Response For Class) демонструє кількість методів, які можуть бути викликані через даний клас; надто велике значення свідчить про перевантаження функціональністю.
- CBO (Coupling Between Object Classes) показує ступінь взаємозалежності між класами: чим вищий цей показник, тим складніше підтримувати та модифікувати код.
- WMC (Weighted Methods per Class) оцінює сумарну складність методів у класі, що дозволяє зробити висновки щодо можливості їх подальшого супроводу та надійності.

У сукупності ці метрики забезпечують цілісне уявлення про якість програмного забезпечення, допомагаючи виявити потенційні ризики, надмірну складність і напрямки для вдосконалення коду.

Для оцінювання якості застосовують лінійні регресійні моделі, що базуються на простих математичних зв'язках між метриками, але часто вони не здатні врахувати нелінійні залежності між метриками RFC, CBO та WMC і виявляються недостатньо точними для складних застосунків з відкритим кодом на Java. У сукупності ці показники формують основу для оцінювання якості коду, проте їх взаємодія має нелінійний характер, який неможливо коректно відобразити за допомогою звичайних лінійних підходів.

З метою підвищення достовірності оцінювання якості застосунків з відкритим кодом на Java, в [5] було запропоновано використовувати довірчі та прогнозні інтервали нелінійних регресій, які враховують взаємозалежність між трьома метриками коду:

- $RFC = f(CBO, WMC)$;
- $CBO = f(WMC, RFC)$;
- $WMC = f(RFC, CBO)$.

Завдяки цьому досягається комплексний опис структури програмного продукту, де кожна метрика виступає одночасно і залежною, і незалежною змінною, що дозволяє точніше виявляти вплив окремих характеристик на загальну якість застосунків.

Метод побудови нелінійних регресій з використанням нормалізуючих перетворень [7] складається з наступних етапів:

1. Перевірка нормальності розподілу даних тестом Мардіа.
2. Нормалізація даних – застосовується одновимірне нормалізуюче перетворення Бокса-Кокса для кожного параметру RFC, CBO та WMC.
3. Розрахунок відстані Махаланобіса для кожного застосунку, що дозволяє виявити аномальні значення (викиди) та видалити їх, якщо відстань перевищує критичний поріг.
4. Побудова по нормалізованим даним лінійних регресій, та довірчих і прогнозних інтервалів для кожної з метрик.
5. Побудова з використанням зворотного нормалізуючого перетворення нелінійних регресій, та довірчих і прогнозних інтервалів нелінійних регресій.

Класифікація рівня якості застосунку відбувається відносно побудованих інтервалів [5]: якщо фактичні значення всіх метрик застосунку розташовані в межах довірчого інтервалу – якість застосунку середня; якщо значення однієї з метрик виходить за верхню межу довірчого інтервалу для цієї метрики – якість застосунку низька; в інших випадках – якість висока.

Запропонований метод дозволяє більш глибоко аналізувати взаємозв'язки між метриками та надає інструмент для комплексного оцінювання архітектурної якості застосунків.

Постановка завдання. Метою дослідження є підвищення достовірності оцінювання якості застосунків з відкритим кодом на Java, що створюються з використанням фреймворку Hibernate.

Для досягнення поставленої мети потрібно вирішити наступні задачі:

- зібрати дані метрик RFC, CBO та WMC із застосунків з відкритим кодом на Java, що створюються з використанням фреймворку Hibernate. Виконати обробку даних, включно з видаленням викидів;

- побудувати три нелінійні регресії RFC(CBO, WMC), CBO(WMC, RFC) та WMC(RFC, CBO) із застосуванням одновимірного нормалізуючого перетворення Бокса-Кокса;

- отримати довірчі та прогнозні інтервали побудованих нелінійних регресій.

Виклад основного матеріалу. Для побудови математичної моделі для оцінювання якості застосунків з відкритим кодом на Java, що створюються з використанням фреймворку Hibernate, було зібрано дані з 70 проєктів аналогічної тематики, доступних на GitHub [8]. Необхідні для побудови моделей метрики RFC, CBO та WMC були отримані завдяки програмному інструменту з відкритим кодом СК [9]. Для кожного проєкту були розраховані середні арифметичні значення кожної з метрик відносно загальної кількості класів.

Набір застосунків та зібраних з них метрик RFC, CBO та WMC наведено в таблиці 1.

Щоб перевірити нормальність багатовимірних даних використовується тест Мардіа [10], що ґрунтується на визначенні параметрів багатовимірної асиметрії та багатовимірного ексцесу. Оцінка аси-

Таблиця 1

Набір застосунків та метрик RFC, CBO та WMC

№	Застосунок	RFC	CBO	WMC
1	adixchen/demo-restWS-spring-jersey-jpa2-hibernate	5,4000	6,5000	6,2000
2	ajanata/PretendYoureXyzzy	9,6290	6,5403	8,6210
3	akshu20791/spring-hibernate-security-react-ant-design-polling-app	6,1087	6,2391	5,9783
4	anakiou/multitenancy	3,0833	4,5833	3,0833
5	bcssp10/multi-tenant	4,4000	5,4000	2,5000
6	bdot/struts-hibernate	2,8571	2,5714	5,5714
7	benssj5/Projet-JEE-Gestion-des-Comptes-Spring-JPA-Hibernate	2,1250	5,5625	4,6250
8	BhartiKhankure/Vehicle_insurance_management_system	8,0000	6,5909	14,2727
9	brishi19791/HealthCare-WebApplication	5,0882	7,4118	7,1765
10	callistaenterprise/blog-multitenancy	5,6842	6,1053	2,6842
11	cassiomolin/jersey-jwt	7,6000	5,6000	5,2571
12	Cepr0/sb-multitenant-db-demo	6,1333	4,1333	4,2000
13	chrissssss/springboot-hibernatespatial-postgis-demo	6,6667	6,0000	13,0000
14	codecode25/HibernateDemo	7,3333	6,0000	4,8333
15	codercz/biubiu	7,5424	4,3390	10,5932
16	CoffeeLatte007/sharebook-jersey-jwt-spring-hibernate	5,7692	5,1923	6,3846
17	colinbut/sales-order-system	3,0476	9,0952	2,1905
18	cr2121/spring	1,6000	3,0000	5,6000
19	cyela/Spring-Web-App	8,5333	5,8667	14,4000
20	dadoonet/hsearch-es-demo	1,7500	4,2500	4,7500
21	daliasheasha/HibernateApp	7,0000	6,6667	11,6667
22	desperatecat/ShoppingPortal_Vue_Springboot_Project	6,3208	7,5849	8,2830
23	devdilip/todo-springboot-mysql-hibernate-rest	2,8000	4,4000	5,2000
24	diegoneuralo/maven	3,5000	8,1667	2,8333
25	discospiff/JavaFullStackEnterpriseWeb	6,4815	5,3704	6,4444
26	dmitrykhramov/Online-Bank-Simulator	4,8378	6,5135	7,0811
27	dobermai/Hibernate-Flyway-Integration	2,6667	6,0000	3,0000
28	F4bwDP6a6W/spring-mvc-REST	4,7500	3,8750	9,7500
29	foysal-mahmud/E-BookShop----Spring-boot	5,7931	6,4310	7,8621
30	frenmanoj/bookstore	3,1667	4,6667	6,6667
31	Fruzenshtein/security-spr	3,0556	5,3333	2,7778
32	IcedSoul/Shopping	6,1842	4,5526	8,6842

Продовження таблиці 1

№	Застосунок	RFC	CBO	WMC
33	jamesward/spring_hibernate_hstore_demo	6,4000	5,5000	3,7000
34	jasenkoh/spring-boot-jersey-hibernate	2,6667	5,5833	4,4167
35	Java-Techie-jt/Payment-Resource	3,1667	5,1667	5,0000
36	jjoe64/shiroexample	8,0000	5,9091	6,4545
37	Joryun/Shop	5,4648	5,3521	7,2254
38	Karszt/sda-zdjava124	10,0930	4,5814	4,2326
39	khozema-nullwala/online-shopping	7,8788	6,9091	10,8182
40	lastfm/musicbrainz-data	4,2368	6,4474	5,6842
41	licyun/SpringMVCgbook	7,1667	4,7500	9,5417
42	michaelisvy/hibernate-4-spring-3.1-samples	4,8000	7,7000	4,2000
43	MiteshPuthran/Shopping-Cart-Spring-MVC-Hibernate-Application	7,4444	7,9259	7,4815
44	mohamedYoussefi/eboutique-jpa-hibernate-spring	4,8333	7,0556	10,0000
45	MoonSulong/Ecommerce	4,8649	5,9459	5,8108
46	OKaluzny/spring-boot-rest-api-postgresql	4,2500	8,0000	2,2500
47	OKaluzny/springboot-rest-api-angularjs-https	5,5000	6,2000	5,0000
48	oregami/dropwizard-guice-jpa-seed	6,1163	6,2791	6,2326
49	pfac/demo-spring-data-jpa-hibernate-h2	4,6667	4,0000	6,5000
50	quarkiverse/quarkus-hibernate-types	2,9444	5,6111	5,1111
51	radoslaw-piwowarczyk/PetClinic-Spring-Thymeleaf-Hibernate	3,3333	4,8974	2,6923
52	rajbhatta/springboot_restapi_hibernate	2,2143	4,0000	4,0000
53	ralscha/eds-starter6-jpa	13,6222	7,5111	8,8000
54	RameshMF/springboot-thymeleaf-crud-pagination-sorting-webapp	3,8333	5,1667	4,6667
55	rampatra/springboot-hibernate	9,2857	8,0000	9,5714
56	rid17pawar/HospitalManagement	7,0189	8,0189	6,0943
57	royvanrijn/graalvm-native-microservice	6,0769	3,6923	5,5385
58	rutujaBacchuwar/MuzixAppHibernate	4,8571	5,8571	5,0000
59	scbushan05/book-api	3,6667	5,3333	4,0000
60	scottescue/dropwizard-entitymanager	12,6471	6,9412	11,0588
61	soulcode-acad/goservice	5,0488	5,9756	7,0244
62	sroysf/webapp-springmvc-jpa-hibernate	4,2857	5,8571	4,1429
63	stevegocoding/SpringMVC-SpringDataJpa-Hibernate-Demo	2,0000	3,4000	4,6000
64	swapnildipankar/java_jersey_spring_hibernate_maven	8,2000	5,9500	11,9500
65	ThxGodNovember/AprilDragonSpring	13,3056	5,1111	16,2500
66	tolitius/money-making-project	2,5000	2,6667	3,6667
67	vasylmalik/project-hibernate-1	6,6923	4,3846	7,3846
68	yannbriancon/spring-hibernate-query-utils	5,8750	2,6250	6,3750
69	ykameshrao/spring-mvc-angular-js-hibernate-bootstrap-java-single-page-jwt-auth-rest-api-webapp-framework	5,3235	5,3088	5,1029
70	ZhuXS/Spring-Shiro-Spark	7,4000	6,1400	9,0400

метрії виконується шляхом розрахунку відстаней Махаланобіса для кожного з проєктів [11]:

$$d_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{S}_N^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}), \quad (1)$$

де: $\mathbf{X}$ – вектор випадкових величин, $\mathbf{x} = \{Y, X_1, X_2\}^T$;
 $\bar{\mathbf{x}}$ – це вектор вибірових середніх, $\bar{\mathbf{x}} = \{\bar{Y}, \bar{X}_1, \bar{X}_2\}^T$;

N – кількість точок даних;

$\mathbf{S}_N^{-1}$ – обернена коваріаційна матриця, $\mathbf{S}_N^{-1} = \frac{1}{\mathbf{S}_N}$;

$\mathbf{S}_N$ – зміщена вибіркова коваріаційна матриця:

$$\mathbf{S}_N = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T. \quad (2)$$

Формули для обчислення багатовимірної асиметрії та ексцесу за методом Мардіа виглядають наступним чином [12]:

$$\beta_1 = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \left[(\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{S}_N^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) \right]^3, \quad (3)$$

$$\beta_2 = \frac{1}{N} \sum_{i=1}^N \left[(\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{S}_N^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) \right]^2, \quad (4)$$

де β_1 – багатовимірна асиметрія; β_2 – багатовимірний ексцес.

Щоб перевірити багатовимірну асиметрію β_1 (3) необхідно обчислити тестову статистику $\frac{1}{6}\beta_1$, яка наближається до розподілу χ^2 зі ступенями свободи $k(k+1)(k+2)/6$ та рівнем значущості $\alpha = 0,005$ [12].

Для багатовимірного ексцесу β_2 (4) застосується тестова статистика z_2 , що відповідає

$1 - \alpha$ -квантилю нормального розподілу із середнім значенням $k(k+2)$ і дисперсією $8k(k+2)/N$.

Перевірка нормальності виконується завдяки порівнянню значення $\frac{N}{6}\beta_1$ із квантилем $\chi^2_{k(k+1)(k+2)/6,\alpha}$ та значення β_2 із статистикою z_2 . Розподіл вважається не нормальним у тому випадку, коли хоча б одна з тестових статистик перевищує відповідний критичний рівень.

Внаслідок розрахунків були отримані наступні значення:

$$\beta_1 = 2,8425; \quad \frac{N}{6}\beta_1 = 33,1619; \quad \chi^2_{k(k+1)(k+2)/6,\alpha} = 25,1882; \\ \beta_2 = 17,4883; \quad z_2 = 18,3726.$$

Відповідно до наведених вище результатів, було виявлено перевищення критичного рівня за асиметрією: $33,1619 > 25,1882$, що свідчить про відсутність нормального розподілу даних. Тому, слід спочатку провести нормалізацію даних за допомогою нормалізуючого перетворення Бокса-Кокса:

$$Z_j = \Psi(\lambda_j) = \begin{cases} (X_j^{\lambda_j} - 1)/\lambda_j, & \text{якщо } \lambda_j \neq 0; \\ \ln(X_j), & \text{якщо } \lambda_j = 0, \end{cases} \quad (5)$$

де Z_j – це випадкова змінна з розподілом Гаусу (нормалізована змінна);

Ψ – вектор, $\Psi = \{\Psi_{RFC}, \Psi_{CBO}, \Psi_{WMC}\}^T$ перетворення $T = \Psi(P)$, $T = \{Z_{RFC}, Z_{CBO}, Z_{WMC}\}^T$ – гаусівський випадковий вектор, $P = \{RFC, CBO, WMC\}^T$ – негаусівський випадковий вектор;

λ_j – параметр перетворення Бокса-Кокса;

X_j – компонента вектору випадкових величин X , $X = \{RFC, CBO, WMC\}^T$;

$j = RFC, CBO, WMC$.

Розраховані значення параметрів перетворення (5) для метрик RFC, CBO та WMC дорівнюють $\lambda_{RFC} = 0,2485$, $\lambda_{CBO} = 1,0492$ та $\lambda_{WMC} = 0,0370$.

Після нормалізації даних за допомогою перетворення Бокса-Кокса з розрахованими значеннями параметрів, було отримано наступні результати:

$$\beta_1 = 1,2705; \quad \frac{N}{6}\beta_1 = 14,8224; \quad \chi^2_{k(k+1)(k+2)/6,\alpha} = 25,1882; \\ \beta_2 = 14,9362; \quad z_2 = 18,3726.$$

Відповідно до розрахованих значень асиметрії та ексцесу, внаслідок нормалізації дані мають нормальний розподіл та проходять тест Мардіа.

Було обчислено квадрат відстані Махаланобіса для нормалізованих даних для кожного спостереження за (1). Жодне із значень не перевищувало критичне значення, а отже можна стверджувати про відсутність викидів у нормалізованих вхідних даних.

Побудова нелінійних моделей виконується з використанням нормалізуючого перетворення (5).

Для реалізації цього етапу, першочергово необхідно побудувати три лінійні регресії для нормалізованих даних за загальною формулою:

$$\hat{Z}_y = b_0 + b_1 Z_{x_1} + b_2 Z_{x_2} + \varepsilon, \quad (6)$$

де $\hat{Z}_y$ – нормалізована залежна випадкова величина (RFC, CBO або WMC);

b_0 , b_1 та b_2 – оцінки параметрів лінійного рівняння регресії;

Z_{x_1} , Z_{x_2} – нормалізовані незалежні змінні X_1 та X_2 ;

μ – випадкова величина з нормальним розподілом.

Оцінки параметрів, отримані методом найменших квадратів [13] наступні.

Для регресії RFC(CBO, WMC):

$$b_0 = -0,2599, b_1 = 0,1469, b_2 = 0,8532.$$

Для регресії CBO(WMC, RFC):

$$b_0 = 4,1336, b_1 = -0,7971, b_2 = 1,1138.$$

Для регресії WMC(RFC, CBO):

$$b_0 = 1,1486, b_1 = 0,4837, b_2 = -0,0596.$$

Розподіл випадкової величини $\varepsilon \in$ нормальним для цих моделей.

Шляхом зворотного нормалізуючого перетворення з лінійних моделей (6) отримаємо нелінійні регресійні моделі:

$$\hat{r}_{RFC} = \left(\left[-0,2599 + 0,1469 \cdot \frac{X_{CBO}^{1,0492} - 1}{1,0492} + 0,8532 \cdot \frac{X_{WMC}^{0,0369} - 1}{0,0369} + \varepsilon \right] \cdot 0,2485 + 1 \right)^{\frac{1}{0,2485}} \\ \hat{r}_{CBO} = \left(\left[4,1336 - 0,7971 \cdot \frac{X_{WMC}^{0,0369} - 1}{0,0369} + 1,1138 \cdot \frac{X_{RFC}^{0,2485} - 1}{0,2485} + \varepsilon \right] \cdot 1,0492 + 1 \right)^{\frac{1}{1,0492}} \\ \hat{r}_{WMC} = \left(\left[1,1486 + 0,4837 \cdot \frac{X_{RFC}^{0,2485} - 1}{0,2485} - 0,0596 \cdot \frac{X_{CBO}^{1,0492} - 1}{1,0492} + \varepsilon \right] \cdot 0,0369 + 1 \right)^{\frac{1}{0,0369}} \quad (7)$$

Моделі побудовано в межах метрик RFC[1,6000; 13,6222], CBO[2,5714; 9,0952] та WMC[2,1905; 16,2500].

Розрахунок показників якості для побудованих нелінійних регресійних моделей (7) виконуємо за допомогою критеріїв MMRE та PRED(0,25).

Для регресії RFC(CBO, WMC):

$$MMRE = 0,2777, PRED(0,25) = 0,5571.$$

Для регресії CBO(WMC, RFC):

$$MMRE = 0,2011, PRED(0,25) = 0,7571.$$

Для регресії WMC(RFC, CBO):

$$MMRE = 0,3085, PRED(0,25) = 0,5714.$$

Відповідно до отриманих результатів було визначено, що усі три побудовані нелінійні моделі демонструють прийнятну точність прогнозування загальної вибірки проєктів за критеріями MMRE та PRED(0,25).

Формули рівнянь границь довірчих та прогнозних інтервалів побудованих нелінійних регресій згідно [5]:

$$\Psi_Y^{-1} \left(\hat{Z}_Y \pm t_{\alpha/2, v} S_{Z_Y} \left\{ \frac{1}{N} + (\mathbf{z}_X^+)^T \mathbf{S}_Z^{-1} (\mathbf{z}_X^+) \right\}^{\frac{1}{2}} \right), \quad (8)$$

$$\Psi_Y^{-1} \left(\hat{Z}_Y \pm t_{\alpha/2, v} S_{Z_Y} \left\{ 1 + \frac{1}{N} + (\mathbf{z}_X^+)^T \mathbf{S}_Z^{-1} (\mathbf{z}_X^+) \right\}^{\frac{1}{2}} \right), \quad (9)$$

де Ψ_Y – перетворення для нормалізації залежної змінної Y ;

$\hat{Z}_Y$ – оцінка відповідно до лінійного рівняння регресії (6) для нормалізованих даних;

$t_{\alpha/2, v}$ – квантиль розподілу Стюдента з рівнем значущості $\alpha/2$ та $v=N-3$ ступенями вільності;

$S_{Z_Y}^2 = \frac{1}{v} \sum_{i=1}^N (Z_{Y_i} - \bar{Z}_Y)^2$ – залишкова дисперсія, $v=N-k-1$;

$\mathbf{z}_X^+$ – матриця нормалізованих регресорів зі значеннями $Z_{X_{1j}} - \bar{Z}_{X_1}, Z_{X_{2j}} - \bar{Z}_{X_2}, \bar{Z}_{X_j} = \frac{1}{N} \sum_{i=1}^N Z_{X_{ij}}, j = 1, 2$;

$$\mathbf{S}_Z = \begin{pmatrix} S_{Z_1 Z_1} & S_{Z_1 Z_2} \\ S_{Z_2 Z_1} & S_{Z_2 Z_2} \end{pmatrix}, \text{ де } S_{Z_q Z_r} = \sum_{i=1}^N [Z_{q_i} - \bar{Z}_q][Z_{r_i} - \bar{Z}_r], q, r = 1, 2.$$

Для регресії RFC(CBO, WMC) параметри рівнянь (8) та (9) є наступними:

$$S_{Z_Y} = 0.2638, \bar{Z}_1 = 4.9326, \bar{Z}_2 = 1.8380, \mathbf{S}_Z^{-1} = \begin{pmatrix} 0.0062 & -0.0011 \\ -0.0011 & 0.0586 \end{pmatrix}.$$

Для регресії CBO(WMC, RFC) параметри рівнянь (8) та (9) є наступними:

$$S_{Z_Y} = 2.0007, \bar{Z}_1 = 2.0328, \bar{Z}_2 = 1.838, \mathbf{S}_Z^{-1} = \begin{pmatrix} 0.0473 & -0.0416 \\ -0.0416 & 0.0951 \end{pmatrix}.$$

Для регресії WMC(RFC, CBO) параметри рівнянь (8) та (9) є наступними:

$$S_{Z_Y} = 0.1496, \bar{Z}_1 = 2.0328, \bar{Z}_2 = 4.9326, \mathbf{S}_Z^{-1} = \begin{pmatrix} 0.0332 & -0.0054 \\ -0.0054 & 0.0071 \end{pmatrix}.$$

Оцінювання якості з використанням побудованої математичної моделі розглянемо на прикладі застосунку MiteshPuthran/Shopping-Cart-Spring-MVC-Hibernate-Application.

Для регресії RFC(CBO, WMC) довірчий інтервал [6,4501; 8,6328], прогнозний [3,7325; 13,5346], показник якості за цією метрикою – середній, тому що фактичне значення RFC 7,4444 знаходиться в середині меж довірчого інтервалу.

Для регресії CBO(WMC, RFC) довірчий інтервал [5,6699; 6,4659], прогнозний [3,4209; 8,6586], показник якості за цією метрикою – низький, тому що фактичне значення CBO 7,9259 більше верхньої границі довірчого інтервалу.

Для регресії WMC(RFC, CBO) довірчий інтервал [5,6469; 7,8745], прогнозний [3,1551; 13,8262], показник якості за цією метрикою – середній, тому що фактичне значення WMC 7,4815 знаходиться в середині меж довірчого інтервалу.

Оскільки фактичне значення однієї з метрик (CBO) застосунку MiteshPuthran/Shopping-Cart-Spring-MVC-Hibernate-Application виходить за верхню межу довірчого інтервалу цієї метрики – якості застосунку низька.

Висновки. У роботі побудовано математичну модель для оцінювання якості застосунків з відкритим кодом на Java, що створюються з використанням фреймворку Hibernate. При побудові було проведено нормалізацію даних за допомогою одновимірного перетворення Бокса-Кокса. Завдяки перевірці квадрата відстані Махаланобіса для кожного рядка даних було встановлено відсутність викидів в даних, а отже всі емпіричні дані були використанні для побудови нелінійних регресій для метрик RFC, CBO та WMC, та довірчих і прогнозних інтервалів цих регресій з використанням методу нормалізуючих перетворень. Всі три побудовані нелінійні регресії демонструють прийнятну точність прогнозування загальної вибірки проєктів за критеріями MMRE та PRED(0,25).

В подальшому планується застосування багатовимірних нормалізуючих перетворень для покращення якості побудованої математичної моделі.

Список літератури:

1. Sellers S. 2025 State of Java Survey & Report. URL: <https://securityboulevard.com/2025/01/state-of-java-survey-confirms-majority-of-enterprise-apps-are-built-on-java> (дата звернення: 20.10.2025).
2. Bagnulo M., Claise B., Eardley P., Morton A., Akhter A. RFC 8911. URL: <https://www.rfc-editor.org/rfc/rfc8911.html> (дата звернення: 21.01.2026).
3. Shyam R., Chris F. Metrics suite for object-oriented design. URL: <https://www.eso.org/~tcsmgr/oowg-forum/TechMeetings/Articles/OOMetrics.pdf> (дата звернення: 21.01.2026).
4. Aanchal S., Kumar. Metrics for Software Components in Object Oriented Environments: A Survey // *International Journal of Scientific Research in Computer Science and Engineering*. 2013. ISSN 2320-7639. URL: https://isroset.org/pub_paper/IJSRCSE/ISROSET-IJSRCSE-00043.pdf (дата звернення: 21.01.2026).
5. Prykhodko S. Evaluating Quality of Software Systems by the Confidence and Prediction Intervals of Regressions for RFC, CBO and WMC Metrics // *WSEAS Transactions on Systems*. 2024. Vol. 23. P. 322–330. DOI: <https://doi.org/10.37394/23202.2024.23.36>.
6. Prykhodko S., Makarova L., Pukhalevych A. Statistical Analysis of the Three-Dimensional Data of Software Metrics RFC, CBO, and WMC that are not Normally Distributed // *International Conference on Applied Innovations in IT (ICAIIIT)*. 2025. P. 127–131. DOI: <https://doi.org/10.25673/119224>.

7. Приходько С.Б. Метод побудови нелінійних рівнянь регресії на основі нормалізуючих перетворень // Тези доповідей міждерж. наук.-методич. конф. “Проблеми математичного моделювання” (Дніпродзержинськ, 13-15 червня 2012 р.). С. 31-33.
8. GitHub. URL: <https://github.com/> (дата звернення: 21.01.2026).
9. Aniche M. Java code metrics calculator (CK). URL: <https://github.com/mauricioaniche/ck> (дата звернення: 21.01.2026).
10. Mardia K. Measures of multivariate skewness and kurtosis with applications // *Biometrika*. 1970. Vol. 57. P. 519–530. DOI: <https://doi.org/10.1093/biomet/57.3.519>.
11. Mahalanobis Distance. URL: <https://www.machinelearningplus.com/statistics/mahalanobis-distance/> (дата звернення: 21.01.2026).
12. Wulandari D., Sutrisno S., Nirwana M. Mardia's Skewness and Kurtosis for Assessing Normality Assumption in Multivariate Regression // *Enthusiastic: International Journal of Applied Statistics and Data Science*. 2021. P. 2–4. DOI: <https://doi.org/10.20885/enthusiastic.vol1.iss1.art1>.
13. Afifi A., Azen S. Statistical analysis: a computer-oriented approach. New York; London: Academic Press, 1972.

Pukhalevych A.V., Makarova L.M., Kovalenko V.V. A MATHEMATICAL MODEL FOR ASSESSING THE QUALITY OF OPEN SOURCE JAVA APPLICATIONS CREATED USING THE HIBERNATE FRAMEWORK

The study is dedicated to the topical issue of quality assessment for open-source Java applications developed using the Hibernate framework. Given the widespread use of the Java programming language and ORM technologies, the approach proposed in this work—based on confidence and prediction intervals of nonlinear regressions for the RFC, CBO, and WMC metrics – contributes to increasing the reliability of such assessments. To construct the mathematical model, metrics from 70 open-source applications utilizing the Hibernate framework were collected and analyzed.

According to Mardia's test, it was found that the multivariate skewness exceeded the critical level, indicating that the empirical data did not follow a normal distribution, making the construction of linear models impossible. Therefore, the data were normalized using univariate Box-Cox transformation. Following this, Mardia's test confirmed that the transformed data followed a normal distribution.

By verifying the squared Mahalanobis distance for each data row, it was determined that there were no outliers in the overall sample. Therefore, all empirical data were used in normalized form to construct confidence and prediction intervals for nonlinear regressions using the method of normalizing transformations. For this purpose, three linear regressions were first constructed for the normalized data; then, using the inverse normalizing transformation, nonlinear regressions and their corresponding confidence and prediction intervals were derived. All three nonlinear regressions demonstrate acceptable prediction accuracy for the project sample based on the MMRE and PRED(0.25) criteria.

The object of the study is the quality assessment process for open-source Java applications built with the Hibernate framework.

The subject of the study is mathematical model for assessment the quality of open-source Java applications created using the Hibernate framework, constructed based on the RFC, CBO, and WMC metrics.

Keywords: *assessment, quality, application, software metric, mathematical model, confidence interval, prediction interval, regression, Java, Hibernate framework, normalizing transformation.*

Дата першого надходження статті до видання: 27.01.2026

Дата прийняття статті до друку після рецензування: 20.02.2026

Дата публікації (оприлюднення) статті 11.05.2026

Рубель Ю.Б.

<https://orcid.org/0009-0003-1891-2446>

Національний університет «Львівська політехніка»

Грицюк Ю.І.

<https://orcid.org/0000-0001-8183-3466>

Національний університет «Львівська політехніка»

ГЕНЕРУВАННЯ СИНТЕТИЧНОГО БАЛІСТИЧНОГО НАБОРУ ДАНИХ НА ПІДСТАВІ ЧИСЕЛЬНОГО МОДЕЛЮВАННЯ ТА РАНДОМІЗАЦІЇ ДОМЕНІВ ДЛЯ СИСТЕМИ УПРАВЛІННЯ АРТИЛЕРІЙСЬКИМ ВОГНЕМ

Наведено результати розроблення методики генерування синтетичного балістичного набору даних на підставі чисельного моделювання польоту снаряда та рандомізації доменів для навчання нейромережових компонентів автоматизованих систем управління вогнем артилерії. Йдеться про формування якісних навчальних наборів даних без проведення ресурсомістких натурних випробувань із одночасним забезпеченням фізичної коректності та репрезентативності згенерованих даних. Запропонована методика реалізована як п'ятиетапний конвеєр, що послідовно поєднує формування систематичної сітки кутів наведення гармати, рандомізацію фізичних параметрів стрільби згідно з принципом Domain Randomization, чисельне моделювання польоту снаряда за моделлю чотирьох ступенів свободи (4-DOF) відповідно до стандарту NATO STANAG 4355, багаторівневу фільтрацію та консолідацію результатів. У роботі адаптовано фреймворк СТТ (англ. Classical Test Theory) для формалізованого багатовимірної оцінювання якості синтетичного набору даних у завданнях регресійного балістичного прогнозування за дев'ятьма метриками у трьох вимірах: достовірність, відтворюваність та відповідність. Сформований набір даних обсягом 67102 записів для 12 артилерійських систем калібру 155 мм охоплює повний практичний діапазон стрільби. Верифікацію якості синтетичного набору даних виконано через їхній аналіз за адаптованими метриками СТТ та глибокого навчання нейронної мережі з розділними моделями для режимів низької та високої стрільби. Зведена оцінка якості синтетичного набору даних за фреймворком СТТ становила 0,875. Навчання нейронної мережі продемонструвало MAE у діапазоні 2,2–2,9° для кута підвищення гармати та 2,5–2,9° для азимута стрільби при $R^2 = 0,86$ – $0,87$, а класифікаційна точність стрільби становила 93,3–96,7 %, що підтверджує придатність згенерованого набору даних для навчання нейронних моделей прогнозування балістичних рішень.

Ключові слова: синтетичний набір даних, рандомізація параметрів симуляції, чисельне моделювання польоту снаряда, управління вогнем артилерії, нейронна мережа, фреймворк СТТ, модель чотирьох ступенів свободи, стандарт NATO STANAG 4355.

Постановка проблеми. Сучасні збройні конфлікти ставлять дедалі вищі вимоги до точності та швидкості виконання артилерійського вогню. Автоматизовані системи управління вогнем (АСУВ) відіграють ключову роль у підвищенні бойової ефективності артилерійських підрозділів, забезпечуючи розрахунок балістичних рішень у реальному часі [1]. Традиційні методи розрахунку траєкторій стрільби артилерійських установок спираються на балістичні таблиці й аналітичні моделі, однак їх точність істотно падає через випадкові метеорологічні умови, знос

ствола, нестабільність характеристик порохових зарядів та складний рельєф місцевості [12]. Це робить застосування методів машинного навчання моделей стрільби, зокрема – на підставі глибоких нейронних мереж, перспективним дослідним підходом до вирішення завдань наведення і коригування артилерійських систем.

Результати останніх досліджень підтверджують, що нейронні мережі здатні ефективно вирішувати завдання моделювання зовнішньої балістики, демонструючи високу точність прогнозування траєкторій руху снарядів за різних умов стрільби [2].

Проте, широке впровадження таких систем стримується обмеженням доступом до навчальних наборів даних з відомими результатами стрільби, отримання яких через реальні випробування є дорогим і технічно складним. Тому формування балістичних наборів даних засобами чисельного моделювання польоту снаряда є одним із найбільш практично доцільних підходів до забезпечення процедури навчання нейронних мереж. У дослідженнях з передачі знань між симуляцією та реальністю обґрунтовано підхід рандомізації доменів DR (англ. *Domain Randomization*) [3], ефективність якого підтверджено і в суміжних завданнях прогнозування траєкторій руху безпілотних літальних апаратів (БПЛА) [4]. Водночас, генерування синтетичного набору даних є тільки першим кроком до його формування, не менш важливим є оцінювання якості згенерованого набору даних, для чого у роботі [5] запропоновано фреймворк СТТ (англ. *Classical Test Theory*) із дев'ятьма метриками у трьох вимірах, структура якого допускає адаптацію до вирішення регресійних завдань балістичного прогнозування.

Отже, у науковій літературі залишається невирішеною проблема розроблення комплексної методики формування та верифікації синтетичних балістичних наборів даних. Наявні роботи або зосереджуються винятково на симуляції набору даних без оцінювання їх якості, або пропонують метрики якості для інших предметних областей без адаптації до фізично орієнтованих регресійних завдань. Тому пропонується підхід, що поєднує DR-конвеєр генерування синтетичного балістичного набору даних з адаптованим СТТ-фреймворком для оцінювання його якості. Водночас, під синтетичним балістичним набором даних (англ. *Synthetic Ballistic Dataset*) слід розуміти штучно згенерований набір даних, що відтворює статистичні закономірності та фізичні взаємозв'язки артилерійської стрільби засобами чисельного моделювання – без проведення натурних випробувань та без використання конфіденційних військових даних.

Аналіз останніх досліджень і публікацій. Дослідження в галузі нейромережевих методів для завдань симуляції зовнішньої балістики охоплюють широкий спектр підходів – від класичних архітектур прямого поширення до складних гібридних моделей. Ще в ранніх роботах [6] показано, що нейронні мережі здатні апроксимувати балістичні функції для артилерійських ракет із достатньою для практичного застосування точністю, що поклало початок систематичному

вивченню цього напрямку. У подальших дослідженнях [7] запропоновано нейромережевий пропагатор траєкторій руху артилерійських снарядів, де нейронна мережа замінює традиційний чисельний інтегратор і забезпечує значне пришвидшення розрахунків за прийнятною точністю. Для прогнозування траєкторій руху балістичних ракет у роботі [8] розроблено алгоритм прогнозування траєкторії руху на основі даних та знань, що поєднує мережу зворотного поширення з фізичною моделлю гравітаційного розвороту цілі. Спільним обмеженням зазначених підходів є залежність від великих навчальних вибірок із відомими результатами стрільби, отримання яких у реальних умовах є складним і ресурсомістким процесом.

Проблема управління вогнем артилерії та архітектурних рішень для відповідних програмних систем досліджена в низці робіт. Наприклад, у роботі [9] запропоновано підхід до інтелектуального коригування стрільби в АСУВ на основі нейронної мережі, що враховує похибки попередніх пострілів для коригування наступних. В дослідженні [10] розроблено модульну архітектуру для гібридної системи управління вогнем, що містить модуль генерування набору даних, балістичний симулятор, нейронну мережу прогнозування кутів наведення гармати та модуль управління вогнем. Алгоритми управління вогнем для модульних артилерійських систем морського базування наведено в роботі [11], де обґрунтовано вимоги до точності та швидкодії програмного забезпечення систем наведення артилерійського вогню. Загальні принципи проєктування артилерійських систем для досягнення точності стрільби систематизовано в роботі [12], а в дослідженні [13] запропоновано метод швидкого коригування кутів наведення гармати на основі модифікованої моделі точкової маси снаряда з матрицею передачі в дискретному часі. Застосування глибокого навчання нейронної мережі для автоматизації завдань розпізнавання та класифікації цілей у системах управління вогнем продемонстровано в роботі [14].

Математична основа балістичного симулювання траєкторій руху снарядів спирається на стандарт NATO STANAG 4355, валідацію якого для систем управління вогнем здійснено в дослідженні [15]. Підтверджено коректність модифікованої моделі точкової маси з чотирма ступенями свободи (4-DOF) при відповідній параметризації аеродинамічних коефіцієнтів, що є основою для її подальшого застосування в системах більших калібрів. У нейронних моделях доцільним є засто-

сування механізму самоуваги [16], який дає змогу моделі динамічно зважувати важливість різних вхідних ознак – метеорологічних і технічних залежно від контексту стрільби.

Оцінювання невизначеності результатів балістичного моделювання польоту снаряда є окремим дослідним напрямом. У роботах [17, 18] порівняно методи квантифікації невизначеності – розкладання поліноміального хаосу РСЕ (англ. *Polynomial Chaos Expansion*) та Монте-Карло, зокрема – з розпаралелюванням обчислень на графічних процесорах, де встановлено, що РСЕ забезпечує вищу обчислювальну ефективність за порівнянної точності при невеликій кількості стохастичних змінних. В роботах [19, 20] досліджено вплив невизначеностей на результати дослідження балістичного моделювання польоту снаряда: запропоновано гібридний підхід до аналізу артилерійської динаміки зі стохастичними та інтервальними параметрами, а також показано, що стохастичні відхилення метеорологічних параметрів і початкової швидкості руху снаряду є основними джерелами невизначеності й мають враховуватись при формуванні балістичного набору даних.

Отже, проведений аналіз останніх досліджень і публікацій підтверджує, що вони охоплюють окремі аспекти задачі – нейромережеве прогнозування траєкторій руху снарядів, архітектуру АСУВ, квантифікацію невизначеності та балістичне моделювання польоту снаряда стрільби, однак у таких дослідженнях бракує методики, яка б поєднувала конвеєр генерування синтетичних наборів даних на підставі рандомізації доменів із верифікацією їх якості за багатовимірними метриками.

Постановка завдання. Об'єкт дослідження: генерування синтетичного балістичного набору даних для навчання нейронних мереж у системах управління артилерійським вогнем.

Предмет дослідження: методи і засоби генерування синтетичного балістичного набору даних на підставі чисельного моделювання польоту снаряда та рандомізації їх параметрів, що дасть змогу використати адаптовані метрики фреймворку СТТ для оцінювання якості згенерованого набору.

Мета роботи: розробити методику генерування синтетичного балістичного набору даних на підставі чисельного моделювання польоту снаряда та рандомізації його параметрів для навчання нейромережевих компонентів систем управління артилерійським вогнем, яка забезпечить верифікацію якості такого набору за адаптованими метриками фреймворку СТТ.

Для досягнення поставленої мети визначено такі завдання дослідження:

1. Розробити методику генерування синтетичного балістичного набору даних із застосуванням принципу систематичної рандомізації фізичних параметрів стрільби на основі чисельного моделювання польоту снаряда за моделлю 4-DOF, що дасть змогу отримати репрезентативний набір даних без проведення натурних артилерійських випробувань.

2. Обґрунтувати вибір параметрів рандомізації набору даних та їх діапазони, що забезпечують відтворення реальних умов артилерійської стрільби в синтетичному наборі даних, що дасть можливість охопити весь практично значущий операційний простір умов стрільби та підвищити здатність нейронної мережі до узагальнення на реальних даних.

3. Адаптувати метрики фреймворку СТТ для оцінювання якості згенерованого синтетичного набору даних, що забезпечить формалізовану багатовимірну верифікацію його фізичної коректності, статистичної стабільності та відповідності вимогам навчання нейромережевих компонентів АСУВ.

4. Верифікувати якість сформованого набору даних шляхом оцінювання результатів навчання нейронної мережі для прогнозування кутів наведення гармати, що допоможе підтвердити практичну цінність розробленої методики генерування даних для застосування у системах управління артилерійським вогнем.

Виклад основного матеріалу

Архітектура методики генерування синтетичного балістичного набору даних. Методику генерування синтетичного балістичного набору даних реалізовано як послідовний процес, що поєднує реальні метеорологічні дані з принципом Domain Randomization [3] та фізично обґрунтованим чисельним моделюванням польоту снаряда. Необхідність формування якісних навчальних наборів даних зумовлена потребою АСУВ у швидкому обчисленні балістичних рішень в реальному часі [1], що вимагає застосування нейромережевих компонентів із відповідними навчальними вибірками. Архітектуру конвеєра генерування синтетичних наборів даних розроблено відповідно до модульної структури гібридної системи управління вогнем [10], де модуль генерування набору даних визначено як окремий компонент системи.

Передумовою для початку процесу генерування балістичного набору даних є два незалеж-

них джерела вхідних даних. Масив метеопрофілів формується на основі архівних погодинних спостережень сервісу Open-Meteo за 2024 рік для 24 обласних центрів України і налічує близько 210000 записів із природним кореляційним зв'язком між атмосферними змінними. База даних балістичних параметрів містить характеристики 12 артилерійських систем калібру 155 мм із двома типами снарядів і формує 24 унікальні конфігурації для моделювання польоту снаряда.

На першому етапі роботи конвеєра формується систематична сітка кутів наведення гармати, що охоплює повний практичний діапазон стрільби: 70 дискретних значень кутів підвищення гармати від 1° до 85° та 60 рівномірних азимутальних напрямків стрільби від 0° до 360° , що в сукупності дає 4200 унікальних комбінацій. Вибір діапазону кутів підвищення гармати обґрунтовується вимогами до точності артилерії [12].

На другому етапі роботи конвеєра для кожної комбінації кутів наведення гармати генерується 30 сценаріїв умов стрільби з випадковим вибором метеорологічного профілю та конфігурації артилерійської системи і снаряда із застосуванням стохастичних варіацій відповідно до принципу Domain Randomization [3].

На третьому етапі роботи конвеєра кожна із 126000 згенерованих конфігурацій передається до балістичного симулятора, реалізованого відповідно до стандарту NATO STANAG 4355 [15]. Для удосконалення обчислювального процесу реалізовано механізм паралельних обчислень з асинхронним розподілом завдань між процесорними ядрами [18].

На четвертому етапі роботи конвеєра результати симуляцій проходять багаторівневу фільтрацію: видаляються записи з відсутніми значеннями та записи з координатами влучання поза межами фізично реалістичних діапазонів, визначених на підставі аналізу впливу невизначеностей балістичної моделі [20]. А на п'ятому фінальному етапі роботи конвеєра відфільтровані результати консоліднуються у фінальний набір даних у форматі CSV, що містить 17 вхідних ознак і дві цільові змінні – координати точки падіння снаряда (x, z).

Балістичний симулятор та рандомізація параметрів набору даних. Балістичний симулятор реалізовано на базі модифікованої моделі точкової маси з чотирма ступенями свободи (4-DOF) відповідно до стандарту NATO STANAG 4355 [15]. Модель описує тривимірний рух снаряда з урахуванням аеродинамічного опору з Мах-залежним коефіцієнтом, що інтерполюється між табличними

вузловими точками окремо для кожного типу снаряда, атмосферної щільності та швидкості звуку за моделлю ISA 1976. Також враховується відхилення траєкторії руху снаряда внаслідок ефекту Коріоліса залежно від географічної широти й азимута стрільби, поперечної сили Магнуса на нарізному снаряді зі швидкістю обертання 260 об/с, а також логарифмічного профілю вітру за висотою над поверхнею землі. Чисельне інтегрування системи диференціальних рівнянь шостого порядку виконується методом Рунге-Кутта 4–5 порядку з адаптивним кроком [21, 22]. Критерієм зупинки є досягнення снарядом заданої висоти цілі, а результатом кожної симуляції – координати точки падіння снаряда (x, z) у декартовій системі координат відносно позиції гармати.

Метеорологічні параметри (температура, тиск, вологість і вітер) не синтезуються штучно, а відбираються з масиву реальних погодинних записів, при цьому висота позиції гармати відповідає реальній висоті рельєфу місцевості в точці відбору метеоданих, що зберігає природний кореляційний зв'язок між атмосферними змінними [19].

Початкова швидкість руху снаряда варіюється в межах $\pm 5\%$ від номінального значення за рівномірним розподілом, охоплюючи варіативність, зумовлену зносом ствола та відмінностями порохових зарядів [20, 23]. Температура порохового заряду визначається відносно температури повітря зі зміщенням від -5 до $+10^\circ\text{C}$. Висота цілі рандомізується за трьома сценаріями рельєфу місцевості: рівнинна (± 200 м, ймовірність 60 %), узвишшя ($+200\dots+800$ м, 25 %) та низина ($-800\dots-200$ м, 15 %). Ефект Коріоліса активується з ймовірністю 70 %, а ефект Магнуса з ймовірністю 30 %. Тип снаряда обирається рівномірно між двома конфігураціями: Standard-HE та ERFB-BB. Вибір методу рандомізації узгоджується з результатами порівняльного аналізу методів квантифікації невизначеності [17].

Оцінювання якості згенерованого синтетичного набору даних. Після виконання процедури генерування набору даних і його фільтрації, фінальний обсяг балістичного набору даних становив 67102 записи, тобто містив 53,3 % від початкового його обсягу зі збалансованим розподілом між режимами стрільби: 58,4 % – для низького ($1^\circ\text{--}45^\circ$) та 41,6 % – для високого ($45^\circ\text{--}85^\circ$). На рис. 1, подано гістограми розподілу дев'яти ключових параметрів синтетичного балістичного набору даних.

Статистичний аналіз параметрів балістичного набору даних підтверджує їх репрезентативність,

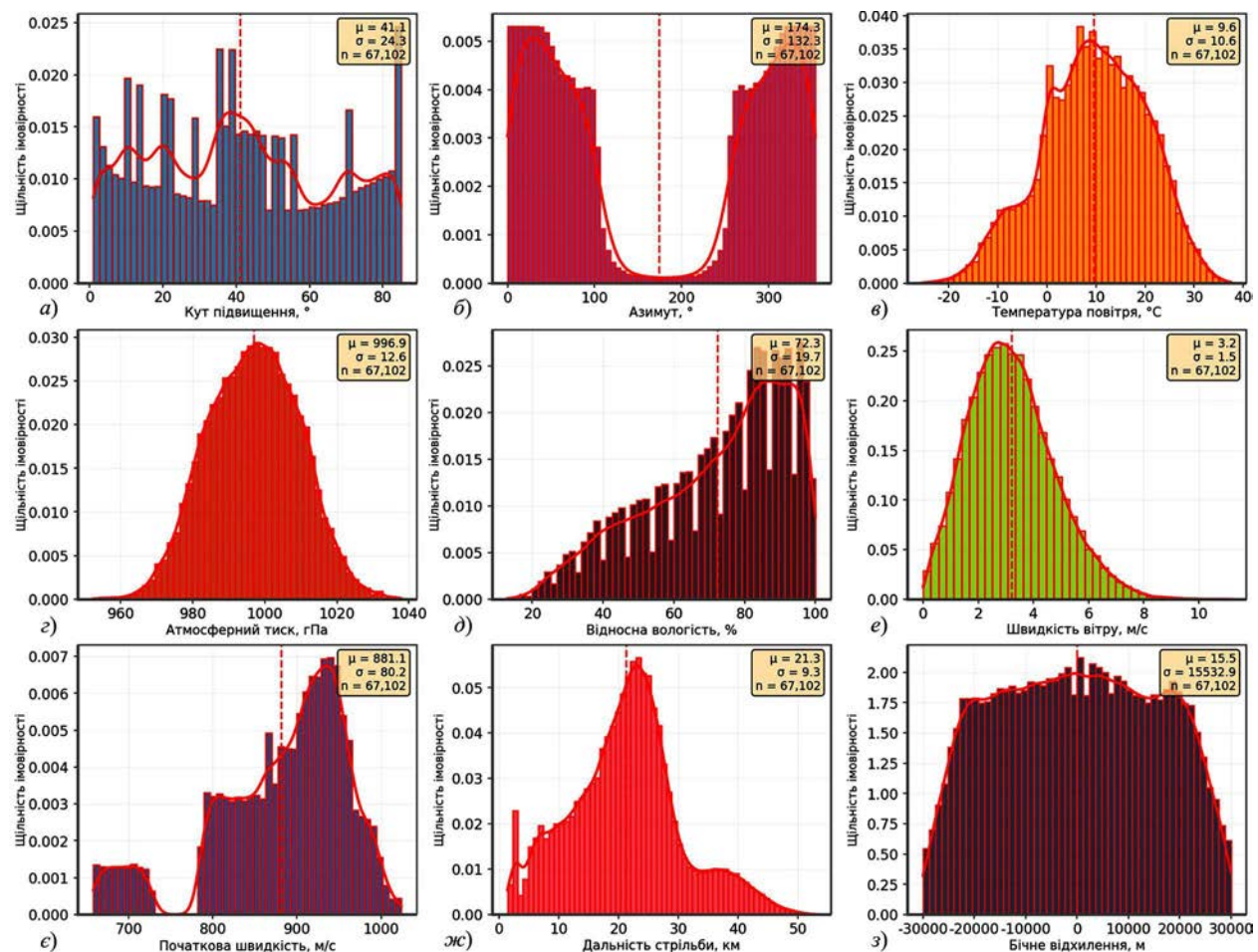


Рис. 1. Гістограми розподілу параметрів синтетичного балістичного набору даних:

- а) – кут підвищення гармати; б) – азимут стрільби; в) – температура повітря; г) – атмосферний тиск; д) – відносна вологість; е) – швидкість вітру; є) – початкова швидкість руху снаряда; ж) – дальність стрільби; з) – бічне відхилення

де μ позначає середнє значення, а σ – стандартне відхилення. Кут підвищення гармати (рис. 1,а) має нерівномірний розподіл із підвищеною густотою у зоні 35° – 45° ($\mu = 41,1^{\circ}$, $\sigma = 24,3^{\circ}$), що є наслідком цілеспрямованого проєктного рішення для кращого охоплення критичної зони максимальної дальності, де відповідно до загальних принципів проєктування артилерії [12] досягається найвище розсіювання снарядів. Азимут стрільби (рис. 1,б) демонструє бімодальний розподіл із підвищеною густотою у зонах 0° – 100° та 280° – 360° і зниженою в центральній зоні, що є наслідком нерівномірного відсіювання записів при фільтрації за координатами влучання ($\mu = 174,3^{\circ}$, $\sigma = 132,3^{\circ}$).

Розподіли метеорологічних параметрів – температури повітря, атмосферного тиску, відносної вологості та швидкості вітру (рис. 1,в–е) – відповідають кліматичним умовам сходу України впродовж повного річного циклу, що підтверджує коректність використання реальних архів-

них даних Open-Meteo. Початкова швидкість руху снаряда (рис. 1,є) є бімодальною з основним піком у зоні 850 – 950 м/с та допоміжним у зоні 700 – 780 м/с ($\mu = 881$ м/с, $\sigma = 80$ м/с), що відображає дві групи артилерійських систем із різними балістичними характеристиками. Дальність стрільби (рис. 1,ж) має асиметричний розподіл із $\mu = 21,3$ км та $\sigma = 9,3$ км і довгим правим хвостом до 50 км, характерним для 155 -мм систем із боеприпасами ERFB-BB підвищеної дальності. Бічне відхилення (рис. 1,з) є симетричним відносно нуля, що підтверджує коректне відтворення азимутальної залежності балістичних функцій, узгоджуючись із результатами моделювання польоту 155 -мм снарядів у роботах [21, 25].

Для формалізованого оцінювання якості згенерованого набору даних використано фреймворк СТТ [5], що об'єднує дев'ять метрик у три виміри: достовірність (C), відтворюваність (R) та відповідність (V). Кожен вимір адаптовано до специфіки синте-

тичного балістичного набору даних: роль анотатора виконує фізичний симулятор, роль інструкцій анотування – стандарт STANAG 4355, а роль лінгвістичного явища – фізичні закономірності польоту снаряда. Результати оцінювання якості синтетичного набору даних наведено на рис. 2, де зведена оцінка якості за дев'ятьма метриками становила 0,875, що вважається відмінним результатом.

Вимір достовірності ($C = 0,807$) охоплює покриття очікуваних фізичних діапазонів вхідних змінних ($C1 = 0,772$), коректність відтворення кліматичних умов через реальні архівні дані Open-Meteo ($C2 = 0,856$) та відповідність симульованих результатів фізичним закономірностям зовнішньої балістики ($C3 = 0,794$), де окремі відхилення спостерігаються у зонах граничних кутів підвищення гармати з підвищеною нелінійністю балістичних функцій.

Вимір відтворюваності ($R = 0,837$) оцінює стійкість статистичних моментів при розбитті на незалежні підвибірки ($R1 = 0,778$), узгодженість розподілів між тренувальною, валідаційною

та тестовою вибірками після поділу у пропорції 78/12/10 % ($R2 = 0,876$) та відсутність систематично незаповнених областей у просторі вхідних ознак ($R3 = 0,857$).

Вимір відповідності ($V = 0,921$) оцінює наявність очікуваних фізичних залежностей: монотонність дальності відносно кутів підвищення гармати та початкової швидкості ($V1 = 0,834$), повноту охоплення операційного простору за всіма типами рельєфних сценаріїв і боєприпасів ($V2 = 0,965$) та відповідність розподілу дальностей очікуваному фізичному розподілу ($V3 = 0,964$).

Верифікація якості синтетичного набору даних через навчання нейронної мережі. Для підтвердження придатності сформованого набору даних до навчання нейромережових компонентів АСУВ виконано його верифікацію шляхом глибокого навчання нейронної мережі з окремими моделями для низького (1° – 45°) та високого (45° – 85°) режимів стрільби. Зведені метрики якості прогнозування нейронної мережі для обох режимів стрільби наведено на рис. 3.

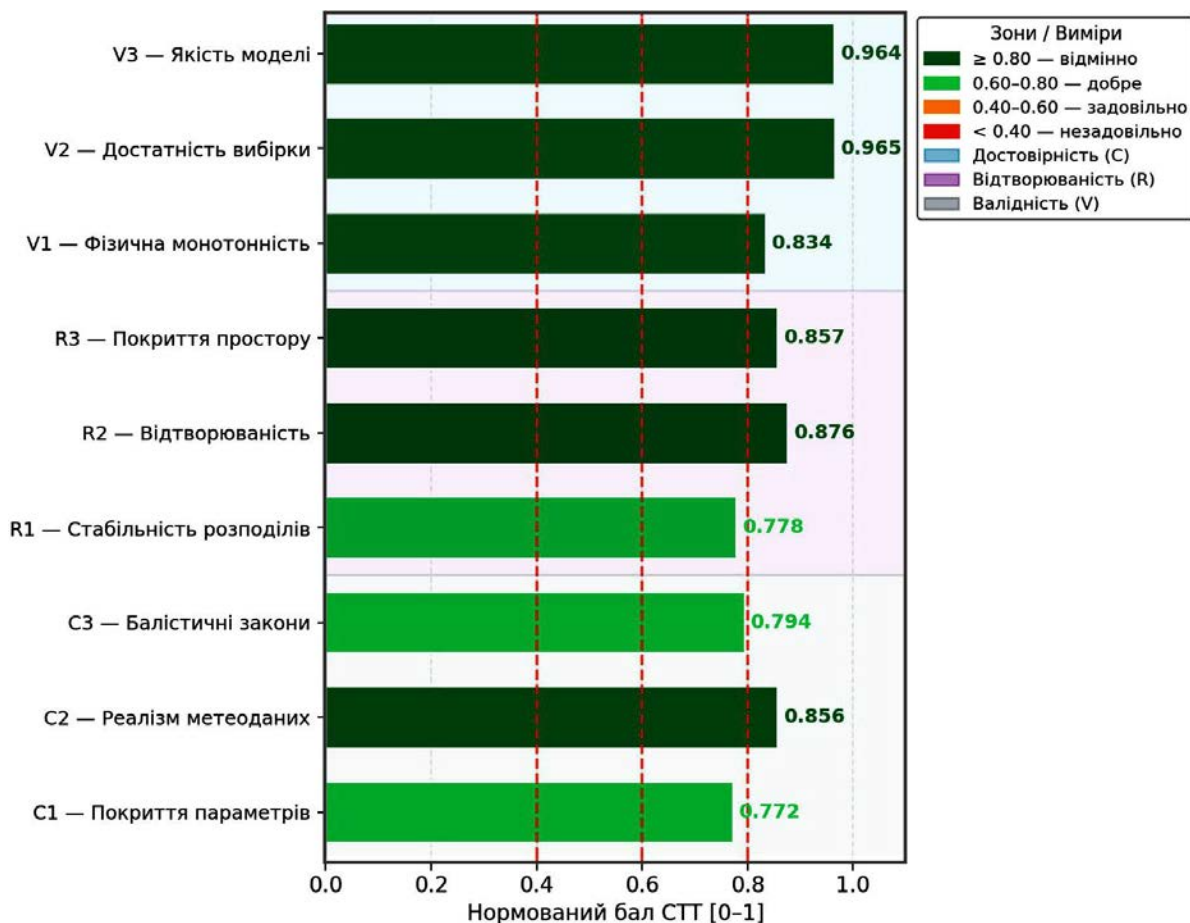


Рис. 2. Результати оцінювання якості синтетичного набору даних за адаптованими метриками фреймворку СТТ

Запропонована архітектура нейронної мережі містить блок розширення ознак, механізм самоуваги [16] та чотири залишкові блоки з окремими гілками для прогнозування кута підвищення гармати та азимута стрільби. Подібний підхід із застосуванням глибокого навчання нейронної мережі винятково на симульованих даних розглянуто в роботі [24]. Набір даних поділено у пропорції 78/12/10 % на тренувальну, валідаційну та тестову вибірки.

Згідно з результатами навчання нейронної мережі, для кута підвищення гармати в низькому режимі стрільби досягнуто $MAE = 2,2^\circ$, $RMSE = 4,9^\circ$ та $R^2 = 0,87$, у високому режимі показники є порівнянними: $MAE = 2,9^\circ$, $RMSE = 4,7^\circ$ та $R^2 = 0,86$. Різниця між MAE та $RMSE$ в обох режимах свідчить про наявність окремих складних конфігурацій із вищою похибкою, переважно в зонах перехідних кутів підвищення гармати. Для азимута стрільби обидва її режими демонструють близькі значення $MAE = 2,5^\circ$ і $2,9^\circ$ та $RMSE = 3,2^\circ$ і $3,6^\circ$. Отримані значення MAE забезпечать звуження простору пошуку кутів наведення гармати до коридору $\pm 3^\circ - 5^\circ$, що гарантуватиме швидку збіжність ітераційного уточнення кутів її наведення в системах управління артилерійським вогнем. Класифікаційний аналіз за матрицями плутанини (рис. 4) підтверджує коректність навчання обох моделей.

Точність класифікації кута підвищення гармати становила 93,3 % для низького та 93,9 % для високого режимів стрільби, при цьому в обох режимах нижчі значення $recall$ спостерігаються у підзонах із вищою нелінійністю балістичних функцій ($1^\circ -$

23° та $65^\circ - 85^\circ$ відповідно). Класифікація азимута стрільби демонструє стабільно вищі показники: 96,7 % для низького режиму із $recall$ 0,96–0,99 та 94,6 % для високого режиму із $recall$ 0,93–1,00 по всіх чотирьох квадрантах.

Отримані метрики якості навчання нейронної мережі підтверджують, що сформований синтетичний набір даних забезпечує достатнє різноманіття та репрезентативність для навчання нейромережових моделей прогнозування балістичних рішень і є придатним для використання в компонентах АСУВ.

Обговорення результатів дослідження.

Отримані результати дослідження підтверджують ефективність розробленої методики генерування синтетичного балістичного набору даних. Сформований набір даних засвідчив спроможність запропонованого п'ятиетапного конвеєру генерувати фізично коректні балістичні траєкторії руху снаряда без залучення даних натурних випробувань із повним охопленням операційного простору артилерійської стрільби [12].

Ключовою перевагою запропонованого підходу є використання реальних архівних метеоданих Open-Meteo замість штучної незалежної рандомізації кожного атмосферного параметра. На відміну від наявних робіт [7, 8], де синтетичні дані було згенеровано зі статистично незалежною рандомізацією, застосування реальних метеопрофілів зберігає природний кореляційний зв'язок між атмосферними змінними, що узгоджується з принципом Domain Randomization [3]. Цілеспрямоване згущення сітки у зоні $35^\circ - 45^\circ$ відповідає

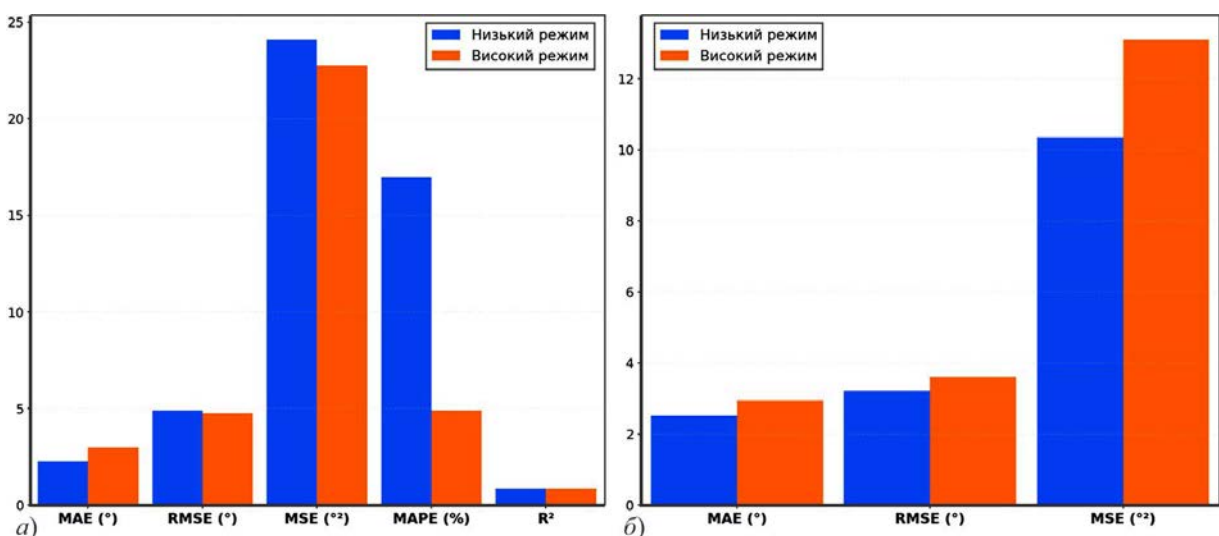


Рис. 3. Зведені метрики якості прогнозування нейронної мережі для низького та високого режимів стрільби: а) – кут підвищення; б) – азимут стрільби

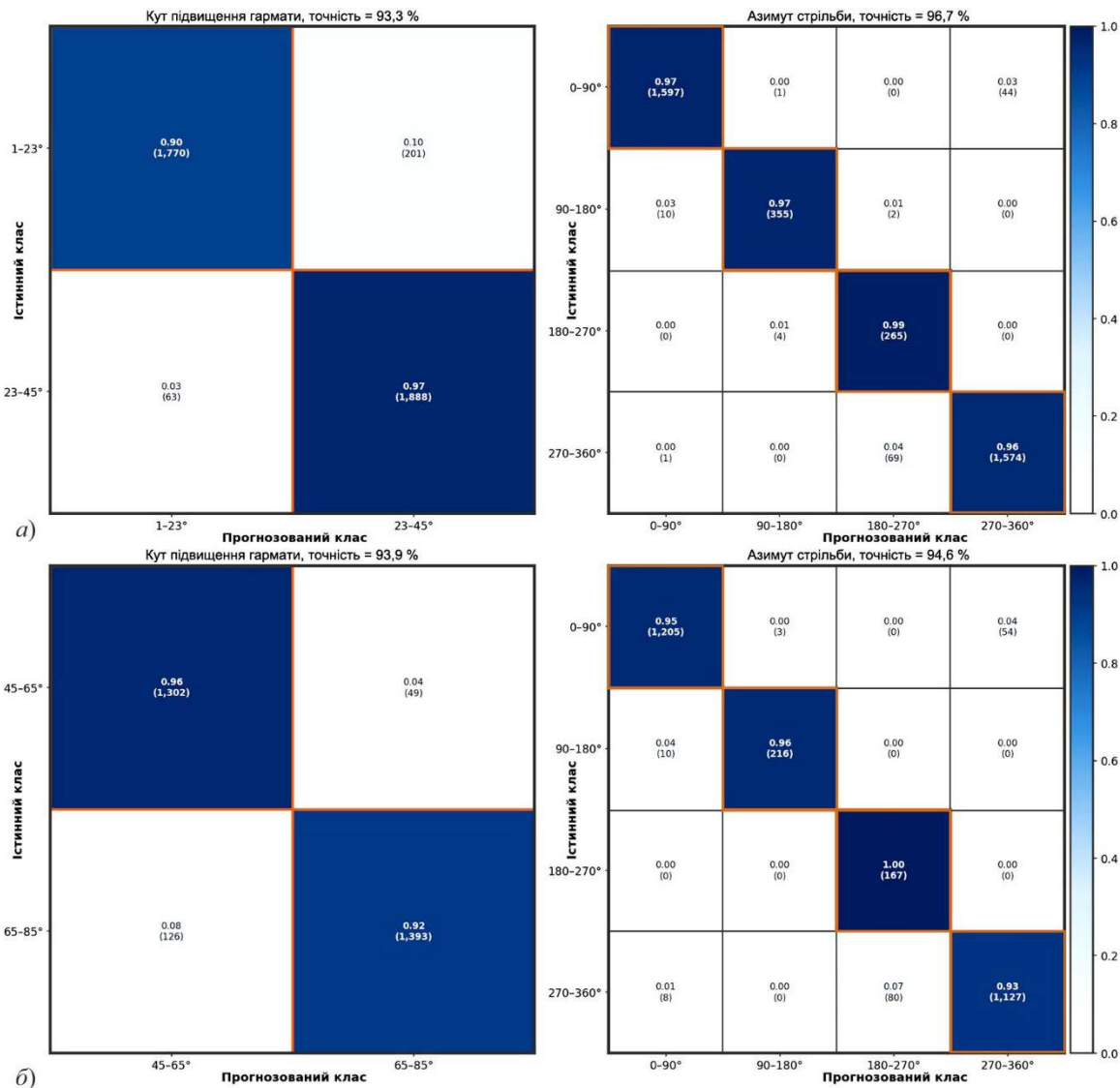


Рис. 4. Матриці плутанини для класифікації кута підвищення та азимута стрільби: а) – низький режим стрільби (1–45°); б) – високий режим стрільби (45–85°)

висновку роботи [5] про необхідність врахування контексту та цільового призначення під час оцінювання якості синтетичного набору даних.

Адаптований фреймворк СТТ [5] продемонстрував практичну придатність для формалізованого оцінювання якості синтетичних балістичних наборів даних. Найвищі показники виміру відповідності ($V2 = 0,965$, $V3 = 0,964$) засвідчують повноту охоплення операційного простору та фізичну коректність згенерованих траєкторій руху снаряда, що корелює з результатами роботи [20]. Дещо нижчі оцінки метрик $C1$ (0,772) та $R1$ (0,778) є результатом цілеспрямованого згущення вибірки у зоні максимальної дальності, а не недоліком методики.

Результати верифікації сформованого набору даних через навчання нейронної мережі демон-

струють, що досягнуті значення MAE ($2,2^\circ$ – $2,9^\circ$ для кута підвищення гармати, $2,5^\circ$ – $2,9^\circ$ – для азимута стрільби) є достатніми для використання моделі як генератора початкового наближення у гібридній системі управління вогнем [10]. Порівняно з підходом [13], запропонований нейромеревий компонент забезпечує істотно вищу швидкість за рахунок деякого зниження точності, яка може компенсуватися подальшим ітераційним уточненням [9]. Класифікаційна точність стрільби 93,3 %–96,7 % підтверджує належне покриття операційного простору для навчання нейронних моделей прогнозування балістичних рішень.

Унаслідок обговорення результатів дослідження було підтверджено доцільність та оригінальність запропонованого підходу. Зведена СТТ-оцінка набору даних 0,875, а також класифі-

каційна точність моделі 93,3–96,7 % засвідчують, що розроблена методика їх генерування забезпечує формування фізично коректного та практично придатного синтетичного балістичного набору даних. Водночас, аналіз наукової літератури показав, що жодна з проаналізованих робіт не поєднує в єдиній методиці систематичне генерування балістичних даних, їх фізично обґрунтоване моделювання та формалізоване оцінювання якості, позаяк наявні дослідження охоплюють тільки окремі з цих складових. Виконане дослідження заповнює цю прогалину, що підтверджує його наукову актуальність і практичну значущість.

Отже, на підставі виконаної роботи сформульовано таку наукову новизну та практичну значущість отриманих результатів дослідження.

Наукова новизна отриманих результатів дослідження: удосконалено методику генерування синтетичного балістичного набору даних, яка, на відміну від наявних, поєднує формування систематичної сітки кутів наведення гармати, рандомізацію фізичних параметрів стрільби згідно з принципом Domain Randomization, чисельне моделювання польоту снаряда за моделлю чотирьох ступенів свободи відповідно до стандарту NATO STANAG 4355, а також багаторівневу фільтрацію та консолідацію результатів, що забезпечує відтворення реалістичних умов стрільби на основі реальних метеоархівів і адаптованого фреймворку СТТ.

Практична значущість результатів дослідження: розроблену методику можна застосувати для навчання нейромережових компонентів систем управління вогнем артилерії, для розширення балістичного набору даних задля прогнозування стрільби нових артилерійських систем через доповнення бази балістичних параметрів без їх натурних випробувань, що скорочує тривалість та вартість розроблення нейромережових компонентів АСУВ.

Висновки. Розроблено методику генерування синтетичного балістичного набору даних на підставі чисельного моделювання польоту снаряда та рандомізації його параметрів для навчання нейромережових компонентів систем управління артилерійським вогнем, яка забезпечить верифікацію якості такого набору за адаптованими метриками фреймворку

СТТ. За результатами проведеного дослідження можна зробити такі основні висновки.

1. Розроблено методику генерування синтетичного балістичного набору даних як п'ятиетапного конвеєра, що поєднує формування систематичної сітки кутів наведення гармати, рандомізацію фізичних параметрів стрільби, чисельне моделювання польоту снаряда за моделлю 4-DOF відповідно до стандарту NATO STANAG 4355, фільтрацію та консолідацію результатів. Сформований набір даних обсягом 67102 записів для 12 артилерійських систем калібру 155 мм підтвердив спроможність розробленої методики генерувати фізично коректні балістичні траєкторії руху снаряда без натурних випробувань.

2. Обґрунтовано вибір параметрів рандомізації набору даних, де ключовим рішенням є використання реальних архівних метеоданих Open-Meteo для збереження природного кореляційного зв'язку між атмосферними змінними. Визначено діапазони рандомізації початкової швидкості руху снаряда ($\pm 5\%$), температури порохового заряду ($-5...+10^\circ\text{C}$) та висоти цілі за трьома рельєфними сценаріями, що забезпечують відтворення реалістичних умов артилерійської стрільби.

3. Адаптовано фреймворк СТТ для оцінювання якості синтетичного набору даних у завданнях регресійного балістичного прогнозування з дев'ятьма метриками у трьох вимірах: достовірність, відтворюваність та відповідність. Зведена оцінка якості 0,875 підтверджує високу якість сформованого набору даних та повноту охоплення операційного простору і фізичну коректність згенерованих траєкторій руху снаряда.

4. Верифіковано якість сформованого набору даних через їхній аналіз за адаптованими метриками СТТ та за навчанням нейронної мережі. Досягнуті значення MAE = $2,2^\circ$ – $29,0^\circ$ для кута підвищення гармати та MAE = $2,5^\circ$ – $2,9^\circ$ – для азимута стрільби, а також класифікаційна точність стрільби 93,3 %–96,7 % підтверджують придатність згенерованого набору даних для навчання нейромережових компонентів АСУВ.

Запропонована методика дає змогу розширювати набір даних для нових артилерійських систем без повторного проведення натурних випробувань, що скорочує тривалість і вартість їх розроблення.

Список літератури:

1. Рубель Ю., Грицюк Ю. Сучасний стан автоматизованих систем управління вогнем артилерійських установок. *Herald of Khmelnytskyi National University. Technical sciences*. 2025. Т. 355, № 4. С. 520–530. DOI: <https://doi.org/10.31891/2307-5732-2025-355-73>

2. Li Y., Bai H., Chen D., Feng Y., Yin J., Zhao J., Yang S., Wang T. Spatiotemporal model for real-time

projectile prediction in digital prototyping of artillery. *Journal of Computational Design and Engineering*. 2025. Vol. 12, no. 8. P. 60–77 DOI: <https://doi.org/10.1093/jcde/qwaf072>

3. Tobin J., Fong R., Ray A., Schneider J., Zaremba W., Abbeel P. Domain randomization for transferring deep neural networks from simulation to the real world. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2017. P. 23–30. DOI: <https://doi.org/10.1109/iros.2017.8202133>

4. Becker S., Hug R., Huebner W., Arens M., Morris B. Generating Synthetic Training Data for Deep Learning-based UAV Trajectory Prediction. *2nd International Conference on Robotics, Computer Vision and Intelligent Systems*. 2021. P. 13–21. DOI: <https://doi.org/10.5220/0010621400003061>

5. Wang C., Dong Q., Wang X., Sui, Z. Statistical dataset evaluation: A case study on named entity recognition. *Natural Language Processing*. 2024. Vol. 31, no. 1. P. 1–21. DOI: <https://doi.org/10.1017/nlp.2024.37>

6. Ghosh A. K., Prakash, O. Neural Models for Predicting Trajectory Performance of an Artillery Rocket. *Journal of Aerospace Computing, Information, and Communication*. 2004. Vol. 1, no. 2. P. 112–115. DOI: <https://doi.org/10.2514/1.3398>

7. Bragado A. C., Solera-Rico A., Gómez M. A. Implementation of Trajectory Propagator for Artillery Projectiles Based on Artificial Neural Networks. *Conference New Technologies and Developments in Unmanned Systems*. 2023. P. 187–192. DOI: https://doi.org/10.1007/978-3-031-37160-8_29

8. Zang H., Gao C., Hu Y., Jing W. Trajectory prediction algorithm of ballistic missile driven by data and knowledge. *Defence Technology*. 2025. Vol. 48. P. 187–203. DOI: <https://doi.org/10.1016/j.dt.2025.02.001>

9. Рубель Ю. Б., Грицюк Ю. І. Інтелектуальне коригування стрільби в автоматизованих системах управління вогнем артилерійських установок. *Collection of Scientific Papers with Proceedings of the 1st International Scientific and Practical Conference. International Scientific Unity*. 2025. P. 192–198. DOI: <https://doi.org/10.70286/isu-14.05.2025>

10. Рубель Ю. Б., Грицюк Ю. І. Архітектурний підхід до реалізації гібридної системи управління вогнем артилерії. *Collection of Scientific Papers with Proceedings of the 4th International Scientific and Practical Conference. International Scientific Unity*. 2026. P. 125–130. DOI: <https://doi.org/10.70286/isu-14.01.2026.006>

11. Rothe H., Kuhrt A., Schroeder S., Trebing S. Fire control algorithms and software for the modular naval artillery concept (MONARC) of the German navy. *Computational Ballistics II, WIT Press*. 2005. P. 406–416. URL: <https://www.witpress.com/elibrary/wit-transactions-on-modelling-and-simulation/40/14941>

12. Qian L., Chen G., Tong M., Tang J. General design principle of artillery for firing accuracy. *Defence Technology*. 2022. Vol.18, no.12. P. 2125–2140. DOI: <https://doi.org/10.1016/j.dt.2022.09.001>

13. Mostafa K., Xiaoting R., Hossam H. A Rapid Gun Firing Correction Algorithm Using Discrete Time Transfer Matrix Method. *16th International Conference on Aerospace Sciences & Aviation Technology*. 2015. DOI: <https://doi.org/10.21608/asat.2015.23008>

14. Chen D., Yang C., Ji H., Jiang B., & Liu Z. Application and Implementation of CNN in Artillery Countermeasure Training System. *IOP Conference Series: Materials Science and Engineering*. 2019. Vol. 612, no. 3, article ID 032015. DOI: <https://doi.org/10.1088/1757-899X/612/3/032015>

15. Corriveau D. Validation of the NATO Armaments Ballistic Kernel for use in small-arms fire control systems. *Defence Technology*. 2017. Vol. 13, no. 3. P. 188–199. DOI: <https://doi.org/10.1016/j.dt.2017.04.006>

16. Ashish V., Noam S., Niki P., Jakob U., Llion J., Aidan G., Łukasz K., Illia P. Attention is All you Need. *31st Conference on Neural Information Processing Systems (NIPS 2017)*. 2017. DOI: <https://doi.org/10.48550/arXiv.1706.03762>

17. Herbut A., Brząkała W. Polynomial chaos expansion vs. Monte Carlo simulation in a stochastic analysis of wave propagation. *Wave Motion*. 2024. Vol. 130, article ID 103390. DOI: <https://doi.org/10.1016/j.wavemoti.2024.103390>

18. Ilg M., Rogers J., Costello M. Projectile Monte-Carlo Trajectory Analysis Using a Graphics Processing Unit. *AIAA Atmospheric Flight Mechanics Conference*, Portland. 2011. DOI: <https://doi.org/10.2514/6.2011-6266>

19. Wang L., Chen Z., Yang G. An Uncertainty Analysis Method for Artillery Dynamics with Hybrid Stochastic and Interval Parameters. *Computer Modeling in Engineering & Sciences*. 2021. Vol. 26, no. 2. P. 479–503. DOI: <https://doi.org/10.32604/cmes.2021.011954>

20. Jacewicz M., Lichota P., Miedziński D., Głębocki R. Study of Model Uncertainties Influence on the Impact Point Dispersion for a Gasodynamically Controlled Projectile. *Sensors*. 2022. Vol. 22, no. 9, article ID 3257. DOI: <https://doi.org/10.3390/s22093257>

21. Elsaadany A., Wen-jun Y. Accurate trajectory prediction for typical artillery projectile. *33rd Chinese Control Conference, IEEE*. 2014. P. 6368–6374. DOI: <https://doi.org/10.1109/chicc.2014.6896037>

22. Tian F., Sun X. W., Zhang X., Shan H. Y., Wang L. J. Research on the Simulation of Naval Gun Exterior Trajectory Differential Equations in Real-Time Solution. *Advanced Materials Research*. 2014. Vol. 981. P. 758–761. DOI: <https://doi.org/10.4028/www.scientific.net/amr.981.758>

23. Wang Y.-W., Zhu W.-F., Di J.-W., Hu X.-H. Study on the analysis method on ballistic performance of detherred propellant with large web size in large caliber artillery. *Defence Technology*. 2018. Vol. 14, no. 5. P. 522–526. DOI: <https://doi.org/10.1016/j.dt.2018.07.027>

24. Roux A., Changey S., Weber J., Lauffenburger J.-P. LSTM-Based Projectile Trajectory Estimation in a GNSS-Denied Environment. *Sensors*. 2023. Vol. 23, no. 6, article ID 3025. DOI: <https://doi.org/10.3390/s23063025>

25. Elsaadany A., Wen-jun Y. Accuracy Improvement Capability of Advanced Projectile Based on Course Correction Fuze Concept. *The Scientific World Journal*. 2014. article ID 273450. DOI: <https://doi.org/10.1155/2014/273450>

Rubel Yu. B., Hrytsiuk Yu. I. GENERATION OF SYNTHETIC BALLISTIC DATASET BASED ON NUMERICAL MODELLING AND DOMAIN RANDOMISATION FOR ARTILLERY FIRE CONTROL SYSTEM

This paper presents the results of developing a methodology for generating a synthetic ballistic dataset based on numerical projectile flight simulation and domain randomization, intended for training neural network components of automated artillery fire control systems. The study addresses the challenge of producing high-quality training datasets without resource-intensive field trials, while simultaneously ensuring the physical correctness and representativeness of the generated data. The proposed methodology is implemented as a five-stage pipeline that sequentially combines the formation of a systematic grid of gun elevation and azimuth angles, randomization of firing parameters in accordance with the Domain Randomization principle, numerical simulation of projectile flight using a four-degree-of-freedom model (4-DOF) compliant with the NATO STANAG 4355 standard, and multi-level filtering and consolidation of results. The Classical Test Theory (CTT) framework is adapted for formalized multidimensional quality evaluation of the synthetic dataset in regression-based ballistic prediction tasks, employing nine metrics across three dimensions: credibility, reliability, and validity. The resulting dataset comprises 67102 records covering 12 artillery systems of 155 mm calibre and spans the full practical range of firing conditions. Dataset quality verification was performed through CTT metric analysis and deep learning of a neural network with separate models for low and high firing modes. The overall CTT quality score of the synthetic dataset reached 0,875. Neural network training yielded a mean absolute error in the range of 2,2°–2,9° for the elevation angle and 2,5°–2,9° for the firing azimuth, with a coefficient of determination $R^2 = 0,86–0,87$. The firing classification accuracy of 93,3%–96,7% confirms the suitability of the generated dataset for training neural network models of ballistic decision prediction.

Keywords: synthetic dataset, simulation parameter randomization, numerical simulation of projectile flight, artillery fire control, neural network, CTT framework, four-degree-of-freedom model, NATO STANAG 4355 standard.

Дата першого надходження статті до видання: 05.03.2026

Дата прийняття статті до друку після рецензування: 30.03.2026

Дата публікації (оприлюднення) статті 11.05.2026

УДК 004.8:004.946.5:37.091.4

DOI DOI DOI DOI <https://doi.org/10.32782/2663-5941/2026.2.1/40>**Руденко Ю.О.**<https://orcid.org/0000-0003-3162-1216>

Сумський національний аграрний університет

Агаджанова С.В.<https://orcid.org/0000-0002-0486-3511>

Сумський національний аграрний університет

В'юненко О.Б.<https://orcid.org/0000-0002-8835-0704>

Сумський національний аграрний університет

Ситник Л.Г.<https://orcid.org/0000-0001-8176-750X>

Сумський національний аграрний університет

Агаджанов-Гонсалес К.Х.<https://orcid.org/0000-0002-1409-4648>

Сумський національний аграрний університет

ОЦІНЮВАННЯ РЕЗУЛЬТАТИВНОСТІ ОПАНУВАННЯ LOW-CODE ТЕХНОЛОГІЙ У МЕЖАХ ПІДГОТОВКИ ФАХІВЦІВ ДО АНАЛІТИКИ ВЕЛИКИХ ДАНИХ

Статтю присвячено порівняльному оцінюванню ефективності опанування студентами low-code технологій у процесі підготовки до аналітики великих даних. Актуальність дослідження зумовлена зростанням ролі інструментів спрощеної розробки в освітньому процесі та необхідністю обґрунтованого вибору аналітичних платформ для формування професійних навичок майбутніх фахівців з інформаційних систем та технологій. Метою роботи є емпіричне порівняння результативності використання платформ Orange Data Mining, Weka та KNIME в однакових навчальних умовах.

Дослідження реалізовано у форматі педагогічного експерименту за участю 131 студента, розподілених на три незалежні групи відповідно до обраної платформи. Порівняння здійснювалося за системою об'єктивних (час виконання завдань, кількість типових помилок, якість побудованих моделей) та суб'єктивних (самооцінка складності інтерфейсу) метрик. Для перевірки статистичної значущості відмінностей застосовано непараметричний критерій Краскала–Уолліса.

Результати засвідчили статистично значущі відмінності за показниками часу виконання завдань та кількості типових помилок ($p < 0,05$). Найкращі результати продемонструвала платформа Orange Data Mining, що характеризувалася меншою кількістю помилок, вищою швидкістю виконання завдань та кращою суб'єктивною оцінкою зручності. Водночас якість побудованих аналітичних моделей виявилася порівнюваною між усіма групами та не мала статистично значущих відмінностей. Отримані результати підтверджують, що вибір low-code платформи впливає насамперед на процесуальні характеристики навчальної діяльності, але не визначає кінцевий рівень аналітичних результатів за умови належної методичної підтримки.

Сформульовано практичні рекомендації щодо поетапної інтеграції low-code інструментів у курси з аналітики великих даних залежно від рівня підготовки студентів.

Ключові слова: Big Data, low-code, Orange Data Mining, Weka, KNIME, аналітика даних.

Постановка проблеми. Швидкий розвиток технологій Big Data актуалізував потребу у формуванні навичок аналітики даних у різних галузях, зокрема в освіті, економіці, технічних та прикладних науках. Аналітика великих даних стала ключовим підходом до вилучення знань з масивних і різномірних наборів даних та підтримки прийняття обґрунтованих рішень у широкому спектрі сфер застосування.

Емпіричні дослідження у сфері освіти підтверджують зростаючу роль навичок аналітики даних у професійній підготовці. Зокрема, показано, що цифрова освіта та сформовані навички аналітики даних можуть позитивно впливати на кар'єрні перспективи майбутніх фахівців, особливо інженерних і технічних спеціальностей.

Водночас практична реалізація навчання аналітиці даних часто гальмується через складність традиційних інструментів та вимогами до програмної підготовки. У цьому контексті low-code та no-code платформи розглядаються як важливий технологічний тренд, спрямований на зниження технічних бар'єрів доступу до аналітики даних. Питання, пов'язані з опануванням low-code технологій майбутніми фахівцями залишаються недостатньо дослідженими. Обмежено представлені наукові розвідки, які б висвітлювали особливості результатів виконання навчальних завдань, кількість типових помилок, якість побудованих моделей та сприйняття складності інтерфейсу під час роботи з різними low-code платформами. Відсутність таких порівняльних оцінок ускладнює обґрунтований вибір інструментів для навчальних курсів з аналітики великих даних у закладах вищої освіти. Саме ця прогалина зумовлює актуальність проведеного дослідження.

Аналіз останніх досліджень і публікацій. Розвиток аналітики великих даних зумовлює трансформацію підходів до підготовки фахівців у сфері інформаційних систем та технологій. У [1] Big Data (великі дані) розглядаються як ключовий ресурс цифрової економіки, що потребує сучасних інструментів збирання, оброблення та аналізу. Дослідження авторів [2] підтверджує наявність статистично значущого зв'язку між рівнем цифрової підготовки здобувачів освіти та їх професійною успішністю, зокрема у сферах, пов'язаних з обробленням даних. У межах концепції Education 5.0 інтеграція інтелектуальних та адаптивних технологій у навчальний процес визначається як стратегічний напрям модернізації освіти [3].

Одним із перспективних напрямів цифрової трансформації розроблення програмних рішень

є використання low-code/no-code (LC/NC) технологій. У джерелах [4], [5], low-code/no-code визначається як підхід до створення програмних продуктів та інформаційних систем із мінімальним або повністю відсутнім ручним програмуванням, що реалізується через використання візуальних середовищ розробки, шаблонних компонентів, декларативних налаштувань та автоматизованої генерації коду.

Окремий напрям досліджень присвячено використанню візуальних платформ як бази low-code технологій. Платформа Orange Data Mining характеризується як доступне та наочне середовище, придатне для початкового та прикладного аналізу даних у навчальних контекстах [6]. Натомість платформа KNIME позиціонується як більш гнучке аналітичне середовище, орієнтоване на побудову складних workflow та застосування методів машинного навчання, що є доцільним у бізнес-освіті та навчанні інформаційних систем [7]. Дослідження також свідчать, що використання платформ на кшталт Orange, Weka та RapidMiner сприяє зниженню складності навчання та когнітивного навантаження студентів під час опанування аналітичних методів [8], [9].

Незважаючи на наявність значної кількості робіт, присвячених окремим платформам та методам, у науковій літературі все ще бракує порівняльних досліджень, що описують особливості опанування платформ.

Постановка завдання. Метою даного дослідження є порівняльне оцінювання ефективності опанування студентами low-code технологій у процесі підготовки фахівців до аналітики великих даних. Дослідження ґрунтується на емпіричному порівнянні результатів роботи студентів з платформами Orange Data Mining, Weka та KNIME в однакових навчальних умовах і спрямоване на формування практичних рекомендацій щодо вибору інструментів для освітніх курсів з аналітики даних.

Для досягнення поставленої мети в роботі передбачено розв'язання таких завдань:

1. Проаналізувати особливості застосування low-code платформ у навчанні аналітиці великих даних студентів.
2. Оцінити суб'єктивне сприйняття складності інтерфейсу та зручності використання low-code інструментів на основі результатів анкетування студентів.
3. Застосувати статистичні методи для перевірки значущості відмінностей між результатами роботи студентів з різними платформами.

4. Сформулювати практичні рекомендації щодо вибору low-code платформ при підготовці фахівців до аналітики великих даних.

Виклад основного матеріалу. Дослідження виконано у форматі педагогічного експерименту, у якому брали участь студенти 3-го курсу українських університетів, які навчаються за освітніми програмами менеджменту, економіки та інформаційних технологій і вивчають дисципліну «Основи дата аналітики».

Формування експериментальних груп здійснювалося на основі готових академічних груп, що зумовлено організаційними особливостями освітнього процесу. Загальна вибірка становила 131 студент, які були розподілені на три незалежні групи відповідно до використаної аналітичної платформи: G1 - Orange Data Mining ($n = 45$); G2 - Weka ($n = 46$); G3 - KNIME ($n = 40$). Кожна група протягом експерименту працювала лише з однією платформою.

Для забезпечення коректності подальшого порівняльного аналізу було здійснено контроль початкової порівняльності експериментальних груп. Як базовий показник використовувалася попередня академічна успішність студентів, визначена за середнім балом з дисциплін аналітичного та інформаційного спрямування (Інформатика, Математика, Економіка), вивчених на попередніх курсах.

Для перевірки нормальності розподілу показника попередньої академічної успішності було застосовано критерій Шапіро–Уїлка як один із найбільш чутливих та рекомендованих тестів для вибірок малого та середнього обсягу [10]. Розрахунки виконано у хмарному статистичному середовищі Jamovi Cloud. Виявлені статистично значущі відхилення від нормального розподілу в усіх досліджуваних групах ($p < 0,05$) зумовили використання непараметричного однофакторного аналізу за критерієм Краскела–Уолліса для перевірки початкової порівняльності груп.

Результати аналізу не виявили статистично значущих відмінностей між групами ($\chi^2 = 1,45$, $df = 2$, $p = 0,483$), що дозволяє вважати їх порівнюваними за рівнем попередньої академічної успішності. Таким чином, подальші відмінності у результатах виконання практичних завдань можуть бути інтерпретовані як наслідок використання різних low-code платформ, а не як результат початкових академічних відмінностей між студентами.

У межах дослідження було використано три low-code платформи аналітики даних:

– Orange Data Mining – візуальне аналітичне середовище зшироким набором інструментів для

описового аналізу, кластеризації й візуалізації даних;

– Weka – інструмент машинного навчання з акцентом на алгоритмічний підхід, параметризацію моделей і класичні методи Data Analytics;

– KNIME – платформа, що дозволяє будувати аналітичні процеси та комбінувати різні методи аналізу.

Для забезпечення порівнюваності результатів усі групи виконували два однакових практичні завдання, які відрізнялися лише використаною аналітичною платформою.

Практичне завдання 1. Описовий та порівняльний аналіз освітніх даних. Завдання було спрямовано на опанування базових методів описової статистики та міжгрупового порівняльного аналізу. Для аналізу використовувалися результати опитування студентів щодо якості освітніх послуг (експорт із LMS Moodle) та офіційні статистичні дані Державної служби статистики України з банку даних «Мережа та діяльність закладів освіти» [11].

Практичне завдання 2. Кореляційний аналіз соціально-освітніх показників. Спрямовано на формування навичок багатовимірного аналізу та інтерпретації взаємозв'язків між показниками на основі відкритого глобального датасету UNICEF JMP WASH (2025) [12].

Для забезпечення об'єктивного та порівнюваного оцінювання ефективності опанування студентами low-code платформ у дослідженні було визначено систему кількісних і якісних метрик, які охоплювали часові характеристики виконання завдань, кількість типових помилок, якість побудованих аналітичних моделей та суб'єктивну оцінку складності інтерфейсу (анкетування). Таблиця 1 описує метрики, що застосовувалися в експерименті.

Якість побудованих моделей оцінювалася викладачем курсу за уніфікованими критеріями. Попередня обробка даних, розрахунок описових статистик і підготовка зведених таблиць здійснювалися в Microsoft Excel. Для підвищення валідності та відтворюваності дослідження всі навчальні завдання, набори даних та критерії оцінювання були уніфіковані для всіх груп, а аналітичні робочі потоки будувалися за однаковою логікою.

Для узагальненого порівняння ефективності опанування платформ було проаналізовано три об'єктивні метрики: час виконання практичних завдань, кількість типових помилок та якість побудованих аналітичних моделей. Узагальнені результати наведено в таблиці 2.

Таблиця 1

Метрики порівняльного оцінювання ефективності опанування low-code платформ

№	Метрика	Опис показника	Спосіб фіксації / оцінювання	Шкала
1	Час виконання завдання	Загальна тривалість виконання завдання від імпорту даних до формулювання аналітичних висновків	Фіксація часу початку та завершення виконання завдання	Низька/середня/висока
2	Кількість типових помилок	Кількість зафіксованих помилок під час виконання аналітичного завдання	Підрахунок помилок за уніфікованим переліком	кількість
3	Якість побудованих моделей	Рівень коректності, логічності та повноти аналітичного рішення	Експертна оцінка за уніфікованою шкалою	0–5 балів
4	Самооцінка складності інтерфейсу	Суб'єктивне сприйняття зручності та інтуїтивності платформи	Анкетування студентів (авторський опитувальник)	шкала (1–5)

Примітка: уніфікований перелік для метрики 2: помилки імпорту даних (неправильний формат, пропуски, дублікати); некоректне очищення або попередня обробка даних; неправильний вибір або налаштування аналітичного методу; некоректна або неповна візуалізація результатів; помилки інтерпретації отриманих залежностей.

Таблиця 2

Порівняльні результати опанування low-code платформ (об'єктивні метрики)

Метрика	G1: Orange Data Mining (n=45)	G2: Weka (n=46)	G3: KNIME (n=40)	Статистичний тест
Час виконання завдань	низька	середня	висока	Kruskal–Wallis, $p = 0.00193$
Кількість помилок	2,1	2,6	3,0	Kruskal–Wallis, $\chi^2 = 18.32$, $p < 0.001$
Якість моделей (0–5)	$\approx 4,6$	$\approx 4,4$	$\approx 4,3$	Kruskal–Wallis, $p = 0.054$

Примітка. Для метрики часу виконання наведено відносні значення (низька / середня / висока)

Таблиця 3

Самооцінка складності інтерфейсу low-code платформ

Група	Платформа	Середній бал (1–5)
G1	Orange Data Mining	4,4
G2	Weka	4,0
G3	KNIME	3,7

Подальша інтерпретація рівня доступності та складності low-code платформ ґрунтується на сукупності об'єктивних (час виконання завдань, кількість типових помилок, якість моделей) та суб'єктивних (самооцінка складності інтерфейсу) показників, отриманих у межах даного експерименту.

Результати свідчать, що вибір low-code платформи істотно впливає на показники навчальної діяльності. Студенти, які працювали з Orange Data Mining, виконували завдання швидше та допускали меншу кількість типових помилок порівняно з групами Weka та KNIME, що підтверджується статистично значущими результатами ($p < 0,05$).

Водночас якість побудованих аналітичних моделей виявилася порівнюваною між усіма трьома групами: отримані відмінності не досягли рівня статистичної значущості ($p = 0,054$). Це свідчить про те, що за умови однакових навчальних завдань усі платформи дозволяють досягти подібного рівня якості кінцевих аналітичних рішень, навіть за різної складності інтерфейсу та робочого потоку.

Окремо було проаналізовано самооцінку складності інтерфейсу, що відображає суб'єктивне

сприйняття студентами зручності та інтуїтивності використаної платформи. Оцінювання здійснювалося за п'ятибальною шкалою на основі авторського опитувальника.

Отримані результати засвідчують, що Orange Data Mining сприймається студентами як найбільш зручна та інтуїтивна платформа, тоді як Weka займає проміжну позицію, а KNIME характеризується найвищим рівнем суб'єктивної складності. Виявлені відмінності є статистично значущими ($p < 0,05$).

Самооцінка складності інтерфейсу узгоджується з об'єктивними показниками часу виконання завдань та кількості типових помилок, що підсилює валідність отриманих результатів і підтверджує взаємозв'язок між інтерфейсною складністю платформи та ефективністю навчальної діяльності.

Аналіз результатів виконання завдань показав, що характер завдання впливає на ефективність використання low-code платформ.

Практичне завдання 1, орієнтоване на описовий та порівняльний аналіз освітніх даних, загалом виконувалася студентами швидше та з меншою кількістю типових помилок, незалежно від використаної платформи. Це пояснюється чіткою структурою завдання, наявністю знайомих показників та меншою кількістю етапів аналітичного робочого процесу. Особливо ефективною виявилася робота з Orange Data Mining, де візуальні

інструменти описової статистики та візуалізації дозволяли швидко отримувати інтерпретовані результати.

Практичне завдання 2, спрямоване на кореляційний аналіз багатовимірних соціально-освітніх даних, виявилася складнішою для студентів усіх груп. Для цього завдання характерними були більший час виконання, зростання кількості типових помилок та підвищене когнітивне навантаження, особливо у групах, що працювали з Weka та KNIME. Це пов'язано з необхідністю попередньої обробки даних, вибору коректних змінних та інтерпретації кореляційних зв'язків.

Інтерпретація рівня доступності та складності платформ Orange Data Mining, Weka та KNIME ґрунтується на сукупності об'єктивних показників (час виконання завдань, кількість типових помилок, якість побудованих моделей) і суб'єктивних оцінок користувачів (сприйняття складності інтерфейсу), отриманих у межах даного експерименту. Orange Data Mining продемонструвала найменший середній час виконання навчальних завдань, найменшу кількість типових помилок і найвищі показники користувацької задоволеності, що дозволяє інтерпретувати її як найбільш доступну платформу для початкового рівня навчання аналітиці даних. Weka характеризується помірними значеннями часу виконання завдань і кількості помилок, що позиціонує її як платформу середньої складності, доцільну для ознайомлення студентів із класичними алгоритмами машинного навчання після формування базових навичок аналітики даних.

KNIME продемонструвала кращі результати у виконанні складних workflow-завдань та високу якість моделей, однак супроводжувалася більшим часом виконання завдань і вищою суб'єктивною складністю інтерфейсу, що свідчить про доцільність її використання на професійному етапі навчання.

Таким чином, результати свідчать, що Orange Data Mining є більш ефективною платформою для виконання базових аналітичних завдань, тоді як Weka та KNIME доцільніше використовувати для

складніших завдань або на наступних етапах підготовки, коли студенти вже володіють базовими навичками аналітики даних.

Висновки. Отримані результати підтвердили, що вибір low-code платформи істотно впливає на процес навчальної діяльності, зокрема на час виконання завдань, кількість типових помилок і суб'єктивне сприйняття складності інтерфейсу. Платформа Orange Data Mining продемонструвала найкращі показники за швидкістю виконання завдань, мінімальною кількістю помилок та найвищим рівнем користувацької задоволеності, що робить її найбільш доступною для початкового етапу навчання аналітиці даних.

На основі отриманих результатів сформульовано такі практичні рекомендації для організації навчання аналітиці даних:

1. Ефективною формою навчання аналітиці великих даних є навчальні learning modules з поступовим зростанням складності завдань, де вибір платформи узгоджується з рівнем підготовки студентів і типом аналітичної діяльності. Orange Data Mining доцільно використовувати як стартову платформу для навчальних курсів з аналітики даних. Weka рекомендується застосовувати після формування базових навичок аналітики даних, ця платформа корисна для ознайомлення студентів із класичними алгоритмами машинного навчання. KNIME доцільно інтегрувати у процес навчання в межах дослідницьких завдань, для побудови складних робочих потоків і комбінування різних аналітичних методів.

2. Для підвищення ефективності навчання доцільно поєднувати декілька low-code платформ у межах одного курсу, використовуючи їх послідовно відповідно до освітніх цілей і очікуваних результатів навчання.

Перспективи подальших досліджень пов'язані з розширенням вибірки, включенням інших low-code інструментів, а також аналізом довготривалого впливу використання різних платформ на формування професійних аналітичних навичок студентів.

Список літератури:

1. Bakdi M., Chadli W. Big Data: An Overview. *Big Data Analytics. First edition.*, 2021. P. 3–13. doi.org/10.1201/9781003129660-2
2. B. Tello, Y.-T. Wu, J. Vilchez-Guizado, and T. P. Hoang, "Professional success mediated by big data: The relationship between digital education and career prospects among engineering students," *Multidisciplinary Science Journal*, 2024, Art. no. 2026377, doi: 10.31893/multiscience.2026377
3. N. Vaidya, K. Solanki, K. Kant, and J. Panchal, "Learning metamorphoses: Revolutionising education with technology integration and customisation in Education 5.0," in *Lecture Notes in Networks and Systems*. Springer, 2024, doi: 10.1007/978-981-96-7505-0_14.

4. M. O. Ajimati, N. Carroll, and M. Maher, "Adoption of low-code and no-code development: A systematic literature review and future research agenda," *Journal of Systems and Software*, vol. 203, Art. no. 111708, 2023, doi: 10.1016/j.jss.2023.111708.
5. Zhong, J. Oruongo, and J. B. J. B. Kim, "LLM-powered low-code/no-code data analytics in education and workforce development," *Computer*, vol. 58, no. 1, pp. 48–57, 2025, doi: 10.1109/MC.2024.3516614.
6. Z. Dobešová, "Evaluation of Orange data mining software and examples for lecturing machine learning tasks in geoinformatics," *Computer Applications in Engineering Education*, 2024, doi: 10.1002/cae.22735.
7. M. Halaweh, "Teaching Tip: AI and Machine Learning for Business and Information Systems Education Using KNIME," *Journal of Information Systems Education*, vol. 36, no. 4, pp. 352–366, 2025, doi: 10.62273/QOIR5976.
8. Domingues and D. Rasteiro, "A paradigm shift in teaching machine learning to sustainable city management master students," in *Proceedings of ICETC 2024*. ACM, 2024, doi: 10.1145/3702163.3702448.
9. Руденко Ю.О., Півторайко В.В., Агаджанова С.В. Аналіз агрономічних даних за допомогою Orange Data Mining: інструменти та методи для агробізнесу. *Digital evolution of modern business: challenges, trends and prospects in the digital economy of Ukraine* : collective monograph / edited by A. V. Cherep, I. M. Dashko, Yu. O. Ohrenych, O. H. Cherep. Riga, Latvia : Baltija Publishing, 2025. С.36-48 <http://www.baltijapublishing.lv/omp/index.php/bp/catalog/view/627/16621/35207-1>
10. Wilk J. Investigation of the applicability of single cantilever beam test for the evaluation of fracture toughness of sandwich composites. *Engineering Fracture Mechanics*. 2023. P. 109075. URL: <https://doi.org/10.1016/j.engfracmech.2023.109075>
11. Державна статистика. Мережа та діяльність закладів освіти <https://stat.gov.ua/uk/datasets/merezha-ta-diyalnist-zakladiv-osvity>
12. Unicef. Progress on household drinking water, sanitation and hygiene 2000-2024: Special focus on inequalities <https://data.unicef.org/resources/jmp-report-2025/>

Rudenko Yu.O., Ahadzhanova S.V., Vyunenko O. B., Sytnyk L.H., Ahadzhanov-Honsales K.H.
EVALUATION OF THE EFFECTIVENESS OF MASTERING LOW-CODE TECHNOLOGIES
IN THE TRAINING OF SPECIALISTS FOR BIG DATA ANALYTICS

The article is devoted to a comparative evaluation of the effectiveness of students' mastery of low-code technologies in the process of preparation for big data analytics. The relevance of the study is обусловено by the growing role of simplified development tools in the educational process and the need for a well-grounded selection of analytical platforms to develop the professional competencies of future specialists in information systems and technologies. The aim of the study is an empirical comparison of the effectiveness of using Orange Data Mining, Weka, and KNIME under identical learning conditions.

The research was conducted in the format of a pedagogical experiment involving 131 students divided into three independent groups according to the selected platform. The comparison was carried out using a system of objective (task completion time, number of typical errors, quality of the constructed models) and subjective (self-assessment of interface complexity) metrics. The nonparametric Kruskal–Wallis test was applied to verify the statistical significance of the differences.

The results demonstrated statistically significant differences in task completion time and the number of typical errors ($p < 0,05$). The best results were shown by Orange Data Mining, which was characterized by fewer errors, higher task completion speed, and better subjective usability assessment. At the same time, the quality of the constructed analytical models was comparable across all groups and showed no statistically significant differences. The obtained results confirm that the choice of a low-code platform primarily affects the procedural characteristics of learning activities but does not determine the final level of analytical outcomes provided that appropriate methodological support is ensured.

Practical recommendations are formulated for the step-by-step integration of low-code tools into big data analytics courses depending on the students' level of preparation.

Keywords: Big Data, low-code, Orange Data Mining, Weka, KNIME, data analytics.

Дата першого надходження статті до видання: 13.03.2026
Дата прийняття статті до друку після рецензування: 08.04.2026
Дата публікації (оприлюднення) статті 11.05.2026

Samoilov O.O.<https://orcid.org/0009-0002-6711-4769>

Kherson State Maritime Academy

Doshchenko H.G.<https://orcid.org/0000-0002-1004-4934>

Kherson State Maritime Academy

ARTIFICIAL INTELLIGENCE TECHNIQUES FOR SHIP ELECTRICAL SYSTEM CONDITION MONITORING

The objective of this article is to analyze and organize AI techniques for predictive analytics of a vessel's technical state. The need comes from these types of technical equipment failures, representing over a quarter of all incidents at sea. The scientific novelty in this review is based on the lack of a comprehensive analysis of the best neural network architectures for predicting the failures of such critical nodes within the maritime industry, for example, including diesel generators and transformers. We consider how deep learning architectures evolved, with an increasing importance to consider how attention is being used to analyze complex time series, from simple Recurrent networks to state-of-the-art transformers. Specifically, we compare the Informer, PatchTST, and Autoformer models, emphasizing that the latter have a 38% enhancement in forecasting accuracy achieved by the addition of trend and seasonality decomposition blocks. The merits of neural network architectures (LSTMs) compared to classical statistical-based models (ARIMA) for modeling high-dimensional data in a precise maritime environment, such as maritime data, were highlighted. Finally, we introduce a new method of diesel engine fault detection based on SqueezeNet and IVY-RF algorithms. The Random Forest ensemble method has an accuracy for dissolved gas analysis (95.2%-100%) in transformer diagnostics. We further describe the employment of multi-task multilayer task networks MTL-DNN for electrical network protection against arcing and a method of analyzing circuit breaker fault contact wear without disassembly. The results are verified using testing. In summary, as a result of the present limitations, the future of marine predictive analytics is to create hybrid systems that will utilize the dynamic aspect that the machine learning method provides and the subject-matter knowledge that the user of the expert system has, in order to enhance safety and competitiveness in shipping.

Keywords: artificial intelligence, predictive analytics, Shipping 4.0, ship electrical equipment, fault diagnosis, technical condition monitoring, deep learning, neural networks.

Formulation of the problem. Over 80% of global cargo movements today are through shipping, and systems operating during shipping are essential for uninterrupted production. In 2024, as reported by EMSA [1], the breakdown of technical equipment accounts for 25.3% of all maritime accidents. Traditional maintenance methods – reactive and planned preventive – often fail, leading to expensive or variable blackouts as a result. Modern electrical equipment takes on many complexities, and certain stresses encountered in the marine environment (vibration, humidity, salt fog) result in the sudden requirement of automatic and predictive analysis to detect early-stage component degradation.

Analysis of recent research and publications. Recent research and publications are analyzed.

Marine predictive analytics research is littered with comprehensive reviews providing a systematic experience with the use of digital technologies. This chapter specifically suggests in [2] that the development of integrated monitoring systems is essential when considering predictive maintenance, yet it lacks unified approaches when considering the complex nature of ship power systems. The current debate of applying modern architectures when studying time series is described in detail in the review of large models [3], emphasizing the advantage of transformer architectures over recurrent networks in tasks of long-term forecasting, crucial for planning repair work planned on the voyage. While dedicated literature reviews [4] of marine internal combustion engines indicate

a trend to leverage machine learning methods for fault detection, they pointed to the dark box aspects of such deep neural networks as an important source of low interpretability. At the same time, research in the field of offshore energy [5] describes interesting wind turbine monitoring installations in wind turbine applications in combination with their implementation into hybrid ship systems. It is also worth mentioning that the work on AI-based techniques to localize damage in distribution networks through AI methods [6]. While there have been many publications, the majority of reviews have either not considered maritime-specific considerations or have only focused on general artificial intelligence algorithms, or have been specific to the ship nodes. The article seeks to contribute to bridging this gap with an extensive review of intelligent diagnostics technology for ship electrical systems.

Task statement. The purpose of this work is a comprehensive analysis and systematization of artificial intelligence methods for predictive analytics of the condition of ship electrical equipment. The study aims to review the most effective neural network architectures for predicting failures of critical components, such as diesel generators and transformers, and to justify the transition to integrated decision-making systems in the maritime industry.

Outline of the primary material of the study. Artificial intelligence is a large domain in computer science that centers its efforts around creating agents that allow them to perform tasks that were previously considered only as human intelligence. These include, for instance, learning, problem solving, perception, language understanding, and decision-making. The modern architecture of intelligent systems is built on a three-level hierarchy, where artificial intelligence serves as a fundamental concept to replicate human cognitive functions. The next level is represented by machine learning (ML), which refers to an algorithm's ability to identify latent patterns in data without explicit programming rules independently. The highest degree of specialization is provided by deep learning (DL), which, using multilayer artificial neural networks, implements mechanisms for detecting complex hierarchical patterns in large arrays of unstructured information [7]

Machine learning is the core of modern "statistical" AI. It differs fundamentally from "symbolic" AI (such as expert systems) in that rules of behavior are not programmed manually, but are derived (learned) directly from data. ML is divided into three main paradigms: supervised, unsupervised, and reinforcement learning. In supervised learning, a model uses labeled

data to perform classification (category prediction) and regression (numerical value prediction). In the case of unsupervised learning, the model uses unlabeled data to cluster things or to reduce their dimensionality (e.g., via PCA). Reinforcement learning involves an agent interacting by trial and error with the environment to maximize reward in control and optimization tasks. [7]

Deep learning is realized by implementing multilayer artificial neural networks (ANNs), the biggest benefit being the automated extraction of relevant features directly from the original datasets. This eliminates the subjectivity and laboriousness of manual feature engineering typical of classical machine learning. The modern typology of DL architectures includes convolutional networks (CNNs), which are deterministic and used for processing spatial grid structures in computer vision systems; recurrent networks (RNNs), which model complex time sequences; and generative adversarial networks (GANs), which focus on synthesizing high-fidelity data. [7]

The technological vanguard of the industry comprises third-generation neuromorphic networks (SNNs), which reproduce the dynamics of biological neurons, and hypernetworks, which, through the meta-learning paradigm, ensure high system adaptability and enable effective compression for operation on mobile devices. [8]

NLP is a branch of AI that enables machines to understand, interpret, analyze, and generate human language (text or speech). A large amount of management data consists of unstructured text (reports, emails, protocols, regulatory documents). Classic methods include syntactic and semantic analyses, such as bag-of-words. Modern techniques based on DL can be used for Sentiment Analysis – assessing the emotional coloring of the text (positive, negative, neutral); large language models (LLM – like ChatGPT) understand complex contexts, respond to queries, summarize meaningful texts, and generate new ones. [9]

Computer vision (CV) provides automated interpretation of visual information and has undergone a technological transformation enabled by deep learning algorithms. The functional spectrum of CV includes image classification for object identification; object detection, which involves simultaneous identification and localization; and semantic segmentation for pixel-by-pixel labeling of visual data, all of which are critical for autonomous navigation and obstacle recognition systems. [10]

Expert systems (ES) are a symbolic AI paradigm in which decisions are based on deterministic rules

rather than learning facts from statistical training. The ES framework consists of one part: a knowledge base (a collection of “if-then” heuristics), and the other part: a logical inference mechanism that runs off the input data. Autonomous agents (AA) serve as software or physical systems capable of interacting independently with the environment to implement goal functions. Alternatively, agent-based modeling reflects the dynamics of complex systems where a number of independent agents act together. [11]

The taxonomy described illustrates a fundamental shift in approaches. Classical systems analysis was well combined with symbolic AI and expert systems, since both required explicit, manual descriptions of rules and structures. However, for the study of complex “soft” management systems, where the rules are unknown or too complex, statistical AI (ML/DL) that learns these rules from data is an incomparably more powerful tool.

AI is most valuable when no single technology is used to analyze an incident, but only when multiple technologies and other systems are synergistically integrated within a single systems-analysis loop. For example, let’s review the complete automated incident analysis cycle:

1. NLP [9] automatically collects and structures operator reports, conversation recordings, and data from log files.
2. Data Mining Algorithms [11] analyze this structured data and find hidden correlations.
3. The ML model [7] builds a risk model from these patterns that predicts the probability of an incident occurring in real time.
4. The system automatically generates notifications and suggests specific corrective actions to the manager when the predicted risk exceeds the threshold.

Thus, AI enables the creation of an end-to-end, automated, and continuous system-analysis cycle that runs in parallel with the real management process.

The fundamental problem with traditional approaches, such as ARIMA (Autoregressive Integrated Moving Average), is that they suffer from the inability to learn the high-dimensional data and the time dependencies in ambiguous situations. In 2024, the studies clearly demonstrated the superiority of neural network-based approaches [12]. In particular, the efficiency of long short-term memory (LSTM) architectures in time-series forecasting tasks was assessed in comparison to the classical ARIMA model, with 85.1% accuracy, and ensemble methods like Random Forest (87.4%) at 89.3%. Moreover, LSTM scores much lower on errors (MAE=2.7,

RMSE=3.8), which makes these models absolutely indispensable when the total cost of errors is high.

In time series analysis, a combination of convolutional neural networks (CNNs) and long short-term memory networks has a synergistic effect. In this setup, the CNN's performance is efficient as a feature extractor for local patterns and invariant structures that have been extracted from raw input data (e.g., sensor spectrograms or multidimensional financial indicators). The resulting feature vectors are fed to LSTM layers that model time dependencies and long-range dynamics.

Most traditional approaches treat time series as a discrete data stream that does not account for the correlations between sources. Graph neural networks (GNNs) can solve this issue by learning the data as a graph and mapping it in such a way that the nodes represent entities (such as sensors, customers, warehouses) and the edges are interactions between them. Using heterogeneous GNNs has been successfully implemented for fraud detection in supply chain finance, which seeks to study complex transactional relationships between counterparties. GNNs allow for the collection of data from distributed sensors while maintaining network topology, and this is a very crucial issue because of the need for load prediction and energy saving optimization in industrial systems.

The Transformer architecture, originally made for natural language processing, has come a long way in the past few years for time series data, arguably the most. Transformers follow a self-attention mechanism; however, unlike RNNs (recurrent neural networks), which usually treat the same data in a sequential manner, they can experience vanishing gradients on long sequences. This makes it possible for the model to treat the relations among all time series' variables, at all distances, concurrently. [7] The development of transformer architecture has led to the creation of a series of specialized models, each with its own unique challenges for computational power and accuracy, as shown in Table 1.

The Autoformer model demonstrates an explicit decomposition and decomposes non-stationarity (the variation in the statistical properties of a series over time). This is a notable milestone in comparison with the “black boxes” of traditional transformers. [13]

Shipping 4.0 is a period of significant transformation of the shipping industry, responsible for more than 80% of the world's commercial cargo transport. This process has been characterized by the integration of computer-based systems in ships with IoT and artificial intelligence. The 2024 report from the United Nations Conference on Trade and Develop-

Table 1

Transformer models architectures

Model	Architectural feature and mechanism	Advantages and scope of application	Limitations and challenges
Informer	Uses the ProbSparse mechanism Self-Attention, which selects only the most significant queries based on the KL-divergence metric, reducing the complexity from $O(n^2)$ to $O(n \log(n))$. [13]	Very efficient for ultra-long time series. That reduces memory requirements and greatly speeds inference.	In high-volatility data, sparse attention leads to loss of local detail.
Autoformer	It directly integrates the Decomposition Block into the architecture. It separates the series into trend and seasonality by using autocorrelation rather than conventional attention. [13]	Provides a 38% increase in long-term forecasting accuracy compared to analogues. Best for highly complex and non-stationary processes.	Greater architectural complexity complicates interpretations of the model's internal weights.
PatchTST	Time-series segmentation is implemented for patches (subseries patches) that are processed as discrete tokens, preserving semantic locality. [14]	The model helps in better capturing local patterns by adding fewer tokens, thus reducing computational cost.	Requires careful tuning of patch size and step hyperparameters.
Pyraformer	Implements a pyramidal attention graph for multiscale dependency and multiscale modeling. [14]	It captures dependencies with effective and linear complexity over short and long timescales.	There are fewer general libraries used for standard libraries, making it difficult to implement quickly.

ment (UNCTAD) states that last year, over 12 million cargo movements took place, and a steady year-on-year growth rate of 2.4% is estimated for the next 10 years up until 2029. This dynamic increase in shipping intensity imposes an increased burden on shipboard technologies, particularly electrical power systems, which play a vital role in guaranteeing the survival, navigation, and safety of the ship. Conventional maintenance methods, like reactive maintenance (maintenance after the accident) and planned preventive maintenance with predetermined intervals to maintain, are no longer very useful to the ever-more intricate ship equipment. They frequently result in either too much maintenance of workable parts - and a raised production and maintenance cost - or a sudden emergency breakdown that endangers crew members and the cargo. [2] Accordingly, a shift to predictive maintenance with machine learning algorithms is more than just progress—it is also a prerequisite for keeping up with environmental standards and remaining competitive. Knowledge of failures (types and prevalence) is the cornerstone of building effective machine learning models. We need some deep historical analysis to solve these problems; otherwise, we will never be able to train an algorithm capable of learning patterns or failures that precede such things. The findings of system and apparatus failures are well ahead of the others, according to the annual review report from the European Maritime Safety Agency (EMSA) entitled “Annual Overview of Marine Casualties and Incidents 2024.” In the years 2014 to 2022, equipment failures accounted for 25.3% of maritime incidents, second only to those due to navigation, but

mainly through vehicular causes [1]. The statistics of failure of dedicated ships are of particular relevance. Analysis of operations of research vessels showed that 76.3% of surveyed specialists took part in expeditions that were broken from measurement and related electrical equipment [15]. This proves that even on vessels with a competent crew and up-to-date equipment, the reliability issue is not solved. The risk analysis established 12 overarching risk factors; despite natural forces, technical ones take precedence, highlighting the importance of implementing automated control systems. Data from offshore and Onshore Reliability reveal that, for underwater control systems, it is most widely reported that the common failure will be a signal failure, whereas electronic module failures should be considered to have the highest criticality ratio. [16] This corresponds to changes in the vulnerability vector from mechanical devices to digital management and information transmission ones.

A study of the reliability of generation systems on offshore patrol vessels revealed a worrying trend. The reliability of a power plant made up of four diesel generators was under 70% for the first 24 months of operation [17]. These results line up with the "bathtub curve," which describes the initial period of operation as a period of higher failure due to installation defects and running-in [18]. Utilization of ANNs found that the threshold load (approximately 200 kVA) dramatically escalates the probability of overheating as well as generator malfunction, allowing optimization of operational modes. The marine environment has unique environmental stresses for electronics, such as vibration, shock, humidity, and salt spray.

There is a need for an in-depth knowledge of these processes in order to interpret sensor data. Research into the impact of shock loads on marine electronics has revealed specific failure mechanisms at the microlevel [19]:

1. Solder materials such as SAC305 exhibit a significant reduction in strength and toughness at high strain rates, typical of wave shocks. This results in rapid initiation and propagation of microcracks in solder joints.

2. Plastic deformation of connectors. Insufficient thickness of the gold coating on the connector contacts can lead to permanent bending of the copper substrate under inertial forces, thereby interrupting the electrical circuit.

3. Stress concentration in packages. The corners of chip packages are sites of mechanical stress concentration, making them susceptible to cracking under storm conditions.

These indicate the necessity of monitoring systems that can detect high-frequency transients that can signal microsecond contact interruptions prior to full system failure. Diesel generators continue to be the main energy source. They are the primary method for diagnosis, usually relying upon thermodynamic parameters and vibration monitoring. But AI allows for a level of accuracy that was not possible before. A novel method to detect faults in marine diesel engines was presented in 2025 [20], which adopted Compressive Sensing and SqueezeNet deep learning models. The dimensionality of vibration data is too big compared to most of the traditional methods. The compressed sensing method allows effective data reduction without any information loss, and SqueezeNet, as a lightweight architecture, provides high performance to maintain high accuracy in fault classification even with varying engine speeds and a smaller base training sample size. This is particularly useful for limited resource-constrained on-board hardware systems. Another method [4] builds recurrence graphs (plots), constructed using vibration signals, and converts a one-dimensional time series into a two-dimensional image. Based on this image, CNN is trained to recognize the various stages of crankshaft bearing wear that could not be accomplished by classical spectral analysis.

Bayesian networks are also used to diagnose thermodynamic processes. A framework, also used in [21], incorporated ML anomaly detectors, EWMA (Exponentially Weighted Moving Average), and Bayesian diagnostic networks. This system permits detection of not just deviations (for instance, for any variation in exhaust-gas temperature rise); it can also

be used to find the cause behind it. For example, the model can distinguish, with a considerable rate of accuracy, if the rise in temperature results from a failure of a fuel injector in a cylinder or from an issue that extends to the entire air supply system (e.g., a charge air cooler contaminated). The IVY-RF algorithm (Random Forest optimized by the ivy algorithm) has demonstrated excellent performance for marine diesel engine fuel systems diagnosis [22]. 12 different fault categories were determined on this model, and 100% accuracy was achieved when using perfect data, even when 5% noise (that is typical of real ship conditions) was applied. The robustness of this method allows IVY-RF to be applied as a promising tool for automating decision-making systems.

The integration of wind turbines and solar panels into ship power systems creates new challenges for diagnostics. In hybrid systems, power converters are critical components. Failure of IGBT transistors can result in an instantaneous loss of generation. To predict such shortcomings, LSTM networks [5] have been proposed. LSTM networks analyze time-series data of three-phase currents and voltages, detecting subtle changes in signal shape that indicate thermal aging of the semiconductor crystal or degradation of the gate insulation.

Ship's transformers are exposed to significant vibration and thermal loads. Their diagnosis is mainly made based on dissolved gas analysis (DGA) and vibration monitoring. Conventional DGA (Duval triangle, Rogers ratio) interpretation methods often fail to make diagnoses on those problems that have multiple defects at the same time. Machine-learning algorithms like Random Forest (RF) and Decision Tree (DT) improve the accuracy significantly. In a study with 628 lubricant samples, the Random Forest model achieved diagnostic accuracies ranging from 95.2% to 100% on different types of thermal and discharge defects [23]. A second method, using correlation analysis and density-based spatial clustering of applications with noise (DBSCAN) to extrapolate DGA data, significantly increased the accuracy in fault clustering by 32.7% compared to the direct method [24]. This allows for the detection of defect initiation (e.g., local oil overheating) well in advance of gas protection.

Recent research suggests using the Transformer architecture and the Attention mechanism for transformer housing vibration analysis [25]. Mechanical deformation of the windings or weakening of the magnetic core layer changes the vibration spectrum. Deep learning models can automatically extract these features from the "raw" data, ignore the background

noise of the vessel, and classify the transformer state with high accuracy, even without prior expert signal processing.

Short circuits and arcing faults in switchgear are among the most dangerous accidents. Traditional current protections are often insensitive to high-resistance arc faults. Artificial intelligence enables the implementation of protection based on arc pattern recognition. In [26], the MTL-DNN (Multi-Task Learning Deep Neural Network) simultaneously addresses the tasks of accident detection, damage type classification (short circuit, overload, ground fault), and situation development prediction. The accuracy of diagnosing overloads and short circuits reaches 96.5% and 94.3%, respectively.

Fault localization is also essential. Algorithms based on traveling waves and condition assessment enable accurate determination of the location of a breakdown in a cable or busbar trunking, thereby minimizing search and repair time. [6]

Failure of the circuit breaker during a short circuit can result in a complete loss of power to the vessel. The leading causes of failure are contact wear and degradation of the drive mechanism. In [27], methods for assessing contact wear are compared. The "arc energy integral" method is commonly used but inaccurate due to the assumption of a constant arc-current shape. A more promising method is to measure dynamic contact resistance during the switching operation. Analysis of the DCR curve using ML algorithms allows for the estimation of the contact thickness and the degree of its erosion without disassembling the arc extinguishing chamber.

The current shape of the control solenoid coil (Trip/Close coil profile) analysis is a useful tool to diagnose mechanical components. Delays in armature movement and changes in the shape of current peaks indicate a sticking mechanism, spring weakening, or lubrication problems. This type of mechanical defect detection occurs through hybrid classifiers (SVM +

ELM), which can be used to determine such defects at an early stage with high accuracy [28]. Most ships still operate magnetron radars. Magnetrons have a limited lifespan and often fail suddenly. AI enables the prediction of the magnetron's remaining lifespan and planning to replace it before failure by measuring the emitted pulse spectrum and frequency stability. [29]

Conclusions. This research systematizes artificial intelligence methods for specific tasks in the diagnostics of ship equipment. Unlike existing general reviews, the work provides a cross-analysis of the physical nature of input signals such as vibration spectra, thermal parameters, oscillograms of arc processes, and the mathematical potential of modern neural network architectures within the Shipping 4.0 concept.

The research confirms that artificial intelligence has evolved from a theoretical concept into an effective tool for improving the reliability of ship equipment. Deep learning methods, specifically CNNs, LSTMs, and various modern Transformer-based ones, have outperformed the traditional methods in terms of technology development by offering a much higher prediction accuracy in marine systems with high noise and nonlinearity. Specialized solutions such as the Autoformer model can also achieve robustness to process non-stationarity. The outcome data tests with algorithms like IVY-RF for diesel plant diagnostics or MTL-DNN for electrical system protection have very high robustness and the detection of very complex fault types with nearly 100% precision. The integration of NLP technologies for unstructured documentation processing and expert systems for decision support creates a dynamic system analysis cycle that can be automated in real time. The ultimate idea for development can be that of hybrid systems, which offer a good balance between the practical aspects (physical models based on expert knowledge) and the freedom (flexible machine learning methods) to lower risk and keep maritime safety on the edge.

Bibliography:

1. Cicek K. Marine engineering systems failure analysis using extended failure modes and effects analysis with quality function deployment: maritime house of reliability. *International Journal of Quality; Reliability Management*, 2025. P. 1–23. URL: <https://doi.org/10.1108/ijqrm-05-2025-0157> (date of access: 13.01.2026)
2. Kalafatelis A. S., Nomikos N., Giannopoulos A., Alexandridis G., Karditsa, A., Trakadas Towards Predictive Maintenance in the Maritime Industry: A Component-Based Overview. *Journal of Marine Science and Engineering*. 2025. Vol. 13(3), P. 425. URL: <https://doi.org/10.3390/jmse13030425> (date of access: 13.01.2026)
3. Kottapalli, Siva, Hubli, Karthik, Chandrashekhara, Sandeep, Jain, Garima, Hubli, Sunayana, Botla, Gayathri, Doddaiiah, Ramesh. Foundation Models for Time Series: A Survey. 2025. URL: <https://doi.org/10.48550/arXiv.2504.04011> (date of access: 13.01.2026)
4. Srinivaas A., Sakthivel N. R., Nair B. B. Machine Learning Approaches for Fault Detection in Internal Combustion Engines: A Review and Experimental Investigation. *Informatics*. 2025. Vol. 12(1), P. 25. URL: <https://doi.org/10.3390/informatics12010025> (date of access: 13.01.2026)

5. Kou L., Li, Y., Zhang F., Gong X., Hu Y., Yuan Q., Ke W. Review on Monitoring, Operation and Maintenance of Smart Offshore Wind Farms. *Sensors (Basel, Switzerland)*. 2022. Vol. 22(8), P. 2822. URL: <https://doi.org/10.3390/s22082822> (date of access: 13.01.2026)
6. Rezapour H., Jamali S., Bahmanyar A. Review on Artificial Intelligence-Based Fault Location Methods in Power Distribution Networks. *Energies*. 2023. Vol. 16(12), P. 4636. URL: <https://doi.org/10.3390/en16124636> (date of access: 13.01.2026)
7. Zhang S. Artificial Intelligence and Applications in Structural and Material Engineering. *Highlights in Science, Engineering and Technology*. 2023. Vol. 75, P. 240-245. URL: <https://doi.org/10.54097/9qknfc57> (date of access: 13.01.2026)
8. Chauhan V.K., Zhou J., Lu P. et al. A brief review of hypernetworks in deep learning. *Artif Intell.* 2024. Rev 57. P. 250 URL: <https://doi.org/10.1007/s10462-024-10862-8> (date of access: 13.01.2026)
9. Domrös-Zoungrana D., Rajaeen N., Boie, S., Fröling E., Lenz C. Medical Education: Considerations for a Successful Integration of Learning with and Learning about AI. *Journal of Medical Education and Curricular Development*, 2024. P. 11. URL: <https://doi.org/10.1177/23821205241284719> (date of access: 13.01.2026)
10. Hassaballah M., Awad A.I. (Eds.). Deep Learning in Computer Vision: Principles and Applications (1st ed.). *CRC Press*. 2020. URL: <https://doi.org/10.1201/9781351003827> (date of access: 13.01.2026)
11. Namdarian L. Towards AI-enabled evidence-based policymaking in science and technology: a conceptual framework. *Technology Analysis & Strategic Management*. 2025. P. 1–16. URL: <https://doi.org/10.1080/09537325.2025.2486003> (date of access: 13.01.2026)
12. Nikitha D., Imran Hussain S., Poornima M. Integrating Power BI with Machine Learning Models for Predictive Analytics. *2024 7th International Conference on Contemporary Computing and Informatics (IC3I)*. 2024. P. 1245–1250. URL: <https://doi.org/10.1109/ic3i61595.2024.10828685> (date of access: 13.01.2026)
13. Arnab Majumdar. Transformers for Time Series Forecasting. 2025. URL: <https://medium.com/@serana.ai/transformers-for-time-series-forecasting-e5e0327e78be> (date of access: 13.01.2026)
14. Oliveira J. M., Ramos P. Evaluating the Effectiveness of Time Series Transformers for Demand Forecasting in Retail. *Mathematics*. 2024. Vol. 12(17), P. 2728. URL: <https://doi.org/10.3390/math12172728> (date of access: 13.01.2026)
15. Kubacka M., Gajewski L., Burchacz M., Matczak M., Janowski P., Piotrowicz J. Many faces of survey equipment failures during marine research at sea—Risk analysis. *PLOS ONE*. 2022. Vol. 17(8), P. e0272960. URL: <https://doi.org/10.1371/journal.pone.0272960> (date of access: 13.01.2026)
16. Mahmoudi J. SIL analysis of subsea control system components based on a typical OREDA database. *Quality and Reliability Engineering International*. 2021. Vol. 37(8). P. 3297–3313. Portico. URL: <https://doi.org/10.1002/qre.2909> (date of access: 13.01.2026)
17. Daya A. A., Lazakis I. Component Criticality Analysis for Improved Ship Machinery Reliability. *Machines*. 2023. Vol. 11(7), P. 737. URL: <https://doi.org/10.3390/machines11070737> (date of access: 13.01.2026)
18. BĂJENESCU T. M. Reliability and Failures of Electronic Components. *Systems. Electrotehnica, Electronica, Automatica*. 2021. Vol. 69(2), P. 76–82. URL: <https://doi.org/10.46904/eea.21.69.2.1108009> (date of access: 13.01.2026)
19. Lang F., Zhou Z., Liu J., Cui M., Zhang Z. Review on the impact of marine environment on the reliability of electronic packaging materials. *Frontiers in Materials*. 2025. P. 12. URL: <https://doi.org/10.3389/fmats.2025.1584349> (date of access: 13.01.2026)
20. Radić J., Šarić M., Rubić A. Spectral-Based Fault Detection Method in Marine Diesel Engine Operation. *Sensors (Basel, Switzerland)*. 2025. Vol. 25(18), P. 5669. URL: <https://doi.org/10.3390/s25185669> (date of access: 13.01.2026)
21. Cheliotis M., Lazakis I., Cheliotis A. Bayesian and machine learning-based fault detection and diagnostics for marine applications. *Ships and Offshore Structures*. 2022. Vol. 17(12), P. 2686–2698. URL: <https://doi.org/10.1080/17445302.2021.2012015> (date of access: 13.01.2026)
22. Ouyang H., Li W., Gao F., Huang K., Xiao P. Research on Fault Diagnosis of Ship Diesel Generator System Based on IVY-RF. *Energies*. 2024. Vol. 17(22), P. 5799. URL: <https://doi.org/10.3390/en17225799> (date of access: 13.01.2026)
23. Al-Sakini S. R., Bilal G. A., Sadiq A. T., Al-Maliki W. A. K. Dissolved Gas Analysis for Fault Prediction in Power Transformers Using Machine Learning Techniques. *Applied Sciences*, 2025. Vol. 15(1), P. 118. URL: <https://doi.org/10.3390/app15010118> (date of access: 13.01.2026)
24. Liu Y., Song B., Wang L., Gao J., Xu R. Power Transformer Fault Diagnosis Based on Dissolved Gas Analysis by Correlation Coefficient-DBSCAN. *Applied Sciences*. 2020. Vol. 10(13), P. 4440. URL: <https://doi.org/10.3390/app10134440> (date of access: 13.01.2026)
25. Chen M., Gan H., Wu H. Research on Fault Diagnosis Method for Marine Diesel Engines Based on Multi-Scale Attention Mechanism Transformer. *Journal of Marine Science and Engineering*. 2024. Vol. 12(12), P. 2348. URL: <https://doi.org/10.3390/jmse12122348> (date of access: 13.01.2026)

26. Wan S., Kong Z. Fault diagnosis and prediction model of ship electrical system based on deep learning. *International Conference on Electrical Engineering and Intelligent Systems (IC2EIS 2025)*. 2025. P. 128. URL: <https://doi.org/10.1117/12.3069269> (date of access: 13.01.2026)
27. Li L., Wang B., Liu Y., Yu H., Zhang S., Huang M. Early Detection of Health Condition Degradation of Circuit Breaker Based on Electrical Quantity Monitoring. *Energies*. 2023. Vol. 16(14), P. 5581. URL: <https://doi.org/10.3390/en16145581> (date of access: 13.01.2026)
28. Sun T., Geng Y., Chen C., Yu Q., Yang L., Lu T. Static and dynamic contact performance of marine circuit breakers consider jolt vibration. *Electrical Engineering*. 2023. Vol. 106(3), P. 2785–2798. URL: <https://doi.org/10.1007/s00202-023-02110-z> (date of access: 13.01.2026)
29. King S. C., Tougher B., Zetterlind V. Estimating Speed Error of Commercial Radar Tracking to Inform Whale-Ship Strike Mitigation Efforts. *Sensors (Basel, Switzerland)*, 2025. Vol. 25(6), P. 1676. URL: <https://doi.org/10.3390/s25061676> (date of access: 13.01.2026)

Самойлов О.О. Дощенко Г.Г. МЕТОДИ ШТУЧОГО ІНТЕЛЕКТУ ДЛЯ ПРЕДИКТИВНОЇ АНАЛІТИКИ СТАНУ СУДНОВОГО ЕЛЕКТРОУСТАТКУВАННЯ

Метою цієї статті є аналіз та систематизація методів штучного інтелекту для прогнозовної аналітики технічного стану судна. Потреба виникає через такі типи відмов технічного обладнання, які становлять понад чверть усіх інцидентів у морі. Наукова новизна цього огляду ґрунтується на відсутності комплексного аналізу найкращих архітектур нейронних мереж для прогнозування відмов таких критичних вузлів у морській галузі, наприклад, включаючи дизель-генератори та трансформатори. Ми розглядаємо, як розвивалися архітектури глибокого навчання, при цьому зростає важливість розгляду того, як увага використовується для аналізу складних часових рядів, від простих рекурентних мереж до сучасних трансформаторів. Зокрема, ми порівнюємо моделі *Informer*, *PatchTST* та *Autoformer*, підкреслюючи, що останні мають 38% підвищення точності прогнозування, досягнуте завдяки додаванню блоків декомпозиції трендів та сезонності. Було виділено переваги архітектур нейронних мереж (LSTM) порівняно з класичними статистичними моделями (ARIMA) для моделювання високовимірних даних у точному морському середовищі, такому як морські дані. Нарешті, ми представляємо новий метод виявлення несправностей дизельного двигуна на основі алгоритмів *SqueezeNet* та *IVY-RF*. Метод ансамблю випадкового лісу має точність аналізу розчиненого газу (95,2%-100%) у діагностиці трансформаторів. Далі ми описуємо використання багатозадачних багатопарових мереж завдань MTL-DNN для захисту електричної мережі від дугового утворення та метод аналізу зносу контактів вимикача без розбирання. Результати перевіряються за допомогою тестування. Підсумовуючи, внаслідок існуючих обмежень, майбутнє морської прогнозовної аналітики полягає у створенні гібридних систем, які використовуватимуть динамічний аспект, що забезпечує метод машинного навчання, та предметні знання, які має користувач експертної системи, з метою підвищення безпеки та конкурентоспроможності судноплавства.

Ключові слова: штучний інтелект, предиктивна аналітика, *Shipping 4.0*, суднове електроустаткування, діагностика несправностей, моніторинг технічного стану, глибоке навчання, нейронні мережі.

Дата першого надходження статті до видання: 15.01.2026
Дата прийняття статті до друку після рецензування: 18.02.2026
Дата публікації (оприлюднення) статті 11.05.2026

Ushakova I.O.<https://orcid.org/0000-0001-8315-0917>

Simon Kuznets Kharkiv National University of Economics

MODERNIZATION OF THE BACKEND OF THE WEB APPLICATION BASED ON MICROSERVICE ARCHITECTURE

The article considers the problem of inefficiency of web applications built on the basis of monolithic architecture, in particular in terms of speed, scalability and support. The relevance of the study is due to the increasing load on information systems and the need to ensure high performance when working with a large volume of user requests. The purpose of the work is to justify the feasibility of transitioning from a monolithic architecture of the backend of the web application to a microservice architecture and increasing the efficiency of the system through its structural decomposition. In the process of the study, an analysis of the existing architecture of the application was carried out, its logical and functional components were determined, which allowed for the decomposition of modules and the transition to separate independent services. The main functional microservices were identified, which are responsible for working with users, authorization, processing text, graphic and video data, working with vacancies and resumes, search engines, social networks, payment systems, mail and SMS services, as well as a security and administration module. The proposed architectural solution allowed to ensure independent scaling of individual services, simplify the process of their modernization and optimize the distribution of computing resources. The study contains a comparative analysis of the performance indicators of the web application before modernization and after the implementation of the microservice architecture. The results of testing the functions of administration, search and integration with social networks are presented, which indicate a significant improvement in the system response time. It is shown that the average increase in speed is from several times to several dozen times depending on the type of request. Special attention is paid to the optimization of the cluster infrastructure by redistributing resources and integrating specialized data storage and processing services. The results obtained confirm the effectiveness of using the microservice architecture to improve the server part of web applications and can be used in the development and modernization of complex information systems focused on high load, flexibility and stability of operation.

Keywords: web application, backend, monolithic architecture, microservice architecture, structural decomposition, cluster infrastructure, performance testing.

Formulation of the problem. Software architecture plays a key role in the software development process. Software architecture approaches are principles for organizing the structure of a system to ensure its reliability, scalability, and ease of maintenance. The choice of architecture primarily depends on the scale of the project: monolithic architecture is better for small applications and is suitable for small businesses, while microservices are more suitable for large, complex systems. Also, when choosing an architectural approach, the requirements for system flexibility, constraints on the duration of development, the qualifications of the development team, and the scale of the project are taken into account.

When choosing the right approach, software developers often do not take into account future risks or problems that they may encounter in the future.

It is the problem of underestimating architectural solutions that can play a key role in the quality of a software product. Therefore, the problem of choosing the right approach to defining software architecture is relevant and requires a comprehensive analysis and justification of approaches to its solution. In addition, there is the problem of technology obsolescence, when software was developed using outdated technologies and therefore does not meet the requirements of modern technologies. In general, owners of such software applications are most often dissatisfied with the speed of the software, its scalability and support.

Microservice architecture allows you not to be limited to one technology, but to develop software applications using different technologies. This architecture allows you to divide the software product into modules that can be written in different program-

ming languages, which also allows you to minimize future problems, dynamically change development approaches and methods to improve the quality of the software product.

Therefore, development using microservice architecture is very promising for implementation, this architecture has been most actively developing recently in the field of software development. The main advantages of such a solution, compared to monolithic architecture and serverless, are the savings of resources, both human and material – during the development process, and computational – during the scaling process. However, it is also worth considering the difficulties that arise when using microservices, mainly related to the complication of transaction mechanisms, as well as the need to implement systems for automated monitoring and logging of program operation. Therefore, the microservice architecture of software development itself was chosen for further development.

The purpose of the work is to justify the feasibility of transitioning from a monolithic architecture of the backend of a web application to a microservice architecture and to increase the efficiency of the system's functioning through its structural decomposition. This will contribute to increasing the speed of the application, reducing the cost of technical resources and expanding their functional capabilities. To achieve the goal, the following tasks must be solved in the work: to analyze the issues of modern server-side software architectures; to review and analyze existing methods for building software architecture; to analyze the functionality and perform structural decomposition of the web application; to improve the web application based on microservice architecture and to analyze the effectiveness of the new architectural solution.

Analysis the last research and publications. The problem of choosing the right approach to building a web application architecture. A significant number of scientific works are devoted to it. It is relevant and requires a comprehensive analysis and development of approaches to its solution. It is noted that microservices architecture is fundamentally different from traditional monolithic architecture, where a software system essentially consists of a single deployable system that encompasses all the functions of the software system [3, 5–7, 11]. This important distinction makes microservices modular, independently scalable, and technologically diverse compared to monoliths.

In the work of Järvinen K. emphasizes that the main principles of microservices include decomposition into bounded contexts, independent deploy-

ment, and decentralized management. Explicitly limiting responsibilities within clear boundaries makes microservices flexible, easy to maintain, and allows teams to work independently of other teams without excessive communication. The author explores how the advantage in flexibility of microservices depends on their scalability. The paper uses empirical and quantitative methodology to analyze different levels of scalability of microservices and their impact on performance and costs of cloud technologies [2].

Paper «From Monolithic Systems to Microservices» notes that microservices implementation still faces challenges in terms of time and computational resources, scalability, orchestration, organization, etc. [11]. There are procedures for migrating from monolithic architectures to microservices, but none of them accurately quantify the performance difference. The paper proposes an approach to evaluate performance and the relationship between different application variables to compare monolithic and microservice architectures.

The problem of shortcomings of existing approaches to microservice architecture, especially for systems operating in dynamic real-time conditions, is raised in the article «Microservices Architecture for Real-Time» [5]. In this study, the authors propose a new scalable microservices architecture optimized for real-time data processing using a modular, event-driven design. The paper describes a system that can consume real-time weather information based on data available through the OpenWeatherMap application interface. The system is built using complementary technologies used in modern data processing and microservices architectures, such as Apache Kafka, Apache Flink, Redis, Kubernetes, and based on adaptive autoscaling through KEDA and HPA in the architecture.

Recently, the number of studies on cloud microservices has been increasing. In the paper of Sigala N. S. examines the architecture of microservices from the perspective of practical issues related to the processes of its design, deployment, and management in cloud computing environments [9]. The results presented in this study provide a deep understanding of how to effectively implement and manage microservices architectures to make software solutions more efficient, reliable, and scalable in cloud computing environments.

In the article by Mendonça N. C., Box C., Manolache C., Ryan L. notes that despite the advantages of microservices, their implementation poses some technical and organizational challenges related to opera-

tional complexity, increased technological diversity, and the need for better coordination between teams [4]. The authors describe design decisions, tradeoffs, and lessons learned from the open source service network project Istio, which had a microservices architecture from the beginning. However, less than three years after the first release of Istio, problems arose and the control plane microservices were merged into a monolith.

Different approaches to security can significantly influence architectural decisions. A monolithic system interacts with the external environment as a single unit and can provide simpler access control. The security of a monolith is provided centrally, simplifying protection but creating the risk of a single point of failure. On the other hand, in microservices, each element can have its own security system, which provides more flexible but also more complex control. In the book Siriwardena P., Dias N. addresses the security issues of microservices [10]. The book covers the basics of securing both the application boundary and the communication between services. The authors explore how to deploy and secure microservices, access microservices from a single page, and web application.

Microservices require more investment in development, but this approach is justified in the following cases: high load on the application, which must stably serve thousands of users simultaneously; a long-term and large-scale project that will constantly develop and grow; complex business processes that are more convenient to divide into services and scale independently; large teams that can work in parallel on different services without conflicts in the code. That is, the problem of choosing the right approach to defining the software architecture requires a comprehensive analysis and development of approaches to its solution. To compare the characteristics of a web application built using different architectural approaches, performance testing is performed using various tools [12, 13]. In the work Selivorstova T., Klishch S., Kyrychenko S., Guda A., Ostrovskaya K. Analysis of monolithic and microservice architectures: a statistical analysis of the main quantitative characteristics of monolithic and microservice architectures was performed using the Kubectl tool for managing Kubernetes clusters. [6]. The authors use the Kubernetes cluster deployment methodology for microservice architecture using Minikube, Kubectl, and Docker. The results confirmed that the microservice architecture has much greater fault tolerance and flexibility compared to the monolithic architecture.

Task statement. The main tasks of the work are to improve the architecture of the backend of the software application and analyze the results of its performance. The improved system should be divided into modules that will work independently of each other and interact using the HTTP protocol. The application built on the basis of monolithic architecture will be improved. The basic system is a web application and has two interfaces: for candidates looking for work; for employers looking for candidates for vacant positions. Employers and candidates can post articles and advertisements. The application also has integration with third-party services, namely social networks and payment systems. To monetize the web application, a decision was made to develop special functionality that should interact with many parts of the system. Before the application was modernized, an analysis was conducted of the number of visits and the directions of its use by users. It turned out that the main problem was the speed of the application. This was a consequence of the fact that the web application was developed quite a long time ago using outdated and no longer supported technologies. To solve this problem, the task of changing the monolithic architecture to a microservice one was set.

Outline of the main material of the study

Decisions on improving the software application. Before improving the software application, a complete analysis of the existing system was carried out, which made it possible to divide it into logical and functional modules. The old system included the following services: user service; authorization service; social network service; article service; marketing service; resume and vacancy service; search service; SMS and mail service; payment system service.

After analyzing the existing structure, work was carried out to create new and decompose old modules. The following main functional modules were identified: user management module; authorization module; image management module; video file management module; text document management module; blog article management module; article comment management module; blog post moderation module; resume management module; job search module; resume search module; social media management module; payment system management module; mail service management module; SMS service management module; security module; advertising management module.

The existing web application, in addition to the main system, also had an additional administration panel. This panel also required architectural changes to ensure correct interaction with the main system. From the very beginning, the web application had one

large monolithic service that managed all actions on the administrator panel. This architectural approach worsened the speed of the application; its structural diagram is presented in Fig. 1.

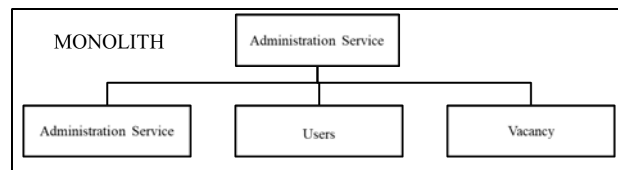


Fig. 1. Structural diagram of the monolithic architecture of the administration panel

As a result of refactoring, the following modules were identified in the system: user administration; vacancy administration; resume administration; search tag administration; statistics viewing. The structural diagram of the changed architecture of the administrator panel is presented in Fig. 2. Decomposition into modules allowed us to encapsulate the logic of the software product, which will speed up the speed of work in the future. Decomposition of the web application into modules allowed us to identify problematic modules, namely the resume and vacancy search module, CRUD queries for working with the administration panel, as well as modules for working with social networks and users. These modules had to be redesigned.



Fig. 2. Structural diagram of the microservice architecture of the administration panel

Thanks to the transition to microservice architecture, it became possible to write modules in different languages, so these modules were rewritten from Java to PHP, which has long proven itself both in terms of greater simplicity and speed of development, and in terms of increased speed of operation of the application itself, written in this language. In addition, the development team had the necessary skills. A module for monetizing the software application was also to be developed. This module should be independent and reusable in another software application. In addition to the software part, the cluster on which the web application was located also needed changes. The cluster consists of two database replications and eight computing servers. The structural diagram of the existing cluster is presented in Fig. 3.

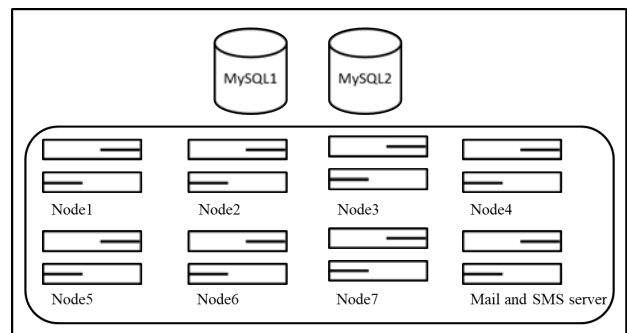


Fig. 3. Structural diagram of the old cluster

The main problem of the software product was the speed of work. Therefore, an analysis of the work of the created modules was carried out. As a result of the analysis, it was found that some modules require a larger amount of resources, and some, on the contrary, scarce resources. Therefore, a decision was made to optimize the cluster structure. Thanks to the redistribution of computing resources between the modules, it was possible to reduce the number of servers to three. Most of the functionality for searching in the web application was transferred to the ElasticSearch search server database, which is often used as a NoSQL database. Elasticsearch is an open-source distributed search and analytical engine built on the basis of the Apache Lucene library. It provides fast full-text search, storage and analysis of large volumes of data (Big Data) in real time using JSON documents. The search server has the following advantages: designed for instant search, scales from a single server to large clusters; has full-text search, which provides efficient search of text information, supports relevance, synonyms and analytics; thanks to the REST architectural style, the API interacts with data using standard HTTP requests, which makes it convenient for developers; it is often used for log analysis and monitoring together with Logstash for data collection and Kibana for visualization. This technology made it possible to store a large amount of data with quick access to it. Services for working with mail and SMS were also allocated. They were transferred to AWS services. The optimized cluster structure is presented in Fig. 4.

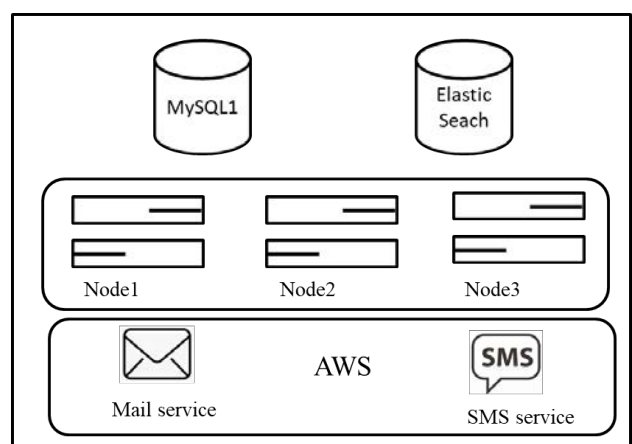


Fig. 4. Optimized cluster structure diagram

Comparative analysis of the effectiveness of a software application for monolithic and microservice architecture. To analyze the effective the microservice approach to web application architecture was tested with server response speed measurements. Fig. 4 shows the response speed measurements of the “Administration” service request in a web application on a monolithic architecture, and Fig. 5 shows measurements of the response speed of the “Administration” microservice request in a web application based on a microservice architecture.

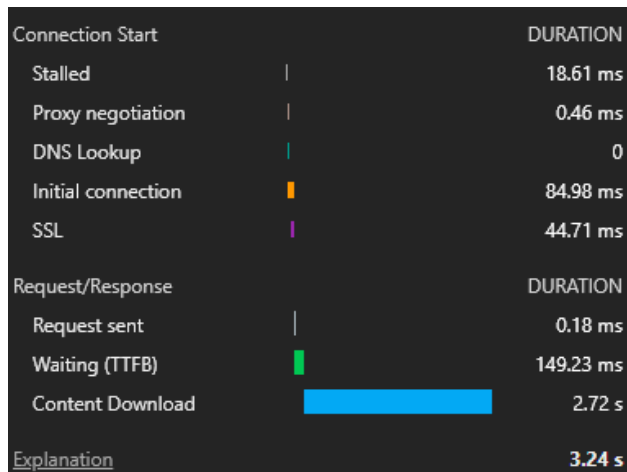


Fig. 5. Measurements of the response speed of the “Administration” service request in a web application on a monolithic architecture

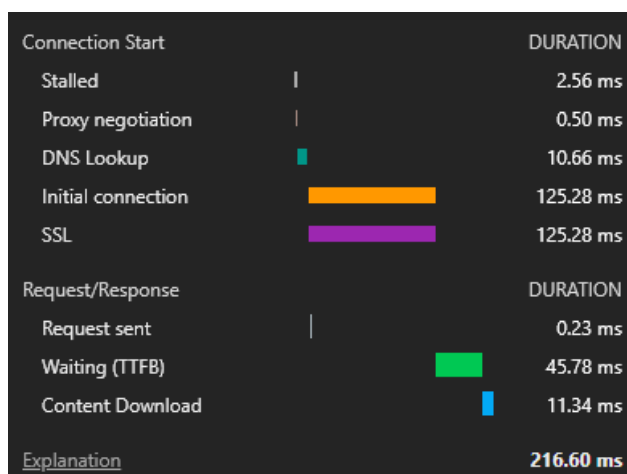


Fig. 6. Measurements of the response speed of the Administration microservice request in a web application on a microservice architecture

A comparative analysis of the results of testing the “Administration” function for monolithic and microservice architectures is presented in Table 1.

Fig. 6 shows measurements of the response speed of the “Vacancy Search” service request in a web application on a monolithic architecture, and Fig. 7 shows measurements of the response speed of the

microservice request “Search for vacancies” in a web application based on a microservice architecture.

Table 1
Results of testing the “Administration” function using different server-side architectures

Indicator	Time, ms		Relative growth
	Monolithic architecture	Microservice architecture	
Resume list page	3410	216	15.79
User list page	2989	197	15.17
Job listing page	3018	203	14.87
Article list page	3704	300	12.34
Statistics page	2050	100	20.5

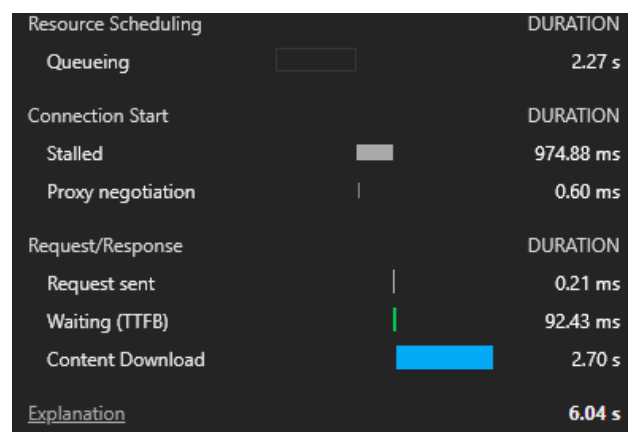


Fig. 7. Measurements of the response speed of the Job Search service request in a web application on a monolithic architecture

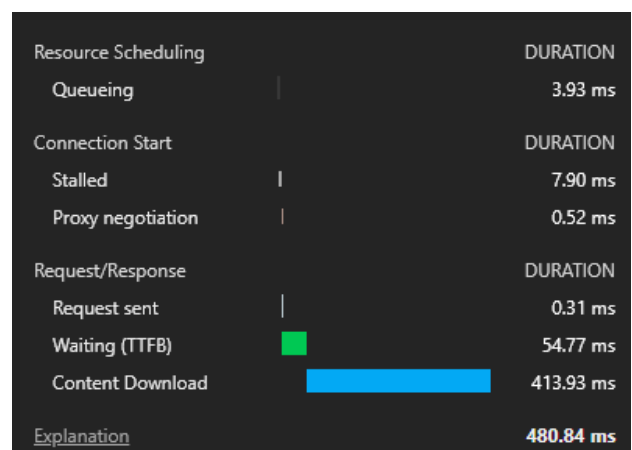


Fig. 8. Measurements of the response speed of the microservice request “Search for vacancies” in a web application on a microservice architecture

A comparative analysis of the results of testing the “Job Search” function for monolithic and microservice architectures is presented in Table 2.

Figure 8 shows measurements of the response speed of a service request for working with social networks in a web application on a monolithic architecture, and Fig. 9 shows measurements of the

response speed of a microservice request for working with social networks in a web application based on a microservice architecture.

Table 2
Results of testing the "Job Search" function using different server-side architectures

Indicator	Time, ms		Relative growth
	Monolithic architecture	Microservice architecture	
Search for vacancies without filters	2179	423	5.15
Search for vacancies with 1 filter	4902	458	10.7
Search for vacancies with 2 filters	6040	480	12.58
Search for vacancies with 4 filters	8298	503	16.09
Search for vacancies with 16 filters	17921	489	36.65

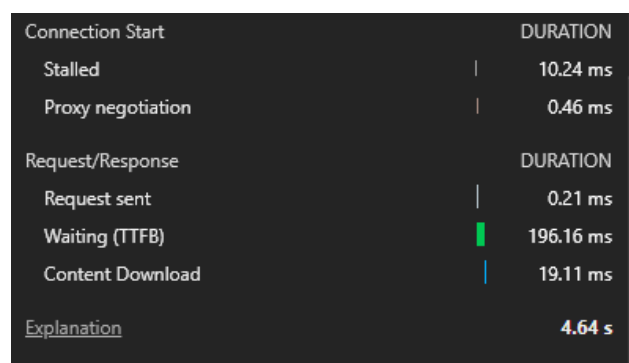


Fig. 9. Measurements of the response speed of a service request for working with social networks in a web application on a monolithic architecture

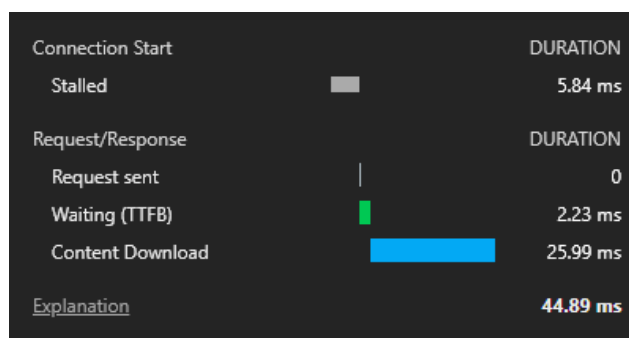


Fig. 10. Measurements of the response speed of a microservice request for working with social networks in a web application based on a microservice architecture

The results of testing the "Working with social networks" function for monolithic and microservice architecture are presented in Table 3. Analysis of the test results showed that the use of microservice architecture is more appropriate, because the service request response speed indicators are better than

monolithic architecture. The average indicators of the tests performed showed a speed increase in the range from 3 to 36 times. These indicators may indicate that the system has been successfully improved and the decision to switch to a new type of architecture was correct.

Table 3
Results of testing the "Working with social networks" function using different server-side architectures

Indicator	Time, ms		Relative growth
	Monolithic architecture	Microservice architecture	
Authorization when working with Facebook	3580	1203	2.98
Data request when working with Facebook	2790	980	2.85
Data processing when working with Facebook	1246	44	28.31
Authorization when working with Google	2987	891	3.35
Data query when working with Google	1790	594	3.01
Data processing when working with Google	988	89	11.1
Authorization when working with LinkedIn	2199	918	2.4
Data request when working with LinkedIn	2123	702	3.02
Data processing when working with LinkedIn	1008	76	13.26

It should also be noted that due to changes in the cluster architecture, both current and future financial costs have been significantly reduced. The main problem with the old approach was the problem of scaling, which took up too many hardware resources. Now the load can be easily distributed between different services, thereby making it a less expensive procedure.

Comparison of the analyzed methods. Development using microservice architecture is very promising for implementation. This architecture has been most actively developing recently in the field of software development. The main advantages of such a solution compared to monolithic architecture are the savings of resources, both human and material – during the development process, and computational – during the scaling process. However, it is also worth considering the difficulties that arise when using

microservices, mainly related to the complication of transaction mechanisms, as well as the need to implement systems for automated monitoring and logging of program operation. Therefore, the microservice architecture of software development itself was chosen for further development.

Conclusions. The article considers the task of improving the server part of a web application by moving from a monolithic architecture to a microservice architecture. The analysis showed that the monolithic approach, despite its simplicity of implementation, has significant limitations in terms of scalability, performance, and support, especially in the context of an increasing number of users and the complexity of the system's business logic.

As part of the study, a structural analysis of the existing software application was carried out and its decomposition into separate logical and functional modules was performed. Based on the results of this analysis, a microservice architecture was formed, which includes services for working with users, authorization and authentication, content management, processing text, graphic and video data, job search and resumes, integration with social networks, interaction with payment systems, as well as security and administration modules. This approach allowed us to clearly delineate the functional responsibilities of the system components, reduce their interdependence and create conditions for the independent development and updating of each service. An important stage of the work was the improvement of the architecture of the cluster infrastructure that ensures the functioning of the web application. The redistribution of computing resources between individual services allowed us to optimize the load on servers, reduce the number of physical or virtual machines and at the same time increase the speed of processing user requests. The integration of specialized tools for searching and processing large data sets helped reduce the load on the main database and ensured more efficient use of hardware resources. Separate services for processing e-mail and SMS messages were outsourced to external cloud services, which further increased the stability and scalability of the system.

A comparative analysis of the performance of monolithic and microservice architectures showed a

significant increase in performance for the main functions of the web application. In particular, a reduction in response time was recorded for the administration, job search, and integration with social networks modules. The average values of the productivity increase ranged from several times to dozens of times, which confirms the effectiveness of the chosen approach. In addition to improving technical characteristics, the proposed architectural solution has important organizational and economic advantages. The decomposition of the system into separate services creates conditions for the parallel work of several development teams, which reduces the time to implement new functions and simplifies the process of supporting the software product. The ability to use different programming languages and technology stacks to implement individual components increases the flexibility of the system and allows you to choose the most appropriate tools for solving specific tasks.

The results obtained indicate the feasibility of using microservice architecture to modernize outdated web applications. The proposed solution allows you to increase the flexibility of the system, simplify its maintenance, and create conditions for further scaling without significant changes in the structure of the program code. In addition, the microservice approach provides the ability to use different technologies to implement individual components, which increases the system's adaptability to changing requirements and the emergence of new development tools.

The scientific significance of the study lies in substantiating practical approaches to refactoring the architecture of the server part of web applications and in forming methodological recommendations for the transition from monolithic solutions to microservices. The practical significance of the work lies in the possibility of using the obtained results in the design and optimization of real information systems focused on working with a large number of users and high performance requirements.

Thus, the study confirmed that the implementation of microservice architecture is an effective means of improving the server part of web applications, allows achieving a significant increase in performance, optimizing resource use and providing conditions for further development of the software product.

Bibliography:

1. Aydemir F., Başçiftçi F. Building a performance efficient core banking system based on the microservices architecture. *Journal of Grid Computing*. 2022. Vol. 20. No. 4. 37 p. Doi: 10.1007/s10723-022-09624-z.
2. Järvinen K. Analysing the Effects of Scalability in Microservices to UserPerceived Performance and Cloud Costs. Master of Science (Tech) Thesis. 2025. 102 p.
3. Kalske M. Transforming monolithic architecture towards microservice architecture: Master of Science (Tech) Thesis. 2018. 76 p.

4. Mendonça N. C., Box C., Manolache C., Ryan L. Why Istio Migrated From Microservices to a Monolithic Architecture *IEEE Software Magazine*. 2021. №5. P. 1 – 22.
5. Reddy D. N., Suryodai R., Kumar S. V., Ambika M., Muniyandy E., Krishna R. V., Abdurasul B A Scalable Microservices Architecture for Real-Time. *International Journal of Advanced Computer Science and Applications (IJACSA)*. Vol. 16. No. 9. 2025. P 31 – 42. DOI: <https://dx.doi.org/10.14569/IJACSA.2025.0160905>.
6. Selivorstova T., Klishch S., Kyrychenko S., Guda A., Ostrovskaya K. Analysis of monolithic and microservice architectures: Features and metrics. *International scientific journal «Computer systems and information technologies»*. 2021. № 3. P. 59 – 60. Doi: 10.31891/CSIT-2021-5-8.
7. Shaik Sh. Technical review: Server-side Composition vs. Client-side Composition. *World Journal of Advanced Research and Reviews*. 2025. 26(02). P. 3037-3046.
8. Shethiya A. S. Scalability and Performance Optimization in Web Application Development. *Integrated Journal of Science and Technology Computer Science & Information Technology*. 2025. Vol. 2. Iss. 1. 7 p.
9. Sigala N. S. Microservices Architecture in Cloud Computing: A Software Engineering Perspective on Design, Deployment, and Management. *International journal of research and innovation in social science (IJRISS)*. 2025. Vol. IX. Iss. XV. P. 215 – 239. Doi: <https://dx.doi.org/10.47772/IJRISS.2025.915EC0013>.
10. Siriwardena P., Dias N. Microservices Security in Action: Design secure network and API endpoint security for Microservices. *Manning Publications Co*. 2020. 616 p.
11. Tapia F., Mora M. A., Fuertes W., Aules H., Aules H., Flores E., Toulkeridis Th. From Monolithic Systems to Microservices: A Comparative Study of Performance. *Applied Sciences*. 2020. №10 (17). 30 p. Doi:10.3390/app10175797.
12. Ushakova I., Plokha O., Skorin Yu. Approaches to Web Application Performance Testing And Real-Time Visualization of Results. *Bulletin of Kharkiv National Automobile and Highway University*. 2022. №96. P.71 – 80.
13. Ushakova, I., Skorin, Y., Shcherbakov, A. Methods of quality assurance of software development based on a systems approach. *CEUR Workshop Proceedings*. 2021. № 3200. P. 158 – 168.

Ушакова І.О. МОДЕРНІЗАЦІЯ СЕРВЕРНОЇ ЧАСТИНИ ВЕБЗАСТОСУНКУ НА ОСНОВІ МІКРОСЕРВІСНОЇ АРХІТЕКТУРИ

У статті розглянуто проблему неефективності функціонування вебзастосунків, побудованих на основі монолітної архітектури, зокрема в аспектах швидкодії, масштабованості та підтримки. Актуальність дослідження обумовлена зростанням навантаження на інформаційні системи та потребою забезпечення високої продуктивності при роботі з великим обсягом користувацьких запитів. Метою роботи є обґрунтування доцільності переходу від монолітної архітектури серверної частини вебзастосунку до мікросервісної архітектури та підвищення ефективності функціонування системи шляхом її структурної декомпозиції. В процесі дослідження проведено аналіз існуючої архітектури застосунку, визначено його логічні та функціональні компоненти, що дозволило здійснити декомпозицію модулів та перехід до окремих незалежних сервісів. Виділено основні функціональні мікросервіси, що відповідають за роботу з користувачами, авторизацію, обробку текстових, графічних та відеоданих, роботу з вакансіями та резюме, пошукові механізми, соціальні мережі, платіжні системи, поштові та SMS сервіси, а також модуль безпеки й адміністрування. Запропоноване архітектурне рішення дозволило забезпечити незалежне масштабування окремих сервісів, спростити процес їх модернізації та оптимізувати розподіл обчислювальних ресурсів. Дослідження містить порівняльний аналіз показників продуктивності вебзастосунку до модернізації та після впровадження мікросервісної архітектури. Наведено результати тестування функцій адміністрування, пошуку та інтеграції із соціальними мережами, які свідчать про суттєве покращення часу відповіді системи. Показано, що середній приріст швидкодії становить від кількох разів до кількох десятків разів залежно від типу запиту. Окрему увагу приділено оптимізації кластерної інфраструктури шляхом перерозподілу ресурсів та інтеграції спеціалізованих сервісів зберігання та обробки даних. Отримані результати підтверджують ефективність застосування мікросервісної архітектури для вдосконалення серверної частини вебзастосунків і можуть бути використані при розробленні та модернізації складних інформаційних систем, орієнтованих на високе навантаження, гнучкість та стабільність функціонування.

Ключові слова: вебзастосунок, монолітна архітектура, мікросервісна архітектура, структурна декомпозиція, кластерна інфраструктура, тестування продуктивності.

Дата першого надходження статті до видання: 17.02.2026

Дата прийняття статті до друку після рецензування: 13.03.2026

Дата публікації (оприлюднення) статті 11.05.2026

Янко А.С.<https://orcid.org/0000-0003-2876-9316>

Національний університет

«Полтавська політехніка імені Юрія Кондратюка»

Гончаренко С.О.<https://orcid.org/0009-0009-3417-6198>

Національний університет

«Полтавська політехніка імені Юрія Кондратюка»

ПОРІВНЯЛЬНИЙ АНАЛІЗ АРХІТЕКТУРНИХ ТА КРИПТОГРАФІЧНИХ ІННОВАЦІЙ У СПЕЦИФІКАЦІЯХ LORAWAN 1.0.4 ТА 1.1

У статті проведено порівняльний аналіз еволюції протоколу LoRaWAN, зосереджений на переході від специфікації версії 1.0.4 до архітектури версії 1.1. Актуальність дослідження зумовлена масштабуванням систем Інтернету речей (IoT) та посиленням вимог до кібербезпеки критичних інфраструктур, де рішення класу LPWAN стали галузевим стандартом. Розглянуто архітектурні особливості LoRaWAN 1.0.4 як завершального етапу гілки 1.0.x, що забезпечує базову функціональність, проте має обмеження у гнучкому управлінні ключами та підтримці глобального роумінгу. Висвітлено вразливості механізмів активації «по повітрю» (OTAA), притаманні раннім версіям, та обґрунтовано необхідність нових підходів до автентифікації.

Основну увагу приділено концептуальним трансформаціям у LoRaWAN 1.1. Досліджено роль виділеного Join Server (JS), який відокремлює процес автентифікації від функцій мережевого сервера. Таке розділення повноважень визначено як критичний елемент для побудови мультиоператорських середовищ, де довіра між власником пристрою та постачальником послуг має бути чітко розмежована. Проаналізовано модель розподілу серверних ролей на обслуговуючий (sNS) та домашній (hNS) сервери, що створює підґрунтя для безшовного роумінгу та підвищення живучості мереж.

Окремий розділ присвячено аналізу криптографічних інновацій. Висвітлено перехід від єдиного мережевого ключа сесії (NwkSKey) до розширеної ієрархії, що включає спеціалізовані ключі: FNwkSIntKey та SNwkSIntKey для перевірки цілісності висхідного й низхідного каналів, а також NwkSEncKey для шифрування MAC-команд. Доведено, що такий підхід підвищує стійкість протоколу до MITM-атак та захищає від компрометації сегментів мережі. Розібрано вдосконалені механізми управління лічильниками кадрів та введення нових нонсів (Nonces), що нівелює ризики replay-атак.

Результати підтверджують, що LoRaWAN 1.1 є стратегічним кроком до створення захищених промислових IoT-систем. Попри складність реалізації, переваги у безпеці та масштабованості роблять версію 1.1 оптимальною для сфер «розумного міста», моніторингу та енергетики. Практична цінність роботи полягає у формулюванні критеріїв доцільності модернізації мереж на базі LoRaWAN 1.0.4 та наданні рекомендацій щодо впровадження нових безпекових стандартів.

Ключові слова: LoRaWAN, Інтернет речей, кібербезпека, криптографічний захист, масштабованість мереж, механізм автентифікації, розподілена серверна архітектура, сесійні ключі.

Постановка проблеми. Стрімкий розвиток концепції Інтернету речей (IoT) зумовив потребу у масштабованих, енергоефективних та довготривалих мережевих рішеннях для передавання великих обсягів даних на великі відстані. Однією з ключових технологій класу LPWAN (Low Power Wide Area Network), що відповідає цим вимогам,

є LoRaWAN – відкритий протокол мережевого рівня, розроблений для організації енергоощадних бездротових мереж зіркоподібної топології. Завдяки підтримці великої кількості кінцевих пристроїв, низькому енергоспоживанню та широкому покриттю LoRaWAN активно застосовується у сферах розумного міста, промислового моніто-

рингу, сільського господарства та екологічного контролю [1, с. 1].

Версія LoRaWAN 1.0.4 стала завершальним етапом розвитку гілки 1.0.x, у якій було вдосконалено механізми сумісності, управління пристроями та окремі аспекти MAC-рівня [2, с. 13]. Водночас зростання вимог до безпеки, масштабованості та підтримки роумінгу між мережами зумовило появу специфікації LoRaWAN 1.1, що передбачає суттєві архітектурні та криптографічні зміни [3, с. 41].

На відміну від попередньої версії, LoRaWAN 1.1 запроваджує розширену модель розподілу ключів безпеки, оновлену процедуру активації пристроїв (Join-процес), а також чіткіше розмежування функцій між мережевим, прикладним та join-серверами. Такі зміни безпосередньо впливають на рівень захищеності мережі, керованість пристроями та можливість масштабування інфраструктури.

Отже, аналіз відмінностей між LoRaWAN 1.0.4 та LoRaWAN 1.1 є актуальним з точки зору оцінки еволюції протоколу, визначення переваг нової специфікації та розуміння її впливу на функціонування сучасних LoRaWAN-мереж [4, с. 8]. Подальший розвиток Інтернету речей супроводжується не лише збільшенням кількості підключених пристроїв, але й ускладненням вимог до інфраструктури зв'язку. Для ринку IoT сьогодні характерні три ключові тенденції: масштабування мереж до десятків і сотень тисяч вузлів, прагнення мінімізувати капітальні та операційні витрати, а також суттєве посилення вимог до кібербезпеки. У таких умовах технології класу LPWAN, зокрема LoRaWAN, залишаються одним із найбільш економічно доцільних рішень для побудови розподілених сенсорних мереж.

Разом із тим, широке впровадження LoRaWAN у критично важливих сферах – енергетиці, інфраструктурному моніторингу, промисловості та розумних містах – підвищує ризики, пов'язані з несанкціонованим доступом, компрометацією ключів шифрування та масштабними мережевими атаками. З огляду на тривалий життєвий цикл IoT-пристроїв і обмежені можливості їх оновлення, питання стійкої криптографічної архітектури та гнучкого управління ключами набувають стратегічного значення.

У цьому контексті поява специфікації LoRaWAN 1.1 є відповіддю на сучасні виклики ринку. Нова версія протоколу пропонує вдосконалену модель безпеки, розмежування ролей мережових компонентів, підтримку роумінгу та підвищену керованість пристроями. Водночас значна частина

існуючих розгортань продовжує працювати на базі LoRaWAN 1.0.4, що зумовлює необхідність системного порівняння обох версій [5, с. 1].

Отже, актуальність дослідження полягає у потребі комплексної оцінки архітектурних і протокольних змін у LoRaWAN 1.1 з позиції їх практичного впливу на безпеку, масштабованість і економічну ефективність мереж. Такий аналіз дозволяє визначити доцільність переходу на нову специфікацію та її відповідність сучасним вимогам розвитку IoT-інфраструктур.

Аналіз останніх досліджень і публікацій.

Питання еволюції протоколу LoRaWAN та порівняння його версій перебувають у центрі уваги багатьох дослідників. Зокрема, у дослідженнях [6, с. 12] та [7, с. 36] детально розглянуто обмеження перших версій LoRaWAN (1.0.x) щодо масштабованості та ємності мережі в умовах щільного розгортання пристроїв. Автори зазначають, що хоча топологія «зірка зірок» є ефективною для енергоощадності, вона створює певні виклики для безпеки під час передачі ключів через мережеві сервери.

Криптографічні аспекти специфікації 1.1 стали об'єктом системного аналізу у статті [8, с. 348], які проаналізували вразливості механізму активації пристроїв «по повітрю» (з англ. Over-the-Air Activation – OTAA) у версії 1.0.4, зокрема ризики атак повторного відтворення (replay attacks). Дослідники підкреслюють, що впровадження окремого Join Server у версії 1.1 є критичним кроком для забезпечення довіри у мульти-операторських мережах [3, с. 45].

У науковому дослідженні [9, с. 8332] увага приділяється практичному впровадженню роумінгу та архітектурним змінам, що стали можливими завдяки чіткому розділенню функцій між серверами домашньої (hNS) та обслуговуючої (sNS) мереж. Проте, попри очевидні переваги версії 1.1, у публікації [10, с. 5] відзначається, що складність реалізації оновленого стека протоколів та підвищені вимоги до обчислювальних ресурсів кінцевих вузлів сповільнюють масовий перехід від старої специфікації 1.0.4 до 1.1.

Незважаючи на значну кількість публікацій, порівняльний аналіз архітектурних інновацій саме в контексті інтеграції безпеки та операційної ефективності між версіями 1.0.4 та 1.1 потребує більш глибокого систематизованого підходу, що і зумовило вибір теми даної статті.

Постановка завдання. Сучасний стан розвитку технологій LPWAN потребує переходу від базової функціональності до побудови архі-

тектурно стійких та криптографічно захищених систем. Незважаючи на стабільність специфікації LoRaWAN 1.0.4, виникає об'єктивна потреба у дослідженні переваг версії 1.1 як стратегічного інструменту для масштабування IoT-інфраструктур.

Метою статті є проведення системного порівняльного аналізу архітектурних рішень та криптографічних інновацій у специфікаціях LoRaWAN версій 1.0.4 та 1.1 для визначення їхнього впливу на рівень захищеності, масштабованість та операційну ефективність сучасних мереж Інтернету речей.

Для досягнення поставленої мети необхідно вирішити такі завдання:

1. Проаналізувати архітектурні особливості специфікації LoRaWAN 1.0.4, виявивши ключові обмеження щодо розмежування довіри та управління сесійними ключами.

2. Дослідити еволюцію компонентів мережі у версії 1.1, зокрема роль окремого Join Server у забезпеченні безпечного роумінгу та активації пристроїв.

3. Здійснити порівняльну оцінку криптографічних моделей, фокусуючись на переході від єдиного мережевого ключа до розширеного набору (FNwkSIntKey, SNwkSIntKey, NwkSEncKey) для посилення цілісності та конфіденційності даних.

4. Вивчити механізми захисту від атак повторного відтворення (replay-attacks) та принципи управління лічильниками кадрів у новій специфікації.

5. Оцінити вплив впроваджених змін на економічну ефективність та можливості інтеграції пристроїв у мультиоператорські середовища.

6. Сформулювати рекомендації щодо доцільності модернізації існуючих розгортань та переходу на стек протоколів LoRaWAN 1.1 у нових проєктах.

Виклад основного матеріалу. *Архітектурні особливості LoRaWAN 1.0.4.* LoRaWAN 1.0.4 є завершальною версією гілки 1.0.x та зберігає класичну зіркоподібну топологію (star-of-stars), у якій кінцеві пристрої (End Devices) взаємодіють із мережею через шлюзи (Gateways), що передають трафік до мережевого сервера (Network Server, NS). Специфікація 1.0.4 уточнює поведінку MAC-рівня, механізми ADR (Adaptive Data Rate), процедури підтвердження/непідтвердження повідомлень та обробку лічильників кадрів (Frame Counters).

Модель безпеки в LoRaWAN 1.0.4. У версії 1.0.x використовується дворівнева схема ключів:

– NwkSKey – для забезпечення цілісності повідомлень MAC-рівня;

– AppSKey – для шифрування корисного навантаження (FRMPayload).

При активації пристрою через OTAA (Over-The-Air Activation) формується сесійний набір ключів на основі AppKey та параметрів процедури Join. Однак у 1.0.x мережевий сервер фактично має доступ до AppSKey (залежно від реалізації), що обмежує розділення довіри між мережевим та прикладним рівнями. Також у версії 1.0.x відсутнє повноцінне розмежування функцій Join Server, а процедура приєднання має спрощену криптографічну модель, що ускладнює реалізацію масштабованого та безпечного роумінгу між мережами різних операторів [11, с. 1].

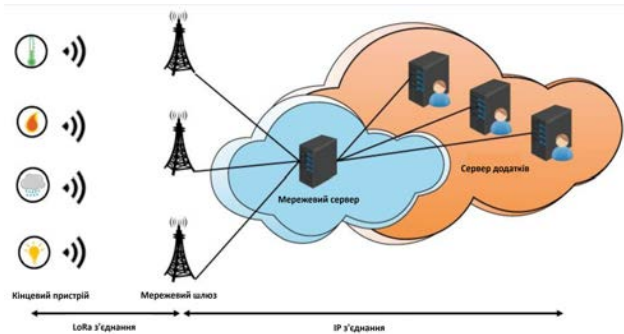


Рис. 1. Архітектура мережі LoRaWAN 1.0.4

Архітектурна еволюція у LoRaWAN 1.1. LoRaWAN 1.1 запроваджує суттєві зміни у структурі мережі та криптографічній архітектурі, що безпосередньо впливають на масштабованість і безпеку.

Розмежування мережевих ролей. У версії 1.1 чітко визначено логічне розділення між Network Server (NS), Application Server (AS) та Join Server (JS).

Join Server стає окремою довіреною сутністю, відповідальною за управління кореневими ключами та генерацію сесійних ключів. Це дозволяє:

- мінімізувати довіру до мережевого оператора;
- реалізувати міжмережевий роумінг;
- централізовано керувати криптографічною інфраструктурою [12, с. 1].

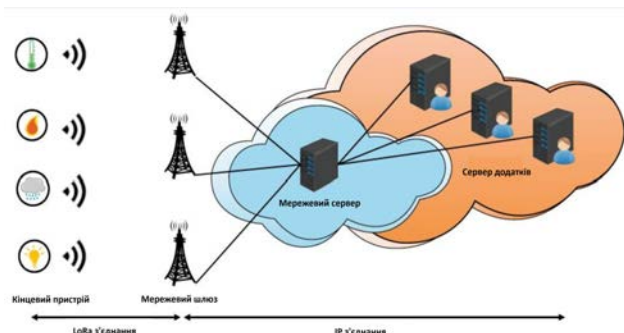


Рис. 2. Архітектура мережі LoRaWAN 1.1

Розширена криптографічна модель LoRaWAN 1.1. Однією з ключових відмінностей є введення нової ієрархії ключів. Кореневі ключі. Замість одного AppKey, у 1.1 використовується:

- NwkKey – для мережевої аутентифікації;
- AppKey – для прикладного рівня.

це забезпечує криптографічне розділення відповідальності між мережею та застосунком.

Сесійні ключі. Суттєвою архітектурною інновацією специфікації LoRaWAN 1.1 є декомпозиція єдиного мережевого ключа сесії (NwkSKey), характерного для версії 1.0.4, на три спеціалізовані вектори безпеки. Такий підхід базується на принципі мінімізації привілеїв та функціонального розділення криптографічних задач:

- FNwkSIntKey – використовується для обчислення та перевірки коду цілісності повідомлення (з англ. Message Integrity Code – MIC) у висхідних лініях зв'язку (uplink), це забезпечує автентифікацію кінцевого пристрою мережевим сервером;

- SNwkSIntKey – призначений для верифікації цілісності даних у низхідних лініях зв'язку (downlink), що гарантує кінцевому пристрою отримання команд саме від легітимного сервера обслуговування;

- NwkSEncKey – застосовується виключно для шифрування та дешифрування корисного навантаження MAC-команд на мережевому рівні.

Завдяки такій диференціації досягається вищий рівень ізоляції функцій безпеки. Зокрема, це унеможливує компрометацію всієї сесії при витоку одного з ключів та значно підвищує стійкість системи до атак повторного відтворення (replay attacks) за рахунок незалежної індексації лічильників повідомлень [13, с. 2]. Окрім того, розділення ключів на «висхідні» та «низхідні» є необхідною умовою для реалізації безпечного роумінгу, оскільки дозволяє мережі обслуговування (Serving Network) виконувати базові операції з трафіком без надання їй повного доступу до конфіденційних даних домашньої мережі (Home Network).

Процедура Join та механізми автентифікації. У LoRaWAN 1.1 процедура OTAA суттєво вдосконалена:

- введено новий тип ідентифікатора JoinEUI (замість AppEUI);

- додано RJoinCount для захисту від повторного використання join-запитів;

- розширено перевірку MIC із використанням окремих ключів.

Ці зміни підвищують захист від replay-атак та атак повторного приєднання, що є критично важливим у масштабних IoT-мережах з тривалим життєвим циклом пристроїв [14, с. 1].

Підтримка роумінгу. Одним із стратегічних нововведень 1.1 є нативна підтримка роумінгу між мережами операторів. Роумінг реалізується через:

- розмежування Serving Network Server (SNS) та Home Network Server (HNS);

- збереження контролю над ключами на стороні домашньої мережі;

- централізований Join Server.

Все це відкриває можливість побудови глобальних LoRaWAN-інфраструктур із збереженням безпеки та контролю над пристроями [15, с. 6].

Управління лічильниками кадрів та захист від replay-атак. У LoRaWAN 1.0.x використовуються 16-бітні або 32-бітні лічильники кадрів, однак їх обробка була менш формалізованою. Удосконалення механізмів контролю послідовності пакетів у LoRaWAN 1.1 базується на таких ключових інноваціях:

- формалізація обробки невідповідних значень (Strict Replay Protection);

- посилена верифікація цілісності (MIC) з урахуванням вектора напрямку;

- захист від скидання лічильників при переключенні;

- посилено контроль повторних пакетів.

Дані нововведення є критично важливими для масштабованих мереж з високою щільністю вузлів. Саме така архітектура лічильників забезпечує довготривалу стабільність мережі, запобігаючи деградації каналів зв'язку через масовані replay-атаки, які в іншому випадку могли б паралізувати роботу шлюзів та серверів обробки.

MAC-рівень та керування мережею. LoRaWAN 1.1 зберігає механізм класів пристроїв (Class A, B, C), проте уточнює поведінку MAC-команд і їх шифрування через NwkSEncKey. Важливими аспектами є:

- вдосконалений ADR;

- точніше управління параметрами каналу;

- краща сумісність із регіональними специфікаціями.

MAC-команди тепер мають чітке криптографічне розмежування, що унеможливує їх перехоплення або модифікацію без відповідного ключа. У таблиці 1 наведено список нових MAC-команд у LoRaWAN 1.1 [16, 1. с]:

Вплив змін на масштабованість та економічну ефективність. З технічної точки зору, розділення серверних ролей і ключів:

- зменшує ризик компрометації всієї мережі;

- дозволяє операторам гнучко масштабувати інфраструктуру;

Таблиця 1

Перелік та функціональне призначення MAC-команд у специфікації LoRaWAN 1.1

CID	Назва MAC-команди	Напрямок передачі	Функціональне призначення
0x01	ResetInd	End-device → Network	Для ABP пристрою: індикація reset (повернення MAC/радіо параметрів до стандартних) та узгодження minor-версії протоколу; має повторюватися в uplink, доки не буде отримано ResetConf.
0x01	ResetConf	Network → End-device	Підтвердження ResetInd.
0x0B	RekeyInd	End-device → Network	Для OTAA: сигналізація оновлення security context (rekeying) після Join-акцепту; ED додає RekeyInd у наступні uplink, доки не прийде RekeyConf.
0x0B	RekeyConf	Network → End-device	Підтвердження RekeyInd.
0x0C	ADRParamSetupReq	Network → End-device	Налаштування параметрів ADR_ACK_LIMIT та ADR_ACK_DELAY на боці ED (точніше керування ADR-реакцією пристрою).
0x0C	ADRParamSetupAns	End-device → Network	Підтвердження ADRParamSetupReq.
0x0E	ForceRejoinReq	Network → End-device	Команда мережі “Rejoin негайно” з параметрами періодичності/DR/типу Rejoin та лімітом повторів (для RekeyInd або роумінгу).
0x0F	RejoinParamSetupReq	Network → End-device	Встановлення параметрів періодичного надсилання RejoinReq Type 0 (за часом або за лічильником uplink).
0x0F	RejoinParamSetupAns	End-device → Network	Підтвердження RejoinParamSetupReq.
0x20	DeviceModelInd	End-device → Network	Class C доповнення в 1.1: ED інформує мережу про поточний режим роботи (Class A або Class C) і має повторювати в uplink, доки не отримає DeviceModeConf.
0x20	DeviceModeConf	Network → End-device	Підтвердження DeviceModelInd.

- спрощує інтеграцію з хмарними сервісами.
- З економічної точки зору:
 - централізований Join Server зменшує витрати на адміністрування ключів;
 - підтримка роумінгу дозволяє будувати мультиоператорські мережі;
 - покращена безпека знижує потенційні втрати від атак.

Таким чином, LoRaWAN 1.1 є не просто інкрементальним оновленням, а структурною еволюцією протоколу, спрямованою на відповідність сучасним вимогам безпеки [17, с. 329], масштабованості та ринкової інтеграції IoT-рішень [18, с. 2118].

Висновки. У межах проведеного аналізу було розглянуто еволюцію протоколу LoRaWAN від версії 1.0.4 до 1.1 з позиції архітектурних, криптографічних та функціональних змін. Встановлено, що LoRaWAN 1.0.4 є логічним завершенням гілки 1.0.x та забезпечує стабільну роботу мережі, сумісність пристроїв і базовий рівень безпеки, достатній для більшості класичних IoT-застосувань. Водночас ця версія має обмеження щодо розмежування довіри між компонентами мережі, управління ключами та реалізації масштабного роумінгу.

На відміну від попередньої специфікації, LoRaWAN 1.1 пропонує концептуально нову модель безпеки із розподілом кореневих та сесій-

них ключів, чітким розмежуванням функцій між Network Server, Application Server та Join Server, а також удосконаленими механізмами автентифікації та захисту від replay-атак. Запровадження окремих мережових ключів (FNwkSIntKey, SNwkSIntKey, NwkSEncKey) та розширеної процедури Join суттєво підвищує криптографічну стійкість системи та зменшує ризики компрометації мережової інфраструктури.

Крім підвищення рівня безпеки, специфікація 1.1 створює передумови для масштабування та інтеграції LoRaWAN-мереж у мультиоператорські середовища завдяки підтримці роумінгу та розподіленої серверної архітектури. Це відповідає сучасним тенденціям розвитку IoT, які характеризуються зростанням кількості пристроїв, підвищенням вимог до кіберзахисту та необхідністю оптимізації операційних витрат.

Таким чином, LoRaWAN 1.1 слід розглядати не як інкрементальне оновлення, а як стратегічний етап розвитку протоколу, спрямований на забезпечення довгострокової стійкості, безпеки та масштабованості мереж LPWAN. Отримані результати можуть бути використані для обґрунтування доцільності переходу з LoRaWAN 1.0.4 на 1.1 у нових і модернізованих IoT-розгортаннях, з урахуванням вимог конкретного сценарію застосування.

Список літератури:

1. What Is LoRaWAN® Specification / LoRa Alliance. 2023. URL: <https://lora-alliance.org/about-lorawan/> (дата звернення: 19.03.2026).
2. LoRaWAN 1.0.4 Specification / LoRa Alliance. 2020. 102 p. URL: <https://resource.lora-alliance.org/technical-specifications/lorawan-1-0-4-specification> (дата звернення: 19.03.2026).
3. LoRaWAN 1.1 Specification / LoRa Alliance. 2017. 98 p. URL: <https://resource.lora-alliance.org/technical-specifications/lorawan-1-1-specification> (дата звернення: 19.03.2026).
4. Analysis of LoRaWAN 1.0 and 1.1 Protocols Security Mechanisms / S. Loukil, L. C. Fourati, A. Nanyar, K.-W.-A. Chee. *Sensors*. 2022. Vol. 22, No. 10. Art. 3717. DOI: <https://doi.org/10.3390/s22103717>.
5. End-to-End IoT Solutions / TEKTELIC. 2026. URL: <https://tektelic.com/products/solutions/> (дата звернення: 01.03.2026).
6. Farrell S. LPWAN Overview. RFC 8376. 2018. 32 p. DOI: <https://doi.org/10.17487/RFC8376>.
7. Understanding the Limits of LoRaWAN / F. Adelantado et al. *IEEE Communications Magazine*. 2017. Vol. 55, No. 9. P. 34–40. DOI: <https://doi.org/10.1109/MCOM.2017.1600613>.
8. Butun I., Pereira N., Gidlund M. Security Analysis of LoRaWAN v1.1. *2019 IEEE 44th Conference on Local Computer Networks (LCN)*. Osnabrueck, Germany, 2019. P. 348–351. DOI: <https://doi.org/10.1109/LCN44234.2019.8990744>.
9. Han J., Wang J. The Architecture and Security Analysis of LoRaWAN v1.1. *IEEE Access*. 2018. Vol. 6. P. 8330–8341. DOI: <https://doi.org/10.1109/ACCESS.2018.2796191>.
10. Lavric A., Popa V. LoRaWAN Gateway: A Comparative Analysis of v1.0.4 and v1.1 Implementations. *Sensors*. 2021. Vol. 21, No. 14. Art. 4872. DOI: <https://doi.org/10.3390/s21144872>.
11. Ivezic M. LoRaWAN Security 101 (Non-5G IoT Connectivity Options). *PostQuantum*. 2019. URL: <https://postquantum.com/5g-security/iot-lorawan-security/> (дата звернення: 07.08.2019).
12. Introducing LoRaWAN 1.1 Support / Mbed. 2026. URL: <https://os.mbed.com/blog/entry/Introducing-LoRaWAN-1-1-support/> (дата звернення: 01.03.2026).
13. Chen X., Lech M., Wang L. A Complete Key Management Scheme for LoRaWAN v1.1. *Sensors*. 2021. Vol. 21, No. 9. Art. 2962. DOI: <https://doi.org/10.3390/s21092962>.
14. LoRaWAN / Smart Parks Wiki. 2026. URL: <https://wiki.smartparks.com/en/concepts/lorawan> (дата звернення: 01.03.2026).
15. Янко А. С., Горчаренко С. О. Практичні рекомендації з налаштування LORAWAN-мережі для стабільної роботи та захищеності даних. *Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління* : тези доповідей п'ятнадцятої міжнар. наук.-техн. конф., 24–25 квітня 2025 р. Баку–Харків–Жиліна, 2025. Т. 3. С. 6–7. DOI: <https://doi.org/10.32620/ICT.25.t3>.
16. LoRaWAN® Specification v1.1 / LoRa Alliance. 2023. URL: <https://resources.lora-alliance.org/technical-specifications/lorawan-specification-v1-1> (дата звернення: 15.09.2023).
17. Formal security analysis of LoRaWAN / M. Eldefrawy, I. Butun, N. Pereira, M. Gidlund. *Computer Networks*. 2019. Vol. 148. P. 328–339. DOI: <https://doi.org/10.1016/j.comnet.2018.11.017>.
18. Naoui S., Elhdhili M. E., Saidane L. A. Novel Enhanced LoRaWAN Framework for Smart Home Remote Control Security. *Wireless Personal Communications*. 2020. Vol. 110. P. 2109–2130. DOI: <https://doi.org/10.1007/s11277-019-06832-x>.

Yanko A.S., Goncharenko S.O. COMPARATIVE ANALYSIS OF ARCHITECTURAL AND CRYPTOGRAPHIC INNOVATIONS IN LORAWAN 1.0.4 AND 1.1 SPECIFICATIONS

This paper provides a comparative analysis of the LoRaWAN protocol evolution, focusing on the transition from the version 1.0.4 specification to the version 1.1 architecture. The relevance of the study is driven by the scaling of Internet of Things (IoT) systems and the escalating cybersecurity requirements for critical infrastructures, where LPWAN solutions have become an industry standard. The architectural features of LoRaWAN 1.0.4 are examined as the final stage of the 1.0.x branch, which provides basic functionality but possesses limitations in flexible key management and global roaming support. The study highlights vulnerabilities in Over-the-Air Activation (OTAA) mechanisms inherent in earlier versions and justifies the necessity for new approaches to authentication.

The primary focus is placed on the conceptual transformations in LoRaWAN 1.1. The role of the dedicated Join Server (JS) is investigated, which decouples the authentication process from network server functions. This separation of concerns is identified as a critical element for building multi-operator environments, where trust between the device owner and the connectivity provider must be clearly demarcated. The paper analyzes the model of server role distribution into Serving (sNS) and Home (hNS) servers, which establishes the framework for seamless roaming and enhancing network survivability.

A dedicated section is devoted to the analysis of cryptographic innovations. It elucidates the transition from a single network session key (NwkSKey) to an extended hierarchy, including specialized keys: FNwkSIntKey and SNwkSIntKey for uplink and downlink integrity verification, and NwkSEncKey for MAC command encryption. It is demonstrated that this approach enhances the protocol's resilience against MITM attacks and protects against the compromise of network segments. The study details improved frame counter management and the introduction of new Nonces, which eliminate the risks of replay attacks.

The results confirm that LoRaWAN 1.1 is a strategic step toward creating secure industrial IoT systems. Despite implementation complexity, the security and scalability benefits make version 1.1 optimal for smart cities, monitoring, and energy sectors. The practical value of the work lies in formulating criteria for assessing the feasibility of modernizing LoRaWAN 1.0.4-based networks and providing recommendations for implementing new security standards.

Keywords: *LoRaWAN, Internet of Things (IoT), cybersecurity, cryptographic, network scalability, authentication mechanisms, distributed server architecture, session keys.*

Дата першого надходження статті до видання: 14.02.2026

Дата прийняття статті до друку після рецензування: 10.03.2026

Дата публікації (оприлюднення) статті 11.05.2026

Відомості про авторів

Агаджанова С.В. – кандидат технічних наук, доцент, доцент кафедри кібернетики та інформатики Сумського національного аграрного університету

Агаджанов-Гонсалес К.Х. – старший викладач кафедри кібернетики та інформатики Сумського національного аграрного університету

Антоняк В.С. – студентбакалавр кафедри комп'ютерних наук та інформаційних систем Карпатського національного університету імені Василя Стефаника

Ахмедов М.Ш. – аспірант кафедри інженерії програмного забезпечення в енергетиці Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Баглай О.П. – аспірант кафедри Інженерії вагонів та якості продукції Українського державного університету залізничного транспорту

Басюк Т.М. – кандидат технічних наук, доцент, доцент кафедри інформаційних систем та мереж Національного університету “Львівська політехніка”

Березюк І.А. – кандидат технічних наук, доцент, доцент кафедри автоматизації виробничих процесів Центральноукраїнського національного технічного університету

Берт С.Д. – незалежний дослідник кафедри кібербезпеки та інформаційних технологій, Університету митної справи та фінансів

Білий М.О. – аспірант кафедри інформаційних систем та технологій Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Бокій А.Р. – аспірант кафедри «Автоматизації виробничих процесів» Центральноукраїнського національного технічного університету

Бондарчук О.В. – кандидат технічних наук, доцент кафедри автоматизації технологічних процесів Київського національного університету будівництва і архітектури

Босько В.В. – кандидат технічних наук, доцент, доцент кафедри автоматизації виробничих процесів Центральноукраїнського національного технічного університету

Вакалюк Т.А. – доктор педагогічних наук, професор, завідувач кафедри інженерії програмного забезпечення Державного університету «Житомирська політехніка»

Варава І.А. – доцент кафедри інженерії програмного забезпечення в енергетиці Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Вдовина А.В. – студентка кафедри інженерії програмного забезпечення в енергетиці Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Власенко О.В. – старший викладач кафедри інженерії програмного забезпечення Державного університету «Житомирська політехніка»

Вовченко Д.С. – аспірант кафедри програмного забезпечення комп'ютерних систем Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Волк М.О. – доктор технічних наук, професор, професор кафедри електронних обчислювальних машин Харківського національного університету радіоелектроніки

В'юненко О.Б. – кандидат економічних наук, доцент, доцент кафедри кібернетики та інформатики Сумського національного аграрного університету

Гаврюков О.В. – доктор технічних наук, професор кафедри автоматизації технологічних процесів Київського національного університету будівництва і архітектури

Гейко О.О. – аспірант кафедри інженерії програмного забезпечення в енергетиці Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Глухова Н.В. – кандидат технічних наук, доцент, доцент кафедри кіберфізичних та інформаційновимірювальних систем Національного технічного університету «Дніпровська політехніка»

Гончаренко С.О. – аспірант кафедри комп'ютерних та інформаційних технологій і систем спеціальності «Комп'ютерні науки» Національного університету «Полтавська політехніка імені Юрія Кондратюка»

Гончаров Д.С. – аспірант кафедри комп'ютерної інженерії Чорноморського національного університету імені Петра Могили

Горєлов В.О. – кандидат технічних наук, доцент кафедри комп'ютерних наук та інформаційних систем Карпатського національного університету імені Василя Стефаника

Грицюк Ю.І. – доктор технічних наук, професор, професор кафедри інженерії програмного забезпечення Інституту комп'ютерних наук та інформаційних технологій Національного університету "Львівська політехніка"

Громський О.О. – асистент кафедри інженерії програмного забезпечення Державного університету «Житомирська політехніка»

Дембіцький В.В. – аспірант кафедри інженерії програмного забезпечення в енергетиці Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Денисенко І.В. – аспірант, факультет програмних систем та прикладної математики, кафедра системного програмування і спеціалізованих комп'ютерних систем Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Дикий С.А. – аспірант кафедри інформаційних систем і технологій Харківського національного університету радіоелектроніки

Діденко Т.В. – магістр, асистент кафедри програмної інженерії Національного університету «Одеська політехніка»

Дощенко Г.Г. – кандидат технічних наук, доцент кафедри експлуатації суднового електрообладнання та засобів автоматики Херсонської державної морської академії

Дудник О.О. – кандидат технічних наук, доцент кафедри програмного забезпечення Вінницького національного технічного університету

Дунєв С.С. – асистент кафедри інженерії програмного забезпечення Державного університету «Житомирська політехніка»

Духов Д.Ю. – магістр кафедри комп'ютерних систем та мереж Криворізького національного університету

Єременко В. С. – доктор технічних наук, завідувач кафедри інформаційновимірjuвальних технологій Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Засадна Х.О. – кандидат фізикоматематичних наук, доцент кафедри математики та комп'ютерних дисциплін Приватного вищого навчального закладу «Європейський університет»

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, старший викладач кафедри конструювання електронно-обчислювальної апаратури Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Зієров О.Р. – аспірант кафедри комп'ютеризованих систем та програмних технологій Національного університету «Одеська політехніка»

Іващенко О.Р. – аспірант кафедри автоматизації, комп'ютерних наук і технологій Криворізького національного університету

Ілуща А.С. – аспірант кафедри комп'ютеризованих систем та програмних технологій Національного університету «Одеська політехніка»

Іляш Ю.Ю. – кандидат технічних наук, викладач кафедри комп'ютерних наук та інформаційних систем Карпатського національного університету імені Василя Стефаника

Іносов С.В. – кандидат технічних наук, доцент кафедри автоматизації технологічних процесів Київського національного університету будівництва і архітектури

Кандиба І.О. – доктор філософії, старший викладач кафедри інженерії програмного забезпечення Чорноморського національного університету імені Петра Могили

Кармазін К.В. – аспірант Інституту кібернетики імені В. М. Глушкова Національної академії наук України

Карплевський Б.В. – спеціаліст, дизайнер світла Glowcraft Installations

Коваленко В.В. – студент 6 курсу, групи 6151м кафедри програмного забезпечення автоматизованих систем Національного університету кораблебудування імені адмірала Макарова

Ковтун Є.І. – аспірант кафедри електронних обчислювальних машин Харківського національного університету радіоелектроніки

Козак Р.Р. – магістр, технічне обслуговування «Lifeway Food»

Коломієць В.В. – здобувач Державного університету «Житомирська політехніка»

Комаров Д.Ю. – аспірант кафедри «Автоматизації виробничих процесів» Центральноукраїнського національного технічного університету

Корабльов М.М. – доктор технічних наук, професор кафедри інформаційних систем і технологій Харківського національного університету радіоелектроніки

Коцун В.І. – кандидат технічних наук, завідувач кафедри математики та комп'ютерних дисциплін Приватного вищого навчального закладу «Європейський університет»

Крилов Є.В. – кандидат технічних наук, доцент, доцент кафедри інформаційних систем та технологій Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Кузнєцов Д.І. – кандидат технічних наук, доцент, доцент кафедри комп'ютерних систем та мереж Криворізького національного університету

Кунгурцев О.Б. – кандидат технічних наук, професор, професор кафедри програмної інженерії Національного університету «Одеська політехніка»

Купін А.І. – доктор технічних наук, професор, завідувач кафедри комп'ютерних систем та мереж Криворізького національного університету

Лакоса С.Л. – кандидат технічних наук, доцент, доцент кафедри комп'ютерно-інтегрованих оптичних та навігаційних систем Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Ложечников В.Ф. – кандидат технічних наук, доцент, доцент кафедри комп'ютеризованих систем та програмних технологій Національного університету «Одеська політехніка»

Ломовацький А.А. – аспірант кафедри інформаційних систем та мереж Національного університету «Львівська політехніка»

Луценко В.Ю. – кандидат технічних наук, доцент кафедри автоматизації технологічних процесів Київського національного університету будівництва і архітектури

Ляшенко О.С. – кандидат технічних наук, доцент, декан факультету комп'ютерної інженерії та інформаційних технологій Харківського національного університету радіоелектроніки

Макарова Л.М. – кандидат технічних наук, доцент, доцент кафедри програмного забезпечення автоматизованих систем Національного університету кораблебудування імені адмірала Макарова

Мацуй А.М. – доктор технічних наук, професор кафедри «Автоматизації виробничих процесів» Центральноукраїнського національного технічного університету

Мацюк Д.М. – аспірант кафедри комп'ютерно-інтегрованих оптичних та навігаційних систем Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», головний конструктор Акціонерного товариства «Науково-технічний комплекс «Електронприлад»

Мелешко Є.В. – доктор технічних наук, професор, доцент кафедри кібербезпеки та програмного забезпечення Центральноукраїнського національного технічного університету

Михалюк А.П. – аспірант кафедри автоматизації та комп'ютерних технологій систем керування імені професора А. П. Ладанюка Національного університету харчових технологій

Міркевич Р.М. – Доцент, кандидат технічних наук, кафедра автоматизації та комп'ютерних технологій систем керування імені професора А. П. Ладанюка Національного університету харчових технологій

Міхав В.В. – доктор філософії з комп'ютерної інженерії, викладач кафедри інформаційних технологій Приватної установи «Університет науки, підприємництва та технологій»

Могильний Г.А. – кандидат технічних наук, доцент, директор Навчальнонаукового інституту математики та інформаційних технологій ДЗ «Луганський національний університет імені Тараса Шевченка»

Монченко О.В. – кандидат технічних наук, доцент, професор кафедри біомедичної інженерії та біомехатроніки Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Монченко Т.М. – здобувач вищої освіти кафедри акустичних та мультимедійних електронних систем Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Музика І.О. – кандидат технічних наук, доцент, декан факультету Інформаційних технологій Криворізького національного університету

Нальотов Д.Р. – магістрант комп'ютерної інженерії Харківського національного університету радіоелектроніки

Невлюдов І.Ш. – доктор технічних наук, професор, завідувач кафедри комп'ютерноінтегрованих технологій, автоматизації, робототехніки та безпекової інженерії Харківського національного університету радіоелектроніки

Нерода С.І. – асистент кафедри інженерії програмного забезпечення Державного університету «Житомирська політехніка»

Новоселов С.П. – доктор технічних наук, доцент, професор кафедри комп'ютерноінтегрованих технологій, автоматизації, робототехніки та безпекової інженерії Харківського національного університету радіоелектроніки

Олещенко Л.М. – кандидат технічних наук, доцент кафедри програмного забезпечення комп'ютерних систем Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Пархоменко М.Д. – доцент, директор інформаційнообчислювального центру Центральноукраїнського національного технічного університету

Пархоменко Ю.М. – кандидат технічних наук, доцент кафедри «Автоматизації виробничих процесів» Центральноукраїнського національного технічного університету

Пелих С.М. – доктор технічних наук, професор, професор кафедри комп'ютеризованих систем та програмних технологій Національного університету «Одеська політехніка»

Переметчик Д.О. – незалежний дослідник кафедри кібербезпеки та інформаційних технологій Університету митної справи та фінансів

Переяславська С.О. – кандидат педагогічних наук, доцент, доцент кафедри інформаційних технологій та систем ДЗ «Луганський національний університет імені Тараса Шевченка»

Писаренко Ю.В. – кандидат технічних наук, Інститут кібернетики імені В. М. Глушкова Національної академії наук України

Піотровський Є.Є. – спеціаліст за спеціальністю «Літаки і вертольоти», старший інженер програміст Throne Labs Inc., United States

Прокопович Ткаченко Д.І. – кандидат технічних наук, доцент, завідувач кафедри кібербезпеки та інформаційних технологій, Університет митної справи та фінансів, старший науковий співробітник Державної наукової установи «Інститут інформації, безпеки і права Національної академії правових наук України», докторант кафедри систем та технологій кібербезпеки Державного університету інформаційнокомунікаційних технологій

Пупена О.М. – кандидат технічних наук, доцент, доцент кафедри автоматизації та комп'ютерних технологій систем управління імені проф. А.П. Ладанюка Національного університету харчових технологій

Пуха Г.С. – молодший науковий співробітник Особливого конструкторського бюро «ШТОРМ» Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Пухалевич А.В. – кандидат технічних наук, доцент, доцент кафедри програмного забезпечення автоматизованих систем Національного університету кораблебудування імені адмірала Макарова

Рубель Ю.Б. – аспірант, асистент кафедри інженерії програмного забезпечення Інституту комп'ютерних наук та інформаційних технологій Національного університету «Львівська політехніка»

Руденко Ю.О. – доктор педагогічних наук, доцент, доцент кафедри кібернетики та інформатики Сумського національного аграрного університету

Самойленко М.І. – асистент кафедри автоматизації технологічних процесів Київського національного університету будівництва і архітектури

Самойлов О.О. – аспірант кафедри експлуатації суднового електрообладнання та засобів автоматики Херсонської державної морської академії

Ситник Л.Г. – кандидат технічних наук, доцент, доцент кафедри кібернетики та інформатики Сумського національного аграрного університету

Сичова О.В. – кандидат технічних наук, доцент, доцент кафедри комп'ютерноінтегрованих технологій, автоматизації, робототехніки та безпекової інженерії Харківського національного університету радіоелектроніки

Смагіна О.О. – кандидат педагогічних наук, доцент, доцент кафедри інформаційних технологій та систем ДЗ «Луганський національний університет імені Тараса Шевченка»

Сорокін Д.А. – аспірант кафедри конструювання електроннообчислювальної апаратури Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Субін А.А. – кандидат технічних наук, доцент, доцент кафедри технології машинобудування Навчальнонаукового механікомашинобудівного інституту Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Суховій С.І. – кандидат технічних наук, доцент, доцент кафедри конструкції авіаційних двигунів Національного аерокосмічного університету «Харківський авіаційний інститут»

Терейковський І.А. – доктор технічних наук, професор, професор кафедри системного програмування і спеціалізованих комп'ютерних систем факультету програмних систем та прикладної математики Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Теслюк С.І. – старший викладач кафедри комп'ютерноінтегрованих технологій, автоматизації, робототехніки та безпекової інженерії Харківського національного університету радіоелектроніки

Торстенссон О. – викладач, Хальмстадський університет, м. Хальмстад, Швеція

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»

Черепанська І.Ю. – доктор технічних наук, професор, професор кафедри Автоматизації та систем неруйнівного контролю Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Черкаський Д.О. – аспірант Національного технічного університету «Дніпровська політехніка»

Якименко М.С. – кандидат фізикоматематичних наук, доцент, завідувач кафедри вищої математики та фізики Центральноукраїнського національного технічного університету

Янко А.С. – кандидат технічних наук, доцент, доцент кафедри комп'ютерних та інформаційних технологій і систем Національного університету «Полтавська політехніка імені Юрія Кондратюка»

Яцишин А.В. – доктор технічних наук, старший науковий співробітник, заступник директора з наукової роботи Центру інформаційноаналітичного та технічного забезпечення моніторингу об'єктів атомної енергетики Національної академії наук України

НОТАТКИ

Науковий журнал

**ВЧЕНІ ЗАПИСКИ
ТАВРІЙСЬКОГО НАЦІОНАЛЬНОГО УНІВЕРСИТЕТУ
ІМЕНІ В.І. ВЕРНАДСЬКОГО**

Серія: Технічні науки

Том 37 (76) № 2 2026

Частина 1

Коректура • *Н. Славогородська*

Комп'ютерна верстка • *Н. Кузнєцова*

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. 34,49. Ум. друк. арк. 38,6.

Зам. № 0426/305. Наклад 150 прим.

Дата розміщення онлайн: 11.05.2026. Дата друку: 18.05.2026

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.