



Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Ministry of Education and Science of Ukraine
Kharkiv National University of Radio Electronics

Автоматизовані системи управління та прилади автоматики

Випуск 188

*Всеукраїнський міжвідомчий
науково-технічний збірник*

Заснований у 1965 р.

Затверджений до друку
Науково-технічною Радою
Харківського національного
університету радіоелектроніки
(протокол № 3 від 10 березня 2026 р.)

ЗАСНОВНИК

Харківський національний
університет радіоелектроніки,

АДРЕСА РЕДАКЦІЇ:

Україна, 61166, м. Харків, проспект Науки, 14
Інформаційний сайт: <https://asu-pa.nure.ua>
E-mail редколегії: asu@nure.ua

Management Information System and Devices

Issue 188

*All-Ukrainian Interdepartmental
Scientific and Technical Collection*

Founded in 1965

Approved for publication
by the Scientific and Technical Council
of the Kharkiv National
University of Radio Electronics
(Protocol No. 3 d.d. 10 March 2026)

ESTABLISHER

Kharkiv National
University of Radio Electronics,

EDITORIAL OFFICE ADDRESS:

14 Nauky Ave., Kharkiv, Ukraine, 61166
Information site: <https://asu-pa.nure.ua>
E-mail of the editorial board: asu@nure.ua

Ідентифікатор медіа R30-03874

Витяг з реєстру суб'єктів у сфері медіа – реєстрантів від 25.04.2024 № 1410

*Журнал включено до Переліку наукових фахових видань України, в яких можуть публікуватися
результати дисертаційних робіт на здобуття наукових ступенів доктора і кандидата наук
наказом Міністерства освіти і науки України від 16.07.2018 №775 (додаток 7)*

Харків – 2026

РЕДАКЦІЙНА КОЛЕГІЯ

Головний редактор

Рубан Ігор Вікторович,
доктор технічних наук, професор

Заступник головного редактора

Левикін Віктор Макарович,
доктор технічних наук, професор

Члени редколегії:

Волк Максим Олександрович,
доктор технічних наук, професор;

Гребеннік Ігор Валерійович,
доктор технічних наук, професор;

Євсєєв Владислав В'ячеславович,
доктор технічних наук, професор;

Єрохін Андрій Леонідович,
доктор технічних наук, професор;

Каргін Анатолій Олексійович,
доктор технічних наук, професор;

Кошовий Микола Дмитрович,
доктор технічних наук, професор;

Невлюдов Ігор Шакирович,
доктор технічних наук, професор;

Новоселов Сергій Павлович,
доктор технічних наук, доцент;

Осадчий Сергій Іванович,
доктор технічних наук, професор;

Пітерська Варвара Михайлівна,
доктор технічних наук, професор;

Удовенко Сергій Григорович,
доктор технічних наук, професор;

Федорович Олег Євгенович,
доктор технічних наук, професор;

Філатов Валентин Олександрович,
доктор технічних наук, професор;

Филипенко Олександр Іванович,
доктор технічних наук, професор;

Варламіс Іракліс,
PhD, Professor;

Девадзе Давид,
Candidate of Sciences in Physics and Mathematics, Professor;

Сергієнко Олег,
доктор технічних наук, професор;

Койфман Олексій Олександрович,
кандидат технічних наук, доцент;

Семенов Сергій,
доктор технічних наук, професор

Відповідальний секретар:

Цимбал Олександр Михайлович,
доктор технічних наук, професор;

EDITORIAL BOARD

Editor in Chief

Ihor Ruban,
Doctor of Technical Sciences, Professor

Deputy Chief Editor

Viktor Levykin,
Doctor of Technical Sciences, Professor

Editorial Board Members:

Maksym Volk,
Doctor of Technical Sciences, Professor;

Igor Grebennik,
Doctor of Technical Sciences, Professor;

Vladyslav Yevsieiev,
Doctor of Technical Sciences, Professor;

Andriy Yerokhin,
Doctor of Technical Sciences, Professor;

Anatolii Kargin,
Doctor of Technical Sciences, Professor;

Mykola Koshovyi,
Doctor of Technical Sciences, Professor;

Igor Nevliudov,
Doctor of Technical Sciences, Professor;

Sergiy Novoselov,
Doctor of Technical Sciences, Associate Professor;

Sergei Osadchy,
Doctor of Technical Sciences, Professor;

Varvara Pitera,ska,
Doctor of Technical Sciences, Professor;

Serhiy Udoenko,
Doctor of Technical Sciences, Professor;

Oleg Fedorovich,
Doctor of Technical Sciences, Professor;

Valentin Filatov,
Doctor of Technical Sciences, Professor;

Oleksandr Filipenko,
Doctor of Technical Sciences, Professor;

Iraklis Varlamis,
PhD, Professor;

David Devadze,
Candidate of Sciences in Physics and Mathematics, Professor;

Oleg Sergiyenko,
DSc, Professor;

Oleksiy Koyfman,
Candidate of Technical Sciences, Associate Professor;

Serhii Semenov,
Doctor of Technical Sciences, Professor

Responsible secretary:

Oleksandr Tsybmal,
Doctor of Technical Sciences, Professor;

ЗМІСТ

Інтелектуальні інформаційні технології

- Ахундов Р. Г., Гашимов Е. Г.* 5
 en
 Моделювання інформаційних процесів та визначення вимірюваних вимог при проектуванні систем фізичного захисту
- Гребеннік І. В., Коваленко О. А.* 17
 en
 Використання штучних нейронних мереж для управління сортуванням посилок у конвеєрному потоці
- Неліпа О. Д., Калита Н. І.* 33
 ua
 Гібридні моделі штучного інтелекту в задачах автоматизованого оцінювання достовірності результатів валідації інформаційних систем
- Панченко В. І., Кучук Г. А.* 44
 en
 Теоретико-алгебраїчні основи еволюційного синтезу тестів компонентів інтелектуальних шлюзів високощільних IoT-систем
- Петров К. Е., Божко О. Ю.* 54
 ua
 Метод автоматичної розмітки ознак структурної складності документів із використанням великих мовних моделей
- Пилипенко А. Г.* 69
 ua
 Модель динамічної активації кореляційних зв'язків в онтологічних графах в умовах опрацювання різномірних даних
- Ткачов В. М.* 78
 en
 Метод відновлення критичних сервісів інформаційної системи на мобільній платформі в умовах керованої деградації
- Чалий С. Ф., Лещинська І. О.* 98
 ua
 Нейросимвольний фреймворк формування ментальних моделей рішень інтелектуальних систем

Комп'ютерні системи, мережі та вбудовані технології

- Корнієнко Є. Д., Ляшенко О. С.* 108
 en
 Дослідження методологічних основ упровадження блокчейну й смарт-контрактів в електроенергетичних мікромережах
- Усик В. В., Федорчак Б. П.* 121
 en
 Аналіз етапів акустичної експертизи замкнених приміщень

CONTENTS

Intelligent information technology

- Akhundov R., Hashimov E.*
 Modeling Information Processes and Deriving Measurable Requirements in Physical Protection System Design
- Grebennik I., Kovalenko O.*
 Use of artificial neural networks for managing the sorting of parcels in a conveyor flow
- Nelipa O., Kalyta N.*
 Hybrid models of artificial intelligence in automated assessment tasks for the reliability of information system validation results
- Panchenko V., Kuchuk H.*
 Theoretical-algebraic basics of evolutionary test synthesis for gateways components of high-density IoT systems
- Petrov K., Bozhko O.*
 Method of automatic labeling for signs of structural complexity in documents using large language models
- Pylypenko A.*
 Model of dynamic activation of correlation relations in ontological graphs when processing heterogeneous data
- Tkachov V.*
 Method of restoring critical information system services on a mobile platform under conditions of controlled degradation
- Chalyi S., Leshchynska I.*
 A neurosymbolic framework for forming mental models of intelligent system decisions

Computer systems, networks and built-in technologies

- Korniienko Y., Liashenko O.*
 A study of the methodological foundations for implementing blockchain and smart contracts in electric power microgrids
- Usik V., Fedorchak B.*
 Analysis of the stages of conducting an acoustic examination of closed rooms

**Програмні системи
та безпека інформаційних технологій**

- Прокопенко О. С., Смеляков С. В.** 129
ua
Розроблення дескриптора зображень
для ефективного пошуку на основі вмісту
у великих сховищах даних
- Репіхов В. М., Чуприна А. С.** 145
ua
Метод превентивного супроводження
програмного забезпечення
на основі компараторної моделі

**Системне моделювання
та обчислювальні методи**

- Левикін В. М., Дмитревський Д. В.** 163
en
Метод розподілу виробничих ресурсів
у кондитерській промисловості
- Пащенко Р. Е.** 177
ua
Автоматизоване виявлення аномалій
на космічних знімках сільськогосподарських
земель із використанням фрактального аналізу
- Полупан Ю. В., Малєєва О. В., Кікоть М. С.** 191
en
Моделювання й аналіз логістичних маршрутів
постачання та збуту високотехнологічної
продукції з огляду на різні види транспорту
- Алфавітний покажчик** 203

**Software systems
and information technology security**

- Prokopenko O., Smelyakov S.**
Development of an image descriptor
for effective content-based retrieval
in large data repositories
- Repikhov V., Chupryna A.**
Method of preventive software
maintenance based
on a comparator model

**Sistem modeling
and computational method**

- Levykin V., Dmytrevskiy D.**
Improving the method of optimizing the distribution
of production resources in confectionery production
- Pashchenko R.**
Automated discovery anomalies
on the agricultural earths spaces pictures
with use fractal analysis
- Polupan Yu., Malyeyeva O., Kikot M.**
Modeling and analysis of logistics supply
and distribution routes for high-tech products
considering different modes of transport
- Alphabetical index**

Ramil Akhundov, Elshan Hashimov

MODELING INFORMATION PROCESSES AND DERIVING MEASURABLE REQUIREMENTS IN PHYSICAL PROTECTION SYSTEM DESIGN

The study examines information processes in the conceptual design of physical protection systems and their role in transforming sensor outputs, operator inputs, and communication flows into timely protective action. Attention is focused on the fact that design sufficiency is constrained not only by sensors, barriers, and response forces, but also by delays and uncertainty arising between detection, decision, and dispatch. The purpose of the study is to formalize information processes as explicit design objects and to develop an approach for deriving measurable information requirements that can be justified and verified at the conceptual design stage. The objectives are to define a reference workflow from event generation to response activation, to decompose the total information-to-action time into operational stages, to incorporate uncertainty related to false alarms, ambiguity, workload, and degraded communications, and to link requirement statements with acceptance evidence. The research uses a conceptual and methodological approach based on systems analysis, scenario decomposition, latency structuring, and traceability mapping between threat scenarios, bottlenecks, requirement targets, and verification evidence. The study develops a structured information-to-action model that makes explicit the stages of sensing, validation, fusion, decision-making, communication, and dispatch. On this basis, a method is proposed for translating scenario-specific bottlenecks into verifiable requirements for timeliness, accuracy, completeness, and resilience. The study also identifies practical forms of acceptance evidence, including timed drills, log-based measurements, stress testing, and simulation-supported assessment. The results show that conceptual design becomes more defensible when information processes are modeled explicitly rather than treated as implicit assumptions. The proposed approach enables designers to justify measurable requirements, reveal critical latency sources, and support revalidation of design sufficiency under changing operational conditions.

Keywords: physical protection system; conceptual design; information processes; alarm validation; data fusion; command and control; communications resilience; acceptance evidence.

1. Introduction

Physical protection systems are typically described in terms of tangible components: sensors, barriers, access control devices, and response forces [1–3]. While these elements are necessary, they are not sufficient to explain why a system succeeds or fails under real adversarial conditions. In practice, operational outcomes are often determined by information processes that connect detection to action. The same sensor coverage can produce radically different security performance depending on how signals are validated, how information is fused across sources, how quickly decisions are made, how reliably communications propagate directives, and how response resources are dispatched under uncertainty and workload [4, 5].

A persistent methodological weakness in conceptual design is the implicit assumption of near-perfect information [6]. Early design stages frequently treat detection as an instantaneous event, decision-making as a negligible delay, and communications as a transparent channel. This assumption is convenient for describing

system architecture, but it is risky for requirement justification because it masks the dominant sources of failure. False alarms can saturate operators and erode trust in alerts. Ambiguous signals can trigger prolonged validation cycles. Degraded communications can disrupt escalation and dispatch. Insider-enabled actions can bypass traditional detection cues. Under such conditions, the bottleneck is not the physical layer but the information-to-action loop that governs the timing and correctness of intervention [6, 7].

The problem addressed in this paper is that information processes are often not treated as first-class design objects at the conceptual stage. As a result, information requirements are either absent or expressed as generic aspirations rather than measurable thresholds. This gap weakens review defensibility. A design may be compliant on paper and still be operationally insufficient because the information chain cannot deliver timely and reliable decision support for the detection–decision–response sequence. Without a structured model of information flows and latencies, it is difficult to justify why particular sensor configurations, staffing levels,

alarm-management policies, or command-and-control procedures are sufficient for representative scenarios.

This study argues that conceptual design must explicitly model information processes, including their latency and uncertainty, and must translate them into verifiable requirements. The core idea is straightforward: sufficiency is time-conditioned, and information delays consume the time margin needed to interrupt an adversary before the critical element is reached. Therefore, conceptual design should specify not only what is installed, but how information is generated, transformed, transmitted, interpreted, and acted upon, with explicit acceptance targets for timeliness, reliability, completeness, and resilience under degraded conditions.

The aim of this paper is to formalize information processes within conceptual physical protection design and to provide a reproducible method for deriving measurable information requirements that support system-level sufficiency. The objectives are to define a reference information workflow from sensing to decision to response, decompose the information-to-action timeline into measurable components, represent key uncertainty sources such as false alarms and degraded communications, and establish a traceability structure that links information requirements to scenario needs and verification evidence.

The remainder of the paper is organized as follows. The next section reviews related approaches in physical protection, command-and-control, and alarm management and clarifies where information processes are under-modeled. The methods section presents the information-process model, latency decomposition, and requirement derivation logic. The results section provides measurable requirement templates and practical artifacts for verification and monitoring. The paper concludes with discussion of implications, limitations, and directions for future work.

2. Background And Related Approaches

Conceptual design of physical protection systems has traditionally been dominated by prescriptive and component-oriented approaches. Normative frameworks provide minimum requirements for barriers, access control, detection devices, and response arrangements, enabling consistency and auditability across portfolios of facilities [8, 9]. This baseline remains essential, particularly for high-consequence settings, because it prevents omission of critical protective functions. However, normative specifications frequently treat

information as an implicit enabler rather than as an explicit design object. Requirements tend to describe what hardware must exist and what procedures must be documented, while the operational performance of the information-to-action loop, how quickly and reliably an alarm becomes a validated decision and an executed response, is under-specified [10–12].

The classical detection–delay–response doctrine provides an operational logic that highlights why information processes matter [13, 14]. Detection must occur early enough to initiate the defender chain, delay must create the time reserve needed for response, and response must be executed before the adversary reaches the target. In this chain, information is not merely a front-end sensor output. It is the continuous input to decision-making and coordination: the defender cannot exploit delay if alarms are not validated promptly, and response effectiveness degrades if communications are unreliable or if command-and-control introduces friction. Yet, in much of the applied practice, information processes are treated as a background assumption, with decision latency and communications reliability either neglected or treated as fixed constants.

Research on situation awareness and decision-making in military and security operations provides a more explicit treatment of information processing, particularly through the concepts of perception, comprehension, and projection as prerequisites for action [15, 16]. Within this literature, performance is strongly influenced by information quality, cognitive workload, ambiguity management, and the design of human–machine interfaces. Alarm fatigue and operator overload are well-known phenomena that can substantially increase validation and decision time, even when sensors technically function as specified. These insights are directly relevant to physical protection design, but they are often not integrated into the conceptual requirement-setting process, which still privileges hardware-centric measures [17, 18].

A related body of work addresses sensor data fusion, alarm management, and security operations center workflows. Multi-sensor fusion can reduce uncertainty and improve classification, but it can also introduce additional processing steps and latency, especially when signals are inconsistent or when the system is configured conservatively to reduce false positives. Alert management policies, such as escalation rules, acknowledgement procedures, and prioritization, largely determine the effective time to take action [19]. The literature also emphasizes that communications and command-and-

control are not transparent pipes: network congestion, degraded links, and procedural bottlenecks can delay dispatch and reduce coordination quality, particularly during multi-event situations or intentional diversion.

Across these approaches [20–22], recurring failure modes can be traced to information processes rather than to physical components alone. False alarms can saturate operators and cause delays in acknowledging genuine events. Ambiguous detections can trigger prolonged verification cycles [23, 24]. Degraded communications can fragment the operational picture and slow or misdirect response. Insider-assisted actions can bypass expected signatures, forcing reliance on procedural checks and internal monitoring. These failure modes are consistent with "compliant but insufficient" outcomes: a facility may meet prescriptive equipment and staffing requirements, yet still fail operationally because the information loop cannot deliver timely, reliable decision support under representative conditions.

Scenario-based and risk-informed design approaches partially address these issues by emphasizing operational context [23]. They encourage designers to consider how different threat mechanisms stress different subsystems, including decision-making. Nevertheless, many scenario-based treatments still stop short of defining measurable information requirements. Scenarios are used to justify security measures, but the information processes that link information discovery to action are not formalized in a way that allows for the establishment of acceptance criteria, such as the maximum allowable verification time for a given alarm load, the minimum communication availability under degraded conditions, or the required classification accuracy for a given threat class [25–27]. The gap addressed by this paper is therefore specific: existing approaches recognize that information matters, but they lack a unified conceptual model that treats information processes as first-class design objects and converts them into verifiable requirements tied to system-level sufficiency. The contribution pursued here is to formalize the information workflow from sensing to decision to response, decompose its latency and uncertainty components, and provide a traceable method to derive measurable information requirements and evidence pathways that can be integrated with physical protection design from the earliest stages.

3. Material and Methods

This study develops a conceptual and methodological framework that treats information

processes as explicit design objects in the conceptual design of physical protection systems. Two analytical layers are distinguished. First, a reference information-process model is introduced to describe how information moves from sensing to response activation in a protected facility. Second, a Scenario-Driven Information Requirement Derivation Method is defined to translate the modeled information process into measurable, verifiable, and traceable requirement statements. The scope is methodological rather than facility-specific: the purpose is not to calibrate a single site, but to provide a reproducible structure that can later be instantiated with site data, operating assumptions, and threat scenarios.

The protected facility is represented as a socio-technical system comprising sensing assets, human operators, procedures, command-and-control functions, communications channels, and response forces. In this study, command-and-control (C2) denotes the set of functions responsible for decision-making, authorization, coordination, and dispatch based on validated security information. Information is modeled as a bounded forward process that links observation to action through a sequence of operational stages. These stages include event generation and sensing, alarm validation, multi-source fusion, threat classification, decision and authorization, communication of directives, dispatch and mobilization, and response activation. A feedback loop is also acknowledged conceptually because logs, after-action review, and parameter updates are necessary for lifecycle revalidation, but the formal part of the framework focuses on the forward information-to-action path.

A reference information-process model used in conceptual physical protection design is presented in Figure 1. The figure describes only the forward information-to-action path and the lifecycle feedback loop; the derivation of measurable requirements is introduced separately below.

Figure 1 shows how information moves from heterogeneous sources to response activation through detection, validation, fusion, decision, communication, and dispatch stages. In this representation, C2 denotes the command-and-control functions responsible for decision-making, authorization, coordination, and dispatch.

Let S denote the set of information sources, including intrusion sensors, access-control events, video analytics, operator reports, and system-health telemetry. Let $m \in M$ denote information messages generated by these sources and transmitted through communications channels under a given operating context. The operating

context includes visibility, weather, staffing posture, communications status, operator workload, and the presence of concurrent incidents. Within this representation,

the reference model serves a descriptive role: it makes explicit where information is created, transformed, delayed, degraded, or lost before a protective action is activated.

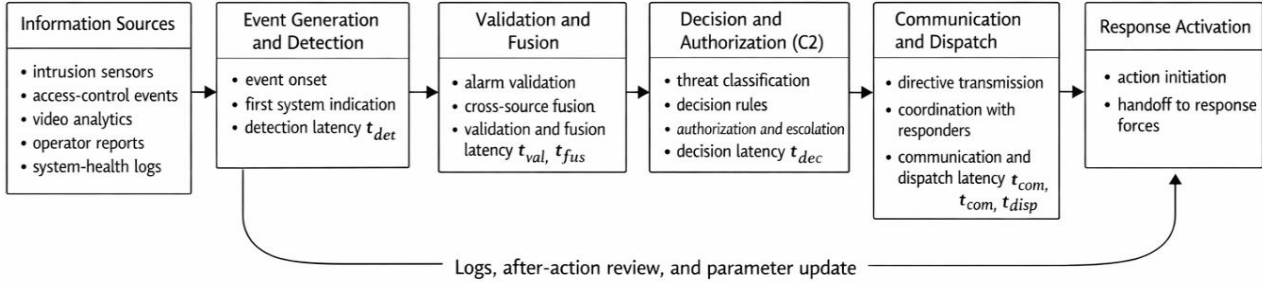


Fig. 1. Reference information-process model for conceptual physical protection system design

To avoid ambiguity, the principal symbols and operational metrics are defined before formal derivation. For a scenario s , let $T_{info,s}$ denote the total information-to-action time. Let $t_{det,s}$ denote detection latency from event onset to first system indication; $t_{val,s}$, alarm validation time from first indication to confirmed event status; $t_{fus,s}$, multi-source fusion time; $t_{dec,s}$, decision and authorization latency; $t_{com,s}$, communications latency for disseminating directives and situational context; and $t_{disp,s}$, dispatch and mobilization time from decision to response activation. In addition, let $P_{D,s}$ denote the probability of timely detection within a specified time window, $P_{C,s}$ the probability of correct classification, FAR_s the false alarm rate, $A_{com,s}$ communications availability, $L_{loss,s}$ message loss rate, and W_s an operator-workload indicator expressed, for example, through alert arrival rate or queue length. The symbol $Pr(\cdot)$ denotes probability.

Information quality is represented through a minimal set of dimensions that are both operationally meaningful and measurable at the conceptual design stage: timeliness, accuracy, completeness, integrity, and availability. Timeliness is captured through the latency structure of the information-to-action chain. Accuracy is represented through timely-detection and correct-classification probabilities. Completeness characterizes whether the decision-maker receives sufficient context to choose and authorize an action without excessive additional verification. Integrity addresses the trustworthiness of information and resistance to corruption or misleading inputs. Availability reflects whether information and communications

remain usable under degraded or adversarial conditions. These dimensions are treated as design constraints rather than as qualitative aspirations because the objective is to justify measurable requirement targets at an early design stage.

The reference model is made operational through an explicit decomposition of the information-to-action time into stage-specific latency components. For scenario s , the total information-to-action time is represented as

$$T_{info,s} = t_{det,s} + t_{val,s} + t_{fus,s} + t_{dec,s} + t_{com,s} + t_{disp,s}. \quad (1)$$

Equation (1) states that the total delay between event onset and response activation is the sum of the detection, validation, fusion, decision, communications, and dispatch latencies. This decomposition is aligned with sufficiency-oriented defender timelines because information delays directly consume the time margin available for interruption. At the conceptual design stage, the key implication is that information processes must not be treated as negligible constants. Their duration is scenario-dependent and may become dominant under ambiguity, high workload, multi-event concurrency, or degraded communications.

Uncertainty is represented in two complementary ways depending on the available evidence base. When historical logs, controlled tests, or validated simulations support probabilistic characterization, latency terms and performance measures are treated as random variables and requirement statements are expressed through chance constraints. A generic probabilistic acceptance form is

$$Pr\left(T_{info,s} \leq T_{info,s}^{max}\right) \geq 1 - \varepsilon_s, \quad (2)$$

where $T_{info,s}^{max}$ is the maximum admissible information-to-action time for scenario s , and ε_s is a scenario-specific tolerance for non-compliance that reflects consequence severity and operational risk posture. When only selected

stages are dominant for a given scenario, the requirement may be stated in stage-specific form, for example,

$$Pr(t_{val,s} + t_{dec,s} + t_{com,s} \leq T_{proc,s}^{max}) \geq 1 - \varepsilon_s, \quad (3)$$

where $T_{proc,s}^{max}$ is the maximum admissible time allocated to the validation, decision, and communication stages under the scenario considered.

Table 1. Information metrics catalogue for conceptual PPS design

Metric / variable	Definition	Measurement source	Typical acceptance target form
t_{det}	Detection latency from event onset to first system indication	Sensor timestamps; event logs	$t_{det} \leq \overline{t_{det}}(s)$
t_{val}	Alarm validation time from first indication to confirmed event status	SOC logs; operator acknowledgement records	$t_{val} \leq \overline{t_{val}}(s)$
t_{fus}	Multi-source fusion time to consolidate signals into a coherent event	Fusion engine logs; analytics pipeline logs	$t_{fus} \leq \overline{t_{fus}}(s)$
t_{dec}	Decision and authorization latency to select and approve an action	C2 system logs; dispatch approvals	$t_{dec} \leq \overline{t_{dec}}(s)$
t_{com}	Communications latency for disseminating directives and situational context	Network telemetry; message delivery logs	$t_{com} \leq \overline{t_{com}}(s)$
t_{disp}	Dispatch and mobilization time from decision to response activation	Dispatch logs; radio logs; unit status logs	$t_{disp} \leq \overline{t_{disp}}(s)$
T_{info}	Total information-to-action time $t_{det} + t_{val} + t_{fus} + t_{dec} + t_{com} + t_{disp}$	Integrated logs across SOC and C2	$T_{info} \leq \overline{T_{info}}(s)$ or $Pr(\cdot) \geq 1 - \varepsilon_s$
P_D	Probability of timely detection within specified window	Test campaigns; labeled data; simulation	$P_D \geq \underline{P}_D(s)$
P_C	Probability of correct classification (threat vs nuisance, class label)	Labeled data; red-team drills; analytics QA	$P_C \geq \underline{P}_C(s)$
FAR	False alarm rate per sensor or per zone per time unit	SOC alarm logs; analytics output	$FAR \leq \overline{FAR}(s)$
A_{com}	Communications availability (successful delivery ratio)	Network monitoring; message receipts	$A_{com} \geq \underline{A}_{com}(s)$
L_{loss}	Message loss rate under nominal/degraded conditions	Network telemetry; stress tests	$L_{loss} \leq \overline{L}_{loss}(s)$
W	Operator workload indicator (alerts per unit time, queue length)	SOC dashboards; ticketing system	$W \leq \overline{W}(s)$ or queue stability criterion
ε_s	Risk tolerance for probabilistic acceptance	Governance parameter	$Pr(T_{info} \leq \overline{T_{info}}) \geq 1 - \varepsilon_s$

When probabilistic estimation is not sufficiently supported by data, uncertainty is handled through conservative bounding rather than by unsupported precision. In such cases, design targets are assigned explicit margins so that the worst plausible latency values under the defined operating context remain within acceptance limits. This approach avoids conceptually fragile designs that are sufficient only under nominal assumptions and provides a practical bridge between early-stage design and later empirical refinement.

The reference information-process model is descriptive. Requirement derivation is performed by a separate procedure, here termed the Scenario-Driven Information Requirement Derivation Method. Its purpose is to convert scenario-specific information bottlenecks into measurable, verifiable, and traceable requirement statements. The method proceeds through five steps.

First, a bounded set of representative scenario classes is defined. These classes are not intended to exhaust all possible attacks, but to capture the dominant operational stressors that shape the information process. Typical classes include covert intrusion, forced entry, insider-assisted action, diversion plus main attack, coordinated multi-point attack, communications degradation, false alarm surge, and sensor degradation or partial failure. Second, for each scenario class, the dominant information bottleneck is identified. This bottleneck may arise from ambiguity-driven validation delay, false alarm saturation, classification difficulty, decision congestion, communications instability, message loss, or prioritization conflict during concurrent events.

Third, the dominant bottleneck is mapped to one or more measurable metrics. For example, ambiguity-driven scenarios primarily map to $t_{val,s}$, $t_{fus,s}$, and $P_{C,s}$; false

alarm surge maps to FAR_s , W_s , and validation delay under load; degraded communications map to $t_{com,s}$, $A_{com,s}$, and $L_{loss,s}$. Fourth, acceptance thresholds are assigned to the selected metrics in a form compatible with design review and later verification. The resulting scenario-specific requirement set may be expressed as

$$R_s = \{T_{info,s}^{max}, P_{D,s}^{min}, P_{C,s}^{min}, FAR_s^{max}, A_{com,s}^{min}, L_{loss,s}^{max}, W_s^{max}\},$$

where R_s is the requirement set for scenario s ; $P_{D,s}^{min}$ is the minimum acceptable timely-detection probability; $P_{C,s}^{min}$ the minimum acceptable correct-classification probability; FAR_s^{max} the maximum acceptable false alarm rate; $A_{com,s}^{min}$ the minimum acceptable communications availability; $L_{loss,s}^{max}$ the maximum acceptable message loss rate; and W_s^{max} the maximum acceptable operator-workload level. The specific content of R_s may vary by scenario, but the principle remains constant: every requirement must be tied to an identified bottleneck and stated in measurable form.

Fifth, each requirement is paired with an evidence pathway that makes later acceptance testing or operational monitoring possible. This step converts requirement statements from abstract design intentions into review-defensible claims. The method therefore links each scenario not only to a bottleneck and a metric, but also to a verification route that can be exercised during testing, commissioning, or lifecycle revalidation.

The final element of the methodology is a traceability structure that connects scenario classes, information bottlenecks, measurable requirements, and acceptance evidence. Time-bound requirements are primarily associated with instrumented timed drills, synchronized event logs, and dispatch records. Accuracy-related requirements are associated with labeled datasets, controlled red-team exercises, and validated classification tests. Resilience requirements under concurrency, false alarm load, or degraded communications are associated with stress testing, injected-alarm campaigns, network degradation exercises, and simulation-supported assessment when repeated empirical trials are impractical.

This traceability structure has two functions. Methodologically, it ensures that every requirement introduced at the conceptual stage has an identifiable operational motivation and an evidentiary pathway. Practically, it supports lifecycle revalidation because

the same metrics that justify the requirement during design can later be monitored through logs, tests, and after-action review. In this way, information-process requirements become explicit, testable, and revisable components of conceptual physical protection design rather than implicit assumptions embedded in hardware-oriented descriptions.

4. Results

This study yields three methodologically connected outputs for conceptual physical protection system design:

- 1) a reference information-process model that makes the information-to-action chain explicit;
- 2) a scenario-specific set of measurable information requirements derived from that model;
- 3) a traceability structure linking each requirement to a preferred acceptance evidence pathway.

Reference information-process model as a design object

The first result is a reference information-process model that represents information not as a background assumption, but as an explicit part of conceptual design. In this representation, the information chain is structured as a bounded sequence of operational stages extending from event generation and detection to validation, fusion, decision, communication, dispatch, and response activation. As described in Section 3.1, this model captures the forward information-to-action path and the associated lifecycle feedback through logs, after-action review, and parameter update.

The practical value of the model lies in its ability to make time consumption and degradation points visible at the conceptual stage. Instead of treating detection as a single instantaneous event, the model distinguishes between the time required to generate an indication, the time needed to validate and interpret that indication, and the additional time consumed by decision-making, authorization, communication, and dispatch. This separation is important because, in representative operational settings, dominant delays frequently occur after the first alarm indication rather than at the point of sensing itself.

When used as a design object, the reference model supports structured reasoning about where information may be delayed, degraded, or lost. It therefore provides a common analytical basis for identifying bottlenecks associated with alarm ambiguity, false alarm load,

concurrent incidents, communication degradation, and operator workload. In this sense, the model is descriptive rather than prescriptive: it defines the information path that must be analyzed before measurable requirements can be assigned.

Scenario-specific measurable information requirements

The second result is a scenario-specific requirement set that expresses information performance in measurable rather than declarative terms. Using the latency decomposition introduced in Eq. (1), the total information-to-action time is treated as a structured sum of stage-specific delays. The probabilistic acceptance forms in Eqs. (2) and (3) then allow these delays to be bounded either at the total-process level or at the level of dominant processing stages, depending on the operational character of the scenario.

On this basis, measurable requirements can be written as explicit performance targets for timeliness, accuracy, and resilience. In practical terms, the requirement set includes bounded total information-to-action time, bounded validation and decision intervals, minimum timely-detection probability, minimum correct-classification probability, maximum false alarm rate, minimum communications availability, and maximum tolerable message loss and workload level, as formalized by Eq. (4). The advantage of this structure is that it transforms conceptual design statements into reviewable and testable claims.

The requirements are scenario-sensitive rather than universal. Different scenario classes stress different parts of the information process and therefore generate different requirement priorities. In ambiguity-dominated scenarios, the primary design concern is not only prompt detection, but also rapid validation and sufficiently accurate classification. In communication-degraded scenarios, communications latency, availability, and message loss become dominant. In scenarios characterized by false alarm surges or concurrent events, operator workload and sustained validation performance under load become central requirement categories. The method therefore does not impose a single fixed target set for all cases. Instead, it produces a structured requirement profile whose content depends on the dominant information bottleneck of the scenario under consideration.

A further result is that these requirements can be stated without collapsing conceptual design into

unsupported numerical precision. Where validated logs, controlled trials, or simulation-supported evidence are available, the requirement set can be expressed probabilistically in the form introduced in Eqs. (2) and (3). Where such support is limited, the same structure remains usable through conservative bounding and explicit safety margins. This makes the framework applicable both in early-stage design and in later refinement phases.

Traceability from scenario class to evidence pathway

The third result is a traceability structure that connects scenario classes, dominant information stressors, measurable requirements, and acceptance evidence. This result is important because conceptual requirements are not credible unless they can be linked both to an operational need and to a feasible verification route. In the proposed framework, each requirement is justified by a scenario-defined bottleneck, quantified through one or more metrics, and paired with a corresponding evidence pathway.

This traceability logic allows requirement statements to remain operationally meaningful. A bounded validation time is linked to ambiguity management and can be assessed through instrumented timed drills or synchronized event logs. A classification requirement is linked to interpretation quality and can be supported by labelled test events, controlled exercises, or validated analytics assessment. A communications availability target is linked to degraded-network scenarios and is supported by telemetry, stress testing, or injected degradation exercises. In this way, the requirement object is not limited to a threshold value, but includes the rationale for the threshold and the means by which later acceptance can be demonstrated.

To make the traceability structure operational, Table 2 maps representative scenario classes to dominant information stressors, corresponding requirement priorities, and preferred acceptance evidence.

Table 2 summarizes this traceability structure in operational form. It shows that scenario classes differ not only by external threat mechanism, but also by the type of information stress they impose on the protection process. This mapping provides a disciplined basis for selecting requirement priorities and matching them with appropriate evidence pathways at the conceptual design stage.

Table 2. Scenario class to information-requirement mapping and preferred evidence

Scenario class	Dominant information stressor	Primary information requirements (examples)	Preferred acceptance evidence
Covert intrusion	Weak signatures, ambiguity, delayed confirmation	Tight bounds on t_{det} and t_{val} ; minimum P_D ; improved P_C for low-signature events	Instrumented detection tests; labeled video/radar datasets; simulation on stealth paths
Forced entry	Clear signatures but short time window; rapid escalation	Bound t_{det} ; constrain $t_{dec} + t_{disp}$ to preserve time margin; reliable communications A_{com}	Timed drills; dispatch logs; communications delivery tests
Insider-assisted	Boundary collapse, atypical signatures, access-event dominance	Strong P_C for credential misuse patterns; low t_{val} for access anomalies; integrity/availability of logs	Access-control audit; red-team drills; log forensics and anomaly testing
Diversion plus main attack	Concurrent alarms, saturation, prioritization conflict	Queue stability ($workload(W)$); bound $t_{val} + t_{dec}$ under load; probabilistic acceptance $Pr(T_{info} \leq \overline{T_{info}}) \geq 1 - \epsilon_s$	Multi-event exercises; stress tests with injected alarms; simulation or hybrid evidence
Coordinated multi-point attack	C2 overload, ambiguity, competing priorities	Bound t_{dec} and t_{com} under concurrency; resilience of dissemination A_{com}	Coordinated drills; C2 workload tests; simulation of simultaneous incidents
Communications degradation	Delayed/failed message delivery, fragmented picture	Availability A_{com} and loss L_{loss} targets; tighter bounds on t_{com} ; conservative margins in T_{info}	Network stress tests; degraded-mode exercises; telemetry-based validation
False-alarm surge	Operator fatigue, trust erosion, long validation	FAR limit; bound t_{val} as a function of W ; improved P_C to reduce nuisance	Log-based measurement; controlled alarm-injection tests; analytics QA
Sensor degradation / partial failure	Increased miss probability, delayed detection	Maintain P_D above threshold under degraded state; redundancy targets; fallback procedures reflected in t_{val} and t_{dec}	Degradation tests; redundancy checks; simulation with failure modes

Practical implications for conceptual design review

Taken together, the three outputs strengthen conceptual design in two ways. First, they make information performance reviewable at the same level of rigor as sensors, barriers, and response arrangements. Second, they provide a structured basis for lifecycle revalidation because the same metrics used to justify conceptual requirements can later be monitored through logs, drills, and stress testing.

As a result, the proposed framework supports a shift from hardware-centered sufficiency claims to evidence-oriented information-process justification. Conceptual design is thereby strengthened not through the addition of complexity for its own sake, but through explicit recognition that information delay, ambiguity, and degradation are often decisive determinants of protective performance.

5. Discussion

The findings of this study indicate that information processes should be treated as a primary design concern in conceptual physical protection system design

rather than as a secondary implementation detail. The proposed reference information-process model shows that operational sufficiency depends not only on the presence of sensors, barriers, and response forces, but also on the ability of the protection system to transform heterogeneous observations into timely and reliable action. This point is important because conceptual designs often assume that once an event is detected, the remaining information chain will function with negligible delay or uncertainty. The present results show that this assumption is methodologically weak, since validation, fusion, decision, communication, and dispatch may consume a substantial share of the available interruption margin.

A central implication of the study is that conceptual design should explicitly connect information performance to the broader detection-delay-response logic. Delay resources create practical protective value only if information is processed quickly enough to support timely authorization and mobilization. In this sense, the reference model contributes more than descriptive clarity: it provides a structured basis for identifying where information bottlenecks arise and how they affect protective sufficiency. The model therefore strengthens

conceptual review by making information latency and degradation visible at the same analytical level as physical barriers and response arrangements.

The results also support the view that information requirements should be differentiated by scenario rather than stated as universal design aspirations. Different scenario classes stress different parts of the information process and therefore produce different requirement priorities. Covert intrusion and ambiguity-dominated situations place greater emphasis on rapid validation and correct classification, whereas communications-degraded scenarios make transmission delay, availability, and message loss more critical. Similarly, false alarm surges and concurrent events highlight the importance of workload tolerance and sustained validation performance under operational load. The scenario-driven requirement derivation method is therefore valuable because it converts these differences into measurable and reviewable requirement profiles rather than leaving them at the level of general design judgment.

Another important implication is that requirement credibility depends on traceability to evidence. Conceptual requirements are stronger when they are not limited to threshold statements, but are explicitly linked to a scenario rationale, a measurable metric, and a feasible acceptance pathway. In this respect, the proposed traceability structure addresses a common weakness of early-stage design, namely the tendency to formulate information-related expectations without specifying how they will later be tested, monitored, or revalidated. By linking requirement categories to instrumented drills, synchronized logs, stress testing, or simulation-supported assessment, the framework improves the defensibility of conceptual design claims and supports later lifecycle governance.

From a practical design perspective, the study highlights several areas that deserve more deliberate treatment in conceptual review. Alarm validation logic, operator decision support, communications robustness, and dispatch coordination should be regarded as measurable engineering objects rather than as procedural background assumptions. This does not mean that all information variables must be estimated with high numerical precision at the conceptual stage. Rather, it means that information performance should be framed in a disciplined way, with explicit targets, justified margins, and a clear pathway for later refinement when richer empirical evidence becomes available. Such an approach is especially useful in environments

where operating conditions, adversary behavior, or communication reliability may change over time.

At the same time, the proposed framework should not be interpreted as a substitute for facility-specific calibration or empirical validation. Its contribution is methodological: it offers a structured way to represent the information-to-action chain, derive scenario-sensitive requirement sets, and connect those requirements to evidence. The framework is therefore most useful as a conceptual design instrument that improves transparency, traceability, and review discipline before detailed site-specific modeling is performed.

Overall, the discussion confirms the main argument of the paper: conceptual physical protection design is strengthened when information processes are modeled explicitly, translated into measurable requirements, and tied to evidence-based acceptance logic. By separating the descriptive role of the reference model from the procedural role of the requirement derivation method, the study provides a clearer basis for integrating information performance into conceptual protection design and for defending those design choices during expert review, testing, and subsequent revalidation.

The proposed framework should also be interpreted within its scope. Its contribution is methodological rather than site-specific, and the requirement structures introduced here still require calibration against facility conditions, operating procedures, and validated empirical evidence. In particular, probabilistic formulations depend on the availability of sufficiently reliable logs, controlled trials, or simulation-supported datasets. These limitations do not reduce the conceptual value of the framework, but indicate the need for future work on site-level instantiation and empirical validation.

6. Conclusion

This study has shown that information processes should be treated as explicit design objects in the conceptual design of physical protection systems. The proposed framework demonstrates that protective sufficiency depends not only on sensing, barriers, and response resources, but also on the timely and reliable transformation of observations into authorized action. By representing the information-to-action chain as a structured process, the study makes visible the operational delays and degradation mechanisms that may otherwise remain implicit in hardware-centered design descriptions.

The main contribution of the study is threefold. First, a reference information-process model was developed to describe the forward path from event generation and detection to response activation. Second, a scenario-driven method was proposed for deriving measurable information requirements from dominant information bottlenecks. Third, a traceability structure was introduced to connect scenario classes, requirement priorities, and preferred acceptance evidence. Taken together, these elements provide a more transparent and reviewable basis for conceptual design decisions.

The results indicate that information requirements should not be formulated as generic aspirations, but as scenario-sensitive and measurable design statements. Depending on the operational character of the scenario, priority may shift toward rapid validation, correct classification, communications robustness, workload tolerance, or bounded end-to-end information latency. This makes conceptual design more defensible because requirement targets are linked not only to abstract performance expectations, but also to identifiable operational stressors and feasible evidence pathways.

At the same time, the framework should be interpreted within its intended scope. Its contribution is methodological rather than site-specific, and the proposed requirement structures still require calibration against facility conditions, operating procedures, and validated empirical evidence. In particular, probabilistic formulations depend on the availability of reliable logs, controlled trials, or simulation-supported datasets. Future research should therefore focus on site-level instantiation,

empirical validation, and comparative assessment across different classes of protected facilities.

Overall, the study supports the conclusion that conceptual physical protection design is strengthened when information processes are modeled explicitly, translated into measurable requirements, and linked to evidence-based acceptance logic. This approach improves transparency, supports expert review, and creates a more rigorous foundation for later testing, commissioning, and lifecycle revalidation.

Conflict of Interest

The authors declares that there is no conflict of interest, particularly of a financial, personal, authorial, or any other nature, that could influence the research or the results published in this article.

Funding

The study was conducted without financial support.

Data Availability

The manuscript has no associated data.

Use of Artificial Intelligence

The authors confirms that they did not use artificial intelligence technologies to write this work.

References

1. Akhundov, R., Hashimov, E. (2025), Enhancing the efficiency of the military environmental security system through the implementation of advanced technical means", *Proceedings of International scientific and practical conference: Modeling, Control and Information Technologies*, No. 8, pp. 348–352. DOI: <https://doi.org/10.31713/MCIT.2025.108>
 2. Akhundov, R., Hashimov, E. G., Islamov, I. (2026), "Methodological limitations of normative design of physical protection systems for critical and military facilities in a dynamic threat environment", *International scientific journal "Grail of Science"*, No.62, pp. 873–889. DOI: <https://doi.org/10.36074/grail-of-science.20.02.2026.096>
 3. Cozens, P., Love, T. (2015), "A Review and Current Status of Crime Prevention through Environmental Design (CPTED)", *Journal of Planning Literature*, No. 30(4), pp. 393–412. DOI: <https://doi.org/10.1177/0885412215595440>
 4. El Wely, I. C., Chetaine. A. (2020), "Analysis of physical protection system effectiveness of nuclear power plants based on performance approach", *Annals of Nuclear Energy*, No. 153, pp. 108051. DOI: <https://doi.org/10.1016/j.anucene.2020.107980>
 5. Garcia, M. L. (2008), *Design and Evaluation of Physical Protection Systems*. 2nd ed. Elsevier. DOI: <https://doi.org/10.1016/C2009-0-25612-1>
 6. Genserik, L. L. Reniers, A. A. (2014), "Preparing for major terrorist attacks against chemical clusters: Intelligently planning protection measures w.r.t. domino effects", *Process Safety and Environmental Protection*, No. 92(6), pp. 583–589. DOI: <https://doi.org/10.1016/j.psep.2013.04.002>
-

8. Akhundov, R., Hashimov, E. G. (2025), Quantitative categorization of facilities and modeling of potential adversaries", *Grail of Science*, No. 60, pp. 469–482.
9. Hashimov, E. et al. (2026), "Research of the efficiency multiservice networks using MIMO technology", *Advanced Information Systems*, No. 10(1), pp. 66–71. DOI: <https://doi.org/10.20998/2522-9052.2026.1.08>
10. Hashimov, E., Akhundov, R. G., Talibov, A. M., Islamov, I. (2026), "Constrained optimization of an integral security indicator for adaptive management of hazardous facilities", *Grail of Science*, No. (62), pp. 1003–1014.
11. Islamov, I. et al. (2025), "Big data analytics and machine learning for predicting radiation and chemical threats in the military sphere", *Theory and practice of modern science: Collection of scientific papers "SCIENTIA" with proceedings of the X International Scientific and Theoretical Conference*, pp. 30–38. DOI: <https://doi.org/10.36074/scientia-26.09.2025>
12. Akhundov, R., Islamov, I. (2025), "Military Environmental Security under Radiation and Chemical Threats", *Control and Information Technologies: Proceedings of International scientific and practical conference*, No. 8, pp. 414–419. DOI: <https://doi.org/10.31713/MCIT.2025.129>
13. Islamov, I. et al. (2025), "Controller-level scalability problems in software-defined networks", *Proceedings of the 13th International Scientific and Technical Conference*, Vol. 1, pp. 70–71. DOI: <https://doi.org/10.13140/RG.2.2.31197.88801>
14. Islamov, I. et al. (2025), "Hybrid communication models for UAV swarms: Towards scalable and energy-aware network optimization", *Scientific guidelines: Theory and practice of research – Proceedings of the VI International Scientific Conference*, pp. 185–195. <https://doi.org/10.62731/mcnd-03.10.2025>
15. Islamov, I. et al. (2025), "Innovative approaches to environmental recovery in conflict-affected areas", *Proceedings of the VII International Scientific Conference*, pp. 180–190. DOI: <https://doi.org/10.62731/mcnd-24.10.2025>
16. Islamov, I. et al. (2025), "The use of unmanned systems and artificial intelligence to enhance radiation and chemical safety in military ecology", *Proceedings of the VII International Scientific Conference*, pp. 183–192. DOI: <https://doi.org/10.62731/mcnd-10.10.2025>
17. Kampova, K., Lovecek, T., Řehák, D. (2020), "Quantitative approach to physical protection systems assessment of critical infrastructure elements: Use case in the Slovak Republic", *International Journal of Critical Infrastructure Protection*, No. 30, pp. 100376. DOI: <https://doi.org/10.1016/j.ijcip.2020.100376>
18. Akhundov, R., Hashimov, E. G., Islamov, I. (2026), "Scenario oriented sufficiency criteria for physical protection systems provide a traceable path from threat classes to design requirements", *Grail of Science*, No. 63. DOI: <https://doi.org/10.36074/grail-of-science.06.03.2026.074>
19. Kaplan, S., Garrick, B. J. (1981), "On the quantitative definition of risk", *Risk Analysis*, Vol. 1, No. 1, pp. 11–27. DOI: <https://doi.org/10.1111/j.1539-6924.1981.tb01350.x>
20. Lovecek, T., Ristvej, J., Simak, L. (2010), "Critical Infrastructure Protection Systems Effectiveness Evaluation", *Journal of Homeland Security and Emergency Management*, Vol. 7, No. 1. DOI: <https://doi.org/10.2202/1547-7355.1613>
21. Mondal, S., Adak, B., Mukhopadhyay, S. (2023), "Functional and smart textiles for military and defence applications", *Smart and functional textiles*, pp. 397–468. DOI: <https://doi.org/10.1515/9783110759747-011>
22. Řehák, D., Senovsky, P., Hromada, M., Lovecek, T. (2019), "Complex approach to assessing resilience of critical infrastructure elements", *International Journal of Critical Infrastructure Protection*, No. 25, pp. 125–138. DOI: <https://doi.org/10.1016/j.ijcip.2019.03.003>
23. Rehak, D., Slivkova, S., Janeckova, H., Stuberova, D., Hromada, M. (2022), "Strengthening Resilience in the Energy Critical Infrastructure: Methodological Overview", *Energies*, Vol. 15, No. 14, pp. 5276. DOI: <https://doi.org/10.3390/en15145276>
24. Shoop, B., et al. (2006), "Mobile detection assessment and response systems (MDARS): A force protection physical security operational success", *Unmanned Systems Technology*, Vol. 6230, pp. 668–678. DOI: <https://doi.org/10.1117/12.665939>
25. Talibov, A. M., Hashimov, E. G., Akhundov, R. G. (2025), "Modeling and forecasting radiological and chemical threats in the military sphere", *Proceedings of the 15th International Scientific and Technical Conference*, Vol. 1, pp. 120–121.
26. Yang, J., Huang, L., Ma, H., Xu, Z., Yang, M., Guo, S. (2022), "A 2D-graph model-based heuristic approach to visual backtracking security vulnerabilities in physical protection systems", *International Journal of Critical Infrastructure Protection*, No 38, pp. 100554. DOI: <https://doi.org/10.1016/j.ijcip.2022.100554>
27. Zou, B., Yang, M., Zhang, Y., Benjamin, E.-R., Tan, K., Wu, W., Yoshikawa, H. (2018), "Evaluation of vulnerable path: Using heuristic path-finding algorithm in physical protection system of nuclear power plant", *International Journal of Critical Infrastructure Protection*, No. 23, pp. 90–99. DOI: <https://doi.org/10.1016/j.ijcip.2018.08.006>

Received (Надійшла) 08.02.2026

Accepted for publication (Прийнята до друку) 01.03.2026

Publication date (Дата публікації) 12.03.2026

About the Author / Відомості про автора

Ахундов Раміль Гурбаналі – доктор філософії з національної безпеки та військових наук, професор, Національний університет оборони; Баку, Азербайджан;

Ramil Akhundov – PhD in National Security and Military Sciences, Professor, National Defense University; Baku, Azerbaijan;
e-mail: mr.axundov1@gmail.com

ORCID ID: <https://orcid.org/0009-0001-8798-8044>

Гашимов Ельшан Гіяс – доктор національної безпеки та військових наук, професор, Азербайджанський технічний університет; професор, Національний університет оборони; Баку, Азербайджан;

Elshan Hashimov – Doctor in National Security and Military Sciences, Professor, Azerbaijan Technical University; Professor, National Defense University; Baku, Azerbaijan;

e-mail: hasimovel@gmail.com

ORCID ID: <http://orcid.org/0000-0001-8783-1277>

МОДЕЛЮВАННЯ ІНФОРМАЦІЙНИХ ПРОЦЕСІВ ТА ВИЗНАЧЕННЯ ВИМІРЮВАНИХ ВИМОГ ПРИ ПРОЕКТУВАННІ СИСТЕМ ФІЗИЧНОГО ЗАХИСТУ

У дослідженні розглядаються інформаційні процеси на етапі концептуального проектування систем фізичного захисту та їхня роль у перетворенні даних, отриманих від датчиків, вхідних даних оператора та потоків комунікації на своєчасні захисні дії. Особлива увага приділяється тому, що ефективність проекту обмежується не лише датчиками, бар'єрами та силами реагування, а й затримками та невизначеністю, що виникають між виявленням, прийняттям рішення та відправленням сил реагування. **Метою** дослідження є формалізація інформаційних процесів як явних об'єктів проектування та розробка підходу до визначення вимірюваних інформаційних вимог, які можна обґрунтувати та перевірити на етапі концептуального проектування. **Завдання** полягають у визначенні еталонного робочого процесу від генерації події до активації реагування, розбитті загального часу від інформації до дії на оперативні етапи, врахуванні невизначеності, пов'язаної з помилковими тривогами, неоднозначністю, навантаженням та погіршенням зв'язком, а також ув'язуванні формулювань вимог із доказами прийнятності. У дослідженні використовується **концептуальний та методологічний підхід**, заснований на системному аналізі, розбитті сценаріїв, структуруванні затримок та відображенні простежуваності між сценаріями загроз, вузькими місцями, цільовими вимогами та доказами перевірки. У дослідженні **розроблено** структуровану модель «від інформації до дії», яка чітко визначає етапи зондування, валідації, злиття, прийняття рішень, комунікації та диспетчеризації. На цій основі запропоновано **метод** перетворення вузьких місць, характерних для конкретних сценаріїв, на перевірявані вимоги щодо своєчасності, точності, повноти та стійкості. Дослідження також визначає практичні форми доказів прийнятності, включаючи навчання з обмеженим часом, вимірювання на основі журналів, стрес-тестування та оцінку з використанням моделювання. **Результати** показують, що концептуальне проектування стає більш обґрунтованим, коли інформаційні процеси моделюються явно, а не розглядаються як неявні припущення. Запропонований підхід дозволяє проектувальникам обґрунтувати вимірювані вимоги, виявити критичні джерела затримки та підтримати повторну перевірку достатності проекту в умовах мінливих експлуатаційних умов.

Ключові слова: система фізичного захисту; концептуальне проектування; інформаційні процеси; перевірка сигналів тривоги; об'єднання даних; управління та контроль; стійкість комунікацій; докази прийнятності.

Бібліографічні описи / Bibliographic descriptions

Ахундов Р. Г., Гашимов Е. Г. Моделювання інформаційних процесів та визначення вимірюваних вимог при проектуванні систем фізичного захисту. *Автоматизовані системи управління та прилади автоматики*. 2026. № 1 (188). С. 5–16. DOI: <https://doi.org/10.30837/0135-1710.2026.188.005>

Akhundov, R., Hashimov, E. (2026), "Modeling Information Processes and Deriving Measurable Requirements in Physical Protection System Design", *Management Information System and Devices*, No. 1 (188), P. 5–16. DOI: <https://doi.org/10.30837/0135-1710.2026.188.005>

Igor Grebennik, Oleksii Kovalenko

USE OF ARTIFICIAL NEURAL NETWORKS FOR MANAGING THE SORTING OF PARCELS IN A CONVEYOR FLOW

The subject matter of the article is the decision-making models of automated sorting lines (ASLs) installed at parcel sorting centers (PSCs). The goal of the work is to develop a decision-making model for the ASLs serving the loading doors of PSC terminals in the form of an ANN that classifies parcels according to their weight and dimensions. The following tasks were solved in the article: an analysis of the design modifications of ASL equipment for implementing three variants of truck loading logic was conducted; the weight and dimensional ranges of parcels associated with loading chute and loading door numbers were determined; the criteria for implementing the specified parcel sorting logic; the parameters of training and testing datasets for ANN models were determined; using MATLAB tools, training and testing of four types of ANNs were carried out: RBFNN, GRNN, PNN, FFNN; a comparative analysis of the ANNs was conducted using the relative precision of correct parcel classification by weight and dimensional ranges and the number of neurons employed as quality metrics. Methods used: systems analysis, analytical methods, computer simulation, ANN architecture training methods, and mathematical and statistical analysis of training effectiveness. Results achieved. Comparative training and testing of four types of ANNs was performed. Based on the test results, the FFNN model trained using the Levenberg–Marquardt method was selected for implementation, as it provides a relative precision (RP) of correct classification 97%. Conclusions. The developed ASL decision-making model in the form of an FFNN enables the classification of parcels within a conveyor flow into three ranges of weight and dimensional parameters, thereby implementing the specified sorting logic across loading doors or chutes. The implemented sorting logic ensures compact loading of truck cargo compartments and reduces the risk of parcel damage when parcels are stacked on top of one another.

Keywords: delivery logistics; decision-making model; classification; sorting line; artificial neural networks.

1. Introduction

Statement of the problem

Industrial automation of object sorting is one of the key areas of development in modern logistics associated with the provision of postal services in the parcel delivery industry (PDI) [1] and cross-docking services [2] for delivering goods from manufacturers to the warehouses of e-commerce retailers. To support the PDI, logistics companies employ a network of intermediate and final parcel sorting centers (PSCs), central parcel consolidation terminals (CPCTs), and parcel hubs (PHs). To ensure the delivery of goods from manufacturers to consumers, a network of intermediate and final cross-docking distribution centers is used. The aforementioned transshipment nodes of delivery logistics networks are designed for unloading inbound trucks, sorting freight, and loading outbound trucks for further delivery to the addresses of subsequent network nodes.

Automated sorting lines (ASLs) are used for object sorting and serve as structural components for constructing automated sorting conveyors (ASCs). Such ASCs are installed at PSCs and cross-docking distribution centres.

A major issue in applying ASLs within the PDI is the limitation of their decision-making models, which arises from the uncertainty of parcel sorting conditions based on categories defined by weight and dimensional parameters. This problem is caused by the limited functional capabilities of ASLs, which sort parcels solely by delivery address without considering their weight and dimensions. This may lead to inefficient utilisation of truck cargo space during loading, as well as the risk of parcel damage when parcels are stacked on top of one another. This problem necessitates research into the development of an ASL decision-making model that enables parcel sorting management based on weight and dimensional classification to implement the specified outbound truck loading logic.

This article is devoted to the development of an ASL decision-making model in the form of a neural network that implements the specified classification of parcels by their weight and dimensions. This choice of model implementation is motivated by recent advances in the use of neural networks, which have led to significant progress in solving complex problems associated with virtually every field of human activity [3–5]. The objective of this study is to compare neural networks and machine learning algorithms for implementing an ASL decision-making model that classifies and sorts

parcels according to specified ranges of their weight and dimensional parameters.

2. Analysis of recent studies and publications

For parcel sorting, PDI transshipment centres are equipped with ASCs that incorporate ASLs. In general, the decision-making model of an ASC at a PSC is implemented as a computer system that monitors parcel unloading operations and their sorting based on barcode label information, followed by loading. An analysis of publications related to the improvement of postal service logistics within the PDI is associated with the overarching task of developing decision-making models for ASLs or ASCs to ensure the continuous (round-the-clock) and uninterrupted operation of intermediate PSCs. The problems addressed can be divided into two research directions.

The first research direction addresses the scheduling of freight transport across transshipment points of the logistics network, including arrival and departure times. In [6], a genetic algorithm was developed to minimise idle time by monitoring the parcel unloading rate to manage the dispatch schedule of outbound trucks, regardless of the loading volume of their cargo compartments. In [1, 7], a PSC model and genetic algorithms were developed to plan the arrival times of inbound trucks in order to minimise unloading time. For planning purposes, the genetic algorithm in [1] uses information from the PSC model regarding the number of available unloading doors, while the genetic algorithm in [7] uses information on the cargo volumes of inbound trucks. In [8], an adaptive genetic algorithm was developed that uses parcel label information to determine an unloading plan for inbound trucks with unequal cargo volumes. The algorithm also enables the determination of a shortened parcel transport route along the conveyor, thereby increasing the throughput of the CPCT. The study in [8] was continued in [9], where a CPCT simulation model and a hybrid genetic algorithm were developed to determine an outbound truck loading plan, taking into account the parcel transport time along shortened routes. In [10], a combinatorial algorithm was developed for planning train interval schedules for their allocation to railway tracks in order to minimise idle time at railway transshipment points.

The second direction is associated with research aimed at increasing PSC throughput, taking into account the configuration of ASC structural elements. In [11],

a genetic algorithm was developed to plan the optimal structure of ASC components that ensures maximum throughput. A CPCT simulation model and parcel label information are used to verify the algorithm's performance. In [12], a mixed-integer programming model for PSC was developed that provides two-stage parcel sorting. A reduction in sorting time is achieved by minimising the number of ASC equipment settings that affect the redistribution of parcels from loading to unloading doors. In [13], two models were developed to address the problem of increasing conveyor throughput at a distribution centre while considering its operational constraints. The first model reduces sorting time by determining the minimum object transport route along the conveyor for a specific unloading–loading scenario. The second model implements a conveyor operating mode with maximum throughput and minimises all object transport routes for each unloading and loading operation. In [14], a deterministic mixed-integer linear model was developed for evaluating PSC operating scenarios. The modelling results were used to modify the conveyor configuration in order to reduce the time required for manual parcel transport. In [15], a model was developed for evaluating PSC conveyor throughput, enabling the assessment of various ASC layout options while considering parcel sorting time as a function of the number of loading and unloading doors.

To ensure the delivery of goods from manufacturers to consumers, a network of intermediate and final cross-docking distribution centers is used. Such distribution centers are designed for receiving trucks at inbound docks, unloading, sorting, and temporary storage of goods, followed by loading trucks at outbound docks for dispatch to destinations (the intermediate storage time for goods is limited to 24 hours). Goods are transported in standard pallets. ASCs (or ASLs) are used when it is necessary to unpack goods from delivered pallets for storage or to consolidate goods onto pallets for dispatch to a specified address. Research on cross-docking can be divided into several directions [2, 16–20]: studies on planning the location of distribution network nodes and goods delivery routing [2, 16]; studies on optimising the design and layout of cross-docking centre equipment [17, 18]; and studies on scheduling inbound and outbound trucks allocated to docks under conditions of limited dock availability and warehouse capacity [19, 20].

As a separate direction, one can identify studies [21–23] on the development of algorithms and decision-making models for ASLs that address the task

of sorting objects with respect to their weight and dimensions. In [21], a nonlinear model for loading transport containers (sub-containers, pallets) was developed that solves the problem of object consolidation planning while considering their weight and geometric parameters (object orientation, dimensional characteristics of object placement within sub-containers). In [22], a fuzzy decision-making model for ASLs was developed, and in [23], an adaptive neuro-fuzzy inference system was developed, both enabling parcel sorting by weight and dimensions within a conveyor flow and implementing the specified loading logic for outbound trucks at the PSC.

This article is a continuation of the studies [22, 23].

3. Goal and objectives of the study

The goal of the study is to develop a decision-making model for the ASLs serving the loading doors of PSC terminals in the form of a neural network that performs flow-based classification of parcels into three ranges of weight and dimensions for implementing the specified sorting logic across loading doors and chutes.

To achieve the stated goals, the following tasks must be solved:

- a) determine the weight and dimensional ranges of parcels associated with loading chute numbers;
- b) determine the parcel sorting logic by weight and dimensions;
- c) determine the parameters of datasets for training and testing neural networks;
- d) conduct training and testing to determine the architecture of the following ANNs:
 - Radial Basis Function Network (RBFNN);
 - Generalized Regression Neural Network (GRNN);
 - Probabilistic Neural Network (PNN);
 - Feed-Forward Neural Network (FFNN);
- e) conduct a comparative analysis of the neural networks using the number of neurons employed, the mean squared error (MSE) of training, and the relative precision of correct parcel classification by the weight and dimensional ranges defined by loading chute (loading door) numbers as quality metrics.

4. Materials and methods

Two PSC conveyor modification schemes [22, 23] are used in this study, enabling three variants of the specified parcel sorting logic for loading outbound trucks.

A simplified PSC conveyor scheme [22] implementing the first and second variants of outbound truck loading logic is presented in Fig. 1. Parcels from inbound trucks are unloaded at the PSC unloading doors designated as "U1", "U2", ..., "UN". By means of sorting lines ASL-A and ASL-B of the first conveyor ASC-1, the primary sorting of parcels is performed across two terminals "A" or "B", each assigned to two subsequent delivery directions. For this purpose, the decision-making model of ASL-A (ASL-B) uses the delivery address information from the parcel label. To facilitate sorting at terminals "A" and "B", sorting lines ASL-1, ASL-2, ..., ASL-N of conveyor ASC-A (ASC-B) are used, which are responsible for feeding parcels into the loading chutes "LC" of loading doors designated as "L1", "L2", ..., "LN". Flow-based loading of parcels into outbound trucks is performed from chutes "LC" by manually stacking them in layers on top of one another, which entails a risk of parcel damage when they are placed on top of each other and results in inefficient utilisation of the vehicle cargo space.

To implement the specified loading logic according to the first and second variants, the decision-making models of sorting lines ASL-1, ASL-2, ..., ASL-N of conveyors ASC-A and ASC-B must ensure parcel classification across N weight and dimensional ranges defined for each of the N loading doors. Figure 1 shows a variant for classifying parcels across 3 ranges. Such sorting provides two variants of loading logic:

– the first variant of loading logic is defined for a single outbound truck at terminal "A". For each loading door, ASL-1,2,3 selects parcels within the specified range. First, parcels whose parameters correspond to the maximum weight and dimensional range values (door "L3") are loaded into the truck. Then, the truck is sequentially loaded at doors "L2" and "L1", corresponding to decreasing weight and dimensional ranges. This ensures the possibility of compact truck cargo loading and reduces the risk of parcel damage when parcels are placed on top of one another;

– the second variant of loading logic is shown in Fig. 1 for several outbound trucks at terminal "B". In this variant, the simultaneous loading of 3 outbound trucks at doors "L1", "L2", ..., "LN" with parcels of a single weight and dimensional parameter range is considered.

To implement the specified loading logic according to the third variant (Fig. 2), all loading doors "L" of terminals "A" and "B" are equipped with not one but several loading chutes "LC". Figure 2 shows a variant

with three chutes: "LC1", "LC2", "LC3". These chutes are designed to receive parcels in 3 weight and dimensional ranges. The range values increase in accordance with the order of loading chute "LC" numbers. In this case, the ASL decision-making

model for loading doors does not differ from the first and second sorting logic variants (Fig. 1) and must ensure parcel classification across 3 weight and dimensional ranges defined for each of the 3 loading chutes "LC" (Fig. 2).

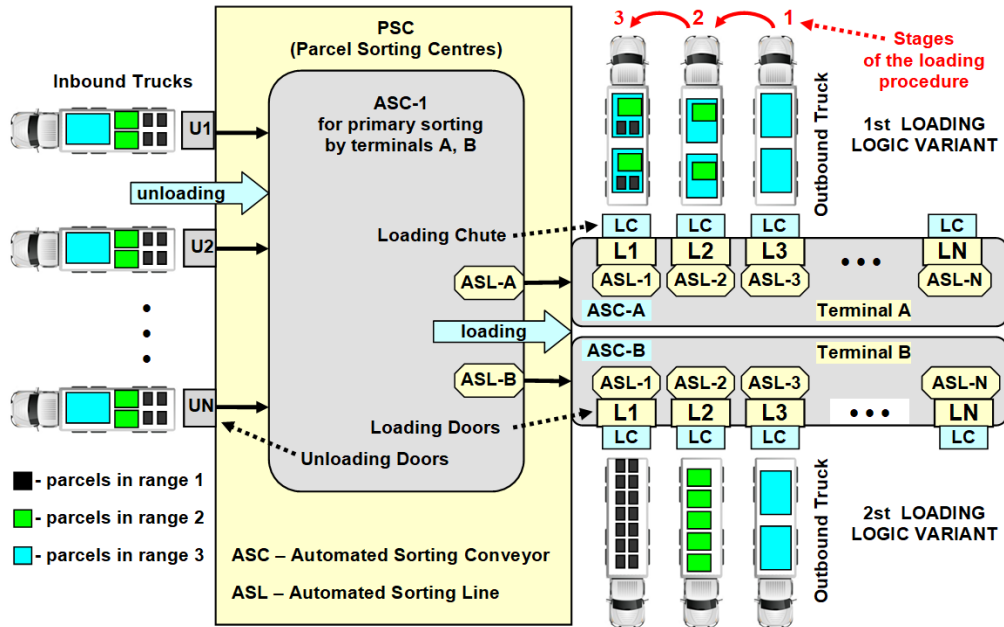


Fig. 1. First and second variants of truck loading logic

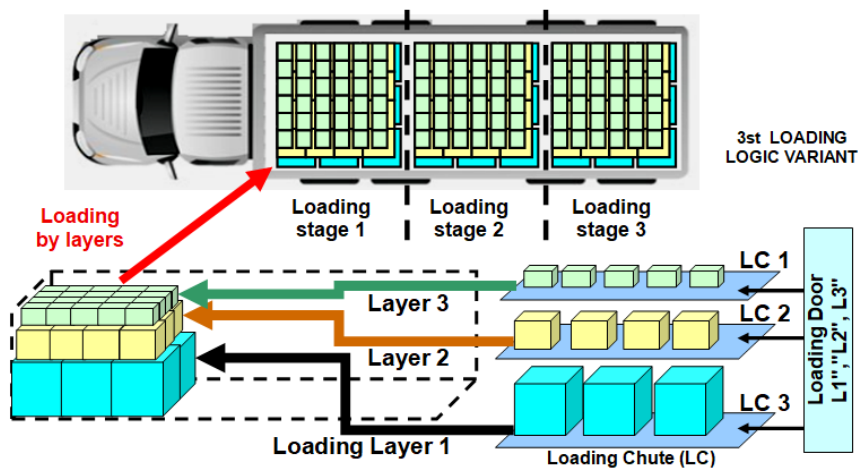


Fig. 2. Third variant of truck loading logic

Sorting across three loading chutes provides the third variant of parcel loading. Loading is performed in three or more stages depending on the volume of the vehicle cargo space. At each loading stage, parcels are arranged in three layers. The first layer consists of parcels with the greatest weight and dimensions from loading chute "LC3". The second layer consists of parcels in the medium range from chute "LC2", and the third layer consists of the lightest and smallest parcels from chute

"LC1". In the second and subsequent loading stages, the process of arranging parcels by layers is repeated. This ensures the possibility of compact loading of the vehicle cargo space and reduces the risk of parcel damage when they are placed on top of one another.

Thus, in accordance with the goal and objectives of the study, it is necessary to develop a decision-making model for the ASLs serving the loading doors of PSC terminals in the form of a neural network that performs

flow-based classification of parcels into three ranges of weight and dimensions for implementing the specified sorting logic across 3 loading doors or chutes.

The MATLAB environment toolkit (version R2016a) was employed for the development and investigation of neural network models using the computer simulation method. Enhanced training effectiveness of the neural networks under study was achieved through the use of a normalisation procedure for the input data of training and validation samples in accordance with the requirements of [24].

For the neural networks under investigation, the input data were specified as a set of column vectors of dimension $4 \times N$: $P_N = (p_{1n}, p_{2n}, p_{3n}, p_{4n})^T$, $n = 1, 2, \dots, N$, where p_i denotes the parcel parameters: p_1 – weight, p_2 – height, p_3 – width, p_4 – depth. The target data (classification output) were defined as a scalar value representing the loading chute (door) number $Y_{Tn} \in \{1, 2, 3\}$, assigned to each vector P from the set P_N . The target data for the validation datasets Y_{Vn} were specified in an analogous manner. According to the problem formulation, the neural networks under investigation were required to perform classification of parcels into three loading chutes "LC" (or doors "L"). Each loading chute (door) number is associated with three value ranges for each of the four parcel parameters p_i . The assignment of one of the three ranges of each parameter of vector P to the loading chute or door numbers $Y_T(Y_V)$ is determined by the specified parcel sorting logic.

For the comparative analysis of the trained neural networks, the following quality assessment metrics were employed: the number of neurons utilized, the mean squared error of training (MSE), and the relative precision (RP) of correct classification of parcels into one of the three loading chute (door) numbers PSC. The RP was calculated during both the training (RP_T) and validation (RP_V) phases of the neural networks. For this purpose, the output values Y_{On} of the trained neural networks were rounded to integers in accordance with standard mathematical rules. Subsequently, the number of true classification (TC) results was recorded during training $TC_T = count\{round(Y_{On}) = Y_{Tn}\}$ and validation $TC_V = count\{round(Y_{On}) = Y_{Vn}\}$. Analogously,

the false classification (FC) results were recorded during training $FC_T = count\{round(Y_{On}) \neq Y_{Tn}\}$ and validation $FC_V = count\{round(Y_{On}) \neq Y_{Vn}\}$, where the $count()$ function was used to calculate the number of satisfied conditions, and the $round()$ function was used to round a number to the nearest integer. The relative precision of correct classification was computed according to the following expression:

$$RP_{T(V)} = \frac{TC_{T(V)}}{TC_{T(V)} + FC_{T(V)}} = \frac{TC_{T(V)}}{N}, \quad (1)$$

where N is the number of input vectors in the training or validation datasets P_N .

5. Research Results and Discussion

5.1. Determination of Parcel Weight and Dimensional Ranges

According to the problem formulation, the neural networks under investigation were required to perform classification of parcels into three loading chutes "LC" (or doors "L"). Each loading chute (door) number is associated with three value ranges for each of the four parcel parameters p_i . The following parcel parameters are defined as neural network input data: p_1 – weight, p_2 – height, p_3 – width, and p_4 – depth. For each input parameter p_i , $i \in \{1, 2, 3, 4\}$, 4 value ranges are defined corresponding to 3 loading chutes "LC" (doors "L"). The parameter value ranges are designated as R_{1q} , R_{2j} , R_{3m} , R_{4s} , where the first subscript is the parameter index p_i , and the second subscript $(q, j, m, s) \in \{1, 2, 3\}$ defines the range index, which coincides with the loading chute "LC" (door "L") number. The values of ranges $R_{i(q,j,m,s)}$ are given in Table 1.

In Table 1, the dimensional parameter ranges of parcels are chosen to be identical. This is done intentionally for the case when the QR code of a parcel label does not contain data on its dimensions. In this case, the ASC determines these parcel parameters from the results of processing their 3D images. However, during unloading, a parcel may be placed on the conveyor belt platform in an arbitrary orientation. Therefore, it is impossible to determine from the processing results which parameters $\{p_2, p_3, p_4\}$ the measured dimensional values correspond to.

Table 1. Parcel parameter value ranges

Input parameters	Range designation	Range of parcel parameter values for loading chutes (doors)		
		"LC1" ("L1")	"LC2" ("L3")	"LC3" ("L3")
p_1 , weight (kg)	R_{1q}	$R_{11} - [0, 14.9]$	$R_{12} - [15, 44.9]$	$R_{13} - [45, 60]$
p_2 , height (cm)	R_{2j}	$R_{21} - [0, 39.9]$	$R_{22} - [40, 119.9]$	$R_{23} - [120, 160]$
p_3 , width (cm)	R_{3m}	$R_{31} - [0, 39.9]$	$R_{32} - [40, 119.9]$	$R_{33} - [120, 160]$
p_4 , depth (cm)	R_{4s}	$R_{41} - [0, 39.9]$	$R_{42} - [40, 119.9]$	$R_{43} - [120, 160]$

5.2. Determination of parcel sorting logic

The parcel sorting logic is defined in accordance with the rules

$$\text{RULE}_{k=1}^K: IF(p_1 \in R_{1q}) \text{ and } IF(p_2 \in R_{2j}) \text{ and } IF(p_3 \in R_{3m}) \text{ and } IF(p_4 \in R_{4s}) \rightarrow L_k, \quad (2)$$

where $k = 1, 2, \dots, K$ are rule indices; $R_{i(q,j,m,s)}$ are the value ranges of parcel parameters p_i corresponding to three loading chutes "LC" (doors "L") with range numbers defined by the indices $(q, j, m, s) \in \{1, 2, 3\}$ (Table 1); $L_k = \{L_1, L_2, L_3\} = \{1, 2, 3\}$ is the loading chute "LC" (door "L") number.

The sorting logic was defined for each k -rule (2). For this purpose, four true conditions of belonging

of a parcel with parameters $\{p_1, p_2, p_3, p_4\}$ to the corresponding ranges with indices $(q, j, m, s) \in \{1, 2, 3\}$ were considered, and the loading chute number L_k was assigned. The number of rules in (2) is determined by the expression: $K = Z^G = 4^3 = 81$, where Z is the number of conditions in the rule, and G is the number of loading chutes. The sorting logic matrix is presented in Table 2 (in abridged form).

Table 2. Parcel sorting logic matrix (abridged)

Rule index k	Indexes of fulfilled conditions in rules (2)				Number "LC" ("L") for the k -th rule L_k
	$p_1 \in R_{1q}$	$p_2 \in R_{2j}$	$p_3 \in R_{3m}$	$p_4 \in R_{4s}$	
	index q	index j	index m	index s	
1	1	1	1	1	1
2	1	1	1	2	
3	1	1	1	3	
4	1	1	2	1	
5	1	1	3	1	
6	1	2	1	1	
7	1	3	1	1	
8	1	1	2	2	2
9	1	1	2	3	
10	1	1	3	2	
11	1	1	3	3	
12	1	2	1	2	
13	2	1	1	1	
...	
47	2	3	2	2	3
48	2	1	3	3	
49	2	2	3	3	
50	2	3	1	3	
51	2	3	2	3	
52	2	3	3	1	
53	3	1	1	1	
...	
81	3	3	3	3	

The sorting logic presented in Table 2 implements two criteria for verifying the truth of conditions for the belonging of parcel weight and dimensional values to the corresponding ranges associated with the three loading chutes (doors):

– first criterion: if the rule contains no more than one true condition with an index from $\{j, m, s\}$ whose value exceeds the index q , then the parcel is transported to the loading chute numbered q ;

– second criterion: if the rule contains two or more indices of true conditions from $\{j, m, s\}$ that exceed the value of index q , then the parcel is transported to the loading chute numbered $q+1$. If $(q+1) > 3$ then $q=3$.

5.3. Determination of dataset parameters for training and testing neural networks

For neural networks, the input data is a column vector P of dimension 4×1 , and the target data Y_T

Table 3. Parameters of datasets for training and testing ANNs

Purpose of data sets	Number of values for ranges $R_{i(q,j,m,s)}$				Total number of tuples in the set (N)	Data dimension		Name of data sets
	R_{1q}	R_{2j}	R_{3m}	R_{4s}		$P_N (4 \times N)$	$Y_T (1 \times N)$	
Training	3	3	3	3	34=81	4×81	1×81	dataset1
	4	4	4	4	44=256	4×256	1×256	dataset2
	5	5	5	5	54=625	4×625	1×625	dataset3
	6	6	6	6	64=1296	4×1296	1×1296	dataset4
	9	9	9	9	94=6561	4×6561	1×6561	dataset5
Validation	13	13	13	13	134=28561	$P_N (4 \times N)$	$Y_V (1 \times N)$	dataset6
						4×28561	1×28561	

5.4. Investigation of Radial Basis Function (RBF) neural network models

The architecture of the RBF neural network model in MATLAB is presented in Fig. 3. The number

is a scalar value representing the loading chute (door) number. For training the neural networks, 5 normalised datasets were used, which were formed as an input data array P of dimension $4 \times N$. The vectors $P = (p_1, p_2, p_3, p_4)^T$ of array P_N contain the parcel parameter values, which were determined in accordance with Table 1. The normalisation of weight data p_1 and dimensional data p_2, p_3, p_4 was performed separately. The output (target) data is represented as a row vector $Y_T = \{L_{k1}, L_{k2}, \dots, L_{kN}\}$ of loading chute (door) numbers, defined in accordance with Table 2 for each vector of array P_N . Validation of trained neural networks was performed using the "dataset6", which was formed in an analogous manner.

The parameters of the datasets for training and testing ANNs are presented in Table 3.

of RBF network inputs is determined by the dimension R of the training data P_N . For the datasets P_N used (Table 3), with dimension $R \times N$, the RBF network has 4 inputs ($R = 4$).

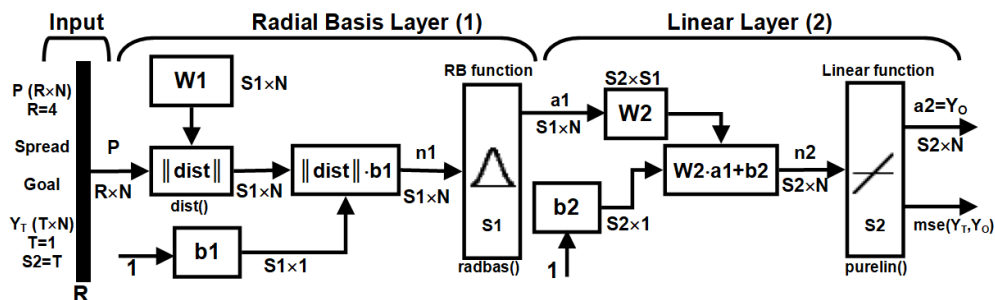


Fig. 3. Architecture of the RBF neural network model in MATLAB

The radial basis layer (Fig. 3) transforms the input data P using RB activation functions, centring their values with the weight vector $W1$ (net.IW{1,1}). The RB neurons compute the euclidean distance using the $dist()$ function, adding the bias $b1$ (net.b{1}):

$$n1 = dist((net.IW\{1,1\},p),net.b\{1\}) = \|W1 - P\| \cdot b1. \quad (3)$$

The resulting RB neuron activation function has the form:

$$a1 = \exp\left(-\left(\|W1 - P\| \cdot b1\right)^2 / spread\right) = radbas(n1), \quad (4)$$

where "Spread" is the smoothing parameter of the gaussian RB activation function that determines its width.

In the second layer, the output values $a1$ are linearly combined using weights $W2$ (net.LW{2,1}) and biases $b2$ (net.b{2}). The formed combinations $n2 = W2 \cdot a1 + b2$, after activation by the linear function, form the RBF network output: $Y_o = a2 = purelin(n2)$. The number of linear layer neurons $S2$ (the number of RBF network outputs) is determined by the dimension T

of the target dataset Y_T (for the problem under consideration, $S2 = T = 1$).

For training the RBF network, the input data P , target data Y_T , the minimum MSE value "Goal", and the "Spread" parameter are specified. The backpropagation method is used for training the RBF network. Training begins with one RB neuron ($S1 = 1$). After determining the output value Y_o , the RBF network calculates the mean squared error between two vectors:

$$MSE = mse(Y_T, Y_o) = \frac{1}{N} \sum_{n=1}^N (Y_{Tn} - Y_{On})^2. \quad (5)$$

MSE minimisation is performed for each subsequent training epoch of the RBF network by iteratively adjusting the values of vectors $W1$, $W2$ and increasing the number of first-layer neurons $S1$ by "1" until the condition $MSE < Goal$ is satisfied.

Training and validation of RBF models for parcel classification research were performed using the datasets given in Table 3. The best training and testing results of RBF models, obtained for datasets "dataset1" and "dataset4", are presented in Table 4.

Table 4. Results of RBF model training and validation

No	Training							Validation			
	Data Set	S1	Spread	Goal	MSE	MSE_{TR}	RP_T	Data Set	MSE_V	MSE_{VR}	RP_V
1	dataset1 (4×81)	74	0.1	0.01	0.0	0.0	1.0	dataset6 (4×28561)	2.0283	2.1186	0.0418
2		67	0.3	0.01	0.0099	0.0	1.0		0.0609	0.0523	0.9477
3		43	0.5	0.01	0.0096	0.0	1.0		0.0739	0.0715	0.9285
4		50	0.7	0.01	0.0098	0.0	1.0		0.0765	0.0806	0.9194
5		40	0.9	0.01	0.0094	0.0	1.0		0.0925	0.1086	0.8914
6		53	1.0	0.01	0.0087	0.0	1.0		0.0944	0.1193	0.8807
7		45	1.5	0.01	0.0097	0.0	1.0		0.3693	0.4860	0.5604
8		45	2.0	0.01	0.0098	0.0	1.0		0.4535	0.5846	0.4850
9		63	0.3	0.02	0.0179	0.0	1.0		0.0639	0.0586	0.9414
10		58	0.3	0.03	0.0295	0.0617	0.9383		0.0672	0.0643	0.9357
11		55	0.3	0.04	0.0383	0.0370	0.9630		0.0695	0.0686	0.9314
12		51	0.3	0.05	0.0461	0.0494	0.9506		0.0696	0.0692	0.9308
13	dataset4 (4×1296)	941	0.1	0.04	0.0400	0.0201	0.9799	dataset6 (4×28561)	0.0813	0.0295	0.9705
14		882	0.1	0.05	0.0495	0.0918	0.9082		0.0797	0.0640	0.9360
15		240	0.3	0.01	0.0096	0.0	1.0		0.0683	0.0655	0.9345
16		137	0.3	0.03	0.0300	0.0046	0.9954		0.0725	0.0808	0.9192

In the course of the research, the values of relative precision of correct parcel classification RP (1) during training (RP_T) and validation (RP_V), the number of first-layer neurons $S1$, and four mean squared

error values (Table 4) were recorded: MSE (5), MSE_{TR} , MSE_V , MSE_{VR} . The MSE_{TR} values were calculated using the formula:

$$MSE_{TR} = mse(Y_T, Y_o) = \frac{1}{N} \sum_{n=1}^N (\text{round}(Y_{Tn}) - \text{round}(Y_{On}))^2, \quad (6)$$

where the round() function defines the rounding operation to the nearest integer according to mathematical rules.

The values $MSE_V = mse(Y_V, Y_O)$ and $MSE_{VR} = mse(Y_V, Y_O)$ were calculated for the validation output Y_O and target Y_V data using (5) and (6).

The use of MSE_V and MSE_{VR} is associated with the fact that the relative precision of parcel classification RP_T , RP_V by loading chute (door) numbers is determined from the RBF network output data after rounding them to integers; therefore, the standard MSE values are uninformative. As the study showed (rows 1–10, 15 of Table 4), achieving zero standard MSE values does not lead to expected results, which is explained by the presence of errors exceeding ± 0.5 . The value of the specified error "Goal" determines the number of first-layer neurons: as it increases, the $S1$ value decreases (rows 2, 9–12 of Table 4). When using large-size samples and decreasing the "Goal" value

(rows 13–16 of Table 4), the number of neurons $S1$ can reach the dimension N of the training sample.

In accordance with the problem statement, the RBF model (row 2 of Table 4) was selected, providing the required relative precision of correct parcel classification $RP_V = 95\%$ with a minimum number of neurons $S1 = 67$.

5.5. Investigation of Generalized Regression Neural Network (GRNN) models

The architecture of the GRNN model in MATLAB is presented in Fig. 4. Its first layer is analogous to the RBF network architecture (Fig. 3) but differs in the principle of forming the weight vector $W1$ (net.IW{1,1}) values. These values exactly coincide with the training sample P_N data values. Each neuron in the first layer computes the Euclidean distance and the radial basis function in accordance with (3) and (4).

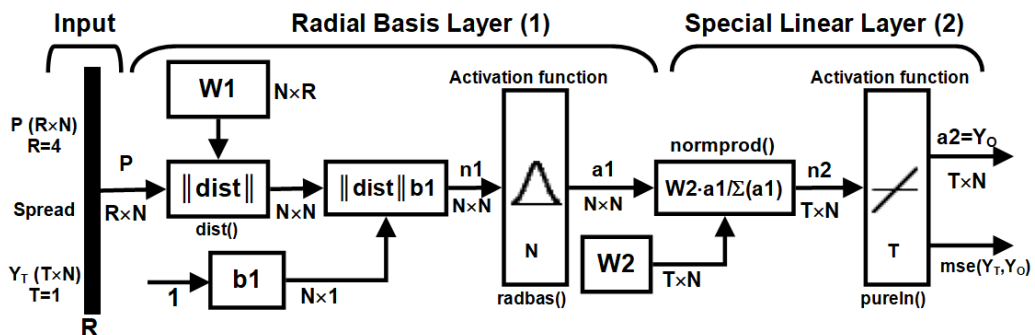


Fig. 4. Architecture of the GRNN model in MATLAB

The GRNN linear neuron layer computes the weighted mean of the outputs $a1$ associated with each input vector P and the weights of vector $W2$

$$n2 = \frac{W2 \cdot a1}{\sum a1} = \frac{\text{net.LW}\{2,1\} \cdot a1}{\text{sum}(a1)} = \text{normprod}(\text{net.LW}\{2,1\}, a1).$$

The formed weighted outputs $n2$, after activation by the linear function, form the GRNN output: $Y_O = a2 = \text{purelin}(n2)$. The number of GRNN linear layer neurons T is determined by the dimension of the target dataset Y_T (for the problem under consideration, $T = 1$).

Training and validation of GRNN models for parcel classification research were performed using

(net.LW{2,1}). For this purpose, the normalised dot product is calculated:

the datasets given in Table 3. GRNN training by the backpropagation method does not include interactive adjustment of the number of first-layer neurons, which is determined by the dimension N of the input training data P_N and weights $W1$. The best training and testing results of GRNN models, obtained for datasets "dataset1", "dataset2", "dataset3", "dataset4" are presented in Table 5.

Table 5. Results of GRNN model training and validation

No	Training						Validation			
	Data Set	N_1	Spread	MSE	MSE_{TR}	RP_T	Data Set	MSE_V	MSE_{VR}	RP_V
1	dataset1 (4×81)	81	0.1	0.0	0.0	1.0	dataset6 (4×28561)	0.0228	0.0126	0.9874
2		81	0.2	0.0	0.0	1.0		0.0459	0.0329	0.9671
3		81	0.3	0.0261	0.0	1.0		0.0831	0.0738	0.9262
4		81	0.5	0.1191	0.1235	0.8765		0.1375	0.1531	0.8469
5		81	0.7	0.1983	0.2222	0.7778		0.1839	0.2576	0.7424
6	dataset1 (4×81)	81	0.9	0.2562	0.3827	0.6173	dataset6 (4×28561)	0.2182	0.3981	0.6019
7		81	1.0	0.2773	0.4568	0.5432		0.2307	0.4333	0.5667
8	dataset2 (4×256)	256	0.1	0.0001	0.0	1.0	dataset6 (4×28561)	0.2358	0.2682	0.7318
9		256	0.3	0.0640	0.0234	0.9766		0.1350	0.2147	0.7853
10		256	0.5	0.1935	0.3594	0.6406		0.1625	0.2447	0.7553
11		256	1.0	0.4490	0.6250	0.3750		0.2816	0.4431	0.5569
12	dataset3 (4×625)	625	0.1	0.1052	0.1744	0.8256	dataset6 (4×28561)	0.1618	0.1989	0.8011
13		625	0.2	0.1291	0.1808	0.8192		0.1219	0.1755	0.8245
14		625	0.3	0.1553	0.1952	0.8048		0.1163	0.1620	0.8380
15		625	0.5	0.2179	0.3456	0.6544		0.1678	0.2951	0.7049
16	dataset4 (4×1296)	1296	0.1	0.0030	0.0	1.0	dataset6 (4×28561)	0.0485	0.0629	0.9371
17		1296	0.3	0.0957	0.1057	0.8943		0.0965	0.1013	0.8987
18		1296	1.0	0.3158	0.5054	0.4946		0.2423	0.4429	0.5571

During the GRNN model research (similarly to Table 4), the values of relative precision of correct parcel classification RP_T , RP_V , the "Spread" parameter value, the number of first-layer neurons N_1 , and four mean squared error values MSE , MSE_{TR} , MSE_V , MSE_{VR} were recorded.

As the study showed, for each training sample the "Spread" parameter must be selected to achieve the maximum value of relative parcel parameter classification precision RP_V by loading chute numbers (ranges). For training datasets "dataset1", "dataset4" (rows 1–7, 16–18 of Table 5), the RP_V value decreases with increasing "Spread" value, and the maximum RP_V value was obtained for $Spread = 0,1$. For training datasets "dataset2", "dataset3" (rows 8–15 of Table 5), the maximum RP_V value was obtained for $Spread = 0,2$. The RP_V values for training datasets "dataset2", "dataset3", "dataset4" are lower than for "dataset1", despite their larger dimension. A deterioration in validation results was also observed (rows 1–2 of Table 5) when zero MSE values were achieved during training, which was caused by the presence of errors exceeding $\pm 0,5$.

In accordance with the problem statement, the GRNN model (row 1 of Table 5) was selected, providing the required relative precision of correct parcel

classification $RP_V = 99\%$ with a minimum number of neurons $N_1 = 81$.

5.6. Investigation of probabilistic neural network (PNN) models

The architecture of the PNN model in MATLAB is presented in Fig. 5. Training data, containing N vectors P for each of which the target data Y_T defines numerical class values in the form of loading chute (door) numbers $L_k = \{1, 2, 3\}$, is fed to the PNN input. To facilitate classification, the function $ind2vec(Y_T)$ converts the dimensionality of the target data Y_T from the $1 \times N$ format into a positional target vector matrix T of dimension $G \times N$. The parameter $G_{max} = 3$ specifies the number of classifiable vectors and is determined by the maximum value of the loading chute (door) number L_k , $k = \{1, 2, 3\}$. The $ind2vec()$ function converts the numerical class values $\{1, 2, 3\}$ into three vectors $\{(1\ 0\ 0)^T, (0\ 1\ 0)^T, (0\ 0\ 1)^T\}$, respectively.

The first PNN layer (Fig. 5) is analogous to the GRNN architecture (Fig. 4), where the weight vector W_1 (net.IW{1,1}) values exactly coincide with the training sample P_N values. Each neuron in the first layer computes the Euclidean distance and the radial basis function in accordance with (3) and (4).

The second-layer weight matrix $W2$ (net.LW{2,1}) corresponds to the target vector matrix T . The product $W2 \cdot a1$ determines N elements of matrix $n2$ and their numerical correspondence to each of the G classes. The competing activation function $compet(n2)$ creates the output data array $a2$. This array contains only one value equal to "1" for the largest element value of

vector $n2$ and "0" in all other cases. The $vec2ind(a2)$ function is used to convert the positional matrix $a2$ to the numerical class representation.

Training and validation of PNN models for parcel classification research were performed using the datasets given in Table 3. The best training and testing results of PNN models, obtained for datasets "dataset1"– "dataset4", are presented in Table 6.

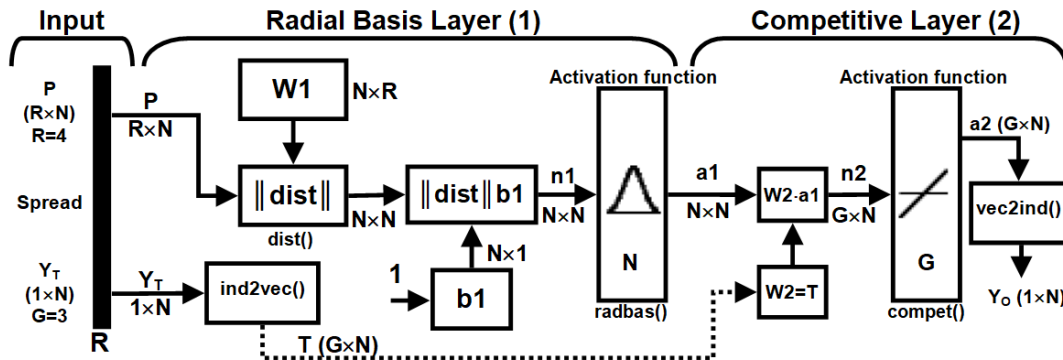


Fig. 5. Architecture of the PNN model in MATLAB

Table 6. Results of PNN model training and validation

No	Training					Validation		
	Data Set	$N1$	Spread	MSE_{TR}	RP_T	Data Set	MSE_{VR}	RP_V
1	dataset1 (4×81)	81	0,1	0.00	1.0	dataset6 (4×28561)	0.0126	0.9874
2		81	0,2	0.00	1.0		0.0329	0.9671
3		81	0,3	0.00	1.0		0.0736	0.9264
4		81	0,5	0.1235	0.8765		0.1303	0.8697
5		81	0,7	0.1605	0.8395		0.1485	0.8515
6		81	0,9	0.1728	0.8272		0.1903	0.8097
7		81	1,0	0.1728	0.8272		0.2203	0.7797
8	dataset2 (4×256)	256	0,1	0.00	1.0	dataset6 (4×28561)	0.2764	0.7236
9		256	0,3	0.00	1.0		0.2712	0.7288
10		256	0,5	0.1484	0.8516		0.2524	0.7476
11		256	1,0	0.3555	0.6445		0.2164	0.7836
12	dataset3 (4×625)	625	0,1	0.1744	0.8256	dataset6 (4×28561)	0.1996	0.8004
13		625	0,2	0.1808	0.8192		0.1838	0.8162
14		625	0,5	0.1904	0.8096		0.1612	0.8388
15		625	1,0	0.4704	0.5296		0.4398	0.5602
16	dataset4 (4×1296)	1296	0,1	0.00	1.0	dataset6 (4×28561)	0.0629	0.9371
17		1296	0,3	0.1057	0.8943		0.1035	0.8965
18		1296	1,0	0.2917	0.7083		0.2411	0.7589

During the PNN model research, the values of relative precision of correct parcel classification during training (RP_T) and validation (RP_V), the "Spread" parameter value, and the number of first-layer neurons $N1$ were recorded. Unlike the previously considered neural networks, the model output data are integers and do not require rounding. Therefore, only MSE_{TR} and MSE_{VR} were calculated.

For training datasets "dataset1", "dataset4" (rows 1–7, 16–18 of Table 6), the RP_V value decreases with increasing "Spread" value, and the maximum RP_V value was obtained for $Spread = 0,1$. The maximum RP_V value for training dataset "dataset2" (rows 8–11 of Table 6) was obtained for $Spread = 1,0$, and for "dataset3" (rows 12–15 of Table 6), $Spread = 0,5$.

The RP_V values for training datasets "dataset2", "dataset3", "dataset4" are lower than for "dataset1", despite their larger dimension. The minor differences in the obtained data between the PNN (Table 6) and GRNN (Table 5) models are explained by the fact that the architectures of their first RB layer are completely analogous.

In accordance with the problem statement, the PNN model (row 1 of Table 6) was selected,

providing the required relative precision of correct parcel classification $RP_V = 99\%$ with a minimum number of neurons $N1 = 81$.

5.7. Investigation of Feed-Forward Neural Network (FFNN) Models

The architecture of the FFNN model in MATLAB is presented in Fig. 6.

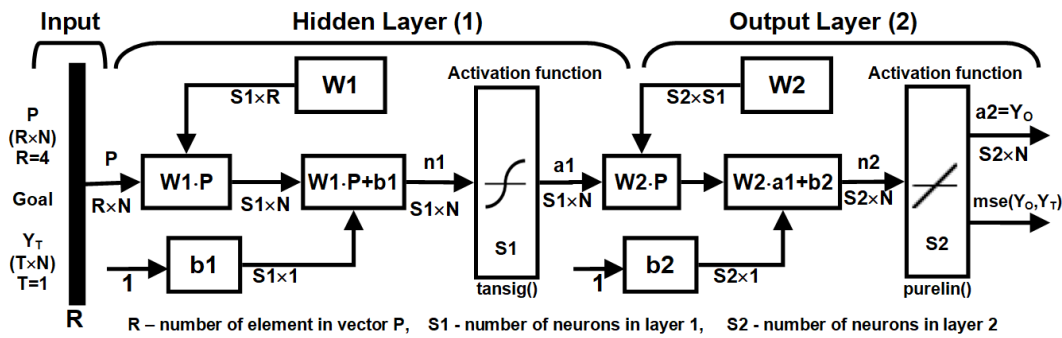


Fig. 6. Architecture of the FFNN model in MATLAB

For the datasets P_N used (Table 3), with dimension $R \times N$, the FFNN has 4 inputs ($R = 4$). In the first (hidden) layer, the P values are linearly combined using weights W_1 (net.IW{1,1}) and biases b_1 (net.b{1}). The formed combinations $n_1 = W_1 \cdot P + b_1$, after activation by the hyperbolic tangent function, form the hidden layer output of the FFNN: $a_1 = \text{tansig}(n_1)$. In the second layer, the values a_1 are linearly combined using weights W_2 (net.LW{2,1}) and biases b_2 (net.b{2}). The combinations $n_2 = W_2 \cdot a_1 + b_2$, after activation by the linear function, form the FFNN output: $Y_0 = a_2 = \text{purelin}(n_2)$. The number of linear layer neurons S_2 (the number of FFNN outputs) is determined by the dimension T of the target dataset Y_T (for the problem under consideration, $S_2 = T = 1$).

The backpropagation method is used for training the FFNN. During FFNN training, the weights W_1 , W_2 (Fig. 6) of the hidden and output FFNN layers are adjusted. The MATLAB toolkit implements 14 training algorithms. Based on the research results, two of them were selected: the Levenberg–Marquardt (LM) algorithm and the Bayesian Regularisation (BR) algorithm.

For conducting parcel classification research using the FFNN, the following were specified: the training algorithm (LM, BR), the number of first-layer neurons S_1 , input data P_N , and target data Y_T (Table 3). Table 7 presents a fragment of the FFNN model training and testing results.

During the FFNN model research, the values of relative precision of correct parcel classification RP_T , RP_V , the number of first-layer neurons S_1 , and four mean squared error values MSE , MSE_{TR} , MSE_V , MSE_{VR} were recorded. As the FFNN modelling demonstrated, for the two training algorithms used and training datasets with dimension N less than 1296, unsatisfactory values at the level of 90% or below were obtained (rows 1–4, 10–12 of Table 7). Increasing the volume of training datasets yields only a slight MSE decrease, with errors still exceeding ± 0.5 , and an RP_V gain of only 0.1–0.3%.

In accordance with the problem statement, the FFNN model (row 9 of Table 7) trained using the Levenberg–Marquardt algorithm was selected, providing the required relative precision of correct parcel classification $RP_V = 97\%$ with a minimum number of first-layer neurons $S_1 = 81$.

Table 7. Results of FFNN model training and validation

No	Training					Validation			
	Data Set	S1	MSE	MSE _{TR}	RP _T	Data Set	MSE _V	MSE _{VR}	RP _V
Levenberg–Marquardt (LM) training method									
1	dataset1 (4×81)	20	0.0416	0.0617	0.9383	dataset6 (4×28561)	0.1118	0.1381	0.8628
2		24	0.0357	0.0247	0.9753		0.1125	0.1517	0.8483
3		28	0.1004	0.1111	0.8889		0.1542	0.1804	0.8196
4	dataset4 (4×1296)	20	0.0305	0.0193	0.9807	dataset6 (4×28561)	0.0628	0.0692	0.9308
5		24	0.0327	0.0270	0.9730		0.0636	0.0710	0.9290
6		28	0.0484	0.0255	0.9745		0.0802	0.1061	0.8939
7	dataset5 (4×6561)	20	0.0449	0.0453	0.9547	dataset6 (4×28561)	0.0540	0.0550	0.9450
8		24	0.0510	0.0590	0.9410		0.0448	0.0396	0.9604
9		28	0.0447	0.0430	0.9570		0.0403	0.0302	0.9698
Bayesian Regularization (BR) training method									
10	dataset1 (4×81)	20	0.0406	0.0370	0.9630	dataset6 (4×28561)	0.1010	0.1106	0.8894
11		24	0.0422	0.0741	0.9259		0.0990	0.0908	0.9092
12		28	0.0368	0.0494	0.9506		0.0985	0.1082	0.8918
13	dataset4 (4×1296)	20	0.0257	0.0039	0.9961	dataset6 (4×28561)	0.0703	0.0668	0.9332
14		24	0.0198	0.0039	0.9961		0.0694	0.0696	0.9304
15		28	0.0171	0.0031	0.9969		0.0754	0.0731	0.9269
16	dataset5 (4×6561)	20	0.0369	0.0346	0.9654	dataset6 (4×28561)	0.0680	0.0647	0.9353
17		24	0.0354	0.0373	0.9627		0.0729	0.0828	0.9172
18		28	0.0344	0.0340	0.9660		0.1732	0.1883	0.8437

5.8. Discussion of Research Results

In accordance with the problem statement, an investigation of neural network models was conducted, using as quality metrics the minimum number of first-layer neurons $N1$, MSE , and relative precision of correct parcel classification RP_V by the weight and dimensional ranges defined by loading chute

(loading door) numbers. The neural networks selected based on the research results are presented in Table 8.

An analysis of the obtained results shows that the GRNN and PNN networks do not satisfy the selection criteria. Their principal disadvantage is that during training, the number of neurons in their hidden (first) layer is determined by the number of elements in the training dataset.

Table 8. Comparative evaluation results of trained neural network models

Вид NN	$N1$	MSE	MSE_{TR}	RP_T	MSE_V	MSE_{VR}	RP_V
RBF	67	0.0099	0.0	1.0	0.0609	0.0523	0.9477
GRNN	81	0.0	0.0	1.0	0.0228	0.0126	0.9874
PNN	81	–	0.00	1.0	1.0	0.0126	0.9874
FFNN	28	0.0447	0.0430	0.9570	0.0403	0.0302	0.9698

It should be noted that the implementation of exponential radial functions for the GRNN, PNN, and RBF networks leads to a slowdown in their operation due to large computational volumes. In this regard, the FFNN is the most promising, as it provides the required relative precision of correct parcel classification ($RP_V = 97\%$), and its hidden layer is implemented using the minimum number of neurons ($N1 = 28$) compared to the other neural networks.

The computational volume of the trained FFNN is associated with vector addition and multiplication operations, which significantly simplifies its software implementation. The values of the weight matrix $W1$ (28×4) and weight vectors $W2$ (28×1), $b1$ (28×1), and $b2$ for the software implementation of the trained FFNN are presented in Table 9 (in the table, the matrix and vector elements are presented in transposed form).

Table 9. Weight coefficient values of the trained FFNN model

$W1^T (4 \times 28)$																											
1,6	-0,4	-0,5	-0,1	-5,0	-3,2	-9,2	-14,9	-6,5	-0,6	-13,1	2,7	1,2	-0,8	-1,8	-0,9	11,5	-1,2	-3,7	-1,2	-1,3	-0,5	0,3	-21,1	-1,5	5,0	1,3	1,5
0,0	-0,2	-0,6	0,6	-1,0	0,0	-0,7	-0,6	7,2	0,4	0,3	-0,6	-0,4	1,9	-2,2	2,0	-0,5	-0,7	0,2	6,8	-0,5	0,3	0,5	-1,6	0,8	-1,0	-0,7	0,7
0,3	1,2	1,6	1,4	-1,1	3,4	-0,8	-0,6	0,1	1,0	0,2	-0,9	0,9	0,0	0,1	0,1	-0,5	-0,1	-4,3	-2,0	-0,2	-1,1	-0,8	-1,7	0,6	-0,6	-0,6	1,2
7,6	-0,1	-0,7	-0,7	-0,8	-0,3	-0,6	-0,5	0,0	-0,4	0,2	-1,6	-1,4	-1,2	0,2	-1,4	-0,4	1,2	-4,5	0,6	1,3	-0,6	-0,2	-1,6	0,0	1,8	0,2	1,5
$b1^T (1 \times 28)$																											
-4,3	1,3	-1,8	2,3	1,0	1,6	3,5	6,6	3,1	0,6	-6,7	-0,7	0,4	-0,3	-0,9	-0,4	5,9	-0,5	-1,8	3,7	-0,5	-0,7	-1,1	-10,7	-2,2	5,3	1,9	2,9
$W2^T (1 \times 28)$																											
0,1	-1,7	-0,2	0,8	-0,5	-0,1	2,5	-2,6	-0,1	0,6	-2,4	0,2	0,3	0,8	0,3	-0,6	-1,9	-1,8	-0,1	0,0	2,2	0,4	-1,8	0,3	-2,7	0,2	-3,2	0,5
$b2 = -0.59$																											

6. Conclusions and prospects for further research

The conducted study made it possible to determine the advantages and limitations of the neural network architectures selected for implementing the ASL decision-making model that classifies parcels into three ranges of weight and dimensions and implements the specified logic of their sorting across PSC loading doors and chutes within a conveyor flow.

A comparative analysis of such two-layer neural network architectures as RBF, GRNN, and PNN demonstrated that the FFNN provides the required relative precision of correct parcel classification within a conveyor flow according to their weight and dimensional ranges defined for PSC loading doors and chutes, with the minimum number of neurons employed in the hidden layer.

It is particularly noteworthy that the GRNN and PNN, while capable of achieving higher levels of relative precision of correct parcel classification, are limited in their applicability by the number of neurons in their hidden (first) layer, which is determined by the number of elements in the training dataset. A further disadvantage of the GRNN, PNN, and RBF is the significant computational volume caused by the large number of neurons and the use of exponential radial functions; however, such neural networks may prove useful in other application scenarios.

Considering the identified shortcomings, the following directions for further research can be defined:

- conducting research into possible scenarios for using decision-making models based on GRNN, PNN, and RBF to implement parcel sorting planning prior to their arrival at the PSC. With a prepared sorting plan, it is possible to implement short parcel transport routes along the PSC conveyor for their loading with consideration of weight and dimensional parameters;

- conducting research into possible scenarios for using FFNN, GRNN, PNN, and RBF at cross-docking transshipment distribution centers to implement models for planning preliminary goods sorting, as well as ASL decision-making models for direct goods sorting. Such use of neural networks enables the planning of goods unloading and their distribution across storage racks. An alternative scenario involves planning goods sorting using their labels, where the ASL is used for consolidating goods onto pallets with consideration of weight and dimensions.

Thus, the comparative analysis not only identified the most effective neural network architectures but also revealed their shortcomings and limitations. The research results demonstrated that solving the problem of parcel sorting within a conveyor flow requires implementing an ASL decision-making model based on the FFNN, which will leverage the advantages of its architecture.

Conflict of interest

The authors declare that they have no conflict of interest with respect to this research, including financial, personal, authorship, or other conflicts that could influence the research and its results presented in this article.

Funding

The study was conducted without financial support.

Data availability

The manuscript has no associated data.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence technology in the creation of this work.

References

1. McWilliams, D., Stanfield, P., Geiger, C. (2005), "The Parcel Hub Scheduling Problem: A Simulation-Based Solution Approach", *Computers & Industrial Engineering*, Vol. 49(3). P. 393–412. DOI: <https://doi.org/10.1016/j.cie.2005.07.002>
2. Akkerman, F., Lalla-Ruiz, E., Mes, M., Spitters, T. (2022), "Cross-Docking: Current Research Versus Industry Practice and Industry 4.0 Adoption", *Smart Industry – Better Management*, Vol. 28, P. 70–104. DOI: <https://doi.org/10.1108/S1877-636120220000028007>
3. Soni, B., Mathur, P., Bora, A. (2021), "In Depth Analysis, Applications and Future Issues of Artificial Neural Network", *In: Enabling AI Applications in Data Science. Studies in Computational Intelligence*, Vol. 911, P. 149–183. Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-52067-0_7
4. Katal, A., Singh, N. (2022), "Artificial Neural Network: Models, Applications, and Challenges", *In: Innovative Trends in Computational Intelligence. EAI/Springer Innovations in Communication and Computing*. P. 235–257. Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-78284-9_11
5. Daniel, G. (2019), "Principles of Artificial Neural Networks. Basic Designs to Deep Learning", *The World Scientific Press*, 440 p. ISBN: 978-981-12-0122-6. DOI: <https://doi.org/10.1142/11306>
6. Bugow, S., Kellenbrink, C. (2023), "The Parcel Hub Scheduling Problem With Limited Conveyor Capacity and Controllable Unloading Speeds", *OR Spectrum*, No 45, P. 325–357. DOI: <https://doi.org/10.1007/s00291-022-00702-y>
7. McWilliams, D., Stanfield, P., Geiger, C. (2008), "Minimizing the completion time of the transfer operations in a central parcel consolidation terminal with unequal-batch-size inbound trailers", *Computers & Industrial Engineering*, Vol. 54 (4), P. 709–720, <https://doi.org/10.1016/j.cie.2007.10.006>
8. Chen J., Chen, T., Ou, T., Lee, Y. (2019), "Adaptive Genetic Algorithm for Parcel Hub Scheduling Problem With Shortcuts in Closed-Loop Sortation System". *Computers & Industrial Engineering*. Vol. 138:106114. DOI: <https://doi.org/10.1016/j.cie.2019.106114>
9. Chen, J., Chen, T., Lee, Y. (2023), "Simulation Optimization For Parcel Hub Scheduling Problem in Closed-Loop Sortation System with Shortcuts", *Simulation Modelling Practice and Theory*, Vol. 124(10):102728. DOI: <https://doi.org/10.1016/j.simpat.2023.102728>
10. Grebennik, I., Dupas, R., Lytvynenko, O., Urniaieva, I. (2017), "Scheduling Freight Trains in Rail-rail Transshipment Yards with Train Arrangements", *International Journal of Intelligent Systems and Applications (IJISA)*, Vol. 9(10), P. 12–19. DOI: <https://doi.org/10.5815/ijisa.2017.10.02>
11. Chen, T., Chen, J., Huang, C., Chang, P. (2021), "Solving the Layout Design Problem by Simulation-Optimization Approach – A Case Study on a Sortation Conveyor System", *Simulation Modelling Practice and Theory*, Vol. 106:102192. DOI: <https://doi.org/10.1016/j.simpat.2020.102192>
12. Jarrah, A., Xiangtong, Q., Bard, J. (2014), "The Destination-Loader-Door Assignment Problem for Automated Package Sorting Centers", *Transportation Science*, No 50(4), P. 1314–1336. DOI: <https://doi.org/10.1287/trsc.2014.0521>
13. Karami, F., Fathi, M., Pardalos, P. (2023), "Conveyor Operations in Distribution Centers: Modeling and Optimization", *Optim Lett*, Vol. 17, P. 1049–1068. DOI: <https://doi.org/10.1007/s11590-022-01912-7>
14. Werners, B., Wülfing, T. (2010), "Robust Optimization of Internal Transports at a Parcel Sorting Center Operated by Deutsche Post World Net", *Eur. J. Oper. Res*, Vol. 201(2), P. 419–426. DOI: <https://doi.org/10.1016/j.ejor.2009.02.035>
15. Fedtke, S., Boysen, N. (2014), Layout Planning of Sortation Conveyors in Parcel Distribution Centers. *Transportation Science*, Vol. 51(1), P: 3–18. DOI: <https://doi.org/10.1287/trsc.2014.0540>
16. Lee, Y., Jung, J., Lee, K. (2006), Vehicle Routing scheduling for Cross-Docking in the Supply Chain, *Computers & Industrial Engineering*, Vol. 51(2), P. 247–256. DOI: <https://doi.org/10.1016/j.cie.2006.02.006>
17. McAree, P., Bodin, L., Ball, M. (2002), "Models for the Design and Analysis of a Large Package Sort Facility. Networks", Vol. 39, P. 107–120. DOI: <https://doi.org/10.1002/net.10017>
18. Bozer, Y., Hsieh, Y. (2004), "Expected Waiting Times at Loading Stations in Discrete-space Closed-loop Conveyors", *European Journal of Operational Research*, Vol. 155(2), P. 516–532. DOI: [https://doi.org/10.1016/S0377-2217\(02\)00886-X](https://doi.org/10.1016/S0377-2217(02)00886-X)
19. Oh, Y., Cha, C., Lee, S. (2006), "A Dock-door Assignment Problem for the Korean Mail Distribution Center", *Computers & Industrial Engineering*, Vol. 51(2), P. 288–296. DOI: <https://doi.org/10.1016/j.cie.2006.02.009>
20. Choy, K., Chow, H., Poon, T., Ho, G. (2012), "Cross-dock Job Assignment Problem in Space-constrained Industrial Logistics Distribution Hubs with a Single Docking Zone", *International Journal of Production Research*, Vol. 50(9), P. 2439–2450. DOI: <https://doi.org/10.1080/00207543.2011.581006>
21. Romanova, T., Litvinchev, I., Grebennik, I., Kovalenko, A., Urniaieva, I. (2020), "Packing Convex 3D Objects with Special Geometric and Balancing Conditions", *Advances in Intelligent Systems and Computing*, Vol. 1072. DOI: https://doi.org/10.1007/978-3-030-33585-4_27
22. Grebennik, I., Kovalenko, O. (2024), "Realisation of a Given Trucks Loading Logic using a Fuzzy Decision Making Model", *14th International Conference on Advanced Computer Information Technologies (ACIT)*, IEEE, P. 27–31. DOI: <https://doi.org/10.1109/ACIT62333.2024.10712615>
23. Grebennik, I., Kovalenko, O. (2025), "Adaptive Neuro-Fuzzy Inference System (ANFIS) for Control of Loading Logic in Parcel Sorting Centers", *15th International Conference on Advanced Computer Information Technologies (ACIT)*, IEEE, P. 46–50, DOI: <https://doi.org/10.1109/ACIT65614.2025.11185826>
24. Nawi, N., Atomi, W., Rehman, M. (2013), "The Effect of Data Pre-Processing on Optimized Training of Artificial Neural Networks, *Procedia Technology*", Vol. 11, P. 32–39. DOI: <https://doi.org/10.1016/j.protcy.2013.12.159>

Received (Надійшла) 09.02.2026

Accepted for publication (Прийнята до друку) 20.02.2026

Publication date (Дата публікації) 12.03.2026

Відомості про авторів / About the Authors

Гребеннік Ігор Валерійович – доктор технічних наук, професор, Харківський національний університет радіоелектроніки, завідувач кафедри комп'ютерного моделювання та інтелектуальних технологій; Харків, Україна;

Igor Grebennik – Doctor of Technical Sciences, Professor, Kharkiv National University of Radio Electronics, Head at the Computer Modelling and Intelligent Technologies Department; Kharkiv, Ukraine;

e-mail: igor.grebennik@nure.ua

ORCID ID: <https://orcid.org/0000-0003-3716-9638>

Коваленко Олексій Андрійович – Харківський національний університет радіоелектроніки, аспірант кафедри комп'ютерного моделювання та інтелектуальних технологій; Харків, Україна;

Oleksii Kovalenko – Kharkiv National University of Radio Electronics, Postgraduate Student at the Computer Modelling and Intelligent Technologies Department; Kharkiv, Ukraine;

e-mail: oleksii.kovalenko3@nure.ua

ORCID ID: <https://orcid.org/0009-0008-4779-6161>

ВИКОРИСТАННЯ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ ДЛЯ УПРАВЛІННЯ СОРТУВАННЯМ ПОСИЛОК У КОНВЕЄРНОМУ ПОТОЦІ

Предметом дослідження є моделі прийняття рішень автоматизованих сортувальних ліній (ASL), якими оснащені перевалкові центри сортування посилок (PSC). **Мета роботи** – розроблення моделі прийняття рішень для ASL завантажувальних дверей терміналів PSC у вигляді штучної нейронної мережі (ANN), що виконує завдання сортування посилок з огляду на їх ваги й габарити. У статті розв'язано такі **завдання**: проаналізовано конструктивні особливості модифікації обладнання ASL для реалізації трьох варіантів логіки завантаження вантажівок; визначено діапазони ваги й габаритів посилок, пов'язаних з номерами завантажувальних лотків і дверей; встановлено критерії для реалізації заданої логіки сортування посилок; визначено параметри наборів даних для навчання й тестування моделей штучних нейронних мереж (ANNs); з використанням інструментарію середовища MATLAB проведено навчання й тестування чотирьох видів ANNs: RBFNN, GRNN, PNN, FFNN; із застосуванням як метрики якості відносної точності правильної класифікації посилок за діапазонами їх ваги й габаритів і кількості використовуваних нейронів проведено порівняльний аналіз ANNs. **Методи**: системний, аналітичний, комп'ютерного моделювання, методи навчання з архітектурою ANN, математичний і статистичний аналіз ефективності навчання. **Досягнуті результати**. Проведено порівняльне навчання й тестування чотирьох видів ANNs: RBFNN, GRNN, PNN, FFNN. За результатами тестування для реалізації обрано модель FFNN, навчену методом Левенберга – Марквардта, який забезпечує відносну точність класифікації ОПВ на рівні 97%. **Висновки**. Розроблена модель прийняття рішень ASL у вигляді FFNN дає змогу здійснити класифікацію посилок у потоці конвеєра за трьома діапазонами їх вагогабаритних параметрів для реалізації заданої логіки сортування за завантажувальними дверима або лотками. Реалізована логіка сортування забезпечує компактне завантаження кузовів вантажівок і знижує ризики пошкодження посилок, коли вони укладаються одна на одну.

Ключові слова: логістика дорправлення; модель прийняття рішень; класифікація; автоматизована сортувальна лінія; штучні нейронні мережі.

Бібліографічні описи / Bibliographic descriptions

Гребеннік І. В., Коваленко О. А. Використання штучних нейронних мереж для управління сортуванням посилок у конвеєрному потоці. *Автоматизовані системи управління та прилади автоматики*. 2026. № 1 (188). С. 17–32. DOI: <https://doi.org/10.30837/0135-1710.2026.188.017>

Grebennik, I., Kovalenko, O. (2026), "Use of artificial neural networks for managing the sorting of parcels in a conveyor flow", *Management Information System and Devices*, No. 1 (188), P. 17–32. DOI: <https://doi.org/10.30837/0135-1710.2026.188.017>

Неліпа О. Д., Калита Н. І.

ГІБРИДНІ МОДЕЛІ ШТУЧНОГО ІНТЕЛЕКТУ В ЗАДАЧАХ АВТОМАТИЗОВАНОГО ОЦІНЮВАННЯ ДОСТОВІРНОСТІ РЕЗУЛЬТАТІВ ВАЛІДАЦІЇ ІНФОРМАЦІЙНИХ СИСТЕМ

Предметом дослідження є автоматизоване оцінювання достовірності результатів валідації в інформаційних системах з огляду на структурні залежності між компонентами й семантичний зміст валідаційних артефактів. Мета роботи – підвищення надійності та відтворюваності валідації за допомогою розроблення гібридної моделі штучного інтелекту, що інтегрує графові та семантичні види подання даних для регресійного оцінювання рівня достовірності. Завдання дослідження: аналіз структурних і структурно-семантичних моделей (AST, PDG, CPG, Vul-LMGNN, ECHO, C-PCFG) у контексті валідації інформаційних систем; формулювання математичної постановки задачі оцінювання достовірності; розроблення архітектури гібридної моделі на основі GNN і трансформерного кодувальника; реалізація механізму інтеграції структурної та семантичної модальностей у спільному векторному просторі; експериментальне дослідження на графовому фінансовому наборі даних DgraphFin V2 й порівняння з базовими моделями. Методи дослідження передбачають графові нейронні мережі, трансформерний кодувальник типу BERT/CodeBERT, регресійне моделювання із сигмоїдною активацією, оптимізацію методом Adam та статистичне оцінювання якості й калібрування (MSE, MAE, ECE, R^2 , Spearman ρ , ROC-AUC). Результати роботи. Запропонована гібридна архітектура забезпечує регресійне оцінювання достовірності в інтервалі [0;1] і демонструє стабільну збіжність без ознак перенавчання, а також кращі значення всіх ключових метрик порівняно з чисто семантичними, чисто структурними та класичними статистичними моделями. Аналіз ваг механізму самоуваги підтвердив коректну ідентифікацію значущих структурних і семантичних залежностей. Висновок: запропонована модель забезпечує більш точне каліброване оцінювання достовірності результатів валідації, що дає змогу автоматизувати відбір ненадійних результатів, зменшити вплив людського фактора й підвищити відтворюваність валідаційних процедур. Перспективним напрямом подальших досліджень є розвиток мультимодальних архітектур із динамічним перерозподілом ваг між модальностями.

Ключові слова: гібридні моделі штучного інтелекту; валідація інформаційних систем; графові нейронні мережі; трансформерні моделі; оцінювання достовірності; DgraphFin V2; регресійне моделювання; калібрування ймовірностей.

1. Вступ

Стабільне підвищення комплексності сучасних інформаційних систем стало причиною важливості етапу валідації як ключового складника їх функціонування. Із зростанням обсягів даних і ускладнення багатокomпонентних архітектур програмного забезпечення підтримка достовірності й точності результатів валідації насамперед визначає загальну цілісність системи. Традиційні методи валідації здебільшого залежать від людського досвіду та інтерпретації вручну, що неминує призводить до суб'єктивізації процесу оцінювання, його непослідовності та обмеженої розширюваності. Крім того, не є явною і перспектива відтворюваності традиційних методів валідації, оскільки підсумкова оцінка часто залежить від конкретного експерта або особливостей контексту тестування.

Технології штучного інтелекту визначаються альтернативною парадигмою в цьому контексті, дотримання якої здатне усунути зазначені обмеження

за допомогою автоматизації оцінювання результатів валідації та зниження впливу людського фактора. З огляду на вищесказане подібні архітектури моделей застосовуються тоді, коли обсяг даних і складність взаємозв'язків у межах інформаційної системи (ІС) роблять мануальну перевірку непрактичною.

Окреслена проблема є особливо актуальною в таких середовищах, як банкові та фінансові платформи, корпоративні бази даних, інфраструктурні варіації DevOps, а також в аналітичних та інженерних обчислювальних комплексах, для яких цільова модель здатна інтерпретувати не тільки структурні залежності між компонентами, але й текстові матеріали, що дає змогу виявляти суперечності й аномалії, на які в іншому разі можна не зважати.

2. Аналіз літературних джерел і визначення проблеми дослідження

Спершу варто звернути увагу на структурно-семантичний аналіз даних і програмного складника

IC, що спирається на сукупність подання, призначену для фіксації як синтаксичної структури, так і логічних залежностей у межах системи. Серед такого подання основоположну позицію мають три базові структури: абстрактні синтаксичні дерева (*Abstract Syntax Tree* – AST) [1], графи програмних залежностей (*Program Dependency Graph* – PDG) [2] і граф властивостей коду (*Code Property Graph* – CPG) [3], які утворюють теоретичне підґрунтя сучасних гібридних інтелектуальних моделей, орієнтованих на структурний аналіз.

AST відтворюють ієрархічну структуру вихідного коду або вхідної інформації, переводячи їх граматичну організацію у формалізовану форму. Кожний вузол у дереві відповідає синтаксичній конструкції, тобто вислову, оператору або певній інструкції, а ребра визначають батьківські відношення між цими елементами.

На відміну від простих послідовних токенів, AST мають таку структуру, в якій спостерігається абстракція від поверхневих деталей типу пунктуації або форматування й водночас зберігається наявна структура програми. Така ієрархічна організація дає змогу виконувати семантичні перевірки, перетворення та аналітичні операції з вищою точністю, оскільки взаємозв'язки між компонентами є чітко закодованими.

Так, у дослідженні [1] описано підхід до кросмовної класифікації програм із використанням уніфікованого абстрактного синтаксичного дерева (UAST). Нейромережева модель була розроблена як поєднання двох підмереж – SAST (послідовної) та GAST (графової) – для одночасного вилучення глобальних синтаксичних і локальних семантичних ознак коду.

Запропоновано механізм уніфікованого словника, за допомогою якого зменшується різниця між мовами програмування способом нормалізації вузлів AST. Продемонстровано вищі показники точності, повноти й міри F1 порівняно з CodeBERT (*Bidirectional Encoder Representations from Transformers*) і InferCode на наборах даних із *Java*, *C++*, *Python*, *C* та *JavaScript*, що підтверджує ефективність моделі в завданні кросмовного семантичного аналізу коду.

Структурі PDG властиве розширення цієї парадигми, що передбачає подання залежностей даних і логіку керування в системі, які сукупно визначають характер роботи програми. У цих графах вузли відповідають операторам або інструкціям, а спрямовані ребра описують зв'язки, що визначають порядок виконання або потік даних. Керувальні

залежності відтворюють логічні умови, від яких залежить виконання певних інструкцій, тоді як залежності даних визначають принцип задавання, поширення та використання даних між операціями. Такий дуальний підхід дає змогу аналізувати як причинно-наслідкові, так і функціональні відношення, що є важливим для розуміння того, як зміни в одній частині процесу впливають на результати в іншій.

У роботі [2] запропоновано розширену модель – паралельно семантичну PDG (*Parallel Semantics Program Dependence Graph* – PS-PDG) – як еволюцію класичного PDG, який описує залежності даних і керування між інструкціями програми. Продемонстровано, що традиційний PDG, орієнтований на послідовний код, обмежує оптимізаційний простір для паралельних програм.

Модель PS-PDG узагальнює попередню структуру способом додавання ієрархічних вузлів, контекстів, направлених і ненаправлених ребер та змінних із паралельною семантикою. Завдяки цьому забезпечується точне збереження семантики паралельного виконання в компіляторах і оптимізація планів розпаралелювання для OpenMP (*Open Multi-Processing*) і Cilk, що сприяє зростанню продуктивності без порушення семантичної еквівалентності.

CPG об'єднують описані структури в єдину репрезентацію, суміщаючи синтаксичну, семантичну та керувальну інформацію, і таким чином є цілісною системою для аналізу та виявлення закономірностей.

У статті [3] розроблено підхід до інтеграції LLVM-IR (*Low Level Virtual Machine – Intermediate Representation*) у структуру CPG для уніфікованого подання вихідного й проміжного коду. Метою є подолання обмежень попередніх реалізацій CPG, що працювали лише з окремими мовами. Автори вперше запропонували повне відтворення інструкцій LLVM-IR у вузли CPG без втрати сумісності зі структурою, підтримали оброблення винятків LLVM і оптимізували граф, зменшивши кількість вузлів без втрати семантики. Експериментально підтверджено, що аналізи, створені для вихідного коду, можуть без змін використовуватись до LLVM-IR. Застосування до бібліотек *Rust* продемонструвало масштабованість і стабільність підходу (менше ніж 1% помилок парсингу), що робить модель придатною для багатомовного аналізу й безпечного виявлення вразливостей у компільованих або відновлених із бінарних файлів програмах.

Унаслідок перенесення цих структурних моделей в область валідації інформаційних систем вони набувають ширшого інтерпретаційного значення. У валідаційних підсистемах окремі елементи даних, правила перевірки й проміжні результати можуть розглядатися як вузли графа залежностей, а відношення, що визначають їх перетворення або взаємодію з ними, – як ребра. Така структурна репрезентація дає змогу автоматично виявляти невідповідності та логічні відхилення.

Отже, гібридність моделей штучного інтелекту в цьому дослідженні розглядається як синтез глибоких мовних видів подання за типом BERT або CodeBERT, що спеціалізуються на семантичному моделюванні та контекстному аналізі, з графовими нейронними архітектурами, зокрема GNN (*Graph Neural Networks*) або їх рекурентними варіаціями GGNN (*Gated Graph Neural Networks*), застосування яких є актуальним для моделювання реляційних залежностей та ієрархічних структур [4]. Таке поєднання дає змогу моделям фіксувати неструктуровані дані з огляду не тільки на їх мовний зміст, але й логіко-топологічну організацію.

Як показовий приклад такого переходу варто згадати архітектуру Vul-LMGNN, що об'єднує структуру CPG з векторними видами подання CodeBERT і шаром GGNN. CodeBERT застосовується для генерації семантично значущими векторними видами подання токенів [5], що відіграють роль вихідних ознак для вузлів графа [6]. Компонент GGNN необхідний для поширення й агрегації інформації по ребрах графа за допомогою механізмів рекурентного обміну сигналами, що дає змогу моделі брати до уваги й вивчати залежності між структурними й контекстними елементами коду.

Емпіричне оцінювання Vul-LMGNN з використанням різних еталонних наборів даних, наприклад DiverseVul [7], Devign [8] та Draper VDSIC (*Vulnerability Detection in Source Code*) [9], показало вищу ефективність моделі за показниками точності й міри F1 порівняно з ізольованими методами, основаними тільки на трансформерних моделях і тільки на графових мережах. Отже, на наборі DiverseVul виявлено показники точності – 93,06% та міри F1 – 23,54%, що вказує на перевершення запропонованою моделлю GraphCodeBERT приблизно на 2 відсоткових пункти. Подібні покращення було зафіксовано й унаслідок виявлення міжмовних проблемних аспектів і невідповідностей у коді, що

підтвердило ефективність суміщення лінгвістичних і структурних видів подання.

Практичне застосування гібридної архітектури також продемонстровано в задачах автоматизованого оцінювання в алгоритмі ECHO (*Efficient Critique Harvesting and Ordering*) [10]. Розроблений для автоматизованого оцінювання освітніх середовищ програмування, цей алгоритм застосовує подання у вигляді синтаксичних дерев для прогнозування коментарів і автоматичної генерації зворотного зв'язку щодо студентського коду. Система передбачає побудову AST для кожної відправленої програми, виокремлення повторних структурних шаблонів і зіставлення їх з базою попередньо анотованих фрагментів коду. Завдяки виявленню структурних і семантичних подібностей ECHO визначає, які зауваги чи виправлення є найбільш релевантними для конкретного варіанта розв'язання задачі.

Такий підхід за своєю сутністю узгоджується з логікою автоматичної валідації в інформаційних системах. Ключовою перевагою моделей типу ECHO є гібридний механізм перетворення, який поєднує структурний аналіз із навченими семантичними асоціаціями. За допомогою алгоритмів аналізу дерев, зокрема *TreeminerD*, система виявляє часті підшаблони в структурі AST, обчислює зважені міри подібності та класифікує можливі варіанти фідбеку залежно від їх контекстної релевантності. Це забезпечує узгоджене відтворення структурних і семантичних властивостей перевірених артефактів [11], що дає змогу автоматизованій системі формувати релевантні та відтворювані оцінки валідаційних результатів без залучення експертної інтерпретації.

Поруч з архітектурами Vul-LMGNN та ECHO важливе значення має модель C-PCFG (*Compound Probabilistic Context-Free Grammar*) [12], що зосереджується на індукції граматики з імовірнісно-генеративного погляду. На відміну від традиційних імовірнісних контекстно-вільних граматики (PCFG), у цій моделі впроваджується прихована безперервна змінна z , що моделює ймовірнісні правила на рівні речення. Це дає змогу в граматиках брати до уваги контекстно-залежні синтаксичні варіації, поєднуючи ригідні символні граматики з нейромережевими контекстними видами подання. З практичного погляду кожне речення моделюється як згенероване з розподілу PCFG, параметризованого змінною z , що забезпечує врахування віддалених синтаксичних

залежностей за умови збереження зрозумілої деревоподібної структури генерації.

Навчання цієї моделі реалізується за допомогою методу згорнутого варіаційного виведення, за якого приховані синтаксичні дерева маргіналізуються з використанням динамічного програмування. Модель визначається високою результативністю в задачах неконтрольованого синтаксичного розбору на англійському корпусі PTB (*Penn Treebank*) і китайському корпусі CTB (*Chinese Treebank*), перевершуючи більш старі нейронні моделі граматичної індукції, як-от PRPN (*Parsing-Reading-Predict Network*) та ON (*Ordered Neurons*).

3. Мета дослідження

Мета роботи полягає в аналізі наявних гібридних моделей штучного інтелекту, оцінюванні їх потенціалу для автоматизації процесу перевірки надійності результатів валідації інформаційних систем, а також у розробленні прототипу нової гібридної моделі, спеціально орієнтованої на цю предметну галузь.

4. Викладення основного матеріалу

Запропонована гібридна модель штучного інтелекту, зображена на рис. 1, здатна автономно оцінювати достовірність результатів валідації в ІС.

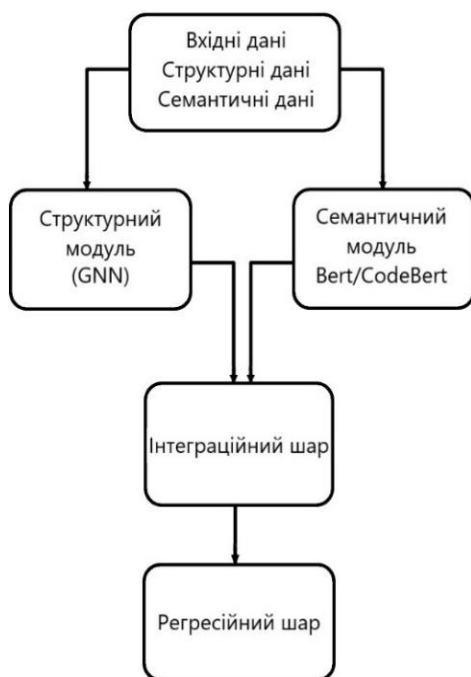


Рис. 1. Архітектура гібридної моделі оцінювання достовірності результатів валідації

Модель ґрунтується на об'єднанні семантичних і структурних аспектів аналізу даних, що дає змогу виявляти невідповідності, логічні відхилення [13] й аномальні результати перевірки [14].

Формально поставлена задача виражається таким чином. Припускається наявність множини результатів валідації:

$$R = \{r_1, r_2, \dots, r_n\}. \quad (1)$$

Кожен елемент $r_i, i = \overline{1, N}$ є структурним об'єктом, що містить два компоненти:

$$r_i = (S_i, T_i), \quad (2)$$

де S_i – множина структурних властивостей результату (операції, залежності, перевірки); T_i – множина семантичних властивостей результату (вимоги, правила, пояснення).

Для кожного результату r_i формується вектор ознак $A_i = q(r_i)$, де $q(\cdot)$ поєднує структурні й семантичні види подання в єдиний вектор фіксованої розмірності.

Задача зводиться до побудови оцінної функції:

$$f_\theta(A_i) \rightarrow y_i, \quad (3)$$

де $y_i \in [0; 1]$ відповідає ступеню достовірності певного результату валідації, а θ – набір параметрів моделі, що підлягають оптимізації під час навчання.

Унаслідок комбінованого навчання структурний і семантичний компоненти формують єдине подання, що відтворює як логічну структуру, так і контекстне значення процесів валідації.

Структурний компонент моделі, що реалізує частину функції $f_\theta(A_i)$, базується на побудові структурного подання валідаційного середовища інформаційної системи, описаний у вигляді орієнтованого графа:

$$G = (V, E); \quad (4)$$

$$V = \{v_1, v_2, \dots, v_m\}; \quad (5)$$

$$E = \{e_{kj} = (V_k, V_j) \mid V_k, V_j \in V\}, \quad (6)$$

де V – множина вершин, що відповідають окремим компонентам, залученим у процес валідації; E – множина орієнтованих ребер, що кодують логічні та інформаційні залежності між ними. Кожна вершина V_k відповідає певному елементу перевірки, тобто джерелу даних, правилу перетворення, обмеженню цілісності, алгоритмічному модулю чи

проміжному обчислювальному результату, а кожне спрямоване ребро e_{kj} задає причинно-наслідковий або інформаційний зв'язок між компонентами.

Отриманий граф відтворює багаторівневу логіку залежностей, що лежить в основі процесу валідації. Якщо обчислювальний вузол залежить як від попереднього етапу агрегації, так і від зовнішнього вхідного параметра, графова структура явно кодує ці зв'язки, що дає змогу моделі відстежувати, яким чином локальні невідповідності поширюються системою.

Побудований граф інтегрується у високорозмірний векторний простір із застосуванням GNN. Подання кожного вузла ітеративно уточнюється завдяки механізму передачі повідомлень (*message passing*), що допомагає мережі інтегрувати інформацію не тільки від безпосередніх сусідів, але й з огляду на більш широкий контекст валідації.

Як наслідок, топологічна структура процесу набуває вигляду числового векторного подання, зберігаючи як реляційні залежності, так і ієрархічну значущість елементів системи.

Прихований стан вузлів оновлюється на кожному етапі поширення повідомлень відповідно до правил:

$$h_v^{(t)} = \sigma \left(W_m AGG \left(\left\{ h_u^{(t-1)} \mid u \in N(v) \right\} \right) + W_s h_v^{(t-1)} \right), \quad (7)$$

де $h_v^{(t)}$ – приховане подання вузла V_k на ітерації t ; $N(V)$ позначає множину сусідніх вузлів, що впливають на V_k ; $AGG(\cdot)$ відповідає функції агрегації (усереднення, сума або *attention-pooling*), що об'єднує інформацію із сусідніх вузлів.

Типи ребер у графі можуть бути анотованими семантичними мітками, що дає змогу GNN розрізняти характер зв'язків у процесі поширення повідомлень [15], зокрема виокремлювати керувальні, інформаційні та причинно-наслідкові залежності, а також адаптивно зважувати їх вплив на оновлення прихованих станів вузлів під час валідації [16]. Крім того, можуть застосовуватися механізми уваги, що визначають значущість вхідних повідомлень. Це допомагає моделі виокремлювати значущі залежності, які найбільш сильно корелюють зі змінами результатів валідації.

Результатом цього етапу є матриця векторних видів подання вузлів:

$$H_G = [h_1^{cmp}, h_2^{cmp}, \dots, h_n^{cmp}], \quad (8)$$

де кожний вектор h_i^{cmp} відтворює структурну значущість відповідних компонентів валідаційної

мережі. Ці векторні види подання утворюють структурний шар моделі та відповідають GNN на рис. 1.

Отже, важливим етапом запропонованої архітектури є семантична інтерпретація текстової інформації. Для реалізації цього механізму в архітектурі застосовується попередньо навчений кодувальник на основі трансформерної моделі (BERT або CodeBERT), що використовує механізм самоуваги (*self-attention*), який дає змогу фіксувати віддалені залежності між словами й фразами, забезпечуючи розуміння як синтаксичних, так і семантичних відношень у межах текстової послідовності [17].

Кодувальник отримує текстову інформацію, розбиту на токени:

$$X^t = \{x_1, x_2, \dots, x_L\}, \quad (9)$$

де кожний токен $x_t, t = \overline{1, L}$ – мовна одиниця, тобто або слово (лексема), або знак пунктуації.

Проходячи крізь декілька шарів трансформера, модель обчислює контекстні векторні види подання, що відтворюють значення кожного токена не ізольовано, а залежно від усієї решти токенів у послідовності.

Результатом є компактне векторне подання:

$$f_{BERT}(X^t) = [h_1^{cem}, h_2^{cem}, \dots, h_L^{cem}], \quad (10)$$

в якому кожний вектор $h_t^{cem}, t = \overline{1, L}$ кодує контекстне значення відповідного токена в межах усієї семантичної структури опису процесу валідації.

Отже, кодувальник здатний розпізнавати неявну логічну структуру, закладену в інформацію, створену людиною. Трансформер фіксує ці залежності за допомогою ваг уваги, виокремлюючи ключові терміни та їх рольові зв'язки в межах речення. Семантичний кодувальник формує семантичну частину відтворення $f_\theta(A_i)$ та відповідає однойменному модулю на рис. 1.

Відтворюючи подібні елементи в єдиний простір векторних видів подання, модель формує семантичний простір, у якому схожі сценарії валідації розташовуються близько один до одного, тоді як аномальні або суперечливі описи займають більш віддалені області. У такий спосіб забезпечується більш "осмислене" розрізнення системою ступенів смислової подібності та відмінності між контекстами.

Варто зазначити, що попередньо навчений кодувальник також може бути донавченим на спеціалізованих корпусах даних, що містять історичні

звіти щодо валідації, зведення помилок і аудиторські записи. Це підвищує мовну чутливість моделі до термінології та властивих формулювань, що застосовуються у сфері валідації інформаційних систем.

Наступний етап, який відповідає шару інтеграції модальностей на рис. 1, полягає в інтеграції структурної та семантичної модальностей в єдиний прихований простір. Це дає змогу системі аналізувати дані на рівні логіки залежностей і контекстного значення.

Ця інтеграція реалізується способом проектування вихідних видів подання двох компонентів у спільний простір векторних типів подання та їх об'єднання за допомогою зваженого механізму агрегації.

Позначимо матрицю структурних видів подання вузлів графа, отриману після етапу GNN, як X_g , тоді формально це буде виражено як:

$$H = \alpha f_{BERT}(X_t) + (1 - \alpha) f_{GNN}(X_g), \quad (11)$$

де $f_{BERT}(X_t)$ – семантичне подання, отримане мовним кодувальником; $f_{GNN}(X_g)$ – структурне подання, сформоване на базі графа залежностей. Гіперпараметр $\alpha \in [0; 1]$ є регульованим коефіцієнтом рівноваги, що визначає вплив кожної модальності на об'єднане гібридне подання H . Більше значення параметра α пов'язане з підвищенням впливу семантичної інтерпретації, а менше – навпаки, з підвищенням впливу структурного подання.

Перед етапом об'єднання обидва типи векторного подання проходять нормалізацію розмірності для забезпечення числової сумісності. Це реалізується через лінійний шар проєкції:

$$\begin{aligned} z_t &= W_t f_{BERT}(X_t) + b_t \\ z_g &= W_g f_{GNN}(X_g) + b_g \end{aligned} \quad (12)$$

Завершальний етап роботи гібридної архітектури, регресійний шар (рис. 1), полягає в кількісному оцінюванні достовірності валідації, під час якого об'єднане приховане подання перетворюється в імовірнісний показник достовірності.

Це перетворення здійснюється за допомогою регресійного шару повнозв'язним нейронним блоком, який навчений виявляти взаємозв'язок між гібридними ознаками й еталонними мітками валідації.

Математичний опис має такий вигляд:

$$y_i = \sigma(WH + b), \quad (13)$$

де W і b – навчені параметри регресійного шару; $\sigma(\cdot)$ – сигмоїдна функція активації, що обмежує вихідне значення інтервалом $[0; 1]$. Отримане скалярне значення y_i інтерпретується як оцінка достовірності.

Отже, замість категоричного розділення результатів валідації на правильні й неправильні, система формує градуйовані оцінки достовірності, що дає змогу ранжувати результати за ступенем надійності та виокремлювати для мануальної перевірки лише ті з них, що мають низький рівень достовірності.

Відтворення за сигмоїдною функцією також спрощує калібрування вихідних значень із використанням методів, наприклад масштабування Платта (*Platt scaling*) або ізотонічної регресії [18].

На цьому етапі навчання застосовується функція втрат регресійного типу, оскільки оцінка $y_i \in [0; 1]$ відтворює ступінь достовірності. Базовою функцією втрат є середньоквадратична помилка (*Mean Squared Error* – MSE) з регуляризациєю за нормою L2:

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 + \lambda \|\theta\|_2^2, \quad (14)$$

де y_i – істинна оцінка достовірності для i -го екземпляра; $\hat{y}_i = \sigma(WH_i + b)$ – передбачене значення моделі; N – загальна кількість навчальних екземплярів.

Для оптимізації параметрів моделі застосовується алгоритм Adam (*Adaptive Moment Estimation*), який поєднує переваги методів *Momentum* і *RMSProp* (*Root Mean Squared Propagation* – стохастичний градієнтний спуск), забезпечуючи адаптивне оновлення швидкості навчання для кожного параметра. Зважаючи на гібридність архітектури, де параметри різних модулів можуть мати неоднакові масштаби градієнтів, застосування цього методу є особливо доцільним.

5. Результати дослідження

У межах експериментальної частини проведено комплексне оброблення даних, спрямоване на подальшу емпіричну перевірку ефективності запропонованої гібридної моделі. Основною експериментальною інформаційною базою застосовувалися масиви часових міток вузлів і ребер з фінансового графового набору даних DgraphFin V2 [19], що забезпечують структурно-часове подання валідаційних залежностей.

Аналіз вмісту показав, що масив часових міток вузлів містить приблизно 3,7 млн записів з діапазоном значень від 793 до 2147483648 із середнім значенням 2138483546 за умови середнього відхилення 138732009. Мінімальне 32-бітне значення цілого числа вказує на наявність пошкоджених або фіктивних часових міток, що за замовчуванням застосовуються як тимчасові заповнювачі (*placeholder value*). Це означає, що частина вузлів не має ініціалізованих часових властивостей або містить неповну інформацію про часову активність.

На противагу цьому, масив часових міток ребер (приблизно 4,3 млн записів) визначається послідовністю додатних цілих значень у діапазоні 0–790 із стандартним відхиленням 221.32.

Така структура свідчить про те, що показники ребер кодують дискретні часові зрізи, які відповідають послідовним інтервалам валідації або змінам стану системи в реальному часі. Перед побудовою графів проведено очищення та імпутацію часових міток вузлів, тобто всі значення, які є меншими ніж нуль або дорівнюють нулю, зокрема заповнювачі, були замасковані та заміщені середнім значенням часових міток ребер, з'єднаних з відповідними вузлами. Така процедура дала змогу усунути розриви в часовій послідовності та правильно застосувати часову схему навчання GNN.

Паралельно виконувалась інтеграція семантичних атрибутів. Кожному вузлу й ребру відповідали текстові описи й події, вилучені із супутніх файлів DGraph (журнали, звіти, контекстні повідомлення). Структурні ембединги h_i^{cmp} навчалися на основі поширення повідомлень між вузлами з огляду на тимчасову динаміку ребер.

Об'єднане подання кожного валідаційного елемента мало такий вигляд:

$$x_i = [h_i^{sem} | h_i^{cmp}]. \quad (15)$$

Для унеможливлення витоку інформації у часі дані були поділені у відношенні 70/15/15 зі збереженням хронологічного порядку:

- інтервали $t \in [0; 550]$ – навчання;
- $t \in [551; 670]$ – валідація;
- $t \in [671; 790]$ – тестування.

Оптимізація за допомогою Adam-оптимізатора проводилася з такими параметрами: швидкість навчання 10^{-4} , $\beta_1 = 0.9$, $\beta_2 = 0.999$; параметр зниження ваг 10^{-5} .

Для окремих модулів моделі задавалися різні темпи навчання:

- трансформерний кодувальник – 10^{-5} ;
- графовий кодувальник – 5×10^{-5} ;
- шар злиття й регресійна головка – 10^{-4} .

Використано ранню зупинку (10 епох без покращень) та зниження швидкості навчання за умови стагнації метрики понад п'ять епох. Для регуляризації параметр вимкнення встановлено на рівні 0.3, а обмеження градієнта здійснювалося за нормою 1.0. Результати тестування подано на рис. 2.

Проводилося порівняння з базовими моделями – семантичною, в якій використовувалася лише трансформерна модель, структурною, тобто лише GNN, та класичними статистичними (логістична регресія, метод випадкового лісу).

Додатково виконано абляційні експерименти зі зміною ваги $\alpha \in \{0.2; 0.5; 0.8\}$, що регулює внесок семантичної та структурної модальностей в об'єднане подання.

Візуальний аналіз передбачав гістограми розподілу активності ребер за часовими зрізами (0–790), оцінку частки валідних часових міток вузлів до і після імпутації, динаміку функції втрат і теплові карти уваги для обох модулів, що показують, які ознаки та зв'язки найбільше впливали на підсумкову оцінку достовірності.

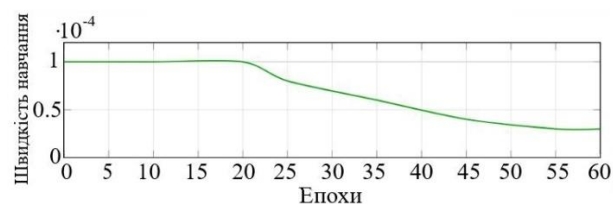


Рис. 2. Динаміка зниження швидкості навчання під час тренування гібридної моделі

Порівняльний аналіз подано на рис. 3.

На рис. 3 видно, що крива швидкості навчання демонструє стабільну збіжність моделі без ознак коливань або деградації.

На початкових етапах (до ~20 епох) зберігалася фіксована швидкість 1×10^{-4} , що забезпечувало ефективну адаптацію ваг регресійного шару й шару злиття в умовах стабільної динаміки втрат.

Починаючи з 23–25 епох, після періоду стагнації валідаційної метрики, було автоматично активовано механізм зниження швидкості навчання.

Подальше тренування супроводжувалося плавним спадом коефіцієнта навчання до рівня 5×10^{-5}

на 40-й епоці та приблизно 3×10^{-5} на фінальному етапі (55–60 епох).

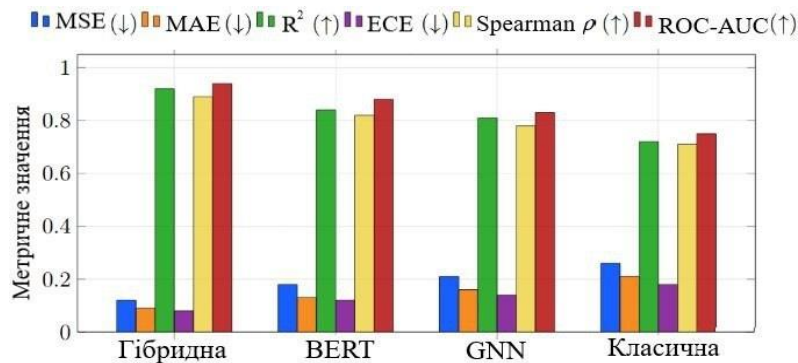


Рис. 3. Порівняння продуктивності гібридної, семантичної, структурної та класичних моделей

Такий профіль відповідає очікуваній поведінці адаптивного оптимізатора Adam із вбудованим контролем швидкості навчання, що дає змогу уникати локальних мінімумів на пізніх стадіях оптимізації.

Зниження швидкості збігалося з фазою стабілізації втрат на валідаційній вибірці, що вказує на коректну роботу механізму ранньої зупинки й ефективне використання регуляризації. У сукупності це забезпечило плавну збіжність моделі й стабільне узгодження між навчальною та валідаційною вибірками без перенавчання.

Отримані показники на рис. 3 свідчать про явну перевагу гібридної моделі над усіма базовими варіантами. Для неї зафіксовано мінімальні значення MSE, MAE (*Mean Absolute Error* – середня абсолютна похибка) та ECE (*Expected Calibration Error*), що вказує на найкращу узгодженість між передбаченими й фактичними оцінками достовірності.

Одночасно спостерігаються найвищі значення коефіцієнтів детермінації R^2 , кореляції Спірмена ρ (*Spearman's Rank Correlation Coefficient*) та ROC-AUC (*Receiver Operating Characteristic Area Under the Curve*), що демонструє здатність моделі коректно відтворювати рангові та ймовірнісні залежності між результатами валідації.

Семантична модель на основі трансформера показала високу кореляцію між оцінками ($\rho \approx 0.88$) та прийнятний ROC-AUC, однак поступалася гібридній моделі за точністю регресійних метрик. Структурна модель продемонструвала сильну стабільність у структурних залежностях ($R^2 \approx 0.85$), але її похибка (MAE, MSE) залишалася вищою через обмежене відтворення семантичного контексту.

Класичні статистичні підходи, як очікувалося, мали найнижчу точність і погану каліброваність, що видно через високу ECE. Це підтверджує їх непридатність для складних валідаційних залежностей.

Отже, для запропонованої гібридної моделі зафіксовано стабільну збіжність із керованим профілем швидкості навчання. Після етапу фіксованої швидкості активується механізм її автоматичного зниження, що супроводжується стабілізацією втрат і відсутністю перенавчання.

На тлі базових підходів гібридна архітектура забезпечує найменші значення MSE та MAE, вищу якість калібрування й одночасно демонструє найкращі показники детермінації, кореляції рангових порядків і класифікаційної здатності за площею під ROC-AUC.

6. Висновки

й перспективи подальших досліджень

У процесі роботи послідовно реалізовано етапи побудови, навчання й оцінювання гібридної моделі штучного інтелекту, призначеної для автономного визначення достовірності результатів валідації в інформаційних системах.

На першому етапі сформульовано математичну постановку задачі, розроблено структурне й семантичне подання даних, визначено механізм об'єднання модальностей.

На другому – реалізовано архітектуру, що поєднує графову нейронну мережу для аналізу топології процесів валідації та трансформерний кодувальник з метою оброблення текстових контекстів. Їх інтеграція в спільному векторному просторі забезпечила комплексне подання залежностей і контекстів перевірок.

У реалізованій системі використано модулі попередньої нормалізації ознак, злиття прихованих станів за зваженим коефіцієнтом α , регресійний шар із сигмоїдною активацією, а також оптимізацію методом Adam з адаптивним регулюванням швидкості навчання. Навчання проводилося на наборі даних DgraphFin V2 із часовими й текстовими атрибутами, зокрема з етапами очищення, імпутації та нормалізації часових міток.

Експериментальні результати засвідчили стабільну збіжність навчання без перенавчання, мінімальні значення MSE, MAE та ECE, а також найвищі показники R^2 , кореляції Спірмена ρ і ROC-AUC порівняно з базовими семантичними, структурними й класичними моделями. Аналіз ваг механізму самоуваги підтвердив коректну ідентифікацію значущих залежностей між вузлами й атрибутами даних.

Отже, розроблена гібридна модель продемонструвала підвищену точність, узгодженість і надійність прогнозів, що підтверджує її ефективність у задачах автоматизованого оцінювання достовірності результатів валідації в інформаційних системах.

Подальші дослідження можуть бути спрямовані на розширення гібридної архітектури в напрямі мультимодального контекстного аналізу, доповнюючи структурні графові й семантичні текстові види подання зовнішніми операційними сигналами (наприклад, системними метриками продуктивності, логами навантаження, параметрами QoS). Інтеграція таких каналів має потенціал підвищити чутливість моделі до латентних кореляцій, що не фіксуються на рівні валідаційних протоколів, але впливають на поведінку системи в реальному часі. Крім того, перспективним є дослідження механізмів динамічного перерозподілу ваг α між семантикою та структурною залежно від контексту валідації,

що може підвищити гнучкість моделі й адаптивність під різні режими даних.

Наявна можливість інтеграції запропонованого підходу в оптимізацію регресійних наборів [20] унаслідок використання гібридної структурно-семантичної моделі, здатної формувати узгоджені векторні види подання взаємозалежностей і контекстів тестових сценаріїв, дає змогу підвищити точність аналізу фактичних способів виконання, автоматично виявляти аномальні або непокриті сценарії, а також більш надійно оцінювати важливість і пріоритет тестів у процесі оптимізації регресійних наборів.

Конфлікт інтересів

Автори статті декларують, що не мають конфлікту інтересів, зокрема фінансового, особистого, авторського чи будь-якого іншого характеру, який міг би вплинути на дослідження, а також на результати, опубліковані в цій роботі.

Фінансування

Дослідження проводилося без фінансової підтримки.

Доступність даних

Рукопис не має пов'язаних матеріалів у сховищі даних.

Використання засобів штучного інтелекту

Автори підтверджують, що не застосовували технології штучного інтелекту для написання цієї роботи.

References

1. Wang, K., Yan, M., Zhang, H., Hu, H. (2022), "Unified abstract syntax tree representation learning for cross-language program classification", *arXiv preprint arXiv:2205.00424*, Ithaca, Cornell University. DOI: <https://doi.org/10.48550/arXiv.2205.00424>
2. Homerding, B., Patel, A., Deiana, E. A., Su, Y., Tan, Z., Xu, Z., Godala, B. R., August, D. I., Campanoni, S. (2024), "The parallel semantics program dependence graph", *arXiv preprint arXiv:2402.00986*, Ithaca, Cornell University. DOI: <https://doi.org/10.48550/arXiv.2402.00986>
3. Kuchler, A., Banse, C. (2022), "Representing LLVM-IR in a code property graph", *arXiv preprint arXiv:2211.05627*, Ithaca, Cornell University. DOI: <https://doi.org/10.48550/arXiv.2211.05627>

4. Le Anh Quan, V., Thuan Phat, C., Van Nguyen, K., The Duy, P., Pham, V.-H. (2023), "XGV-BERT: Leveraging contextualized language model and graph neural network for efficient software vulnerability detection", *arXiv preprint arXiv:2309.14677*. DOI: <https://doi.org/10.48550/arXiv.2309.14677>
5. Liu, R., Wang, Y., Xu, H., Sun, J., Zhang, F., Li, P., Guo, Z. (2024), "Vul-LMGNNs: Fusing language models and online-distilled graph neural networks for code vulnerability detection", *Information Fusion*, Vol. 115, Article 102748. DOI: <https://doi.org/10.1016/j.inffus.2024.102748>
6. Sun, Y., Yan, J., Lu, L., Zhang, H., Shang, Y. (2025), "Semantic fusion-oriented bi-typed multi-relational heterogeneous graph neural network", *IEEE Access*, Vol. 13, P. 64584–64593. DOI: <https://doi.org/10.1109/ACCESS.2025.3559720>
7. Chen, Y., Ding, Z., Alowain, L., Chen, X., Wagner, D. (2023), "DiverseVul: A new vulnerable source code dataset for deep learning-based vulnerability detection", *Proceedings of the 26th International Symposium on Research in Attacks, Intrusions and Defenses (RAID '23)*, P. 654–668, New York, Association for Computing Machinery. DOI: <https://doi.org/10.1145/3607199.3607242>
8. Zhou, Y., Liu, S., Siow, J., Du, X., Liu, Y. (2019), "Devign: Effective vulnerability identification by learning comprehensive program semantics via graph neural networks", *arXiv preprint arXiv:1909.03496*, Ithaca, Cornell University. DOI: <https://doi.org/10.48550/arXiv.1909.03496>
9. Kim, L., Russell, R. (2018), "Draper VDISC Dataset – Vulnerability Detection in Source Code", *OSF (Open Science Framework)*, Charlottesville, Center for Open Science. Available at: <https://osf.io/d45bw>
10. Van Petegem, C., Demeyere, K., Maertens, R., Strijbol, N., De Wever, B., Mesuere, B., Dawyndt, P. (2024), "Mining patterns in syntax trees to automate code reviews of student solutions for programming exercises", *arXiv preprint arXiv:2405.01579*, Ithaca, Cornell University. DOI: <https://doi.org/10.48550/arXiv.2405.01579>
11. Guo, X., Zhang, R., Zhou, L., Lu, X. (2022), "Precise code clone detection with architecture of abstract syntax trees", *Lecture Notes in Computer Science*. DOI: https://doi.org/10.1007/978-3-031-19211-1_10
12. Kim, Y., Dyer, C., Rush, A. M. (2019), "Compound probabilistic context-free grammars for grammar induction", *arXiv preprint arXiv:1906.10225*, Ithaca, Cornell University. DOI: <https://doi.org/10.48550/arXiv.1906.10225>
13. Li, J., Pang, G., Chen, L., Namazi-Rad, M.-R. (2023), "HRGCN: Heterogeneous graph-level anomaly detection with hierarchical relation-augmented graph neural networks", *arXiv preprint arXiv:2308.14340*, Ithaca, Cornell University. DOI: <https://doi.org/10.48550/arXiv.2308.14340>
14. Nelipa, O., Kalyta, N. (2025), "Machine learning-driven anomaly detection in information system verification and validation", *Grail of Science*, No. 51, P. 560–568. DOI: <https://doi.org/10.36074/grail-of-science.18.04.2025.077>
15. Yang, T., Hu, L., Shi, C., Ji, H., Li, X., Nie, L. (2021), "HGAT: Heterogeneous graph attention networks for semi-supervised short text classification", *ACM Transactions on Information Systems*, Vol. 39, No. 3, Article 32, New York, Association for Computing Machinery. DOI: <https://doi.org/10.1145/3450352>
16. Santos, N., Minatel, D., Valejo, A. D., Lopes, A. (2024), "Semi-supervised coarsening of bipartite graphs for text classification via graph neural network", *Proceedings of DSAA 2024*, P. 1–10. DOI: <https://doi.org/10.1109/DSAA61799.2024.10722822>
17. Feng, Z., Guo, D., Tang, D., Duan, N., Feng, X., Gong, M., Shou, L., Qin, B., Liu, T., Jiang, D., Zhou, M. (2020), "CodeBERT: A pre-trained model for programming and natural languages", *arXiv preprint arXiv:2002.08155*, Ithaca, Cornell University. DOI: <https://doi.org/10.48550/arXiv.2002.08155>
18. Guo, C., Pleiss, G., Sun, Y., Weinberger, K. Q. (2017), "On calibration of modern neural networks", *arXiv preprint arXiv:1706.04599*, Ithaca, Cornell University. DOI: <https://doi.org/10.48550/arXiv.1706.04599>
19. Huang, X., Yang, Y., Wang, Y., Wang, C., Zhang, Z., Xu, J., Chen, L., Vazirgiannis, M. (2022), "DGraph: A large-scale financial dataset for graph anomaly detection", *arXiv preprint arXiv:2207.03579*. DOI: <https://doi.org/10.48550/arXiv.2207.03579>
20. Nelipa, O., Kalyta, N., Ponomarova, S. (2025), "A machine learning approach to regression suite optimization for verification and validation of information systems", *Proceedings of the 15th International Conference on Advanced Computer Information Technologies (ACIT)*, Sibenik, Croatia, P. 719–722. DOI: <https://doi.org/10.1109/ACIT65614.2025.11185751>

Received (Надійшла) 01.12.2025

Accepted for publication (Прийнята до друку) 26.12.2025

Publication date (Дата публікації) 12.03.2026

Відомості про авторів / About the Authors

Неліпа Олександр Дмитрович – Харківський національний університет радіоелектроніки, аспірант кафедри системотехніки; Харків, Україна;

Oleksandr Nelipa – Kharkiv National University of Radio Electronics, Postgraduate Student at the System Engineering Department; Kharkiv, Ukraine;

e-mail: oleksandr.nelipa1@nure.ua

ORCID ID: <https://orcid.org/0000-0003-1387-1414>

Калита Надія Іванівна – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, професор кафедри системотехніки; Харків, Україна;

Nadiia Kalyta – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Professor at the System Engineering Department; Kharkiv, Ukraine;

e-mail: nadiia.kalyta@nure.ua

ORCID Author ID: <https://orcid.org/0000-0001-6181-732X>

HYBRID MODELS OF ARTIFICIAL INTELLIGENCE IN AUTOMATED ASSESSMENT TASKS FOR THE RELIABILITY OF INFORMATION SYSTEM VALIDATION RESULTS

The subject of the study is the automated assessment of the reliability of validation results in information systems, taking into account the structural dependencies between components and the semantic content of validation artifacts. The purpose of the work is to improve the reliability and reproducibility of validation by developing a hybrid artificial intelligence model that integrates graph and semantic data representations for regression assessment of reliability. Research objectives: analysis of structural and structural-semantic models (AST, PDG, CPG, Vul-LMGNN, ECHO, C-PCFG) in the context of information system validation; formulation of a mathematical formulation of the reliability assessment problem; development of a hybrid model architecture based on GNN and a transformer encoder; implementation of a mechanism for integrating structural and semantic modalities in a common vector space; experimental research on the DgraphFin V2 graph financial dataset and comparison with baseline models. Research methods include graph neural networks, BERT/CodeBERT-type transformer encoders, regression modeling with sigmoid activation, Adam optimization, and statistical quality assessment and calibration (MSE, MAE, ECE, R^2 , Spearman ρ , ROC-AUC). Results. The proposed hybrid architecture provides regression reliability estimation in the interval [0;1] and demonstrates stable convergence without signs of overfitting, as well as better values of all key metrics compared to purely semantic, purely structural, and classical statistical models. Analysis of the weights of the self-attention mechanism confirmed the correct identification of significant structural and semantic dependencies. Conclusion: The proposed model provides a more accurate calibrated assessment of the reliability of validation results, which makes it possible to automate the selection of unreliable results, reduce the influence of the human factor, and increase the reproducibility of validation procedures. A promising direction for further research is the development of multimodal architectures with dynamic weight redistribution between modalities.

Keywords: hybrid artificial intelligence models; information system validation; graph neural networks; transformer models; reliability assessment; DgraphFin V2; regression modeling; probability calibration.

Бібліографічні описи / Bibliographic descriptions

Неліпа О. Д., Калита Н. І. Гібридні моделі штучного інтелекту в задачах автоматизованого оцінювання достовірності результатів валідації інформаційних систем. *Автоматизовані системи управління та прилади автоматики*. 2026. № 1 (188). С. 33–43. DOI: <https://doi.org/10.30837/0135-1710.2026.188.033>

Nelipa, O., Kalyta, N. (2026), "Hybrid models of artificial intelligence in automated assessment tasks for the reliability of information system validation results", *Management Information System and Devices*, No. 1 (188), P. 33–43. DOI: <https://doi.org/10.30837/0135-1710.2026.188.033>

Volodymyr Panchenko, Heorhii Kuchuk

THEORETICAL-ALGEBRAIC BASICS OF EVOLUTIONARY TEST SYNTHESIS FOR GATEWAYS COMPONENTS OF HIGH-DENSITY IOT SYSTEMS

Relevance. The modern Internet of Things (IoT) paradigm is undergoing a fundamental transformation, shifting from simple telemetry collection networks to complex, High-Density IoT ecosystems, where Intelligent IoT Gateways play a pivotal role. The architectural heterogeneity creates a "state explosion" problem, where the space of possible configurations and failure scenarios exceeds the capabilities of traditional deterministic testing methods, which are unable to effectively detect deeply hidden vulnerabilities. **Object of research:** processes of automated synthesis of diagnostic tests and reliability verification for multi-layer heterogeneous components of intelligent IoT gateways. **Purpose of the article:** development of theoretical foundations and a mathematical model for evolutionary test synthesis for intelligent IoT gateways. **Research objectives:** formalization of the adaptation of genetic algorithms to the specifics of IoT architecture; improving the efficiency of critical defect detection. **Research methods:** apparatus of the theory of universal algebras and category theory. **Research results.** The article proposes and mathematically substantiates a generalized model of test synthesis based on the use of a pair of universal algebras describing the test scenario space and evolutionary operators. An extended conceptual apparatus is introduced and systematized, mapping population genetics terminology to the field of technical diagnostics of cyber-physical systems. The existence of a homomorphism between algebraic models of classic genetic algorithms and vulnerability search processes in IoT devices is proven. A test synthesis method based on the mathematical apparatus of modern algebra has been developed. The application of category theory to describe morphisms between gateway state spaces and evolutionary operators is substantiated, guaranteeing the correctness of test set transformations. **Conclusions.** The proposed approach allows to create a universal testing methodology that provides a significant increase in code coverage and defect detection. Scope of application of the obtained results: computer-aided design and diagnostics systems for IoT, testing platforms for cyber-physical systems, development of toolkits for QA engineers in the Embedded Systems and Edge AI fields.

Keywords: Internet of Things; intelligent gateway; genetic algorithms; universal algebra; category theory; test synthesis.

Introduction

Statement of the problem

The evolution of digital ecosystems over the past decade has been characterized by unprecedented growth in the number of connected devices that form the global infrastructure of the Internet of Things (IoT). According to analysts' forecasts, the number of active IoT connections continues to grow exponentially, transforming industry, healthcare, and urban infrastructure within the concepts of Industry 4.0 and 5.0. The central hub that provides functional compatibility, security, and manageability in these heterogeneous networks is the Intelligent IoT Gateway. Unlike traditional network bridges, modern gateways are complex computing platforms that integrate routing, protocol conversion, cryptographic protection, and edge analytics (Edge AI) functions [1].

The architectural complexity of such devices creates unique challenges for ensuring their reliability. The gateway must simultaneously handle thousands of concurrent connections through various physical

interfaces (Ethernet, Wi-Fi, BLE, Zigbee, LoRa, etc.), manage message queues (MQTT, AMQP), maintain data integrity during power outages, and withstand cyberattacks. This multidimensionality of the system state space leads to a phenomenon known as "state explosion", which makes exhaustive testing physically impossible and economically impractical. As noted in [2], traditional test generation methods focused on deterministic digital circuits are powerless against the dynamic nature of the IoT.

Research confirms that critical failures in such systems often arise from complex temporal correlations and resource contention when processing asynchronous interrupts [3], or require the reproduction of specific, deep sequences of network packets that cannot be generated by stateless fuzzing methods [4].

Therefore, there is a pressing scientific need to develop new, intelligent methods for automated test synthesis that can effectively explore the solution space and identify critical vulnerabilities within an acceptable time frame. The most promising direction is the use of evolutionary computing, in particular genetic

algorithms (GA), which demonstrate high efficiency in optimization tasks where the device under investigation is presented as a "black box". However, the direct, unadapted application of classical GAs for testing gateways faces problems of premature convergence, generation of syntactically incorrect tests, and low learning speed. Solving these problems requires the creation of a rigorous theoretical basis that will allow formalizing the process of adapting biological metaphors of evolution to engineering tasks of technical diagnostics.

Analysis of recent research and publications

The issue of testing IoT systems and the application of evolutionary methods is the subject of active research by the global scientific community. Systematic literature reviews in recent years indicate a transition from testing individual components to comprehensive end-to-end testing that covers devices, gateways, and cloud platforms.

Researchers Minani et al. [2] propose testing taxonomies that take into account the heterogeneity and distributed nature of IoT, but note the lack of formalized methods for automatically generating test data that would take into account the semantics of protocols.

Considerable attention is paid to the use of genetic algorithms for cybersecurity tasks in IoT, in particular for intrusion detection. Works [5–8] demonstrate the successful application of GA for optimizing neural network parameters (CNN, LSTM) and feature selection in attack detection systems. However, these studies focus primarily on traffic classification rather than on the generation of test scenarios capable of causing system failure, which is the goal of stress testing and fuzzing (the introduction of invalid, inappropriate, or randomly generated inputs).

The theoretical aspects of formalizing complex systems are considered through the prism of algebraic approaches and category theory. Studies [9, 10] demonstrate the effectiveness of algebraic methods for verifying arithmetic schemes and synthesizing program invariants. The application of category theory to model-based systems engineering (MBSE) and architecture modeling is discussed in [11]. The authors argue that category theory provides a powerful toolkit for describing the composability of systems, allowing the interaction of components to be modeled as morphisms. However, as noted in the work, there is a gap between abstract theory and engineering testing practice, especially in the context of evolutionary methods.

Thus, the task of creating a mathematical model that combines the heuristic power of evolutionary algorithms with the rigor of algebraic structures for the purposeful synthesis of tests for complex IoT gateways remains unresolved.

The aim of this work is to develop the theoretical foundations and mathematical model of evolutionary test synthesis for intelligent IoT gateways based on the use of universal algebra and category theory to formalize the processes of adapting genetic algorithms to the multilayer architecture of cyber-physical systems.

Main material

1. Correspondence of evolutionary theory conceptual frameworks and technical diagnostics

Genetic algorithms are a class of stochastic search and optimization methods based on the mechanisms of natural selection and genetics. For the correct application of this mathematical apparatus in the field of technical diagnostics, it is necessary to establish a clear isomorphism between biological entities and the objects of testing of intelligent gateways.

Classic GAs, which operate on fixed-length bit strings, are ineffective for IoT gateways due to their complex heterogeneous structure. The gateway operates with data at different levels of abstraction: from physical parameters (RSSI signal level, battery voltage, etc.) to high-level data structures (JSON objects in the MQTT protocol, binary CBOR format in the CoAP protocol [12]). Therefore, the concept of an "individual" in a test population must be transformed from a simple vector to a complex, hierarchical data structure that describes the scenario of interaction with the system.

Analysis of the structure of smart gateways and the specifics of their testing [1, 2] allows us to form an extended list of terminology correspondences, which is the basis for further formalization, the main concepts of which are given in Table 1.

The specificity of IoT requires the introduction of new concepts that are absent in classical digital automata theory. In particular, the concept of a "multi-protocol chromosome" describes a scenario that includes a sequence of actions in different protocols (for example, receiving data via Zigbee and sending via MQTT), while an "energy phenotype" reflects an energy consumption profile that is critical for battery-powered devices.

Table 1. Systematization of the correspondence between the concepts of population genetics and test diagnostics of IoT systems

The concept of genetics	The concept of IoT technical diagnostics	Interpretation and implementation in the context of the gateway
Allele	Permissible set of parameter values	A set of valid and invalid values for a specific gateway component
Gene	Elementary test impact parameter	Atomic configuration within a scenario
Chromosome	Parameter vector / Test set	Complete, ordered specification of a single test case
Genotype	Coded representation of the test	Data structure subject to genetic operations
Phenotype	System behavior under load	Observed gateway response to the test
Population	Set of tests	A set of test scenarios executed in parallel or sequentially at the current stage of evolution (generation)
Fitness	Test effectiveness / Coverage	Numerical metric that aggregates: code coverage percentage, number of detected failures, degree of resource load
Mutation	Parameter modification	Stochastic change of values: bit inversion, change of numerical values, substitution of data types
Crossover	Combination of scenarios	Creation of a new test by recombining parts of parent scenarios
Locus	Parameter location in the structure	Semantic binding of a gene to a specific gateway component

2. Development of evolutionary test synthesis mathematical model

To ensure the rigour of transferring evolutionary search methods to the field of diagnostics, it is proposed to use the apparatus of universal algebra. This allows us to abstract from the specific physical nature of signals and work with generalised operations on sets, which is particularly relevant for heterogeneous systems [10, 13].

Let us define an algebra as a pair (M, D) , where M is the algebra carrier (a non-empty set of elements), and D is the signature (a set of operations of a given arity). It is proposed to describe the process of evolutionary test synthesis using a system of two interrelated algebras:

1. Algebra of scenario generation (individual level)

$A_1^T(M_1^T, D_1^T)$, where $M_1^T = \{c_1, \dots, c_i, \dots, c_n\}$ – is the set of all n acceptable and unacceptable input vectors (test sets) for the gate; the elements of this set are chromosomes c_i ; $D_1^T = \{d_{1r}^{T1}, d_{2r}^{T1}, \dots, d_{kr}^{T1}\}$ – signature containing generation operations (d_{1r}^{T1}) , unary mutation operations (d_{2r}^{T1}) , binary crossover operations (d_{3r}^{T1}) and other specific operators (e.g., translocation).

2. Algebra of population evolution (population level):

$A_2^T(M_2^T, D_2^T)$, where M_2^T – the set of all possible subsets of tests (populations); $D_2^T = \{d_{1r}^{T2}, d_{2r}^{T2}, d_{3r}^{T2}\}$ – selection, reproduction, and reduction (selection of the best) operations.

Similarly, a classical genetic algorithm can be described by a pair of algebras $A_3^r(M_3^r, D_3^r)$ та $A_4^r(M_4^r, D_4^r)$.

The key idea of the proposed approach is to establish homomorphism. $f_1: A_3^r \rightarrow A_1^T$ та $f_2: A_4^r \rightarrow A_2^T$.

The presence of homomorphism guarantees that applying an abstract genetic operator (e.g., bit string crossover) to the test model corresponds to a correct operation in the real space of gateway parameters (e.g., correct merging of protocol configurations), which generates a valid test for execution. If this principle is violated (e.g., crossover creates a packet with an incorrect checksum that is rejected by the driver), then homomorphism is absent, and the efficiency of the algorithm degrades to a random search.

For a more in-depth modeling of structural relationships, we suggest using category theory [11, 14, 15]. This allows us to consider the testing process as a mapping between categories of states.

Let K_T is a category where $Ob K_T$ objects are the states of the gate during testing (memory configurations, queues, registers, etc.), and the morphisms $Mor K_T$ are transitions between states caused by test influences. Let K_{Ev} – be the category of the evolutionary process, where the objects $Ob K_{Ev}$ are test populations, and the morphisms $Mor K_{Ev}$ are evolutionary transformations. The existence of a functor $F=(K_{Ev}, K_T)$ allows us to automatically translate fitness function optimization strategies into strategies for finding paths to erroneous system states.

The conclusion about the nesting of categories is particularly important. If we consider the category of simple tests K_1^* (changing one parameter) and the category of complex stress tests K_2^* (simultaneous change of a group of parameters), then K_1^* is

a subcategory of K_2^* . This theoretically justifies the advantage of genetic algorithms operating in space K_2^* (with macro-mutation operators) over local search methods, as they are capable of detecting complex, correlated defects that are unattainable for simple morphisms.

3. Application of universal algebras and category theory for creating test synthesis methods

Based on the developed mathematical foundation, a number of modified test synthesis methods adapted to the specifics of IoT gateways are proposed.

Heuristic combination method. Given the high uncertainty of IoT environment behavior, no single heuristic (e.g., bit inversion alone) can guarantee effective vulnerability discovery. The method of combining heuristics is described by an extended signature D^E of $A^E(M^E, D^E)$ algebra, which includes a set of heterogeneous heuristics $d_1^E, d_2^E, \dots, d_n^E$ based on mutation, crossover, and other

$$D^E = \{d_1^E, d_2^E, \dots, d_n^E\}. \quad (1)$$

The algorithm adaptively selects an operator at each step based on its current effectiveness (probabilistic approach). For the specifics of gateways, it is proposed to use a specialized combined crossover operator with translocation. The principle of its operation is as follows: during crossover, one of the offspring receives part of the genes from the first parent (inheritance), and the other part – inverted or logically opposite values from the second parent (entropy introduction). When applied in the IoT field, this allows for the automatic generation of "contrast" load scenarios. For example, the system simultaneously attempts to transmit valid critical priority data via MQTT and a stream of damaged packets via the CoAP interface. This approach effectively detects errors in the prioritization and isolation mechanisms of processes in the gateway operating system.

A test synthesis method that takes into account a priori information about the structure. One of the main problems with applying classical GAs to network protocols is the destruction of data structure. Simple crossover operators break bit strings at arbitrary locations, leading to a violation of the integrity of JSON/XML structures or checksums (CRC). Such tests are rejected by the gateway's input parsers even before

the main logic is executed, which makes testing ineffective.

To solve this problem, a hierarchical chromosome structure with the introduction of control genes is proposed. The chromosome is represented as a tuple:

$$C = \langle G_{ctl}, G_{data} \rangle, \quad (2)$$

where G_{ctl} – higher-level control genes that determine testing strategy; G_{data} – information genes containing specific parameters.

A structurally adaptive crossover operator d_{StrAw}^I is also introduced, which allows breaking a chromosome exclusively at the boundaries of logical blocks (between the header and body of a packet, between configuration records, etc.). This guarantees that the test remains syntactically correct when its semantics change, ensuring that the test penetrates deeply into the logic of the device.

Evolutionary method on parallel computing structures. Testing IoT gateways is an extremely time-consuming process. A single test can take from seconds to minutes due to the need to wait for network timeouts and system stabilization. Thus, sequential execution of a population of 100 tests can take hours.

The algebraic model is extended by introducing parallel operations. Let algebra $A^II(M^II, D^II)$ be given, the operation of evaluating the fitness $\mu(x)$ of an individual x , which is the most computationally complex, is transferred to parallel streams. In the proposed method, the population P_i is divided into subsets $\{P_{i,1}, \dots, P_{i,k}\}$, which are processed by independent nodes in a cloud or fog environment.

Each node emulates a virtual instance of the gateway or interacts with a separate physical interface of the test bench. The central orchestrator performs only selection and reproduction operations. This allows for nearly linear acceleration of the test synthesis process, which is critical for ensuring continuous integration (CI/CD) in IoT development [16].

Method based on evolutionary strategies. Many parameters of gateway operation are continuous rather than discrete in nature: supply voltage level, processor temperature, time intervals for maintaining persistent connections, sensor thresholds, etc. For such parameters, the bit representation of classical GA is unnatural and inefficient. A method based on evolutionary strategies is used, where a chromosome is a vector of real numbers.

The mutation operator is implemented by adding a random variable x_i with a normal distribution to each parameter:

$$x'_i = x_i + N(0, \sigma_i), \quad (3)$$

where σ_i – the mean square deviation (mutation step), which is also part of the genotype and adapts during evolution (self-adaptation).

This allows for hardware-level stress testing by finding the limit values of parameters (for example, the minimum permissible voltage at which the gateway still functions but begins to allow errors in writing to flash memory), which is impossible to do using combinatorial logic methods [17].

Test synthesis using non-homologous chromosomes. Gateways work with variable-length data streams (message queues, fragmented IP packets). Classic crossover works with chromosomes of fixed length (homologous), which limits its applicability. The concept of non-homologous chromosomes with different lengths and structures is introduced.

In this case, it is proposed to apply a modified crossover algorithm with a correction phase:

1. Selection of breakpoints at arbitrary (even asymmetrical) locations on the parent chromosomes.
2. Exchange of sections, which leads to a change in the length of the offspring.
3. Correction phase:
 - if the length of the offspring exceeds the permissible MTU (Maximum Transmission Unit), truncation or fragmentation of the packet is performed;

– if the length is insufficient (mandatory fields are lost), neutral fillers are added. This approach allows for the evolutionary study of specific classes of vulnerabilities, such as buffer overflows and attacks on packet fragmentation mechanisms.

4. Test synthesis based on the mathematical apparatus of modern algebra

Integrating all the methods we've looked at lets us build a single, generalized test synthesis scheme for smart gateways (Fig. 1).

The process is described as an iterative procedure in the multidimensional space of gateway states, controlled by a multifactorial fitness function $\mu(x)$.

The formula for the fitness function for the IoT gateway looks like this:

$$\mu(x) = w_1 \cdot S_{cov} + w_2 \cdot E_{detect} + w_3 \cdot R_{usage} - w_4 \cdot T_{exec} \quad (4)$$

where S_{cov} – code coverage or protocol state space coverage; E_{detect} – the weight of the detected errors (depending on their criticality); R_{usage} – peak resource usage (memory, central processing unit), which stimulates the search for DoS attacks; T_{exec} – penalty for test execution time (to optimize the length of the set); $w_i, i = \overline{1,4}$ – weight coefficients that can be adjusted depending on the testing objective.

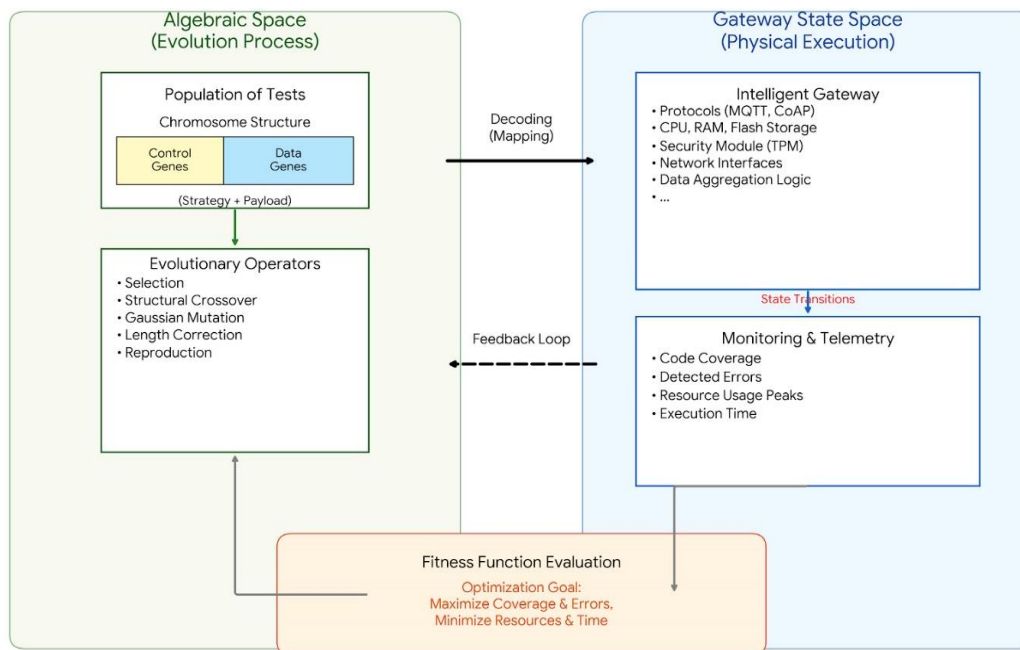


Fig. 1. Generalized scheme of test synthesis based on an algebraic approach

As a result of the analysis, the following generalized algorithm for synthesizing tests for IoT gateway components is proposed:

1. Initialization. Formation of the initial population of scenarios using templates and taking into account the structure (2).

2. Evaluation. Parallel launch of tests on digital twins or physical devices. Collection of metrics via telemetry interfaces (4).

3. Selection. Selection of tests that caused anomalies or reached deep branches of the code.

4. Recombination and mutation. Application of an adaptive set of operators (1) from algebra A^E , including combined crossover and mutations of evolutionary strategies (3) for physical parameters.

5. Correction. Restoration of the structural integrity of tests for non-homologous chromosomes.

6. Termination condition. Achievement of the target reliability level or exhaustion of the time budget.

Research results and their discussion

To experimentally verify the proposed model, a software package was developed that implements adapted genetic algorithms. Testing was performed on a model of an intelligent gateway with heterogeneous interfaces operating under realistic load conditions [18].

The effectiveness of the proposed method was evaluated in comparison with three existing approaches:

- classical random testing [19–21];
- testing based on rules created by experts (heuristic expert) [22, 23];
- methods based on state-of-the-art FSM models [4, 24, 25].

Fig. 2 shows comparative histograms of the final error coverage for five key gateway components.

Analysis of the diagrams shows a significant advantage of the proposed method for all system components.

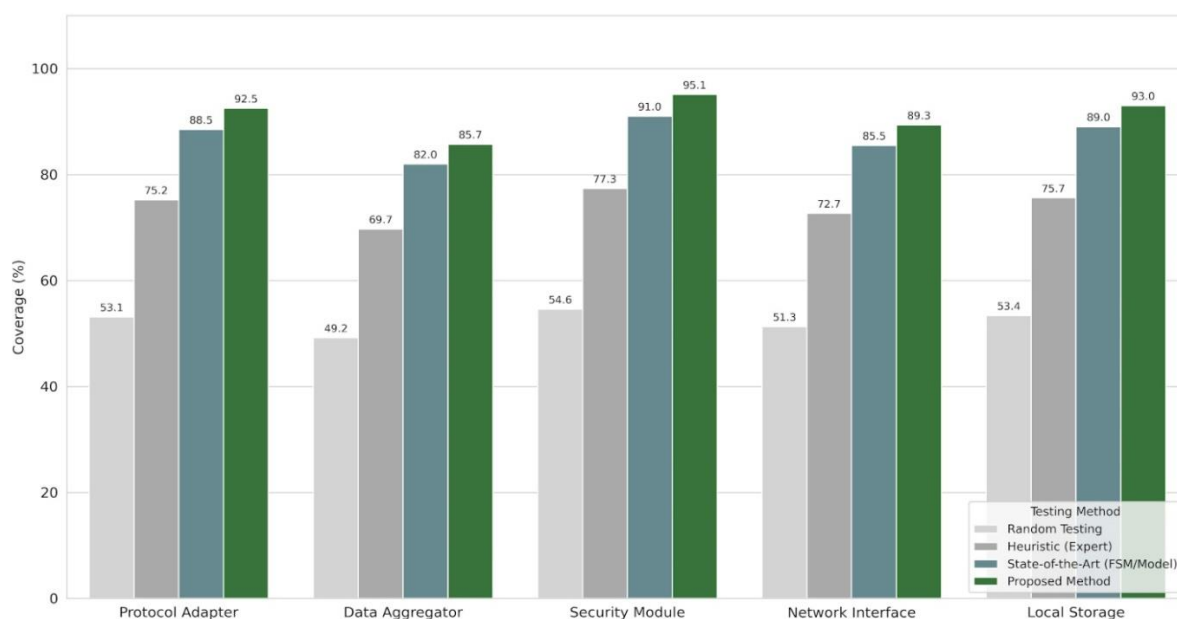


Fig. 2. Comparison of gateway component coverage

Thus, for the protocol adapter, the best coverage is 92.5%, which is explained by the use of a structurally adaptive crossover that preserves the syntactic correctness of packets. For local storage and network interface, the coverage also exceeds 89%, which indicates the successful operation of evolutionary strategies with real numbers when testing resource constraints.

An important aspect is not only the final result, but also the speed at which it is achieved. Figure 3 shows the generalized dynamics of the testing process, averaged

across the entire system. The graph shows that the proposed method is characterized by the steepest slope in the initial stages (0–40% of the process duration). This indicates the high convergence speed of the algorithm due to the use of a priori information about the gateway structure. While random testing methods show almost linear, slow growth, the proposed approach quickly reaches a coverage level of over 90%. The use of non-homologous chromosomes made it possible to avoid the population stagnation observed in the heuristic method graph.

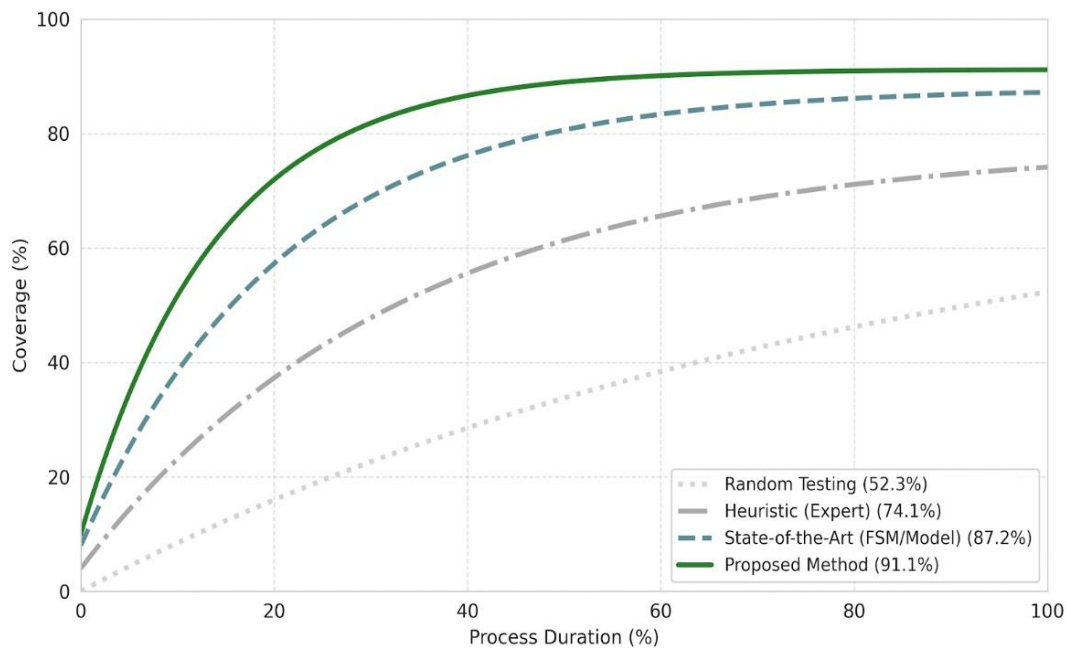


Fig. 3. Generalized dynamics of gateway component coverage

Thus, experiments confirm that algebraic formalization of the test synthesis process provides a significant increase in code coverage and defect detection compared to random search and static analysis methods.

Conclusions and prospects for further development

1. A theoretical algebraic model for synthesizing tests for intelligent IoT gateways has been developed and substantiated, based on the use of a pair of universal algebras (scenario generation and population evolution). This made it possible to formalize the process of testing complex heterogeneous systems and prove the mathematical correctness of transferring evolutionary computing methods to the field of technical diagnostics.

2. An extended classification of concepts (gene-parameter, chromosome-scenario, phenotype-behavior) has been proposed, which, unlike classical approaches, takes into account the specifics of IoT: multi-protocol, energy dependence, real-time operation, and variable data formats.

3. The effectiveness of using a combination of heuristics and combined crossover to overcome local extrema when searching for vulnerabilities has been proven. It has been shown that the introduction of structural constraints (control genes) and chromosome

length correction mechanisms significantly increases the percentage of valid tests in the population, allowing computational resources to be focused on verifying the deep logic of the gateway.

4. The practical implementation of the proposed methods on parallel computing structures solves the problem of critical time constraints when testing high-density networks, ensuring the scalability of the diagnostic process and the possibility of integration into modern development pipelines.

The theoretical and algebraic solutions obtained in this work and the proven effectiveness of evolutionary test synthesis create a fundamental basis for the transition from disparate testing tasks to the creation of a comprehensive methodology. The priority direction for further scientific research is the development of methods for component and integral diagnostics of the intelligent gateway of the peripheral layer of a multi-density IoT. The implementation of this direction involves solving two interrelated classes of problems:

1. At the level of individual gateway subsystems (components), further work will focus on deepening the specialization of evolutionary operators to take into account hardware specifics (development of methods for generating stress tests for the gateway's neural processing units (NPUs), creation of adaptive scenarios for testing the physical level in conditions of high signal interference characteristic of dense LoRaWAN, ZigBee, Wi-Fi 6, etc. networks).

2. Development of integral diagnostics methods (detection of defects that arise exclusively during the interaction of correctly functioning separate components (taking into account resource collisions, inter-level tracking of error propagation).

Thus, the proposed algebraic approach creates a reliable foundation for the implementation of new directions, since its abstract nature allows expanding the signature of operations without changing the overall architecture of the method.

Conflict of interest

The authors declare that they have no conflict of interest with respect to this research, including financial, personal, authorship, or other conflicts that

References

1. Beniwal, G., Singhrova, A. (2022), "A systematic literature review on IoT gateways", *Journal of King Saud University – Computer and Information Sciences*, No. 34(10), P. 9541–9563. DOI: <https://doi.org/10.1016/j.jksuci.2021.11.007>
2. Minani, J. B., Sabir, F., Moha, N., Guéhéneuc, Y.-G. (2024), "A Systematic Review of IoT Systems Testing: Objectives, Approaches, Tools, and Challenges", *IEEE Transactions on Software Engineering*, No. 50(4), P. 785–815. DOI: <https://doi.org/10.1109/TSE.2024.3363611>
3. Trimananda, R., Aqajari, S.A.H., Chuang, J., Demsky, B., Xu, G. H., Lu, S. (2020), "Understanding and automatically detecting conflicting interactions between smart home IoT applications", *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, P. 1215–1227. DOI: <https://doi.org/10.1145/3368089.3409682>
4. Shu, Z., Yan, G. (2022), "IoTIInfer: Automated Blackbox Fuzz Testing of IoT Network Protocols Guided by Finite State Machine Inference", *IEEE Internet of Things Journal*, No. 9(22), P. 22737–22751. DOI: <https://doi.org/10.1109/JIOT.2022.3182589>
5. Hakiki, R.I., Azerine, A., Tlemsani, R., Golabi, M., Idoumghar, L. (2025), "Enhancing IoT intrusion detection with genetic algorithm-optimized convolutional neural networks", *The Journal* DOI: <https://doi.org/10.1007/s11227-025-07626-8>
6. Jain, V., Agrawal, M. (2020), "Applying Genetic Algorithm in Intrusion Detection System of IoT Applications", *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI) (48184)*, P. 284–287. DOI: <https://doi.org/10.1109/ICOEI48184.2020.9143019>
7. Dong, J., Li, Z., Zheng, Y., Luo, J., Zhang, M., Yang, X. (2024), "Real-time fault detection for IIoT facilities using GA-Att-LSTM based on edge-cloud collaboration", *Frontiers in Neurorobotics*, No. 18, P. 1499703. DOI: <https://doi.org/10.3389/fnbot.2024.1499703>
8. Katsura, Y., Endo, A., Arai, I., Fujikawa, K. (2025), "Efficient IDS for IoT Networks Using Host-Based Data Aggregation and Multi-Entropy Analysis", *IEEE Access*, No. 13, P. 125406–125419. DOI: <https://doi.org/10.1109/ACCESS.2025.3589057>
9. Sampathkumar, B., Das, R., Martin, B., Enescu, F., Kalla, P. (2025), "An Algebraic Approach to Partial Synthesis of Arithmetic Circuits", *Proceedings of the 30th Asia and South Pacific Design Automation Conference*, P. 1097–1103. DOI: <https://doi.org/10.1145/3658617.3697724>
10. Humenberger, A., Amrollahi, D., Bjørner, N., Kovács, L. (2022), "Algebra-Based Reasoning for Loop Synthesis", *Formal Aspects of Computing*, No. 34(1), P. 1–31. DOI: <https://doi.org/10.1145/3527458>
11. Vidalie, J., Batteux, M., Mhenni, F., Choley, J.-Y. (2022), "Category Theory Framework for System Engineering and Safety Assessment Model Synchronization Methodologies", *Applied Sciences*, No. 12(12), P. 5880. DOI: <https://doi.org/10.3390/app12125880>
12. Molina Araque, S., Martinez, I., Papadopoulos, G.Z., Montavont, N., Toutain, L. (2023), "Yet Another Compact Time Series Data Representation Using CBOR Templates (YACTS)", *Sensors*, No. 23(11), P. 5124. DOI: <https://doi.org/10.3390/s23115124>

could influence the research and its results presented in this article.

Funding

The study was conducted without financial support.

Data availability

The manuscript has no associated data.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence technology in the creation of this work.

13. Tao, T. (2024), *A pilot project in universal algebra to explore new ways to collaborate and use machine assistance?* URL: <https://terrytao.wordpress.com/2024/09/25/a-pilot-project-in-universal-algebra-to-explore-new-ways-to-collaborate-and-use-machine-assistance/> (дата звернення: 21.01.2026).
14. Mabrok, M.A., Ryan, M J. (2017), "Category Theory as a Formal Mathematical Foundation for Model-Based Systems Engineering", *Applied Mathematics & Information Sciences*, No. 11(1), P. 43–51. DOI: <https://doi.org/10.18576/amis/110106>
15. Breiner, S., Subrahmanian, E., Sriram, R.D. (2023), "Category Theory". In A. M. Madni, N. Augustine, & M. Sievers (Eds.), *Handbook of Model-Based Systems Engineering*, P. 1259–1299. DOI: https://doi.org/10.1007/978-3-030-93582-5_85
16. Jamil, M.A. (2025), "Evolutionary Algorithm Behavior to Optimize the IoT Scheduling Problem", *2025 IEEE International Conference on E-Business Engineering (ICEBE)*, P. 288–291. DOI: <https://doi.org/10.1109/ICEBE68123.2025.00052>
17. Surayya, A., Muzakkir Hussain, M., Reddy, V.D., Abdul, A., Gazi, F. (2025), "Evolutionary Algorithms for Edge Server Placement in Vehicular Edge Computing", *IEEE Access*, No. 13, P. 79030–79052. DOI: <https://doi.org/10.1109/ACCESS.2025.3566172>
18. De Benedictis, A., Flammini, F., Mazzocca, N., Somma, A., Vitale, F. (2023), "Digital Twins for Anomaly Detection in the Industrial Internet of Things: Conceptual Architecture and Proof-of-Concept", *IEEE Transactions on Industrial Informatics*, No. 19(12), P. 11553–11563. DOI: <https://doi.org/10.1109/TII.2023.3246983>
19. Muench, M., Stijohann, J., Kargl, F., Francillon, A., Balzarotti, D. (2018), "What You Corrupt Is Not What You Crash: Challenges in Fuzzing Embedded Devices", *Proceedings 2018 Network and Distributed System Security Symposium*. Network and Distributed System Security Symposium, San Diego, CA. DOI: <https://doi.org/10.14722/ndss.2018.23166>
20. Al-Garadi, M.A., Mohamed, A., Al-Ali, A.K., Du, X., Ali, I., Guizani, M. (2020), "A Survey of Machine and Deep Learning Methods for Internet of Things (IoT) Security", *IEEE Communications Surveys & Tutorials*, No. 22(3), P. 1646–1685. DOI: <https://doi.org/10.1109/COMST.2020.2988293>
21. Aldysty, A.R., Moustafa, N., Lakshika, E. (2025), "A Holistic Review of Fuzzing for Vulnerability Assessment in Industrial Network Protocols". *IEEE Open Journal of the Communications Society*, No. 6, P. 4437–4461. DOI: <https://doi.org/10.1109/OJCOMS.2025.3569291>
22. Qureshi, A.-H., Larijani, H., Ahmad, J., Mtetwa, N. (2019), "A Heuristic Intrusion Detection System for Internet-of-Things (IoT)". *Intelligent Computing*, Vol. 997, P. 86–98. DOI: https://doi.org/10.1007/978-3-030-22871-2_7
23. Latha, R., Thangaraj, J.J. (2025), "IoT security using heuristic aided symmetric convolution-based deep temporal convolution network for intrusion detection by extracting multi-cascaded deep attention features", *Expert Systems with Applications*, No. 269, P. 126363. DOI: <https://doi.org/10.1016/j.eswa.2024.126363>
24. Németh, G.Á. (2025), "Model-based mutation testing for Finite State Machine specifications with MTR", *Infocommunications Journal*, No. 17(3), P. 84–91. DOI: <https://doi.org/10.36244/ICJ.2025.3.10>
25. Pan, Z., Zhang, L., Hu, Z., Li, Y., Chen, Y. (2022), "SATFuzz: A Stateful Network Protocol Fuzzing Framework from a Novel Perspective", *Applied Sciences*, No. 12(15), P. 7459. DOI: <https://doi.org/10.3390/app12157459>

Received (Надійшла) 22.01.2026

Accepted for publication (Прийнята до друку) 29.01.2026

Publication date (Дата публікації) 12.03.2026

Відомості про авторів / About the Authors

Панченко Володимир Іванович – Національний технічний університет "Харківський політехнічний інститут", старший викладач кафедри комп'ютерної інженерії та програмування; Харків, Україна;

Volodymyr Panchenko – National Technical University "Kharkiv Polytechnic Institute", Senior Lecturer at the Computer Engineering and Programming Department; Kharkiv, Ukraine;

e-mail: Volodymyr.Panchenko@khp.edu.ua

ORCID ID: <https://orcid.org/0000-0003-3364-3398>

Кучук Георгій Анатолійович – доктор технічних наук, професор, Національний технічний університет "Харківський політехнічний інститут", професор кафедри комп'ютерної інженерії та програмування; Харків, Україна;

Heorhii Kuchuk – Doctor of Technical Sciences, Professor, National Technical University "Kharkiv Polytechnic Institute", Professor at the Computer Engineering and Programming Department; Kharkiv, Ukraine;

e-mail: mkuchuk56@ukr.net

ORCID ID: <http://orcid.org/0000-0002-2862-438X>

ТЕОРЕТИКО-АЛГЕБРАЇЧНІ ОСНОВИ ЕВОЛЮЦІЙНОГО СИНТЕЗУ ТЕСТІВ КОМПОНЕНТІВ ІНТЕЛЕКТУАЛЬНИХ ШЛЮЗІВ ВИСОКОЦІЛЬНИХ ІОТ-СИСТЕМ

Актуальність. Сучасна парадигма Інтернету речей (IoT) переживає фундаментальну трансформацію, переходячи від простих мереж збору телеметрії до складних, високоцільних екосистем, де ключову роль відіграють інтелектуальні шлюзи. Архітектурна гетерогенність шлюзів породжує проблему "комбінаторного вибуху станів", коли простір можливих конфігурацій та сценаріїв відмов перевищує можливості традиційних детермінованих методів тестування, не здатних ефективно виявляти глибоко приховані вразливості. **Об'єктом дослідження** є процеси автоматизованого синтезу діагностичних тестів і верифікації надійності для багатошарових гетерогенних компонентів інтелектуальних IoT-шлюзів. **Мета статті** – розроблення теоретичних засад і математичної моделі еволюційного синтезу тестів для інтелектуальних шлюзів IoT. **Завдання дослідження:** формалізація адаптації генетичних алгоритмів до особливостей архітектури IoT; підвищення ефективності виявлення критичних дефектів. **Застосовані методи:** апарат теорії універсальних алгебр і теорії категорій. **Результати дослідження.** У роботі запропоновано й математично обґрунтовано модель синтезу тестів, яка базується на використанні пари універсальних алгебр, що описують простір тестових сценаріїв та еволюційні оператори. Упроваджено й систематизовано розширений понятійний апарат, що відтворює впровадження термінології популяційної генетики у сферу технічної діагностики кіберфізичних систем. Доведено наявність гомоморфізму між алгебраїчними моделями класичних генетичних алгоритмів і процесами пошуку вразливостей у IoT-шлюзах. Розроблено метод синтезу тестів на основі математичного апарату сучасної алгебри. Обґрунтовано застосування теорії категорій для опису морфізмів між просторами станів шлюзу й еволюційними операторами, що дає змогу гарантувати коректність перетворень тестових наборів. **Висновки.** Запропонований підхід сприяє створенню універсального підходу до тестування, що забезпечує суттєве підвищення покриття коду й виявлення дефектів. Сфера використання досягнутих результатів: системи автоматизованого проєктування й діагностики для IoT, платформи тестування кіберфізичних систем, розроблення інструментарію для QA-інженерів у сфері Embedded Systems та Edge AI.

Ключові слова: Інтернет речей; інтелектуальний шлюз; генетичні алгоритми; універсальна алгебра; теорія категорій; синтез тестів.

Бібліографічні описи / Bibliographic descriptions

Панченко В. І., Кучук Г. А. Теоретико-алгебраїчні основи еволюційного синтезу тестів компонентів інтелектуальних шлюзів високоцільних IoT-систем. *Автоматизовані системи управління та прилади автоматики*. 2026. № 1 (188). С. 44–53. DOI: <https://doi.org/10.30837/0135-1710.2026.188.044>

Panchenko, V., Kuchuk, H. (2026), "Theoretical-algebraic basics of evolutionary test synthesis for gateways components of high-density IoT systems", *Management Information System and Devices*, No. 1 (188), P. 44–53. DOI: <https://doi.org/10.30837/0135-1710.2026.188.044>

Петров К. Е., Божко О. Ю.

МЕТОД АВТОМАТИЧНОЇ РОЗМІТКИ ОЗНАК СТРУКТУРНОЇ СКЛАДНОСТІ ДОКУМЕНТІВ ІЗ ВИКОРИСТАННЯМ ВЕЛИКИХ МОВНИХ МОДЕЛЕЙ

Предметом дослідження є методи автоматизованого аналізу й визначення ознак структурної складності слабкоструктурованих ділових документів із застосуванням сучасних мультимодальних великих мовних моделей. Мета – розробити й експериментально перевірити працездатність методу автоматичної розмітки ознак структурної складності документів, який забезпечує можливість прогнозування якості подальшої екстракції даних, а також побудувати математичну модель для оптимізації вибору конфігурації мовних моделей у просторі критеріїв "якість – вартість". Завдання передбачають формалізацію таксономії ознак, формування експериментального корпусу, розроблення уніфікованої схеми промптингу й розв'язання задачі багатокритеріальної оптимізації вибору моделей. Методологічну основу дослідження становлять методи системного аналізу, емпіричне профілювання великих мовних моделей, методи інженерії підказок (prompt engineering), методи математичної статистики (кореляційний аналіз, розрахунок метрик Precision, Recall, F1) для оцінювання якості класифікації, а також методи дискретної оптимізації для знаходження компромісних рішень. Результати. Основним результатом досліджень є розроблений оригінальний метод автоматичної розмітки ознак структурної складності документів, що дає змогу автоматично генерувати структурований профіль складності документа й обробляти його оригінальне візуальне PDF-подання. Також запропоновано зважену функцію якості, яка бере до уваги ступінь впливу кожної ознаки на помилки екстракції, та виділено Парето-оптимальні конфігурації, що уможливають мінімізацію витрат з огляду на задані вимоги до надійності. На підставі досягнутих результатів можна зробити певні висновки. Експериментально доведено, що економічно ефективні моделі в режимі з використанням прикладів забезпечують високу точність розмітки, яка конкурує з результатами значно вартісних моделей. Установлено, що застосування режимів із розширеним міркуванням для задачі бінарної класифікації є економічно недоцільним через диспропорційне зростання вартості без суттєвого приросту якості. Запропонований метод розв'язує проблему відсутності інструментів попереднього оцінювання складності в системах інтелектуального опрацювання документів. Він забезпечує прогнозованість і керованість процесів екстракції, даючи змогу реалізувати адаптивну маршрутизацію документів залежно від їх складності. Це створює підґрунтя для побудови ефективних промислових систем із збалансованими показниками точності та вартості експлуатації, усуваючи необхідність у трудомісткій ручній розмітці.

Ключові слова: слабкоструктурований документ; екстракція даних; промпт-інжиніринг; Парето-оптимальність; мультимодальні моделі.

1. Вступ

Автоматизація опрацювання ділових документів є важливим напрямом розвитку сучасних інформаційних систем управління.

Попри досягнення у сфері інтелектуального аналізу текстів, процес екстракції структурованих даних із документів залишається нетривіальним завданням через різноманіття форматів, неузгодженість розмітки й наявність прихованих контекстуальних залежностей. Оцінювання структурної складності документа може слугувати основою для прогнозування якості подальшої екстракції даних і вибору відповідної стратегії їх оброблення. Традиційні підходи до визначення ознак складності передбачають експертний аналіз, що є трудомістким

процесом, має елемент суб'єктивності й погано масштабується для великих документних корпусів.

Застосування великих мовних моделей (LLM) для розв'язання цієї проблеми є перспективним напрямом досліджень і створює передумови для автоматизованого розпізнавання таких ознак без попереднього навчання на спеціалізованих даних. Проте відсутність апробованих методів автоматичної розмітки ускладнює оцінювання надійності отриманих результатів і порівняння ефективності використання різних моделей.

Тому актуальним є розроблення методу автоматичної розмітки ознак структурної складності документів із використанням мультимодальних мовних моделей. Установлення кореляційних зв'язків між автоматично виявленими ознаками складності та

фактичними показниками якості екстракції даних (зокрема мірою F1) дасть змогу оцінити практичну придатність таких моделей для попереднього прогнозування складності опрацювання документів.

2. Аналіз літературних джерел і визначення проблеми дослідження

Вивчення структурної складності документів має тривалу історію, яка охоплює період від систем на основі правил до сучасних методів глибокого навчання. У ранніх роботах [1, 2] розглянуто методи, що ґрунтуються на застосуванні експертних правил і граматичних підходів для опису логічної структури документів. Однак суттєвими недоліками таких методів є чутливість до "шуму" й складність масштабування на нові типи макетів через необхідність ручного створення наборів правил. Подальші статистичні методи [3] й класичне машинне навчання [4] дали змогу частково автоматизувати розпізнавання структурних елементів, проте й надалі потребували великих обсягів навчальних даних і ручної розмітки.

З появою спеціалізованих моделей глибокого навчання, що використовуються для аналізу документів, дослідники досягли значного покращення точності аналізу складних макетів. Прикладом таких моделей є *TableNet* [5], призначена для виявлення таблиць, або серії моделей *LayoutLM* [6], де був реалізований механізм самоуваги з огляду на просторові відношення. Ці архітектури об'єднали текстові, візуальні та просторові ознаки, що дало змогу розуміти слабкоструктуровані документи, зокрема рахунки-фактури чи форми. Проте навіть найкращі моделі залишалися залежними від великих обсягів анотованих даних, а ручна розмітка таких корпусів залишається однією з основних проблем. Процес ручної розмітки вимагає значних ресурсів, а надійність розмітки залишається низькою для складних випадків.

У сучасних студіях [7–9] автори доводять, що структурна складність документів безпосередньо впливає на продуктивність систем інформації, зокрема на показник якості екстракції даних F1-міри (F1-score). Для документів із багаторівневою структурою, неоднозначними підписами або вкладеними таблицями спостерігається зменшення значення F1-міри до 50%, якщо порівнювати з простими форматами.

Це створює об'єктивну потребу в методах, які могли б попередньо визначати рівень складності документа ще до етапу екстракції даних.

Попри значні успіхи у сфері використання засобів штучного інтелекту для роботи з документами, сучасні підходи не розв'язують проблеми автоматичного оцінювання ознак структурної складності документів. Наявні системи аналізу макетів або потребують повної розмітки, або зосереджуються на виявленні елементів макета (заголовки, таблиці, списки), але не на концептуальних властивостях, що впливають на якість розпізнавання.

Отже, навіть найсучасніші методи не забезпечують універсального засобу кількісного прогнозування складності документа.

Останні дослідження у сфері синтезу великих мовних моделей відкрили нові можливості для когнітивного аналізу документів без попереднього навчання. Так, автори праць [10, 11] засвідчили, що LLM здатні узагальнювати контекст, відстежувати структуру та ідентифікувати логічні зв'язки в документі навіть у режимі без прикладів (*zero-shot mode*). Проте питання застосування LLM для автоматичної розмітки ознак складності документів залишається практично недослідженим. Особливо це стосується бізнес-документів, де різноманітність форматів і мовних структур створює суттєві труднощі для класичних моделей.

Крім того, у науковій літературі відсутні дослідження, що оцінюють узгодженість рішень між різними мовними моделями або ефективність ансамблевих (*ensemble*) підходів до розмітки документів. Тоді як мультиагентні архітектури [12, 13] демонструють переваги в задачах логічного міркування, їх застосування до завдань оцінювання складності ділових документів ще не було предметом системного аналізу.

Можемо підсумувати, що результати попередніх досліджень дають змогу сформулювати основну проблему: попри наявність розвинених методів аналізу макетів і візуально-текстового подання документів, не існує ефективних методів автоматичного визначення ознак структурної складності, які б поєднували переваги LLM, можливість роботи без ручної анотації та відтворюваність результатів. Це створює передумови для розроблення й порівняння ефективності нових методів, зокрема ансамблевих підходів із використанням розширеного мислення (*reasoning*

models) як арбітра між рішеннями кількох моделей нижчого рівня.

3. Мета й завдання дослідження

Метою дослідження є розроблення та експериментальна перевірка методу автоматичної розмітки ознак структурної складності бізнес-документів у форматі PDF із використанням великих мовних моделей, а також побудова на цій основі формальної моделі вибору конфігурації моделей у просторі "якість – вартість".

Запропонований метод має забезпечити ймовірність отримання для кожного документа структурованого подання його складності, яке може використовуватися для прогнозування очікуваної якості екстракції слабкоструктурованих даних для адаптивного вибору моделей екстракції та для керованого компромісу між якістю і вартістю оброблення.

Для досягнення поставленої мети в роботі необхідно розв'язати такі завдання:

- уточнити й формалізувати таксономію ознак структурної складності бізнес-документів, виокремити підмножину ознак, найбільш релевантних для задачі екстракції слабкоструктурованих даних, й обґрунтувати критерії їх відбору для експериментальної перевірки методу;

- сформувати експериментальний корпус бізнес-документів у форматі PDF з ручною еталонною розміткою обраних ознак структурної складності та прив'язкою до результатів екстракції даних із попереднього дослідження, забезпечити цим достатнє різноманіття структур і типів складності;

- розробити метод автоматичної розмітки ознак структурної складності з використанням мультимодальних великих мовних моделей, визначити формат вхідної інформації (передача PDF-документів у вигляді бінарних вкладень), уніфіковану схему вихідних даних (рішення про наявність ознаки, текстове пояснення й числова оцінка впевненості), а також принципи побудови запитів і налаштування режимів роботи моделей;

- експериментально оцінити якість автоматичної розмітки ознак складності для кількох мовних моделей різних провайдерів у різних режимах роботи (без прикладів, з прикладами, з поглибленим внутрішнім міркуванням), порівняти результати

з ручною розміткою й проаналізувати вплив окремих ознак на якість екстракції даних;

- побудувати зважену функцію якості для множини моделей, яка агрегує значення якості автоматичної розмітки за всіма розглянутими ознаками з огляду на сили їх зв'язку з результатами екстракції слабкоструктурованих даних, та на цій підставі сформулювати задачу оптимізації вибору підмножини моделей у просторі "якість – вартість";

- розв'язати поставлену задачу оптимізації на основі емпіричних даних про якість і вартість використання мовних моделей, виокремити Парето-оптимальні конфігурації моделей і описати правило спеціалізованого призначення моделей окремим ознакам структурної складності, яке забезпечує практично доцільний баланс між точністю розмітки, вартістю та прогнозованістю процесу екстракції даних.

4. Матеріали й методи дослідження

Для розв'язання задачі автоматичної розмітки ознак структурної складності документів запропоновано оригінальний метод на основі емпіричного профілювання ефективності великих мовних моделей.

Ключова гіпотеза методу полягає в тому, що різні моделі демонструють неоднакову ефективність на різних типах ознак складності, тому призначення кожній ознаці оптимальної моделі забезпечує вищу загальну точність за контрольованої вартості порівняно з універсальним підходом використання однієї моделі для всіх ознак.

Метод передбачає реалізацію трьох послідовних етапів.

1. Профілювання моделей на експериментальному корпусі з еталонною розміткою для визначення ефективності кожної конфігурації моделі на кожній ознаці складності та вимірювання обчислювальної вартості.

2. Оптиміальне призначення моделей до ознак з огляду на точність, вартість і детермінізм результатів.

3. Застосування сформованої системи призначень до нових документів з паралельним обробленням різних ознак різними моделями.

Реалізація першого етапу полягає у формуванні експериментального корпусу з еталонною розміткою та в систематичному оцінюванні конфігурацій моделей за метриками якості та вартості. Такий корпус необхідний для профілювання ефективності моделей і визначення оптимальних конфігурацій.

В ідеальних умовах метод має бути перевірений на повному наборі ознак складності, виявлених у попередніх дослідженнях оброблення документів. Однак комплексна перевірка всіх можливих ознак пов'язана зі значними обчислювальними витратами й вимагає суттєвих ресурсів для профілювання моделей. Тому доцільним є відбір репрезентативної підмножини ознак, що забезпечує достатню валідацію методу за умови контрольованих витрат.

Для експериментальної перевірки методу ознаки обираються за такими критеріями:

1) емпірично підтверджений вплив на якість екстракції – ознака має демонструвати статистично значущу кореляцію з метриками якості екстракції даних;

2) покриття різних типів когнітивної складності – відібрані ознаки мають демонструвати різні аспекти складності документів (структурні, семантичні, контекстуальні), щоб забезпечити репрезентативність висновків;

3) достатнє подання в корпусі – кожна ознака має бути присутня в необхідній кількості документів для надійного оцінювання якості її автоматичної розмітки різними моделями;

4) обмеження дослідницьких ресурсів – кількість відібраних ознак має бути збалансована з доступними обчислювальними ресурсами й бюджетом на використання API комерційних моделей.

Розмір експериментальної вибірки документів з базового корпусу визначається на основі принципу стратифікованої випадкової вибірки зі збереженням розподілу ознак складності, що забезпечує репрезентативність результатів за умови скорочення обчислювальних витрат.

Конкретні параметри експерименту, як-от: кількість виокремлених ознак, розмір корпусу, властивості документів – подано в п'ятому розділі цієї статті.

Для профілювання ефективності автоматичної розмітки ознак складності використовуються великі мовні моделі, які задовольняють перелічені нижче вимоги. Моделі мають підтримувати безпосереднє оброблення PDF-файлів без попереднього перетворення в текстовий формат.

Ця вимога зумовлена тим, що значна частина ознак структурної складності (зокрема об'єднання комірок, візуальне розташування елементів, просторові залежності) закодована у візуальному поданні документа.

Попередня конвертація PDF у текст або HTML що ускладнює або унеможливує коректну ідентифікацію окремих ознак складності. Мультимодальні моделі, що обробляють PDF нативно, зберігають повний контекст документа, зокрема візуальну структуру й макет, що є критичним для об'єктивного оцінювання їх здатності до розмітки ознак.

Моделі мають бути доступними через програмний інтерфейс (API) без необхідності локального розгортання інфраструктури. Ця вимога забезпечує практичну застосовність методу: результати можуть бути легко відтворені іншими дослідниками або імплементовані у виробничих системах без складного технічного налаштування. API-доступ також гарантує використання актуальних версій моделей і забезпечує масштабованість оброблення.

Окремий інтерес становить експериментальна перевірка моделей у режимах міркування (*thinking modes*), які забезпечують покрокове обґрунтування прийнятих рішень.

Попри те, що такі режими зазвичай не дають змоги встановлювати температуру на рівні нуля й, відповідно, не гарантують повної детермінованості результатів, вони ефективні для оцінювання впливу експліцитних процесів міркувань на точність розмітки складних ознак. Порівняння продуктивності стандартних режимів і режимів міркувань дає змогу виявити типи ознак, для яких покрокове міркування забезпечує суттєве покращення якості.

Якість автоматичної розмітки кожної бінарної ознаки складності оцінювалася способом порівняння результатів роботи моделей з еталонною ручною розміткою.

Для кожної ознаки $f_i (i = 1, \dots, m)$, де m – загальна кількість ознак складності, та кожної конфігурації моделі обчислювалися стандартні метрики бінарної класифікації.

Нехай для ознаки f_i та документа d відомі:

– $y_i \in \{0,1\}$ – еталонне значення (ручна розмітка);

– $\hat{y}_i \in \{0,1\}$ – передбачене значення (результат роботи моделі).

На основі порівняння еталонних і передбачених значень для всіх документів корпусу визначаються базові величини:

– TP (*True Positives*) – кількість документів, де ознака правильно визначена як присутня ($y_i = 1, \hat{y}_i = 1$);

– *FP* (*False Positives*) – кількість документів, де ознака хибно визначена як присутня ($y_i = 0, y_i = 1$);

– *TN* (*True Negatives*) – кількість документів, де ознака правильно визначена як відсутня ($y_i = 0, y_i = 0$);

– *FN* (*False Negatives*) – кількість документів, де ознака хибно визначена як відсутня ($y_i = 1, y_i = 0$).

На основі перелічених величин обчислюються чотири метрики якості.

1. *Precision* (точність) – частка правильних позитивних передбачень серед усіх позитивних передбачень:

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i} \quad (1)$$

2. *Recall* (повнота) – частка правильно визначених позитивних випадків серед усіх еталонних позитивних випадків:

$$\text{Recall}_i = \frac{TP_i}{TP_i + FN_i} \quad (2)$$

3. F1-міра – гармонічне середнє точності та повноти:

$$F1_i = 2 \cdot \frac{\text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \quad (3)$$

4. *Accuracy* (точність класифікації) – частка правильних передбачень (як позитивних, так і негативних):

$$\text{Accuracy}_i = \frac{TP_i + TN_i}{N} \quad (4)$$

Для інтегрального оцінювання якості роботи конфігурації моделі на всіх ознаках використовувався показник макросередньої F1-міри. Відповідно до загальноприйнятої методології оцінювання мультикласової класифікації [14, 15] ця метрика розраховується як середнє арифметичне F1-мір для окремих класів (ознак), що дає змогу рівноправно брати до уваги внесок як поширених, так і рідкісних класів без зміщення, властивого зваженим методам.

$$F1_{macro} = (1/n) \sum_{i=1}^n F1_i, \quad i=1, \dots, n. \quad (5)$$

Цей показник забезпечує збалансовану увагу до якості розмітки всіх ознак незалежно від їх подання в корпусі й використовується як основний критерій для порівняння ефективності різних конфігурацій моделей.

Другий етап методу полягає у формалізації та розв'язанні задачі оптимального вибору

підмножини моделей для автоматичної розмітки ознак складності. На відміну від тривіального підходу, що передбачає використання однієї моделі для всіх ознак, запропонований метод дає змогу призначити кожній ознаці найбільш ефективну модель, формуючи спеціалізовану систему, що забезпечує оптимальне співвідношення між якістю розмітки й обчислювальними витратами.

На основі результатів профілювання маємо:

– множину моделей $M = \{m_1, m_2, \dots, m_k\}$,

де k – кількість моделей;

– множину ознак складності $C = \{c_1, c_2, \dots, c_n\}$,

де n – кількість ознак;

– значення $F1_{ij}$ – F1-міра, досягнута моделлю m_i на ознаці c_j ($i=1 \dots k; j=1 \dots n$);

– вектор вартості $V = (v_1, v_2, \dots, v_k)$, де v_i – вартість використання моделі m_i для оброблення одного документа.

Критична особливість задачі полягає в тому, що модель m_i за одну операцію (один виклик API), яка має вартість (коштує) v_i , здійснює розмітку всіх n ознак одночасно, але з різною якістю $F1_{ij}$ для кожної ознаки c_j . Це принципово відрізняє цю задачу від класичних задач призначення, де кожен ресурс використовується для одного об'єкта.

Запропонуємо бінарний вектор рішень $x = \{x_1, x_2, \dots, x_k\}$, де $x_i = 1$, якщо модель m_i застосовується, та $x_i = 0$ – в іншому випадку. Якість розмітки ознаки c_j за умови вибору підмножини моделей x визначається як максимум серед якостей усіх обраних моделей:

$$F1_j(x) = \max_{i \in \{1, \dots, k\}, x_i = 1} F1_{ij}. \quad (6)$$

Для порівняння різних конфігурацій моделей використовується скалярна функція якості $Q(S)$, яка агрегує значення F1 для всіх ознак структурної складності. У найпростішому випадку $Q(S)$ може задаватися як звичайне середнє арифметичне значення F1 за ознаками, що фактично передбачає рівну важливість усіх ознак.

Однак попередній аналіз кореляцій між ознаками структурної складності та загальною F1 екстракції даних показав, що окремі ознаки по-різному впливають на якість екстракції. Тому доцільно

використовувати зважений інтегральний показник, у якому ознаки із сильнішим зв'язком із F1 екстракції мають більший внесок у функцію якості.

Нехай C – множина ознак структурної складності. Для кожної ознаки $c \in C$ на попередньому етапі обчислюється коефіцієнт кореляції r_c між бінарною наявністю цієї ознаки та F1 якості екстракції даних (за результатами зовнішнього експерименту).

На основі модулів цих кореляцій задаються ваги:

$$w_c = \frac{|r_c|}{\sum_{g \in C} |r_g|}, \quad c \in C, \quad (7)$$

які відтворюють відносну важливість кожної ознаки щодо її впливу на якість екстракції.

Позначаючи як $F1_{S,c}$ якість розмітки ознаки f для множини моделей S (визначену як максимум F1 серед моделей із S), функцію якості для множини моделей $S \subseteq M$ визначатимемо таким чином:

$$Q(S) = \sum_{c \in C} w_c \cdot F1_{S,c}. \quad (8)$$

Така конструкція відповідає практичному механізму роботи системи: для кожної ознаки $c \in C$ на етапі налаштування обирається модель із множини S , яка демонструє найвище значення $F1_{m,c}$, і надалі саме ця модель використовується для автоматичної розмітки відповідної ознаки. Ваги w_c , отримані з таблиці кореляцій, забезпечують узгодженість цільового критерію $Q(S)$ з реальною значущістю ознак для задачі екстракції слабкоструктурованих даних: ознаки, що сильніше пов'язані зі зміною F1 екстракції, мають більший вплив на прийняття рішень щодо вибору конфігурації моделей.

Загальна вартість оброблення одного документа дорівнює сумі вартостей усіх обраних моделей:

$$\text{Cost}(x) = \sum_{i=1}^k v_i \cdot x_i. \quad (9)$$

У наведених ознаках припускається, що для кожного документа розглядається фіксований перелік ознак структурної складності, а кожна модель з множини M під час одного запуску формує

$$\max_x \text{Cost}(x) | F1_j(x) \geq \tau_j, \quad j=1, \dots, n, \quad x_i \in \{0, 1\}, \quad i=1, \dots, k, \quad (11)$$

де τ_j – мінімально прийнятне значення F1-міри для j -ї ознаки.

Пошук Парето-оптимальних рішень. Загалом, коли важливо знайти компроміс, розглядається задача

оцінки для всіх цих ознак одночасно. Це вказує на те, що вартість використання моделі не залежить від кількості ознак, які фактично будуть закріплені за цією моделлю: якщо бодай одна ознака призначена моделі m , документ обробляється нею повністю, і до функції вартості $\text{Cost}(S)$ додається внесок c_m .

Отже, $\text{Cost}(S)$ інтерпретується як сумарні витрати на оброблення одного документа всіма моделями, які містить множина S . Функція якості $Q(S)$, задана як середнє арифметичне значення максимумів $\max_{m \in S} F1_{m,k}$

за ознаками, відтворює найкращий досяжний рівень F1 для кожної ознаки за умови, що для її розмітки застосовуються лише моделі з множини S . Максимум за моделями означає, що на етапі налаштування системи для кожної ознаки обирається відповідальна модель із S , яка демонструє найвище значення F1 для цієї ознаки, і надалі саме ця модель використовується для автоматичної розмітки відповідної ознаки в робочому режимі.

Сформульована задача є задачею багатокритеріальної оптимізації з двома конфліктуєчими цілями: максимізацією якості $Q(x)$ та мінімізацією вартості $\text{Cost}(x)$. Залежно від особливостей прикладної галузі та наявних обмежень ресурсів задачу вибору оптимальної підмножини моделей можна сформулювати в кількох варіантах. Розглянемо три основних.

Максимізація якості за умови обмеження на вартість. Застосовується, коли бюджет фіксований і необхідно отримати найкращу можливу якість розмітки. Математична модель має такий вигляд:

$$\max_x Q(x) | \text{Cost}(x) \leq V_{\{\max\}}, \quad x_i \in \{0, 1\}, \quad i=1, \dots, k, \quad (10)$$

де $V_{\{\max\}}$ – максимально допустима вартість оброблення одного документа.

Мінімізація вартості за умови гарантії якості. Цей варіант доцільний, коли існують суворі вимоги до точності розмітки й необхідно досягти їх з мінімальними витратами:

багатокритеріальної оптимізації. Множина рішень x^* називається Парето-оптимальною, якщо не існує іншого рішення x , що одночасно покращує обидва критерії, тобто $Q(x) \geq Q(x^*)$

та $Cost(x) \leq Cost(x^*)$. Крім того, хоча б одна з нерівностей є строгою.

Задача належить до класу задач комбінаторної оптимізації з бінарними змінними й нелінійною цільовою функцією. За своєю структурою вона є узагальненням відомої задачі про покриття множини (*Set Cover Problem*) [16], ускладненої додатковим критерієм якості та "м'яким" характером покриття (через функцію максимуму).

Зважаючи, що отримана математична модель є типовою задачею дискретної оптимізації, її розв'язують відомими методами з використанням стандартних бібліотек (*OR-Tools*, *PuLP* тощо) без необхідності розроблення спеціалізованих алгоритмів.

Вихідним матеріалом для експерименту взято матеріал попереднього дослідження [11], в якому було сформовано корпус із 115 синтетичних інвойсів. Для них вручну виконано розмітку 11 бінарних ознак структурної складності документа та обчислено метрику якості екстракції даних F1 для кількох великих мовних моделей.

Аналіз кореляцій між наявністю кожної ознаки й значенням F1 показав, що лише частина критеріїв демонструє стійкий і статистично значущий зв'язок із якістю екстракції даних.

Таблиця 1. Коефіцієнти кореляції між ознаками структурної складності та F1-міри якості екстракції даних

Ознака структурної складності документа	Значення F1-міри			Середнє $ r $
	<i>GPT-4.1-nano</i>	<i>GPT-4o-mini</i>	<i>Gemini 2.5 Flash</i>	
<i>has_ambiguous_fields</i>	-0.298	-0.257	-0.469	0.341
<i>has_spanning_data</i>	-0.462	-0.431	-0.449	0.447
<i>has_unlabeled_data</i>	-0.477	-0.563	-0.597	0.546
<i>requires_calculations</i>	-0.331	-0.131	-0.277	0.246

Від'ємні значення r відповідають зменшенню F1-міри в присутності відповідної ознаки. Значущість коефіцієнтів оцінювалась за t -критерієм Стюдента: для обсягу вибірки $N=115$ та рівня значущості $\alpha=0.05$ критичне значення кореляції становить $|r|_{crit} \approx 0.183$. Відповідно, усі коефіцієнти, що перевищують це значення за модулем, є статистично значущими.

Усі чотири ознаки мають помітний зворотний зв'язок з якістю екстракції: у присутності цих ознак значення F1, як правило, зменшується. Найбільше за модулем середнє значення $|r|$ спостерігається для ознаки непідписаних табличних даних і для ознаки даних, які охоплюють кілька рядків або

Для експериментального дослідження обрано чотири ознаки структурної складності, які є найбільш релевантними для задачі екстракції слабкоструктурованих даних з PDF-документів облікового призначення:

- наявність непідписаних даних у табличній частині (*has unlabeled data*);
- наявність неоднозначних або потенційно суперечливих полів (*has ambiguous fields*);
- наявність даних, що охоплюють кілька рядків чи стовпців таблиці (*has spanning data*);
- необхідність виконання арифметичних обчислень для відновлення відсутніх значень (*requires calculations*).

Вибір саме цієї підмножини ознак зумовлений такими міркуваннями: по-перше, ці ознаки демонструють найсильніший вплив на якість екстракції для всіх досліджуваних моделей; по-друге, вони покривають різні категорії складності (структурні, семантичні, контекстуальні); по-третє, їх автоматичне визначення становить задачу, що потребує когнітивного аналізу структури й змісту документа.

Узагальнені значення коефіцієнтів кореляції Пірсона між обраними ознаками й загальною F1-мірою якості екстракції даних для трьох моделей наведено в табл. 1.

стовпців, що свідчить про їх особливо несприятливий вплив на стабільність автоматизованого витягання реквізитів. Важливо, що частина обраних ознак принципово не може бути коректно визначена лише на основі лінійного текстового подання документа.

Ознаки, пов'язані з комірками, що охоплюють кілька рядків або стовпців, зі складеними комірками з кількома логічними полями чи непідписаними числовими даними, залежать від просторового розміщення елементів таблиці на сторінці. Для їх надійної розмітки модель має опрацьовувати структуроване або мультимодальне подання PDF-документа (табличну розмітку, координати, ієрархію блоків), а не лише "плоский" текст.

Отже, обраний набір ознак дає змогу не тільки оцінити вплив структурної складності на якість екстракції, але й протестувати здатність великих мовних моделей інтерпретувати повну візуально-структурну організацію документа.

Для експериментальної перевірки запропонованого методу автоматичної розмітки було сформовано підвибірку з 44 документів базового корпусу. До неї увійшли інвойси з еталонною (ручною) розміткою всіх 11 ознак структурної складності та відомими значеннями F1-міри якості екстракції для кожної моделі.

Підвибірка формувалася таким чином, щоб забезпечити достатню кількість прикладів позитивних і негативних значень для кожної з чотирьох обраних ознак та їх поєднань. Саме на цих 44 документах проводилося профілювання моделей та оцінювалася якість автоматичної розмітки ознак структурної складності.

5. Результати дослідження та їх обговорення

Розглянемо отримані результати автоматичної розмітки чотирьох ознак структурної складності для підвибірки з 44 PDF-документів.

Для кожного документа застосовувалися кілька конфігурацій великих мовних моделей трьох

постачальників (*Anthropic Claude*, *OpenAI GPT-5*, *Google Gemini*) у різних режимах роботи: без прикладів у підказці, з використанням прикладів і за наявності підтримки так званих режимів міркування (*thinking mode*). Оцінювання виконувалося окремо для кожної з чотирьох ознак складності, а також за інтегральним показником – середнім арифметичним значенням F1-міри (*macro-F1*) за всіма ознаками для кожної конфігурації моделі.

У табл. 2 подано зведені значення F1-міри для кожної моделі, режиму застосування та ознаки, а також відповідні *macro-F1*, що дає змогу порівнювати якість розмітки як між різними постачальниками, так і між режимами для одних і тих самих моделей.

В експерименті розмітка ознак структурної складності виконувалася з використанням єдиного стандартизованого промпта. Усі документи передавалися до моделей як бінарні PDF-вкладення через API: окремо передавався PDF-файл, що підлягає розмітці, а також один або декілька PDF-файлів із прикладами. Для прикладів додатково надавалися окремі файли з еталонною розміткою ознак, що застосовувалися в промпті для демонстрації цільового формату відповіді та інтерпретації критеріїв складності.

Таблиця 2. Оцінка якості автоматичної розмітки ознак структурної складності для різних моделей і режимів

Режим	Модель	Значення F1-міри				
		<i>macro</i>	<i>has unlabeled data</i>	<i>has ambiguous fields</i>	<i>has spanning data</i>	<i>requires calculations</i>
ZeroShot	<i>claude-haiku-4.5</i>	0.456	0.065	0.708	0.764	0.286
	<i>claude-sonnet-4.5</i>	0.594	0.359	0.864	0.653	0.500
	<i>gemini-2.5-flash</i>	0.710	0.846	0.619	0.488	0.889
	<i>gemini-3-pro</i>	0.781	0.909	0.629	0.588	1.000
	<i>gpt-5-mini</i>	0.700	0.462	0.898	0.439	1.000
	<i>gpt-5.1</i>	0.618	0.500	0.780	0.692	0.500
	<i>gpt-5.1codex</i>	0.594	0.333	0.905	0.138	1.000
FewShot	<i>gemini 2.5 flash</i>	0.828	0.929	0.652	0.731	1.000
	<i>gemini3 pro</i>	0.771	0.966	0.414	0.703	1.000
	<i>gpt-5-mini</i>	0.660	0.400	0.846	0.562	0.833
	<i>gpt-5.1</i>	0.619	0.903	0.757	0.816	0.000
Thinking	<i>claude sonnet 4.5</i>	0.615	0.286	0.800	0.375	1.000
Thinking	<i>claude sonnet 4.5</i>	0.793	0.897	0.800	0.474	1.000
FewShot	<i>gemini 3 pro</i>	0.725	0.929	0.364	0.606	1.000

У всіх конфігураціях без використання режимів міркування (*thinking mode*) параметр температури (*temperature*) встановлювався на рівні 0, що забезпечувало максимально можливу детермінованість результатів для фіксованого набору вхідних файлів і тексту підказки. Це налаштування є принциповим,

оскільки дає змогу інтерпретувати розбіжності в показниках якості як наслідок відмінностей між моделями та режимами, а не стохастичних варіацій генерації.

У тексті промпта явно задавалася вимога повертати результат розмітки у вигляді JSON-об'єкта

фіксованої структури. У відповіді формувався запис, що містив ідентифікатор документа (*document_id*) і словник критеріїв *criteria*, у якому для кожної ознаки структурної складності зберігався трійковий опис: *label*, *evidence* та *confidence*. Поле *label* могло набувати одного з трьох значень фіксованого словника (*present*, *absent*, *unsure*), тобто задавало тринарне рішення щодо наявності ознаки.

Для обчислення бінарних метрик якості (TP, FP, TN, FN, *Precision*, *Recall*, F1-міри) ці значення додатково відображалися у бінарну змінну: значення *present* тлумачилося як позитивний клас, а значення *absent* і *unsure* – як негативний.

Поле *confidence* містило числову оцінку впевненості моделі в діапазоні [0;1], а поле *evidence* – стисле текстове пояснення із посиланням на конкретні елементи PDF-документа, на підставі яких було прийнято відповідне рішення.

Такий формат відповіді, жорстко визначений промптом, є важливим як для інтерпретованості результатів (особливо в разі режимів міркування), так і для подальшої перевірки еталонної розмітки та ітеративного вдосконалення самого промпта.

Окрему групу конфігурацій становили так звані режими міркування (*thinking mode*) моделей із розширеним мисленням. Їх використання виявило низку практичних обмежень. По-перше, для режиму міркування *Anthropic Claude* мінімально доступне значення параметра *temperature* становило 1, а для *reasoning*-режиму *Gemini* – 0.2, що не дає змогу реалізувати повністю детермінований сценарій оброблення, аналогічний до режимів із *temperature* = 0. За таких налаштувань навіть для фіксованого набору вхідних PDF-файлів, прикладів і тексту підказки відповіді моделі залишаються стохастичними: повторні запуски можуть повертати різні варіанти значень *label*, *evidence* та *confidence* для тих самих документів.

По-друге, експериментальні результати продемонстрували, що використання режимів міркування (*thinking mode*) не забезпечує суттєвого виграшу в якості для задачі бінарної розмітки чотирьох ознак структурної складності. Значення F1-міри окремих ознак та *macro-F1* для конфігурацій типу *Thinking* і *Thinking + FewShot* залишаються на рівні найкращих *FewShot*-конфігурацій або лише незначно їх перевищують, тоді як споживання токенів і, відповідно, вартість виконання запитів зростають у кілька разів.

У поєднанні з підвищеною недетермінованістю це робить *thinking*-режими менш привабливими для промислового застосування саме в розглянутій постановці задачі, де критичними є відтворюваність результатів і прогнозовані витрати, а структура рішення (формат JSON з *label*, *evidence* і *confidence*) уже забезпечує достатню інтерпретованість без необхідності додаткового багатоетапного міркування.

Водночас результати спостереження щодо роботи режимів міркування мають окрему методологічну цінність. Аналіз проміжних міркувань моделі, структури сформованих пояснень і властивих помилок дає змогу краще зрозуміти, як саме інтерпретуються формулювання критеріїв у промпті. Це розуміння може використовуватися для цілеспрямованого вдосконалення підказок: уточнення визначень ознак, добору більш репрезентативних прикладів, зміни інструкцій щодо оброблення приграничних випадків.

Отже, незважаючи на те, що режими міркування в розглянутій задачі не є економічно доцільними для продуктового застосування, вони можуть розглядатися як корисний інструмент діагностики й налагодження промптів на етапі дослідження.

Для оцінювання економічної доцільності застосування різних конфігурацій моделей було проаналізовано споживання токенів і вартість виконання розмітки корпусу із 44 документів. Для кожної моделі окремо бралися до уваги кількість вхідних і вихідних токенів, тарифи постачальника за 1 млн токенів відповідного типу, а також підсумкова вартість оброблення всієї вибірки.

На підставі цих показників було обчислено повну вартість запуску кожної конфігурації моделі в межах проведеного експерименту. Зведені результати наведено в табл. 3, що дає змогу безпосередньо порівнювати моделі не лише за якістю розмітки ознак, а й за витратами, необхідними для досягнення такої якості.

У механізмі ціноутворення сучасних хмарних LLM принциповим є не тільки сам факт застосування тієї чи іншої моделі, а й спосіб її експлуатації: обсяг вхідних і вихідних токенів, наявність режимів міркування, використання кешування, а також під'єднання старших версій моделей із розширеним функціоналом.

У табл. 3 подано оцінку вартості для кожної конфігурації в припущенні застосування верхнього тарифного рівня для відповідного постачальника

та без огляду на механізми кешування. Це означає, що наведені значення відтворюють консервативну (максимальну) оцінку витрат у процесі оброблення

44 документів у межах експерименту, тоді як реальна вартість для малих обсягів може бути нижчою через дію тарифних градацій.

Таблиця 3. Споживання токенів і вартість використання моделей

Режим	Модель	Вхідні токени			Вихідні токени			Загальна вартість, USD
		кільк.	ціна, USD/1M	вартість, USD	кільк.	ціна, USD/1M	вартість, USD	
ZeroShot	claude-haiku-4.5	459680	0.100	0.046	25658	5.000	0.128	0.174
	claude-sonnet-4.5	459680	0.300	0.138	21579	15.000	0.324	0.462
	gemini-2.5-flash	130930	1.000	0.131	17194	2.500	0.043	0.174
	gemini-3-pro	301498	4.000	1.206	14380	18.000	0.259	1.465
	gpt-5-mini	340290	0.250	0.085	64232	2.000	0.128	0.214
	gpt-5.1	318090	1.250	0.398	16777	10.000	0.168	0.565
	gpt-5.1codex	318090	1.250	0.398	58229	10.000	0.582	0.980
FewShot	gemini 2.5 flash	432754	1.000	0.433	17512	2.500	0.044	0.477
	gemini3 pro	667336	4.000	2.669	16055	18.000	0.289	2.958
	gpt-5-mini	827852	1.250	1.035	60548	10.000	0.605	1.640
	gpt-5.1	740998	1.250	0.926	16554	10.000	0.166	1.092
Thinking	claude sonnet 4.5	461000	0.300	0.138	128791	15.000	1.932	2.070
Thinking	claude sonnet 4.5	852173	0.300	0.256	156324	15.000	2.345	2.601
FewShot	gemini 3 pro	657584	4.000	2.630	16325	18.000	0.294	2.924

Для моделей Google й Anthropic ціна за одиницю токена формально не залежить від того, використовується стандартний режим чи режим міркування. Однак доступ до такого режиму надається переважно в старших версіях моделей, що вже саме по собі підвищує базовий тариф, якщо порівнювати з "легкими" конфігураціями.

Крім того, у режимах міркування суттєво зростає кількість вихідних токенів: до звичайної відповіді моделі додаються токени, пов'язані з проміжними кроками міркування (ланцюжок думок, планування, розгортання структурованих пояснень тощо). У білінгу такі "токени міркування" розглядаються як вихідні, і для них застосовується тариф на вихідні токени, який у більшості комерційних планів істотно вищий, ніж для вхідних. Як наслідок, сумарні витрати для reasoning-конфігурацій можуть зростати в рази навіть за незмінної кількості оброблених документів.

Важливим елементом гнучкості є здатність обмежувати максимальну кількість токенів, що можуть бути витрачені під час оброблення одного запиту, зокрема й на етапі "обміркування". Це дає змогу контролювати верхню межу вартості "міркувальних" запитів і запобігати неконтрольованому розширенню вихідної послідовності, однак водночас створює ризик передчасного обрізання ланцюжка міркувань і, зрештою, деградації якості відповіді.

Отже, налаштування лімітів для режимів міркування перетворюється на оптимізаційну задачу: необхідно досягти прийняттого компромісу між глибиною міркування, стабільністю якості та прогнозованістю витрат.

Додатковим чинником є тарифні градації великих постачальників: вартість 1 млн токенів зазвичай залежить від сумарного обсягу використання за певний період.

У табл. 3 для простоти порівняння взято до уваги максимальні (найгірші, на думку користувача) ставки. На практиці в сценаріях оброблення поодиноких документів або незначних партій реальна ціна за токен може бути нижчою, якщо обсяг застосування не досягає найвищих тарифних коридорів.

Водночас за умови масового оброблення значної кількості документів сумарні витрати різко зростають, і вплив вибору режиму (*ZeroShot*, *FewShot* або *Thinking*) і конкретної моделі стає критичним для економічної ефективності системи.

Окремо важливо наголосити на значенні механізму кешування токенів. У низці реалізацій LLM API передбачено можливість повторно використовувати попередньо зафіксовані фрагменти запитів (наприклад, системні інструкції, спільну частину промпта або незмінні приклади). До того ж вартість "кешованих" вхідних токенів значно нижча за повну ціну стандартного введення.

Для задачі розмітки ознак структурної складності це особливо актуально в сценаріях із застосуванням прикладів, де один і той самий набір еталонних прикладів багаторазово використовується для розмітки різних документів. Перенесення цих прикладів до кешу дає змогу суттєво зменшити сумарні витрати на вхідні токени. Водночас кешування практично не впливає на вартість вихідних tokenів, а саме вихідні токени (зокрема й токени міркування) є найбільш вартісними.

Отже, механізм кешування ефективно знижує витрати в конфігураціях із значним обсягом повторюваного вхідного контексту, але майже не допомагає в разі, коли основна частина вартості зумовлена великою довжиною вихідних відповідей.

Загалом наведені фактори показують, що оцінювання економічної доцільності використання тієї чи іншої моделі не може обмежуватися лише порівнянням "номінальних" тарифів за 1 млн tokenів. Необхідно зважати на розподіл навантаження між вхідними й вихідними токенами, наявність або відсутність режимів міркування, обсяги застосування, можливість кешування повторюваних фрагментів промпта, а також практичні обмеження на температурні налаштування й максимальну довжину відповіді.

Тільки з огляду на перелічені аспекти вартісний аналіз уможливорює коректне порівняння моделей між собою та обґрунтоване обрання конфігурації, що забезпечують найкращий баланс між якістю розмітки ознак структурної складності й витратами на обчислювальні ресурси.

Практичне використання системи автоматичної розмітки ознак структурної складності вимагає уваги не лише до досяжної якості, а й до вартості застосування великих мовних моделей.

Вище запропоновано зважену функцію якості $Q(S)$, яка агрегує F1-міри за всіма ознаками з використанням ваг w_c , пропорційних середнім за моделі абсолютним значенням коефіцієнтів кореляції $|r_c|$ між наявністю ознаки $c \in C$ і загальною F1 якістю екстракції даних.

Отже, ознаки, що демонструють сильніший зв'язок із деградацією або покращенням якості екстракції, мають більший внесок у цільовий показник $Q(S)$.

Для кожної моделі $m \in M$ та ознаки $c \in C$ за результатами експерименту було обчислено

значення $F1_{m,c}$, що визначає якість автоматичної розмітки відповідної ознаки, а також емпіричну вартість повного прогону моделі c_m , встановлену на основі фактичного споживання вхідних і вихідних tokenів. Для кожної підмножини моделей $S \subseteq M$ значення $F1_{S,c}$, зваженої функції якості $Q(S)$ та функції вартості $Cost(S)$ обчислювалися за формулами (7)–(9), тобто як максимум F1 за моделями для кожної ознаки, зважена сума значень якості з використанням ваг w_c , отриманих із таблиці кореляцій, і сума вартостей c_m для всіх моделей, які містить множина S .

Така конструкція відповідає сценарію, у якому для кожної ознаки структурної складності на етапі налаштування системи обирається відповідна модель із множини S , що демонструє найвище значення $F1_{m,c}$. Саме цю модель надалі застосовано для автоматичної розмітки відповідної ознаки, тоді як вартість використання конфігурації визначається сумою вартостей усіх моделей, які хоча б для однієї ознаки беруть участь у розмітці: $Cost(S) = \sum_{m \in S} c_m$. На завершальному етапі дослідження побудовано зважену функцію якості $Q(S)$ та функцію вартості $Cost(S)$, які обчислюються за формулами (7)–(9), і сформульовано задачу вибору підмножини моделей у просторі "якість – вартість".

Для заданої множини моделей M необхідно знайти підмножину $S \subseteq M$, яка максимізує $Q(S)$ за умови, що загальна вартість $Cost(S)$ не перевищує припустимого бюджетного порогу.

Після цього з множини всіх конфігурацій було виокремлено Парето-оптимальні (ефективні) набори моделей, для яких не існує іншого набору S' , що має не гіршу якість за не більшу вартість, тобто $Q(S') \geq Q(S)$, $Cost(S') \leq Cost(S)$. Крім цього, хоча б одна з нерівностей є строгою.

Сукупність таких конфігурацій описує фронт Парето в просторі "якість – вартість" і задає практично релевантні компроміси між якістю автоматичної розмітки й витратами на використання моделей.

У табл. 4 наведено Парето-оптимальні конфігурації відповідно до зростання вартості. Для кожного набору моделей S подано сумарну

вартість $Cost(S)$ (за результатами оброблення 44 документів) та зважену якість $Q(S)$, обчислену за описаною вище схемою (значення наведено з округленням до трьох знаків).

Отримані результати демонструють, що оптимально підібрані невеликі набори недорогих моделей допомагають забезпечити майже максимальну

досягнути якість розмітки складності за значно нижчої вартості, ніж просте використання всіх доступних моделей, і без залучення вартісних режимів міркування.

Для кожної Парето-оптимальної конфігурації S , що наведені в табл. 4, відповідна модель для ознаки $c \in C$ визначається за правилом:

$$m^*(c; S) = \arg \max_{m \in S} F1_{m,c} \quad (12)$$

Таблиця 4. Парето-оптимальні набори моделей у просторі "якість – вартість"

№	Набір моделей, S	Вартість $Cost(S)$, USD	Зважена якість, $Q(S)$
1	{ gemini-2.5-flash }	0.174	0.702
2	{ claude-haiku-4-5, gemini-2.5-flash }	0.348	0.800
3	{ gemini-2.5-flashF }	0.477	0.824
4	{ claude-haiku-4-5, gemini-2.5-flash, gpt-5-mini }	0.562	0.858
5	{ gemini-2.5-flashF, gpt-5-mini }	0.690	0.877
6	{ claude-haiku-4-5, gemini-2.5-flashF, gpt-5-mini }	0.864	0.887
7	{ gpt-5-mini, gpt-5.1F }	1.305	0.892
8	{ gemini-2.5-flashF, gpt-5-mini, gpt-5.1F }	1.782	0.901
9	{ gemini-2.5-flashF, gpt-5.1-codex, gpt-5.1F }	2.548	0.903
10	{ gemini-3-pro-previewF, gpt-5-mini, gpt-5.1F }	4.264	0.914
11	{ gemini-3-pro-previewF, gpt-5.1-codex, gpt-5.1F }	5.030	0.916

Іншими словами, для кожної ознаки обирається модель із множини S , яка демонструє найвище значення F1-міри для цієї ознаки, і саме ця модель використовується для її автоматичної розмітки.

Отже, табл. 4 дає не лише розуміння щодо оптимальних наборів моделей у просторі "вартість – якість", а й однозначно задає розподіл ознак між моделями для кожної обраної конфігурації.

6. Висновки

й перспективи подальших досліджень

У межах проведеного дослідження розроблено метод автоматичної розмітки ознак структурної складності документів у форматі PDF з використанням великих мовних моделей. Метод може бути впроваджений для довільної множини ознак і довільного набору моделей і не обмежується певною кількістю критеріїв чи конкретними архітектурами. Ключовою ідеєю є те, що документ подається до мовної моделі у вигляді цілісного файлу, а на виході формується структурований опис його складності, який містить для кожної ознаки чітке рішення про її наявність, стисле текстове пояснення та числову оцінку впевненості в цьому рішенні.

Таке подання є універсальним і може безпосередньо використовуватися на подальших

етапах оброблення документів. Запропонований метод робить процес екстракції слабкоструктурованих даних з PDF-документів не лише більш точним, а й прогнозованим і керованим. Оскільки для ознак структурної складності заздалегідь досліджено їх зв'язок із якістю екстракції, профіль складності конкретного документа може використовуватися як окрема вхідна властивість: за ним можна оцінити очікуваний рівень якості вилучення даних, визначити, чи достатньо застосувати недорогу модель, чи доцільно залучити більш потужну модель або передати документ на додаткову перевірку оператору.

Отже, розмітка ознак складності перетворюється на інструмент, за допомогою якого система екстракції набуває властивостей прогнозованості (очікуваний рівень якості відомий наперед) і керованості (користувач або розробник може явно обирати компроміс між якістю та витратами).

В експериментальній частині роботи метод застосовано до підмножини з чотирьох ознак структурної складності, для яких попередньо було встановлено найбільш виразний вплив на якість екстракції даних. Це лише один із можливих варіантів налаштування методу. Сам метод не прив'язаний ані до конкретної кількості ознак, ані до конкретних типів документів: за наявності

нових ознак або нових форматів документів достатньо повторити етап аналізу зв'язку між ознаками та якістю екстракції і оновити ваги в інтегральній функції якості. Автоматична розмітка ознак складності за допомогою великих мовних моделей у цьому разі дає змогу отримувати потрібні дані для будь-якого обсягу документів без різкого зростання витрат на ручну працю, тоді як повністю ручна розмітка такого самого обсягу була б значно вартісною та практично нереальною для підтримки в актуальному стані.

Експериментальні результати продемонстрували, що розроблений метод сприяє досягненню узгодженості автоматичної розмітки з ручною на рівні, достатньому для практичного використання в реальних конвеєрах оброблення документів. Запропонована зважена функція якості, яка бере до уваги різний вплив окремих ознак на підсумкову точність екстракції, у поєднанні з розв'язанням оптимізаційної задачі в просторі "якість – вартість" дала змогу побудувати Парето-оптимальні конфігурації мовних моделей.

Показано, що невеликі набори відносно недорогих моделей у режимах без прикладів і з незначною кількістю прикладів забезпечують майже максимально досягну зважену якість автоматичної розмітки. Водночас конфігурації з режимами поглибленого внутрішнього міркування не продемонстрували переваги, яка б виправдовувала суттєве зростання витрат.

Це підтверджує, що автоматичну розмітку ознак складності доцільно будувати не на найскладніших і найдорожчих моделях, а на продуманій комбінації кількох більш доступних моделей, оптимально підібраних з огляду на співвідношення "якість – вартість".

Перспективи подальших досліджень пов'язані з розширенням таксономії ознак структурної складності, залученням ширшого кола типів документів та інтеграцією модуля автоматичної розмітки до промислових систем екстракції слабкоструктурованих даних. Профіль складності документа може слугувати

основою для побудови моделей прогнозування якості екстракції, для адаптивного вибору конфігурації моделей залежно від складності конкретного документа й для моніторингу зміни потоку документів у часі.

Подальші дослідження можуть також бути присвячені розробленню багатокритеріальних моделей оптимізації, які одночасно беруть до уваги якість, вартість, затримку оброблення та стабільність результатів. Загалом це створює підґрунтя для побудови систем, у яких екстракція слабкоструктурованих даних з PDF-документів досягне необхідного рівня якості, а їх поведінка залишиться прозорою, прогнозованою та керованою навіть за умови постійних змін структури вхідних документів.

Конфлікт інтересів

Автори статті декларують, що не мають конфлікту інтересів, зокрема фінансового, особистого, авторського чи будь-якого іншого характеру, який міг би вплинути на дослідження, а також на результати, опубліковані в цій роботі.

Фінансування

Дослідження проводилося без фінансової підтримки.

Доступність даних

Рукопис не має пов'язаних матеріалів у сховищі даних.

Використання засобів штучного інтелекту

Автори підтверджують, що не застосовували технології штучного інтелекту для написання цієї роботи.

References

1. Document Parsing Unveiled: Techniques, Challenges, and Prospects for Structured Information Extraction. *arXiv*. 2024. URL: <https://arxiv.org/html/2410.21169v1>
 2. Fisher, J. L. (1991), "Logical Structure Descriptions of Segmented Document Images", *Proceedings of International Conference on Document Analysis and Recognition (ICDAR)*, Saint-Malo, France, P. 302–310.
 3. Vinay, V. et al. (2006), "Measuring the Complexity of a Collection of Documents", *Advances in Information Retrieval*, Vol. 3936, P. 107–118. DOI: https://doi.org/10.1007/11735106_11
-

4. Pembe, F. C., Güngör, T. (2015), "A Tree-Based Learning Approach for Document Structure Analysis and Its Application to Web Search", *Natural Language Engineering*, Vol. 21, No. 4, P. 569–605. DOI: <https://doi.org/10.1017/S1351324914000023>
5. Paliwal, S. et al. (2020), "TableNet: Deep Learning model for end-to-end Table detection and Tabular data extraction from Scanned Document Images", *arXiv*. DOI: <https://doi.org/10.48550/arXiv.2001.01469>
6. Huang, Y. et al. (2022), "LayoutLMv3: Pre-training for Document AI with Unified Text and Image Masking", *arXiv*. DOI: <https://doi.org/10.48550/arXiv.2204.08387>
7. Atagong, S. D. et al. (2025), "A review on knowledge and information extraction from PDF documents and storage approaches", *Frontiers in Artificial Intelligence*, Vol. 8, P. 1466092. DOI: <https://doi.org/10.3389/frai.2025.1466092>
8. Meuschke, N. et al. (2023), "A Benchmark of PDF Information Extraction Tools using a Multi-Task and Multi-Domain Evaluation Framework for Academic Documents", *arXiv*. DOI: <https://doi.org/10.48550/arXiv.2303.09957>
9. Bozhko, O. (2025), "Development of an iterative method for data extraction from unstructured documents based on the use of large language models", *Transactions of Kremenchuk Mykhailo Ostrohradskyi National University*, Iss. 1, P. 119–124. DOI: <https://doi.org/10.32782/1995-0519.2025.1.15>
10. Wang, W. et al. (2025), "Document Intelligence in the Era of Large Language Models: A Survey", *arXiv*. DOI: <https://doi.org/10.48550/arXiv.2510.13366>
11. Ding, Y. et al. (2025), "Deep Learning based Visually Rich Document Content Understanding: A Survey", *arXiv*. DOI: <https://doi.org/10.48550/arXiv.2408.01287>
12. Kupin, A. I., Kosei, M. P. (2024), "Overview of Multi-Agent System Architectures and Swarm Intelligence Algorithms", *Scientific notes of Taurida National V.I. Vernadsky University. Series: Technical Sciences*, Iss. 2, P. 98–104. DOI: <https://doi.org/10.32782/2663-5941/2024.2/14>
13. Петров, К. Е., Боков, І. П., Кобзев, І. В. (2025), "Розробка комбінованого методу аналізу емоційної забарвленості текстів", *АСУ та прилади автоматики*, Вип. 186, С. 5–16. DOI: <https://doi.org/10.30837/0135-1710.2025.186.005>
14. Jurafsky, D., Martin, J. H. (2024), "Speech and Language Processing (3rd ed. draft)". URL: <https://web.stanford.edu/~jurafsky/slp3/> (accessed 20.08.2024).
15. Yamasaki, Chihiro, et al. (2025), "Function-based Labels for Complementary Recommendation: Definition, Annotation, and LLM-as-a-Judge", *arXiv*. DOI: <https://doi.org/10.48550/ARXIV.2507.03945>
16. Karp, R. M. (1972), "Reducibility Among Combinatorial Problems", *Complexity of Computer Computations*, P. 85–103.

Received (Надійшла) 30.11.2025

Accepted for publication (Прийнята до друку) 28.12.2025

Publication date (Дата публікації) 12.03.2026

Відомості про авторів / About the Authors

Петров Костянтин Едуардович – доктор технічних наук, професор, Харківський національний університет радіоелектроніки, завідувач кафедри інформаційних управляючих систем; Харків, Україна;

Konstantin Petrov – Doctor of Technical Sciences, Professor, Kharkiv National University of Radio Electronics, Head at the Information Control Systems Department; Kharkiv, Ukraine;

e-mail: kostiantyn.petrov@nure.ua

ORCID ID: <https://orcid.org/0000-0003-1973-711X>

Божко Олександр Юрійович – Харківський національний університет радіоелектроніки, аспірант кафедри інформаційних управляючих систем; Харків, Україна;

Oleksandr Bozhko – Kharkiv National University of Radio Electronics, Postgraduate Student at the Information Control Systems Department; Kharkiv, Ukraine;

e-mail: oleksandr.bozhko@nure.ua

ORCID ID: <https://orcid.org/0009-0004-6820-1228>

METHOD OF AUTOMATIC LABELING FOR SIGNS OF STRUCTURAL COMPLEXITY IN DOCUMENTS USING LARGE LANGUAGE MODELS

The subject of the study is methods of automated analysis and determination of structural complexity features in weakly structured business documents using modern multimodal large language models. The purpose of the work is to develop and experimentally

test the performance of a method for automatic marking of signs of structural complexity in documents, which provides the ability to predict the quality of further data extraction, as well as to build a mathematical model for optimizing the selection of language model configurations in the "quality-cost" criteria space. **The tasks** involve formalizing the taxonomy of features, forming an experimental corpus, developing a unified prompting scheme, and solving the problem of multi-criteria optimization of model selection. **The methodological basis** of the study consists of methods of system analysis, empirical profiling of large language models, prompt engineering methods, methods of mathematical statistics (correlation analysis, calculation of Precision, Recall, F1) for evaluating classification quality, as well as discrete optimization methods for finding compromise solutions. **Results.** The main result of the research is the development of an original method for automatic tagging of document structural complexity features, which allows automatically generating a structured document complexity profile and processing its original visual PDF representation. A weighted quality function is also proposed, which takes into account the degree of influence of each feature on extraction errors, and Pareto-optimal configurations are identified that enable cost minimization given the specified reliability requirements. Based on the results achieved, certain **conclusions** can be drawn. It has been experimentally proven that cost-effective models in example-based mode provide high marking accuracy that competes with the results of significantly more expensive models. It has been established that the use of extended reasoning modes for binary classification tasks is economically unfeasible due to a disproportionate increase in cost without a significant increase in quality. The proposed method solves the problem of the lack of tools for preliminary assessment of complexity in intelligent document processing systems. It ensures the predictability and controllability of extraction processes, enabling adaptive routing of documents depending on their complexity. This creates the basis for building effective industrial systems with balanced accuracy and operating costs, eliminating the need for labor-intensive manual marking.

Keywords: weakly structured document; data extraction; prompt engineering; Pareto optimality; multimodal models.

Бібліографічні описи / Bibliographic descriptions

Петров К. Е., Божко О. Ю. Метод автоматичної розмітки ознак структурної складності документів із використанням великих мовних моделей. *Автоматизовані системи управління та прилади автоматики*. 2026. № 1 (188). С. 54–68. DOI: <https://doi.org/10.30837/0135-1710.2026.188.054>

Petrov, K., Bozhko, O. (2026), "Method of automatic labeling for signs of structural complexity in documents using large language models", *Management Information System and Devices*, No. 1 (188), P. 54–68. DOI: <https://doi.org/10.30837/0135-1710.2026.188.054>

Пилипенко А. Г.

МОДЕЛЬ ДИНАМІЧНОЇ АКТИВАЦІЇ КОРЕЛЯЦІЙНИХ ЗВ'ЯЗКІВ В ОНТОЛОГІЧНИХ ГРАФАХ В УМОВАХ ОПРАЦЮВАННЯ РІЗНОРІДНИХ ДАНИХ

Предмет дослідження – процеси інтерпретації структурних залежностей у корпоративних онтологічних графах в умовах опрацювання різномірних даних (текстових новинних і кількісних технічних сигналів). Описано перехід від статичного трактування графа знань як фіксованого набору зв'язків до динамічної моделі, де структурні відношення є простором потенційних кореляцій. **Мета** статті полягає у формуванні концептуальної та формалізованої моделі умовної активації залежностей у корпоративних онтологічних графах, що дає змогу відокремити сталі топологічні зв'язки від їх ситуативної впливовості в конкретному часовому контексті. **Завдання дослідження:** формалізація структурного шару графа як системи потенційних залежностей; визначення критеріїв і функції активації для переходу залежності в активний стан; розроблення механізму інтеграції подієвих сигналів як тригерів активації; алгоритмічна реалізація моделі в інтелектуальних системах підтримки прийняття рішень. **Методи дослідження** основані на теорії графів, онтологічному моделюванні та семантичному міркуванні (semantic reasoning). Процедурну реалізацію моделі здійснено в межах багатоагентного пайплайну з упровадженням великих мовних моделей (LLM-агентів) для інтерпретації настрою новин і агрегації різномірних ознак без параметричного навчання. **Результати** емпіричної перевірки, проведеної в режимі суворої часової узгодженості (forward-only), підтвердили ефективність моделі. У режимі close→close використання графового контексту забезпечило зростання кумулятивної дохідності з +3.58% до +13.42%, що свідчить про стійкий ефект підсилення сигналу. У режимі close→open досягнуто максимального результату +19.80%, що демонструє домінуючу роль новинно-семантичного компонента в нічному режимі за умови коригувального впливу структури графа. **Висновки.** Запропонована модель дає змогу усунути некоректне отождошення наявності структурного зв'язку з фактичним впливом. Установлено, що інтеграція формалізованого механізму активації (на основі структурного шляху до 2-го, релевантного сигналу та часової валідності) забезпечує систематичне покращення якості прогнозів у задачах опрацювання різномірних даних.

Ключові слова: модель активації; граф знань; різномірні дані; LLM-агенти; інформаційні системи; прийняття рішень.

Постановка проблеми

У сучасних дослідженнях онтологічні графові моделі відіграють ключову роль у моделюванні складних систем, де зв'язки між об'єктами змінюються в часі (наприклад, у соціальних мережах, біологічних системах, фінансових ринках), а також у корпоративних інформаційних системах і системах підтримки прийняття рішень [1–3]. Наявні підходи до аналізу часового графового подання спрямовані на вивчення структури графа та її впливу на подання вузлів. Але ці підходи зазвичай трактують зв'язки як активні постійно або не формалізують умови їх динамічної активації в контексті зовнішніх подієвих сигналів.

У корпоративних онтологічних графах структурні зв'язки (наприклад, "постачальник – клієнт" або "фірма – конкурент") традиційно розглядаються як статичні або як апіорно задані для моделювання впливу в задачах прогнозування. Але в реальних економічних системах наявність

структурного зв'язку не гарантує його фактичного впливу в будь-який момент часу [4]. Вплив конкретної кореляційної залежності може активуватися лише за наявності певних умов або подій.

Невідповідність між фіксованою структурою графа знань і динамічною актуальністю впливу зв'язків створює фундаментальну проблему: відсутність формального механізму, що відокремлює потенційний структурний зв'язок від його часової активації як впливового фактора в конкретному контексті або моменті часу. Наявні дослідження динамічних графів розглядають зміну топології чи часові "скріншоти" графа, але не забезпечують формальної умовної активації залежностей на основі сигнальних і часових критеріїв. У практичних задачах, наприклад, для прогнозування напрямку руху цін або динамічної оцінки ризиків необхідне моделювання, що бере до уваги не лише "наявність" зв'язків, але й умови їх наявного активаційного впливу. Подібна потреба інтеграції різномірних джерел інформації спостерігається також

у дослідженнях фінансового прогнозування, де поєднання технічних, фундаментальних і текстових даних сприяє підвищенню якості інтерпретації ринкових сигналів [5]. Реалізація такого механізму дає змогу більш точно подавати події й часові контексти, що визначають зміну станів у складних системах, які описуються графами знань. Використання подієвих графів знань для опису процесів і взаємодій між елементами складних систем також висвітлюється в сучасних працях [6].

Отже, актуальність дослідження зумовлена необхідністю формалізації механізмів динамічної активації кореляційних залежностей у корпоративних онтологічних графах і розроблення моделей, здатних брати до уваги часові й сигнальні умови, що визначають фактичну дію зв'язків у конкретні моменти часу. Це має важливе значення для побудови більш точних і адаптивних моделей і систем підтримки прийняття рішень, що інтегрують структурні знання й методи інтелектуального аналізу даних [3, 7].

Аналіз останніх досліджень і публікацій

Як було зазначено, динамічні графи як моделі змінюваних відношень між об'єктами досліджуються в різних напрямках сучасної науки. У класичних підходах до аналізу часових графів широко застосовуються моделі на основі графових мереж із часовим компонентом, а також підходи до міркування в темпоральних графах знань (*temporal knowledge graph reasoning*), які беруть до уваги часові залежності між сутностями та їх відношеннями [4, 8, 9]. Поряд із дослідженнями темпоральних і динамічних графів у сучасних роботах активно розвивається напрям онтологічно керованих графових моделей, де особливу увагу приділено семантичній узгодженості подання сутностей і відношень. Такі підходи розглядають онтологію як концептуальний рівень інтеграції знань, що забезпечує формалізацію змісту зв'язків, уніфікацію даних з різномірних джерел і підвищення інтерпретованості аналітичних процедур у складних інформаційних системах [10, 11]. Подібну інтерпретацію ролі графів знань як інфраструктури аналітичного оброблення даних також висвітлено в сучасних дослідженнях *knowledge graph-driven* підходів до підтримки прийняття рішень, де наголошено на значенні семантично узгодженого подання зв'язків для побудови інтелектуальних

аналітичних систем [12]. Так, у вивченні графових моделей значну увагу приділено побудові багаторівневого подання графів знань і використанню структурної інформації різних рівнів для задач міркування й аналізу залежностей між сутностями [13]. Окремим напрямом досліджень є аналіз структурних залежностей у графових моделях даних з огляду на часові обмеження. Зокрема в роботах, присвячених темпоральним графовим функціональним залежностям (*temporal graph functional dependencies*), залежності між сутностями розглянуто відповідно до часових обмежень та еволюції графових даних [14]. Однак такі підходи здебільшого спрямовані на виявлення й перевірку узгодженості даних у часових графах, а не на формалізацію умовної активації впливу структурних зв'язків у конкретному контексті подій.

У подібному контексті в низці студій запропоновано подієво-орієнтовані підходи до використання графів знань, у межах яких події є механізмами інтерпретації або уточнення взаємозв'язків між сутностями. Зокрема в роботі [15] запропоновано підхід до побудови подієво-орієнтованих графів знань для аналізу складних виробничих систем, де події – це тригери для інтерпретації залежностей між елементами системи. Подієві графи знань також застосовуються для моделювання інженерних процесів і взаємодій між їх елементами [6]. Проте подібні моделі здебільшого орієнтовані на аналіз причинно-наслідкових зв'язків або діагностику подій, тоді як задача формалізації умовної активації структурних залежностей у графі знань у контексті зовнішніх сигналів залишається недостатньо вивченою.

У розвитку досліджень графових моделей наступним концептуальним кроком стало формування напряму онтологічно керованих графових моделей, де структурні зв'язки розглядаються не лише як топологічні елементи графа, але і як носії формалізованого семантичного змісту. Це дає змогу розмежувати опис потенційних залежностей між сутностями й механізми їх інтерпретації в аналітичних задачах. У дослідженнях з онтологічно керованої інтеграції бізнес-даних наголошено, що використання онтологій сприяє гнучкості побудови аналітичних моделей та підвищує прозорість і відтворюваність процесів підтримки прийняття рішень у багатоджерельних інформаційних середовищах, що особливо важливо для побудови

пояснюваних аналітичних моделей на основі графів знань [10–12].

Автори студій, які розглядають корпоративні графи знань, найчастіше трактують структурні відношення як фіксовані й завжди потенційно активні для задач прогнозування. Такі моделі не забезпечують явного механізму умовної активації – наявність зв'язку еквівалентна його впливовості в будь-який момент. Якщо говорити про часове графове навчання (*temporal graph learning*), то існують підходи, що використовують події як тригери для зміни ваг або активації підграфів. Але ці підходи орієнтовані на внутрішню еволюцію графа, а не на інтеграцію зовнішніх подієвих сигналів як умов, що активують кореляційні залежності.

Отже, сучасні дослідження демонструють, що поєднання графових моделей з онтологічним концептуальним рівнем створює методологічні передумови для формалізованого опису умов інтерпретації структурних залежностей між сутностями. Саме онтологічно визначений підхід сприяє розмежуванню потенційної наявності зв'язку й умов його фактичної актуалізації в конкретному подієвому контексті.

Попри значну кількість робіт у галузі динамічних графів, часових залежностей і графових знань, завдання формалізації механізмів умовної активації структурних залежностей у корпоративних онтологічних графах на основі зовнішніх подієвих сигналів залишається недостатньо вивченим. Саме ця нерозв'язана частина визначає предмет цього дослідження, у якому онтологічна модель розглядається як концептуальна основа формалізованого опису структурних залежностей, що дає змогу відокремити їх потенційну наявність від умов фактичної активації в конкретному подієвому контексті.

Визначення мети й завдань дослідження

Метою статті є розроблення моделі умовної активації залежностей у корпоративних онтологічних графах у процесі опрацювання різномірних даних, яка дає змогу: інтерпретувати структурні відношення як потенційні залежності; формально визначати умови їх динамічної активації; розмежовувати постійні структурні зв'язки та їх контекстно-часову впливовість; інтегрувати подієві сигнали як тригери активації залежностей; забезпечувати можливість

алгоритмічної реалізації механізму активації в інформаційних системах.

Для досягнення поставленої мети необхідно виконати такі завдання:

- 1) формалізувати структурний шар корпоративного онтологічного графа як систему потенційних залежностей із типізованими відношеннями;
- 2) визначити набір умов і функцію активації, за допомогою яких потенційна залежність переходить у стан активної;
- 3) побудувати формальний механізм функції активації залежностей;
- 4) показати можливість інтеграції моделі в системи прогнозування, що працюють із різномірними даними.

Запропонований підхід демонструє використання онтологічного концептуального рівня як основи для формалізованого аналізу умовної активності структурних залежностей у корпоративних графах знань. Отже, перед авторами статті постає завдання побудови формальної моделі, яка дасть змогу перейти від статичної інтерпретації графа знань до умовно-динамічної інтерпретації залежностей, що активуються подієвими й часовими факторами.

Запропонована модель

Корпоративну модель знань подано у вигляді онтологічного графа, який фіксує множину сутностей (компаній) і типізовані відношення між ними (постачальник, конкурент, партнер, дочірня структура, клієнт тощо). Така модель задає структурний простір корпоративних взаємозв'язків і є статичною: наявність ребра відображає факт існування економічного відношення, однак не визначає його актуальну значущість у конкретний момент часу. У деяких підходах до моделювання динамічних графів розглядається можливість умовної зміни стану або появи зв'язків залежно від контекстних факторів і часових умов [16]. Проте в корпоративних онтологічних графах структурні відношення зазвичай залишаються фіксованими, тоді як змінюється не сама структура, а інтерпретація їх впливу в конкретному контексті. У класичній інтерпретації графових структур наявність зв'язку часто неявно ототожнюється з наявністю залежності, що узгоджується з підходами до формалізації так званих графових функціональних залежностей (*graph functional dependencies*) у дослідженнях

графових даних [17, 18]. Проте в корпоративному середовищі структурний зв'язок і фактичний вплив не є еквівалентними: економічні відносини можуть зберігатися тривалий час, але ставати значущими лише за настання певних подій, інформаційних сигналів або ринкових умов.

У межах запропонованого підходу запропоновано розрізнення між структурним відношенням як елементом онтологічної моделі знань і активною залежністю як динамічним станом цього відношення, що виникає за виконання додаткових умов. Відповідно, онтологічний граф розглядається як простір потенційних залежностей, а не як фіксована система діючих впливів. Подібне тлумачення узгоджується з підходами до формального опису залежностей у графових моделях даних [17]. Таке розмежування також співвідноситься з різницею між структурними й функціональними зв'язками в складних мережах [4]. Перехід від статичної структури до динамічної інтерпретації реалізується через модель умовної активації залежностей, яка формалізує умови, за яких структурний зв'язок набуває статусу активного в заданому часовому контексті. Така інтерпретація дає змогу використовувати граф знань не як механізм прямої передачі сигналів, а як структурне обмеження простору допустимих залежностей у процедурі прийняття рішень, що узгоджується із сучасними підходами до використання графів знань як структурної основи інтелектуальних систем підтримки прийняття рішень [7].

Формально корпоративний онтологічний граф задається як

$$G=(V,E,\tau V,\tau E), \quad (1)$$

де V – множина сутностей (корпоративних вузлів); $E \subseteq V \times V$ – множина типізованих структурних відношень; $\tau E(e)$ – тип відношення (*supplier*, *competitor*, *partner* тощо); $\tau V(v)$ – тип вершини (тип економічного суб'єкта в корпоративному графі).

Кожне структурне відношення сутностей $e \in E$ інтерпретується як потенційна залежність. Наявність відношення не означає фактичного впливу між сутностями. Для пари сутностей (u,v) потенційна залежність існує, якщо $(u,v) \in E$ або є типізований шлях $P_{u \rightarrow v} = (u = v_0, v_1, \dots, v_k = v)$, де $(v_i, v_{i+1}) \in E$. Шлях $P_{u \rightarrow v}$ розглядається як орієнтований ланцюг

структурних відношень графа, у якому кожне ребро $(v_i, v_{i+1}) \in E$ зберігає свій тип $\tau E(e)$. Отже, потенційна залежність відтворює структурну можливість передачі впливу.

Для переходу до динамічної інтерпретації впроваджено сигнальну функцію $S(u,t)$, що відображає наявність релевантного сигналу, пов'язаного із сутністю u у момент часу t . Сигнал може мати довільну природу (текстову, числову або агреговану), модель не накладає обмежень на спосіб його обчислення. Сигнал вважається релевантним, якщо $S(u,t) \neq 0$. Час t_u відповідає моменту виникнення сигналу, пов'язаного із сутністю u , тоді як t_v визначає момент прийняття рішення для цільової сутності v . Додатково впроваджено параметр часової валідності $\Delta t > 0$. Передача впливу від u до v допускається лише за умови часової узгодженості $0 \leq (t_v - t_u) \leq \Delta t$.

Активна залежність визначається функцією $D_{active}(u,v,t_v) \in \{0,1\}$. Значення $D_{active} = 1$ має місце тоді й лише тоді, коли одночасно виконуються три умови:

- 1) існує потенційна залежність між u та v (пряме відношення або структурний шлях);
- 2) існує релевантний сигнал $S(u,t_u)$;
- 3) реалізується умова часової валідності. В усіх інших випадках $D_{active}(u,v,t_v) = 0$.

Отже, структурне відношення є статичним елементом моделі знань, тоді як активна залежність зумовлена його динамічним станом, що залежить від сигналу й часу. Модель є абстрактною та відокремленою від конкретних алгоритмів аналізу даних.

Процедурна реалізація моделі здійснюється в межах фіксованого багатоагентного LLM-пайплайну без параметричного навчання. Формування сигналів відбувається у два етапи.

По-перше, для кожної новинної статті, що стосується компанії u у день t , застосовується агент аналізу тональності (*Sentiment Agent*), який генерує стислий опис події та категоріальний сентимент. Далі агент-агрегатор (*Aggregator Agent*) об'єднує всі статті за день, усуває дублікати, формує узагальнений підсумок, підсумковий сентимент, попередній

напрямок прогнозу й рівень впевненості. Як наслідок, формується агрегований новинний сигнал $S_{news}(u, t)$.

По-друге, формується квантитативний (технічний) сигнал $S_{TA}(u, t)$. Технічний агент отримує структурований словник ознак $X(u, t)$, розрахованих до моменту прогнозування, нормалізованих, дискретизованих, приведених до фіксованих доменів і розділених на функціональні групи; подібне поєднання різнорідних фінансових ознак, зокрема технічних, фундаментальних і текстових, використовується також у задачах прогнозування ринкової динаміки [5]. Агент працює за жорстко заданою нормативною специфікацією, яка визначає правила інтерпретації вхідних ознак і узгодження сигналів різних груп. У межах цієї процедури перевіряється логічна узгодженість ознак, попередньо визначається напрямок сигналу, а також оцінюється рівень впевненості. У разі суперечливих або недостатніх сигналів система допускає нейтральне рішення (*Neutral*).

LLM у цьому процесі не здійснює числових обчислень, а виконує інтерпретаційну процедуру відповідно до формалізованого протоколу семантичного міркування (*semantic reasoning*).

На етапі визначення активних залежностей для цільової компанії v у момент прийняття рішення t_v з графа знань вилучається підграф завглибшки до двох кроків (*2-hop neighborhood*), який містить всі сутності u , пов'язані з v безпосередньо або через один проміжний вузол. Кожен *2-hop*-ланцюжок подано як типізовану послідовність $P_{u \rightarrow v}$ довжини $k \leq 2$. Для кожної сутності u перевіряється наявність потенційної залежності, релевантного сигналу $S(u, t_u)$ та реалізації умов часової валідності. За виконання цих умов залежність вважається активною та $D_{active}(u, v, t_v) = 1$. Обмеження глибини до *2-hop* зумовлено обмеженням контекстного вікна LLM, припущенням про згасання впливу зі збільшенням довжини ланцюжка й необхідністю збереження інтерпретованості процесу семантичного міркування.

Для оцінювання внеску структурного контексту реалізовано два варіанти фінальної агрегації.

У варіанті AGG фінальний агент отримує лише агрегований новинний і технічний сигнали цільової компанії; графовий контекст не використовується.

У варіанті GRAPH додатково передаються *2-hop*-ланцюжки зв'язків, типи відношень і агреговані сигнали пов'язаних компаній. Граф у цьому разі не генерує сигнал безпосередньо, а є структурним обмеженням простору потенційних залежностей, визначаючи допустимі шляхи передачі впливу між сутностями.

Експериментальне оцінювання проводилося в режимі тільки вперед (*forward-only*) без використання майбутньої інформації. Для кожного торгового дня t сигнали формувалися на основі даних, доступних на момент закриття дня, а прогноз формувався на період $t+1$ у режимах *Close-to-Close* й *Close-to-Open*. Для кожного тикера генерувався один із сигналів *Up*, *Down* або *Neutral*. Потім формувався рівноваговий портфель за всіма активними сигналами без огляду на транзакційні витрати.

Основними метриками були кумулятивна дохідність, середня денна дохідність, коефіцієнт Шарпа (*Sharpe ratio*) й частка прибуткових угод. Додатково досліджувався розподіл прибутковості за рівнями *confidence*, чутливість результатів до режимів *Close-to-Close* та *Close-to-Open* і різниця між AGG і GRAPH. Точність прогнозу не розглядалася як базова метрика, оскільки задача інтерпретується як задача формування торгового рішення.

Обмеженнями експерименту можна вважати статичність графа знань, обмеження глибини структурного контексту до двох кроків (*2-hop*), ігнорування транзакційних витрат і параметричного навчання, а також використання однієї LLM-моделі без тонкого налаштування (*fine-tuning*).

Для експерименту порівняно три підходи до формування торгового сигналу на горизонт $T+1$: агента технічного аналізу (ТА), моделі агрегації новин і технічних ознак без застосування графа (AGG) і моделі з огляду на структурний графовий контекст (GRAPH).

Агент технічного аналізу формує сигнал без параметричного навчання, використовуючи фіксований алгоритм семантичного маркування поверх попередньо розрахованих технічних ознак. На відміну від класичних підходів, індикатори не застосовуються як механічні тригери або входи моделі машинного навчання, а подаються в нормалізованій і дискретизованій формі, адаптованій для інтерпретації LLM. Ознаки структуровано за функціональними групами (тренд, імпульс, діапазони / пробої, обсяг, волатильність тощо), кожна з яких має фіксовану роль у процедурі

прийняття рішення. Для уникнення подвійного врахування корельованих індикаторів використовується правило вибору одного репрезентативного сигналу в межах сімейства, а шумові значення відсікаються через зони нейтральності.

Ключовою властивістю агента ТА є консервативність за замовчуванням: у разі суперечностей між групами ознак або слабкої

роздільності сигналів повертається прогноз *Neutral* з низьким рівнем упевненості. Це забезпечує інтерпретованість, відтворюваність і стійкість до випадкових флуктуацій.

У режимі стандартної денної дохідності $close \rightarrow close$ $T+1$ отримано такі значення кумулятивної дохідності рівновагового портфеля (рис. 1): TA -4.35% ; AGG $+3.58\%$; GRAPH $+13.42\%$.

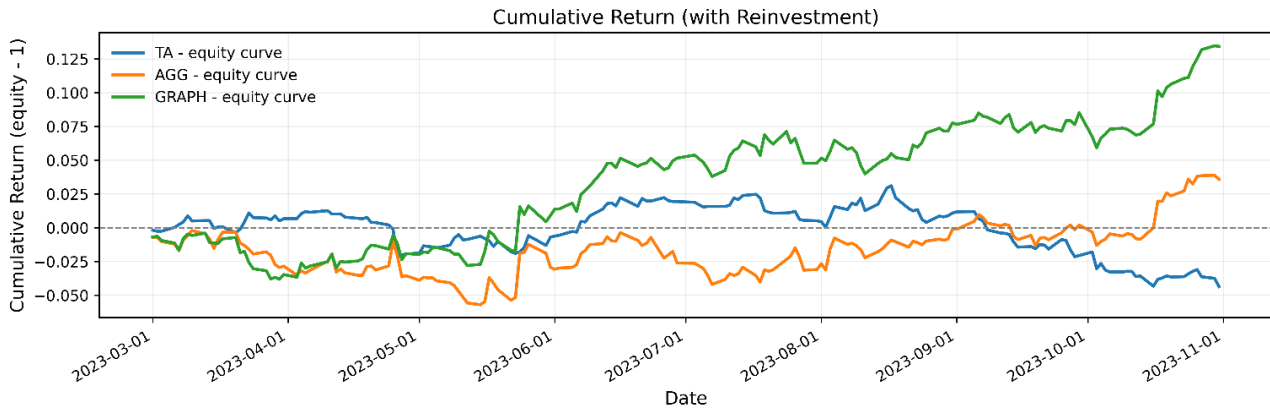


Рис. 1. Кумулятивна дохідність рівновагового портфеля в режимі $close \rightarrow close$

Агент технічного аналізу (ТА) демонструє від'ємну дохідність упродовж більшої частини періоду, що свідчить про обмежену ефективність ізольованих технічних сигналів у певному ринковому режимі. Стратегія AGG, яка поєднує семантичну інтерпретацію новин із технічним сигналом, забезпечує помірно позитивний результат, однак її кумулятивна дохідність залишається відносно низькою та супроводжується підвищеною варіативністю.

Додавання графового контексту призводить до суттєвого покращення результатів: стратегія

GRAPH досягає кумулятивної дохідності $+13.42\%$, що приблизно на 10 відсоткових пунктів перевищує AGG. Важливо, що ця перевага не є наслідком поодиноких екстремальних днів. Аналіз траєкторії кумулятивної дохідності свідчить про поступове й відносно стабільне накопичення результату, що вказує на стійкість ефекту інтеграції структурного контексту.

У режимі $close \rightarrow open$ отримано такі значення кумулятивної дохідності рівновагового портфеля (рис. 2): TA -1.64% ; AGG $+17.82\%$; GRAPH $+19.80\%$.

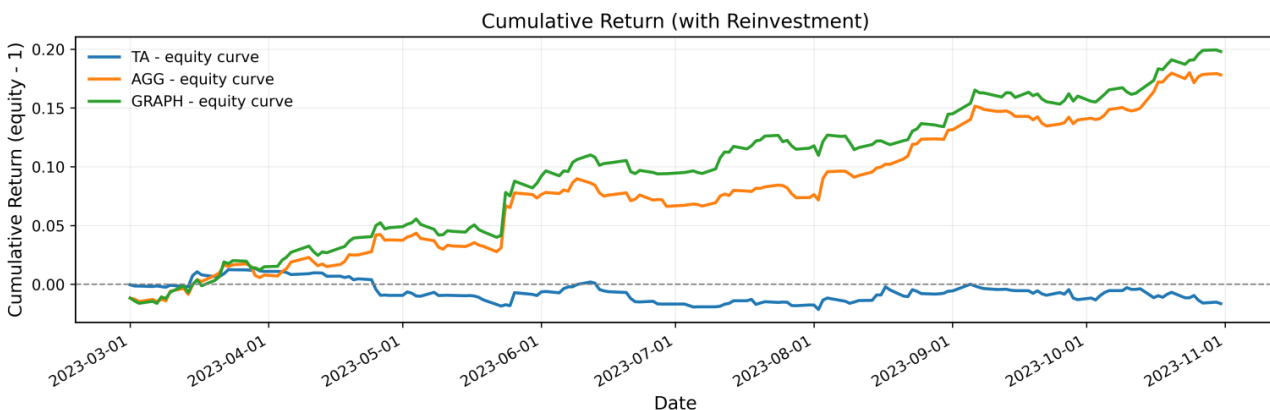


Рис. 2. Кумулятивна дохідність рівновагового портфеля в режимі $close \rightarrow open$

Результати демонструють суттєву перевагу стратегій, що використовують новинний контекст

(AGG і GRAPH), над чистим технічним аналізом і ринковим бенчмарком. Це свідчить про високу

чутливість режиму нічного цінового розриву до семантичної інтерпретації новин. Динаміка кумулятивної дохідності свідчить про близькість траєкторій AGG і GRAPH протягом більшої частини періоду. Підсумкова різниця між ними становить приблизно +2 відсоткових пунктів, що є істотно меншою, ніж у режимі *close*→*close*. Отже, у нічному режимі основним джерелом сигналу є новинно-семантичний компонент, тоді як графовий контекст відіграє коригувальну роль, уточнюючи базовий сигнал.

Висновки

У статті запропоновано модель умовної активації залежностей, її процедурну реалізацію в LLM-архітектурі та протокол експериментальної оцінки в режимі суворої часової узгодженості (*forward-only*, $T+1$). Модель спрямована на дослідження механізму динамічної інтерпретації структурних залежностей у корпоративному онтологічному графі.

Запропоновано зміну підходу до інтерпретації графа знань: від статичної семантики до динамічної моделі. Тут граф є не джерелом безпосереднього сигналу, а простором потенційних залежностей. Активація зв'язків відбувається лише за виконання формально визначених умов, що усуває некоректне ототожнення наявності ребра з наявністю фактичного впливу. Такий підхід забезпечує чітке розмежування структурного й динамічного рівнів [4]. Структурний шар онтологічного графа залишається фіксованим, тоді як контекстний шар сигналів має часову природу. Модель умовної активації виконує роль формального механізму переходу між цими рівнями.

Було впроваджено строгий критерій активації, оснований на одночасному виконанні трьох умов: наявності структурного шляху (до *2-hop*), релевантного сигналу й часової валідності. Така формалізація забезпечує відтворюваність, інтерпретованість і відокремленість моделі від конкретних алгоритмів обчислення сигналів.

Результати проведених експериментів підтверджують практичну значущість структурного контексту. У режимі *close*→*close* варіант GRAPH продемонстрував кумулятивну дохідність +13.42% проти +3.58% у варіанті AGG та -4.35% у технічного агента (TA), що означає приріст приблизно на 10 відсоткових пунктів щодо моделі без графового

контексту. У режимі *close*→*open* перевага GRAPH над AGG зберігається (+19.80% проти +17.82%), хоча її масштаб і зменшується приблизно до двох пунктів, що свідчить про домінування новинно-семантичного компонента в нічному інформаційному режимі. В обох випадках TA демонструє від'ємний або нестійкий результат (-1.64% та -4.35%), що свідчить про обмеженість ізольованих технічних сигналів у досліджуваній постановці.

Загалом досягнуті результати підтверджують, що інтеграція формалізованого механізму умовної активації в межах корпоративного онтологічного графа забезпечує систематичне покращення якості сигналу порівняно з моделлю без структурного контексту. Внесок графового шару насамперед полягає в стабілізації та посиленні новинно-технічного сигналу, що свідчить про доцільність подальших досліджень у напрямі динамічних моделей інтерпретації структурних залежностей.

Конфлікт інтересів

Автор статті декларує, що не має конфлікту інтересів, зокрема фінансового, особистого, авторського чи будь-якого іншого характеру, який міг би вплинути на дослідження, а також на результати, опубліковані в цій роботі.

Фінансування

Дослідження проводилося без фінансової підтримки.

Доступність даних

Дані будуть надані за обґрунтованим запитом.

Використання засобів штучного інтелекту

Під час підготовки цього рукопису автор застосовував модель *Gemini 3 Flash (Thinking mode)*. За допомогою ШІ виконано стилістичне редагування авторського тексту й структуровано попередньо зібрані тези. Крім цього, штучний інтелект використано на етапі збору інформації для пошуку актуальних джерел у відкритих наукових базах даних. У процесі пошуку інформації інструмент застосовувався для ідентифікації ключових тем і переліку публікацій, які згодом автор опрацьовував

вручну. Усі згенеровані мовні виправлення ретельно перевірено на предмет збереження оригінального змісту. Фактологічна інформація та посилання, знайдені за допомогою ШІ, верифіковано через першоджерела й офіційні бази даних.

Використання штучного інтелекту жодним чином не вплинуло на інтерпретацію інформації, формування наукової гіпотези або кінцеві висновки дослідження. Весь науковий зміст і аналіз результатів є продуктом інтелектуальної праці автора.

References

1. Wu, Z., Zhang, H. (2025), "Research on the construction of dynamic knowledge graph and intelligent decision-making for enterprise human resource information based on federated hyper graph neural network", *Discover Applied Sciences*, Vol. 7, No. 1. DOI: <https://doi.org/10.1007/s42452-025-08006-w>
2. Wang, T. (2026), "A causal discovery-based adaptive fusion algorithm for multi-source heterogeneous knowledge graphs", *Scientific Reports*, Vol. 16. DOI: <https://doi.org/10.1038/s41598-025-34507-0>
3. Wang, S., Yang, H., Bai, G. (2025), "Construction of intelligent decision support systems through integration of retrieval-augmented generation and knowledge graphs", *Scientific Reports*, Vol. 15. DOI: <https://doi.org/10.1038/s41598-025-19257-3>
4. Voutsas, V., Van Mieghem, P., Kooij, R. (2021), "Two classes of functional connectivity in dynamical processes in networks", *Journal of the Royal Society Interface*, Vol. 18, No. 181. DOI: <https://doi.org/10.1098/rsif.2021.0486>
5. Zhong, S., Hitchcock, D. (2021), "S&P 500 stock price prediction using technical, fundamental and text data", *Statistics, Optimization & Information Computing*, Vol. 9, P. 769–788. DOI: <https://doi.org/10.19139/soic-2310-5070-1362>
6. Chen, G., Auwal, A., Chen Y., Li, L., Khandaker N., Yongbo L., Khalid E., (2025), "Event Knowledge Graph for a Knowledge-Based Design Process Model for Additive Manufacturing", *Machines*, Vol. 13. 112. DOI: <https://doi.org/10.3390/machines13020112>
7. Napoli, R., Morabito, G., Celesti, A., Villari, M., Fazio, M. (2025), "Knowledge Graphs-Driven Intelligence for Distributed Decision Systems", *Proceedings of the 18th IEEE/ACM International Conference on Utility and Cloud Computing (UCC 2025)*, P. 1–10. DOI: <https://doi.org/10.1145/3773274.3774268>
8. Feng, S., Zhou, C., Liu, Q., Ji, X., Huang, M. (2024), "Temporal Knowledge Graph Reasoning Based on Entity Relationship Similarity Perception", *Electronics*, Vol. 13, No. 12. DOI: <https://doi.org/10.3390/electronics13122417>
9. Zeng, Y., Hou, X., Wang, X., Li, J. (2025), "Towards a Unified Temporal and Event Logic Paradigm for Multi-Hop Path Reasoning in Knowledge Graphs", *Electronics*, Vol. 14, No. 3, P. 516. DOI: <https://doi.org/10.3390/electronics14030516>
10. Roman, D., Alexiev, V., Paniagua, J., Elvæsæter, B., von Zernichow, B. M., Soylu, A., Simeonov, B., Taggart, C. (2022), "The euBusinessGraph ontology: A lightweight ontology for harmonizing basic company information", *Semantic Web*, Vol. 13, No. 1, P. 41–68. DOI: <https://doi.org/10.3233/SW-210424>
11. Blankenberg, C., Gebel-Sauer, B., Schubert, P. (2022), "Using a graph database for the ontology-based information integration of business objects from heterogenous Business Information Systems", *Procedia Computer Science*, Vol. 196, P. 314–323. DOI: <https://doi.org/10.1016/j.procs.2021.12.019>
12. Dey, L. (2024), "Knowledge graph-driven data processing for business intelligence", *WIREs Data Mining and Knowledge Discovery*, Vol. 14, e1529. DOI: <https://doi.org/10.1002/widm.1529>
13. Shang, Z., Wang, P., Ke, W., Liu, J., Huang, H., Li, G., Wu, C., Liu, J., Chen, X., Li, Y. (2025), "Learning Multi-Granularity and Adaptive Representation for Knowledge Graph Reasoning", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 37. DOI: <https://doi.org/10.1109/TKDE.2025.3579774>
14. Noronha, J., Fan, W., Geerts, F., Lu, J. (2021), "Discovery of Temporal Graph Functional Dependencies", *Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM)*, P. 1366–1375. DOI: <https://doi.org/10.1145/3459637.3482087>
15. Cen, W., Su, C., Chen, G., Xie, L. (2025), "Event-driven causal knowledge graphs for bottleneck root cause analysis in manufacturing systems", *Expert Systems with Applications*, Vol. 259. DOI: <https://doi.org/10.1016/j.eswa.2025.129117>
16. Goetz, M., Humm, B. (2025), "Process Anomaly Detection in Cyber-Physical Production Systems Based on Conditional Discrete-Time Dynamic Graphs", *Applied Sciences*, Vol. 15, No. 21. DOI: <https://doi.org/10.3390/app152111354>
17. Fan, W., Wu, Y., Xu, J. (2016), "Functional dependencies for graphs", *Proceedings of the 2016 International Conference on Management of Data (SIGMOD)*, P. 1843–1857. DOI: <https://doi.org/10.1145/2882903.2915232>
18. Wenfei, F., Chunming, H., Xueli, L., Ping, L., (2020), "Discovering Graph Functional Dependencies", *ACM Transactions on Database Systems*, Vol. 45. DOI: <https://doi.org/10.1145/3397198>

Received (Надійшла) 09.02.2026

Accepted for publication (Прийнята до друку) 02.03.2026

Publication date (Дата публікації) 12.03.2026

Відомості про авторів / About the Authors

Пилипенко Андрій Геннадійович – Харківський національний університет радіоелектроніки, аспірант кафедри електронних обчислювальних машин; Харків, Україна;

Andrii Pylypenko – Kharkiv National University of Radio Electronics, Postgraduate Student of Electronic Computers Department; Kharkiv, Ukraine;

e-mail: andrii.pylypenko@nure.ua

ORCID ID: <https://orcid.org/0009-0008-2861-7906>

MODEL OF DYNAMIC ACTIVATION OF CORRELATION RELATIONS IN ONTOLOGICAL GRAPHS WHEN PROCESSING HETEROGENEOUS DATA

Subject of the study is the processes of interpreting structural dependencies in corporate ontological graphs when processing heterogeneous data (textual news and quantitative technical signals). The paper considers the transition from a static interpretation of the knowledge graph as a fixed set of connections to a dynamic model, where structural relations act as a space of potential correlations. The **goal** of the article is to form a conceptual and formalized model of conditional activation of dependencies in corporate ontological graphs, which allows separating stable topological connections from their situational influence in a specific time context. The **research tasks** include: formalization of the structural layer of the graph as a system of potential dependencies; determination of the criteria and activation function for the transition of the dependency to the active state; development of a mechanism for integrating event signals as activation triggers; algorithmic implementation of the model in intelligent decision-making support systems. The **research methods** are based on graph theory, ontological modeling, and semantic reasoning. The procedural implementation of the model is carried out within a multi-agent pipeline using large language models (LLM agents) to interpret news sentiment and aggregate heterogeneous features without parametric learning. The **results** of empirical testing conducted in the forward-only mode confirmed the effectiveness of the model. In the close→close mode, the use of the graph context provided an increase in cumulative profitability from +3.58% to +13.42%, which indicates a stable signal amplification effect. In the close→open mode, the maximum result of +19.80% was obtained, which demonstrates the dominant role of the news-semantic component in the night mode under the corrective influence of the graph structure. **Conclusions.** The proposed model allows eliminating the incorrect identification of the presence of a structural connection with the actual impact. It is established that the integration of a formalized activation mechanism (based on the structural path to 2-hop, relevant signal, and temporal validity) provides a systematic improvement in the quality of predictions in heterogeneous data processing problems.

Keywords: activation model; knowledge graph; heterogeneous data; LLM agents; information systems; decision making.

Бібліографічні описи / Bibliographic descriptions

Пилипенко А. Г. Модель динамічної активації кореляційних зв'язків в онтологічних графах в умовах опрацювання різномірних даних. *Автоматизовані системи управління та прилади автоматики*. 2026. № 1 (188). С. 69–717. DOI: <https://doi.org/10.30837/0135-1710.2026.188.069>

Pylypenko, A. (2026), "Model of dynamic activation of correlation relations in ontological graphs when processing heterogeneous data", *Management Information System and Devices*, No. 1 (188), P. 69–77. DOI: <https://doi.org/10.30837/0135-1710.2026.188.069>

Vitalii Tkachov

METHOD OF RESTORING CRITICAL INFORMATION SYSTEM SERVICES ON A MOBILE PLATFORM UNDER CONDITIONS OF CONTROLLED DEGRADATION

The subject of this work is methods for restoring critical services of an information system on a mobile platform under intermittent connectivity and controlled degradation. *The objective of the study* is to develop a service-oriented recovery method that ensures a measurable reduction in the shortfall in meeting service quality threshold requirements within connectivity windows, taking into account resource constraints and the ability to complete service state reconciliation within the current connectivity window. The study used the following **methods**: system analysis, mathematical modeling of service states and profiles, formalization of the shortfall in meeting service quality threshold requirements, a predicate description of the window feasibility of recovery steps, and a computational experiment to evaluate recovery performance metrics. The article aimed to solve the following **tasks**: formalize a recovery step as a transition between service profiles that reduces the shortfall in meeting service quality threshold requirements; introduce a condition for the window feasibility of a recovery step as a combination of resource feasibility and the possibility of completing service state coordination within the current connectivity window; to construct a rule for forming a set of recovery steps within a window that ensures a reduction in the deficit provided at least one step is executed and is consistent with permissible managed degradation actions as a peripheral resource release mechanism. **Achieved results** consist in the implementation of a window feasibility formalism and a system of metrics for evaluating compliance with feasibility conditions, the proportion of windows with deficit reduction, the cumulative deficit of critical services, and recovery time, which makes it possible to distinguish environmental constraints from the properties of control decisions. **Conclusions**: Ignoring the possibility of completing service state coordination within the current connectivity window leads to the emergence of incomplete recovery transitions and discrepancies between the expected recovery effect and the actual deficit dynamics, whereas the proposed method identifies the steps taken within the current connectivity window and ensures a reduction in the deficit of critical services in windows where recovery is possible. This enhances survivability by reducing the duration of degradation episodes of critical functions and decreasing the cumulative deficit. The practical value lies in the possibility of using the method for onboard information systems of UAV groups and other mobile nodes with connectivity windows of variable duration and bandwidth.

Keywords: information system; survivability; critical service; controlled degradation; connectivity window; service recovery.

1. Introduction

Information systems on mobile platforms (ISMP) operate in environments where available computing, power, and network resources are subject to fluctuations in availability, and connectivity between system components is intermittent [1, 2]. Such MISPs are characterized by short operational control intervals, connectivity windows of variable duration and bandwidth, as well as non-stationary load and disruptive influences [3]. Under these conditions, critical functions of ISMPs are implemented by services, the requirements for which are formulated in terms of service quality metrics, and violations of these requirements manifest as a failure to meet service quality thresholds.

The survivability of an information system on a mobile platform is understood as an integral property consisting of the system's ability to maintain or restore the execution of its primary function in real time despite resource shortages, intermittent connectivity, and incomplete observability of states [4–6]. In the context of mobile platforms, ensuring this resilience is closely

linked to risk-oriented management and robust task distribution under dynamically changing operating conditions [7, 8]. At the operational management level, this means that in each short control interval, the ISMP must ensure the minimum acceptable level of performance for the critical service core and, following a degradation episode, perform recovery transitions of critical services to profiles that ensure the fulfillment of threshold requirements for service quality metrics.

Situations of failure in functional capability that are significant for mobile environments often take the form of controlled degradation, where the ISMP maintains operability by reducing the functionality of some non-critical services or by simplifying data processing, thereby freeing up resources to maintain the minimum acceptable level of operability of the critical service core. However, after the degradation episode ends, a separate task of restoring critical services arises, which cannot be reduced to a general redistribution of resources [9]. Restoration requires the allocation of resources and the completion of service state synchronization within available connectivity windows. If reconfigurations are

initiated without verifying the possibility of completing service state synchronization within the current connectivity window, incomplete recovery transitions occur, the volume of unfulfilled state synchronization accumulates, repeated violations of service quality threshold requirements are recorded after partial recovery, and the time to return to an operational state increases.

An additional complication is that the recovery of critical services often relies on auxiliary actions of controlled degradation of non-critical components and data [10–12]. Such actions must remain permissible under correctness and consistency invariants; otherwise, the short-term release of resources leads to long-term consequences in the form of data degradation, state conflicts, or cascading secondary failures. Therefore, it is relevant to develop a recovery method that operates at the service level due to the failure to meet service quality threshold requirements, generates recovery steps only if it is possible to complete service state coordination within available connectivity windows, and aligns with permissible managed degradation actions as a peripheral resource release mechanism.

2. Analysis of recent research and publications

Modern approaches to ensuring the sustainable operation of edge computing systems are evolving toward the integration of mechanisms for adaptation, redundancy, fault tolerance, and coordination between computing and connectivity. A summary of the state of the art is presented in [13], where risk factors characteristic of edge computing are identified, including high node mobility, degraded or intermittent connectivity, and power constraints. A classification of measures to improve resilience is also provided, including adaptive redundancy, monitoring and anomaly detection, joint optimization of computation and connectivity, and multi-level resource coordination.

A separate area consists of studies devoted to self-adaptive mechanisms in edge computing environments. In particular, in [14], the principles of adaptation design are systematized, a taxonomy of adaptive behavior properties is formed, and the limitations of existing approaches are outlined, including the specifics of reactive schemes, insufficient attention to adaptation at the communication infrastructure level, and the limited suitability of centralized control for decentralized network conditions. This directly reinforces the need for methods capable of generating and executing actions

within short connectivity control intervals, implemented via connectivity windows of variable duration and bandwidth, without assuming full synchronization of states in each interval.

In cloud container environments, the field of controlled functionality reduction is actively developing as a means of preserving the availability of critical components during resource shortages. In [15], a cooperative model of controlled functionality degradation is proposed, where the infrastructure operator performs degradation actions based on the criticality of containers. The goal is to maximize the availability of critical services during large-scale failures.

At the same time, the authors of [13–15] largely assume constant availability of control within the data center and do not separately highlight the requirement to complete service state synchronization within available connectivity windows, which is critical for information systems operating on mobile platforms.

For edge computing infrastructures and fog computing environments, methods for recovering from failures of network chains of functions and service routes in resource-constrained multi-layer topologies are important. In [16], post-failure recovery of service function chains in fog networks under multiple and simultaneous failures is considered. Heuristics for rerouting and remapping requests are proposed, taking into account delays and node load. This increases the proportion of restored chains and reduces recovery time without designating separate backup paths. These results are important as an infrastructure foundation; however, they are focused on restoring network links and do not formalize service-oriented recovery steps due to the lack of fulfillment of threshold requirements for quality of service metrics, with control over the ability to complete state synchronization within the current connectivity window. In other words, the survivability metrics of such systems are not formalized.

The study [17] considers the joint control of handover and service migration as a problem of maximizing long-term utility, taking into account service delay and the resource cost of migration. However, this approach formulates the problem as the optimization of system component placement and migration processes. In the task of restoring critical services in a controlled degradation mode, decisions regarding the relocation or reallocation of resources and the preliminary verification of the possibility of completing service state synchronization within the current connectivity window for each recovery transition are fundamental.

Thus, the works [13–17] under consideration cover the general framework of resilience and adaptation of information systems in edge computing environments, infrastructure mechanisms for controlled functionality reduction, post-failure recovery of network links, and the tasks of service migration under mobility. Beyond their immediate focus lies the service-oriented formulation of critical service recovery within the connectivity windows of a mobile platform as a sequence of discrete recovery steps.

For such steps, the possibility of completing service state synchronization within available connectivity windows must be verified, and effectiveness must be assessed by reducing the shortfall in meeting threshold requirements for service quality indicators, taking into account permissible controlled degradation actions as a peripheral tool for freeing up ISMP resources.

3. Research objective and tasks

The objective of the research is to develop a method for restoring critical services of an information system on a mobile platform under controlled degradation conditions, which, in real time under intermittent connectivity and resource constraints, ensures a measurable reduction in the shortfall in meeting service quality threshold requirements.

To achieve this objective, the following tasks have been set:

- formalize the critical service recovery step as a transition between service profiles, leading to a reduction in the shortfall in meeting service quality threshold requirements;
- introduce a window feasibility condition for the recovery step that simultaneously accounts for resource feasibility on the mobile platform and the possibility of completing service state synchronization within the current connectivity window; – develop a rule for forming recovery steps in each connectivity window that guarantees a reduction in the shortfall in meeting service quality threshold requirements provided there is at least one feasible step and does not result in incomplete recovery transitions.

4. Materials and Methods

We consider the set of services S and the subset of critical services $S_c \subseteq S$.

For each service $s \in S$, a finite ordered set of operational profiles is specified:

$$P_s = \{0, 1, \dots, L_s\}, \quad (1)$$

where the elements of the set $0, 1, \dots, L_s$ define the permissible levels of functionality and service quality: 0 corresponds to the minimum permissible level, and L_s to the maximum level. The value L_s determines the number of levels of functionality and service quality available for the service s .

The current service profile at a given point in time t is denoted by $p_s(t) \in P_s$, which is consistent with the definition of the set P_s (1).

For each service, an aggregated service quality metric $q_s(t)$ and a threshold value θ_s are recorded. Subsequently, the shortfall in meeting the threshold requirement for the service quality metric is defined as:

$$d_s(t) = \max\{0, \theta_s - q_s(t)\}, \quad s \in S. \quad (2)$$

If $d_s(t) = 0$, then the requirement is met; if $d_s(t) > 0$, then it is violated. Formula (2) is used as the baseline coordinate of the service state when constructing recovery steps.

To link the quality metric to profiles, a piecewise-constant model of the dependence of the service quality metric [18] on the profile is introduced:

$$q_s(t) = \bar{q}_s(p_s(t)). \quad (3)$$

In (3), $\bar{q}_s(\cdot)$ is a tabular characteristic of the service profile that records the expected quality level for the profile p_s . Subsequently, formula (3) allows us to interpret the change in the profile $p_s(t)$ as a controlled change in $q_s(t)$ and, accordingly, in the deficit $d_s(t)$ according to formula (2).

Intermittent continuity is defined by a sequence of continuity windows

$$W_k = [\tau_k, \tau_k + \Delta_k], \quad k = 1, 2, \dots, \quad (4)$$

where τ_k is the start time of the k -th window; Δ_k is the duration of the k -th window.

The windows (4) define the time intervals within which exchanges for service state coordination are permitted.

The capacity of the n th window (k) is defined as the maximum allowable amount of data for service

state synchronization that can be transmitted within the window:

$$C_k = \int_{\tau_k}^{\tau_k + \Delta_k} b(t) dt, \quad (5)$$

where $b(t)$ is the instantaneous (ideal) channel capacity at the moment t .

Subsequently, the value C_k (5) is used as a constraint on the possibility of completing service state synchronization during the execution of recovery steps within the window (4).

A recovery step is defined as a discrete transition of the service profile one level upward:

$$a \equiv (s, p \rightarrow p+1), \quad s \in S_c, \quad p \in \{0, \dots, L_s - 1\}. \quad (6)$$

Formula (6) specifies that recovery steps are considered specifically for critical services, and the transition is defined between adjacent profiles.

Each step a from equation (6) is associated with:

- a vector of additional platform resources required $\bar{r}(a) \in \mathbb{R}_+^m$;
- the amount of data required to synchronize the state of the services in $g(a) \in \mathbb{R}_+$, which must be

$$\Delta d(a) = \max \left\{ 0, \max \left\{ 0, \theta_s - \bar{q}_s(p) \right\} - \max \left\{ 0, \theta_s - \bar{q}_s(p+1) \right\} \right\}. \quad (8)$$

In (8), the expected reduction in the $d_s(t)$ deficit resulting from the transition between adjacent profiles is determined.

Within the k -th window, a set of recovery steps A_k is selected. Its feasibility is defined by the following total constraints:

$$\sum_{a \in A_k} \bar{r}(a) \leq \bar{R}_k, \quad \sum_{a \in A_k} g(a) \leq C_k. \quad (9)$$

In (9), the first constraint is consistent with the resource part of predicate (7), and the second – with the possibility of completing the coordination of service states within the current capacity-based connectivity window (5). Subsequently, the conditions in (9) are used as the basic rule for the permissible simultaneous execution of multiple recovery steps.

The operational requirement for window-based deficit reduction is formulated as the presence of at least one critical service for which the deficit in meeting the threshold requirements for service quality metrics is reduced within the window. For a set of recovery steps A_k , this requirement is expressed by the condition:

$$\sum_{a \in A_k} \Delta d(a) > 0. \quad (10)$$

transmitted or received to complete state synchronization during the transition.

The window feasibility of the recovery step in the k window is determined by the predicate

$$\varphi(a, k) = 1 \Leftrightarrow \bar{r}(a) \leq \bar{R}_k \wedge g(a) \leq C_k, \quad (7)$$

where \bar{R}_k is the vector of available platform resources at the beginning of the window W_k .

In (7), the inequality $\bar{r}(a) \leq \bar{R}_k$ is evaluated component-by-component. Mechanism (7) is key to filtering out steps for which, within the current connectivity window, it is impossible to complete service state synchronization or the required platform resources are unavailable.

The expected reduction in the deficit for the recovery step (the recovery step effect) is estimated through the reduction in the deficit of fulfilling the threshold requirement (2). For the step $a = (s, p \rightarrow p+1)$, the expected reduction in the deficit is introduced:

Formula (10) is directly related to the definition of $d(a)$ (8) and the deficit $d_s(t)$ (2) and is used as a criterion that the window W_k has ensured an actual reduction in the deficit of critical services. In the context of survivability, condition (10) specifies an operational requirement for recovery as the process of returning the ISMP to the region of admissible states of critical services. In particular, the fulfillment of (10) within connectivity windows means that the ISMP changes its configuration and ensures a measurable reduction in the shortfall of meeting threshold requirements for service quality metrics, i.e., it shortens the intervals during which survivability is achieved solely through controlled degradation.

In the case where the set of recovery steps that can be performed within the current connectivity window according to (7) is empty due to a resource shortage, it is permissible to employ subordinate controlled degradation actions of non-critical components to free up resources. Let B_k be the set of degradation actions in the k -th window; each action $b \in B_k$ is associated with a resource release vector $\bar{u}(b) \in \mathbb{R}_+^m$.

Then the resource constraint in (9) is refined as follows:

$$\sum_{a \in A_k} \bar{r}(a) \leq \bar{R}_k + \sum_{b \in B_k} \bar{u}(b). \quad (11)$$

Formula (11) allows us to formally describe that degradation actions are used solely to ensure the feasibility of recovery and are not considered an independent goal.

To ensure that degradation does not compromise data processing correctness, a predicate of degradation action admissibility is introduced: $\psi(b) = 1$, which is interpreted as the satisfaction of correctness invariants and error constraints. In general form, this is specified by a system of error budgets:

$$\psi(b) = 1 \Leftrightarrow \forall j \in \{1, \dots, J\} : e_j(b) \leq E_j, \quad (12)$$

where $e_j(b)$ is the estimate of the increase in the j -th component of the error or the risk of violating invariants caused by the execution of the degradation action b ; E_j is the permissible limit value of $e_j(b)$.

Formula (12) is used as a constraint on the selection of the set B_k in each window.

5. Research results and their discussion

5.1. Research scenarios and initial experimental conditions

The method is evaluated in a discrete operational cycle with intermittent connectivity, described by a sequence of windows W_k (4) and their capacities C_k (5). Within each window, steps a to restore critical services defined as profile increases (6) may be performed, provided that window feasibility $\varphi(a, k) = 1$ is satisfied according to (7). Service quality is modeled via a profile characteristic $q_s(t) = \bar{q}_s(p_s(t))$ (3), and a violation of the threshold requirement for the service quality indicator is recorded as a deficit $d_s(t)$ (2). When forming the set of recovery steps A_k , aggregate resource constraints and state coordination (9) are used, and the criterion for the actual reduction of the deficit within the window is specified by condition (10).

For the study, a set of scenarios was developed that replicate combinations of resource shortages and window connectivity typical of mobile platforms:

- scenarios with critical connectivity constraints.

The platform's resource vector \bar{R}_k is relatively stable;

however, the windows W_k are short and of variable duration, and capacity C_k varies from window to window. Under these conditions, the key factor is the limit on the amount of data required for state reconciliation, which is specified by the second part of predicate $g(a) \leq C_k$ (7). The goal of this group of scenarios is to verify that the method generates recovery steps only if it is possible to complete service state synchronization within the current connectivity window and ensures a reduction in the deficit provided that at least one step is performed within the current connectivity window;

- scenarios with critical resource constraints. Windows W_k are sufficiently long and C_k generally do not constitute a bottleneck; however \bar{R}_k decrease due to load or degradation of the mobile platform on which the information system operates. In this group, the resource component of predicate (7) and the satisfaction of the aggregate constraint (9) are decisive. Separately, scenarios of a sharp decrease in available resources and scenarios of gradual resource degradation over time are simulated, allowing us to evaluate the method's behavior as the proportion of windows increases in which the set of steps performed within the current connectivity window narrows;

- combined scenarios. At the same time, C_k and \bar{R}_k change, and a shortfall in the threshold requirements for service quality metrics occurs in a cascading manner as a result of a degradation episode. Such modes are closest to real-world operation: within connectivity windows, only small amounts of state coordination are possible, and within platform windows, the platform has limited resources for performing recovery transitions. For this group of scenarios, the fulfillment of condition (10) is verified without the accumulation of unfinished recovery transitions, as well as the reduction of the integral deficit of critical services, determined based on $d_s(t)$ (2) and $\Delta d(a)$ (8).

Figure 1 shows the implementation of window-based connectivity for a mobile platform: the intervals W_k have variable durations and are spaced apart in time, and the capacity C_k varies significantly between windows, reflecting the non-stationarity of throughput within the connectivity windows. It is precisely this variability that imposes operational constraints on service state synchronization and necessitates verifying the feasibility of completing service state synchronization within each current connectivity window in accordance with (5) and (7).

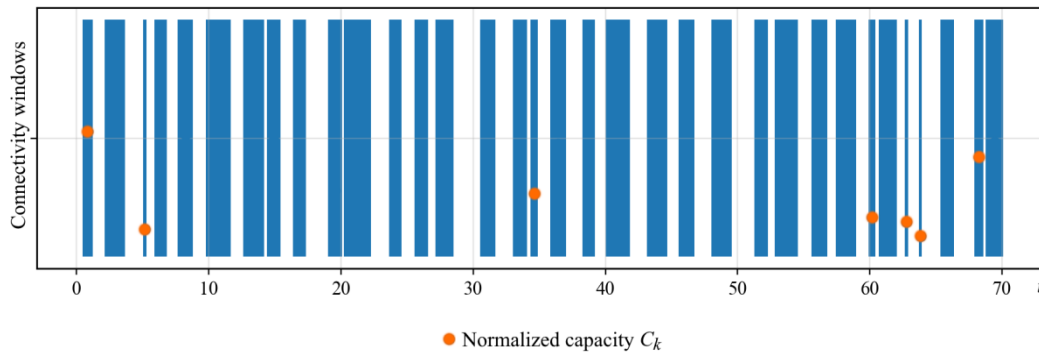


Fig. 1. Connectivity windows W_k and normalized capacity C_k for service state coordination

The initial conditions for all scenarios are defined as the state following a degradation episode. At the start of the simulation, at least one critical service is in a profile $p_s(0)$ for which the deficit $d_s(0)$ (2) is positive, indicating a violation of the service quality threshold requirement. Non-critical services can be transferred to degraded profiles, which is interpreted as a controlled degradation that has already been performed, freeing up part of the resource to maintain the minimum acceptable level of performance of the critical service core. Subsequently, the recovery method operates by forming sets A_k of steps of the form (6), taking into account feasibility (7) and aggregate constraints (9). The initial conditions for all scenarios are defined as the state following a degradation event.

In combined scenarios, the possibility of secondary resource release through controlled degradation is additionally provided for, which is formally accounted for by refining the resource balance (11). Such actions are permitted provided that the feasibility predicate $\psi(b)=1$, defined by the constraints (12), is satisfied. This ensures that degradation is used as an auxiliary mechanism to ensure the feasibility of restoration and excludes the use of degradation actions for the purpose of deliberately altering restoration performance indicators.

5.2. Performance indicators for restoration and methods for calculating them

The method is evaluated based on metrics that directly rely on the shortfall in meeting the threshold requirements for service quality metrics $d_s(t)$ (2), the profile-based quality model $q_s(\cdot)$ (3), the definition of the recovery step (6), the window feasibility

predicate $\varphi(a,k)$ (7), the step effect $\Delta d(a)$ (8), and the condition for window-based deficit reduction within the window (10). A finite simulation period is considered, defined by the sequence K of connectivity windows W_k (4), $k = \overline{1, K}$.

The total deficit of critical services over the simulation period is defined as an integral (in discrete form – cumulative) indicator:

$$D_\Sigma = \sum_{k=1}^K \sum_{s \in S_c} d_s(\tau_k), \quad (13)$$

where $d_s(\tau_k)$ is calculated according to (2) and using $q_s(\tau_k)$ or the tabulated relationship (3). Formula (13) is used to compare methods based on the cumulative deficit of critical services across a sequence of windows.

To account for the varying importance of critical services, a weighted variant is introduced:

$$D_w = \sum_{k=1}^K \sum_{s \in S_c} w_s d_s(\tau_k), \quad w_s > 0, \quad (14)$$

where the weights w_s are set according to the priority scale of critical services. Metric (14) is applied if it is necessary to emphasize the contribution of individual critical services without changing the definition of the deficit (2).

Next, it is necessary to formalize the window feasibility metrics and the satisfaction of feasibility conditions during recovery execution. For each window W_k , a set of potentially feasible recovery steps is determined:

$$F_k = \{a : (s, p \rightarrow p+1) \text{ and } \varphi(a,k)=1\}. \quad (15)$$

Formula (15) establishes that feasibility is defined as a combination of resource feasibility and the ability to complete service state coordination within the current connectivity window.

The proportion of windows in which there is at least one feasible recovery step is defined by the recovery availability metric:

$$\rho_F = \frac{1}{K} \sum_{k=1}^K \mathbf{1}(|F_k| > 0), \quad (16)$$

where $\mathbf{1}(\cdot)$ is an indicator function. Metric (16) characterizes the extent to which environmental conditions (resources and connectivity) allow recovery to be performed, regardless of which actions are selected by the method.

Let A_k be the set of recovery steps actually performed by the method within the window W_k . Compliance with the window feasibility conditions during recovery execution is evaluated by the fraction of steps that are feasible within the current connectivity window as defined in (7):

$$\rho_\varphi = \frac{\sum_{k=1}^K \sum_{a \in A_k} \mathbf{1}(\varphi(a, k) = 1)}{\sum_{k=1}^K |A_k|}. \quad (17)$$

For a method that generates A_k taking into account (7) and (9), $\rho_\varphi = 1$ is expected; whereas for the basic comparison modes, indicator (17) provides a quantitative measure of the occurrence of non-terminating recovery transitions.

Separately, an indicator of state coordination capacity excess is introduced, which is relevant specifically for modes that do not account for state coordination completion conditions within the current connectivity window:

$$G_\varphi = \sum_{k=1}^K \max \left\{ 0, \sum_{a \in A_k} g(a) - C_k \right\}. \quad (18)$$

Indicator (18) is not a substitute for (17), but complements it with a quantitative assessment of the volume of data required for service state coordination that cannot be transmitted within the connectivity windows.

Since the reduction of the deficit within the window is determined by $(\Delta d(a))$ (8) and condition (10), an indicator of the reduction of the deficit within the window is introduced:

$$I_k = \mathbf{1} \left(\sum_{a \in A_k} \Delta d(a) > 0 \right), \quad (19)$$

which is a formalized representation of criterion (10). The proportion of windows in which a deficit reduction is observed is defined as

$$\rho_P = \frac{1}{K} \sum_{k=1}^K I_k. \quad (20)$$

Indicator (20) is applied together with (16): if ρ_F is high but ρ_P is low, this means that steps were taken, but the method did not ensure a reduction in the deficit in the corresponding windows.

For an operational interpretation of the deficit reduction, the reduction in the total deficit of critical services within the window is also used:

$$\Delta D_k = \sum_{s \in S_c} (d_s(\tau_k) - d_s(\tau_k + \Delta_k)). \quad (21)$$

Indicator (21) is used as a consistency check: in a correctly constructed model and when determining $\Delta d(a)$ (8), indicator (19) must correlate with the sign of ΔD_k .

The recovery time is estimated as the moment when the zero-deficit state $d_s(t)$ is reached for each critical service. For the service $s \in S_c$, the recovery time is defined as:

$$T_s = \min \{ \tau_k : d_s(\tau_k) = 0 \}, \quad (22)$$

where τ_k represents the start of the windows (4).

For the ISMP as a whole, an aggregate indicator is used:

$$T_{\max} = \max_{s \in S_c} T_s, \quad (23)$$

which reflects the time required to restore the full critical set of services. Formulas (22)–(23) are used in conjunction with the integral indicators (13)–(14): the former evaluate the restoration rate, while the latter evaluate the cumulative deficit characteristic over the entire simulation period.

The method for calculating the above indicators is as follows. Indicators (13)–(23) are calculated according to the simulation protocol, in which the following are recorded for each window W_k : initial profiles $p_s(\tau_k)$, selected recovery steps A_k , degradation actions if present B_k , available resources \bar{R}_k , capacity C_k , and values $q_s(\tau_k)$ or tabulated values $\bar{q}_s(p_s(\tau_k))$. Next:

Step 1. Deficits $d_s(\tau_k)$ are determined $d_s(t)$ (2) using $q_s(t)$ (3).

Step 2. For each step $a \in A_k$, $\varphi(a, k)$ (7) and $\Delta d(a)$ (8) are calculated.

Step 3. F_k (15) is formed and ρ_F (16) is calculated.

Step 4. Using A_k , ρ_φ (17) and G_φ (18) are calculated.

Step 5. I_k (19), ρ_p (20), and ΔD_k (21) are calculated.

Step 6. Determine T_s (22), T_{\max} (23), and the integrated indicators D_Σ (13) and D_w (14).

The presented system of indicators allows us to distinguish three aspects of recovery quality: the presence of feasible opportunities in the environment ρ_F ; compliance with feasibility conditions and conditions for completing state alignment during the execution of recovery actions ρ_φ , G_φ ; reduction of the deficit and recovery speed ρ_p , ΔD_k , T_{\max} with a direct link to the deficit $q_s(t)$ and the step effect $\Delta d(a)$.

5.3. Basic comparison modes and conditions for correct matching

The comparison was performed using a set of basic modes that differ in the rule for forming sets of recovery steps A_k in the window W_k (4) and in the method of accounting for the deficit $d_s(t)$ (2), the step effect $\Delta d(a)$ (8), and window feasibility $\varphi(a,k)$ (7). In all modes, the recovery steps take the form (6), i.e., they correspond to raising the critical service profile by one level.

Thus, the basic modes are:

1. Mode R0 (passive control). In each window, recovery steps are not performed, i.e., $A_k = \emptyset$. This mode is used as the lower bound for deficit reduction metrics (19)–(21), recovery time (22)–(23), and for estimating cumulative deficit (13)–(14).

2. Mode R1 (resource-constrained recovery without considering the possibility of completing state coordination within the connectivity window). The set A_k is formed from the steps (6) with a focus on maximizing the total effect (8) under the constraint of only the resource part (9). The window capacity C_k and the condition $g(a) \leq C_k$, which is part of predicate (7), are intentionally disregarded. The consequences of this simplification are captured by indicators (17)–(18), where G_φ (18) characterizes the accumulated volume of data for state synchronization that cannot be transmitted within the connectivity windows.

3. Mode R2 (restoration with a constraint on the extent of state coordination and a priority for minimal

coordination). The set A_k consists only of steps for which the condition $g(a) \leq C_k$ holds, in accordance with the second part of (7). Selection is performed with a priority for minimizing $\sum_{a \in A_k} g(a)$, provided that the deficit does not increase; that is, steps with a small $g(a)$ are preferred even if $\Delta d(a)$ is small (8). The resource feasibility is controlled via the first constraint (9). This mode contrasts with R1 and demonstrates the effect of the priority of minimum coordination in the absence of a focus on reducing the deficit on ρ_p (20), ΔD_k (13), T_{\max} (23).

4. Mode R3 (fixed priority of critical services). A fixed order of critical services is specified: s_1, s_2, \dots by importance. In each window, steps (6) are sequentially considered for s_1 , then for s_2 , and so on, with the first step added to A_k for which $\varphi(a,k)=1$ is satisfied and the aggregate constraints (9) are not violated. Mode R3 does not optimize the total effect (8) and provides a reproducible behavioral benchmark where the decision is determined by a fixed order of services.

These modes cover three typical simplifications: ignoring the constraint on the possibility of completing state coordination within the connectivity window (R1), the absence of a focus on reducing the deficit subject to the coordination constraint (R2), and the rejection of adaptive selection in favor of a fixed order of services (R3). Their comparison with the method using deficit (2), window feasibility (7), and condition (10) is performed according to indicators (13)–(23).

5.4. Results for the window feasibility of recovery steps

This subsection analyzes compliance with the window feasibility conditions of recovery steps in terms of the fulfillment of $\varphi(a,k)$ (7) and consistency with the aggregate constraints (9). The main focus is on three aspects: the presence in the mobile platform's computing environment of at least one feasible step F_k according to (15) and the proportion of such windows ρ_F according to (16); the proportion of actually executed steps that are feasible within the current connectivity window, ρ_φ (17); the presence or absence of coordination capacity exceedances, G_φ (18).

Compliance with the feasibility conditions of recovery steps is defined as follows. For the proposed method, the sets of steps A_k are formed such that each selected step belongs to the set of feasible steps F_k (15), and the set of steps satisfies constraint (9). Given the definition of the predicate $\varphi(a, k)$ (7), this means that for all k and all $a \in A_k$, $\varphi(a, k) = 1$ holds. For the indicator (17), this yields $\rho_\varphi = 1$, and for the accumulated capacity excess (18) – the absence of excesses, i.e., $G_\varphi = 0$.

Fig. 2 shows the relationship between the window capacity C_k and the actual amount of state coordination used for two control rules. For the proposed method,

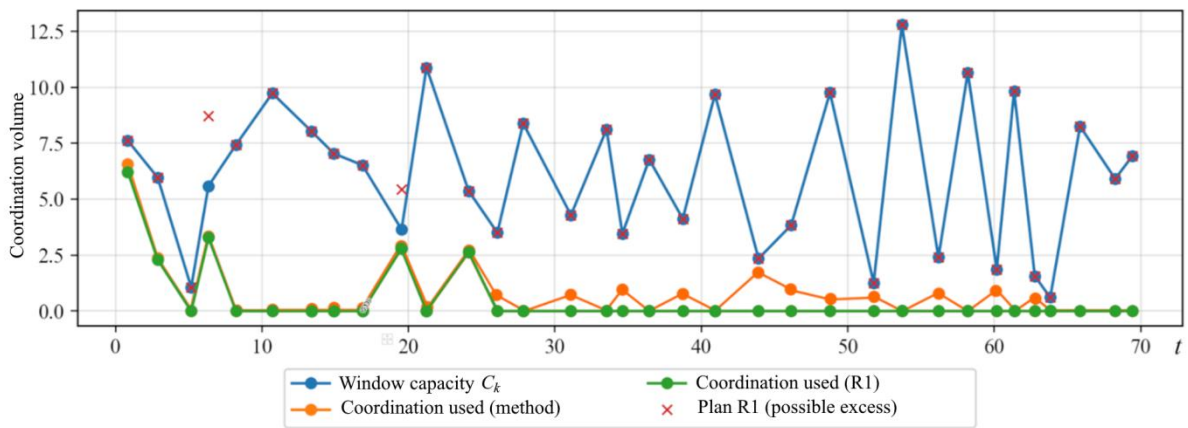


Fig. 2. Comparison of the capacity C_k use for state coordination

This result is a direct consequence of constructing the A_k taking into account (7) and (9) and is confirmed in experiments by the consistent absence of unfinished state-consistency transitions.

For mode R1, in which the window capacity C_k (5) and the synchronization constraint are not taken into account during the formation of A_k , a distinct property is observed: immediately after the execution of the resource part (9), windows arise for which

$$\sum_{a \in A_k} g(a) > C_k,$$

which is directly captured by the indicator $G_\varphi > 0$ in (18) and reduces ρ_φ in (17).

Thus, a comparison of (17)–(18) demonstrates that ignoring the condition $g(a) \leq C_k$, which determines the possibility of completing service state synchronization within the current connectivity window, generates errors that are critical for a mobile platform with intermittent connectivity.

the coordination volume is consistent with the constraint $\sum_{a \in A_k} g(a) \leq C_k$ (9). Therefore, capacity usage does not exceed the window limits. For the base mode R1, the planned coordination volume is additionally shown, which may exceed C_k . This illustrates the consequence of approaches that do not account for the condition $g(a) \leq C_k$, which determines the possibility of completing service state coordination within the current connectivity window during the action formation phase: a portion of the planned recovery transitions remains incomplete within the window, which is related to the second component of the predicate $\varphi(a, k)$ (7).

Modes R2 and R3, unlike R1, explicitly check the condition $g(a) \leq C_k$ in the step selection procedure and adhere to the aggregate constraints (9). Therefore, for them, the metrics ρ_φ (17) and G_φ (18) also indicate correct fulfillment of the feasibility conditions. The differences between the method and these regimes manifest in how the feasibility steps translate into a reduction in the deficit within the window, which is consistent with the indicator I_k (19) and the ratio ρ_p (20).

Next, we consider the availability of recovery within connectivity windows as a property of the environment. The indicator ρ_F (16) characterizes the presence of a recovery possibility in the computational environment: does there exist at least one step a of the form (6) for which $\varphi(a, k) = 1$ (7), i.e., does the ISTM allow recovery to be performed in this window. Analysis of the previously introduced scenarios reveals two consistent patterns.

First, in scenarios with determining connectivity constraints, ρ_F is predominantly determined by the condition $g(a) \leq C_k$ as a component of predicate (7), i.e., by the ratio between the data volume required to reconcile the state $g(a)$ and the capacity C_k (5). Even with sufficient resources \bar{R}_k , the sets F_k (15) are reduced to steps with small $g(a)$, and for small C_k , they may become empty, which reduces ρ_F in (16).

Second, in scenarios with deterministic resource constraints, ρ_F is determined primarily by the condition $\bar{r}(a) \leq \bar{R}_k$ as a component of predicate (7) and by the resource constraints in (9): the set F_k (15) may be non-empty for $g(a) \leq C_k$, but recovery steps become unavailable due to a shortage of \bar{R}_k for any candidate a . In combined scenarios, these two effects manifest simultaneously, underscoring the appropriateness of precisely predicate (7) as a common condition for window feasibility.

To interpret which constraint actually determines the composition of the set A_k , it is advisable to evaluate the residual capacity after selecting steps. For each window, a residual state-consistency capacity is introduced:

$$S_k^g = C_k - \sum_{a \in A_k} g(a), \quad (24)$$

which must be non-negative for modes that satisfy the second constraint (9), and characterizes how fully the capacity of the window C_k (5) is utilized.

Similarly, the vector residual state-consistency capacity is introduced:

$$\bar{S}_k^r = \bar{R}_k - \sum_{a \in A_k} \bar{r}(a), \quad (25)$$

which is non-negative component-wise under the first constraint (9) and indicates whether the solution was constrained by the platform resource.

To reduce capacity utilization to a dimensionless quantity, the window utilization coefficient is applied:

$$u_k = \frac{\sum_{a \in A_k} g(a)}{C_k}, \quad C_k > 0. \quad (26)$$

$u_k \leq 1$ applied to modes that satisfy constraint (9). $u_k > 1$ is possible for modes that do not consider this constraint during formation A_k .

The obtained values (24)–(26) provide an unambiguous interpretation of the modes of the

previously introduced scenarios. In scenarios with determining connectivity constraints, the proposed method is characterized by binding the coordination constraint itself: the residual state coordination capacity S_k^g (24) is small in a significant portion of the windows, while u_k (26) approaches unity. This means that the method utilizes the capacity of the state coordination window and generates recovery transitions within the bounds of constraint (9). In contrast, in scenarios with resource-constrained conditions, the resource component is more frequently constrained: the residual resource vector \bar{S}_k^r (25) has small components, while S_k^g (24) may remain relatively large, i.e., the window capacity is not fully utilized due to a lack of resources for steps (6).

For mode R1, the coefficient u_k (26) exceeds 1 in some windows, and the magnitude of this excess is accumulated by the indicator G_φ (18). This is direct quantitative confirmation that the formation of the A_k without accounting for constraint (9) leads to incomplete coordination exchanges within the connectivity windows and repeated violations of the threshold requirements for service quality indicators.

The results obtained show that the proposed method ensures compliance with the window feasibility conditions for recovery steps: the executed steps are feasible within the current connectivity window according to (7) and satisfy the aggregate constraints (9), as reflected in ρ_φ (17) and the absence of violations of G_φ (18). The indicator ρ_F in (16) distinguishes situations where recovery is possible within the window from situations where the environment has no feasible steps according to (15). The residual state coordination capacity (24), the residual resource vector (25), and the coefficient (26) specify which constraint is decisive in each scenario class.

5.5. Results for window-based reduction of critical service shortages

This subsection analyzes the method's ability to convert the recovery capabilities available on the mobile platform into an actual reduction in the deficit of critical services. To this end, the deficit reduction indicator within the window I_k (19), the proportion of windows in which a reduction in the deficit is observed ρ_p (20), and the magnitude of the reduction in the total deficit of

critical services within the window ΔD_k (21) are used. Additionally, the recovery availability metric from ρ_F (16) is used, which allows distinguishing environmental constraints from the quality of management decisions.

Since ρ_F (16) represents the proportion of windows in which there is at least one recovery step that can be performed within the current connectivity window, and ρ_p (20) reflects the proportion of windows in which the method resulted in an actual reduction of the deficit (19), it is appropriate to estimate the conversion rate of feasibility into deficit reduction:

$$\eta = \frac{\rho_p}{\rho_F + \varepsilon}, \quad (27)$$

where $\varepsilon > 0$ is a small number to avoid division by zero in the case of $\rho_F = 0$.

Formula (27) is used as a generalized indicator of how effectively the method utilizes windows in which recovery is possible.

In scenarios with determining connectivity constraints, the set F_k (15) is often restricted to steps with a small amount of coordination $g(a)$ due to the constraint $g(a) \leq C_k$, which is part of predicate (7). Under such conditions, the indicator ρ_F (16) decreases, reflecting the environmental constraints. The key point is that windows with $|F_k| > 0$ create an opportunity to reduce the deficit, which is realized through the selection A_k satisfying condition (10), and thus $I_k = 1$ (19). This is why η (27) is interpreted as a characteristic of the formation rule A_k .

For the proposed method, the reduction of the deficit in the window is ensured by the selection construct

A_k , which combines the feasibility requirement (7), the total constraints (9), and condition (10). This means that if there is at least one step a from $\varphi(a, k) = 1$ and $\Delta d(a) > 0$ (8), the method forms A_k such that $\sum_{a \in A_k} \Delta d(a) > 0$ (10), i.e., ensures $I_k = 1$ (19). In this sense, progress is a controlled property of the solution, related to deficits (2) and step effects (8).

Next, we consider the magnitude of the deficit reduction within the window and its stability. The indicator I_k (19) captures the fact of the deficit reduction but does not reflect its magnitude. Therefore, the analysis is supplemented by the average magnitude of the positive deficit reduction:

$$\overline{\Delta D^+} = \frac{\sum_{k=1}^K \max\{0, \Delta D_k\}}{\sum_{k=1}^K I_k + \varepsilon}. \quad (28)$$

Formula (28) characterizes the average magnitude of the reduction in the total deficit of critical services in those windows where a deficit reduction is recorded.

Fig. 3 shows the dynamics of the total deficit of critical services and its accumulation over time, reflecting the integral characteristic of the system's operation in degraded modes. It can be seen that after a series of windows in which completion steps of recovery are possible, the total deficit decreases to zero and remains there, while the accumulated deficit reaches a plateau, corresponding to the completion of recovery. In this scenario implementation, the curves for the method and R1 are visually similar, which is interpreted as the decisive influence of environmental constraints over differences in control rules. That is why the indicators ρ_F and ρ_p in Fig. 4 are additionally analyzed.

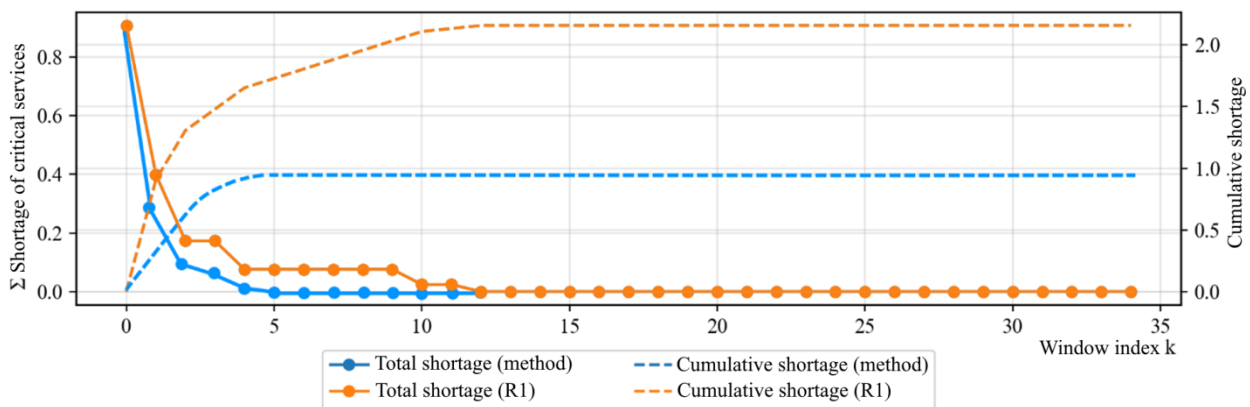


Fig. 3. Total and cumulative deficit of critical services (comparison)

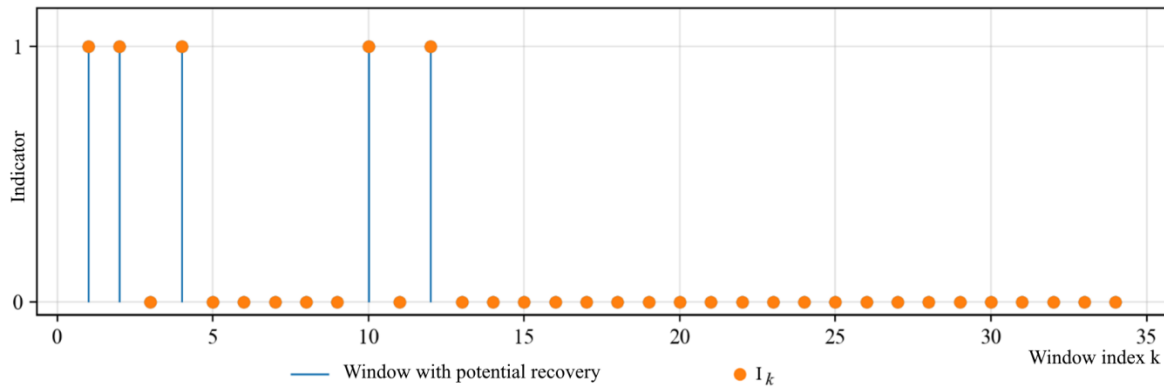


Fig. 4. Indicators of the separation between the possibility of recovery and the actual reduction of the deficit in connectivity windows

Fig. 4 demonstrates the separation of the possibility of recovery and the actual reduction of the deficit in connectivity windows. The vertical markers correspond to windows in which there is at least one recovery step ($\exists a: \varphi(a, k) = 1$), i.e., the environment allows recovery to be performed. The points I_k reflect the actual reduction of the deficit in the window, i.e., the realization of the reduction of the critical service deficit. In this experiment, ρ_F is small (of the order of 0.15), which implies rare windows with recovery potential; accordingly, ρ_p is bounded from above and turns out to be close to ρ_F , which is consistent with the structural constraint $\rho_p < \rho_F$ and explains the shape of the shortage curves in Fig. 3.

In modes that do not account for the constraint on the ability to complete service state coordination within the current connectivity window, a discrepancy arises between the estimated reduction in the deficit based on the aggregate effect (8) and the actual change in the deficit (21). If the synchronization constraint $\sum_{a \in A_k} g(a) \leq C_k$ (9) is violated within the window, as reflected by the excess $G_\varphi > 0$ (18), part of the planned synchronization exchanges cannot be completed. In such a situation, $\sum_{a \in A_k} \Delta d(a)$ may remain positive, but the actual reduction in the deficit ΔD_k (21) becomes unstable and may approach zero in individual windows. This is why it is advisable to analyze the reduction in the deficit using the pair of indicators I_k (19) and ΔD_k (21), rather than solely through step effects (8).

For the proposed method, this discrepancy is eliminated by the feasibility constraint (7) and the total constraints (9), which is consistent with the previous results via ρ_φ (17) and G_φ (18). Therefore, the progress

value in the window according to (21) becomes consistent with the cumulative effect of the selected steps $\sum_{a \in A_k} \Delta d(a)$ in (10), and the average progress value $\overline{\Delta D^+}$ in (28) reflects precisely the effectiveness of the step selection.

The reduction in the deficit is distributed among critical services as follows. Since ΔD_k (21) aggregates the contributions of all critical services, it is important to ensure that recovery does not focus on a single service at the expense of keeping others in a deficit state for an extended period. To achieve this, individual recovery times T_s (22) and the aggregated time T_{\max} (23) are used. If the method reduces the deficit at I_k (19) and ΔD_k (21), but T_{\max} (23) remains high, this means that the reduction in the deficit across the windows is achieved unevenly across services, and certain critical services may be restored significantly later.

For the R3 fixed-priority mode, this situation is typical by design: deficit reduction is prioritized for services with the highest priority, which may increase ρ_p (20) in early windows, but increases the T_{\max} (23) and the cumulative deficit D_Σ (13) due to the prolonged persistence of deficits in lower-priority critical services. In the proposed method, this is addressed by basing the step selection criterion on the deficit (2) and the step effect (8), while priority is determined by the step's contribution to reducing the deficit and its window feasibility (7). This is reflected in an increase in ρ_p (20) and a decrease in T_{\max} (23) in the absence of coordination violations G_φ (18).

Next, the relationship between the reduction of the deficit within the connectivity windows and the integral deficit indicators is formalized. The reduction of the deficit according to (19), (21) is local in

nature, whereas the integral estimate is given by the accumulated deficit D_{Σ} (13) and the weighted deficit D_w (14). The reduction of D_{Σ} (13) is determined by ρ_p (20) and $\overline{\Delta D^+}$ (28). When comparing modes, it is important to consider modes in which the deficit reduction is achieved through frequent small steps, and modes in which steps with a large expected effect are applied without controlling the possibility of completing state convergence within the window, leading to the occurrence of overshoots G_{ρ} (18) and unstable dynamics ΔD_k (21).

As a result, the results are consistent with the logic of constructing the indicators. Compliance with the window feasibility conditions (17)–(18) ensures that the planned effect of the steps (8) corresponds to the actual change in the deficit (21). The presence of opportunities in the environment is captured via ρ_f (16), and the effectiveness of their use is assessed via ρ_p (20) and the coefficient η (27). The total impact on the quality of recovery is reflected through the integral indicators (13)–(14) and the recovery time (22)–(23), which are direct functions of the deficit (2) and are therefore correctly linked to condition (10).

5.6. Sensitivity analysis of connectivity parameters and state coordination volume

The sensitivity of critical service recovery under intermittent connectivity is determined by how the sets of recovery steps performed within the current connectivity window F_k (15) and, accordingly, the indicators ρ_f (16), ρ_p (20), η (27), D_{Σ} (13) and T_{\max} (23) while varying the window parameters W_k (4), their capacity C_k (5), and the matching characteristics $g(a)$ included in (7) and (9). The following analysis is conducted at the level of parameterized series, in which individual factors are varied while other scenario parameters remain constant.

To isolate the effects, scaling factors are introduced that are applied to the environment parameters and recovery steps:

1. Scaling of the connectivity window sizes:

$$C_k(\alpha) = \alpha C_k, \quad \alpha > 0. \quad (29)$$

Equation (29) models the change in the capacity of connectivity windows caused by a change in the

window's throughput or duration, assuming constant recovery profiles.

2. Scaling of the volume of recovery step state coordination:

$$g_{\beta}(a) = \beta g(a), \quad \beta > 0. \quad (30)$$

Formula (30) is interpreted as a change in the volume of data required for state synchronization during the execution of recovery steps, for example, due to varying state compactness, different incremental synchronization policies, or varying levels of detail during recovery.

3. Scaling of the available mobile platform resource:

$$\bar{R}_k(\gamma) = \gamma \bar{R}_k, \quad \gamma > 0. \quad (31)$$

Formula (31) reflects different mobile platform load modes or different hardware configurations.

In the subsequent sensitivity analysis, all dependencies are derived as a consequence of defining window feasibility (7) and constructing the sets F_k (15).

Next, the monotonicity of recovery availability within connectivity windows is formalized. The set of feasible steps F_k (15) is considered with respect to the scaling parameters (29)–(31). The following monotonicities directly follow from the definition of the predicate (7):

1. For $\alpha_2 \geq \alpha_1$, the following is enabled:

$$F_k(C(\alpha_1)) \subseteq F_k(C(\alpha_2)), \quad (32)$$

since the condition $g(a) \leq C_k(\alpha)$ in (7) becomes weaker as α increases according to (29), whereas the resource part of (7) remains unchanged. Equation (32) implies that increasing the window size cannot reduce the set of available recovery steps.

2. For $\beta_2 \geq \beta_1$, the opposite inclusion holds:

$$F_k(g\beta_2) \subseteq F_k(g\beta_1), \quad (33)$$

because the condition $g_{\beta}(a) \leq C_k$ in (7) becomes stricter as β increases according to (30). Equation (33) establishes that increasing the scope of recovery step coordination narrows the set of feasible steps.

3. For $\gamma_2 \geq \gamma_1$, the following holds

$$F_k(\bar{R}(\gamma_1)) \subseteq F_k(\bar{R}(\gamma_2)), \quad (34)$$

since the component-wise condition $\bar{r}(a) \leq \bar{R}_k(\gamma)$ in (7) weakens as γ increases according to (31).

Since the recovery feasibility indicator ρ_f (16) is defined by the fact that the set F_k (15) is non-empty, it

follows from (32) and (34) that ρ_F is non-decreasing with respect to α and γ , and from (33) the ρ_F does not increase with respect to β . This is important for interpretation: the variation of ρ_F is, first and foremost, a characteristic of the presence of conditions for recovery in the environment.

The actual reduction in the deficit within the window is estimated by the indicator I_k (19) and the ratio ρ_p (20), while the conversion is estimated by the coefficient η (27). The definitions imply the constraint

$$\rho_p \leq \rho_F, \quad (35)$$

since a reduction in the deficit within the window is possible only if there is at least one feasible recovery step (15) with $\varphi(a,k)=1$ (7), i.e., when $|F_k| > 0$ (16). Inequality (35) is structural and does not depend on the specific selection method A_k .

At the same time, ρ_p and η are determined by the value of ρ_F and the rule for using the set of candidates F_k in the window. The key parameter is the ratio between $g(a)$ and C_k , which defines the recovery resolution. If for most candidates a satisfies $g(a) \approx C_k$, then the set F_k in (15) often contains only 1–2 steps, and the choice of A_k is close to the deterministic bound (7). In this mode, η (27) approaches 1 provided that these steps satisfy $\Delta d(a) > 0$ (8), since then the satisfaction of condition (10) becomes inevitable.

Another limiting regime arises when $g(a) \ll C_k$ for many a . Then F_k (15) increases, and the total constraint (9) becomes dominant. In this case, the sensitivity of ρ_p and η is determined by the ability of the formation rule A_k to ensure $\sum_{a \in A_k} \Delta d(a) > 0$ (10) while satisfying (9). In this mode, the influence of the parameter β (30) is most noticeable: an increase in $g_\beta(a)$ narrows F_k (33) and increases competition between steps for window capacity, which may reduce ρ_p (20) for a constant ρ_F (16) if steps with a small $\Delta d(a)$ (8) remain predominantly feasible.

The cumulative deficit D_Σ (13) and the recovery time T_{\max} (23) are integral consequences of the reduction in the deficit within connectivity windows and its magnitude. From the definition ΔD_k (21) and summation

over windows, it follows that the decrease in D_Σ is ensured simultaneously by an increase in the proportion of windows with a decrease in the deficit ρ_p (20) and an increase in the average positive decrease in the deficit $\overline{\Delta D^+}$ (28). Therefore, the parameters α , β , γ (29)–(31) affect D_Σ and T_{\max} primarily through two channels:

- by changing the proportion of windows in which recovery is possible (via ρ_F (16) based on the monotonicities (32)–(34));

- by changing the magnitude of the deficit reduction within the window (via which specific steps remain feasible in F_k (15) and how they are formed in A_k under constraint (9)).

This implies the following. A decrease in α (29) or an increase in β (30) leads to the constraint $g(a) \leq C_k$ (7) being satisfied less frequently, the sets F_k becoming narrower, ρ_F (16) decreasing, and the recovery skipping some connectivity windows, which increases D_Σ (13) and T_{\max} (23). Conversely, increasing α or decreasing β expands the sets F_k (32)–(33), which potentially reduces D_Σ and T_{\max} , provided that the formation rule A_k ensures the fulfillment of (10) without violating (9).

Analysis of (29)–(35) yields the following operational interpretation. If the environment is characterized by small C_k (5) or high variability Δ_k (4), then the assurance of recovery is determined by the ratio between the available resource \bar{R}_k and the scope of state coordination $g(a)$ (7). In this context, it is advisable to construct recovery profiles such that there are steps with small $g(a)$ and a non-zero effect $\Delta d(a)$ (8). Such steps form non-empty sets F_k (15) under strict window conditions, increase ρ_F (16), and ensure the possibility of satisfying condition (10) while maintaining the feasibility conditions (7), (9).

Thus, the sensitivity of the method to connectivity parameters and coordination volumes is structural in nature and directly follows from the definition of window feasibility (7), the sets of feasible steps (15), and the progress indicators (19)–(21). This allows the obtained results to be interpreted as a consequence of the relationship between window capacity (5), step

coordination volumes ($g(a)$), and available resources (\bar{R}_k), which together determine the domain of admissible recovery actions in each connectivity window.

5.7. Discussion of the Research Results

It is appropriate to interpret the results obtained by distinguishing between two reasons for the lack of a reduction in the shortage of critical services: environmental constraints and the characteristics of management decisions. Environmental constraints are reflected by the non-emptiness of the set of feasible steps F_k (15) and the indicator ρ_F (16), which is directly determined by the window feasibility predicate (7). The properties of a control decision are manifested in how often and by how much the method ensures a reduction in the deficit within the connectivity windows, which is recorded by I_k (19), ρ_p (20), the conversion coefficient η (27), and the values ΔD_k (21) and $\overline{\Delta D}^+$ (28). The decrease in ρ_p is interpreted in conjunction with the change in ρ_F , since recovery within the window is possible provided that $|F_k| > 0$.

From a survivability perspective, the results obtained should be interpreted as confirmation that the recovery of critical services in the ISMP is a controllable component of maintaining survivability. The metrics ρ_F (16) and ρ_p (20) reflect, respectively, the potential and realized ability of the ISMP to restore critical functions within window connectivity. The metrics D_Σ (13) and T_{\max} (23) characterize the duration and intensity of the degradation mode during which the failure to meet the threshold requirements for service quality metrics persists. In this context, compliance with the window feasibility conditions specified by $\varphi(a, k)$ (7) and the constraints (9) is a necessary condition for ensuring real-time recovery, as it eliminates incomplete recovery transitions that prolong degradation intervals and increase the risk of secondary violations of critical requirements.

It is important that the ability to complete service state synchronization within the current connectivity window, defined by the condition $g(a) \leq C_k$ as a component of $\varphi(a, k)$ (7) and the second constraint (9), is a determining constraint for the mobile platform. Ignoring this constraint manifests as the occurrence

of G_φ (18) and a decrease in ρ_φ (17) in modes that form A_k without accounting for the constraint $\sum_{a \in A_k} g(a) \leq C_k$ (9).

In a service-oriented interpretation, this implies an accumulation of uncompleted balancing exchanges and a discrepancy between the expected effect $\sum_{a \in A_k} \Delta d(a)$ (8) and the actual reduction in the deficit ΔD_k (21). In contrast, satisfying conditions (7) and (9) ensures consistency between the local estimate of the effect $\Delta d(a)$ (8) and the actual reduction in the deficit ΔD_k , which is essential for mobile environments with short connectivity windows.

The issue of recovery separability is also important. Sensitivity analysis shows that the relationship between capacity C_k (5) and state synchronization volumes $g(a)$ (7) determines the range of possible recovery steps within the window. The monotonicities (32)–(34) and the structural constraint (35) reflect that an increase in environment availability does not guarantee a reduction in the cumulative deficit D_Σ in (13): if the set of feasible steps F_k (15) grows, but most steps have a small effect $\Delta d(a)$ (8), then the reduction in the deficit may occur in a significant proportion of windows ρ_p (20), but be small in magnitude $\overline{\Delta D}^+$ (28), which increases T_{\max} in (23). If the steps are characterized by $g(a) \approx C_k$, then the choice within the window is determined by the constraint $g(a) \leq C_k$ (7), and the effectiveness of the recovery is determined by the presence of steps with $\Delta d(a) > 0$ (8). It is advisable to construct recovery profiles such that the set of steps of the form (6) contains steps with small $g(a)$ (7) and a non-zero effect $\Delta d(a)$ (8), since they ensure the non-emptiness of F_k (15) and increase ρ_F (16) in strict windowing modes.

Separately, it is worth noting the limits of applicability arising from the previously adopted abstractions. First, service quality is represented by the aggregate metric $q_s(t)$ and the threshold θ_s , while the deficit is defined by formula (2). Such a model is suitable for formalizing deficit reduction; however, in real systems, service may have a multidimensional SLA profile, and the deficit may be determined by a more complex composition of metrics. This does not alter the

feasibility logic (7) or condition (10), but requires refinement of the definition $d_s(t)$ (2) and the step effect $\Delta d(a)$ (8) for the multidimensional case.

Second, discrete profiles P_s (1) and a tabular correspondence of profile quality (3) are used. In ISMPs with continuous control or a large number of modes, a procedure for quantizing profiles and evaluating $\bar{q}_s(\cdot)$ is required. The accuracy of such a procedure will directly affect the correctness of the evaluation $\Delta d(a)$ (8).

Third, the method relies on an estimate of the window capacity C_k (5) and the state alignment volume $g(a)$ (7). In practical implementation, these values are determined via telemetry and may be subject to error. Overestimation C_k or underestimation $g(a)$ leads to steps for which state coordination completion within the current connectivity window is impossible being incorrectly identified as feasible, which manifests as an increase in G_ρ (18) and a discrepancy between $\sum_{a \in A_k} \Delta d(a)$ (10) and the actual ΔD_k (21). Underestimating C_k or overestimating $g(a)$, conversely, narrows F_k (15) and reduces ρ_F (16), which may artificially increase D_Σ (13) and T_{\max} (23) under constant environmental conditions. In applied scenarios, it is advisable to monitor the quality of estimation C_k and $g(a)$, as well as to use the residual state alignment capacity (24) and the residual resource vector (25) as indicators of approaching the feasibility boundaries.

Fourth, this study does not introduce specific constraints on the intensity of reconfigurations as a separate control mechanism. The key elements in this work are the feasibility conditions for steps within the window, defined by $(\varphi(a, k))$ (7) and the constraints (9), and condition (10). If the applied control loop requires additional stabilization regarding switching frequency, such constraints can be introduced as external regulations based on known solutions [19], [20].

The scope of application of the proposed method corresponds to architectures in which a mobile platform combines the functions of an edge computing node and a coordinator for a group of subordinate unmanned aerial vehicles (UAVs). A typical ISMP is the onboard information system of the carrier vehicle, which provides mission planning and task distribution among vehicles,

situational awareness based on the aggregation of sensor streams, navigation correction and reference map alignment, the generation of control commands and execution monitoring, and the secure exchange of telemetry for critical flight safety parameters. Intermittent connectivity in such a system is determined by flight geometry, shielding, radio channel conditions, and electronic countermeasures. This defines connectivity windows with variable duration and bandwidth for communication between the carrier aircraft, subordinate aircraft, and the ground control station, and limits the window capacity for state synchronization. During a degradation episode, the ISMP performs a controlled reduction of non-critical service profiles, specifically by reducing the update frequency of secondary telemetry streams, switching to coarser video processing models, deferring the synchronization of non-critical logs, and limiting map detail. These actions free up computational resources and bandwidth to maintain the minimum acceptable level of performance for the critical service core. Once conditions have stabilized, the task of restoring critical services to their normal profiles arises, with the feasibility of restoration determined by the availability of onboard resources and the ability to complete service state synchronization within the current connectivity window. This applies to the synchronization of map increments, updated task parameters, integrity check data, sensor aggregation parameters, and other state components that must be transmitted or received before the end of the window. Under such conditions, the application of the recovery step window feasibility predicate (7) and the condition of deficit reduction within the window (10) ensures the formation of only those transitions that elevate the critical service profile, for which state synchronization can be completed within the available window. This reduces the likelihood of accumulating incomplete reconfigurations, repeated violations of service quality thresholds, and erroneous recovery, which in a UAV group translates into a loss of coordination and reduced mission stability.

Thus, service-oriented recovery, based on the deficit (2) and the step effect (8) and subject to window feasibility (7) and aggregate constraints (9), ensures a measurable reduction of the deficit within connectivity windows, as reflected by the metrics (19)–(23) and integral measures (13)–(14). The limits of applicability are determined by the accuracy of estimating the parameters of windows (4) and (5), the characteristics of state coordination within (7), as well as the level

of detail of service profiles (1)–(3). This provides a basis for refining profile models and procedures for evaluating feasibility parameters without altering the logic of window-based recovery, as defined by equations (7), (9), and (10).

6. Conclusions and prospects for further research

This paper solves the problem of restoring critical services of an information system on a mobile platform under conditions of controlled degradation with intermittent connectivity and resource constraints. The proposed method presents recovery as a sequence of discrete steps to improve service profiles, with performance monitoring through the reduction of the shortfall in meeting threshold requirements for service quality indicators. This formulation distinguishes recovery transitions that ensure an actual reduction in the deficit from reconfigurations that do not lead to a reduction in the deficit within connectivity windows.

It is shown that the correctness of the recovery of critical ISMP services is determined by the resource feasibility of actions and the possibility of completing service state coordination within the current connectivity window. It has been established that ignoring the constraint on state coordination within the window leads to the accumulation of unfinished coordination exchanges, discrepancies between the expected effect of recovery actions and the actual dynamics of the deficit, as well as to repeated violations of threshold requirements after partial recovery.

In the context of improving the survivability of the ISMP, the method provides an operational mechanism for maintaining and restoring critical functions during a connectivity window with variable duration and bandwidth. Reducing the cumulative deficit of critical services and shortening the recovery time of the critical set correspond to a reduction in the duration of degradation episodes of critical services and an increase in the proportion of control intervals in which service quality threshold requirements are met. The method generates recovery steps based on window feasibility and on aggregate resource and state coordination constraints, ensuring recovery without accumulating unfinished transitions.

A distinctive feature of the proposed method is the introduction of a window feasibility condition for a step, which accounts for resource feasibility on the mobile platform and the possibility of completing service state coordination within the current connectivity window. This ensures the pruning of steps for

which completion of coordination within the window is impossible, and consistency between the expected reduction in the deficit and the actual change in the deficit. The evaluation metric system ensures the separation of computational environment constraints and control decision properties, as well as the comparison of recovery modes based on cumulative deficit, the magnitude of deficit reduction within the window, and recovery time.

The value of the results is determined by the method's suitability for implementation as an operational recovery control module in a window-based ISMP, particularly in onboard information systems of UAV groups. The method ensures the return of critical services to their normal profiles without accumulating unfinished coordination exchanges and while maintaining the acceptability of controlled degradation as a peripheral tool for freeing up resources. This reduces the likelihood of repeated violations of service quality thresholds following partial recovery and shortens the system stabilization time after a degradation episode.

Prospects for further research cover three areas. First, it is advisable to generalize the quality deficit from a scalar metric to a multidimensional profile of requirements for critical services with a formalized rule for aggregating deficits, while maintaining the criterion of deficit reduction within the connectivity window. Second, it is necessary to extend the window feasibility model to account for errors in estimating window capacity and the scope of state coordination [21], including robust feasibility estimates and adaptive adjustment of coordination parameters [22]. Third, it is promising to account for dependencies between services and partial state coordination modes of the ISMP [23], which will allow for the formation of recovery steps with varying granularity and improve the efficiency of using short connectivity windows while preserving the correctness of ISMP recovery.

Conflict of interest

The author declare that they have no conflict of interest with respect to this research, including financial, personal, authorship, or other conflicts that could influence the research and its results presented in this article.

Funding

The study was conducted without financial support.

Data availability

The manuscript has no associated data.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence technology in the creation of this work.

References

1. Dodonov, O., Gorbachyk, O., Kuznietsova, M. (2023), "Critical Infrastructure Resilience and Cybersecurity of Information Management Systems", *XXIII International Scientific and Practical Conference Information Technologies and Security (ITS 2023)*, pp. 1–10, available at: <https://ceur-ws.org/Vol-3887/paper1.pdf>
2. Fesenko, H., Iliashenko, O., Kharchenko, V., Leichenko, K., Sachenko, A., Scislo, L. (2024), "Methods and Software Tools for Reliable Operation of Flying LiFi Networks in Destruction Conditions", *Sensors*, Vol. 24, No. 17, 5707. DOI: <https://doi.org/10.3390/s24175707>
3. Рубан, І.В., Ткачов, В.М. (2025), "Багаторівнева модель інформаційної системи на мобільній платформі та формалізація критеріїв її живучості", *Вісник Херсонського національного технічного університету*, Т. 2, № 3(94), С. 399–409. DOI: <https://doi.org/10.35546/kntu2078-4481.2025.3.2.51>
4. Dodonov, O., Jiang, B., Dodonov, V. (2019), "Survivability Mechanisms of Complex Computer Systems Based on Common Information Space", *Computer Science*, Vol. 45, No. 7, pp. 41–45. DOI: <https://doi.org/10.19678/j.issn.1000-3428.0053440>
5. Churyumov, G., Tokarev, V., Tkachov, V., Partyka, S. (2018), "Scenario of Interaction of the Mobile Technical Objects in the Process of Transmission of Data Streams in Conditions of Impacting the Powerful Electromagnetic Field", *2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)*, pp. 183–186. DOI: <https://doi.org/10.1109/DSMP.2018.8478539>
6. Рубан, І.В., Ткачов, В.М. (2025), "Модель міжрівневих порушень та політик відновлення інформаційної системи на мобільній платформі", *Телекомунікаційні та інформаційні технології*, №. 3 (88), С. 204–221. DOI: <https://doi.org/10.31673/2412-4338.2025.038721>
7. Yu, C., Rosendo, A. (2021), "Risk-Aware Model-Based Control", *Frontiers in Robotics and AI*, Vol. 8, 617839. DOI: <https://doi.org/10.3389/frobt.2021.617839>
8. Zeng, Y., Wu, L., Li, J., Zhuang, X., Wu, C. (2025), "Resilient Task Allocation for UAV Swarms: A Bilevel PSO-ILP Optimization Approach", *Drones*, Vol. 9, No. 9, 623. DOI: <https://doi.org/10.3390/drones9090623>
9. Та, P.B., Mafeni, V., Kim, Y. (2025), "A Design of Workflow-Based Automated Failure Recovery Framework in IoT Edge Environment", *IEEE Access*, Vol. 13, pp. 101885–101899. DOI: <https://doi.org/10.1109/ACCESS.2025.3578739>
10. Trung, K.N., Tran, M.-N., Kim, Y. (2026), "Enabling Service Continuity for Stateful Service Segmentation in Mobile Edge Computing Toward 6G", *IEEE Access*, Vol. 14, pp. 20589–20604. DOI: <https://doi.org/10.1109/ACCESS.2026.3661972>
11. Casini, D., Pazzaglia, P., Becker, M. (2025), "Managing real-time constraints through monitoring and analysis-driven edge orchestration", *Journal of Systems Architecture*, Vol. 164, 103403. DOI: <https://doi.org/10.1016/j.sysarc.2025.103403>
12. Agrawal, K., Jyothi, S.A. (2025), "Cooperative Graceful Degradation in Containerized Clouds", *Proceedings of ASPLOS 2025*, pp. 214–232. DOI: <https://doi.org/10.1145/3669940.3707244>
13. Ergenç, D., Memedi, A., Fischer, M., Dressler, F. (2025), "Resilience in Edge Computing: Challenges and Concepts", *Foundations and Trends in Networking*, Vol. 14, No. 4, pp. 254–340. DOI: <https://doi.org/10.1561/13000000074>
14. Golpayegani, F., Chen, N., Afraz, N., Gyamfi, E., Malekjafarian, A., Schäfer, D., Krupitzer, C. (2024), "Adaptation in Edge Computing: A Review on Design Principles and Research Challenges", *ACM Transactions on Autonomous and Adaptive Systems*, Vol. 19, No. 3, pp. 1–43. DOI: <https://doi.org/10.1145/3664200>
15. Chu, S., Koe, J., Garlan, D., Kang, E. (2024), "Integrating Graceful Degradation and Recovery through Requirement-driven Adaptation", *Proceedings of the 19th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS '24)*, pp. 122–132. DOI: <https://doi.org/10.1145/3643915.3644090>
16. Siasi, N., Jasim, M.A., Ghani, N. (2024), "Post-fault restoration of service function chains in fog networks", *Computer Networks*, Vol. 251, 110580. DOI: <https://doi.org/10.1016/j.comnet.2024.110580>
17. Guo, F., Peng, M. (2023), "Efficient Mobility Management in Mobile Edge Computing Networks: Joint Handover and Service Migration", *IEEE Internet of Things Journal*, Vol. 10, No. 20, pp. 18237–18247. DOI: <https://doi.org/10.1109/JIOT.2023.3279842>
18. Li, B., Nahrstedt, K. (1998), "An Open Task Control Model for Quality of Service Adaptation", *Proceedings of the 14th International Conference of Advanced Science and Technology (ICAST 98)*, pp. 29–41.

19. Zhang, X., Debroy, S. (2023), "Resource Management in Mobile Edge Computing: A Comprehensive Survey", *ACM Computing Surveys*, Vol. 55, No. 13s, Article 291, 37 pp. DOI: <https://doi.org/10.1145/3589639>
20. Zhang, J., Ning, Z., Waqas, M., Alasmary, H., Tu, S., Chen, S. (2024), "Hybrid edge-cloud collaborator resource scheduling approach based on deep reinforcement learning and multiobjective optimization", *IEEE Transactions on Computers*, Vol. 73, No. 1, pp. 192–205. DOI: <https://doi.org/10.1109/TC.2023.3326977>
21. Colarusso, C., Falco, I., Zimeo, E. (2025), "Business continuity of Cloud-based IoT applications through a seamless continuum", *Internet of Things*, Vol. 33, 101723. DOI: <https://doi.org/10.1016/j.iot.2025.101723>
22. Фролов, Д.Є. (2025), "Модель забезпечення відмовостійкості інформаційних систем на основі багаторівневих структур", *Вісник Херсонського національного технічного університету*, Т. 2, №. 3(94). С. 474–483. DOI: <https://doi.org/10.35546/kntu2078-4481.2025.3.2.61>
23. Zolotukhin O., Filatov, V., Kudryavtseva, M., Shaptala, S. (2025), "Using Artificial Intelligence Methods In Tasks Of Decentralized Control Of A Group Of Unmanned Aerial Vehicles", *Military Strategy and Technology*, №. 2(2), pp. 59–73. DOI: <https://doi.org/10.63978/3083-6476.2025.2.2.06>

Received (Надійшла) 08.02.2026

Accepted for publication (Прийнята до друку) 27.02.2026

Publication date (Дата публікації) 12.03.2026

Відомості про авторів / About the Authors

Ткачов Віталій Миколайович – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, докторант кафедри електронних обчислювальних машин; Харків, Україна;

Vitalii Tkachov – Phd (Technical Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Doctoral candidate at the Department of Electronic Computers; Kharkiv, Ukraine;

e-mail: vitalii.tkachov@nure.ua

ORCID ID: <https://orcid.org/0000-0002-6524-9937>

МЕТОД ВІДНОВЛЕННЯ КРИТИЧНИХ СЕРВІСІВ ІНФОРМАЦІЙНОЇ СИСТЕМИ НА МОБІЛЬНІЙ ПЛАТФОРМІ В УМОВАХ КЕРОВАНОЇ ДЕГРАДАЦІЇ

Предметом роботи є методи відновлення критичних сервісів інформаційної системи на мобільній платформі за переривчастої зв'язності та керованої деградації. **Мета дослідження** – розробити сервісно-орієнтований метод відновлення, який забезпечує вимірюване зменшення дефіциту виконання порогових вимог до показників якості обслуговування в межах вікон зв'язності з огляду на ресурсні обмеження й можливості завершення узгодження стану сервісів у поточному вікні зв'язності. У дослідженні застосовано такі **методи**: системний аналіз, математичне моделювання станів і профілів сервісів, формалізація дефіциту виконання порогових вимог до показників якості обслуговування, предикатний опис віконної здійсненості відновлювальних кроків і обчислювальний експеримент для оцінювання показників результативності відновлення. У статті передбачалося розв'язати такі **завдання**: формалізувати крок відновлення як перехід між сервісними профілями, що зменшує дефіцит виконання порогових вимог до показників якості обслуговування; впровадити умову віконної здійсненості кроку відновлення як поєднання ресурсної здійсненості та можливості завершення узгодження стану сервісів у межах поточного вікна зв'язності; побудувати правило формування множини кроків відновлення у вікні, яке забезпечує зменшення дефіциту за наявності хоча б одного здійсненого кроку й узгоджується з допустимими діями керованої деградації як периферійного інструмента вивільнення ресурсів. **Досягнуті результати** полягають у впровадженні формалізму віконної здійсненості та системи показників для оцінювання дотримання умов здійсненості, частки вікон зі зменшенням дефіциту, накопичувального дефіциту критичних сервісів і часу відновлення, що дає змогу відокремлювати обмеження середовища від властивостей керувальних рішень. **Висновки**: ігнорування можливості завершення узгодження стану сервісів у межах поточного вікна зв'язності

призводить до появи незавершених відновлювальних переходів і розбіжності між очікуваним ефектом відновлення й фактичною динамікою дефіциту, тоді як запропонований метод формує кроки, здійснені в поточному вікні зв'язності, та забезпечує зменшення дефіциту критичних сервісів у вікнах, де відновлення є можливим. Це підвищує живучість унаслідок скорочення тривалості деградаційних епізодів критичних функцій та зменшення накопичувального дефіциту. Практична цінність полягає в можливості використання методу для бортових інформаційних систем груп БпЛА та інших мобільних вузлів із вікнами зв'язності зі змінною тривалістю та пропускну здатністю.

Ключові слова: інформаційна система; живучість; критичний сервіс; керована деградація; вікно зв'язності; відновлення сервісу.

Бібліографічні описи / Bibliographic descriptions

Ткачов В. М. Метод відновлення критичних сервісів інформаційної системи на мобільній платформі в умовах керованої деградації. *Автоматизовані системи управління та прилади автоматики*. 2026. № 1 (188). С. 78–97. DOI: <https://doi.org/10.30837/0135-1710.2026.188.078>

Tkachov, V. (2026), "Method of restoring critical information system services on a mobile platform under conditions of controlled degradation", *Management Information System and Devices*, No. 1 (188), P. 78–97. DOI: <https://doi.org/10.30837/0135-1710.2026.188.078>

Чалий С. Ф., Лещинська І. О.

НЕЙРОСИМВОЛЬНИЙ ФРЕЙМВОРК ФОРМУВАННЯ МЕНТАЛЬНИХ МОДЕЛЕЙ РІШЕНЬ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ

Предметом роботи є процес побудови ментальних моделей рішень інтелектуальних систем для користувачів з різним рівнем компетентності та відтворення послідовностей їх взаємодії з інтелектуальними системами в структуроване ментальне подання. **Мета дослідження** – розробити нейросимвольний підхід, що інтегрує концептуальну ментальну модель, систему принципів побудови й нейросимвольну реалізацію ментальної моделі для забезпечення персоналізованих пояснень в інтелектуальних системах. **Завдання:** розробити нейросимвольний фреймворк побудови ментальних моделей; розширити систему принципів для персоналізації пояснень; експериментально перевірити нейросимвольний фреймворк формування ментальних моделей рішень інтелектуальних систем. **Методи дослідження** ґрунтуються на застосуванні нейросимвольного штучного інтелекту, що поєднує машинне навчання з причинно-наслідковим символьним міркуванням, використанні операторів лінійної темпоральної логіки для верифікації каузальних залежностей і адаптивному механізмі відбору концептів за рівнем компетентності користувача. **Результати дослідження.** Розроблено нейросимвольний фреймворк формування ментальних моделей рішень інтелектуальних систем, який містить: ментальну модель, що інтегрує множини позитивних і негативних властивостей рішення, каузальний граф причинно-наслідкових залежностей і лінгвістичні мітки для інтерпретації структури моделі природною мовою; запропоновано розширену систему принципів побудови ментальних моделей, яка до наявних принципів відповідності, множинності, неповноти й доповнення долучає нові принципи деталізації, темпоральної узгодженості та поведінкового виведення; створити адаптивну порогову функцію відбору концептів, що визначає рівень деталізації ментальної моделі відповідно до рівня компетентності конкретного користувача. Доведено, що система з чотирьох принципів є необхідною, але не достатньою умовою коректності нейросимвольної ментальної моделі. Експериментальна перевірка підтвердила суттєве підвищення темпоральної узгодженості каузальних залежностей порівняно з наявними методами. **Висновки.** Запропонований фреймворк забезпечує побудову персоналізованих ментальних моделей з верифікованими каузальними залежностями, відповідає вимогам прозорості моделей штучного інтелекту й забезпечує масштабовану персоналізацію пояснень для інтелектуальних систем.

Ключові слова: нейросимвольний штучний інтелект; ментальна модель; інтелектуальна інформаційна система; пояснювальний штучний інтелект; персоналізація; темпоральна узгодженість; каузальне міркування; формування пояснень.

Вступ

Інтелектуальні інформаційні системи (ІС) широко застосовуються для підтримки прийняття рішень у таких сферах, як медицина, фінанси, освіта, промисловість. Зростання складності внутрішніх моделей ІС, зокрема внаслідок використання глибоких нейронних мереж, призводить до формування рішень такими системами в режимі непрозорості "чорної скриньки", що суттєво ускладнює інтерпретацію цих рішень кінцевими користувачами. Нормативні європейські вимоги EU AI Act [1] та GDPR [2] орієнтують розробників ІС на забезпечення прозорості прийняття рішень за допомогою формування зрозумілих для користувачів пояснень. Формування пояснень досліджують у межах наукового напрямку пояснювального штучного інтелекту (ХАІ, *Explainable Artificial Intelligence*).

Традиційно методи ХАІ аналізують внесок вхідних ознак у рішення ІС і на цій основі генерують узагальнені пояснення. Такі пояснення не беруть до уваги особливості сприйняття конкретного користувача, що спричиняє зниження рівня довіри до ІС, а також ускладнює практичне застосування досягнутих результатів.

Для подолання вказаного обмеження й забезпечення персоналізованих пояснень можуть бути впроваджені ментальні моделі. Останні є внутрішнім когнітивним поданням, яке користувачі використовують для інтерпретації поведінки складних систем і прогнозування результатів їх роботи [3]. Тобто ментальна модель визначає розуміння користувачами ІС причинно-наслідкових зв'язків у отриманому рішенні, тому є основою для формування адаптивних персоналізованих пояснень [4]. Побудова ментальних моделей

на основі інформації щодо поведінки інтелектуальної системи є перспективним напрямом персоналізації ХАІ. Проте наявні підходи до побудови ментальних моделей орієнтовані насамперед на експертне виявлення когнітивних схем за результатами інтерв'ю [5].

Перспективним напрямом для розв'язання задач автоматизованого формування ментальних моделей є використання нейросимвольного штучного інтелекту (NeSy, Neurosymbolic AI). Останній поєднує традиційне машинне навчання з причинно-наслідковим міркуванням у межах символьних систем, що створює умови для побудови інтерпретованих моделей рішень ІС [6, 7]. Останнім часом розроблено низку складників NeSy-підходу до побудови ментальних моделей рішень ІС, зокрема концептуальну модель, що описує структуру ментального подання рішення через знання, концепти й досвід користувача [8]; систему принципів, що визначають вимоги до побудови ментальних моделей зовнішніх користувачів [9]; двошарову NeSy-архітектуру, яку можна застосовувати в побудові ментальної моделі [10], а також метод розроблення цієї нейросимвольної архітектури [11]. Перелічені складники є елементами, на основі яких може бути побудовано єдину систему формування ментальної моделі рішення ІС з використанням нейросимвольного підходу, яка має задовольняти вимоги щодо прозорості систем штучного інтелекту, визначених EU AI Act Article 13 [1].

Зазначене свідчить про актуальність розроблення комплексного нейросимвольного фреймворку, що інтегрує в єдину систему концептуальну ментальну модель, її нейросимвольну реалізацію, а також принципи побудови ментальних моделей з огляду на рівень компетентності користувачів на основі адаптивного механізму відбору концептів для побудови персоналізованих пояснень.

Аналіз літературних джерел і постановка проблеми дослідження

Концепцію ментальних моделей як когнітивного механізму інтерпретації складних систем запропоновано в роботі [3]. Відповідно до цієї концепції користувач формує внутрішній спрощений образ системи, з якою він взаємодіє. Цей спрощений образ дає змогу прогнозувати поведінку системи й оцінювати якість отриманих рішень. Тому ментальна

модель безпосередньо визначає ступінь довіри до рекомендацій ІС і ефективність застосування останніх для розв'язання практичних завдань користувача [4]. Парадигма пояснювального ХАІ 2.0 обґрунтована в дослідженні [12]. Як актуальний напрям розвитку пояснювального штучного інтелекту ХАІ 2.0 розглядає персоналізацію пояснень, що може бути реалізована на основі ментальних моделей.

Методи ХАІ першого покоління містять LIME та SHAP. Метод локальних агностичних пояснень (LIME, *Local Interpretable Model-Agnostic Explanations*) [13] апроксимує поведінку складного класифікатора через інтерпретовану локальну модель, що формується в межах досліджуваного прикладу. Пояснення не залежить від архітектури ІС. Проте LIME генерує неперсоналізоване спільне для всіх користувачів пояснення без огляду на рівень підготовки останніх. Цей метод не бере до уваги темпоральний аспект причинно-наслідкових зв'язків, тому не призначений для побудови ментальної моделі рішення ІС.

Метод Шеплі (SHAP, *SHapley Additive exPlanations*) [14] розроблений на основі теорії кооперативних ігор. Він розподіляє внесок кожної входної ознаки в отримане рішення. Метод SHAP реалізує глобальні та локальні пояснення. Однак він формує уніфіковані пояснення, ігноруючи особливості сприйняття цих пояснень користувачем. Це обмежує можливості його застосування для побудови ментальних моделей.

Зазначені методи призначені для побудови пояснень до рішень вже наявної ІС. Можливості інтеграції функціональності традиційної архітектури ІС з безпосередньою інтерпретацією рішень досягаються з упровадженням нейросимвольних методів.

Нейросимвольний штучний інтелект [6, 7] призначений для подолання обмеження традиційних нейронних мереж, пов'язаного з відсутністю символьного виведення для верифікації процесу досягнення результату. Поєднання традиційного машинного навчання з причинно-наслідковим міркуванням символьних систем створює умови для побудови моделей формування рішень, які можуть бути безпосередньо інтерпретовані.

Моделі концептного вузького місця (CBM, *Concept Bottleneck Models*) [15] інтегрують у архітектуру нейронної мережі явний концептний шар перед виходом класифікатора. Концепти забезпечують інтерпретацію зв'язку між ознаками й рішенням.

Проте множина концептів задається розробником апіорно. Відповідно, персоналізація в межах цього підходу не забезпечується.

Логічні тензорні мережі (LTN, *Logic Tensor Networks*) [16] реалізують нейросимвольне виведення через тензорне подання логічних аксіом і предикатів. LTN дають змогу задавати обмеження у вигляді логічних формул і навчати мережу, зважаючи на ці обмеження. Однак LTN не підтримують персоналізацію та не беруть до уваги темпоральний аспект у причинно-наслідкових залежностях, що не забезпечує узгодженості знань у ментальній моделі.

Нейросимвольне ймовірнісне логічне програмування (*DeepProbLog*) [17] поєднує нейронні компоненти через ймовірнісні предикати, що уможливує навчання мережі для розв'язання логічних завдань. Проте персоналізація відповідно до рівня компетентності користувача й перевірка темпоральної узгодженості в межах цього підходу не підтримується.

Нейросимвольне навчання концептів [18] забезпечує виведення абстрактних концептів зі спостережень у неструктурованих сценах. Однак цей підхід спрямований на аналіз статичних сцен, що не дає змоги застосувати його для побудови ментальної моделі.

Принципи побудови ментальних моделей рішень ІС для зовнішнього користувача уточнено в роботі [9]. Традиційні принципи відповідності, коли модель структурно має відповідати поясненням рішень ІС, множинності, коли ментальні моделі різних користувачів є унікальними, неповноти, коли модель є когнітивно спрощеним поданням рішення, розширено принципом доповнення, згідно з яким модель має зважати на негативні властивості рішення ІС. Проте питання персоналізації ментальних моделей у цьому дослідженні не розглянуто.

Отже, наявні принципи й підходи до побудови ментальних моделей, які забезпечують пояснюваність рішень інтелектуальних систем, не беруть до уваги можливості побудови моделей відповідно до рівня розуміння користувача, а також не забезпечують темпоральну верифікацію застосованих для побудови пояснень причинно-наслідкових залежностей.

Мета й завдання роботи

Метою дослідження є розроблення нейросимвольного підходу, що інтегрує

концептуальну ментальну модель рішення інтелектуальної системи, принципи побудови ментальних моделей, а також нейросимвольну реалізацію ментальної моделі, щоб забезпечити можливість побудови персоналізованих пояснень в інтелектуальних системах.

Для досягнення окресленої мети необхідно виконати такі завдання: розробити нейросимвольний фреймворк побудови ментальних моделей; запропонувати принципи побудови ментальних моделей для персоналізації пояснень в інтелектуальних системах; експериментально перевірити нейросимвольний фреймворк формування ментальних моделей рішень інтелектуальних систем.

Нейросимвольний фреймворк формування персоналізованої ментальної моделі

Нейросимвольний фреймворк призначений для відтворення послідовностей взаємодії користувача з ІС у ментальну модель рішення інтелектуальної системи:

$$\Omega: B_i \rightarrow M_i, \quad (1)$$

де B_i – упорядкована темпоральна послідовність подій, що відображають взаємодію i – користувача з ІС; M_i – ментальна модель, що охоплює: множину V_i^+ позитивних властивостей рішення ІС; множину V_i^- негативних властивостей рішення ІС; граф G_i , що відтворює каузальні залежності між властивостями рішення; множину лінгвістичних міток L_i , призначених для інтерпретації структури моделі природною мовою.

Визначення 1

Нейросимвольною ментальною моделлю рішення інтелектуальної системи називається чотирикомпонентна структура M_i , яка містить множини позитивних і негативних властивостей рішення системи, каузальні зв'язки між властивостями рішення, а також лінгвістичні мітки для інтерпретації структури моделі. Модель формується на основі даних щодо послідовності взаємодії користувача з інтелектуальною інформаційною системою.

Персоналізація цієї моделі в межах розробленого фреймворку потребує розширення запропонованої в праці [9] системи принципами відповідності, множинності, неповноти й доповнення.

Твердження

Система принципів відповідності, множинності, неповноти й доповнення є необхідною, але недостатньою умовою коректності нейросимвольної ментальної моделі рішення інтелектуальної системи в тому сенсі, що може існувати модель M_i , яка задовольняє ці чотири принципи, але порушує щонайменше одну з вимог: автоматизованого виведення ментальної моделі з послідовності взаємодії користувача з інтелектуальною системою (а); адаптації деталізації моделі до рівня компетентності користувача (б); темпоральної несуперечливості каузальних залежностей графа G_i (в).

Проведемо доказ від супротивного на основі трьох відповідних контрприкладів.

Контрприклад щодо вимоги (а). Нехай експерт предметної галузі побудував ментальну модель традиційно вручну, незважаючи на поведінкові дані користувача. Тоді вона задовольняє всі чотири принципи відповідності, множинності, неповноти й доповнення, проте не може бути відтворена й верифікована на основі інформації щодо взаємодії користувача з інтелектуальною системою, оскільки модель не містить механізму реалізації відображення (а).

Контрприклад щодо вимоги (б). Ментальна модель M_i , яка відповідає чотирьом принципам, може містити кількість концептів, що перевищує когнітивну ємність користувача-початківця, визначену законом Міллера: 7 ± 2 унікальних елементи [19, 20]. Отже, ця модель не може бути використана з метою адаптації пояснення для такого користувача. Ця суперечність пов'язана з тим, що принципи відповідності, множинності, неповноти й доповнення визначають склад і структуру властивостей рішення, але не обмежують кількість ключових властивостей, доступних для сприйняття користувачем.

Контрприклад щодо вимоги (в). Нехай каузальний граф у складі ментальної моделі містить дугу, де наслідок передуює причині в часі. Такий граф задовольняє розглянуті чотири принципи, оскільки останні визначають склад властивостей рішення, а не їх упорядкованість у часі. Тобто вимога (в) не є наслідком принципів відповідності, множинності, неповноти й доповнення.

Три розглянуті контрприклади є попарно незалежними, оскільки жоден з них не використовує умови іншого прикладу, що і свідчить про істинність цього твердження.

Отже, нейросимвольний фреймворк визначає реалізацію відображення (1) на основі принципів відповідності, множинності, неповноти й доповнення. Проте в побудові персоналізованих ментальних моделей ця система принципів потребує розширення.

Розширена систем принципів побудови ментальних моделей

У межах нейросимвольного фреймворку наявна система принципів побудови ментальних моделей має бути розширена з огляду на вимоги (а)–(в). Відповідно, розширена система містить принципи деталізації, темпоральної узгодженості та поведінкового виведення.

Принцип деталізації визначає, що ментальна модель повинна мати деталізацію відповідно до рівня підготовки користувача. Початківець зазвичай не здатний проаналізувати стільки ж концептів у складі рішення, що й фахівець. Надлишкова деталізація спричиняє перевантаження й пояснення стає "непрозорим" для користувача. Згідно з принципом деталізації, кількість активних концептів V_i^+ у ментальній моделі рішення має монотонно зростати з підвищенням компетентності користувача. Тобто

$$\Delta(M_i, u) = f(u) : \Delta_{\text{expert}} > \Delta_{\text{practic}} > \Delta_{\text{junor}}, \quad (2)$$

де $\Delta(M_i, u) \in [0, 1]$ – рівень деталізації ментальної моделі M_i відповідно до рівня u підготовки користувача; $f(u)$ – монотонна зростаюча функція компетентності, яка має найбільше значення Δ_{expert} для користувача-експерта й найменше значення Δ_{junor} для новачка.

Принцип темпоральної узгодженості визначає обов'язкову темпоральну упорядкованість каузального графа в ментальній моделі M_i , тобто причина завжди має передувати наслідку.

Важливість цього принципу пов'язана з тим, що побудований на основі аналізу взаємодії користувача з ПС граф може містити дуги з некоректним поданням часу, які не мають каузальної інтерпретації. Тому цей принцип задає обмеження в побудові причинно-наслідкового графа. Вимога темпоральної узгодженості задається з використання операторів \mathbf{G} та \mathbf{F} лінійної темпоральної логіки (LTL). Оператор \mathbf{G} (*globally*) вимагає виконання умови в усіх майбутніх станах

системи. Оператор \mathbf{F} (future) задає виконання умови колись у майбутньому стані. Тоді для довільної дуги (d_l, d_m) з графа G_i , для якої обов'язково буде істинним стан c_l , завжди в майбутньому має бути істинним стан d_m :

$$\forall (d_l, d_m) \in G_i : s(d_l, d_m) \Rightarrow \mathbf{G}d_l \rightarrow \mathbf{F}d_m. \quad (3)$$

Цей принцип дає змогу вилучити в процесі побудови графа G_i дуги (d_l, d_m) , що не задовольняють умову (3).

Принцип поведінкового виведення передбачає, що персоналізована ментальна модель має зважати

на послідовність взаємодії з ІС конкретного користувача (а не лише представлення експерта щодо об'єктів предметної галузі). Тобто ментальна модель рішення має бути функцією поведінкових даних за умови достатнього розміру n набору цих даних:

$$M_i = \Omega(B_i) \parallel |B_i| \geq n. \quad (4)$$

Запропонований принцип забезпечує відтворюваність ментальної моделі, що створює умови для побудови персоналізованих пояснень.

Відповідність розробленої системи принципів і описаної в дослідженні [10] нейросимвольної архітектури ментальної моделі подано в табл. 1.

Таблиця 1. Реалізація принципів побудови ментальної моделі в межах нейросимвольної архітектури

Принцип	Модуль архітектури	Механізм забезпечення
1) Відповідність структури	NeSy-транслятор	Перетворює вектор латентного простору на символічний концепт із онтології, тобто забезпечує зв'язок між нейромережним і символічним рівнями архітектури
2) Множинність	VAE-кодувач	Кодує індивідуальну послідовність взаємодії користувача з ІС у латентний прихований простір
3) Неповнота	Механізм уваги	Фільтрує концепти за релевантністю до поточного контексту рішення, тобто в ментальну модель додаються лише елементи, для яких перевищено поріг значущості для певного типу користувача
4) Доповнення	Модуль негативних ознак	Формує негативний контур ментальної моделі, що передбачає вилучення властивостей, що обмежують використання рішення
5) Деталізація	Адаптивний класифікатор типу користувача	Налаштовує рівень деталізації ментальної моделі під конкретний тип користувача, тобто новачок отримує узагальнене подання рішення, а фахівець – деталізоване подання
6) Темпоральна узгодженість	LTL-верифікатор	Перевіряє відповідність послідовності концептів у ментальній моделі порядку використання (аналізу) рішення користувачем і вилучає інвертовані в часі каузальні ланцюжки
7) Поведінкове виведення	Модуль збору поведінкових даних користувача	Логування даних про взаємодію користувача із ІС

Реалізація принципу адаптивної деталізації потребує формування механізму адаптації моделі. Такий механізм імплементовано через адаптивний поріг відбору концептів, що описують властивості рішення ІС. Експерименти, які провели автори в попередніх дослідженнях, дали змогу визначати базовий поріг відбору на рівні $\theta^* = 0,85$. Однак такий фіксований поріг може спричинити перенавантаження новачків. З іншого боку, набір понять у ментальній моделі може бути недостатнім для фахівців.

Для побудови адаптивного порогу пропонується лінійна модель уточнення значення порогу θ залежно від u – рівня підготовки користувачів:

$$\theta(u) = \theta^* - k \cdot \Delta(M_i, u), \quad (5)$$

де k – коефіцієнт адаптації.

Семантика формули (5) полягає в такому: вищий рівень компетентності означає збільшення $\Delta(M_i, u)$,

що приводить до зниження $\theta(u)$ за умови заданого базового рівня адаптації θ^* . Залежність є монотонно спадною в разі збільшення $\Delta(M_i, u)$, що задовольняє умову (2).

Експериментальна перевірка

Експериментальну перевірку розробленого фреймворку виконано із застосуванням набору відгуків щодо ноутбуків на українській маркетплейс-платформі. Відповіді щодо ноутбуків подано в текстовій формі. Тобто ідея експерименту полягає в тому, щоб на підставі аналізу відгуків оцінити рекомендації щодо товарів і послуг у системі електронної комерції.

На основі текстових відповідей сформовано датасет із відгуків користувачів. Особливість

датасету полягає в тому, що відгуки містять послідовність обговорення характеристик ноутбуків із мітками часу, наприклад обговорення щодо завантаження драйверів із відповідями служби підтримки виробника. Такий формат відгуків дає змогу проаналізувати послідовність взаємодії користувача із ПС.

У процесі формування датасету з кожного відгуку із застосуванням ЛЛІМ вилучено набір концептів, зокрема кількість слів у відгуку, терміни, застосовані у відгуку, числові значення показників у відгуку. Також відгук користувача структурується, тобто виокремлюються у відповіді переваги й недоліки і надається їх кількісна оцінка (кількість переваг і недоліків). На основі виділених концептів у відгуку розраховується його складність. Остання оцінюється як нормована сума кількості термінів, числових параметрів, а також переваг і недоліків. За складністю відгуку оцінюється рівень підготовки користувача. Застосовано емпіричні порогові правила. Отриманий датасет містить 44% фахових відгуків, 25% відгуків новачків та 31% відгуків досвідчених користувачів. Фахові відгуки містять не менше ніж п'ять технічних термінів і зазвичай наданих для ігрових ноутбуків. Покупцями цієї категорії комп'ютерів є, як правило, геймери з технічною підготовкою, які залишають розгорнуті відгуки. Ці відгуки, наприклад, містять виміри температур, fps, порівняння конфігурацій комп'ютера тощо. Новачки залишають короткі емоційні оцінки, наприклад "рекомендую", "купив і не шкодую" тощо. Такі оцінки не містять технічних термінів. Досвідчені користувачі залишають відгуки середньої довжини, що містять типові сценарії використання ноутбуків, наприклад "для роботи й навчання", "підходить для ігор". Однак і ці відгуки не мають технічних термінів.

Нейромережна частина ментальної моделі відтворює ключові ознаки рішення ("продуктивність", "автономність", "екран", "охолодження" тощо) в латентний простір ознак, як було продемонстровано в роботі [10].

Символьна частина формується через проміжний шар із нейромережної [10, 11] і може бути безпосередньо інтерпретована й проаналізована.

Оцінки складності отриманої ментальної моделі залежно від рівня підготовки користувачів подано в табл. 2.

Онтологія рішення ПС (рекомендованого продукту в системі електронної комерції) за результатами

експерименту містить 12 прийнятних для користувача концептів (властивостей рекомендованого продукту), зокрема: "продуктивність", "fps", "TDP", "автономність", "охолодження", "шум". Кількість концептів залежить від рівня підготовки (типу) користувача.

Таблиця 2. Оцінки структурної складності нейросимвольної ментальної моделі

Оцінка	<i>u нов</i>	<i>u пр</i>	<i>u фах</i>
V^+ – прийняті концепти	4	7	12
V^- – негативні концепти	8	5	0
$ G $ – кількість дуг каузального графа	3	6	18
$\Delta(M_i, u)$ – рівень деталізації моделі	0,30	0,60	1,00

Безрозмірний коефіцієнт $\Delta(M_i, u)$ має значення в діапазоні від 0 до 1 та кодує рівень деталізації відповідно до рівня компетентності користувача. Загальна ідея щодо вибору значень цього коефіцієнта полягає в такому. Користувач-новачок сприймає лише ключові поняття, які характеризують рекомендований товар або послугу в системі електронної комерції. Аналіз відгуків демонструє, що набір таких ключових понять становить приблизно 30% повного набору концептів. Просунутий користувач, крім базових понять, бере до уваги й сценарії застосування рекомендованого рішення, що зрештою становить понад половину набору концептів. А користувач-фахівець оперує концептами, що відтворюють усі, зокрема й технічні, властивості рекомендованого продукту.

З огляду на різний рівень деталізації каузальний граф у моделі нового користувача, яка має чотири концепти, містить лише три дуги, що відтворюють зв'язок між "продуктивність", "охолодження", "шум", "автономність". Граф у моделі фахового користувача містить, наприклад, дуги "DDR5 – {fps, теплопакет}", "CPU → {fps, температура, шум}".

Значення адаптивного порогу й темпоральної узгодженості, що встановлюють можливості персоналізації ментальної моделі, подано в табл. 3.

Адаптивний поріг $\theta(u)$ задає мінімальне значення подібності між концептом і послідовністю взаємодії користувача із ПС.

Базове значення коефіцієнта 0,85 було використано в попередніх роботах для побудови нейросимвольної моделі. Таке значення дає змогу не перевантажувати новачка додатковими деталізованими концептами, що відтворюють

властивості рішення ПС. Мінімальне значення коефіцієнта становить 0,72. Експерименти на різномісних наборах даних продемонстрували, що зниження коефіцієнта менше ніж 0,7–0,72 призводить до долучення несуттєвих концептів у ментальну моделі. Відповідно, за умови значення $\theta(u) = 0,81$ відбираються лише концепти, що відтворюють очевидні властивості рішення, а в разі $\theta(u) = 0,72$ в ментальну модель долучаються технічні властивості рішення.

Таблиця 3. Персоналізація ментальної моделі

Параметр	<i>u нов</i>	<i>u пр</i>	<i>u фах</i>
$\theta(u)$	0,81	0,77	0,72
Зниження θ щодо базового	-0,04	-0,08	-0,13
TСI	1,0	0,97	0,95
Зміна TСI щодо базового значення	+0,33	+0,3	+0,28

Показник TСI (*Temporal Consistency Index*) визначає частку дуг каузального графа, що відповідає темпоральним обмеженням. Значення 1,0 для

новачків демонструє, що модель містить лише ключові причинно-наслідкові зв'язки, в яких причина передре наслідку. Значення 0,95 для фахових користувачів свідчить, що одна з дуг може не задовольняти темпоральні обмеження. Значення приросту TСI щодо базового відтворює покращення моделі внаслідок переходу від фіксованого $\theta(u) = 0,85$ до адаптивного порогу. Порівняння отриманих коефіцієнтів (від 0,34 до 0,28) показує, що фіксований поріг не забезпечує деталізацію ментальної моделі насамперед для користувачів-новачків. Тобто такий поріг приводить до долучення до моделі технічних концептів, які не є релевантними для цих користувачів.

Крім цього, необхідно зауважити, що під час експерименту значення темпоральної узгодженості для методів LIME і SHAP визначались на основі порівняння з каузальним графом розробленої моделі.

Порівняння результатів роботи запропонованого фреймворку з позицій формування персоналізованих пояснень і наявних методів побудови пояснень за оцінкою TСI зображено на рис. 1.

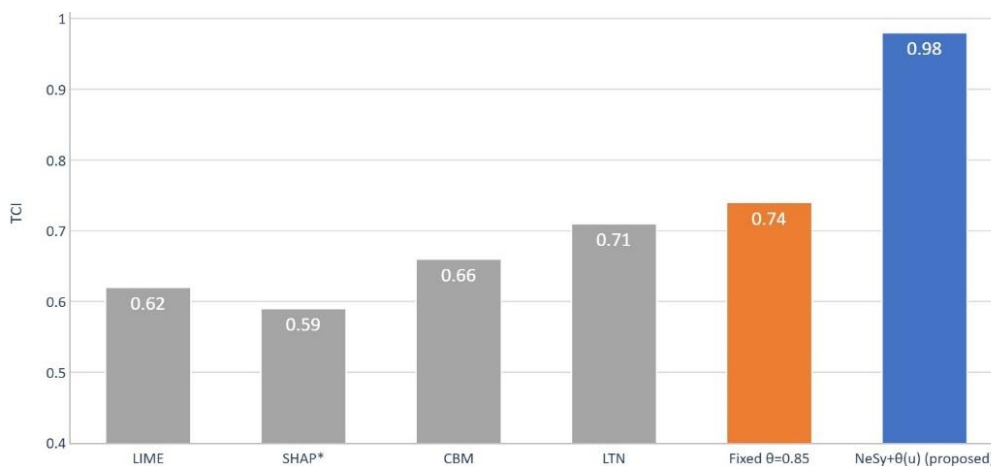


Рис. 1. Порівняльна оцінка TСI у процесі побудови ментальної моделі

Порівняння виконано в двох аспектах. По-перше, встановлено наявність суперечностей у поясненні з огляду на темпоральні залежності. По-друге, проаналізовано застосування розроблених принципів у процесі побудови пояснень.

Методи LIME і SHAP формують локальні пояснення без причинно-наслідкових залежностей. Вони беруть до уваги лише статистичні кореляції між властивостями рішення. Такий підхід приводить до зниження показника TСI до 0,62 для LIME та 0,59 для SHAP. Тобто користувач має сам

відібрати 62% та 59% ознак, відповідно, із запропонованого цими методами пояснення.

Метод CBM (*Concept Bottleneck Models*) додає концепти, але не застосовує деталізації концептів, що й визначає показник TСI на рівні 0,66. Метод LTN (*Logic Tensor Networks*) використовує темпоральні оператори для формування обмежень, унаслідок чого й підвищується значення темпоральної узгодженості. Однак метод не персоналізує пояснення.

Обґрунтування запропонованих на рис. 1 результатів способом аналізу реалізованих цими

методами принципів побудови ментальних моделей подано в табл. 4.

Як видно з табл. 4, перші п'ять методів не використовують повний набір принципів, що й зумовлює різницю в значеннях TCI.

Таблиця 4. Використання принципів побудови ментальних моделей для персоналізації пояснень

Метод	Реалізовані принципи
LIME	Відповідність структури (1), множинність (2)
SHAP	Відповідність структури (1), множинність (2)
CBM	Відповідність структури (1), множинність (2), неповнота (3)
LTN	Відповідність структури (1), множинність (2), неповнота (3), доповнення (4), темпоральна узгодженість (6)
NeSy з фіксованим порогом $\theta^* = 0,85$	Відповідність структури (1), множинність (2), неповнота (3), доповнення (4)
NeSy з адаптивним порогом $\theta(u)$	Відповідність структури (1), множинність (2), неповнота (3), доповнення (4), деталізація (5), темпоральна узгодженість (6), поведінкове виведення (7)

Обговорення результатів

Експериментально перевірені можливості розробленого нейросимвольного фреймворку зумовлені розбіжністю у поставці завдання, коли замість побудови типового пояснення для всіх користувачів формується персоналізована ментальна модель, поточна структура якої визначається за принципом деталізації згідно з рівнем підготовки користувача. Тобто кількість характеристик рішення і специфіка застосованих в описі рішення концептів залежать від рівня компетентності користувача інтелектуальної системи.

Застосування принципів деталізації, темпоральної узгодженості та поведінкового виведення підвищує релевантність отриманої моделі в темпоральному аспекті. Така перевага зумовлена тим, що використання заданого рівня деталізації обмежує множину можливих дуг у каузальному графі, а перевірка темпоральної узгодженості приводить до вилучення потенційних причинно-наслідкових залежностей, у яких наслідок передуює причині.

Обмеження цього фреймворку пов'язані з особливостями подання вхідних даних. Зокрема українськомовні відгуки в системі електронної комерції містять переважно короткі емоційні оцінки, водночас можна використовувати різні мови в межах одного відгуку, що утруднює визначення рівня підготовки користувача.

Ключовою особливістю методу є повторюваність персоналізованого пояснення, тобто за незмінних вхідних даних у межах фреймворку буде створено одну й ту саму ментальну модель, що забезпечує відповідність вимогам щодо прозорості ПС та можливість масштабування для платформ, які генерують значну кількість відгуків користувачів.

Напрями подальших досліджень пов'язані з розширенням онтології, що визначає характеристики рішення, а також з оцінюванням зрозумілості отриманих моделей на основі аналізу відгуків користувачів.

Висновки

Запропоновано нейросимвольний фреймворк формування ментальної моделі, який інтегрує концептуальне подання, систему принципів побудови, а також нейросимвольну реалізацію ментальної моделі. Це дає змогу сформуванню персоналізованих пояснень щодо рішення інтелектуальної системи з огляду на поточний рівень підготовки користувача.

Удосконалено систему принципів побудови ментальних моделей рішення інтелектуальної системи, яка відрізняється від наявної принципами деталізації, темпоральної узгодженості та поведінкового виведення ментальної моделі. Це дає змогу в межах нейросимвольного підходу формувати ментальну модель із заданим рівнем деталізації з можливістю верифікації отриманих причинно-наслідкових залежностей у темпоральному аспекті.

Виконана експериментальна перевірка розробленого фреймворку підтвердила суттєве підвищення темпоральної узгодженості каузальних залежностей порівняно з наявними методами, а також використання повної системи принципів побудови ментальних моделей для персоналізації пояснень щодо рішень інтелектуальних систем.

Конфлікт інтересів

Автори декларують, що не мають конфлікту інтересів, зокрема фінансового, особистого,

авторського чи будь-якого іншого характеру, який міг би вплинути на дослідження, а також на результати, опубліковані в цій статті.

Доступність даних

Рукопис не має пов'язаних даних.

Фінансування

Дослідження проведено без фінансової підтримки.

Використання засобів штучного інтелекту

Автори підтверджують, що не застосовували технології штучного інтелекту для написання статті.

References

1. Regulation (EU) 2024/1689 (2024), "Artificial Intelligence Act: laying down harmonised rules on artificial intelligence", *Official Journal of the European Union*, OJ L, 12.7.2024. URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689>
2. Regulation (EU) 2016/679 (2016), "General Data Protection Regulation: on the protection of natural persons with regard to the processing of personal data and on the free movement of such data", *Official Journal of the European Union*, L 119, P. 1–88.
3. Johnson-Laird, P. N. (1983), *Mental models: Towards a cognitive science of language, inference, and consciousness*, Harvard University Press, Cambridge, 513 p.
4. Bansal, G., Nushi, B., Kamar, E., Weld, D. S., Lasecki, W. S., Horvitz, E. (2019), "Updates in human-AI teams: Understanding and addressing the performance/compatibility tradeoff", *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, No. 1, P. 2429–2437. DOI: <https://doi.org/10.1609/aaai.v33i01.33012429>
5. Brown, O., Power, N., Gore, J. (2024), "Cognitive task analysis: Eliciting expert cognition in context", *Organizational Research Methods*. DOI: <https://doi.org/10.1177/10944281241271216>
6. Sarker, M. K., Zhou, L., Eberhart, A., Hitzler, P. (2021), "Neuro-symbolic artificial intelligence: Current trends", *AI Communications*, Vol. 34, No. 3, P. 197–209. DOI: <https://doi.org/10.3233/AIC-210084>
7. Hitzler, P., Eberhart, A., Ebrahimi, M., Sarker, M. K., Zhou, L. (2022), "Neuro-symbolic approaches in artificial intelligence", *National Science Review*, Vol. 9, No. 6. DOI: <https://doi.org/10.1093/nsr/nwac035>
8. Chalyi, S. F., Leshchynska, I. O. (2023), "Conceptual mental model of explanation in an artificial intelligence system", *Bulletin of the National Technical University "KhPI". Series: System Analysis, Control and Information Technology*, No. 1 (9), P. 70–75. DOI: <https://doi.org/10.20998/2079-0023.2023.01.11>
9. Chalyi, S. F., Leshchynska, I. O. (2024), "Principles of constructing mental models of decisions for an external user in the task of generating explanations in an intelligent system", *Automated Control Systems and Instruments of Automation*, No. 181, P. 82–90. DOI: <https://doi.org/10.30837/0135-1710.2024.181.082>
10. Chalyi, S. F., Leshchynska, I. O. (2026), "Integrated neuro-symbolic architecture of users' mental models for personalised explanations of intelligent system decisions", *Aerospace Technic and Technology*, No. 1 (208), P. 133–140. DOI: <https://doi.org/10.32620/akt.2026.1.12>
11. Chalyi, S. F., Leshchynska, I. O. (2025), "Method of constructing a neuro-symbolic representation of the mental model of an intelligent system's decision", *Bionics of Intelligence*, No. 2 (103), P. 108–115. DOI: [https://doi.org/10.30837/bi.2025.2\(103\).14](https://doi.org/10.30837/bi.2025.2(103).14)
12. Longo, L., Brcic, M., Cabitza, F., Choi, J., Confalonieri, R., Del Ser, J. et al. (2024), "Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions", *Information Fusion*, Vol. 106, P. 102301. DOI: <https://doi.org/10.1016/j.inffus.2024.102301>
13. Ribeiro, M. T., Singh, S., Guestrin, C. (2016), "Why should I trust you?: Explaining the predictions of any classifier", *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, USA, August 13–17, P. 1135–1144. DOI: <https://doi.org/10.1145/2939672.2939778>
14. Lundberg, S. M., Lee, S.-I. (2017), "A unified approach to interpreting model predictions", *Advances in Neural Information Processing Systems*, Vol. 30, P. 4765–4774.
15. Koh, P. W., Nguyen, T., Tang, Y. S., Pierson, E., Koh, J., Liang, P. (2020), "Concept bottleneck models", *Proceedings of the 37th International Conference on Machine Learning (ICML 2020)*, PMLR, Vol. 119, P. 5338–5348.
16. Badreddine, S., d'Avila Garcez, A., Serafini, L., Spranger, M. (2022), "Logic Tensor Networks", *Artificial Intelligence*, Vol. 303, P. 103649. DOI: <https://doi.org/10.1016/j.artint.2021.103649>
17. Manhaeve, R., Dumancic, S., Kimmig, A., Demeester, T., De Raedt, L. (2018), "DeepProbLog: Neural probabilistic logic programming", *Advances in Neural Information Processing Systems*, Vol. 31, P. 3753–3763.
18. Mao, J., Gan, C., Kohli, P., Tenenbaum, J. B., Wu, J. (2019), "The neuro-symbolic concept learner: Interpreting scenes, words, and sentences from natural supervision", *Proceedings of the 7th International Conference on Learning Representations (ICLR 2019)*, New Orleans, USA, May 6–9.
19. Miller, G. A. (1956), "The magical number seven, plus or minus two: Some limits on our capacity for processing information", *Psychological Review*, Vol. 63, No. 2, P. 81–97. DOI: <https://doi.org/10.1037/h0043158>

20. Cowan, N. (2001), "The magical number 4 in short-term memory: A reconsideration of mental storage capacity", *Behavioral and Brain Sciences*, Vol. 24, No. 1, P. 87–114. DOI: <https://doi.org/10.1017/S0140525X01003922>

Received (Надійшла) 11.02.2026

Accepted for publication (Прийнята до друку) 05.03.2026

Publication date (Дата публікації) 12.03.2026

Відомості про авторів / About the Authors

Чалий Сергій Федорович – доктор технічних наук, професор, Харківський національний університет радіоелектроніки, професор кафедри інтелектуальних управляючих систем; Харків, Україна;

Serhii Chalyi – Doctor of Technical Sciences, Professor, Kharkiv National University of Radio Electronics, Professor of the Intelligent Control Systems Department; Kharkiv, Ukraine;

e-mail: serhii.chalyi@nure.ua

ORCID ID: <https://orcid.org/0000-0002-9982-9091>

Лещинська Ірина Олександрівна – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри програмної інженерії; Харків, Україна;

Iryna Leshchynska – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Associate Professor at the Software Engineering Department; Kharkiv, Ukraine;

e-mail: iryna.leshchynska@nure.ua

ORCID ID: <https://orcid.org/0000-0002-8737-4595>

A NEUROSymbOLIC FRAMEWORK FOR FORMING MENTAL MODELS OF INTELLIGENT SYSTEM DECISIONS

The subject of the article is the process of constructing mental models of intelligent system decisions for users with varying levels of competence and mapping their interaction sequences with intelligent information systems into structured mental representations. **The purpose of the work** is to develop a neurosymbolic approach integrating a conceptual mental model, a system of construction principles, and a neurosymbolic implementation of the mental model to enable personalized explanation generation in intelligent systems. **Tasks:** development of a neurosymbolic framework for constructing mental models of intelligent system decisions; extension of the system of principles to support explanation personalization based on user competence level; experimental verification of the neurosymbolic framework for constructing mental models of intelligent system decisions. **Research methods** are based on neurosymbolic artificial intelligence combining machine learning with causal symbolic reasoning, the application of linear temporal logic operators for verifying causal dependencies, and an adaptive concept selection mechanism based on user competence level. **Results achieved.** A neurosymbolic framework for constructing mental models of intelligent system decisions has been developed, which includes: a mental model integrating sets of positive and negative decision properties, a causal graph of cause-and-effect dependencies, and linguistic labels for natural-language interpretation of the model structure; an extended system of mental model construction principles that supplements the existing principles of correspondence, multiplicity, incompleteness, and complementation with the new principles of detailing, temporal consistency, and behavioral inference; an adaptive concept selection threshold function that determines the detailing level of the mental model according to the competence level of a specific user. It has been formally proved that the existing four-principle system is a necessary but not sufficient condition for the correctness of a neurosymbolic mental model. Experimental verification confirmed a substantial improvement in the temporal consistency of causal dependencies compared to existing methods. **Conclusions.** The proposed framework enables construction of personalized mental models with verified causal dependencies, satisfies the transparency requirements of artificial intelligence models, and provides scalable explanation personalization for intelligent systems.

Keywords: neurosymbolic artificial intelligence; mental model; intelligent information system; explainable AI; personalization; temporal consistency; causal reasoning; explanation generation.

Бібліографічні описи / Bibliographic descriptions

Чалий С. Ф., Лещинська І. О. Нейросимвольний фреймворк формування ментальних моделей рішень інтелектуальних систем. *Автоматизовані системи управління та прилади автоматики*. 2026. № 1 (188). С. 98–107. DOI: <https://doi.org/10.30837/0135-1710.2026.188.098>

Chalyi, S., Leshchynska, I. (2026), "A neurosymbolic framework for forming mental models of intelligent system decisions", *Management Information System and Devices*, No. 1 (188), P. 98–107. DOI: <https://doi.org/10.30837/0135-1710.2026.188.098>

Yehor Korniienko, Oleksii Liashenko

A STUDY OF THE METHODOLOGICAL FOUNDATIONS FOR IMPLEMENTING BLOCKCHAIN AND SMART CONTRACTS IN ELECTRIC POWER MICROGRIDS

The subject of research covers the theoretical, methodological, and applied aspects of implementing blockchain technology and smart contracts into microgrid management systems, as well as the automation processes of energy resource exchange between participants of a distributed energy system. *The purpose of this work* is to investigate the methodological foundations for the application of blockchain and smart contracts in microgrids through the analysis of contemporary scientific research, systematization of approaches to consensus algorithm implementation, classification of smart contracts by application areas, and experimental verification of the proposed solutions. To achieve this goal, the following *tasks* were addressed: analyzing existing microgrid architectures and management methods; conducting a comparative analysis of consensus algorithms (PoW, PoS, PoA, PBFT, RAFT, etc.) regarding their applicability in private and public energy grids; developing a classification of smart contracts based on their application areas; and investigating software tools for implementing decentralized applications. **Research Methods.** The study employs system analysis methods to investigate microgrid architecture, comparative analysis to evaluate the efficiency of consensus algorithms, and classification methods for grouping smart contracts. For the practical part, computer modeling and experimental verification methods were used: smart contract development in Solidity, testing in the Remix IDE environment, and simulation of a local blockchain network using the Hardhat toolkit. **Research results.** The research systematized the methodological foundations for integrating blockchain into microgrids. It was determined that hybrid or private consensus models are most effective for energy trading within local communities. A classification of smart contracts was developed and justified, covering four levels: energy trading, monitoring, distributed management, and cybersecurity. The practical result is the implementation of the EnergyTrading smart contract, which successfully automates the process of listing offers and purchasing electricity, as confirmed by experiments in a local environment. The implementation of smart contracts allows for the creation of a reliable P2P platform for electricity trading without intermediaries, increasing economic efficiency for households. The functionality of the automated settlement mechanism was experimentally confirmed. At the same time, key challenges were identified: the limited scalability of existing blockchain solutions and the need to improve cyber defense against vulnerabilities in contract code. Further development requires adaptation of the legislative framework and modernization of the hardware components of energy grids.

Keywords: microgrid; blockchain; smart contract; P2P energy trading; consensus algorithm; Solidity; renewable energy sources; decentralization.

1. Introduction

In today's world, blockchain technologies play a significant role in various areas, ensuring greater transparency, security, and decentralization of processes. Blockchain, as a technology that implements the concept of a distributed database storing an ordered chain of records, is widely used in many modern areas of human activity, such as the financial sector [1], healthcare [2], supply chain logistics [3], and others.

The relevance of microgrids in the modern energy system stems from the need to improve energy efficiency in existing transmission lines [4], reducing carbon emissions [5], the effective use and integration of renewable energy sources such as solar, wind, hydro, and wave energy, and other factors related to the storage, generation, and transmission of electricity [6–8]. Microgrids in Ukraine represent a highly promising and

necessary direction in the context of energy sector development, given the instability of the system in recent years. In October 2022, the Government approved the Concept for the Implementation of "Smart Grids" in Ukraine and adopted a detailed Action Plan for its implementation by 2035 [9].

Blockchain can be implemented in several key aspects of microgrids, including decentralized energy management, where it ensures transparency and security in data exchange among participants. As a result of the expansion and digitalization of electricity distribution infrastructures, peer-to-peer energy trading has become a new paradigm for electricity trading within the microgrid system [10]. Blockchain-based energy system management can offer several advantages: data is protected against loss, tampering, and single points of failure, while security and privacy are enhanced through business rules in smart contracts [11]. Blockchain

technology enables secure transactions between parties that do not have prior trust in one another, eliminating the need for intermediaries [12].

2. Analysis of Recent Scientific Publications

Research on blockchain technologies in energy systems has been actively developing in recent years. A significant number of scientific papers are devoted to analyzing the possibilities of integrating distributed ledgers into various sectors of the economy and industry. In particular, Karadag et al. [1] systematized the application of blockchain in the financial sector, while Kasyapa and Vanmathi [2] conducted a comprehensive study of the implementation of this technology in healthcare, including the management of the pharmaceutical supply chain and the regulation of health insurance. Prakash [3] explored the potential of blockchain to enhance the transparency and efficiency of logistics systems.

In the energy sector, researchers are particularly interested in the efficiency of power grids and the integration of renewable energy sources. Ayaz and co-authors [4] analyzed methods for improving energy efficiency in smart grids, while Wang et al. [5] investigated the potential for reducing carbon emissions from microgrids through optimized management. The issue of integrating various types of renewable energy sources is addressed in the works of Dixit [6], Wang [7], and Pani [8], which cover solar, wind, and wave energy, respectively.

Regarding the application of blockchain directly in microgrids, Umar et al. [10] investigated decentralized energy trading using battery systems in community microgrids. Han et al. [11, 12] proposed a smart contract architecture for decentralized energy management and trading based on blockchain. Aghmadi et al. [13] conducted a systematic review of architecture, communications, and cybersecurity in networked microgrids.

The issue of organizing multi-microgrids has been investigated in the works of Inamdar [14] on nested microgrids, Alam [15] on networked microgrids, Ayir [16] on interconnected microgrids, and Lasseter [17] on coupled microgrids. Bullich-Massagué et al. [18] systematized micro-network clustering architectures.

Azbeq and co-authors [20] made a significant contribution to the study of consensus algorithms by conducting a comparative analysis of various consensus mechanisms in blockchain networks. Bach et al. [21]

compared the efficiency of consensus algorithms for different types of applications. Individual studies are devoted to specific algorithms: Islam [22] – Proof of Authority, Zheng and Feng [23] – PBFT, Bowman [24] – Proof of Elapsed Time, Xu [25] – RAFT, Moniz [26] – IBFT.

In the field of smart contracts for energy systems, Junaidi et al. [31] conducted a systematic review of blockchain solutions for demand management in power grids. Wang et al. [32] investigated the application of smart contracts for energy demand management. Practical implementations are represented by platforms such as Power Ledger [33], SunContract [34], and GridPlus [35].

An analysis of scientific publications indicates growing interest in integrating blockchain into microgrids. However, despite a significant amount of research, the issue of systematizing the methodological foundations for the application of smart contracts in microgrids remains underdeveloped. Most studies focus on individual aspects – consensus algorithms, network architecture, or practical implementations – while a comprehensive approach to the classification and methodology of smart contract implementation requires further research.

3. Research Objectives and Tasks

The objective of this work is to investigate the methodological foundations for the application of blockchain and smart contracts in microgrids by analyzing current scientific research, systematizing approaches to the use of consensus algorithms, classifying smart contracts by application areas, and experimentally verifying the proposed solutions.

To achieve this objective, the following tasks must be addressed:

- analyze existing approaches to the use of smart contracts in energy microgrids;
 - evaluate the advantages and limitations of various consensus algorithms for private and public blockchain systems;
 - identify key aspects of security and data management efficiency;
 - investigate practical scenarios for the application of smart contracts to automate energy trading, monitor resources, and enhance cybersecurity;
 - conduct experimental verification of the developed solutions using a local blockchain model as an example.
-

4. Microgrids and Their Architecture: Definitions, Advantages, and Management Methods

The term "microgrid" refers to the concept of single power supply subsystems connected to a small number of distributed energy resources (DERs), which can be either renewable or conventional sources, including photovoltaic panels, wind power, hydropower, internal combustion engines, gas turbines, and microturbines, which may belong to either the microgrid itself or to individual households. The use of distributed energy resources can lead to issues such as voltage fluctuations, short-term self-sufficiency challenges, high capital costs, and others. In a microgrid, distributed energy resources must be equipped with appropriate electronic interfaces and control systems to ensure operational flexibility while maintaining power quality and energy production.

Microgrids demonstrate compelling advantages over traditional power grids. These include increased reliability, reduced energy costs through integration with renewable energy sources, improved energy security due to self-sufficiency, environmental benefits from reduced carbon emissions, and enhanced flexibility. From a grid perspective, the main advantage of microgrids is that a microgrid is treated as a controllable system within the existing power grid, capable of functioning as a single consumer. From a household perspective, microgrids meet consumers' electricity needs at the local level, ensuring an uninterrupted power supply while reducing transmission losses and providing voltage support. Microgrids are particularly useful in remote areas or in conflict zones where the main power grid may be absent or unreliable. Additionally, microgrids can provide energy independence and resilience, which is especially important in areas prone to natural disasters such as hurricanes or earthquakes.

Blockchain is a type of distributed ledger technology, which is a special form of database maintained by a network of peer-to-peer nodes. The ledger consists of a linear chain of cryptographically linked "blocks", each containing a set of transactions, making these records immutable [19]. Copies of this ledger are stored by all network participants, ensuring its integrity through a "consensus" process, during which nodes collectively verify transactions and add them to the ledger. Therefore, one of the key aspects of blockchain systems is the choice of a consensus algorithm, which determines how nodes reach agreement on each data block. Consensus algorithms in blockchain networks

differ in their operating principles, security level, energy consumption, and productivity [20]. One of the best-known is Proof of Work (PoW), which provides high security but requires significant computational resources. An alternative to it is Proof of Stake (PoS), which reduces energy consumption by allowing nodes to validate transactions based on the number of tokens they hold. Further development led to the emergence of Delegated Proof of Stake (DPoS), where validators are elected by voting, which increases transaction speed. Meanwhile, Proof of Authority (PoA) uses a limited number of trusted participants to validate blocks, making it effective for private blockchains. To optimize storage resources, Proof of Capacity (PoC) / Proof of Space was proposed, where participants use disk space for mining.

Another approach is implemented by Proof of Burn (PoB), which involves burning coins to obtain the right to create new blocks. Distributed systems often use Practical Byzantine Fault Tolerance (PBFT), which effectively coordinates the network state among nodes, while Proof of Elapsed Time (PoET) is used in enterprise solutions due to its cost-effectiveness and low energy consumption. Also, among modern algorithms, it is worth noting Istanbul Byzantine Fault Tolerance (IBFT), which ensures fast block finalization, and Proof of Weight, which uses various participant weight parameters to achieve consensus.

Thus, the choice of algorithm depends on the type of blockchain and its purpose. Public networks most often use PoW, PoS, DPoS, PoC, PoB, and PoWeight, while private networks predominantly use Raft, PoA, IBFT, PoET, PBFT, and HotStuff. Depending on the requirements for security, energy efficiency, and transaction processing speed, the appropriate mechanism is implemented. For example, for a permissionless network with an emphasis on security and significant computational resources, Proof of Work is a suitable option.

Conversely, PoS or DPoS can be used when block validation speed is a priority. Additionally, PoS and DPoS provide speed, efficiency, and a sufficient level of security with lower energy consumption, which is critical for microgrids, especially in the case of autonomous (island) networks. The main differences between PoS and DPoS are that the latter provides higher speed and scalability due to a smaller number of delegates participating in consensus, but this may reduce the level of decentralization. PoS offers a more even distribution of power among network participants [21]. However, despite the fact that a permissionless (public) blockchain

is truly decentralized and open (with nearly zero risk of information being altered, edited, or deleted) unlike a non-public blockchain system, in reality, the participants in the microgrid – household owners – are known in advance, and trading occurs based on the identification of parties by territorial, physical, and other characteristics of households. Therefore, it is necessary to identify and highlight the advantages and disadvantages of

consensus algorithms specific to private blockchains, namely Proof of Authority [22], Practical Byzantine Fault Tolerance [23], Proof of Elapsed Time [24], RAFT [25], and Istanbul Byzantine Fault Tolerance [26] (one of the BFT algorithm family: Dynamic BFT [27], Democratic BFT [28], R-PBFT [29]). A comparative analysis of leading consensus algorithms is presented in Table 1.

Table 1. Comparative analysis of leading consensus algorithms

Consensus algorithm	Advantages	Restrictions
RAFT	simplicity and clarity; coordination based on "leaders"; fault tolerance; no risk of attack; high throughput under stable conditions.	single point of failure (dependence on the leader); scalability issues; delay during leader election; the node is limited.
PBFT	high fault tolerance; high scalability; access control; deterministic finality.	complexity; susceptibility to network delays.
PoET	energy efficiency; scalability; fast consensus.	relies on trusted execution environments (TEEs); security issues; limited adoption.
HotStuff	energy efficiency; scalability; simplicity and modularity; security and reliability.	delay; complex implementation; high resource consumption; centralization issues.
PoA	fast decision-making; low power consumption; ease of configuration; accountability and transparency.	dependence on trust; vulnerability to trust-based attacks.
IBFT	guaranteed finality; defined finality; unambiguous finality.	risks of centralization; complex implementation; susceptibility to network delays.

For the operation of microgrids that use blockchain to manage energy transactions, ensuring data security and integrity is critical. The primary tool for achieving this level of security is cryptography, specifically digital signatures and hash functions. When a household in a microgrid performs a transaction, a digital signature mathematically links it to that transaction. If the digital signature verifies the authenticity and integrity of the transaction data, this guarantees that the transaction has not been altered. Cryptographic hash functions play a key role in this process. They operate as mathematical algorithms that convert variable-length data into a fixed size, enabling efficient searching and minimizing the risk of collisions – situations where different data produce the same hash result. Thanks to this, blockchain is capable of maintaining a high level of security and stability even in decentralized networks, such as microgrids. Thus, cryptographic mechanisms form the foundation for ensuring trust and security in energy transactions within blockchain-based microgrids. Among the large number of

existing hash functions, the leaders are SHA-2, SHABAL, SHAVITE, Keccak, and Blake. The distinctive features of each algorithm include security characteristics, requirements for resistance, and computational resources [30]. Productivity metrics for hash functions show that there is a natural balance between computational speed and the level of security they provide. For example, although SHA-256 demonstrates better productivity in most scenarios, the choice of SHA-512 may be motivated by increased security requirements.

5. Analysis, Development, and Integration of Smart Contracts in Microgrids: Relevance and Applications

Smart contracts are programmable distributed systems that operate in a decentralized environment. They combine several key aspects of computer engineering, including software development, security, cryptography, distributed computing, and efficient

resource management. In addition to recording transactions or any other sequential information, a blockchain can also store smart contracts (SCs) that execute when certain conditions are met. A smart contract is added to the blockchain similarly to a transaction record. The compiled code and associated data are sent to the blockchain, where they are included in a block and added to the ledger via a consensus mechanism. Like transactions, smart contracts are secured by cryptographic hashing, making it impossible to alter or forge them.

Once deployed on the blockchain, a smart contract acts as a software process that executes when certain conditions arise, such as energy consumption, energy production, or any other processes on which the activation of the smart contract depends. The execution of the smart contract code takes place in a virtual environment distributed across all nodes of the blockchain.

From a methodological standpoint, the integration of smart contracts into microgrids requires a systematic approach involving several key stages. First, a domain analysis is conducted to identify business processes that can be automated. Next, a smart contract model is developed, taking into account security, execution efficiency, and compatibility with the selected blockchain protocol. The next step is testing and verifying the smart contract in a microgrid simulation environment, which allows for evaluating its behavior under real-world load conditions. The final stage of the methodology involves integrating the smart contract into the operational microgrid and continuously monitoring its execution, with the ability to make adjustments based on the data obtained. This approach ensures not only technical implementation but also a scientifically grounded methodology for applying blockchain technologies in microgrids, enhancing transparency, security, and the efficiency of energy resource management.

Smart contracts can be written in general-purpose programming languages such as Golang, Node.js, and Java, or in specialized languages such as Solidity, which is the first and most popular language for creating smart contracts to build decentralized applications (DApps). Smart contract features such as self-execution and automation, tamper-resistance, reliability, transparency and accessibility, security, speed and reliability, self-verification, computational productivity and costs, and dependence on a specific programming language make smart contracts a powerful tool for automating processes and ensuring security in various blockchain applications, including the energy sector.

Smart contracts automate various aspects of the execution and management of relationships between consumers, electricity producers, and other participants in the microgrid. For example, in the context of energy systems, smart contracts are used to ensure secure and transparent transactions between energy producers and consumers, as well as to control the reading, writing, and processing of data in the blockchain ledger. Blockchain-based smart contracts automate energy trading and Demand Response (DR) processes, enabling more efficient real-time balancing of supply and demand [31].

The application of smart contracts in microgrids is widespread; for example, one of the most common uses of smart contracts when building a blockchain-based microgrid is a situation where a homeowner (household) has a surplus of generated energy after installing solar panels or other renewable energy sources. In this case, a smart contract can establish the terms of the agreement between the owner and the energy consumer. The details of the smart contract may include the price and amount of energy, the duration of the agreement, as well as the conditions for terminating such a contract. Once the terms of the smart contract agreement have been agreed upon, it is deployed on the blockchain ledger. And in the future, when the solar panel generates excess energy and the smart contract conditions are met, the contract automatically initiates a transaction with the entity that needs this resource, such as the nearest household. Afterward, payment will be made in accordance with the terms of the contract. Since the transaction is recorded in the blockchain ledger, it is secure and transparent for both the seller and the buyer, and is available for further inspection and monitoring if necessary.

In addition, smart contracts can dynamically change the network configuration and the blockchain consensus algorithm. For example, a smart contract can be created that adjusts the block size, consensus mechanism, or transaction fees depending on current network dynamics. In particular, another application of smart contracts for decentralized energy systems is the development of a contract that can verify and prepare energy supply data for a database, and then analyze the stored data [32].

For example, the use of smart contracts in microgrids allows for the automation of the energy exchange process between producers and consumers, eliminating the need for centralized intermediaries. One of the key components of this approach is a mechanism for registering offers to sell electricity, which is implemented via a smart contract on a blockchain network. For example, let's consider

the listEnergy() method (Listing 1), which can be used in microgrids to register offers to sell electricity. This mechanism allows producers (sellers) to set the price

and available volume of energy, ensuring an automated process for concluding transactions in a smart contract implemented in the Solidity language.

Listing 1

```
// Adding an energy offer for sale
function listEnergy(uint256 _pricePerKWh, uint256 _availableKWh) external {
    require(availableKWh > 0, "Кількість енергії має бути більшою за 0");
    require(pricePerKWh > 0, "Ціна має бути більшою за 0");
    offerCount++;
    offers[offerCount] = EnergyOffer(msg.sender, _pricePerKWh, _availableKWh);
    emit EnergyListed(offerCount, msg.sender, _pricePerKWh, _availableKWh);
}
```

The method's operation involves several key steps. First, the input parameters are checked: the function takes two arguments – `_pricePerKWh`, which specifies the price per 1 kWh, and `_availableKWh`, which indicates the amount of energy available for sale. The contract verifies that both values are greater than zero, which prevents invalid operations. After that, the offer is registered: a new entry is added to the general list (offers), containing information about the seller, the specified price, and the available amount of energy. Additionally, the offer counter (`offerCount`) is incremented, ensuring a unique identifier for each transaction. The final step is the generation of an event (`emit EnergyListed`), which records the creation of a new offer on the blockchain. This allows all participants in the microgrid to track the current list of available energy, ensuring transparency and efficiency in trading operations.

It should be noted that there are various implementations of smart contracts in distributed energy systems around the world. One such example is Power Ledger [33], an Australian startup that has developed a blockchain platform for peer-to-peer energy trading using smart contracts to ensure secure and transparent transactions between energy producers and consumers. SunContract [34] allows users to set energy prices and choose suppliers, giving them greater control and flexibility in managing their energy needs. The platform is based on a peer-to-peer (P2P) energy trading model, which allows users to trade energy generated from renewable sources, such as solar and wind power. By utilizing smart contracts on the blockchain, SunContract ensures transparency and automation of transactions, making the energy trading process more efficient and accessible to all participants. GridPlus [35] is also actively working on solutions for integrating

blockchain into the energy sector, including smart contracts to automate processes and ensure transaction transparency. The platform aims to promote the use of renewable energy sources and improve the efficiency of energy systems.

6. Classification of Smart Contracts in Microgrids

The use of smart contracts is broad in scope, addressing the tasks and operational scenarios faced by microgrid engineers and designers.

After analyzing the variety of smart contract implementation options, the following layers were identified that characterize the application of smart contracts: energy trading and energy resource redistribution, as well as inspection, monitoring, and cybersecurity to ensure the stable operation of the electricity market and distribution processes within the microgrid. Figure 1 illustrates the relationship between contract types based on logical affiliation and application characteristics. Within the methodological approach, the classification of smart contracts ensures their orderly implementation in microgrids. The methodology includes identifying the functional zones of the microgrid, defining scenarios for the application of smart contracts, selecting the contract type according to the set tasks and execution conditions, as well as evaluating the effectiveness and security of the chosen solution. This approach not only formalizes the process of developing and integrating smart contracts but also creates a foundation for the scalable and scientifically sound implementation of blockchain technologies in energy systems.

Energy trading is one of the most common applications of blockchain in energy systems. A large number of studies analyze blockchain approaches based

on smart contracts for energy markets, which eliminate the need for third parties to create, regulate, or manage these markets. Such approaches can improve the efficiency of power systems, reduce peak loads, facilitate the participation of small sellers in the market, lower energy costs for residential users, and stimulate investment in renewable energy sources.

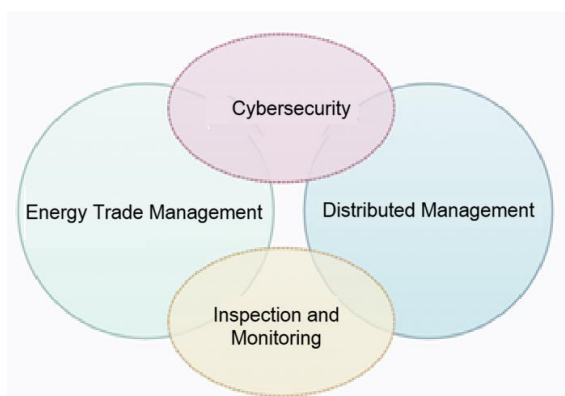


Fig. 1. Venn diagram illustrating the logical relationship between types of smart contract applications

Managing resource allocation in microgrids is a key challenge that can be significantly improved through smart contracts. Smart contracts can implement optimization algorithms. This allows for balancing load distribution among households and energy producers, ensuring a fair distribution of revenues. Smart contracts also help reduce peak load by automatically redistributing energy, leveraging the flexibility of distributed energy assets such as batteries or electric vehicles. In the case of decentralized battery management, smart contracts at this level can analyze information about battery charge status and capacity using "Inspection and Monitoring" level smart contracts, as well as send practical recommendations to synchronize charging and discharging processes. Additionally, such contracts can be used to automatically coordinate decisions among various network assets, thereby enhancing the efficiency and reliability of the energy system.

Inspection and monitoring are essential tools for the operation of any system, including the electricity trading

market and distribution management. Current metrics or metrics over a specified period play a key role in forecasting, management, and the stability of the microgrid. Monitoring electricity traffic, financial transactions, conflicts or malfunctions, and other indicators of the microgrid's operation will help other smart contracts make decisions and "activate" when certain levels or values are reached.

Smart contracts also play a vital role in ensuring cybersecurity within energy systems, including microgrids. They enable the automation and protection of processes related to data exchange and information access.

Smart contracts in the "cybersecurity" layer aim to ensure reliable user authentication and authorization, preventing unauthorized access to critical data. Additionally, they ensure transparency and the immutability of records in the distributed ledger, reducing the risk of data tampering or cyberattacks. An important aspect is the ability to create smart contracts that automatically monitor anomalies or disruptions in system operation and can activate threat response protocols. This ensures an increase in the overall level of cybersecurity in decentralized energy networks.

7. Ethereum-Based Energy Exchange

The goal of the experimental phase was to confirm the possibility of a secure, transparent, and automated exchange of energy resources between participants using smart contracts on the Ethereum platform. To achieve this goal, two experimental environments were implemented: the Remix IDE environment for initial testing and the local Hardhat simulated network for advanced analysis.

In the first phase, the EnergyTrading smart contract, written in Solidity, was implemented. This contract models the exchange's basic functionality: the registration of energy sale offers and the process of the buyer acquiring the corresponding volumes. The contract includes basic protections against overpayment, processing of available resource quantity checks, and logging events (Listing 2).

Listing 2

```
pragma solidity ^0.8.0;
contract EnergyTrading {
    event EnergyListed(uint id, address seller, uint pricePerKWh, uint availableKWh);
    event EnergyBought(uint id, address buyer, uint kWh);
    struct Offer {
        address seller;
```

```

    uint pricePerKWh;
    uint availableKWh;
  }

  mapping(uint => Offer) public offers;
  uint public nextOfferId;
  function listEnergy(uint pricePerKWh, uint availableKWh) public {
    offers[nextOfferId] = Offer(msg.sender, pricePerKWh, availableKWh);
    emit EnergyListed(nextOfferId, msg.sender, pricePerKWh, availableKWh);
    nextOfferId++;
  }
  function buyEnergy(uint offerId, uint kWh) public payable {
    Offer storage offer = offers[offerId];
    require(kWh <= offer.availableKWh, "Not enough energy available");
    require(msg.value >= kWh * offer.pricePerKWh, "Not enough ETH sent");
    offer.availableKWh -= kWh;
    payable(offer.seller).transfer(msg.value);
    emit EnergyBought(offerId, msg.sender, kWh);
  }
}

```

Further testing was conducted in two modes. The first involved using the Remix IDE, where the contract was successfully compiled, deployed, and manually tested under typical scenarios. The second involved the local Hardhat environment, which emulates a full-fledged blockchain network. Within Hardhat, a local chain was deployed using the built-in Proof-of-Work (PoW) consensus mechanism; a corresponding interact.js script was created; automated simulation of user interactions (seller/buyer) was implemented; and execution logs, transaction volumes, and changes in participants' balances were collected.

To perform an initial verification of the EnergyTrading smart contract's functionality, the Remix IDE environment – a tool for online development and testing of Solidity contracts – was used. This allowed for a quick verification of the basic logic without the need to set up a local blockchain.

The following steps were performed as part of the experiment:

1) the contract was compiled – the contract's bytecode was successfully generated using the Solidity 0.8.0 compiler. Next, the contract was deployed on the virtual machine (VM) built into Remix, which emulates the behavior of the Ethereum network with support for multiple accounts and balances;

2) an energy sale offer was created – the listEnergy(pricePerKWh, availableKWh) function was called with the initial parameters. As a result, a new offer was saved in the contract's storage, and the corresponding EnergyListed event was logged in the console; a purchase was made – the buyEnergy (offerId, kWh) function was called, specifying the amount of energy and the corresponding amount in wei. This resulted in the transfer of funds to the seller, an update to the balance, and the triggering of the EnergyBought event (Fig. 2).

The results showed that the contract's business logic functions as expected. Function calls were accompanied by corresponding transactions and events. Additionally, Remix provided a visual interface for debugging and tracking changes in the contract repository, which helped identify logical errors at an early stage. This phase served as validation of the smart contract prior to its deployment in a more complex environment simulating real blockchain execution.

After the initial validation of the contract in Remix, interaction was simulated in a local test network created using the Hardhat toolkit. This provided a flexible and controlled environment supporting multiple accounts, automated scripts, and the ability to emulate the consensus process.

```

[Block:5 txIndex:-] from: 0x8d4...95122 to: EnergyTrading.buyEnergy(uint256,uint256) 0x632...09895 value: 5000 wei data: 0xcdb...00005 logs: 1
hash: 0x9d5...5f55c
status 0x1 Transaction mined and execution succeed
transaction hash 0x14aeb16dbcc4da1cc7a2c9c2c0b446a2e9b6ca5882f3571a198cfadb35759269
block hash 0x9d5d368a15f968772ed698baf64ce50c45005bf13a8c54204f4e65b07415f55c
block number 5
from 0x8D41Ea6F2f80Ef648fA41a0c2a0c166612395122
to EnergyTrading.buyEnergy(uint256,uint256) 0x632a864b5dd70e20502c5356585c6de58c709895
gas 41158 gas
transaction cost 40786 gas
input 0xcdb...00005
decoded input {
  "uint256 offerId": "1",
  "uint256 kWh": "5"
}

```

Fig. 2. Purchasing electricity via the EnergyTrading smart contract

The EnergyTrading smart contract was compiled and deployed to the built-in Hardhat network, which by default simulates the Proof-of-Work (PoW) mechanism. Following this, interaction was implemented via a script that executed the typical behavior of energy market participants – the seller and the buyer. The scenario consisted of the following steps:

1) initiation of an energy sale offer. The seller's account called the listEnergy method, which registers the offer in the contract with the specified price and amount of energy;

2) simulation of an energy purchase. Another account (the buyer) called the buyEnergy method, transferring the corresponding amount of energy and ETH within the contract logic;

3) reading the state. After the transactions, the remaining available energy was retrieved by the offer ID, and the seller's balances before and after the transaction were calculated to verify the correctness of the transfer.

As a result, the functionality of the contract's core logic was confirmed. The system responded correctly to typical use cases, including edge cases (e.g., exceeding the available energy volume). The use of Hardhat ensured the determinism of the experiment, and the PoW consensus allowed us to simulate conditions similar to those of public Ethereum networks prior to the transition to Proof-of-Stake. This lays the groundwork for further research on the application of alternative consensus algorithms, such as PoA or PoS, in private microgrids.

8. Problems and Challenges

There are a number of significant challenges to implementing blockchain technology in microgrids,

one of which is scalability. Currently, blockchain solutions cannot efficiently execute the complex logic of a large number of smart contracts in a short time, requiring more and more resources as the blockchain system grows. A particular problem arises in the context of the overall power grid, where a large number of transactions per second is expected due to the diversity of end consumers, power generation points, and market regulation rules for resources. As of mid-2025, the integration of blockchain systems into Ukraine's overall power grid is impossible due to outdated equipment, limited support and management resources, and the absence of relevant legal regulations not only in Ukraine but globally.

The second issue is security and cybersecurity in the deployment of blockchain within the energy system, specifically in microgrids. Potential external interference with the energy system poses a threat at the national level and could cause serious problems for users, industry, and other consumers. Blockchain-based transactions must be secured at all levels of execution. Controlled management of consumer finances must be a priority when implementing blockchain in any power system. Therefore, improving software and hardware for user identification, encryption, and other means of protecting digital data offers a solution to cybersecurity challenges.

In turn, the deployment of smart contracts on the blockchain poses significant challenges in the areas of cybersecurity and privacy protection, due to the risk of losing access keys or their exposure for any reason. Smart contracts responsible for market control or resource allocation, despite being fully protected against forgery and having their defined methods and management remain unchanged after deployment, can

have negative consequences in the event of code vulnerabilities resulting from intentional or accidental errors on the part of the developer. This can lead to catastrophic consequences for the operation of the microgrid and the improper management of energy resources and finances of end consumers, electricity producers, and other microgrid participants.

Conclusions and Prospects

This article examined the methodological foundations for applying blockchain technologies in microgrids through a comprehensive analysis of the current state of energy systems, an assessment of the relevance of microgrids, and the potential of decentralized solutions. The architectural principles of microgrids and the possibilities for blockchain integration were examined based on scientific sources, practical implementations, and modern technical solutions. A comparison of consensus algorithms and a classification of blockchain access models allowed for the formulation of general principles for selecting technologies for energy applications.

An important component of the study was a review of smart contracts and their capabilities in the context of microgrids. Based on an analysis of the literature, existing solutions, and development trends, four key areas of their application were identified: inspection and monitoring, energy trading management, distributed management, and cybersecurity.

The proposed classification can serve as a methodological basis for further standardization of approaches to the implementation of smart contracts, taking into account regional requirements, regulatory constraints, and the specifics of power systems in different countries.

The study also identifies the main challenges that complicate the implementation of blockchain solutions in microgrids. Technical limitations include scalability issues and high demands on computing resources, which become critical in large systems with high transaction frequencies. Security challenges include the risks of

losing or compromising access keys, the possibility of errors in the immutable code of smart contracts, and vulnerabilities that could affect the operation of energy systems. The decentralized architecture provides protection against forgery, but at the same time exacerbates risks associated with developer errors or hidden software defects.

Given these limitations, a promising direction for further research is the development of methods for automated verification and testing of smart contracts prior to their deployment on the blockchain. Particular attention should be paid to identifying logical errors, malicious activation conditions, and potential vulnerabilities. The use of machine learning and artificial intelligence methods can ensure effective risk detection, increase the reliability of blockchain solutions, and accelerate the development of a standardized methodology for their application in microgrids.

Thus, the results obtained can serve as a basis for further applied developments, technology standardization, and the improvement of blockchain-based energy management systems.

Conflict of interest

The authors declare that they have no conflicts of interest regarding this study, including financial, personal, authorship, or other conflicts that could influence the study and its results presented in this article.

Funding

The study was conducted without financial support.

Data availability

The manuscript has no associated data.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence technology in the creation of this work.

References

1. Karadag, B., Zaim, A. H., Akbulut, A. (2024), "Blockchain in Finance: A Systematic Literature Review", *Journal of Business, Economics and Finance (JBEF)*, Vol. 13, No. 2, P. 113–130. DOI: <https://doi.org/10.17261/Pressacademia.2024.1945>
 2. Kasyapa, M. S. B., Vanmathi, C. (2024), "Blockchain integration in healthcare: a comprehensive investigation of use cases, performance issues, and mitigation strategies", *Frontiers in Digital Health*, Vol. 6, P. 1359858. DOI: <https://doi.org/10.3389/fgth.2024.1359858>
-

3. Prakash, A. (2024), "Blockchain Technology for Supply Chain Management: Enhancing Transparency and Efficiency", *International Journal for Global Academic & Scientific Research (IJGASR)*, Vol. 3, No. 2, P. 01–11. DOI: <https://doi.org/10.55938/ijgasr.v3i2.73>
4. Ayaz, K., Sulemani, M. S., Ahmed, N. (2017), "Efficient Energy Performance within Smart Grid", *Smart Grid and Renewable Energy*, Vol. 8, No. 3, P. 75–86. DOI: <https://doi.org/10.4236/sgre.2017.83005>
5. Gao, S., Wang, Z., Yang, Y., Li, C., Fan, J., Kou, J. (2024), "Economic Cost and Carbon Emission Reduction of Microgrid via Bi-Objective Optimization", *Proceedings of the 2024 43rd Chinese Control Conference (CCC), Kunming, China*, 28–31 July 2024, IEEE, P. 1–6. DOI: <https://doi.org/10.23919/CCC63176.2024.10662760>
6. Dixit, S., Singh, P., Ogale, J., Bansal, P., Sawle, Y. (2023), "Energy Management in Microgrids with Renewable Energy Sources and Demand Response", *Computers and Electrical Engineering*, Vol. 110, Article 108848. DOI: <https://doi.org/10.1016/j.compeleceng.2023.108848>
7. Wang, Y., Wang, Z., Sheng, H. (2024), "Optimizing wind turbine integration in microgrids through enhanced multi-control of energy storage and micro-resources for enhanced stability", *Journal of Cleaner Production*, Vol. 444, Article 140965. DOI: <https://doi.org/10.1016/j.jclepro.2024.140965>
8. Kodanda Pani, N. K., Ravi, A., Bai, L., Qiu, F., Zhao, S., Zhang, Y. (2023), "Enhancing Microgrid Resilience through Wave Energy Integration", *Proceedings of the 2023 North American Power Symposium (NAPS), Asheville, NC, USA*, 15–17 October 2023, IEEE. DOI: <https://doi.org/10.1109/NAPS58826.2023.10318531>
9. Верховна Рада України (2024), Проект Закону про внесення змін до деяких законів України щодо врегулювання окремих питань використання термінології у сфері впровадження "розумних мереж", URL: <https://itd.rada.gov.ua/billInfo/Bills/Card/43837>
10. Umar, A., Kumar, D., Ghose, T. (2022), "Blockchain-based decentralized energy intra-trading with battery storage flexibility in a community microgrid system", *Applied Energy*, Vol. 322, Article 119544. DOI: <https://doi.org/10.1016/j.apenergy.2022.119544>
11. Han, D., Zhang, C., Ping, J., Yan, Z. (2020), "Smart contract architecture for decentralized energy trading and management based on blockchains", *Energy*, Vol. 199, Article 117417. DOI: <https://doi.org/10.1016/j.energy.2020.117417>
12. Hahn, A., Singh, R., Liu, C.-C., Chen, S. (2017), "Smart contract-based campus demonstration of decentralized transactive energy auctions", *2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, Washington, DC, USA. DOI: <https://doi.org/10.1109/ISGT.2017.8086092>
13. Aghmadi, A., Hussein, H., Polara, K. H., Mohammed, O. (2023), "A Comprehensive Review of Architecture, Communication, and Cybersecurity in Networked Microgrid Systems", *Inventions*, Vol. 8, No. 4, Article 84. DOI: <https://doi.org/10.3390/inventions8040084>
14. Inamdar, S., Mohanty, R., Chen, P., Majumder, R., Bongiorno, M. (2019), "On Benefits and Challenges of Nested Microgrids", *2019 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, Macao, China, 01–04 December 2019. DOI: <https://doi.org/10.1109/APPEEC45492.2019.8994363>
15. Alam, M. N., Chakrabarti, S., Ghosh, A. (2019), "Networked Microgrids: State-of-the-Art and Future Perspectives", *IEEE Transactions on Industrial Informatics*, Vol. 15, No. 3, P. 1238–1250. DOI: <https://doi.org/10.1109/TII.2018.2881540>
16. Ayris, W., Helmi, A. M., Ramadan, H. S. (2024), "Interconnected microgrids optimization via reconfiguration-based modular approach", *Applied Energy*, Vol. 375, Article 124050. DOI: <https://doi.org/10.1016/j.apenergy.2024.124050>
17. Lasseter, R. H. (2011), "Smart Distribution: Coupled Microgrids", *Proceedings of the IEEE*, Vol. 99, No. 6, P. 1074–1082. DOI: <https://doi.org/10.1109/JPROC.2011.2114630>
18. Bullich-Massagué, E., Díaz-González, F., Aragüés-Peñalba, M., Girbau-Llistuella, F., Olivella-Rosell, P., Sumper, A. (2018), "Microgrid clustering architectures", *Applied Energy*, Vol. 212, P. 340–361. DOI: <https://doi.org/10.1016/j.apenergy.2017.12.048>
19. Liu, B. (2021), "Overview of the Basic Principles of Blockchain", *2021 International Conference on Intelligent Computing, Automation and Applications (ICAA), Nanjing, China*, 25–27 June 2021. DOI: <https://doi.org/10.1109/ICAA53760.2021.00108>
20. Azbeg, K., Ouchetto, O., Jai Andaloussi, S., Fetjah, L. (2021), "An Overview of Blockchain Consensus Algorithms: Comparison, Challenges and Future Directions", *Advances on Smart and Soft Computing. Advances in Intelligent Systems and Computing*, Vol. 1188, Springer, Singapore. DOI: https://doi.org/10.1007/978-981-15-6048-4_31
21. Bach, L. M., Mihaljević, B., Zagar, M. (2018), "Comparative Analysis of Blockchain Consensus Algorithms", *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija, Croatia, 21–25 May 2018, P. 1545–1550. DOI: <https://doi.org/10.23919/MIPRO.2018.8400278>
22. Islam, Md. M., Merlec, M. M., In, H. P. (2022), "A Comparative Analysis of Proof-of-Authority Consensus Algorithms: Aura vs Clique", *2022 IEEE International Conference on Services Computing (SCC)*, Barcelona, Spain, 10–16 July 2022, P. 327–332. DOI: <https://doi.org/10.1109/SCC55611.2022.00054>
23. Zheng, X., Feng, W. (2021), "Research on Practical Byzantine Fault Tolerant Consensus Algorithm Based on Blockchain", *Journal of Physics: Conference Series*, Vol. 1802, No. 3, Art. no. 032022. DOI: <https://doi.org/10.1088/1742-6596/1802/3/032022>

24. Bowman, M., Das, D., Mandal, A., Montgomery, H. (2021), "On Elapsed Time Consensus Protocols", *Progress in Cryptology – INDOCRYPT 2021. Lecture Notes in Computer Science*, Vol. 13143, Springer, Cham, 12–15 December 2021, Jaipur, India. DOI: https://doi.org/10.1007/978-3-030-92518-5_25
25. Xu, X., Hou, L., Li, Y., Geng, Y. (2021), "Weighted RAFT: An Improved Blockchain Consensus Mechanism for Internet of Things Application", *2021 7th International Conference on Computer and Communications (ICCC)*, Chengdu, China, 10–13 December 2021. DOI: <https://doi.org/10.1109/ICCC54389.2021.9674683>
26. Moniz, H. (2020), "The Istanbul BFT Consensus Algorithm", *arXiv preprint*, arXiv:2002.03613. DOI: <https://doi.org/10.48550/arXiv.2002.03613>
27. Xu, H., Long, Y., Liu, Z., Liu, Z. H., Gu, D. (2018), "Dynamic Practical Byzantine Fault Tolerance", *2018 IEEE Conference on Communications and Network Security (CNS)*, Beijing, China, 30 May – 1 June 2018, P. 1–8. DOI: <https://doi.org/10.1109/CNS.2018.8433150>
28. Crain, T., Gramoli, V., Larrea, M., Raynal, M. (2018), "DBFT: Efficient Leaderless Byzantine Consensus and its Application to Blockchains", *2018 IEEE 17th International Symposium on Network Computing and Applications (NCA)*, 1 November 2018, P. 1–8. DOI: <https://doi.org/10.1109/NCA.2018.8548057>
29. Liu, S., Wang, X., Hui, L., Wu, W. (2023), "Blockchain-Based Decentralized Federated Learning Method in Edge Computing Environment", *Applied Sciences*, Vol. 13, No. 3, Article 1677. DOI: <https://doi.org/10.3390/app13031677>
30. Kuznetsov, O., Peliukh, O., Poluyanenko, N., Bohucharskyi, S., Kolovanova, I. (2023), "Comparative Analysis of Cryptographic Hash Functions in Blockchain Systems", *CPITS-2023-II: Cybersecurity Providing in Information and Telecommunication Systems*, CEUR Workshop Proceedings, Vol. 3550, October 26, 2023, Kyiv, Ukraine, P. 81–94. URL: <https://ceur-ws.org/Vol-3550/paper7.pdf>
31. Junaidi, N., Abdullah, M. P., Alharbi, B., Shaaban, M. (2023), "Blockchain-based management of demand response in electric energy grids: A systematic review", *Energy Reports*, Vol. 9, P. 5075. DOI: <https://doi.org/10.1016/j.egyvr.2023.04.020>
32. Wang, X., Yang, W., Noor, S., Chen, C., Guo, M., van Dam, K. H. (2019), "Blockchain-based smart contract for energy demand management", *Energy Procedia*, Vol. 158, P. 2719–2724. DOI: <https://doi.org/10.1016/j.egypro.2019.02.028>
33. Power Ledger. Офіційний сайт. URL: <https://www.powerledger.io/> (дата звернення: 01.09.2025).
34. SunContract. Офіційний сайт. URL: <https://suncontract.org/> (дата звернення: 01.09.2025).
35. GridPlus. Офіційний сайт. URL: <https://gridplus.io/> (дата звернення: 01.09.2025).

Received (Надійшла) 08.11.2025

Accepted for publication (Прийнята до друку) 17.12.2025

Publication date (Дата публікації) 12.03.2026

Відомості про авторів / About the Authors

Корнієнко Єгор Дмитрович – Харківський національний університет радіоелектроніки, аспірант кафедри електронних обчислювальних машин; Харків, Україна;

Yehor Korniienko – Kharkiv National University of Radio Electronics, Postgraduate Student at the Department of Electronic Computers; Kharkiv, Ukraine;

e-mail: yehor.korniienko@nure.ua

ORCID ID: <https://orcid.org/0009-0002-7274-2815>

Ляшенко Олексій Сергійович – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри електронних обчислювальних машин; Харків, Україна;

Oleksii Liashenko – Phd (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Associate Professor at the Department of Electronic Computers; Kharkiv, Ukraine;

e-mail: oleksii.liashenko@nure.ua

ORCID ID: <https://orcid.org/0000-0002-0146-3934>

ДОСЛІДЖЕННЯ МЕТОДОЛОГІЧНИХ ОСНОВ УПРОВАДЖЕННЯ БЛОКЧЕЙНУ Й СМАРТ-КОНТРАКТІВ В ЕЛЕКТРОЕНЕРГЕТИЧНИХ МІКРОМЕРЕЖАХ

Предметом дослідження є теоретико-методологічні й прикладні аспекти впровадження технології блокчейн і смарт-контрактів у системи управління мікромережами (Microgrids), а також процеси автоматизації обміну енергетичними ресурсами між учасниками розподіленої енергосистеми. Мета роботи – дослідити методологічні основи застосування

блокчейну й смарт-контрактів у мікромережах за допомогою аналізу сучасних наукових студій, систематизації підходів до використання алгоритмів консенсусу, класифікації смарт-контрактів за сферами застосування та експериментальної перевірки запропонованих рішень. Для досягнення окресленої мети необхідно було виконати такі **завдання**: проаналізувати наявні архітектури мікромереж і методи їх управління; здійснити порівняльний аналіз алгоритмів консенсусу (PoW, PoS, PoA, PBFT, RAFT тощо) щодо їх застосовності в приватних і публічних енергетичних мережах; розробити класифікацію смарт-контрактів за сферами застосування; дослідити програмні засоби реалізації децентралізованих застосунків. **Методи дослідження.** Системний аналіз застосовано для вивчення архітектури мікромереж; порівняльний аналіз – для оцінювання ефективності алгоритмів консенсусу; класифікація – для групування смарт-контрактів. У практичній частині використано методи комп'ютерного моделювання та експериментальної перевірки: розроблення смарт-контракту мовою Solidity, тестування в середовищі Remix IDE й симуляція локальної блокчейн-мережі за допомогою інструментарію Hardhat. **Результати дослідження.** Систематизовано методологічні основи інтеграції блокчейну в мікромережі. Визначено, що для торгівлі енергією в межах локальних спільнот найбільш ефективними є гібридні або приватні моделі консенсусу. Розроблено та обґрунтовано класифікацію смарт-контрактів, що передбачає чотири рівні: торгівля енергією, моніторинг, розподілене управління й кібербезпека. Практичним результатом стала реалізація смарт-контракту EnergyTrading, який успішно автоматизує реєстрацію пропозицій та купівлю електроенергії, що підтверджено експериментом у локальному середовищі. Упровадження смарт-контрактів дає змогу створити надійну P2P-платформу для торгівлі електроенергією без посередників і водночас підвищити економічну ефективність для домогосподарств. Експериментально підтверджено працездатність механізму автоматизованих розрахунків. Крім цього, виявлено ключові виклики: обмежена масштабованість наявних блокчейн-рішень і необхідність удосконалення кіберзахисту від вразливостей у коді контрактів. Подальший розвиток потребує адаптації законодавчої бази й модернізації апаратної частини енергомереж.

Ключові слова: мікромережа; блокчейн; смарт-контракт; P2P-торгівля енергією; алгоритм консенсусу; Solidity; відновлювані джерела енергії; децентралізація.

Бібліографічні описи / Bibliographic descriptions

Корнієнко Є. Д., Ляшенко О. С. Дослідження методологічних основ упровадження блокчейну й смарт-контрактів в електроенергетичних мікромережах. *Автоматизовані системи управління та прилади автоматики*. 2026. № 1 (188). С. 108–120. DOI: <https://doi.org/10.30837/0135-1710.2026.188.108>

Korniienko, Y., Liashenko, O. (2026), "A study of the methodological foundations for implementing blockchain and smart contracts in electric power microgrids", *Management Information System and Devices*, No. 1 (188), P. 108–120. DOI: <https://doi.org/10.30837/0135-1710.2026.188.108>

Victoriya Usik, Bohdan Fedorchak

ANALYSIS OF THE STAGES OF CONDUCTING AN ACOUSTIC EXAMINATION OF CLOSED ROOMS

Relevance. The work considers the process of conducting acoustic expertise of closed premises for various purposes, taking into account the features of sound content, creating technical and design recommendations for obtaining optimal properties of the sound field. It is shown that for different types of premises (concert halls, conference halls, press centers, office premises, meeting rooms, etc.), there are specific requirements for the necessary properties of the sound field, sequences of stages of expertise, and features of each stage. Therefore, the current task is to develop recommendations for conducting acoustic expertise depending on the type of premises in order to formalize and further automate this process. **The subject of the study** is methods for determining the acoustic characteristics of closed premises and their optimization. **The purpose** of the work is to develop recommendations for creating optimal acoustic properties depending on the purpose of the premises and the features of sound content by analyzing the features of conducting acoustic expertise for different types of premises and adapting the stages of expertise for each type of premises. **Research objectives:** to analyze the stages of conducting acoustic expertise and the tasks that are solved at each stage; consider mathematical methods for determining acoustic characteristics, their areas of application and limitations; analyze the features of conducting an acoustic examination for different types of premises and adapt the stages of conducting the examination depending on the purpose of the premises and the characteristics of the sound content. **Research methods.** The paper uses statistical theory methods to calculate the reverberation time. Geometric theory considers sound as a set of rays and is used to analyze the structure of reflections from surfaces and detect acoustic defects in rooms with complex shapes. Wave theory is used to analyze the natural modes (resonances) of the premises. **The following results** were obtained. The stages of conducting an acoustic examination, the characteristics of the sound field that are investigated at each stage, mathematical methods for determining acoustic characteristics, and recommendations for their optimization were considered. The features of conducting each stage depending on the type of premises were noted. **Conclusions.** The analysis of the features of performing an acoustic examination for different types of premises will allow formalizing this process, reducing the subjective component and increasing the reliability of the results. The results obtained indicate the prospects for further research in the direction of further formalization and construction of a computer decision support system for conducting acoustic expertise.

Keywords: acoustic expertise; stages; sound content; acoustic properties; optimization; technical and design recommendations.

1. Introduction

Statement of the problem

The task of creating optimal acoustic conditions in a room as a whole, depending on its purpose, involves implementing recommendations aimed at creating and optimizing the sound field in the room.

The acoustic properties of a room are determined by the following factors: the volume and shape of the room; the number and presence of furniture; the size, shape, and construction of the enclosing surfaces; the materials used to finish the surfaces of the room and their distribution on the surfaces.

All of the above acoustic properties are interrelated with objective and subjective criteria for assessing room acoustics, such as: loudness-liveliness; spaciousness; intelligibility and clarity; loudness; warmth; absence of echo.

Thus, the tasks of acoustic expertise can be divided into stages:

- 1) verification of existing architectural and construction solutions for the interior surfaces of the room (geometry, finishing materials);
- 2) development of recommendations for increasing the diffusivity of the sound field and ensuring the optimal structure of the reverberation process (stage of creating technical specifications for the development of enclosing surface structures);
- 3) modeling of room acoustics with recommended finishing materials and enclosing surfaces recommended in the technical specifications;
- 4) measurement of the objective characteristics of the sound field synthesized after the implementation of the recommended measures to ensure optimal acoustic conditions.

The sequence and scope of the above stages may vary depending on the purpose of the room and the

characteristics of the sound content: rooms in which the characteristics of the sound field formation are taken into account; special-purpose rooms; rooms with special requirements for the objective and/or subjective characteristics of the sound field; rooms with special requirements for the type and quality of content transmission.

2. Analysis of recent studies and publications

The acoustic examination process is carried out in accordance with regulatory documents [1–3]. The peculiarities of sound content perception by listeners are analyzed in [4–6]. In [6–8], the general theory of sound field formation, boundary conditions, and methods for calculating objective sound field indicators, such as reverberation time (RT60), content clarity indicators (C7, C50, C80), and speech content intelligibility indicators (%ALcons, STI, RASTI) [9, 10]. An analysis of the features of each stage of acoustic expertise is given in [11–13]. At the same time, the specifics of conducting the stages of expertise for the following types of premises are noted: concert halls, conference rooms, press centers, and meeting rooms. Despite the sufficient volume of regulatory documents and publications on the subject of acoustic expertise, today there is an influence of subjective factors, experience, and qualifications of experts on the results of acoustic expertise, and there is no single formalized approach that takes into account the specific requirements for sound field parameters for each type of room when conducting acoustic expertise. Thus, the urgent task is to develop recommendations for conducting acoustic expertise depending on the type of premises in order to formalize all stages of acoustic expertise of premises and further automate this process.

The scientific novelty of the work is the improvement of the acoustic examination process by adapting the sequence and content of each stage to the functional purpose of the premises, the nature of the content, and the design and architectural constraints, which will formalize this process, reduce the subjective component, and increase the reliability of the results.

3. Purpose and objectives of the study

The purpose of the study is to develop recommendations for creating optimal acoustic properties depending on the purpose of the room and the

characteristics of the sound content by analyzing the features of conducting acoustic expertise for different types of rooms and adapting the stages of conducting expertise for each type of room. To achieve this goal, the following tasks must be solved:

- analyze the stages of acoustic expertise and the tasks that are solved at each stage;
- to consider the mathematical methods used to determine acoustic characteristics, their range of definition, and limitations;
- to analyze the features of acoustic expertise for different types of rooms and adapt the stages of expertise depending on the purpose of the room and the characteristics of the sound content.

4. Analysis of the acoustic examination stages

4.1. First stage

The first stage of acoustic expertise involves checking architectural and construction solutions for the configuration of enclosing surfaces of walls and ceilings in rooms using geometric theory of sound propagation in rooms [14–16]. The main purpose of the check is to identify surfaces that focus sound rays, which can lead to a significant reduction in the unevenness (diffuseness) of the sound field, as well as to check for the presence of parallel surfaces that create conditions for the formation of stable interference patterns (standing waves) for a number of resonant frequencies and, as a result, echoes, reverberation, and excessive spaciousness. There will be multiple reflections of sound.

The first stage of acoustic expertise also includes determining the frequency dependence of the reverberation time T60 when treating enclosing surfaces with materials proposed by architects.

4.2. Second stage

At this stage, measures are developed to correct the acoustic deficiencies of the room that were identified in the first stage. The following are developed: recommendations on the geometry of enclosing surfaces to increase the diffusivity of the sound field; recommendations on the selection of sound-reflecting and sound-absorbing materials for finishing surfaces in the room in order to optimize the frequency dependence of the reverberation time and the structures of the reverberation process [11–13].

4.3. Third stage

Provides the opportunity to: verify the design solutions for the geometry of wall and ceiling panels proposed in the first and second stages of acoustic expertise; analyze the frequency dependence of reverberation time using the interior finishing materials proposed at the second stage; analyze the structures of reverberation processes at control points [11–13].

4.4. Fourth stage

The results of the final stage of acoustic testing are the characteristics of the sound field measured at control points in the room after completion of all construction and finishing works. To perform measurements, a hardware and software complex based on Easera 1.1.3 is usually used [11, 13, 15].

5. Mathematical methods for determining acoustic characteristics

When conducting acoustic testing of enclosed spaces, three fundamental mathematical theories are used, each of which is applied depending on the frequency range and size of the space.

5.1. Statistical theory (Sabine's theory)

It is used to analyze diffuse sound fields where sound is considered to be evenly distributed. Based on statistical theory, the reverberation time T_{60} is calculated – the main indicator of the acoustic quality of a room. The statistical formulas of Sabine and Eiring are typically used for this purpose.

1. Sabine's formula is used for "boomy" rooms with low absorption (when the average sound absorption coefficient $\alpha_{cp} < 0.2$

$$T_{60} = \frac{0,161V}{A} = \frac{0,161V}{\sum_i S_i \alpha_i}, \quad (1)$$

where V – room volume (m^3); A – equivalent absorption area (m^2 or "Sabines"), which is equal to the sum of the products of surface areas and their absorption coefficients.

2. The Eiring formula is considered more accurate for rooms with high absorption levels (studios, cinemas), as it takes into account the attenuation of sound with each reflection.

$$T_{60} = \frac{0,161V}{-S \ln(1 - \alpha_{cp})}, \quad (2)$$

where S – total area of interior surfaces of the premises (m^2); α_{cp} – average sound absorption coefficient, calculated as $\alpha_{cp} = A/S$; \ln – natural logarithm.

Sabine's formula (1) gives an error in rooms with a lot of soft materials, while Eiring's formula (2) better models the situation when sound is almost completely absorbed (at $\alpha = 1$, the reverberation time according to Eiring is zero, which is physically correct).

For very large rooms (with a volume of more than $1000\text{--}2000 m^3$), the term $4\mu V$ is added to the formula, where μ is the sound attenuation coefficient in the air itself.

5.2. Geometric (radial) theory

It considers sound as a set of rays that propagate in a straight line inside a room and are reflected from surfaces according to the laws of geometric optics. The mathematical basis of geometric theory is the laws of reflection (the angle of incidence equals the angle of reflection) and the methods of mirror sources or ray tracing. Geometric theory is used to analyze the structure of early reflections and detect acoustic defects (echoes, sound focusing) in complex rooms.

5.3. Wave theory

It is based on solving differential equations of the sound field and takes into account the wave nature of sound (diffraction, interference). The mathematical basis is the wave equation (Helmholtz equation) and the analysis of the room's eigenmodes (resonances).

The propagation of sound in space is described by the classical **wave equation** for sound pressure

$$\nabla^2 p - \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} = 0, \quad (3)$$

where ∇^2 – Laplace operator; p – sound pressure; c – speed of sound; t – time.

For stationary processes (when the source emits a pure tone with frequency f), the wave equation (3) transforms into Helmholtz's equation:

$$\nabla^2 p + k^2 p = 0,$$

where $k = 2\pi f/c$ – wave number.

Solving this equation, taking into account the boundary conditions (room walls), allows us to find the natural frequencies at which resonance occurs.

Analysis of the premises' own modes

Natural modes are standing waves that arise between parallel or angular surfaces. For a rectangular room, natural frequencies are calculated using Rayleigh's formula:

$$f_{n_x, n_y, n_z} = \frac{c}{2} \sqrt{\left(n_x/L\right)^2 + \left(n_y/W\right)^2 + \left(n_z/H\right)^2}, \quad (4)$$

where L, W, H – length, width, and height of the room; n_x, n_y, n_z – integers (0, 1, 2...), which indicate the order of the mode.

Classification of waves (modes)

1) **Axial:** the wave is reflected between two parallel walls (for example, $n_x = 1, n_y = 0, n_z = 0$). They carry the most energy and have the greatest impact on the "color" of the sound.

2) **Tangential:** formed after reflections from two pairs of parallel surfaces (for example, $n_x = 1, n_y = 1, n_z = 0$). The energy of such waves is approximately 2 times less than that of axial waves.

3) **Oblique:** formed after reflections from three pairs of parallel surfaces. They have the lowest energy (at least 4 times less than axial ones).

Wave theory is used to analyze the acoustic field in small rooms at low frequencies, where standing waves occur (room "hum").

It should be noted that the application of wave analysis is critical only up to a certain frequency, above which the modes become so dense that the field can be considered diffuse. This limit is called the *Schreder frequency* (f_s):

$$f_s = 2000 \sqrt{T_{60}/V}.$$

At frequencies below f_s , wave theory must be used. Bass traps and geometry correction are needed to combat "booming".

At frequencies above f_s , statistical theory (Sabine and Eiring formulas) is used.

In addition, in modern expertise of rooms with complex configurations, methods of random process theory for noise diagnostics and mathematical modeling are actively being introduced, in particular using the finite

element method (FEM), which allows you to visualize nodes and pressure points at each point in the room.

6. Features of conducting acoustic expertise for different types of premises

A) Concert halls. To conduct the first stage of the examination in the auditorium, software modules [4, 11, 12] are used to construct ray patterns on plans and sections of the auditorium, which makes it possible to check the geometry of wall structures on all floors, the geometry of the ceiling, balcony railings, etc.

The main reason for using different wall panel geometry options is to try to eliminate the parallelism of surfaces, which will reduce the likelihood of standing waves forming and eliminate excessively high amplitude reflections in the reverberation process.

B) Conference rooms. For small premises (meeting rooms, conference rooms, meeting halls, press centers), the first stage of acoustic expertise is conducted differently than for auditoriums [14]. A distinctive feature of these premises is that they are usually rectangular in shape and all surfaces are parallel to each other. In such conditions, the reflection of waves from parallel surfaces and the formation of standing waves leads to the appearance of echoes and an increase in reverberation time [4, 15]. In addition, with small room sizes, it is practically impossible to change the geometry of the surfaces in order to reduce the number of parallel surfaces. With large glass surfaces, there is a problem of developing effective measures to reduce reflection amplitudes due to the low sound absorption coefficient of such surfaces.

Since the rooms are small in volume, it is not always possible to change the geometry of the enclosing surfaces for a number of reasons:

- the surfaces are made of glass and cannot be changed in shape;
- since the volume of the room is small, the dimensions of the enclosing surfaces are significantly smaller than the wavelength, especially in the low-frequency range.

These features of the rooms impose corresponding restrictions on the correction of sound field parameters, namely, we can only make corrections in a limited frequency range where the wavelength is commensurate with the size of the room. However, this limitation is not critical because the main sound material in the rooms

under study is speech content, for which the frequency range corresponds to the size of the room.

To optimize the frequency dependence of the reverberation time, finishing materials that work in the speech frequency range are used [7, 16]. In the presence of glass wall surfaces, it is practically impossible to change the acoustic properties by replacing the material on them, and then one of the options for increasing the sound absorption of glass surfaces is the use of curtains. Curtains are made from special "acoustic" fabrics with different sound absorption properties, or from fabrics used to make stage costumes.

Another option used to correct the acoustic properties of the rooms in question is the use of additional suspended ceiling structures such as "frameless sound-absorbing ceiling islands" and vertical sound-absorbing "Baffle" panels.

7. Adaptation of the acoustic examination process to the functional purpose of the premises

Based on the above, it should be noted that acoustic expertise is implemented as a sequence of stages that is sensitive to the functional purpose of the room, the nature of the content, and the structural and architectural limitations, which significantly affects both the content of individual stages and the possibility of changing the sequence of stages in order to adapt the acoustic expertise process to a specific type of room.

Thus, it is possible to generalize such adaptation for the four most representative types of rooms:

1. *For auditoriums, acoustic expertise is structured as a sequence of stages* [11, 12]: first, the existing architectural and construction solutions are checked (to find surfaces that focus sound and impair diffusivity), while at the same time the frequency dependence of the reverberation time and its compliance with the optimal values are assessed. Next, if any inconsistencies are found, recommendations are formulated – requirements for the geometry of the enclosing surfaces are formulated and reflective and absorbent materials are selected or placed to obtain the desired reverberation structure. After that, acoustic modeling of the hall is performed taking into account the proposed solutions, and as a final step, measurements of the objective characteristics of the sound field can be carried out after the completion of construction and finishing works.

2. *For small-volume rooms with predominantly speech content* [12, 14], the general sequence of the examination stages does not change compared to concert halls, but the content of the key steps changes. At the first stage, instead of focusing on the geometric verification of enclosing surfaces, the main action is to analyze the structures of the reverberation process at listening points in order to identify deficiencies in the formation of the sound field (for example, insufficient field diffusivity). The second stage focuses on recommendations for improving sound absorption and eliminating multiple reflections between parallel surfaces (in particular through local changes in geometry, suspended structures, curtains, etc.), while the subsequent stages are performed without fundamental changes.

3. For rooms such as press centers, the methodology is described as typically consisting of three stages, but depending on the initial conditions, the "role and place" of the stages may vary, in particular due to the purpose of the room (high requirements for speech clarity/intelligibility) or the inability to significantly change the geometry due to small dimensions. In the examples given [12, 17], the list of tasks was shifted towards the tools of the second and third stages (recommendations/modeling), because it was during the modeling and analysis of reverberation structures that "problematic" parallel surfaces were identified. Because of this, it is noted that the shortcomings, which in the usual sequence would have been found in the first stage, were actually found later, and as a result, the first stage became the final one, while no stage of the examination was excluded.

4. For special rooms [13, 17], where there are special requirements for the quality of speech content and/or a sound amplification system is already in use, a new sequence is proposed: to start with the "final" fourth stage as the first. This approach involves first determining the conditions for research, taking into account the specifics of the premises and content, selecting the hardware and software complex and informative evaluation criteria, and then processing and analyzing the results. The advantage is that it is possible to identify "problematic" areas of sound field formation and possible ways of correction early on, even before the stage of checking architectural and construction solutions and developing technical specifications for enclosing structures.

Thus, the specific sequence and content of the stages of acoustic expertise are determined by a specialist

individually for each object, taking into account the functional purpose of the room, the type of dominant content, geometric and structural limitations, as well as the possibilities for making changes to planning and finishing solutions. In fact, the choice of methods, priorities, and "control points" is largely based on the professional experience of the expert and accumulated practical experience, which provides flexibility but at the same time increases subjectivity and complicates the reproducibility of results by different performers. In this context, it is advisable to formalize and, if possible, automate the process of forming a sequence of stages and corresponding assessment procedures based on the type of premises, initial conditions, and target quality criteria. Automation (in the form of a decision support system) would allow standardizing the approach to selecting stages, reducing the influence of the human factor, shortening the time required to prepare the assessment, and increasing the comparability of results, while retaining the possibility of expert adjustment in non-standard cases.

In view of the above, a promising direction for further development is to expand the functionality of such a system through intelligent mechanisms for working with data and queries: in particular, the introduction of semantic search in the knowledge/measurement base and typical cases, the individualization of results (for example, by adapting recommendations to a specific type of room and nature of use based on accumulated behavioral data), as well as the use of external NLP modules to improve the accuracy of processing complex text queries and generating more relevant recommendations. This will not only automate the selection of the sequence of examination stages, but also increase the accuracy, convenience, and practical value of the results for the end user.

9. Conclusions and prospects for further research

The paper showed that the main goal of acoustic expertise in enclosed spaces is to create optimal acoustic properties of the sound field. At the same time, acoustic expertise is implemented as a sequence of stages that is sensitive to the functional purpose of the room, design constraints, and the characteristics of the sound content, which has a significant impact on both the content of individual stages and the possibility of changing the sequence of stages. The specific sequence and content of the stages of acoustic expertise for each object are

usually determined by a specialist based on their professional experience.

To achieve the set goal, the study analyzed the four stages of acoustic expertise. The tasks to be solved at each stage were formulated, in particular:

- at the first stage, architectural and construction solutions are checked in order to identify surfaces that focus sound rays, which can lead to a significant reduction in the diffusivity of the sound field, as well as checking for the presence of parallel surfaces that create conditions for the formation of standing waves;

- at the second stage, measures are developed to correct the acoustic deficiencies of the room (recommendations on the selection of sound-reflecting and sound-absorbing materials);

- at the third stage, the design solutions proposed at the first and second stages are checked by means of acoustic modeling;

- the results of the fourth stage are the characteristics of the sound field measured at control points in the room.

Three fundamental mathematical theories were considered, namely: statistical theory, geometric (ray) theory, and wave theory, which are used to determine acoustic characteristics, their scope of application, and limitations.

The features of conducting acoustic expertise for different types of rooms were analyzed, and recommendations were formed on adapting the sequence and content of the stages of conducting the expertise depending on the purpose of the room and the characteristics of the sound content. Thus, the goal of the work has been achieved, since the conduct of acoustic expertise for different types of premises has been scientifically and technically substantiated and systematized, and recommendations have been formulated for adapting its stages depending on the purpose of the premises and the characteristics of the sound content, which creates the basis for further formalization and automation of the process.

Prospects for further research include formalization and further automation of the process of forming the sequence and content of stages depending on the type of room. The implementation of a decision support system for acoustic expertise will allow standardizing the approach to the selection of stages, reducing the influence of the human factor, and increasing the effectiveness of expertise, while maintaining the possibility of expert correction in non-standard cases.

Conflict of interest

The authors declare that they have no conflict of interest with respect to this research, including financial, personal, authorship, or other conflicts that could influence the research and its results presented in this article.

Funding

The study was conducted without financial support.

References

1. SE "UkrNDNTs" (2022), *DSTU EN ISO 3382-1:2022. Akustyka. Vymiryuvannia akustychnykh parametriv prymishchennia (EN ISO 3382-1:2009, IDT; ISO 3382-1:2009, IDT)*. Kyiv: SE "UkrNDNTs". Available at: <https://uas.gov.ua/natsionalnyi-fond-nd/kataloh-natsionalnykh-standativ-ta-k> (last accessed: 24.01.2026).
2. European Committee for Standardization (CEN) (2009), *EN ISO 3382-1:2009. Acoustics – Measurement of room acoustic parameters – Part 1: Performance spaces (ISO 3382-1:2009)*. Brussels: CEN, 26 p. Available at: <https://www.iso.org/standard/40979.html> (last accessed: 26.01.2026).
3. CEN (2008) *EN ISO 3382-2:2008. Acoustics – Measurement of room acoustic parameters – Part 2: Reverberation time in ordinary rooms (ISO 3382-2:2008)*. Brussels: CEN, 26 p. Available at: <https://www.iso.org/standard/36201.html> (last accessed: 24.01.2026).
4. Nilsson, E., Engström, B. and Fritze, D. (2021), "An Energy Model for the Calculation of Room Acoustic Parameters in Rectangular Rooms with Absorbent Ceilings", *Applied Sciences*, 11(14), 6607. DOI: <https://doi.org/10.3390/app11146607>
5. Keränen, J., Hongisto, V. and Hakala, J. (2023), "Prediction accuracies of ray-tracing and regression models of room acoustics in open-plan offices", *Building and Environment*, 239, 110406. DOI: <https://doi.org/10.1016/j.buildenv.2023.110406>
6. Speaks, C.E. (2024), *Introduction to sound: acoustics for the hearing and speech sciences*. 5th edn. San Diego, CA: Plural Publishing, Inc. ISBN 978-1-63550-759-1. Available at: <https://www.pluralpublishing.com/publications/introduction-to-sound-acoustics-for-the-hearing-and-speech-sciences-1> (last accessed: 20.01.2026).
7. Shtrepi, L. and Astolfi, A. (2024), "Ten questions concerning Architectural Acoustics", *Building and Environment*, 262, 111921. DOI: <https://doi.org/10.1016/j.buildenv.2024.111921>
8. Hölter, L. (2024), "Finite difference time domain discretization for room acoustic simulations", *Acta Acustica*, 8, 75. DOI: <https://doi.org/10.1051/aacus/2024071>
9. CENELEC (2020), *EN IEC 60268-16:2020. Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index*. Brussels: CENELEC, 219 p. Available at: <https://webstore.iec.ch/en/publication/26771> (last accessed: 25.01.2026).
10. Kuttruff, H. (2016), *Room acoustics*. 6th edn. Boca Raton: CRC Press, 322 p. DOI: <https://doi.org/10.1201/9781315372150>
11. Wittebol, L. et al. (2024), "A hybrid room acoustic modelling approach for room impulse responses", *Applied Acoustics*, 223, 110068. DOI: <https://doi.org/10.1016/j.apacoust.2024.110068>
12. Poroshin, S., Usik, V. and Butova, O. (2019), "Features of the application of the methodology for assessing the acoustic properties of the sound field for spectator rooms", in *XXIX International Scientific Symposium "Metrology and Metrology Assurance 2019 (MMA 2019)"*, Sozopol, Bulgaria, 6–10 September 2019, pp. 145–149. DOI: <https://doi.org/10.1109/MMA.2019.8935952>
13. Poroshyn, S.M. and Usik, V.V. (2021), "Osoblyvosti vykorystannia chetvertoho etapu akustychnoi ekspertyzy dlia prymishchen spetsialnoho pryznachennia", *Suchasnyi stan naukovykh doslidzhen ta tekhnologii v promyslovosti*, 3(17), pp. 41–61. DOI: <https://doi.org/10.30837/ITSSI.2021.17.041>
14. Cairolì, M. (2021), "The architectural acoustic design for a multipurpose auditorium: Le Serre hall in the Villa Erba Convention Center", *Applied Acoustics*, 173, 107695. DOI: <https://doi.org/10.1016/j.apacoust.2020.107695>
15. Ahnert, W., Schmidt, W., (2005), *Fundamentals to perform acoustical measurements*, P. 1–49.
16. Foteinou, A., Murphy, D. and Cooper, J.P.D. (2023), "An acoustic reconstruction of the House of Commons, c. 1820–1834", *Acoustics*, 5(1), pp. 193–215. DOI: <https://doi.org/10.3390/acoustics5010012>
17. Poroshyn, S.M., Usik, V.V. and Belikov, I.S. (2021), "Metody formuvannia optymalnykh akustychnykh umov peredachi movnoho kontentu v prymishchenniakh malykh obsiahiv", *Suchasni informatsiini systemy*, 5(4), pp. 34–41. DOI: <https://doi.org/10.20998/2522-9052.2021.4.05>

Data availability

The manuscript has no associated data repository.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence technology in the creation of this work.

Received (Надійшла) 03.02.2026

Accepted for publication (Прийнята до друку) 18.02.2026

Publication date (Дата публікації) 12.03.2026

Відомості про авторів / About the Authors

Усик Вікторія Валеріївна – кандидат технічних наук, доцент, Національний технічний університет "Харківський політехнічний інститут", професор кафедри мультимедійних та інтернет-технологій і систем; Харків, Україна;

Victoriya Usik – PhD (Engineering Sciences), Associate Professor, National Technical University "Kharkiv Polytechnic Institute", Professor at the Department of Multimedia and Internet Technologies and Systems; Kharkiv, Ukraine;

e-mail: usik.viktory@gmail.com

ORCID ID: <http://orcid.org/0000-0002-3515-4849>

Федорчак Богдан Петрович – Національний технічний університет "Харківський політехнічний інститут", аспірант кафедри мультимедійних та інтернет-технологій і систем; Харків, Україна;

Bohdan Fedorchak – National Technical University "Kharkiv Polytechnic Institute", Postgraduate Student at the Department of Multimedia and Internet Technologies and Systems; Kharkiv, Ukraine;

e-mail: fedorchakbogdan@gmail.com

ORCID ID: <https://orcid.org/0009-0003-7335-9276>

АНАЛІЗ ЕТАПІВ АКУСТИЧНОЇ ЕКСПЕРТИЗИ ЗАМКНЕНИХ ПРИМІЩЕНЬ

Актуальність. У роботі розглянуто акустичну експертизу замкнених приміщень різного призначення з огляду на особливості звукового контенту, сформовано технічні та проектні рекомендації для отримання оптимальних властивостей звукового поля. Продемонстровано, що для різних типів приміщень (концертні зали, конференц-зали, прес-центри, офіси, переговорні кімнати тощо) існують специфічні вимоги щодо необхідних властивостей звукового поля, послідовностей етапів експертизи та особливостей проведення кожного етапу. Тому актуальним завданням є розроблення рекомендацій щодо акустичної експертизи залежно від типу приміщення з метою формалізації та подальшої автоматизації цього процесу. **Предметом дослідження** є методи визначення акустичних властивостей замкнених приміщень та їх оптимізація. **Мета роботи** полягає в розробленні рекомендацій щодо створення оптимальних акустичних властивостей з огляду на призначення приміщення та специфіку звукового контенту за допомогою аналізу особливостей акустичної експертизи для різних типів приміщень і на адаптацію етапів експертизи для кожного типу приміщень. **Завдання дослідження:** проаналізувати етапи акустичної експертизи та завдання, які розв'язуються на кожному етапі; розглянути математичні методи визначення акустичних властивостей, сфери їх застосування й обмеження; визначити особливості акустичної експертизи для різних типів приміщень й адаптувати етап експертизи залежно від призначення приміщення й особливостей звукового контенту. **Методи дослідження.** У роботі застосовано методи статистичної теорії для розрахунку часу реверберації. Геометрична теорія розглядає звук як сукупність променів і застосовується для аналізу структури відбиттів від поверхонь і виявлення акустичних дефектів у складних за формою приміщеннях. Хвильову теорію впроваджено для аналізу власних мод (резонансів) приміщення. **Досягнуті результати.** Розглянуто етапи акустичної експертизи, властивості звукового поля, які досліджено на кожному етапі, математичні методи визначення акустичних властивостей і рекомендації щодо їх оптимізації. Наголошено на особливостях кожного етапу з огляду на тип приміщення. **Висновки.** Аналіз акустичної експертизи для різних типів приміщень дасть змогу формалізувати цей процес, зменшити суб'єктивний складник і підвищити достовірність результатів. Досягнуті результати свідчать про перспективність подальших досліджень у напрямі формалізації та побудови комп'ютерної системи підтримки прийняття рішень для проведення акустичної експертизи.

Ключові слова: акустична експертиза; етапи; звуковий контент; акустичні властивості; оптимізація; технічні та проектні рекомендації.

Бібліографічні описи / Bibliographic descriptions

Усик В. В., Федорчак Б. П. Аналіз етапів акустичної експертизи замкнених приміщень. *Автоматизовані системи управління та прилади автоматики*. 2026. № 1 (188). С. 121–128. DOI: <https://doi.org/10.30837/0135-1710.2026.188.121>

Usik, V., Fedorchak, B. (2026), "Analysis of the stages of conducting an acoustic examination of closed rooms", *Management Information System and Devices*, No. 1 (188), P. 121–128. DOI: <https://doi.org/10.30837/0135-1710.2026.188.121>

Прокопенко О. С., Смеляков С. В.

РОЗРОБЛЕННЯ ДЕСКРИПТОРА ЗОБРАЖЕНЬ ДЛЯ ЕФЕКТИВНОГО ПОШУКУ НА ОСНОВІ ВМІСТУ У ВЕЛИКИХ СХОВИЩАХ ДАНИХ

Об'єктом дослідження є пошук зображень на основі вмісту. **Предметом** є моделі та методи пошуку зображень на основі вмісту (CBIR). **Мета дослідження** – розроблення дескриптора зображень, що використовує передові моделі комп'ютерного зору для виявлення об'єктів. Запропонований дескриптор підвищує ефективність і точність процесів пошуку зображень і управління ними. **Завдання передбачає:** аналіз сучасних підходів і рішень для створення дескрипторів зображень та їх використання в пошуку зображень на основі вмісту; виявлення обмежень наявних дескрипторів зображень і визначення вимог до дескриптора; розроблення нового дескриптора зображень, що ефективно застосовує інформацію про виявлені об'єкти; проведення експериментів для оцінювання запропонованого дескриптора й порівняння його ефективності з наявними рішеннями. **Методологія дослідження:** усебічний огляд передових методів створення дескрипторів зображень; аналіз використання наявних дескрипторів у системах CBIR; розроблення дескриптора, що застосовує сучасні моделі виявлення об'єктів; валідація ефективності запропонованого дескриптора порівняно з наявними дескрипторами на задачах пошуку. **Досягнуто таких результатів:** розроблено дескриптор зображення, оснований на об'єктах, виявлених за допомогою сучасних моделей машинного навчання; проведено серію експериментів для оцінювання ефективності та якості пошуку в великих системах зберігання зображень за допомогою запропонованого дескриптора. Експерименти виявили певні переваги, зокрема: швидше створення й порівняння дескрипторів, якщо порівнювати з гешованими, створеними вручну та дескрипторами, що ґрунтуються на глибокому навчанні; ефективне фільтрування зображень у сховищі на основі вмісту об'єктів, але ефективність дескриптора залежить від якості моделі й даних, оскільки зображення без виявлених об'єктів не з'являються в результатах пошуку, що може обмежувати його повноту. **Висновки:** розроблений дескриптор зображень є ефективним інструментом для розв'язання низки завдань пошуку зображень на основі вмісту завдяки швидкості побудови й пошуку. Перспективним напрямом дослідження є побудова системи пошуку зображень на основі вмісту з використанням розробленого дескриптора, посилене використання паралельних і розподілених обчислень.

Ключові слова: машинне навчання; дескриптори зображень; оброблення зображень; оптимізація пошуку зображень.

Вступ

Люди отримують більшість інформації зі свого оточення за допомогою органів зору. Тому з появою комп'ютерних технологій значні ресурси були спрямовані на витягування корисної інформації із зображень, її формалізацію та використання для аналізу ситуацій і прийняття рішень у системах комп'ютерного зору, моніторингу, підтримки рішень і штучного інтелекту [1].

Спочатку через обмеження ранніх обчислювальних ресурсів зображення могли мати лише низьку якість і займати лише кількості байтів або кілобайт. Однак нині зображення можуть сягати 32 гігабайти за секунду, якщо йдеться про відео з надвисокою роздільною здатністю, що використовується в спеціалізованих кінопродукціях. Навіть якщо не брати до уваги такі високоякісні знімки, повсякденні зображення зазвичай займають від кількох до десятків мегабайтів. Це дає змогу наблизитися до реалістичного візуального

відтворення, але водночас створює суттєві труднощі для розроблення ефективних алгоритмів, здатних обробляти такі великі масиви даних у реальному часі. Окреслена проблема ще більше ускладнюється стрімким зростанням обсягів і кількості сучасних систем зберігання інформації. Крім того, зображення часто зазнають дублювання й перетворень протягом їх життєвого циклу незалежно від того, чи це покращення, виконане користувачем, чи автоматичні сервіси оброблення, що стискають зображення за допомогою форматів із втратами (наприклад, JPEG).

Сучасні тенденції в пошуку зображень на основі вмісту (CBIR) [2] демонструють необхідність розроблення передових моделей і алгоритмів для витягування, порівняння та пошуку інформації, що ґрунтується на вмісті зображень [3–5]. Щоб спеціалізовані сервіси могли шукати, аналізувати зображення та приймати рішення в реальному часі, зазвичай створюється база даних дескрипторів. У цьому підході дескриптори генеруються й зберігаються в окремому сховищі.

Коли зображення надходить у систему пошуку, створюється його дескриптор. Безпосередній пошук і порівняння зображень у сховищі здійснюються за допомогою цих дескрипторів, що суттєво скорочує час пошуку й економить системні ресурси. Використання дескрипторів є значно ефективнішим рішенням, ніж виконання операцій безпосередньо над повнорозмірними файлами зображень. Однак вимоги до дескрипторів постійно підвищуються, оскільки від них залежать якість результатів пошуку й адекватність подання зображення. Це зі свого боку впливає на складність операцій пошуку [6].

Визначення проблеми

Наявні дескриптори мають знаходити компроміс між швидкістю та якістю і надавати пріоритет високій швидкості оброблення, спираючись на примітивні кольорові властивості, ігноруючи об'єкти на зображенні; або використовують високорозмірні векторні ознаки, отримані за допомогою глибокого навчання, що надають глибоке розуміння контексту й високу якість пошукових результатів, але мають відносно низьку швидкість обчислень, що робить таке рішення неефективним для великих сховищ зображень та інтенсивно завантажених потоків даних.

Аналіз останніх досліджень і публікацій

Будь-які параметри, статистичні показники або перетворення зображень, які дають змогу ефективно ідентифікувати й розрізнити зображення в сховищі, можуть використовуватися як їх дескриптори. Однак одним із найбільш поширених типів дескрипторів для пошуку подібних зображень є геші, що генеруються за допомогою алгоритмів перцептивного гешування. Ці алгоритми є функціями, що створюють порівняні геші [7, 8]. Прикладами є *Average Hash*, *DCT-Based Hash*, *Radial Variance-Based Hash* та *Marr-Hildreth Operator-Based Hash*. Узагальнений робочий процес для цих алгоритмів передбачає [9–11]:

- застосування специфічних трансформацій до зображення з метою усунення шуму;
- нормалізація зображення до фіксованого розміру;
- математичні трансформації, визначені алгоритмом;
- обчислення гешу.

Для порівняння дескрипторів гешів використовуються метрики, як-от відстань Хеммінга, нормалізована відстань Хеммінга та функція пік-кроскореляції. Деякі алгоритми гешування є стійкими до певних трансформацій зображень, наприклад розмиття, масштабування, стиснення, затемнення й незначні обертання [12, 13]. Основною перевагою геш-дескрипторів є швидкість їх порівняння, оскільки геші зазвичай займають до 128 байтів. Однак значним недоліком є те, що вони не беруть до уваги об'єкти в зображенні. Це може призвести до ситуацій, коли абсолютно різні реальні об'єкти з подібними кольорними схемами помилково ідентифікуються як подібні, і зіткнення не є рідкісним явищем [10, 14].

Іншим широко використовуваним типом дескриптора зображень є алгоритми, що базуються на ключових точках [15]: *SIFT (Scale-Invariant Feature Transform)*, *SURF (Speeded-Up Robust Features)*, *GLOH (Gradient Location and Orientation Histogram)*, *ORB (Oriented FAST and Rotated BRIEF)*, *BRISK* та *AKAZE (Accelerated KAZE)*. Ці алгоритми зазвичай передбачають двоетапний процес:

- виявлення ключових точок – ділянок, що суттєво відрізняються від навколишнього середовища, як-от контури чи кути об'єктів [16, 17];
- генерація дескрипторів для цих точок, що дає змогу порівнювати точки на різних зображеннях для визначення, чи належать вони до одного й того самого об'єкта [18–20].

Колекція дескрипторів ключових точок утворює дескриптор зображення. Порівняння цих дескрипторів дає змогу ідентифікувати однакові об'єкти на різних зображеннях, навіть під різними кутами, масштабами, умовами освітлення та іншими трансформаціями [21, 22]. Порівнюючи основані на ключових точках дескриптори з геш-дескрипторами, можна наголосити на значних перевагах останніх. Адже вони дають змогу виявляти об'єкти, здійснювати 3D-реконструкцію сцени, зшивати зображення й виконувати інші складні завдання комп'ютерного зору [23, 24]. Однак побудова й використання дескрипторів ключових точок є обчислювально складним процесом. Це робить їх менш придатними для порівняння великих обсягів зображень, оскільки час порівняння значною мірою залежить від кількості основних точок, виявлених на кожному зображенні [25, 26].

Дескриптори зображень, побудовані на глибокому навчанні, демонструють високу результативність у завданнях пошуку зображень, розпізнавання об'єктів чи семантичної сегментації, оскільки дають змогу узагальнювати різноманітні дані та ефективно аналізувати складні візуальні структури. Проте їх застосування пов'язано зі значними витратами обчислювальних потужностей як на етапі навчання, так і під час практичного використання. Якість роботи цих моделей, наприклад VGG, ResNet і EfficientNet, критично залежить від ретельного підбору, обсягу й різноманіття навчальної інформації. Незважаючи на це, саме здатність таких дескрипторів відтворювати глибокі, багаторівневі ознаки зображень викликають значну потребу в них у сучасних технологіях комп'ютерного зору. Останні дослідження продемонстрували зростання інтересу до екстракції дрібних ознак і крос-модальних підходів. Ці методи зосереджуються на навчанні дрібних семантичних ознак, щоб зберегти високий рівень семантичної релевантності й одночасно усунути особливі для модальності властивості.

Це приводить до більш точної та адекватної репрезентації оригінального зображення [11, 27].

Пошук зображень за вмістом (CBIR) зазнав суттєвого розвитку протягом останніх десятиліть, перейшовши від традиційних методів ручного виділення ознак до сучасних підходів, основаних на глибокому навчанні. Сутність CBIR полягає в тому, щоб знаходити в базі даних зображення, візуально схожі на задане, аналізувати безпосередньо їх вміст, а не текстові метадані. Ранні системи CBIR ґрунтувалися на низькорівневих ознаках, як-от кольорові гістограми, текстури й краї. Хоча ці методи були обчислювально ефективними, вони часто виявляли нездатність адекватно відтворювати високорівневу семантичну інформацію, що призводило до низької точності в реальних застосуваннях. Наприклад, зображення зі схожим розподілом кольорів, але абсолютно різними об'єктами могли бути помилково ідентифіковані як схожі. Такі системи містили значну кількість недоліків, але ці базові техніки заклали початок для створення більш складних пошукових систем. На рис. 1 зображено загальне подання такої системи.

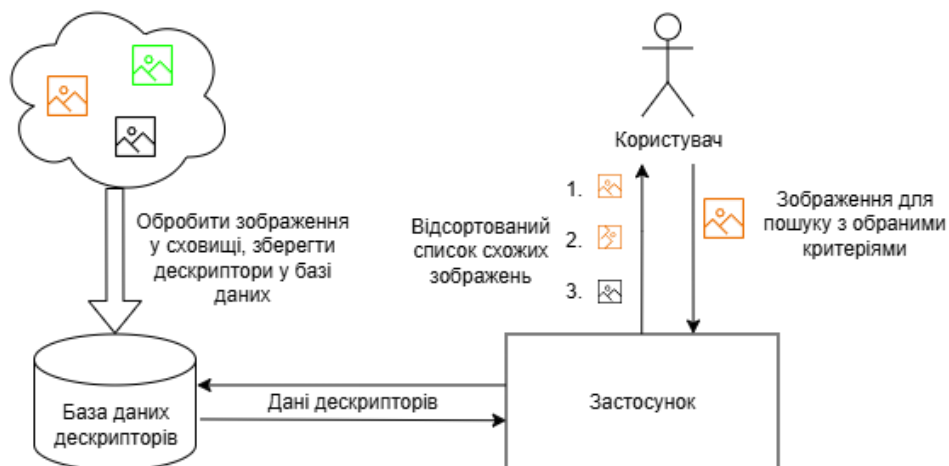


Рис. 1. Подання системи пошуку зображень за вмістом

Поява глибокого навчання сприяла революційним змінам у CBIR, надавши системам можливість самостійно навчатися складному, ієрархічному поданню ознак безпосередньо з даних. Сучасні системи CBIR використовують згорткові нейронні мережі (CNN) та інші архітектури глибокого навчання для отримання високовимірних векторів ознак або ембедингів, що інкорпують як низькорівневу, так і високорівневу візуальну інформацію. Попередньо навчені моделі, зокрема VGG, ResNet або EfficientNet, часто відіграють роль

екстракторів ознак, тоді як спеціалізовані архітектури, зокрема сіамські та триплетні мережі, застосовуються для навчання векторних ознак, оптимізованих саме для оцінювання схожості. Ці досягнення суттєво підвищили точність пошуку, даючи змогу системам CBIR ефективно обробляти складні запити, пов'язані з конкретними об'єктами, сценами чи навіть дрібними деталями. Наприклад, запит із зображенням певного виду птаха тепер може успішно знаходити зображення цього самого виду, зняті з багатьох ракурсів і за різних умов освітлення,

саме завдяки семантичній багатогранності дескрипторів, побудованих на основі глибокого навчання.

Наявні дескриптори для порівняння й пошуку зображень часто надають пріоритет маленькому розміру й високій швидкості оброблення, але здебільшого вони орієнтовані на колірні параметри, ігноруючи об'єкти на зображеннях. Хоча цей підхід добре працює в застосунках з унікальними зображеннями, він страждає від зіткнень, коли зображення мають подібні загальні властивості. Це призводить до недостатньо точних результатів порівняння й пошуку, оскільки дескриптори основані лише на загальних властивостях яскравості й кольору [28, 29].

Окремою значною тенденцією в сучасних CBIR є впровадження методів крос-модального пошуку, що дають змогу шукати співвідношення між різними типами даних, як-от пошук зображень за текстовим описом чи текстом за зображенням. Такий функціонал реалізується завдяки використанню моделей на кшталт CLIP (*Contrastive Language-Image Pretraining*), які навчаються подавати зображення й текст у єдиному спільному векторному просторі. Ці системи виявляються особливо цінними в таких сферах, як електронна комерція, де користувачі можуть шукати товари, описуючи їх словами, або в медичній діагностиці, де текстові висновки застосовуються для пошуку відповідних знімків. Крім того, у системах CBIR дедалі ширше використовуються механізми уваги (*attention*) й архітектури-трансформери, що допомагають моделі зосереджуватися на найбільш важливих ділянках зображення, що суттєво підвищує точність і релевантність результатів пошуку.

Попри ці вдосконалення, у сфері CBIR залишаються нерозв'язані завдання. Одним з основних є масштабованість систем пошуку, оскільки обчислювальна вартість порівняння високорозмірних векторів ознак значно зростає з розміром сховищ. Пошук прискорюють способом гешування й квантування векторів, але ці методи призводять до погіршення точності результуючої вибірки зображень, знайденої системою. Іншою проблемою є інтерпретованість систем CBIR на основі глибокого навчання, оскільки їх внутрішні механізми прийняття рішень часто залишаються "чорною скринькою". Крім цього, продуктивність таких систем критично залежить від даних, на яких вони навчаються. Недостатня репрезентативність або упередженість цих даних може спричинити систематичні помилки

й значно обмежувати здатність моделей коректно працювати з новими, незнайомими зображеннями.

Сучасні застосунки, де обсяги візуальної інформації (наприклад, у соціальних мережах, медичних базах даних чи системах спостереження) зростають експоненційно, наявні методи пошуку на основі вмісту часто є повільними або ресурсомісткими. Необхідний метод, що використовує обмежену кількість високорівневих даних і дає змогу швидко обробляти мільйони зображень, що критично для хмарних сервісів чи мобільних застосунків.

Визначення нерозв'язаних частин загальної проблеми. Мета й завдання роботи

Метою цього дослідження є розроблення дескриптора зображень, який використовує об'єкти на зображенні, що визначаються сучасними моделями виявлення об'єктів (наприклад, YOLO, Faster R-CNN), і застосовує типи об'єктів, їх розміри й місце розташування як ознаки зображення. Такий підхід забезпечує швидкість обчислень завдяки обмеженій кількості використовуваних даних, водночас надаючи глибоку контекстуальну інформацію для порівняння. Об'єкти, як складні сутності, несуть значну частину інформаційного вмісту зображення, що дає змогу забезпечити більш значущі й точні результати пошуку. Цей підхід пропонує збалансоване рішення для пошуку зображень на основі вмісту, значно покращуючи точність традиційних дескрипторів, і є набагато швидшим за високорозмірні вектори ознак, які використовуються в дескрипторах, що ґрунтуються на глибокому навчанні. Топові результати, обрані запропонованим методом, можна потім додатково аналізувати за допомогою обчислювально більш інтенсивних технік, забезпечуючи як ефективність, так і точність пошуку.

Ключова ідея розробленого дескриптора полягає в поєднанні низької обчислювальної вартості дескрипторів на основі низькорівневих ознак зображень із високорівневими даними об'єктів, що надаються згортковими нейронними мережами (CNN), які зазвичай потребують значних обчислювальних ресурсів. Цей підхід має на меті забезпечити баланс між ефективністю й точністю, використовуючи переваги обох методів для створення надійного й масштабованого рішення для завдань пошуку зображень.

На відміну від "чорних скриньок" глибокого навчання з високорозмірними векторами, цей дескриптор базується на зрозумілих властивостях об'єктів (тип, розмір, розташування), що полегшує пояснення, чому певні зображення вважаються подібними. Це важливо для галузей, де необхідна прозорість для прийняття рішень.

Поєднання низькорівневих ознак із високорівневими даними від нейронних мереж пропонує нові гібридні архітектури систем пошуку зображень на основі вмісту. Дослідження в цьому напрямі потрібні для того, щоб наблизитися до точки рівноваги, де ефективність і точність максимізуються, сприяючи створенню більш стійких до шумів чи варіацій зображень систем пошуку.

Важливою перевагою такого дескриптора є робота з об'єктами як абстрактними сутностями. Це дає змогу не прив'язуватися до конкретних моделей згорткових нейронних мереж для пошуку об'єктів, а використовувати найбільш сучасні та ефективні моделі. Залежно від розв'язуваного завдання й даних зображень, з якими необхідно працювати, можна обрати найбільш доцільний метод пошуку об'єктів щодо наявних обчислювальних ресурсів і часу, що не обов'язково має ґрунтуватися на машинному навчанні. Хоча в статті застосовується згорткова нейронна мережа для пошуку звичайних об'єктів на зображеннях, гнучкість дає змогу впроваджувати цей підхід для особливих зображень і наборів об'єктів (наприклад, медичних або військових) з використанням будь-яких методів для ефективного пошуку на основі вмісту у великих сховищах даних.

Розроблений дескриптор має відповідати таким критеріям:

- **ефективність** – застосування якомога менше даних про об'єкти для забезпечення обчислювальної та сховищної ефективності;
- **адаптивність** – сумісність із різними моделями виявлення об'єктів і водночас забезпечення гнучкості у використанні;
- **налаштуваність** – допущення налаштування під конкретні потреби різних наборів даних і випадків застосування;
- **баланс** – забезпечення оптимального балансу між швидкістю пошуку та якістю пошуку для забезпечення як швидкого, так і точного результату;
- **легкість побудови** – простота в створенні, що дає змогу ефективно будувати базу дескрипторів і безперешкодно інтегрувати її в системи пошуку.

Для досягнення окресленої мети необхідно виконати такі завдання:

- визначити поля даних для збережених об'єктів, щоб дескриптор був легким, але містив достатньо інформації для ефективного пошуку;
- розробити дизайн зберігання для дескриптора, який буде адаптивним до різних моделей виявлення об'єктів і налаштовуваним відповідно до конкретних вимог застосування;
- обрати відповідну модель виявлення об'єктів для експериментів, що дасть змогу провести справедливе порівняння з іншими дескрипторами;
- обрати набір даних для проведення серії експериментів, що визначать переваги й недоліки розробленого дескриптора;
- створити програмне забезпечення для реалізації побудови бази дескрипторів для зберігання зображень і забезпечення ефективного порівняння дескрипторів.

З розробленим дескриптором і реалізованим підтримувальним програмним забезпеченням можемо використовувати обраний набір даних і модель виявлення об'єктів та виконувати точний, але швидкий пошук у великих сховищах зображень.

Матеріали й методи

Обрання моделі для пошуку об'єктів

Для обрання моделі, що задовольнить потреби дескрипторів у поставленому завданні, було визначено певні критерії. Зокрема необхідно, щоб модель:

- широко й часто застосовувалася для пошуку об'єктів, що має гарантувати її ефективність за різних умов;
- була простою у використанні, здатною виявляти значну кількість об'єктів і уможлилювала просте перенавчання для адаптації до нових класів об'єктів або спеціалізації на певних типах об'єктів;
- мала кілька версій, які відрізняються кількістю параметрів, швидкістю і точністю, що уможлилює застосування одної й тої самої архітектури як у ресурсозабезпечених, так і в обмежених середовищах (наприклад, мобільні пристрої).

Важливо зауважити, що різні модифікації нейронної мережі можуть бути оптимізовані для зображень конкретних розмірів. Це має вирішальне значення, оскільки швидкість оброблення залежить від цього; нейронні мережі зазвичай добре

обробляють зменшення великих зображень, але вони не завжди можуть адекватно збільшувати маленькі зображення [30–32].

З огляду на перелічені вимоги до моделі для виявлення об'єктів, яку планується інтегрувати в систему керування масштабними сховищами зображень, обрано модель YOLOv8. На момент написання цього дослідження YOLOv8 є одним із найсучасніших рішень у сфері комп'ютерного зору для завдань детекції, що пропонує оптимальний баланс між швидкістю, точністю та адаптивністю. Ця модель є найновішою ітерацією успішної серії YOLO, яка інкорпорує низку нових функцій і вдосконалень для ще більшого підвищення продуктивності й гнучкості. YOLOv8 розроблена як швидкий, точний і зручний у використанні інструмент, що робить її ідеальним вибором для широкого спектра завдань, зокрема детекції та відстеження об'єктів, сегментації, класифікації зображень і оцінювання поз.

Основною перевагою моделей сімейства YOLO для розв'язання поставленого завдання є швидкість, яка відіграє вирішальну роль. Вона працює у 2–2,5 рази швидше за інші популярні моделі, такі як Faster R-CNN, SSD чи RetinaNet, демонструючи порівнянню, а іноді навіть вищу точність. Це досягається завдяки її одноетапній архітектурі детекції, на відміну від дворівневих підходів, властивих для сімейства мереж R-CNN.

Ще одним вагомим аргументом на користь YOLOv8 є її зручна підтримка апаратного прискорення. Модель дає змогу легко долучити графічний процесор (GPU) для пришвидшення детекції, часто за допомогою лише одного додаткового параметра в разі виклику, що може збільшити швидкість оброблення зображень у 5–10 разів.

Архітектурно модель трансформує вхідне зображення в кілька карт ознак, використовуючи пропускання з'єднання та інші сучасні техніки. Ці карти ознак потім об'єднуються та спрощуються до єдиного вихідного шару внаслідок комбінації операцій апсемплінгу та об'єднання. На цьому етапі прогноуються класи об'єктів та їх обмежувальні прямокутники, після чого для кожного об'єкта за допомогою алгоритму *Non-Maximum Suppression* (NMS) обирається найбільш вірогідний прямокутник. Кожен такий прямокутник описується шістьма значеннями: ідентифікатор класу, координати центру (X , Y), ширина, висота й рівень впевненості прогнозу.

Методологія пошуку

На першому етапі виявляються об'єкти на кожному зображенні в репозиторії за допомогою обраної моделі. Дескриптори створюються з використанням об'єктів, виявлених на зображенні, що після перетворення всієї необхідної інформації зберігається як база даних дескрипторів зображень.

Ця база даних стає основою для подальшого пошуку зображень. Вона ефективно порівнює класи, положення та інші ознаки об'єктів. Детальні описи нейронних мереж, дескрипторів і методів їх побудови надано нижче в цьому розділі.

Далі використовуємо просту метрику для оцінювання схожості між зображеннями (точніше, їх дескрипторами) й алгоритм пошуку, що застосовує ці дескриптори й метрики для визначення подібності зображень.

Загальний процес пошуку передбачає такі дії:

- 1) побудова дескриптора – його генерування для запитуваного зображення;
- 2) налаштування пошуку – визначення його критеріїв і налаштування параметрів (наприклад, пороги схожості, типи об'єктів);
- 3) безпосереднє виконання пошуку:
 - перебір дескрипторів у репозиторії;
 - порівняння кожного дескриптора з дескриптором запитуваного зображення за допомогою метрики схожості;
 - сортування зображень за балом подібності;
 - виведення результатів за схожістю.

За ідеальних умов відсортовані результати пошуку мають показати саме запитуване зображення (якщо воно є в репозиторії) як перший результат, далі трансформації або варіації запитуваного зображення (наприклад, змінене за розміром, обрізане або відредаговане зображення), а потім інші зображення з репозиторію, відсортовані за їх схожістю із запитуваним.

Для перевірки алгоритмів і програмного забезпечення розробляються допоміжні моделі, методи, критерії, метрики й програмні інструменти, зокрема метрики для оцінювання якості пошуку. Експериментальне налаштування, результати й детальний аналіз подано у відповідних розділах нижче.

Моделі дескриптора

Дескриптор зображення – це компактне подання, що інкапсулює основну інформацію про зображення в репозиторії. Він містить як технічні метадані (місце

зберігання, роздільну здатність, формат файлу), так і семантичні показники про виявлені на зображенні об'єкти. Останні охоплюють координати обмежувальних прямокутників, геометричні розміри (площу), а також оцінку впевненості моделі комп'ютерного зору в належності об'єкта до певного класу (наприклад, класу i). Такий мінімалістичний, але вичерпний набір атрибутів забезпечує основу для ефективного розроблення алгоритмів пошуку, фільтрації та порівняння зображень на базі їх візуального вмісту, зокрема за присутністю й властивостями об'єктів.

Кожен об'єкт подано вектором із семи чисел, як показано в табл. 1.

Зверніть увагу, що дескриптор об'єкта не зберігає номер класу, до якого належить об'єкт. Для зберігання номеру класу кожного об'єкта використовується структура дескриптора класу, що містить перелік знайдених об'єктів.

Таблиця 1. Модель дескриптора об'єкта

Назва	Опис
x	Відносна координата x -осі центра обмежувальної рамки об'єкта (у діапазоні 0–1)
y	Відносна координата y -осі центра обмежувальної рамки об'єкта (у діапазоні 0–1)
w	Відносна ширина об'єкта (у діапазоні 0–1)
h	Відносна висота об'єкта (у діапазоні 0–1)
$area$	Відносна площа об'єкта
$ratio$	Відношення меншої до більшої сторони обмежувальної рамки об'єкта
$conf$	Впевненість нейронної мережі в тому, що рамка містить об'єкт певного класу (у діапазоні 0–1)

Дескриптор класу описує сукупність об'єктів, що належать до певного класу на зображенні. Він містить набір дескрипторів для кожного виявленого об'єкта цього класу, а також додаткову інформацію, що характеризує групу об'єктів загалом.

Ця додаткова інформація містить кількість об'єктів певного класу, сумарну площу, яку вони займають, а також центр мас (центроїд) всієї групи на зображенні (див. табл. 2).

Таблиця 2. Модель дескриптора класу

Назва	Опис
$class$	Ідентифкатор класу
$number$	Кількість об'єктів цього класу на зображенні
$area$	Сума відносних площ зайнятими об'єктами цього класу на зображенні
$center$	Арифметичний центр групи об'єктів цього класу на зображенні
$objects$	Колекція дескрипторів об'єкта цього класу на зображенні

Хоча ці показники можна обчислити безпосередньо з дескрипторів окремих об'єктів, їх попереднє обчислення й зберігання дає змогу уникнути повторних розрахунків під час операцій порівняння. Це суттєво скорочує час виконання таких операцій і підвищує загальну ефективність системи.

Дескриптор зображення охоплює основну інформацію, необхідну для локалізації та опису оригінального зображення, а саме:

- місце розташування зображення, зокрема шлях до файлу (локальний або мережевий), URI або інший унікальний ідентифікатор, що дає змогу в разі необхідності отримати повнорозмірне зображення;
- габарити зображення, тобто абсолютну ширину й висоту в пікселях. Ці параметри є необхідними для перерахунку відносних координат і розмірів обмежувальних прямокутників об'єктів у абсолютні значення, оскільки всі дані про об'єкти зберігаються у нормованому, відносному вигляді;
- дескриптори класів – колекцію дескрипторів для кожного класу об'єктів, виявлених на зображенні.

Важливо зауважити, що структура дескриптора зображення є гнучкою та може бути розширена для долучення додаткової інформації про зображення або його об'єкти відповідно до конкретних вимог завдання. Незважаючи на можливість адаптації, основним критерієм залишається збереження компактності дескриптора, що є критично важливим для забезпечення високої швидкості оброблення під час роботи з масштабними сховищами зображень. Дані, що формують модель дескриптора зображення, узагальнені в табл. 3.

Таблиця 3. Модель дескриптора зображення

Назва	Опис
$path$	Шлях до файлу зображення
$width$	Ширина зображення
$height$	Висота зображення
$classes$	Дескриптори класів, знайдених на зображенні

Пошукові моделі, алгоритми й метрики

Дані, отримані від нейронної мережі та використані для формування дескрипторів, можуть застосовуватися в різних пошукових моделях.

Фільтрація за класами об'єктів

Дескриптори дають змогу відбирати зображення в репозиторії, визначаючи ті, що містять задану кількість об'єктів певних класів. Залежно від налаштувань, система може шукати зображення,

що містять точний набір об'єктів або ж хоча б зазначений набір. Результати такого пошуку можуть бути корисними самі по собі або слугувати попереднім фільтром для звуження простору пошуку. Це значно зменшує обсяг інформації для наступних, більш складних та ресурсомістких процедур, і підвищує загальну ефективність системи.

Пошук за зображеннями

Дескриптори також забезпечують можливість пошуку за зображеннями, допомагаючи користувачам знаходити конкретні зображення або визначати схожі на основі присутніх об'єктів та їх розташування. Ця модель пошуку може знаходити точні відповідності в репозиторії та ранжувати схожі зображення за їх візуальним вмістом. Вона вважається основною (за замовчуванням).

На цьому етапі пошук за зображенням реалізується за допомогою найпростіших порівняльних операцій: зіставляємо кількість об'єктів одного класу на двох різних зображеннях. Для кожного класу обчислюється різниця в кількості об'єктів, яка потім ділиться на більше з двох значень (між порівнюваними дескрипторами). Отримані значення для всіх класів, присутніх хоча б в одному з дескрипторів, підсумовуються, а сума ділиться на загальну кількість унікальних класів у обох дескрипторах. Такий підхід забезпечує комутативність операції порівняння (результат не залежить від порядку дескрипторів), дає змогу виявляти повні копії зображень у сховищі й гарантує, що зображення з ідентичним набором об'єктів повертатимуться як найбільш релевантні результати пошуку.

$$S = \sum_{s=1}^n w^* \frac{|O_1^i - O_2^i|}{\max(O_1^i, O_2^i)}, \quad (1)$$

де S – коефіцієнт подібності; n – кількість різних типів об'єктів на зображенні; w – ваговий коефіцієнт типу (розраховується як $1/n$); O_1^i – кількість об'єктів i -го типу на першому зображенні; O_2^i – кількість об'єктів i -го типу на другому зображенні.

Ось приклад алгоритму, який можна використати для порівняння двох дескрипторів, що дає змогу зіставити час виконання пошуку з іншими дескрипторами.

Опис програмного забезпечення

Для втілення запропонованих методів пошуку зображень на основі дескрипторів розроблено

застосунки з мінімалістичним вебінтерфейсом. Ці інструменти полегшують проведення експериментів і наочно відтворюють результати пошукових запитів. Для комплексного тестування запропонованої моделі пошуку, метрик, алгоритмів і програмного забезпечення створено чотири компоненти.

1. **Служба YOLO.** Реалізована на *Python*, абстрагує взаємодії з моделлю YOLO. Приймає шлях до зображення або набору зображень і генерує дескриптори, повертаючи їх безпосередньо або зберігаючи в базі даних. Використовує прискорення за допомогою GPU (якщо доступне) для збільшення швидкості обчислень.

2. **База даних.** Для експериментальної установки застосовано *MongoDB*, обраний за свою гнучкість як база даних NoSQL. Під час досліджень і розроблення моделі дескрипторів часто розширюються й модифікуються, і *MongoDB* дає змогу безперешкодно коригувати їх без постійної міграції даних. *MongoDB* також підтримує масштабованість завдяки таким технікам, як шардинг, що робить її придатною для роботи з великими репозиторіями інформації зображень. Створення гнучкого рішення для бази даних є критичним для передбачення й розв'язання проблем, що виникають під час роботи з масштабованими системами зберігання зображень [33–35].

3. **Сервіс на сервері.** Постійно оновлюється, щоб підтримувати актуальність моделей дескрипторів. Відповідає за виконання операцій пошуку й порівняння. Обробляє запити користувачів на детектування об'єктів на зображеннях і додавання фотоархівів до бази даних дескрипторів. Реалізовано в .NET, зберігає дескриптори в пам'яті з метою підвищення швидкості виконання та використовує *ImageMagick* для трансформацій зображень.

4. **UI-сервіс.** Легкий фронтенд, побудований на *Angular* для демонстраційних цілей. Дає змогу користувачам візуалізувати можливості системи, додавати фото до репозиторію, створювати дескриптори для зображень у визначених директоріях і здійснювати пошукові запити (наприклад, за тегами або схожістю до запитуваного зображення). Крім того, створено простий набір інструментів для генерування трансформованих версій вхідних зображень (наприклад, зміна масштабу, стиснення, гамма-корекція тощо). Ці трансформації використовуються для тестування стійкості й точності методу пошуку зображень у великих репозиторіях.

Результати досліджень та їх обговорення

Опис даних

Для експериментів використовуємо зображення з набору даних СОСО 2017. Цей набір містить понад 163 тис. зображень, анотованих для 80 категорій об'єктів. Такий великий обсяг даних дає змогу оцінити швидкодію пошуку в репозиторії, що налічує сотні тисяч елементів.

Повна назва набору – *Common Objects in Context* ("Загальні об'єкти в контексті") – відтворює властивості зображень, які переважно є повсякденними фотографіями: на них зображені люди, природа, міські пейзажі, спортивні події тощо. Серед розпізнаваних класів об'єктів – люди, транспортні засоби, тварини, пристрої, предмети побуту й аксесуари.

Окрім оригінальних зображень із СОСО 2017, додаємо до сховища модифіковані копії тестових зображень, які використовуватимуться як цільові запити. Список трансформацій, застосованих до оригіналів, передбачає: зниження яскравості на 20%, стиснення якості до 70%, підвищення контрастності, збільшення розміру зображення на 20%, зменшення розміру на 20%, перетворення в чорно-білу палітру, квантування кольорів до 128 відтінків, підвищення насиченості на 30% і використання ефекту різкості. Очікується, що в результатах пошуку для кожного тестового запиту серед перших позицій буде наявне як оригінальне зображення, так і всі його трансформовані версії.

Побудова бази даних дескрипторів

Репозиторій дескрипторів створено способом розміщення набору даних на робочій станції та його оброблення створеним програмним забезпеченням. Це програмне забезпечення приймає шляхи до каталогів зі зображеннями, послідовно обробляє всі зображення в цих каталогах, виявляє об'єкти на кожному зображенні, трансформує дані про об'єкти й зображення в описану модель дескриптора й зберігає отримані дескриптори в базі даних. Набір даних містить приблизно 164 тис. зображень, а процес генерації дескрипторів для всього масиву зайняв майже п'ять з половиною годин. Цей час передбачає виявлення об'єктів, побудову дескрипторів та їх індексацію в базі даних.

Підготовка експериментів

Експерименти зосереджені на швидкості побудови дескрипторів, швидкості фільтрації датасету й швидкості виконання порівняно з аналогами.

По-перше, порівняємо швидкість побудови дескрипторів із дескрипторами на основі глибокого навчання й перцептуальними гешами та побудуємо всі три дескриптори на зображеннях датасету.

По-друге, виконаємо тест пошуку за тегами. Цей тест фільтрує загальну вибірку, а саме лише екземпляри, що містять зазначений набір класів об'єктів. Проведемо пошукові запити, вказуючи по черзі від одного до п'яти класів об'єктів, які мають бути присутні в зображенні, виконаємо п'ять запитів із кожною конфігурацією та виміряємо час виконання.

Далі проведемо пошук для десяти тестових зображень, використовуючи розроблені дескриптори, перцептуальні геші та дескриптори на основі глибокого навчання, отримані за допомогою видобутку ознак із зображень датасету, а також згенеровані за допомогою сучасної моделі *Vision Transformer*, яку наприкінці 2020 р. запропонували О. Досовіцький, Л. Баєр, О. Колесніков та ін. [36, 37]. Виміряємо час виконання, а також визначимо кількість оригінальних і трансформованих зображень із цільової групи, які постійно з'являються в топових результатах вибору. Це дасть практичні результати для розроблених дескрипторів і моделей пошуку.

Побудова дескрипторів

Дескриптори побудовано для понад 163 тис. зображень у датасеті. Визначено середню швидкість побудови дескрипторів для кожного типу дескриптора. Варто зауважити, що дескриптори, зокрема на основі глибокого навчання, створені як на CPU, так і на GPU, тоді як перцептуальні геші обчислювалися лише на CPU. Результати подано в табл. 4.

Таблиця 4. Порівняння часу побудови дескрипторів

Тип дескриптора	Середня швидкість (зображень за секунду)
Перцептивний геш (CPU)	4.39
Об'єктний дескриптор (CPU)	2.56
Дескриптор ознак зображення (CPU)	2.14
Об'єктний дескриптор (GPU)	12.89
Дескриптор ознак зображення (GPU)	12.68

Результати експериментів демонструють, що побудова дескрипторів на основі об'єктів виявилася трохи швидшою, ніж генерування дескрипторів зображень, оснований на ознаках (наприклад, SIFT або інші класичні дескриптори), але, як і очікувалося, більш повільною за обчислення перцептивного гешу за однакових умов. Однак у процесі застосування апаратного прискорення на GPU час створення як об'єктних дескрипторів, так і дескрипторів на основі глибокого навчання скоротився у 5–6 разів у нашому тестовому середовищі. Ці результати свідчать про те, що час побудови об'єктних дескрипторів є порівняним, а іноді навіть меншим за час створення аналогічних дескрипторів.

Пошук за тегами

Цей експеримент передбачає запити на пошук присутніх на них зображень із кількістю класів, що коливається від одного до п'яти, із загальною кількістю понад 163 тис. зображень у репозиторії.

Як можна побачити з табл. 5, час пошуку не збільшується – навпаки, він зменшується із зростанням кількості класів на зображенні. Це здебільшого зумовлено порівнянням векторів, що забезпечує стабільне оброблення незалежно від кількості класів. Однак із збільшенням кількості класів кількість зображень, що задовольняють критерії пошуку, зменшується, що знижує витрати на виділення пам'яті за умови повернення результатів.

Середній час виконання запиту (50–100 мс для 164 тис. зображень) вказує на те, що фільтрація за складом об'єктів може обробляти понад мільйон

зображень за секунду, залишаючи вибірку з кількох тисяч зображень для більш детальних порівнянь.

Таблиця 5. Пошук зображень за тегами, час виконання

Кількість класів у запиті	Тест № 1	Тест № 2	Тест № 3	Тест № 4	Тест № 5	Середній час
1	121 мс	78 мс	79 мс	99 мс	106 мс	96.6 мс
2	92 мс	84 мс	80 мс	99 мс	89 мс	88.8 мс
3	71 мс	76 мс	57 мс	71 мс	80 мс	71 мс
4	67 мс	92 мс	64 мс	64 мс	61 мс	69.6 мс
5	54 мс	52 мс	50 мс	46 мс	59 мс	52.2 мс

Такі результати доводять, що розроблений об'єктно-орієнтований дескриптор може використовуватися у великих сховищах зображень як швидкий інструмент фільтрації на основі вмісту.

Пошук за зображенням

У межах цього експерименту здійснено пошук для 10 тестових зображень у великому репозиторії із застосуванням різних типів дескрипторів для порівняльного аналізу. Основним критерієм оцінювання був час виконання пошукових запитів, що дало змогу встановити, наскільки продуктивність розробленого дескриптора відповідає показникам відомих альтернатив. Отримані результати доводять ефективність запропонованого підходу, а розвиток і оптимізація алгоритмів пошуку залишатимуться пріоритетом для подальших наукових досліджень.

Кілька скриншотів результатів пошуку за зображенням подано на рис. 2–4. Усі скриншоти можна знайти в репозиторії <https://github.com/alex-prokopenko-nure/image-search-results>.

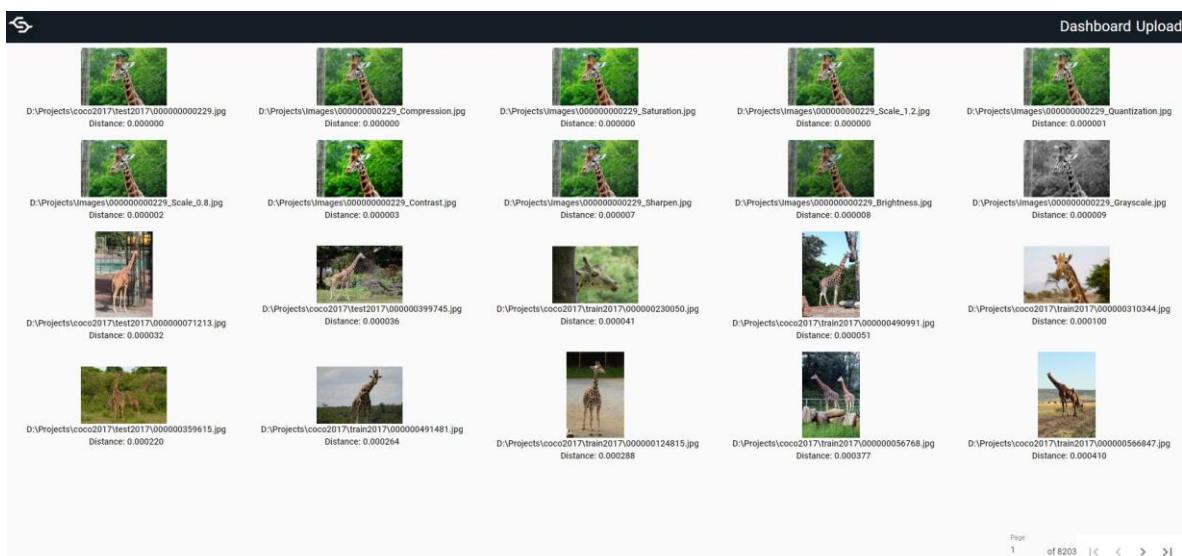


Рис. 2. Результат пошуку для тестового зображення № 3

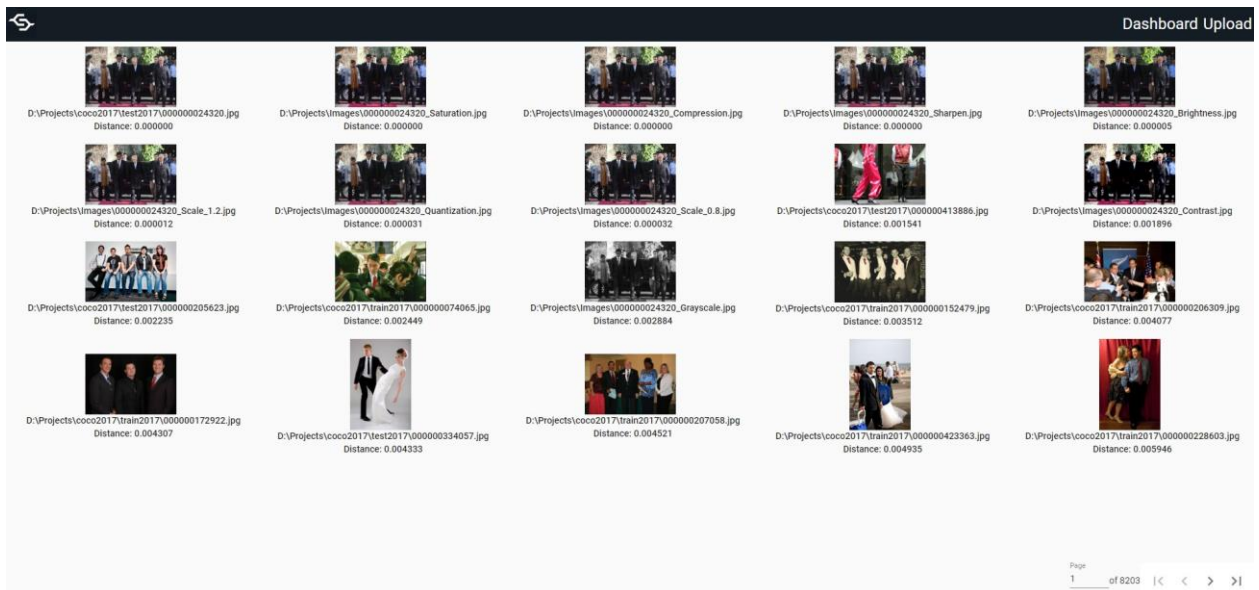


Рис. 3. Результат пошуку для тестового зображення № 6

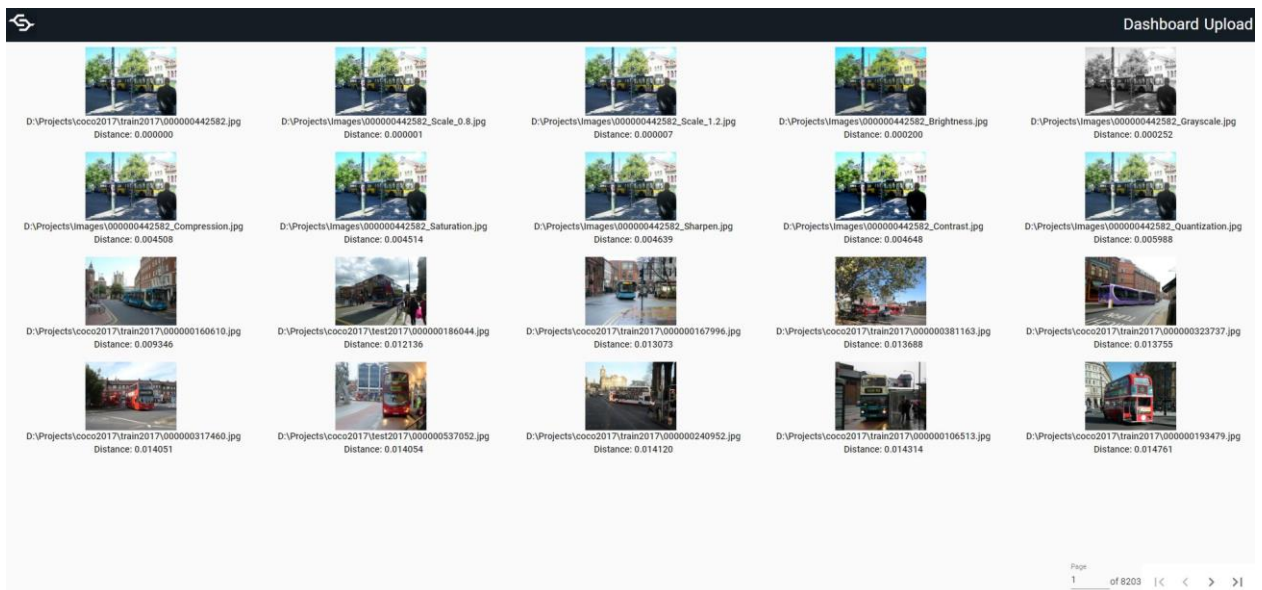


Рис. 4. Результат пошуку для тестового зображення № 10

Пошук зображень у цій роботі реалізовано за допомогою трьох різних методів: перцептивних гешів, дескрипторів на основі глибокого навчання й розроблених об'єктно-орієнтованих дескрипторів. Для перцептивних гешів порівняння відбувається способом обчислення суми квадратів різниць між бінарним поданням гешів. Дескриптори на основі глибокого навчання порівнюються за допомогою обчислення косинусної подібності між їх високовимірним векторним поданням (ембедингом). Об'єктно-орієнтовані дескриптори порівнюються за допомогою примітивного алгоритму, детально описаного в попередніх розділах, який бере

до уваги кількість і розподіл об'єктів різних класів на зображеннях.

У табл. 6 подано результати експерименту. Час пошуку наведено для перцептивних гешів (PH), дескрипторів ознак зображень (FD) та об'єктних дескрипторів (OD).

Деякі нетипові зображення можуть знижувати точність пошуку, зрештою опиняючись вище за трансформовані версії цільового зображення. Це відбувається, коли модель детекції об'єктів виявляє на них інший набір об'єктів, якщо порівнювати з оригіналом. Найбільш суттєве відхилення виникає, коли на трансформованому зображенні (наприклад,

затемненому або зі зміненими кольорами) не виявлено жодного об'єкта з певного класу, наявного на оригіналі. Іноді відстань до такого нетипового зображення зі схожим, але неідентичним набором об'єктів виявляється меншою, ніж до трансформованої версії з неповним або спотвореним набором об'єктів.

Таблиця 6. Час виконання пошуку за зображенням

	Час пошуку (PH)	Час пошуку (FD)	Час пошуку (OD)
1	214 мс	69.7 с	144 мс
2	221 мс	62.6 с	122 мс
3	230 мс	63.2 с	123 мс
4	242 мс	66.4 с	130 мс
5	215 мс	58.1 с	154 мс
6	226 мс	68 с	132 мс
7	223 мс	61.7 с	132 мс
8	252 мс	64.9 с	131 мс
9	235 мс	62.7 с	155 мс
10	228 мс	62.8 с	142 мс

Пошук на основі перцептивних гешів виконується за аналогічний проміжок часу. Однак, хоча 6–8 трансформацій з кожного тестового запиту потрапляють до верхніх позицій результатів, перетворення в чорно-білу палітру для всіх тестових зображень опиняється поза першою тисячею знайдених елементів. Крім того, для 8 з 10 тестових запитів ще 1–2 інші трансформації (пов'язані з маніпуляціями кольоровим діапазоном, як-от зміна насиченості або яскравості) також випадають за межі першої тисячі результатів.

Пошук за допомогою косинусної подібності між векторами ознак, отриманими за допомогою моделі *Vision Transformer* (ViT), значно довший за часом, якщо порівнювати з іншими методами. Аналогічно до експерименту з перцептивними гешами, перетворення зображень у чорно-білу палітру часто не потрапляє до найрелевантніших результатів, опиняючись за межами першої двадцятки в трьох випадках. Однак інші види трансформацій (як-от зміна яскравості, контрасту, розміру) переважно посідають місця в першій десятці. Якщо вилучити з оцінки перетворення в чорно-біле, то якість пошуку для частини зображень за цим методом вища, ніж у запропонованого об'єктно-орієнтованого підходу, але досягається це ціною значно більших витрат обчислювального часу.

Пошук за допомогою порівняння векторів ознак зображень з подальшим доопрацюванням або комбінацією з іншими методами може використовуватися для більш точного порівняння зображень у попередньо відфільтрованій вибірці із

сотень чи тисяч зображень, спочатку відфільтрованих за допомогою запропонованого в цій роботі методу.

Обговорення

Досягнуті результати підтверджують, що створений дескриптор успішно виконує поставлені завдання, зокрема ефективний пошук схожих зображень у великих сховищах даних. Він поєднує низькі обчислювальні витрати, властиві для дескрипторів на основі низькорівневих ознак, із високорівневою семантичною інформацією про об'єкти, яку зазвичай надають згорткові нейронні мережі (CNN), що вимагають значно більших обчислювальних ресурсів.

Дескриптори на основі об'єктів будуються швидше, ніж генерується векторне подання (ембединг) на базі глибокого навчання, але, як і передбачалося, поступається за швидкістю обчисленню перцептивних гешів. Така продуктивність у створенні дескрипторів робить їх практично застосовними і надає певну перевагу перед більш повільними методами на основі глибокого навчання.

Швидкість фільтрації, досягнута завдяки розробленому дескриптору, дає змогу формувати підмножини зображень, ґрунтуючись на їх об'єктному складі, з продуктивністю майже в один мільйон зображень за секунду. Ця можливість відкриває шлях до ефективної комбінації запропонованого дескриптора з більш складними методами пошуку, використовуючи попередню фільтрацію для різкого звуження ділянки пошуку і, як наслідок, отримання точних і релевантних результатів високої якості.

Швидкість порівняння зображень за допомогою об'єктних дескрипторів у сотні разів перевищує швидкість порівняння з використанням векторів ознак, отриманих за допомогою глибокого навчання, зберігаючи в цьому разі ключовий семантичний контекст зображення в пошуковій операції. Якість пошуку в цьому дослідженні не вимірювалася, але можна припустити, що вона може бути дещо нижчою за дескриптори на основі глибокого навчання. Саме це робить поданий дескриптор ідеальним інструментом для попереднього фільтрування та різкого зменшення обсягу вхідної інформації для більш складних і точних моделей пошуку на базі глибокого навчання, які зможуть провести фінальне уточнення порядку результатів для заданого запиту.

Подальші дослідження необхідно зосередити на створенні швидкого й точного алгоритму пошуку для розробленого дескриптора. Такий алгоритм

використовуватиме інформацію про розмір і розташування об'єктів для пошуку подібних зображень, а також трансформацій зображень, як-от зміни яскравості, кольору, масштабу, ступеня стиснення над оригіналом у сховищі.

Ще одним логічним удосконаленням є використання розподілених обчислювальних систем, оскільки поставлене завдання має високу автономність. Результати обчислень, виконаних для одного зображення, не потрібні для обчислень для іншого зображення. Тому навіть у великих репозиторіях, що містять мільярди зображень, швидкий пошук за зображенням відповідно до розробленого алгоритму можна виконати, оскільки складне завдання припустімо поділити на підзавдання й одночасно їх розв'язувати на різних частинах набору даних десятками або сотнями різних машин.

6. Висновки

й перспективи подальшого дослідження

У статті порушено проблему компромісу між швидкістю й точністю дескрипторів зображень, що застосовуються для пошуку за вмістом у великих сховищах даних, і способи підвищення ефективності таких систем за допомогою використання нових видів дескрипторів зображень.

Для розв'язання окресленої проблеми запропоновано об'єктно-орієнтований дескриптор зображень, який застосовує дані, отримані за допомогою моделей детекції об'єктів, з метою створення ефективного й збалансованого рішення для пошуку за вмістом (CBIR). Цей підхід значно покращує точність традиційних дескрипторів, залишаючись набагато швидшим за високовимірне векторне подання, що використовується в дескрипторах на основі глибокого навчання. Для розроблення дескриптора на базі об'єктів було виконано такі завдання: визначено поля даних для збережених об'єктів, створено дизайн зберігання для дескриптора, обрано відповідну модель виявлення об'єктів, обрано набір даних і розроблено програмне забезпечення для порівняння дескрипторів у сховищах зображень.

Результати експериментів, що порівнюють об'єктний дескриптор з методами на основі перцептивних гешів і глибокого навчання, демонструють переваги запропонованого підходу. Зокрема побудова об'єктних дескрипторів відбувається швидше, ніж створення векторного подання за допомогою глибокого

навчання. Крім того, запропонований дескриптор забезпечує найвищу швидкість пошуку у великих сховищах, перевершуючи навіть швидкі обчислення на основі перцептивних гешів. Такі порівняння в 100 разів швидші за порівняння зображень за допомогою дескрипторів на основі глибокого навчання, тоді як вони також зберігають контекстну інформацію про зображення. Як і очікувалося, якість пошуку об'єктного дескриптора є дещо нижчою, ніж у передових методів глибокого навчання, але значно вищою, ніж у методів на основі перцептивних гешів. Отже, запропонований дескриптор пропонує аналогічний час побудови й значно швидший час виконання пошуку порівняно з дескрипторами на основі глибокого навчання, які останнім часом вважаються передовими у сфері пошуку зображень за вмістом. Це робить його ідеальним рішенням для ефективною попередньої фільтрації та швидкого пошуку в масштабних архівах.

Крім того, визначено такі напрями для подальших досліджень:

- розроблення метрик і алгоритмів пошуку для створення системи CBIR з використанням розробленого дескриптора на основі об'єктів;
- поєднання розробленого дескриптора з більш складними алгоритмами пошуку для уточнення топових результатів пошуку;
- виконання завдання пошуку зображень з порівнянням кожного об'єкта найбільш схожих зображень для більш точного ранжування близьких результатів;
- покращення використання паралельних і розподілених обчислень у побудові та порівнянні дескрипторів.

Дескриптор на основі об'єктів балансує ефективність і точність, застосовуючи переваги наявних методів з метою створення надійного й масштабованого рішення для завдань пошуку зображень.

Конфлікт інтересів

Автори заявляють, що не мають конфлікту інтересів, зокрема фінансового, особистого, авторського чи будь-якого іншого характеру, який міг би вплинути на дослідження, а також на результати, опубліковані в цій статті.

Фінансування

Це дослідження було проведено без фінансової підтримки.

Доступність даних

Робота має супутні матеріали в репозиторії даних. Вихідний код програмного забезпечення буде доступний за обгрутованим запитом.

Використання засобів штучного інтелекту

Автори підтверджують, що не застосовували технології штучного інтелекту для написання цієї роботи.

References

- Gonzalez, R. C., Woods, R. E. (2018), "Digital Image Processing", 4th. ed. Pearson/Prentice Hall. 1168 p. DOI/ISBN: 9780133356724
- Liu, X. et al. (2018), "Prior-Based Quantization Bin Matching for Cloud Storage of JPEG Images", *IEEE Transactions on Image Processing*, Vol. 27, No. 7, P. 3222–3235. DOI: <https://doi.org/10.1109/TIP.2018.2799704>
- Zheng, Y. et al. (2018), "Size-Scalable Content-Based Histopathological Image Retrieval From Database That Consists of WSIs", *IEEE Journal of Biomedical and Health Informatics*, Vol. 22, No. 4, P. 1278–1287. DOI: <https://doi.org/10.1109/JBHI.2017.2723014>
- Padma, Y. (2024), "Advancements in Non-Linear Content-Based Image Retrieval (CBIR) Systems for Image Analysis", *Communications on Applied Nonlinear Analysis*, Vol. 31, No. 2s, P. 253–265. DOI: <https://doi.org/10.52783/cana.v31.639>
- Wang, X. et al. (2020), "A Storage Method for Remote Sensing Images Based on Google S2", *IEEE Access*. Vol. 8, P. 74943–74956. DOI: <https://doi.org/10.1109/ACCESS.2020.2988631>
- Yang, W. et al. (2020), "Design of Intelligent Search Engine Service Performance Evaluation System", *5th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS)*, P. 86–91. DOI: <https://doi.org/10.1109/ACIRS49895.2020.9162611>
- Žižakić, N., Pižurica, A. (2022), "Efficient Local Image Descriptors Learned With Autoencoders", *IEEE Access*, Vol. 10, P. 221–235. DOI: <https://doi.org/10.1109/ACCESS.2021.3138168>
- Amorós, F. et al. (2020), "Holistic Descriptors of Omnidirectional Color Images and Their Performance in Estimation of Position and Orientation", *IEEE Access*, Vol. 8, P. 81822–81848. DOI: <https://doi.org/10.1109/ACCESS.2020.2990996>
- Staszewski, P. et al. (2022), "A New Approach to Descriptors Generation for Image Retrieval by Analyzing Activations of Deep Neural Network Layers", *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 33, No. 12, P. 7913–7920. DOI: <https://doi.org/10.1109/TNNLS.2021.3084633>
- Suárez, I., Buenaposada, J. M., Baumela, L. (2021) "Revisiting Binary Local Image Description for Resource Limited Devices", *IEEE Robotics and Automation Letters*, Vol. 6, No. 4, P. 8317–8324. DOI: <https://doi.org/10.1109/LRA.2021.3107024>
- Wang, B. et al. (2023), "Multi-level adversarial attention cross-modal hashing", *Signal Processing: Image Communication*, Vol. 117, 117017, ISSN 0923-5965. DOI: <https://doi.org/10.1016/j.image.2023.117017>
- Reato, T., Demir, B., Bruzzone L. (2019), "An Unsupervised Multicode Hashing Method for Accurate and Scalable Remote Sensing Image Retrieval", *IEEE Geoscience and Remote Sensing Letters*, Vol. 16, No. 2, P. 276–280. DOI: <https://doi.org/10.1109/LGRS.2018.2870686>
- Zhu, Y., Li, Y., Wang, S. (2019), "Unsupervised Deep Hashing With Adaptive Feature Learning for Image Retrieval", *IEEE Signal Processing Letters*, Vol. 26, No. 3, P. 395–399. DOI: <https://doi.org/10.1109/LSP.2019.2892233>
- Bai, J. et al. (2019), "Deep Progressive Hashing for Image Retrieval", *IEEE Transactions on Multimedia*, Vol. 21, No. 12, P. 3178–3193. DOI: <https://doi.org/10.1109/TMM.2019.2920601>
- Zhou, Z. et al. (2020), "Integrating SIFT and CNN Feature Matching for Partial-Duplicate Image Detection", *IEEE Transactions on Emerging Topics in Computational Intelligence*, Vol. 4, No. 5, P. 593–604. DOI: <https://doi.org/10.1109/TETCI.2019.2909936>
- Akita, K., Ukita, N. (2023) "Context-Aware Region-Dependent Scale Proposals for Scale-Optimized Object Detection Using Super-Resolution", *IEEE Access*, Vol. 11, P. 122141–122153. DOI: <https://doi.org/10.1109/ACCESS.2023.3329302>
- Rao, Y. et al. (2023), "Learning Enriched Feature Descriptor for Image Matching and Visual Measurement", *IEEE Transactions on Instrumentation and Measurement*, Vol. 72, P. 1–12. DOI: <https://doi.org/10.1109/TIM.2023.3249237>
- Migliorati, A. et al. (2021), "Learnable Descriptors for Visual Search", *IEEE Transactions on Image Processing*, Vol. 30, P. 80–91. DOI: <https://doi.org/10.1109/TIP.2020.3031216>
- Zhang, Y. (2024), "Graph Convolution Based Efficient Re-Ranking for Visual Retrieval", *IEEE Transactions on Multimedia*, Vol. 26, P. 1089–1101. DOI: <https://doi.org/10.1109/TMM.2023.3276167>
- Iida, K., Kiya, H. (2020), "Privacy-Preserving Content-Based Image Retrieval Using Compressible Encrypted Images", *IEEE Access*, Vol. 8, P. 200038–200050. DOI: <https://doi.org/10.1109/ACCESS.2020.3035563>
- Brogan, J. et al. (2021), "Fast Local Spatial Verification for Feature-Agnostic Large-Scale Image Retrieval", *IEEE Transactions on Image Processing*, Vol. 30, P. 6892–6905. DOI: <https://doi.org/10.1109/TIP.2021.3097175>
- Imbriaco, R. et al. (2022), "Toward Multilabel Image Retrieval for Remote Sensing", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 60, P. 1–14. DOI: <https://doi.org/10.1109/TGRS.2021.3095957>

23. Zhao, Z. -Q. et al. (2019), "Object Detection With Deep Learning: A Review", *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 30, No. 11, P. 3212–3232. DOI: <https://doi.org/10.1109/TNNLS.2018.2876865>
24. Xu, Y., Zhao, X., Gong, J. (2019), "A Large-Scale Secure Image Retrieval Method in Cloud Environment", *IEEE Access*, Vol. 7, P. 160082–160090. DOI: <https://doi.org/10.1109/ACCESS.2019.2951175>
25. Cheng, M. -L., Matsuoka, M. (2020), "An Enhanced Image Matching Strategy Using Binary-Stream Feature Descriptors", *IEEE Geoscience and Remote Sensing Letters*, Vol. 17, No. 7, P. 1253–1257. DOI: <https://doi.org/10.1109/LGRS.2019.2943237>
26. Clark, C. N. et al. (2023), "Investigating the Resolution-Performance Trade-off of Object Detection Models in Support of the Sustainable Development Goals", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 16, P. 5695–5713. DOI: <https://doi.org/10.1109/JSTARS.2023.3284309>
27. Sun, C. et al. (2023), "Multi-label adversarial fine-grained cross-modal retrieval", *Signal Processing: Image Communication*, Vol. 117, 117018, ISSN 0923-5965, DOI: <https://doi.org/10.1016/j.image.2023.117018>
28. Radenović, F., Tolias, G., Chum, O. (2019), "Fine-Tuning CNN Image Retrieval with No Human Annotation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 41, No. 7, P. 1655–1668. DOI: <https://doi.org/10.1109/TPAMI.2018.2846566>
29. Vynokurova O., Peleshko D., Peleshko M. (2020), "Hybrid Deep Convolutional Neural Network with Multimodal Fusion", *Data Stream Mining & Processing. DSMP 2020. Communications in Computer and Information Science*, Vol. 1158, P. 62–78. DOI: https://doi.org/10.1007/978-3-030-61656-4_4
30. Galvez, R. L. et al. (2018), "Object Detection Using Convolutional Neural Networks", *TENCON 2018 – 2018 IEEE Region 10 Conference*, P. 2023–2027. DOI: <https://doi.org/10.1109/TENCON.2018.8650517>
31. Fang, F. et al. (2020), "Combining Faster R-CNN and Model-Driven Clustering for Elongated Object Detection", *IEEE Transactions on Image Processing*, Vol. 29, P. 2052–2065. DOI: <https://doi.org/10.1109/TIP.2019.2947792>
32. Sun, C.-Y. et al. (2021), "Cascade Faster R-CNN Detection for Vulnerable Plaques in OCT Images", *IEEE Access*, Vol. 9, P. 24697–24704. DOI: <https://doi.org/10.1109/ACCESS.2021.3056448>
33. Dhayne, H. et al. (2021), "In Search of Big Medical Data Integration Solutions – A Comprehensive Survey", *IEEE Access*, Vol. 7, P. 91265–91290. DOI: <https://doi.org/10.1109/ACCESS.2019.2927491>
34. Kathiravelu, P., Sharma, A., Sharma, P. (2021), "Understanding Scanner Utilization With Real-Time DICOM Metadata Extraction", *IEEE Access*, Vol. 9, P. 10621–10633. DOI: <https://doi.org/10.1109/ACCESS.2021.3050467>
35. Ebrahimian, A. et al. (2020), "Class-Aware Image Search for Interpretable Cancer Identification", *IEEE Access*, Vol. 8, P. 197352–197362. DOI: <https://doi.org/10.1109/ACCESS.2020.3033492>
36. Wolf, T. et al. (2020), "Transformers: State-of-the-Art Natural Language Processing", *Natural Language Processing: System Demonstrations*, P. 38–45. DOI: <https://doi.org/10.18653/v1/2020.emnlp-demos.6>
37. Dosovitskiy, A. et al. (2021), "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale", *International Conference on Learning Representations*, DOI: <https://doi.org/10.48550/arXiv.2010.11929>

Received (Надійшла) 22.01.2026

Accepted for publication (Прийнята до друку) 14.02.2026

Publication date (Дата публікації) 12.03.2026

Відомості про авторів / About the Authors

Прокопенко Олександр Сергійович – Харківський національний університет радіоелектроніки, аспірант кафедри програмної інженерії; Харків, Україна;

Oleksandr Prokopenko – Kharkiv National University of Radio Electronics, Postgraduate Student at the Department of Software Engineering; Kharkiv, Ukraine;

e-mail: oleksandr.prokopenko1@nure.ua

ORCID ID: <http://orcid.org/0000-0003-0489-6820>

Смеляков Сергій Вячеславович – доктор фізико-математичних наук, професор, Харківський національний університет радіоелектроніки, професор кафедри програмної інженерії; Харківський національний університет Повітряних Сил імені Івана Кожедуба, професор кафедри математичного та програмного забезпечення АСУ; Харків, Україна;

Sergii Smelyakov – Doctor of Physical and Mathematical Sciences, Professor, Kharkiv National University of Radio Electronics, Professor at the Software Engineering Department; Ivan Kozhedub Kharkiv National Air Force University, Professor at the Department of Mathematics and Software for Automated Control Systems; Kharkiv, Ukraine;

e-mail: serhii.smeliakov@nure.ua

ORCID ID: <http://orcid.org/0000-0002-5791-2479>

DEVELOPMENT OF AN IMAGE DESCRIPTOR FOR EFFECTIVE CONTENT-BASED RETRIEVAL IN LARGE DATA REPOSITORIES

The object of the research is content-based image retrieval. *The subject of this research* is models and methods of content-based image retrieval (CBIR). *The purpose of the research* is the development of an image descriptor that uses advanced computer vision models for object detection. The proposed descriptor increase the efficiency and accuracy of image search and management processes. *The objectives of the study* include: analysis of modern approaches and solutions for creating image descriptors and their use in content-based image retrieval; identification of the limitations of existing image descriptors and determination of requirements for a new, improved descriptor; development of a new image descriptor that effectively uses information about detected objects; conducting experiments to evaluate the proposed descriptor and compare its effectiveness with existing solutions. *The methodology* includes conducting a comprehensive review of state-of-the-art methods for creating image descriptors; analysis of the use of existing descriptors in CBIR systems; development of a descriptor that utilizes modern object detection models; validation of the effectiveness of the proposed descriptor in comparison with existing descriptors on retrieval tasks. The following *results* were obtained: an image descriptor based on objects detected using modern machine learning models was developed; a series of experiments was conducted to evaluate the efficiency and quality of retrieval in large image storage systems using the proposed descriptor. The experiments revealed the following advantages: faster creation and comparison of descriptors than hashed, handcrafted, and deep learning-based descriptors; effective filtering of images in the repository based on object content. However, the effectiveness of the descriptor depends on the quality of the model and the data, since images without detected objects do not appear in the search results, which may limit recall. *Conclusions*: the developed image descriptor is an effective tool for solving a range of content-based image retrieval tasks due to the speed of construction and search. A promising direction for further research is the development of a content-based image retrieval system using the proposed descriptor, with enhanced use of parallel and distributed computing.

Keywords: machine learning; image descriptors; image processing; image search optimization.

Бібліографічні описи / Bibliographic descriptions

Прокопенко О. С., Смеляков С. В. Розроблення дескриптора зображень для ефективного пошуку на основі вмісту у великих сховищах даних. *Автоматизовані системи управління та прилади автоматики*. 2026. № 1 (188). С. 129–144. DOI: <https://doi.org/10.30837/0135-1710.2026.188.129>

Prokopenko, O., Smelyakov, S. (2026), "Development of an image descriptor for effective content-based retrieval in large data repositories", *Management Information System and Devices*, No. 1 (188), P. 129–144. DOI: <https://doi.org/10.30837/0135-1710.2026.188.129>

Репіхов В. М., Чуприна А. С.

МЕТОД ПРЕВЕНТИВНОГО СУПРОВОДЖЕННЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ НА ОСНОВІ КОМПАРАТОРНОЇ МОДЕЛІ

Предметом дослідження є метод превентивного супроводження програмного забезпечення на основі компараторної моделі, який забезпечує формалізоване оцінювання експлуатаційного стану й підтримку прийняття рішень щодо експлуатаційної реакції. **Мета роботи** – підвищення ефективності превентивного супроводження програмного забезпечення способом формалізованого оцінювання його експлуатаційного стану, що забезпечує раннє виявлення деградаційних процесів, класифікацію експлуатаційної реакції та автоматизоване формування інцидентів супроводження за умов обмеженої історії інцидентів. **Основні завдання:** обґрунтувати вимоги до формалізованого оцінювання експлуатаційного стану програмного забезпечення в умовах превентивного супроводження та обмеженої доступності історичних даних; побудувати компараторну модель ідентифікації експлуатаційних станів способом формалізації взаємозв'язків між ознаками у вигляді системи логічних рівнянь і перевірки її функціональної визначеності; розробити метод компараторної ідентифікації експлуатаційних станів на основі побудованої моделі; виконати експериментальну валідацію запропонованого методу на експлуатаційних даних і порівняти з традиційними пороговими механізмами сповіщення. **Методи дослідження.** Для ідентифікації експлуатаційного стану застосовано метод компараторної ідентифікації в межах компараторної моделі з використанням предикатного подання стану. Взаємозв'язки між ознаками формалізовано системою логічних рівнянь із перевіркою функціональної визначеності. Формування ознак ґрунтується на поєднанні метрик тестування, агрегованих експлуатаційних показників, трендових характеристик і ознак аномальної поведінки. Емпіричну перевірку виконано на симуляційних експлуатаційних даних із використанням ковзних вікон агрегації та порівнянням із пороговими механізмами сповіщення. **Результати.** Продемонстровано, що запропонований метод забезпечує більш раннє виявлення деградаційних процесів порівняно з класичними пороговими алертами. Використання трендових і комбінованих ознак підвищує стабільність класифікації станів і зменшує кількість хибних спрацьовувань у дослідженому сценарії. На основі ідентифікованого класу стану реалізовано модель превентивного супроводження, що забезпечує автоматизоване формування інцидентів із зазначенням типу проблеми, локалізації та рівня терміновості. **Висновки.** Досягнуті результати розвивають формалізовані підходи до оцінювання експлуатаційного стану програмного забезпечення й демонструють можливість застосування компараторної моделі для превентивного супроводження за відсутності повної та репрезентативної історії інцидентів. Практична значущість роботи полягає в можливості впровадження інтерпретованих механізмів раннього виявлення деградації та автоматизації супроводження, зокрема формування інцидентів і підтримки вибору експлуатаційної реакції, без необхідності накопичення значних обсягів історичних даних.

Ключові слова: програмне забезпечення; превентивне супроводження; компараторна модель; компараторна ідентифікація; експлуатаційний стан; деградація програмного забезпечення; автоматизоване формування інцидентів.

Вступ

Забезпечення стабільної експлуатації програмних продуктів упродовж усього життєвого циклу є одним із ключових завдань сучасної інженерії програмного забезпечення [1]. В умовах тривалої експлуатації програмні системи зазнають поступових змін, пов'язаних із модифікаціями коду, змінами конфігурацій, оновленням інфраструктури й зростанням навантажень, що призводить до появи деградаційних процесів. Традиційні підходи до супроводження програмного забезпечення, як правило, мають реактивний характер і ґрунтуються на усуненні вже зафіксованих інцидентів, що знижує ефективність експлуатації та підвищує ризик критичних відмов [2, 3].

З огляду на сказане актуальним є розвиток підходів превентивного супроводження програмних продуктів, орієнтованих на раннє виявлення ознак деградації та потенційних проблем ще до їх переходу в проблемний стан [1]. Реалізація такого підходу потребує формалізованого оцінювання експлуатаційного стану програмного забезпечення на основі сукупності різномірних показників, що відтворюють поведінку системи в умовах експлуатації.

Водночас ідентифікація стану програмного забезпечення на практиці ускладнюється низкою обмежень. Зокрема здебільшого відсутні репрезентативні історичні дані експлуатації, а також апіорні оцінки проблемності можливих станів програмного продукту. Це унеможливує використання класичних

статистичних або навчальних підходів і зумовлює необхідність формування інформативної множини ознак у процесі поетапного накопичення інформації про систему.

Перспективним інструментом для розв'язання окресленого завдання є метод компараторної ідентифікації [4, 5], що дає змогу оцінювати стан об'єкта способом порівняння векторів ознак без обов'язкової наявності повних навчальних вибірок. У контексті превентивного супроводження програмного забезпечення це відкриває можливість поєднання апріорних властивостей, отриманих на етапах тестування, з апостеріорними експлуатаційними показниками, як-от тренди й ознаки аномальної поведінки.

Отже, ключовою науково-практичною проблемою є розроблення формалізованого методу превентивного супроводження програмного забезпечення, що поєднує поетапне формування множини ознак експлуатаційного стану з компараторною моделлю ідентифікації. Такий підхід має забезпечувати інтерпретовану класифікацію стану програмної системи з погляду експлуатаційної реакції та підтримувати автоматизоване формування інцидентів супроводження із зазначенням типу проблеми, її локалізації, порушеної умови або тенденції до її порушення, рівня терміновості та відповідальної команди. Зважаючи на сказане, сформулюємо основні питання дослідження.

RQ1. Яким чином може бути побудована компараторна модель формалізованої ідентифікації експлуатаційних станів програмного забезпечення в умовах обмеженої історії інцидентів?

RQ2. Яка структура й послідовність поетапного формування алфавіту ознак є достатньою для раннього виявлення деградаційних процесів і забезпечення функціональної визначеності моделі?

RQ3. Яким чином результати компараторної ідентифікації можуть бути використані для автоматизованого формування інцидентів супроводження з визначенням типу проблеми, локалізації та рівня терміновості?

Аналіз сучасних публікацій щодо проблеми превентивного супроводження програмного забезпечення

Превентивне супроводження програмного забезпечення в сучасних дослідженнях висвітлюється

як стратегія проактивного втручання в життєвий цикл програмних систем з метою виявлення та усунення потенційних проблем до їх переходу у фазу відмов або критичних дефектів. У стандартах життєвого циклу превентивне супроводження визначається як модифікація програмного продукту після впровадження в експлуатацію, яка спрямована на ідентифікацію латентних дефектів і зниження ризику їх прояву в майбутньому [6]. Незважаючи на формалізацію цього поняття на нормативному рівні, його практична реалізація залишається неоднорідною та недостатньо систематизованою. Аналіз сучасних публікацій свідчить, що превентивне супроводження все частіше розглядається не як завершальна фаза, а як безперервний процес [7], тісно пов'язаний з розробленням і експлуатацією програмного продукту [8, 9]. Однак така інтеграція породжує нові методологічні й організаційні виклики, що досі не мають усталених рішень.

Однією з ключових проблем превентивного супроводження є складність раннього виявлення латентної деградації програмних систем. Дослідники доводять, що більшість відмов має накопичувальну властивість і супроводжується поступовими змінами метрик продуктивності, якості або структури системи, які не фіксуються традиційними механізмами тестування й моніторингу [1, 10]. У роботах, присвячених аналізу журналів подій, наголошено, що аномалії в логах можуть бути ранніми індикаторами майбутніх інцидентів, проте їх ідентифікація потребує складних методів оброблення даних і машинного навчання [11, 12].

Водночас огляди самонавчальних і самоадаптивних систем свідчать, що наявні підходи часто не здатні відрізнити тимчасові флуктуації від стабільних трендів деградації, що обмежує їх застосовність у превентивному супроводженні [13]. Як наслідок, превентивні дії ініціюються надто пізно, коли система вже потребує коригувального, а не профілактичного втручання.

Також суттєвою проблемою є відсутність сталого зв'язку між сигналами ризику, отриманими в процесі моніторингу, та управлінськими рішеннями щодо супроводження. Навіть коли потенційні проблеми ідентифікуються автоматизованими засобами, вони часто не трансформуються в конкретні завдання для команди розроблення або супроводження [1, 14]. У дослідженнях з управління супроводженням програмного забезпечення зазначено, що аналітичні

результати існують ізольовано від систем планування робіт, що призводить до ігнорування превентивних рекомендацій [15]. Систематичні огляди, присвячені поєднанню супроводження й машинного навчання, вказують на фрагментарність інтеграції аналітичних моделей у процеси прийняття рішень [16]. Це зумовлює домінування реактивного підходу, коли ресурси спрямовуються переважно на усунення вже наявних дефектів, а не на запобігання їх виникненню.

Проблема обґрунтованої пріоритизації превентивних заходів тісно пов'язана з управлінням технічним боргом. У низці систематичних оглядів наголошено, що, незважаючи на наявність численних метрик технічного боргу, відсутні універсальні методи визначення того, які превентивні дії мають найвищий пріоритет у конкретний момент часу [17]. Це ускладнює планування профілактичних робіт і знижує їх привабливість для менеджменту. Дослідження, присвячені прогнозуванню дефектів на основі історії змін, демонструють потенціал кількісного оцінювання ризиків, однак їх застосування потребує високої якості інформації та значних обчислювальних ресурсів [18]. Як наслідок, превентивні заходи часто розглядаються як другорядні порівняно з функціональним розвитком системи.

Хоча методи машинного навчання активно використовуються для підтримки превентивного супроводження, огляд джерел вказує на низку їх обмежень. Зокрема складні моделі глибокого навчання, застосовувані для виявлення аномалій або прогнозування дефектів, визначаються низькою інтерпретованістю, що ускладнює їх практичне впровадження [11, 13]. У студіях з експлуатації *ML*-систем також наголошено на проблемі деградації якості моделей у часі, зумовлену дрейфом даних і зміною контексту застосування [19]. У цьому контексті формування практик *MLOps* розглядається як відповідь на виклики превентивного супроводження *ML*-орієнтованих систем, однак систематичні огляди засвідчують відсутність єдиних стандартів і зрілих процесів у цій сфері [16]. Це обмежує можливість широкого впровадження превентивних підходів у проєктах з інтенсивним використанням машинного навчання.

Окрему групу проблем становлять організаційні та процесні чинники. У дослідженнях продемонстровано, що ефективність превентивного супроводження здебільшого залежить від структури команди й рівня взаємодії між розробленням і експлуатацією [20].

Попри існування міжнародних стандартів, вони мають переважно декларативну властивість і не пропонують детальних механізмів реалізації превентивного супроводження на практиці [6, 9].

Систематичний аналіз літератури демонструє, що превентивне супроводження програмного забезпечення є актуальним, але має низку методологічних прогалин. Основні питання зосереджені навколо раннього виявлення деградації, інтеграції аналітичних сигналів у процеси прийняття рішень, пріоритизації превентивних дій, а також організаційних бар'єрів. Аналіз наукових досліджень дав змогу обґрунтувати актуальність розроблення формалізованих моделей превентивного супроводження.

Мета й завдання дослідження

Метою статті є розроблення та обґрунтування методу превентивного супроводження програмного забезпечення на основі компараторної моделі, що уможливує формалізоване оцінювання його експлуатаційного стану, раннє виявлення деградаційних процесів, класифікацію експлуатаційної реакції та автоматизоване формування інцидентів супроводження в умовах обмеженої історії інцидентів.

Для досягнення поставленої мети в роботі необхідно розв'язати певні завдання.

1. Обґрунтувати вимоги до формалізованого оцінювання експлуатаційного стану програмного забезпечення в умовах превентивного супроводження та обмеженої доступності історичних даних.

2. Побудувати компараторну модель ідентифікації експлуатаційних станів способом формалізації взаємозв'язків між ознаками у вигляді системи логічних рівнянь і перевірки її функціональної визначеності.

3. Розробити метод компараторної ідентифікації експлуатаційних станів програмного забезпечення, оснований на побудованій компараторній моделі.

4. Провести емпіричну валідацію запропонованого методу на експлуатаційних даних і виконати порівняння з традиційними пороговими механізмами сповіщення.

Матеріали й методи

Центральною проблемою розв'язання задачі превентивного супроводження програмного забезпечення є побудова процедур, здатних виявляти потенційні деградаційні стани до настання відмови або порушення якості. Класичні підходи машинного

навчання, зокрема класифікатори й регресійні моделі прогнозування, зазвичай спираються на наявність історії інцидентів або принаймні на приклади негативних наслідків, що утворюють навчальну вибірку. У контексті превентивного супроводження ця передумова є центральною проблемою: ефективне превентивне супроводження програмної системи полягає в тому, що потенційно ризикова ситуація визначається завчасно й очікуваний негативний розвиток ситуації переривається вчасним утручанням, унаслідок чого спостережувані дані не містять оцінки негативних наслідків. За такого підходу ознаки, основані на подіях (інцидент / неінцидент, порушення / непорушення *SLO*) перестають бути об'єктивним відтворенням латентного стану системи. Це знижує коректність прямого перенесення класичного *ML*-підходу на задачу превентивного супроводження й мотивує розгляд альтернативних ідентифікаційних парадигм, які не вимагають повної та репрезентативної історії.

Метод компараторної ідентифікації, розроблений у працях [4, 5, 21], пропонує таку альтернативу, замінюючи вимогу наявності числової цільової функції або "істинних" міток на використання лише бінарної інформації про попарну нерозрізненість або, навпаки, розрізненість ознак. Його вихідним положенням є те, що для відновлення латентної структури станів об'єкта достатньо мати доступ до компаратора, який на парі ознак повертає значення предиката, інтерпретованого як "еквівалентність" цих ознак з погляду внутрішнього стану [4]. Такий підхід принципово відрізняється від класичного *ML*: він не апроксимує безпосередньо ймовірність небажаної події та не вимагає спостереження цієї події в даних, натомість ідентифікує класи станів як фактор-множину за відношенням, що задається компаратором.

Формалізація методу здійснюється апаратом алгебри скінченних предикатів і систем логічних рівнянь. Метою формалізації є побудова такої системи логічних рівнянь, яка однозначно відтворює спостережувані значення предикатів у внутрішні стани системи. Ця система будується ітеративно способом послідовного уточнення алфавіту предикатів до досягнення функціональної визначеності рішення.

Нехай задано універсум об'єктів U . Предикатом називається відображення

$$P_i : U \rightarrow \{0,1\}, i = 1, \dots, n.$$

Сукупність предикатів $A = \{P_1, \dots, P_n\}$ утворює алфавіт опису об'єктів. Кожному об'єкту $u \in U$ відповідає предикатний образ

$$P(u) = (P_1(u), \dots, P_n(u)) \in \{0,1\}^n.$$

Алфавіт предикатів A формує на множині U відношення еквівалентності

$$\sim_A \subseteq U \times U,$$

$$u \sim_A v \Leftrightarrow \bigwedge_{i=1}^n (P_i(u) \leftrightarrow P_i(v)), u, v \in U.$$

Це відношення розбиває універсум U на класи еквівалентності, кожен з яких визначається унікальним предикатним образом. Нехай $D = \{d_1, \dots, d_m\}$ – скінченна множина можливих станів та існує невідома функція класифікації (прийняття рішення щодо певного стану) $Y : U \rightarrow D$. Для формалізації додано булеві предикати $D_j(u) = 1 \Leftrightarrow Y(u) = d_j$, $j = 1, \dots, m$. На ці предикати накладено структурні аксіоми:

1) аксіома повноти:

$$\bigvee_{j=1}^m D_j(u) = 1;$$

2) аксіома взаємонесумісності:

$$D_i(u) \wedge D_j(u) = 0, i \neq j.$$

Алфавіт предикатів є придатним для ідентифікації, якщо виконується умова функціональної визначеності:

$$u \sim_A v \Rightarrow Y(u) = Y(v). \quad (1)$$

Умова (1) означає, що рішення $Y(u)$ однозначно визначається предикатним образом $P(u)$: якщо $P(u) = P(v)$, то $Y(u) = Y(v)$. Порушення цієї умови інтерпретується як неповнота алфавіту предикатів.

Алгебра предикатів дає змогу будувати логічні формули з предикатних змінних за допомогою операцій кон'юнкції, диз'юнкції та заперечення. Інформація, отримана від компаратора, фіксується у вигляді логічних обмежень двох основних типів: імплікації допустимості

$$\Phi(P_1, \dots, P_n) \Rightarrow D_j,$$

яка означає, що за виконання умови $\Phi(P_1, \dots, P_n) = 1$ можливе лише рішення d_j , та імплікації заборони

$$\Psi(P_1, \dots, P_n) \Rightarrow \neg D_j,$$

де $\Phi, \Psi : \{0,1\}^n \rightarrow \{0,1\}$ – булеві функції над предикатами $P_1(u), \dots, P_n(u)$. Сукупність усіх таких

імплікацій разом з аксіомами повноти й взаємнесумісності утворює систему логічних рівнянь, що визначає допустимі значення предикатів рішень:

$$\begin{cases} \Phi(u) \Rightarrow D_j(u), \\ \Psi(u) \Rightarrow \neg D_j(u), \\ \bigvee_{j=1}^{j=1} D_j(u) = 1, \\ D_i(u) \wedge D_j(u) = 0, i \neq j. \end{cases}$$

За умови (1) задача зводиться до побудови булевих функцій f_j на просторі $\{0,1\}^n$:

$$D_j(u) = f_j(P_1(u), \dots, P_n(u)), \quad j \in \{1, \dots, m\},$$

тобто множини

$$X_j = \{x \in \{0,1\}^n \mid f_j(x) = 1\}$$

утворюють розбиття простору предикатних образів.

У реальних умовах експлуатації алфавіт ознак A не є інформаційно повним. Тому поряд із функціональною моделлю прийняття рішення $Y(u) = f(P(u))$ доцільно розглядати парну постановку, яка дає змогу визначати подібність нової ситуації до вже відомих. У парній постановці додається предикат порівняння

$$E(u, v) = 1 \Leftrightarrow Y(u) = Y(v)$$

або через предикати рішень

$$E(u, v) \Leftrightarrow \bigvee_{j=1}^m (D_j(u) \wedge D_j(v)).$$

Сукупність алфавіту A , предиката порівняння E та відповідне розбиття множини об'єктів на класи еквівалентності утворюють компараторну модель. Інформація про латентні стани задається відношенням еквівалентності між об'єктами. Якщо рішення однозначно визначається предикатним образом $P(u)$, то існує булева функція

$$\Theta: \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1\},$$

така, що

$$E(u, v) = \Theta(P(u), P(v)).$$

У загальному вигляді задача компараторної ідентифікації формулюється так [4, 21]. Нехай на множині M задано предикат E . Тоді існує відображення

$$F: M \rightarrow N,$$

таке, що

$$E(x, y) = 1 \Leftrightarrow F(x) = F(y),$$

тоді й лише тоді, коли E є відношенням еквівалентності на M , тобто для всіх $x, y, z \in M$ виконуються властивості рефлексивності, симетричності та транзитивності [4]:

$$E(x, x) = 1,$$

$$E(x, y) = 1 \Rightarrow E(y, x) = 1,$$

$$E(x, y) = 1 \wedge E(y, z) = 1 \Rightarrow E(x, z) = 1.$$

За виконання цих умов множина M факторизується на класи еквівалентності $[x] = y \in M : E(x, y) = 1$, і тоді можна покласти $F(x) = [x]$, отримуючи реалізацію рівності $E(x, y) \Leftrightarrow F(x) = F(y)$. Ця теорема одночасно визначає фундаментальне обмеження методу: якщо E не є еквівалентністю, то не існує жодного відображення F , узгодженого з E в наведеному сенсі, а отже, задача ідентифікації є некоректно поставленою незалежно від обсягу даних або складності алгоритмів.

Зазначене обмеження має принциповий характер. Зокрема предикати, що виражають відношення порядку або переваги (наприклад "гірше / краще", "більш ризикове / менш ризикове"), загалом не задовольняють симетричності й часто порушують транзитивність у присутності шуму, дрейфу та контекстної залежності. У цьому разі інформація компаратора не може бути інтерпретована як ознака рівності латентних станів, і метод компараторної ідентифікації в класичному вигляді не застосовується [1].

Другим, часто неявним, обмеженням є вимога нетривіальної розрізненості ознак. Навіть якщо E формально є еквівалентністю, можлива виродженість, коли

$$\forall x, y \in M \quad E(x, y) = 1.$$

Тоді факторизація M містить рівно один клас і відображення F існує, але є інформативно тривіальним: воно не відображає жодної внутрішньої структури режимів об'єкта. Звідси випливає, що для отримання практично корисної моделі необхідна наявність принаймні двох експлуатаційно розрізнених класів, тобто нетривіальний компаратор, який інколи повертає 0 і цим формує непорожню структуру фактор-множини.

З огляду на ці формальні засади застосування компараторної ідентифікації до превентивного супроводження програмного забезпечення вимагає адаптації. По-перше, необхідно формувати

й обґрунтовувати такі відношення попарної еквівалентності станів, які відображають операційно значущі інваріанти, зокрема експлуатаційні реакції. По-друге, важливо побудувати такий предикат, що забезпечує нетривіальну розрізненість станів (з погляду експлуатаційних реакцій).

Результати дослідження

Побудова системи ознак

Основою підходу до превентивного супроводження програмного забезпечення є набір вхідних ознак стану програмної системи. Ця система ознак містить різноманітні показники якості та поведінки програмного продукту, отримані під час його експлуатації й тестування: метрики тестування, агреговані показники, сигнали аномалій, динаміка й тренди показників тощо.

Метрики тестування відбивають результати й ефективність тестування (наприклад, відсоток пройдених тестів, кількість виявлених дефектів до релізу, показники продуктивності під тестовим навантаженням тощо). Ці метрики слугують об'єктивними індикаторами правильності функціонування системи відповідно до визначених вимог і можливих сценаріїв та можуть сигналізувати про відхилення ще на етапі контролю якості.

Агреговані показники за групами властивостей – це інтегральні індикатори, обчислені для основних якісних аспектів програмного забезпечення, зокрема функціональності, умов експлуатації, безпеки й зручності використання. Кожна група об'єднує низку конкретних метрик (ознак), які відповідають ключовим властивостям якості продукту. Агрегування за категоріями дає змогу отримати узагальнену оцінку стану за кожною важливою сферою якості.

Аномалією вважається будь-яка поведінка системи або значення показника, яке істотно відхиляється від очікуваного або нормального. Кожний сигнал аномалії є ознакою потенційної проблеми, що потребує уваги.

Аналіз динаміки дає змогу розрізнити одноразові відхилення від стійких тенденцій. На основі історичних даних відстежується тенденція (тренд) ключових параметрів, що допомагає сигналізувати про можливі подальші проблеми ще до того, як будуть перевищені граничні допустимі значення метрик.

Набір ознак забезпечує комплексний профіль поточного експлуатаційного стану програмної системи. Цей профіль формалізується у вигляді скінченного алфавіту булевих предикатів, де кожний предикат відображає наявність або відсутність певної ознаки поточного стану програмної системи. Важливо, що вхідні ознаки формалізуються у вигляді скінченного набору показників, які далі можуть бути перетворені для реалізації процедури компараторної ідентифікації.

Формалізація компараторної моделі

Алгебра скінченних предикатів (АСП) надає універсальну мову для опису скінченних множин ознак і відношень між ними у вигляді предикатних формул і рівнянь. На відміну від класичної булевої алгебри, яка оперує бінарними змінними й виразами, АСП працює з предикатами над скінченними універсумами, даючи змогу описувати будь-які відношення між об'єктами. У нашому випадку універсумом є множина можливих значень вектора ознак, а предикат відповідає класифікаційній функції, що встановлює належність стану до певного класу еквівалентності.

Реалізація методу превентивного супроводження на основі компараторної моделі передбачає такі етапи: 1) трансформація спостережуваних ознак у двійкові індикатори; 2) формування вектора стану програмної системи (предикатного образу); 3) побудова системи рівнянь для визначення класів еквівалентності; 4) порівняння векторів стану і встановлення еквівалентності (застосування компаратора); 5) вибір і реалізація відповідної експлуатаційної реакції системи.

На першому етапі кожен вхідний показник (метрика або сигнал), значення яких збираються під час моніторингу, перетворюється на логічний (двійковий) індикатор, що набуває значення 1 або 0 залежно від виконання певної умови. Для кількісних метрик додаються порогові значення або допустимі діапазони: якщо метрика перебуває в межах норми – відповідний індикатор 0 (немає відхилення), у разі перевищення порогу або виходу за межі норми – індикатор 1 (фіксується проблема). Для якісних або категоріальних ознак визначаються логічні умови або правила. Наприклад, сигнал аномалії безпосередньо є бінарним індикатором (1 – виявлено аномалію, 0 – ні), агрегований показник групи може розбиватися на кілька індикаторів (наприклад,

$Functionality_OK = 1$, якщо всі функціональні тести пройдено, $Security_Risk = 1$, якщо знайдено критичну вразливість, тощо). Унаслідок цього етапу вихідний різномірний вектор даних перетворюється на набір логічних змінних $\{x_1, x_2, \dots, x_n\}$, кожна з яких набуває значення 1 за наявності відповідного негативного симптому або 0 у разі його відсутності.

Отже, двійкові індикатори згруппуються у впорядкований набір – вектор стану $x = (x_1, x_2, \dots, x_n)$, кожна компонента якого відповідає окремому індикатору. Вектор x є формальним відображенням поточного стану програмної системи в просторі ознак. Розмірність n вектора стану визначається повним переліком контрольованих ознак, таким чином метод легко розширюється

$$F_i(x_1, \dots, x_n) = (a_1^{(1)} \wedge a_2^{(1)} \wedge \dots \wedge a_n^{(1)}) \vee (a_1^{(2)} \wedge \dots \wedge a_n^{(2)}) \vee \dots \vee (a_1^{(k)} \wedge \dots \wedge a_n^{(k)}),$$

де $a_j^{(q)}$ дорівнює або x_j , або $\neg x_j$ (залежно від того, яким має бути значення ознаки j у q -му варіанті стану, що належить класу C_i). Отже, кожен кон'юнкт відповідає окремому поєднанню ознак, що належить цьому класу, а диз'юнкція таких кон'юнктив перелічує всі можливі варіанти в межах класу. Теоретично доведено, що виділення класів еквівалентності через ДНФ забезпечує повноту опису всіх можливих комбінацій ознак для кожного класу.

Компаративно-орієнтований підхід реалізує процедуру ідентифікації класу поточного стану способом порівняння вектора стану x , отриманого з поточних даних, із умовами (еталонами) для кожного класу C_i . Формально це означає перевірку виконання логічного рівняння $F_i(x) = 1$ для кожного i . Якщо для деякого класу C_k підставлення компонент x у формулу F_k задовольняє її (значення $F_k(x)$ істинне), то поточний вектор стану належить цьому класу еквівалентності. Практично таке зіставлення здійснюється як попарне порівняння бітів двох векторів: вектора x та еталонного вектора (або векторів) класу C_i . Еквівалентність двох станів встановлюється, якщо всі відповідні компоненти збігаються. За більш складних умов (коли клас визначений через диз'юнкцію кількох кон'юнктив) перевірка еквівалентності зводиться до порівняння x з кожним допустимим шаблоном класу: якщо x збігається з хоча б одним шаблоном (варіантом

додаванням нових індикаторів без зміни загального підходу).

На основі експертних знань і аналізу даних будується формальна логічна модель, що описує кожний допустимий клас стану системи. Для цього визначається множина класів еквівалентності C_1, C_2, \dots, C_m , які відповідають різним ситуаціям експлуатації програмної системи. Для визначеності в межах цього дослідження запропоновано обрати три класи експлуатаційних реакцій: "норма", "проблема", "попередження". Кожний клас еквівалентності математично задається логічним виразом щодо компонент вектора x . Зручним канонічним способом подання таких умов є диз'юнктивна нормальна форма (ДНФ). Наприклад, формула для класу C_i може мати вигляд

кон'юнкту), вважаємо, що x належить класу. Таким чином здійснюється компараторна ідентифікація поточного стану: модель порівнює стан системи із заздалегідь визначеними еталонами нормального чи аномального функціонування й обирає еквівалентний клас. З погляду АСП, система рівнянь у ДНФ, побудована для класів стану, фактично є предикатною моделлю об'єкта (програмної системи), яка відтворює простір станів у термінах істинних значень предикатів (індикаторів).

Кожний виділений клас еквівалентності станів програмної системи пов'язаний із певною реакцією або заходом з боку експлуатаційної команди супроводження. Крім належності до категорії, для кожного класу може визначатися рівень критичності, який деталізує пріоритет реагування. Наприклад, клас "проблема" може мати підкласи за критичністю: високий (негайно виправити, падіння системи), середній (значна функціональна деградація), низький (некритичний збій окремого модуля). Критичність визначається як функція від значень індикаторів у векторі стану (наприклад, аномалія безпеки може автоматично підвищувати критичність). Отже, еквівалентні стани об'єднано в класи щодо їх впливу на функціонування програмного продукту, і кожному класу відповідає наперед визначена реакція підтримки й рівень пріоритетності. Це дає змогу автоматизувати ухвалення рішень: формально порівнявши вектор стану з класами, система супроводження одразу визначає, чи потрібні дії та як швидко.

**Метод компараторно-орієнтованого
превентивного супроводження
програмного забезпечення**

Основна ідея методу полягає в побудові компараторної моделі, де кожне рівняння описує окремий клас еквівалентності. Клас еквівалентності визначається як множина станів програмного забезпечення, що потребують однакової реакції з боку експлуатаційної команди. Отже, метод зводить задачу превентивного супроводження до задачі логічної ідентифікації класу, до якого належить поточний стан системи.

У процесі експлуатації алгоритм періодично формує вектор індикаторів для поточного вікна спостереження й виконує компараторну ідентифікацію способом підставлення цього вектора в систему

Input :

- T – специфікація тестів і критерії приймання (*acceptance rules*);
- T – результати тестування й тестові метрики;
- $M(t)$ – потік експлуатаційного моніторингу;
- $A(t)$ – потік зафіксованих аномалій;
- \mathcal{W} – параметри вікон агрегування (розмір, крок);
- \mathcal{R} – модель ризику / критичності (правила домену, безпека, *SLO/SLA*);
- \mathcal{O} – правила локалізації (ознаки / артефакт \rightarrow команда);
- Π – політика ревізії моделі (період / умови).

Output :

$$COPM = (I, P, F, \text{Map}),$$

- де I – множина бінарних індикаторів (скінчених предикатів);
 P – правила / предикати бінаризації;
 $F = \{F_c\}_{c \in C}$ – система ДНФ-рівнянь для класів еквівалентності;
 Map – відображення класу \rightarrow (*reaction, priority, owner*);
 Γ – потік превентивних завдань / алертів.

Phase A. Побудова індикаторів із тестування (апріорна база)

1. $I \leftarrow \emptyset, P \leftarrow \emptyset$.
2. $S_{test} \leftarrow \text{ExtractTestFeatures}(T, T)$.
3. Для кожної ознаки $s \in S_{test}$:
 - 3.1. Побудувати предикат $p_s(\cdot)$ на основі *acceptance criteria* й \mathcal{R} .
 - 3.2. Визначити індикатор $i_s := 1[-p_s]$ (1 – порушено, 0 – виконується).
 - 3.3. $I \leftarrow I \cup \{i_s\}, P \leftarrow P \cup \{p_s\}$.

Phase B. Додавання експлуатаційних ознак (агрегати, тренди, аномалії)

4. Задати групи агрегування $G = \{Func, Env, Sec, Usab\}$.
5. Для кожного вікна $w \in \mathcal{W}$:
 - 5.1. $Z_w \leftarrow \text{Aggregate}(M(t), w, G)$ (агреговані показники за групами).
 - 5.2. $D_w \leftarrow \text{Trends}(Z_w)$ (дрейф / стабільність / повернення до еталона).

ДНФ-рівнянь. Як наслідок, визначається множина класів, предикати яких виконуються для поточного стану. Якщо спрацьовує рівно один клас, він вважається ідентифікованим. У разі спрацювання кількох класів застосовується правило розв'язання конфліктів. Якщо жоден клас не спрацьовує, стан вважається новим або раніше неописаним і обробляється як стан, що потребує додаткової уваги.

Після ідентифікації класу формується експлуатаційна реакція, що передбачає тип дії, рівень критичності та відповідальну команду. Додатково генерується пояснення результату у вигляді переліку індикаторів і диз'юнкта ДНФ, який спрацював. За необхідності автоматично створюється превентивне завдання або алерт для відповідної команди.

Алгоритм методу містить такі етапи й кроки.

5.3. $A_w \leftarrow Anomalies(A(t), w)$ (наявність / тип / оцінка аномалії).

5.4. $S_{ops} \leftarrow ComposeFeatures(Z_w, D_w, A_w)$.

5.5. Для кожної нової ознаки $s \in S_{ops}$, що поки не модельована:

- побудувати $p_s(\cdot)$ із \mathcal{R} і еталонів тестування;
- визначити $i_s := 1[-p_s]$;
- додати i_s до I , а p_s до P .

Phase C. Побудова класів еквівалентності (компараторна модель)

6. Визначити множину реакцій $R = \{NORM, ATTENTION, PROBLEM\}$.

7. Сформувані множини класів еквівалентності C за принципом:

стани еквівалентні моді й лише моді, коли потребують однакової експлуатаційної реакції.

8. Для кожного класу $c \in C$:

8.1. Визначити допустимі патерни індикаторів $X_c \subseteq 0, 1^{|I|}$ (з тестів, аналізу ризиків тощо).

8.2. Побудувати ДНФ-предикат класу:

$$F_c(x) = \bigvee_{k=1}^{K_c} \left(\bigwedge_{l \in L_{c,k}} \lambda_l(x) \right),$$

де $\lambda_l(x)$ – літерал x_j або $\neg x_j$.

9. Визначити $Map(c) = (reaction, priority, owner)$ за \mathcal{R} та \mathcal{O} .

10. (Рекомендовано) Провести валідацію:

- неперетинність $F_c \wedge F_d = 0$ для $c \neq d$ на множині відомих станів;
- покриття "відомих" станів $\bigvee_{c \in C} F_c = 1$.

Phase D. Онлайн-ідентифікація та формування превентивних завдань

11. Для кожного нового вікна w у процесі експлуатації:

12. Обчислити вектор індикаторів

$$x_w \leftarrow EvalIndicators(I, P, M(t), A(t), w) \in 0, 1^{|I|}.$$

13. Знайти множину спрацьованих класів

$$S \leftarrow c \in C :: F_c(x_w) = 1.$$

14. Обрати клас c^*

якщо $|S| = 1$, то $c^* \leftarrow S$;

якщо $|S| > 1$, то $c^* \leftarrow Resolve(S)$ (правило: *safety-first, max priority, max specificity*);

якщо $|S| = 0$, то $c^* \leftarrow c_{new}$ і реакція за замовчуванням ATTENTION (*triage*).

15. $(r^*, p^*, o^*) \leftarrow Map(c^*)$.

16. Сформувані пояснення $E \leftarrow ExplainDNF(F_{c^*}, x_w)$ (який диз'юнкт ДНФ спрацював + активні індикатори).

17. Якщо $r^* \neq NORM$, сформувані завдання:

$$\gamma \leftarrow CreateTask(o^*, p^*, r^*, E, w), \quad \Gamma \leftarrow \Gamma \cup \{\gamma\}.$$

Phase E. Ревізія моделі (адаптація)

18. Якщо політика Π ініціює ревізію (періодично або за умови накопичення c_{new} / дрейфу / нових релізів):

19. Зібрати фідбек (результати реакцій, нові релізи, нові тестові метрики).

20. Оновити P (пороги / еталони), оновити I (вилучити шумові, додати нові індикатори).

21. Уточнити C і перебудувати відповідні F_c у ДНФ; повторити крок 10.

22. Повернути оновлену модель $COPM = (I, P, F, Map)$.

Отже, на відміну від методів машинного навчання, які ґрунтуються на статистичному узагальненні історичних даних, запропонований метод використовує апріорні знання, зафіксовані на етапі тестування, та формалізує їх у вигляді скінченної системи предикатів. Це дає змогу застосовувати метод за відсутності або недостатності статистики відмов, а також забезпечує інтерпретованість результатів ідентифікації.

Оскільки програмні системи еволюціонують у процесі експлуатації, метод передбачає регулярну ревізію компараторної моделі. Підставами для ревізії можуть бути накопичення нових невідомих станів, зміни вимог, нові релізи або результати аналізу інцидентів. У процесі ревізії оновлюються предикати, порогові значення й склад індикаторів, після чого уточнюються класи еквівалентності й перебудовуються відповідні ДНФ-рівняння. Це дає змогу підтримувати актуальність моделі та зберігати її здатність до превентивної ідентифікації деградацій програмного забезпечення.

Отже, метод превентивного супроводження програмного забезпечення на основі компараторної ідентифікації уможливорює формалізований, інтерпретований і незалежний від статистики підхід до раннього виявлення проблем у програмних системах. Поєднання тестових і експлуатаційних ознак у межах компараторної моделі створює науково обґрунтований базис для превентивного супроводження й автоматизованого формування експлуатаційних рішень.

Експериментальні дослідження: симуляційний приклад

Метою експерименту є емпірична перевірка здатності запропонованого методу компараторної ідентифікації виявляти деградаційні процеси раніше, ніж класичні порогові алерти, та формувати інтерпретовані стани експлуатаційної реакції: норма; ознаки деградації; проблема. Для забезпечення контрольованої "істини" про момент деградації застосовано симуляційний сценарій з керованим дрейфом метрик.

Як об'єкт спостереження розглядається мобільний застосунок, що

– зчитує вимірювання глюкози з глюкометра (через канал передачі даних);

– передає показники на сервер або виконує локальне обчислення;

– отримує / обчислює прогноз (наприклад, горизонтом 30–60 хв);

– демонструє результат користувачу.

Експлуатаційні проблеми, релевантні для превентивного супроводження в цьому прикладі: погіршення зв'язку з пристроєм, зростання затримок отримання прогнозу, деградація якості даних (пропуски або аномальні значення), що зрештою призводить до збоїв синхронізації або погіршення UX.

В експерименті використано три метрики, які є реалістичними для мобільного застосунку й спостережувані в *production* без доступу до медичних "істинних значень".

M1. Затримка отримання прогнозу (Prediction latency)

Визначення: час від моменту отримання нового вимірювання глюкози застосунком до моменту отримання прогнозу (або завершення локального *inference*). Познака – L_t (мс).

Очікувана поведінка: у нормі – стаціонарний розподіл із незначними коливаннями; деградація проявляється як тренд росту L_t або зростання хвостів розподілу.

M2. Частка невдалих синхронізацій або передач даних (Sync failure rate)

Визначення: відношення кількості невдалих спроб зчитування або передачі даних за певний інтервал до загальної кількості спроб. Познака – $F_t \in [0,1]$.

Очікувана поведінка: у нормі близька до нуля; деградація – поступове збільшення через проблеми каналу збору первинних даних, мережі або серверних залежностей.

M3. Частка пропусків вимірювань (Missing measurement rate)

Визначення: частка очікуваних вимірювань, що не надійшли у встановлене вікно (наприклад, якщо сенсор або пристрій має очікувану періодичність). Познака – $M_t \in [0,1]$.

Очікувана поведінка: у нормі мала; деградація може проявлятися як стабільне зростання або серії "провалів".

Необхідно зауважити, що метрика M3 не є медичною оцінкою коректності, а суто експлуатаційним індикатором якості потоку даних.

З метою зниження шуму й формування ознак для компаратора застосовано агрегацію на ковзних вікнах. Базове вікно спостереження визначено: $W = 60$ хв, що є достатнім для виявлення деградаційних трендів, але не надто великим, щоб ускладнити визначення переходу до проблемного стану. Крок оновлення $\Delta = 5$ хв (оновлення агрегатів кожні 5 хв) забезпечує оперативність реакції для мобільного застосунку.

Наступним етапом методу є формування агрегатних ознак і трендів. Для кожного вікна W обчислюються:

– $p95(L)$ та $median(L)$ – для фіксації "хвоста" затримок;

– \bar{F} – середнє значення F_t на вікні;

– \bar{M} – середнє значення M_t на вікні.

Для раннього виявлення деградації додаються трендові властивості на двох масштабах часу:

коротке вікно: $W_s = 30$ хв;

довге вікно: $W_l = 6$ год.

Для кожної агрегованої метрики X (наприклад, $p95\ latency$) оцінюється проста тенденція: $\Delta X = X(W_s) - X(W_l)$ (різниця короткого й довгого середнього / квантиля), або знак / величина нахилу лінійної апроксимації на W_s (за потреби).

Як порівняльний базовий підхід в експерименті застосовано статичні пороги, типові для реактивного моніторингу.

Baseline-Alert-1 (Latency): спрацьовує, якщо $p95(L) > 1700$ мс на вікні W .

Baseline-Alert-2 (Sync failures): спрацьовує, якщо $\bar{F} > 0.05$ (тобто понад 5% невдалих синхронізацій) на W .

Baseline-Alert-3 (Missing measurements): спрацьовує, якщо $\bar{M} > 0.10$ (понад 10% пропусків) на W .

Важливо зауважити, що числові пороги наведено як реалістичні стартові значення для мобільного застосунку; вони можуть бути параметризовані під конкретну систему.

Для класу "ознаки деградації" упроваджено правила, що спрацьовують до перетину жорстких порогів, використовуючи тренди й комбіновані умови.

Rule-D1 (ранній дрейф затримок прогнозу). Стан "ознаки деградації", якщо виконується

$$p95(L) \in (1200, 2000] \text{ мс}$$

та одночасно

$$\Delta p95(L) > 300 \text{ мс,}$$

де $\Delta p95(L)$ оцінюється як різниця між коротким і довгим вікном.

Інтерпретація: затримка ще не критична, але помітно "повзе" вгору.

Rule-D2 (комбінована деградація каналу даних). Стан "ознаки деградації", якщо

$$\bar{F} \in (0.02, 0.05] \text{ та } \bar{M} \in (0.05, 0.10].$$

Інтерпретація: ще немає масового фейлу, але канал збору / передачі стає нестабільним і з'являються пропуски.

Rule-D3 (аномальна нестабільність без порогового перевищення). Стан "ознаки деградації", якщо протягом останніх $K = 3$ оновлень (тобто ~ 15 хв) спостерігаються "спайки" $p95\ latency$: $p95(L)$ перевищує власну медіану на W_l понад 1.5 раза у двох із трьох кроків.

Інтерпретація: нестабільність зростає, хоча середній рівень може ще бути прийнятним.

Класифікація експлуатаційної реакції реалізується ієрархічно:

"проблема", якщо спрацював будь-який *baseline-alert* (перетин жорсткого порогу);

"ознаки деградації", якщо *baseline-alert* не спрацював, але істинне хоча б одне правило *Rule-D1–D3*;

"норма", якщо не спрацювали ні *baseline-alert*, ні правила деградації.

Ця схема відповідає практичному процесу: "проблема" – це вже інцидент, "деградація" – підстава для превентивного тикета (або задачі на дослідження), "норма" – відсутність дій (продовжується експлуатація).

Для симуляційного дослідження задається момент початку деградації t_0 і керовані зміни:

– повільний дрейф $p95(L)$ вгору протягом 2–6 год;

– поступове збільшення F_t (наприклад, через погіршення зв'язку);

– зростання M_t (пропуски вимірювань).

Оцінювання проводиться за такими показниками:

- *Lead time* – різниця між моментом, коли система вперше класифікувала "ознаки деградації", та моментом першого *baseline-alert* ("проблема");
- *False alarms* – кількість хибних спрацювань "ознаки деградації" на стабільних сегментах без деградації;
- *Churn* – кількість перемикань класу стану (стабільність класифікації).

Очікуваний результат експерименту: запропоновані правила деградації забезпечують позитивний *lead time* (раннє попередження) за

прийняттю рівня *false alarms* та без надмірної нестабільності класифікації.

Для симуляційного експерименту згенеровано псевдоісторичні дані за два дні моніторингу з дискретністю вимірювання 1 хв. Приклад показників наведено на рис. 1. На цьому періоді штучно задано період деградації та параметри симуляції (інтенсивність деградації, базові рівні, а також алерти). На рис. 2 зображено графік станів (для наочності обрано фрагмент, де експлуатаційні стани явно змінюються). На рис. 3–5 продемонстровано динаміку агрегованих ознак навколо періоду деградації.

	latency_ms	sync_fail_rate	missing_rate
00:00:00	700.098412	0.008409	0.013660
00:01:00	724.074175	0.006765	0.008826
00:02:00	678.418033	0.007744	0.017609
00:03:00	629.276237	0.007835	0.014891
00:04:00	664.324433	0.008510	0.015530

Рис. 1. Генерація псевдоісторичних даних (фрагмент)



Рис. 2. Графік експлуатаційних станів (фрагмент)

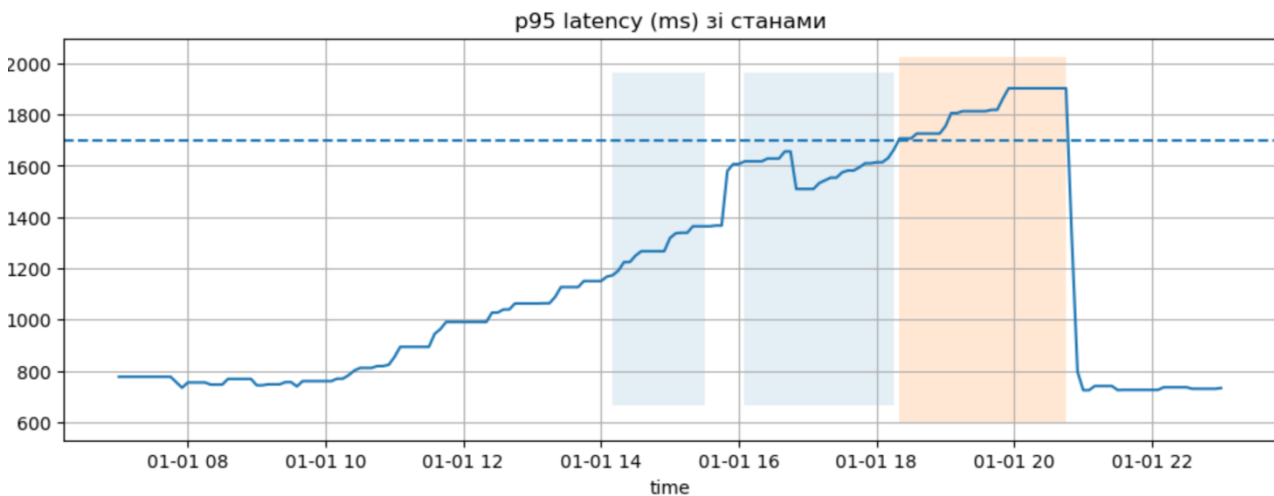


Рис. 3. Тренд затримки отримання прогнозу

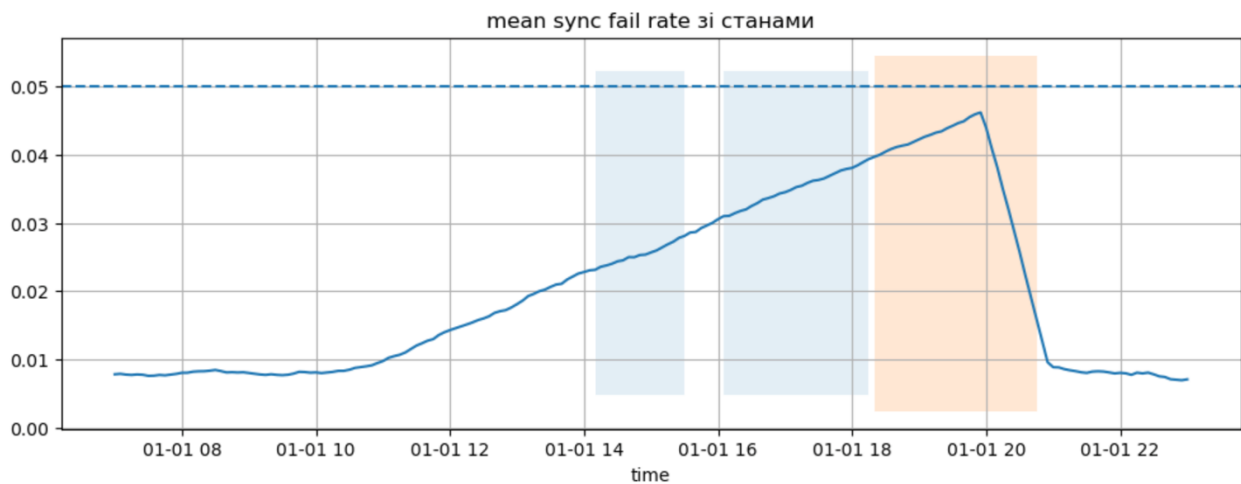


Рис. 4. Тренд частки невдалих синхронізацій або передач даних

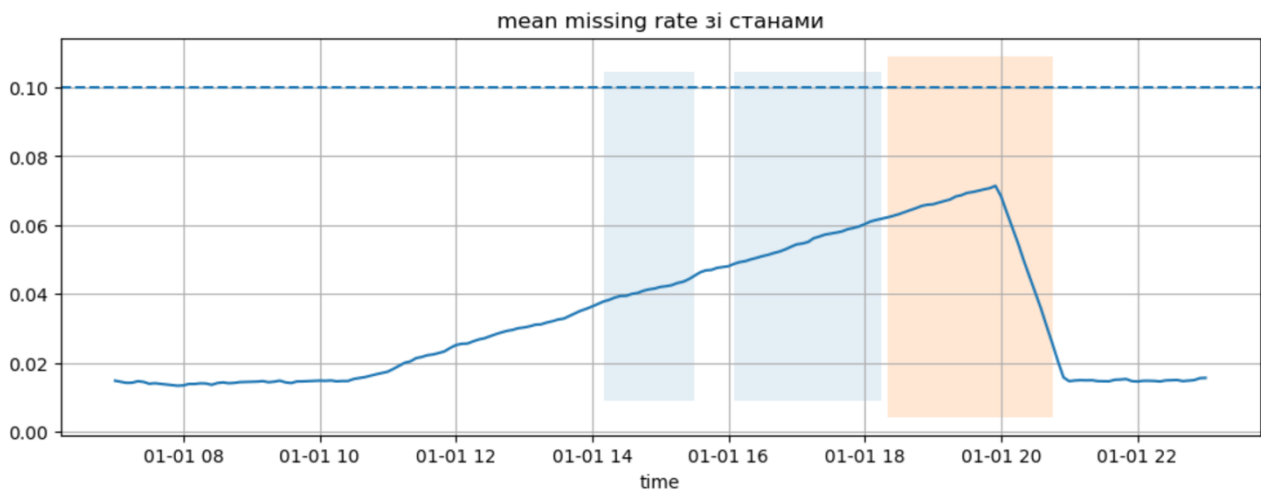


Рис. 5. Тренд частки пропусків вимірювань

Унаслідок експерименту:

– перша деградація (за визначеними правилами й ознаками) зафіксована після 14 год 10 хв від початку симуляції;

– початок проблемного стану – після 18 год 20 хв симуляції;

– *Lead time (min)*: 250.0;

– *False alarms* (кількість): 0;

– *Churn* (кількість змін): 5.

Отже, отримане значення *Lead time* кількісно підтверджує здатність запропонованого методу виявляти деградаційні процеси до моменту досягнення критичних порогів окремих показників. Випередження в часі зумовлено використанням комбінованих і трендових ознак, що дають змогу фіксувати латентні зміни динаміки стану ще до їх явного прояву. Це свідчить про реалізацію саме превентивного, а не реактивного підходу до супроводження програмного забезпечення та

підтверджує практичну ефективність компараторної моделі в задачах раннього реагування.

Обговорення

Запропонований метод превентивного супроводження програмних систем спрямований на раннє виявлення деградації. Цей метод надає формальну основу розв'язання задачі проактивного моніторингу, що актуально з огляду на сучасні тенденції переходу від реактивного до прогнозувального обслуговування програмних систем [5, 21]. Зазначимо, що в традиційних *DevOps*-підходах спостерігається нестача систематичного контролю якості програмного забезпечення на етапі експлуатації [20]. Це створює розрив між етапом тестування й реальними умовами експлуатації, який і покликаний заповнити запропонований підхід. Його новизна полягає

в поєднанні інформації, отриманої під час тестування із результатами моніторингу в експлуатації (агреговані показники, тренди, аномалії) для формування цілісної моделі станів програмної системи. Отже, компараторна модель дає змогу класифікувати стани програмної системи на основі еквівалентності експлуатаційної реакції.

З погляду науки, використання методу компараторної ідентифікації є перспективним, оскільки він забезпечує повне математичне відображення об'єкта дослідження [4, 5]. Виявлені закономірності поведінки системи записуються у вигляді системи логічних рівнянь [21], що підвищує об'єктивність і точність моделі. На відміну від статистичних чи машинних методів, які потребують великого масиву історичних даних для навчання [16, 19], компараторний підхід спирається на експертно визначені критерії та предикати. Це набуває особливого значення за відсутності розгорнутої історії збоїв, де методи машинного навчання є менш ефективними. Замість побудови "чорної скриньки" прогнозування формуються явні логічні умови належності стану до того чи іншого класу. Отже, модель є пояснюваною: команда експлуатації може зрозуміти, чому стан належить до класу "проблема" чи "потребує уваги", оскільки це впливає з порушення конкретних метрик або умов. У цьому сенсі запропонований підхід відповідає сучасним вимогам щодо пояснюваності рішень [22].

Очікуване наукове значення запропонованого методу полягає в об'єднанні теорії компараторної ідентифікації, метрик якості програмного забезпечення та *DevOps*-практик для проактивного супроводження. Наукова новизна методу – це формалізація процесу раннього виявлення проблем: за допомогою використання апріорно визначених класів еквівалентності станів можна математично обґрунтувати момент, коли нормальний стан переходить у аномальний чи проблемний на основі порушення логічних умов.

Практична цінність превентивного моніторингу полягає в здатності виявляти деградацію на ранній стадії – ще до того, як вона переросте в повномасштабний збій. Це досягається внаслідок контролю трендів і відхилень від еталонних значень. Такий проактивний підхід відповідає загальній концепції прогнозувального обслуговування, яка покликана мінімізувати незаплановані простої та відмови [3]. По-друге, розбиття потенційно

проблемних ситуацій на класи з огляду на критичність (низька, середня, висока важливість) і локалізацію (відповідальний домен чи команда) забезпечує пріоритизацію реакції на інциденти. Замість однакової реакції на всі тривожні сигнали команда отримує структуровану інформацію: де виникла проблема, наскільки вона серйозна і хто має втрутитися. Це дає змогу спрямувати зусилля лише туди, де є необхідність, уникаючи зайвих утручань у незначних випадках, що відповідає принципу зниження витрат унаслідок точкових інтервенцій [9]. Крім того, опис умов, за яких система перейшла в проблемний стан, надає команді експлуатації фактичний контекст для діагностики. Такий рівень обґрунтованості дій підвищує довіру до системи моніторингу й полегшує комунікацію між командами.

Запропонована компараторна модель фактично є розвитком *rule-based*-підходу, але на якісно новому рівні: замість ізольованих порогів вона використовує систему логічних умов для всього вектора ознак стану, що дає змогу фіксувати більш складні закономірності. Методи машинного навчання, зокрема алгоритми виявлення аномалій, визначають складні багатовимірні відхилення [11, 12]. Якщо порівнювати з *ML*-моделями, запропонований підхід не потребує великих масивів даних для навчання, може працювати відразу й виграє в інтерпретованості. З іншого боку, гібридні підходи намагаються виявляти нові закономірності й одночасно використовують експертні правила. У сучасних системах діагностики відмов, зокрема в промисловості, вже впроваджують багаторівневі архітектури з використанням нейронних мереж і нечіткої логіки для самодіагностики обладнання [23]. Порівняно з ними, запропонований підхід менш "автоматичний" – його адаптація до нових умов відбувається не способом машинного самонавчання, а внаслідок регулярного перегляду моделі на основі накопичених даних.

Критично оцінюючи запропонований підхід, можемо визначити кілька потенційних обмежень. По-перше, виродженість класифікації: якщо програмна система тривалий час працює безвідмовно й усі спостережувані стани еквівалентні "нормі", модель може не мати достатньої чутливості для виокремлення передвісників проблем. Частково цю проблему усуває долучення показників динаміки й аномалій. Однак налаштування порогів чутливості є

складним завданням: надто вимогливі критерії призведуть до надлишкових сповіщень, тоді як надто м'які – до пропуску реальних проблем. По-друге, повнота еталонного класу "норма" залежить від якості тестування. Якщо сценарії тестів не охопили певні режими роботи або конфігурації системи, під час експлуатації можуть виникнути стани, які не вписуються у визначені класи. Це вказує на необхідність безперервного процесу валідації моделі: нові показники моніторингу й інциденти мають періодично аналізувати експерти для уточнення меж класів і умов. Підхід передбачає такий циклічний перегляд моделі, проте на практиці це вимагатиме налагодження відповідного процесу. По-третє, інтеграція запропонованого рішення в наявну інфраструктуру моніторингу може спричинити складнощі. Необхідно забезпечити збір усіх потрібних метрик з експлуатаційного середовища, їх коректну агрегацію та вчасний аналіз.

Отже, метод компараторної ідентифікації в застосуванні до превентивного супроводження програмних систем є перспективним, оскільки поєднує строго формалізовану модель станів із практичною спрямованістю на раннє виявлення й класифікацію проблем. Очікується, що впровадження такого підходу сприятиме підвищенню надійності та безпеки програмних систем завдяки вчасній ідентифікації проблемних ситуацій і цілеспрямованому плануванню дій з їх усунення.

Висновки

У роботі розглянуто задачу превентивного супроводження програмного забезпечення з позицій формалізованого оцінювання його експлуатаційного стану в умовах обмеженої доступності історичних даних та відсутності апріорних оцінок проблемності можливих станів. Продемонстровано, що за таких умов доцільним є впровадження методів ідентифікації, орієнтованих на аналіз багатовимірних ознак і порівняння поточних станів із формалізованими еталонами.

Запропоновано поетапний підхід до формування множини ознак експлуатаційного стану програмного продукту, який базується на поєднанні метрик тестування, агрегованих експлуатаційних показників, трендових компонентів і ознак аномальної поведінки. Формалізація взаємозв'язків між ознаками у вигляді системи логічних рівнянь дає змогу аналізувати достатність і повноту сформованої множини ознак для задачі ідентифікації стану.

Продемонстровано можливість застосування методу компараторної ідентифікації для оцінювання експлуатаційного стану програмного забезпечення та класифікації експлуатаційної реакції програмного продукту з виокремленням нормального стану, стану з ознаками деградації та проблемного стану. Це створює основу для раннього виявлення деградаційних процесів і вчасного реагування на потенційні проблеми під час експлуатації.

Запропонований метод превентивного супроводження забезпечує автоматизоване формування інцидентів супроводження з визначенням типу проблеми, її локалізації, порушеної умови або тенденції до її порушення, рівня терміновості та відповідальної команди. Отже, досягається підвищення ефективності супроводження програмного забезпечення внаслідок переходу від реактивного до превентивного підходу.

Результати дослідження можуть бути використані для побудови систем моніторингу й супроводження програмних продуктів різного призначення. Подальші випробування доцільно спрямувати на адаптацію запропонованого підходу до різних класів програмних систем, а також на експериментальне вивчення його ефективності в умовах промислової експлуатації.

Конфлікт інтересів

Автори не мають конфлікту інтересів, зокрема фінансового, особистого, авторського чи будь-якого іншого характеру, який міг би вплинути на дослідження, а також на результати, опубліковані в цій статті.

Фінансування

Дослідження проводилося без фінансової підтримки.

Доступність даних

Дані будуть надані за обґрунтованим запитом.

Використання засобів штучного інтелекту

Автори застосовували технології штучного інтелекту (*GPT-5.2*, *DeepL*) для перевірки граматики й орфографії. Після використання цих інструментів автори перевірили й відредагували контент і несуть

повну відповідальність за зміст публікації. Авторі штучного інтелекту для досягнення результатів, підтверджують, що не застосовували технології поданих у роботі.

References

1. Chupryna, A., Repikhov, V. (2025), "Reference model for preventive software maintenance", *Management Information Systems and Devices*, No. 4 (187), P. 254–277. DOI: <https://doi.org/10.30837/0135-1710.2025.187.254>
2. Thakur, P. S., Chouhan, S. S., Rathore, S. S., Parmar, J. (2026), "Systematic literature review on software code smell detection approaches", *Journal of Systems and Software*, Vol. 235, P. 112784. DOI: <https://doi.org/10.1016/j.jss.2026.112784>
3. Liu, P. S., Chin, J. F., Ab-Samat, H., Muhammad, N. A. (2025), "Simulation-based maintenance systems: A systematic review", *Journal of Simulation*, P. 1–39. DOI: <https://doi.org/10.1080/17477778.2025.2579125>
4. Cherednichenko, O., Vovk, M., Sharonova, N., Vorzhevitina, A. (2025), "Comparator-based identification of food edibility from natural language description", *Proceedings of the 9th International Conference on Computational Linguistics and Intelligent Systems (COLINS), CEUR Workshop Proceedings*, Vol. 4015, P. 185–199. DOI: <https://doi.org/10.31110/COLINS/2025-3/014>
5. Karataiev, O., Sitnikov, D., Sharonova, N. (2023), "A method for investigating links between discrete data features in knowledge bases in the form of predicate equations", *CEUR Workshop Proceedings*, Vol. 3387, P. 224–235. Available at: <https://ceur-ws.org/Vol-3387/paper17.pdf>
6. ISO/IEC/IEEE 14764:2022 (2022), Software engineering – Software life cycle processes – Maintenance, International Organization for Standardization.
7. Shahin, M., Babar, M. A., Zhu, L. (2017), "Continuous integration, delivery and deployment: A systematic review on approaches, tools, challenges and practices", *IEEE Access*, Vol. 5, P. 3909–3943. DOI: <https://doi.org/10.1109/ACCESS.2017.2685629>
8. Giordano, G., Della Porta, A., Ferrucci, F., Palomba, F. (2025), "An evidence-based study on the relationship of software engineering practices on code smells in Python ML projects", *Proceedings of the 51st Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, P. 105–120. DOI: https://doi.org/10.1007/978-3-032-04207-1_8
9. Oliveira Carvalho, L. de, Biazotto, J. P., Feitosa, D., Nakagawa, E. Y. (2024), "Technical debt in continuous software engineering: An overview of the state of the art and future trends", *Proceedings of the 27th Ibero-American Conference on Software Engineering (CIBSE)*, P. 313–326.
10. Jiang, L., Xu, G. (2007), "Modeling and analysis of software aging and software failure", *Journal of Systems and Software*, Vol. 80, No. 4, P. 590–595. DOI: <https://doi.org/10.1016/j.jss.2006.06.029>
11. Du, M., Li, F., Zheng, G., Srikumar, V. (2017), "DeepLog: Anomaly detection and diagnosis from system logs through deep learning", *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS)*, P. 1285–1298. DOI: <https://doi.org/10.1145/3133956.3134015>
12. Jia, T., Li, Y., Yang, Y., Huang, G., Wu, Z. (2022), "Augmenting log-based anomaly detection models to reduce false anomalies with human feedback", *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, P. 3081–3089. DOI: <https://doi.org/10.1145/3534678.3539106>
13. Saputri, T. R. D., Lee, S.-W. (2020), "The application of machine learning in self-adaptive systems: A systematic literature review", *IEEE Access*, Vol. 8, P. 205948–205967. DOI: <https://doi.org/10.1109/ACCESS.2020.3036037>
14. Buse, R. P. L., Zimmermann, T. (2012), "Information needs for software development analytics", *Proceedings of the 34th International Conference on Software Engineering (ICSE)*, P. 987–996. DOI: <https://doi.org/10.1109/ICSE.2012.6227122>
15. Yalçiner, A., Dikici, A., Gökalp, E. (2024), "Data-driven software engineering: A systematic literature review", *Systems, Software and Services Process Improvement (EuroSPI 2024), Communications in Computer and Information Science*, Vol. 2179. DOI: https://doi.org/10.1007/978-3-031-71139-8_2
16. Serban, A., van der Blom, K., Hoos, H., Visser, J. (2024), "Software engineering practices for machine learning – Adoption, effects, and team assessment", *Journal of Systems and Software*, Vol. 209, P. 111907. DOI: <https://doi.org/10.1016/j.jss.2023.111907>
17. Codabux, Z., Williams, B. (2013), "Managing technical debt: An industrial case study", *Proceedings of the 4th International Workshop on Managing Technical Debt (MTD)*, P. 8–15. DOI: <https://doi.org/10.1109/MTD.2013.6608672>
18. Hall, T., Beecham, S., Bowes, D., Gray, D., Counsell, S. (2012), "A systematic literature review on fault prediction performance in software engineering", *IEEE Transactions on Software Engineering*, Vol. 38, No. 6, P. 1276–1304. DOI: <https://doi.org/10.1109/TSE.2011.103>
19. Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., Dennison, D. (2015), "Hidden technical debt in machine learning systems", *Proceedings of the 29th International Conference on Neural Information Processing Systems (NIPS)*, Vol. 2, P. 2503–2511.

20. Lwakatare, L. E., Kilamo, T., Karvonen, T., Sauvola, T., Heikkilä, V., Itkonen, J., Kuvaja, P., Mikkonen, T., Oivo, M., Lassenius, C. (2019), "DevOps in practice: A multiple case study of five companies", *Information and Software Technology*, Vol. 114, P. 217–230. DOI: <https://doi.org/10.1016/j.infsof.2019.06.010>

21. Sharonova, N., Doroshenko, A., Cherednichenko, O. (2018), "Issues of fact-based information analysis", *Proceedings of the 2nd International Conference on Computational Linguistics and Intelligent Systems (COLINS), CEUR Workshop Proceedings*, Vol. 2136, P. 11–19. Available at: <http://ceur-ws.org/Vol-2136/10000011.pdf>

22. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Pedreschi, D., Giannotti, F. (2018), "A survey of methods for explaining black box models", arXiv:1802.01933. Available at: <https://arxiv.org/abs/1802.01933>

23. Lei, Z., Shi, J., Luo, Z., Cheng, M., Wan, J. (2024), "Intelligent manufacturing from the perspective of Industry 5.0: Application review and prospects", *IEEE Access*, Vol. 12, P. 167436–167451. DOI: <https://doi.org/10.1109/ACCESS.2024.3496697>

Received (Надійшла) 19.01.2026

Accepted for publication (Прийнята до друку) 25.02.2026

Publication date (Дата публікації) 12.03.2026

Відомості про авторів / About the Authors

Репіхов Вадим Миколайович – Харківський національний університет радіоелектроніки, аспірант кафедри програмної інженерії; Харків, Україна;

Vadym Repikhov – Kharkiv National University of Radio Electronics, Postgraduate Student, Department of Software Engineering; Kharkiv, Ukraine;

e-mail: vadym.repikhov@nure.ua

ORCID ID: <https://orcid.org/0000-0002-1274-4205>

Чуприна Анастасія Сергіївна – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри програмної інженерії; Харків, Україна;

Anastasiya Chupryna – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Associate Professor at the Department of Software Engineering; Kharkiv, Ukraine;

e-mail: anastasiya.chupryna@nure.ua

ORCID ID: <https://orcid.org/0000-0003-0394-9900>

METHOD OF PREVENTIVE SOFTWARE MAINTENANCE BASED ON A COMPARATOR MODEL

The subject of the study is a method of preventive software maintenance based on a comparator model that provides formalized evaluation of the operational state and supports decision-making regarding the corresponding operational response. The purpose of the study is to improve the efficiency of preventive software maintenance through a formalized assessment of the operational state, enabling early detection of degradation processes, classification of the operational response, and automated generation of maintenance incidents under conditions of limited historical incident data. The main objectives are: to substantiate the requirements for the formalized assessment of the operational state of a software system under conditions of preventive maintenance and limited availability of historical data; to construct a comparator-based model for identifying operational states by formalizing the relationships between features in the form of a system of logical equations and verifying its functional determinacy; to develop a comparator-based method for identifying operational states based on the constructed model; and to perform experimental validation of the proposed method using operational data with comparison against traditional threshold-based alerting mechanisms. Research methods. Operational state identification is performed using the comparator-based identification method within a comparator model with predicate-based state representation. Relationships between features are formalized as a system of logical equations with verification of functional determinacy. Feature construction is based on a combination of testing metrics, aggregated operational indicators, trend characteristics, and anomaly-related features. Empirical validation is conducted on simulated operational data using sliding-window aggregation and comparison with traditional threshold-based alerting mechanisms. Results. The proposed method provides earlier detection of degradation processes compared to classical threshold-based alert mechanisms. The use of trend-based and combined features increases the stability of state

classification and reduces false positives in the studied scenario. Based on the identified state class, a preventive maintenance model is implemented to automatically generate maintenance incidents specifying the type of problem, its localization, and priority level. **Conclusions.** The obtained results advance formalized approaches to evaluating the operational state of software systems and demonstrate the applicability of the comparator model to preventive maintenance tasks in the absence of comprehensive historical incident data. The practical significance lies in enabling interpretable mechanisms for early degradation detection and automation of maintenance processes, including incident generation and support of operational response selection, without the need for large volumes of historical data.

Keywords: software; preventive maintenance; comparator model; comparator identification; operational state; software degradation; automated incident generation.

Бібліографічні описи / Bibliographic descriptions

Репіхов В. М., Чуприна А. С. Метод превентивного супроводження програмного забезпечення на основі компараторної моделі. *Автоматизовані системи управління та прилади автоматики*. 2026. № 1 (188). С. 145–162. DOI: <https://doi.org/10.30837/0135-1710.2026.188.145>

Repikhov, V., Chupryna, A. (2026), "Method of preventive software maintenance based on a comparator model", *Management Information System and Devices*, No. 1 (188), P. 145–162. DOI: <https://doi.org/10.30837/0135-1710.2026.188.145>

Viktor Levykin, Dmytro Dmytrevskyi

IMPROVING THE METHOD OF OPTIMIZING THE DISTRIBUTION OF PRODUCTION RESOURCES IN CONFECTIONERY PRODUCTION

The object of the research is the processes of production planning and utilization of production resources in confectionery production under conditions of limited production and technological capacities and variability of the product assortment. *The subject of the research* is methods and mathematical models for optimizing the allocation of material, energy, and production resources aimed at improving the efficiency of enterprise operation. *The purpose of the research* is to improve the method and mathematical model for optimizing the allocation of production resources of an enterprise, ensuring profit maximization under given production and technological conditions, resource constraints, and product assortment structure, as well as enhancing the effectiveness of managerial decision-making in production planning. To achieve the stated purpose, the following **objectives** are addressed in the study: analysis of modern methods and approaches to resource allocation in the confectionery industry; identification of key factors influencing the efficiency of material, energy, and production resource utilization; improvement of the mathematical model for enterprise profit formation; formulation of an optimization objective function considering economic and technological constraints; improvement of the method for solving the optimization problem to determine rational resource allocation parameters. The research employs the following **methods**: system analysis and generalization of scientific sources; linear and nonlinear programming methods; economic and mathematical modeling; discrete-analytical optimization methods; methods of algorithm design and software implementation. The main **results** of the study consist in improving the mathematical model of profit optimization by considering the interrelationships between technological production parameters, resource constraints, and product assortment structure. An optimization objective function has been formulated, and the method for solving it has been improved, enabling the generation, evaluation, and selection of the optimal resource allocation alternative. The proposed approach has been implemented in the form of a software and algorithmic tool suitable for integration into production management information systems. In the **conclusions**, it is established that the implementation of the improved method enhances the efficiency of production planning, rationalizes resource utilization, and increases the profitability of a confectionery enterprise, confirming its practical feasibility and effectiveness.

Keywords: optimization; resources; production; profit; modeling; algorithm; planning; automation; system; enterprise.

1. Introduction

Confectionery enterprises operate in the conditions of high competition, unstable demand, rising costs of raw materials and energy resources, as well as increased requirements for the efficient use of production resources in the confectionery area. In view of this, the rational allocation of material, labor, and production resources becomes one of the key factors in ensuring the profitability and sustainable development of an enterprise [1]. Inefficient production planning leads to overuse of raw materials, equipment downtime, and a decline in economic performance.

In practice, a significant number of confectionery enterprises use traditional approaches to planning and resource allocation based on normative methods, static planning models, or expert assessments [2]. Such approaches do not provide sufficient flexibility in conditions of changing demand, seasonal fluctuations in product range, and limited resources. As noted in works on operational management, the use of static models

reduces the adaptability of production systems and does not allow for a quick response to changes in external and internal factors [3].

The development of mathematical modeling and optimization methods has created the conditions for the application of formalized approaches to resource allocation problems in the food industry [4]. Linear and nonlinear programming methods, heuristic and metaheuristic algorithms, as well as hybrid optimization approaches have been widely implemented in scientific research [5]. However, universal optimization models often do not take into account the specifics of confectionery production, in particular, complex recipe dependencies, multi-stage technological processes, and equipment limitations [6].

Particular attention in modern research is paid to the integration of optimization methods into production management information systems [7]. Such systems provide automation of information collection, formation of production plans, and support for management decision-making. At the same time, existing software

solutions are often expensive or require significant adaptation to the specific conditions of the enterprise, which limits their practical use, especially in medium-sized enterprises [8].

In this regard, there is a pressing scientific and practical need to improve methods for allocating production resources at confectionery enterprises in the confectionery industry by developing mathematical models of profit and optimization algorithms capable of taking into account the dynamics of the production process and the variability of resource constraints. Solving this task contributes to improving the accuracy of planning, the efficiency of resource use, and the growth of the enterprise's profitability, which determines its importance from both a scientific and practical point of view.

2. Analysis of literature sources and formulation of the research issue

The problem of rational distribution of production resources is one of the key issues in the theory and practice of production management, particularly in the food and confectionery industries. In scientific research, resources are considered as a combination of material, labor, energy, and production factors, the effective use of which directly affects the cost of production, output volumes, and financial results of the enterprise [9]. In conditions of growing competition and market instability, the relevance of optimal resource planning tasks is significantly increasing.

In classical works on operations management, the main focus is on production planning methods based on deterministic models that use normative indicators and average parameter values [10]. Such approaches ensure formal simplicity and transparency of calculations, but they have limited flexibility and do not take into account the dynamics of demand changes, seasonality of the product range, and variability of resource constraints that are characteristic of confectionery enterprises.

A significant body of scientific research is devoted to the application of mathematical programming methods to resource allocation problems. In particular, linear programming is widely used to maximize profit or minimize costs when resources are limited [11]. Within such models, the objective function is usually formed on the basis of marginal income, and the variables are the production volumes of individual types of products. However, linear models do not always adequately

reproduce complex technological dependencies and the nonlinear nature of production processes in the food industry.

Today, there are models that allow for the discreteness of production decisions, equipment constraints, and interrelationships between operations [12]. Works devoted to production optimization in the food area emphasize the effectiveness of such approaches for assortment planning and capacity utilization tasks. At the same time, the complexity of implementation and high computational costs often complicate their practical application.

A separate group of studies consists of heuristic and metaheuristic methods, in particular genetic algorithms, particle swarm algorithms, and ant algorithms [13]. They demonstrate a high ability to find optimal solutions in complex multidimensional problems. However, the results of such methods can be unstable, they depend on the choice of parameters and do not always provide the necessary interpretability of management decisions, which is important for production managers [14].

Modern research is increasingly focused on integrating optimization models into production management information systems. Such systems combine the functions of information gathering, production process modeling, and decision support [15]. The works emphasize that the use of information systems makes it possible to increase the accuracy of planning and reduce the influence of the human factor. However, universal software solutions often do not take into account the specific features of confectionery production, which reduces their effectiveness without additional adaptation [16].

An analysis of the literature shows that most of the existing resource allocation methods are based on static profit models that do not take into account the variability of technological parameters, seasonal fluctuations in demand, and the possibility of operational adjustments to the production program [17]. For confectionery enterprises, which are characterized by a wide range of products and limited resources, such simplification of models is a significant drawback.

Therefore, despite a significant number of scientific works in the field of production optimization, the issue of improving resource allocation methods, taking into account the peculiarities of the confectionery area and the possibilities of their practical implementation in information systems, remains insufficiently researched [18]. This necessitates the development of mathematical profit models that combine economic and technological

parameters, as well as the improvement of algorithms for solving optimization problems with a focus on real production conditions [19].

In view of the above, the research problem lies in improving the method of distributing production resources of a confectionery enterprise by developing a mathematical model and algorithmic support, which will increase the efficiency of resource use and create conditions for maximum profit under given production constraints.

3. Research objectives and tasks

The purpose of the study is to improve the method and mathematical model for optimizing the distribution of enterprise production resources, which ensures profit maximization under given production and technological conditions, resource constraints, and product range structure. The outlined goal is aimed at improving the effectiveness of management decisions in the process of production planning and resource utilization.

To achieve this goal, the following tasks must be performed: analyze current methods and approaches to resource allocation in the confectionery area; identify key factors affecting the efficiency of the use of material, energy, and production resources; improve the mathematical model of enterprise profit formation; determine the target optimization functions, taking into account economic and technological constraints; improve the method of solving the optimization problem to find the optimal parameters for resource allocation.

4. Materials and methods of the study

This study uses a combination of theoretical, formalized, and experimental methods aimed at substantiating and verifying the effectiveness of an improved method for optimizing the distribution of production resources at a confectionery enterprise. The methodological basis of the work is a systematic approach that allows us to consider the process of resource allocation as a complex multifactorial system with close interrelationships between economic and technological parameters [20].

System analysis methods were used to analyze the subject area and identify key factors affecting the efficiency of resource use. This made it possible to structure production processes, determine the role of raw material, energy, time, and capacity constraints,

and assess their impact on the formation of the enterprise's profit [21]. The identified dependencies were formalized using mathematical modeling methods, in particular, the construction of an analytical profit model that takes into account the cost of production, technological coefficients, resource consumption standards, and production constraints.

To solve the problem of profit maximization, optimization methods were applied, based on a discrete-analytical approach to finding acceptable alternatives [22]. This approach involves the discretization of variables, the generation of a set of possible production programs, the verification of each of them for compliance with the specified constraints, and the calculation of the value of the objective function. A comparative analysis was introduced to evaluate the effectiveness of the improved method in relation to traditional production planning approaches used in the confectionery area [23].

The developed algorithm for the distribution of production resources was implemented using software engineering methods [24]. Based on them, a software and algorithmic complex was created that automates the processes of preparing input indicators, forming a mathematical model, performing optimization calculations, and presenting the results in a form convenient for analysis. To verify the performance of the model and method, as well as to confirm the theoretical propositions, experimental studies were conducted on sets of realistic production indicators that simulate the operating conditions of a typical confectionery enterprise.

The experimental results made it possible to justify the feasibility of applying the improved method in production management practice and confirmed its effectiveness in increasing profitability and rational use of production resources.

5. Research results

The study analyzes modern methods of planning and optimizing the distribution of production resources used in food industry enterprises, particularly in the confectionery area. It has been established that the vast majority of enterprises use a traditional deterministic approach based on linear programming, normative planning, or expert assessment. This approach makes it possible to formally determine optimal production volumes, but it has a number of significant drawbacks. In particular, it does not take into account the dynamic nature of the production environment, seasonal

fluctuations in demand, changes in raw material prices, and the limited flexibility of technological processes.

Production management at enterprises involves ensuring maximum efficiency of resource use while maintaining stable product quality and regular output. To this end, a mathematical model has been developed

$$P = (b_1 - c_1) \cdot x_1 + (b_2 - c_2) \cdot x_2 + \dots + (b_i - c_i) \cdot x_i + \dots + (b_n - c_n) \cdot x_n, \quad (1)$$

where b_i – selling price per unit of the i -th type of product; c_i – production cost per unit of the i -th type of product; x_i – quantity of manufactured products.

The indicators determine the economic impact of each type of product on the total profit of the enterprise. The main idea is that each product brings the enterprise a certain marginal income $(b_i - c_i)$. The production of each unit of output consumes certain resources, which forms a system of constraints.

The main goal of optimization is to find such values x_i , under which profits are maximized and restrictions are not violated.

Marginal income is a key criterion that allows determining the feasibility of producing a certain type of product.

If:

$$(b_i - c_i) > 0, \quad (2)$$

production is profitable, and provided that

$$(b_i - c_i) < 0 \quad (3)$$

production causes losses.

Effective management of production processes in the confectionery area is impossible without purposeful consideration of external market environment factors.

$$\max P = (b_1 - c_1) \cdot x_1 \cdot k_1 \cdot s_1 + (b_2 - c_2) \cdot x_2 \cdot k_2 \cdot s_2 + \dots + (b_i - c_i) \cdot x_i \cdot k_i \cdot s_i + \dots + (b_n - c_n) \cdot x_n \cdot k_n \cdot s_n. \quad (4)$$

The mathematical optimization model should include a system of constraints that take into account the availability of resources, production capacity, and actual demand.

Constraints on the use of resources (raw materials, materials):

$$\sum_{i=1}^n a_{ij} \cdot x_i \leq R_j, \quad j = 1, 2, \dots, m, \quad (5)$$

where x_i – production volume of the i -th type of product, kg; a_{ij} – consumption of the j -th type of raw material per unit of the i -th type of product, kg/kg; R_j – available amount of j -th type raw material; n – number of product types; m – number of raw material types.

that makes it possible to determine the optimal output of different types of products, taking into account the limited resources of the enterprise, and to obtain profit from their sale [25].

The profit model is defined as the difference between the price of sales and the total cost of production:

One of the most influential factors is seasonality, which determines changes in purchasing activity and affects the demand for various groups of confectionery products. Therefore, the optimization model must include parameters that reproduce these patterns.

To improve the optimization model, two types of adjustment parameters have been introduced, namely: the sales coefficient k_i and seasonality coefficient is s_i .

Coefficient k_i – Sales coefficient, the share of production volume that is actually sold. It takes into account specific market conditions, consumer behavior, and the competitive advantages of the enterprise (brand recognition, market share, product quality, etc).

Seasonality coefficient s_i – adjusting demand depending on the season.

This helps to take into account changes in purchasing activity: $s_i > 1$ – season of increased demand (winter, holidays); $s_i = 1$ – average annual demand, $s_i < 1$ – season of reduced demand (summer, vacations).

Such coefficients enable the optimization model to respond adequately to real market dynamics.

Taking into account the implemented coefficients, the model of maximized profit of the enterprise is defined as

Production capacity rconstraints:

$$\sum_{i=1}^n t_i \cdot x_i \leq T, \quad (6)$$

where t_i – labor intensity of production of 1 kg of the i -th type of product, man-hours/kg; T – available working time fund, man-hours; x_i – volume of production of the i -th type of product, kg; n – number of product types; $i = 1, 2, \dots, n$ – product type index.

Demand constraints due to seasonality:

$$x_i \leq D_i \cdot k_i \cdot s_i, \quad i = 1, 2, \dots, n, \quad (7)$$

where D_i – projected demand for the i -th type of product, kg; k_i – sales coefficient; s_i – seasonality coefficient.

This helps to avoid overproduction during periods of low demand and maximize profits during peak seasons.

The optimization of production processes in the confectionery industry is aimed at achieving maximum economic effect through the rational use of resources, improving product quality, and reducing raw material costs. The main task is to determine the optimal ratio between the costs of raw materials, labor, and energy resources while ensuring the necessary level of production productivity.

Among the key methods of production optimization are classical linear programming methods, which allow determining the optimal output volumes with limited resources of raw materials, energy, and capacity.

In the practice of managing the production activities of enterprises in the food area, and in particular the confectionery area, the classical method of resource allocation optimization based on economic and mathematical modeling and linear programming methods has become widely used [25]. This approach is aimed at determining the production volumes that ensure maximum economic results under conditions of limited resources and compliance with technological standards. This approach makes it possible to formally determine optimal production volumes, but it has a number of significant drawbacks, in particular, it does not take into account the dynamic nature of the production environment, seasonal fluctuations in demand, volatility in raw material prices, and limited flexibility of technological processes.

The current method is based on the assumption of relative stability of production conditions, the deterministic nature of input indicators, and the possibility of formalizing production processes in the form of a system of linear dependencies.

The method is implemented as a sequence of interrelated stages.

Stage 1. Formation of an information base. At this stage, the following information is collected: product cost; cost of raw materials and energy; production standards; planned demand volumes; available production capacity. This information forms the basis for parameterizing the model.

Stage 2. Construction of the profit target function according to expression (4).

Stage 3. Formation of a system of constraints (for example, the amount of flour, cocoa products, sugar, fat components, equipment operating time, workshop capacity, labor resources, etc.).

Stage 4. Normalization of indicators and verification of correctness. The purpose of this stage is to identify illogical information, eliminate anomalies, and verify compliance with production standards.

Stage 5. Solving the optimization problem. To do this, use: the simplex method, the branch and bound method, gradient methods, weighted optimization methods (in the case of multi-objective problems).

Stage 6. Analysis of the results achieved. Based on the solution, the following are determined: optimal production volumes, expected profit, product range structure, and economic effects of applying the model.

Stage 7. Testing the model in production conditions.

The described method is basic and is widely implemented in production planning and resource management systems. It serves as a theoretical and methodological basis for further improvements that take into account dynamic factors, seasonal fluctuations in demand, and changes in resource constraints in the current conditions of confectionery enterprises.

This task can be accomplished using an improved method that allows for a quick response to changes in demand, resource base, and economic parameters, ensuring stable growth in the company's profitability. This method is based on the principle of profit maximization, taking into account relevant technological and market factors. To implement the improved method, a target function has been formed that integrates: the selling price of products, production cost, seasonal demand coefficient, demand adjustment coefficient, and production volume. The purpose of implementing the method is to find the following production volume values $x_1, x_2, \dots, x_i, \dots, x_n$, which maximize efficiency within the limits of available resources. An improved method for determining the rational allocation of resources in the production system of a confectionery enterprise is based on the use of an economic-mathematical model of profit maximization (P).

The method involves step-by-step execution of operations to process input parameters, form an acceptable solution space, calculate the optimality criterion, and select the best option for resource allocation.

Stage 1. Collection and formalization of necessary information. At the first stage, it is necessary to collect, verify, and structure all the necessary parameters. They form the initial basis for further calculations using the profit maximization formula. In a structured form, we obtain the following groups of parameters: economic indicators, technological standards,

resource constraints, and demand forecasts. All this information forms a parametric set that is used in the following stages.

Stage 2. Formation of a set of acceptable production volume options.

Step 1. Construction of a space of possible options. Based on the defined variables $x_1, x_2, \dots, x_i, \dots, x_n$, denoting the production volumes of each product type, the initial space of options is formed. This space can be specified in the following ways: by searching through discrete values of variables or by generating solutions according to a specific algorithm (grid search, stochastic generation, etc). This forms the initial set of possible combinations of production volumes.

Step 2. Checking combinations of options for compliance with the system of constraints. Each formed combination of quantities $x_1, x_2, \dots, x_i, \dots, x_n$. It is necessary to check for compliance with the set of model constraints. It is determined whether the available values exceed the demand for raw materials, labor resources, equipment capacity, and energy. It is important to take into account the conditions for compliance with recipes, ingredient proportions, the duration of technological operations, and the maximum permissible production volumes. It is necessary to control the compliance of budget expenditures with cost constraints and other economic parameters established by the model.

Step 3. Formation of a set of feasible solutions. Combinations that fully satisfy all constraints must be included in the set of feasible solutions.

This set contains all variants of the production program that do not violate resource limits, meet technological requirements, are economically feasible, and can be implemented at the enterprise.

Stage 3. Calculation of profit for each production option. For each feasible set of production volumes, it is necessary to calculate the value of the optimization criterion – the enterprise's profit. The following are taken into account in the calculation process: adjustment of the demand value using coefficients k_i ; seasonal fluctuations in demand, indicated by coefficients s_i ; the quantity of each type of product that can be manufactured within the specified limits.

Stage 4. If the solution meets all criteria, it is considered optimal and the implementation of the method is completed. If inconsistencies or instability are found, it is necessary to return to the previous stages to adjust the input parameters.

The optimal solution found should be implemented in specific recommendations for planning the production activities of the enterprise. In accordance with the selected set $x_1, x_2, \dots, x_i, \dots, x_n$. The following are formed: an optimal production plan that determines the volume of production of each type of product, the rational distribution of resources between technological operations and production lines, recommendations for equipment schedules, in particular with load adjustment, elimination of downtime, and optimization of the sequence of operations, automatic updating of production program parameters if there are changes in prices, costs, seasonal coefficients, or the volume of available resources.

Traditionally, the task of allocating production resources in food industry enterprises is solved using the linear programming method, which involves the formation of a profit objective function and a system of linear constraints, followed by a search for the optimal solution using the simplex method or its modifications. This approach is effective in the case of a linear model, continuous variables, and the absence of complex technological dependencies.

At the same time, in the practical conditions of confectionery production, discrete constraints, nonlinear technological relationships, recipe ratios, and combinatorial dependencies often arise, which significantly complicate the application of classical linear programming methods.

Unlike the basic approach, the improved optimization method is based on the formation of a set of acceptable production volume options with their subsequent selection according to the criterion of profit maximization. At the second stage of the method, a space of possible combinations of variables that determine the production volumes of individual types of products is constructed.

This space can be formed by discrete enumeration or algorithmic generation of options, which allows the integer nature of production decisions to be taken into account directly.

Each generated combination must be checked for compliance with resource, technological, and economic constraints. This approach ensures strict adherence to real production conditions along with restrictions on raw materials, equipment, recipes, and acceptable production volumes. Only those options that fully satisfy all constraints are included in the set of feasible solutions, from which the optimal one is selected. Thus, the key difference between the improved method and classical

linear programming is the transition from analytical search for extremes to discrete-analytical analysis of a set of feasible alternatives. This makes it possible to increase the flexibility of the model, ensure that complex technological constraints are correctly taken into account, and form economically sound production programs in the conditions of real confectionery production.

The improved method ensures greater accuracy in determining the rational distribution of resources; takes into account seasonal and market fluctuations in production; and increases planning flexibility thanks to dynamic coefficients s_i and k_i ; the ability to adapt the production program in case of changes in available resources or restrictions; the formation of economically sound decisions on the distribution of resources between competitive areas of production. The result of the method is an optimal production program that determines the rational and balanced distribution of resources between product types in order to achieve the highest possible target value. The method can be implemented as

1) a separate module (or subsystem) within the production management information system,

2) an element of the ERP system. The IS helps to automatically solve the optimization problem of resource allocation, providing the user with a ready-made set of management decisions.

This method and model enable the rapid updating of the production plan and adaptive response to market changes. The advantages of using the improved method for production are: reduction of total production costs; increased equipment utilization; and ensuring sustainable production planning taking into account seasonal factors. To demonstrate the modularity of the system and the clear division of responsibilities between its components, a class diagram of the resource allocation optimization task was developed (Fig. 1).

The UML class diagram is an element of modeling the architecture of the software and algorithmic complex that implements the method and model of resource allocation optimization within the enterprise production management information system.

The construction of a class diagram is necessary for several reasons. It reproduces the static structure of the software, i.e., it shows which entities (classes) are involved in the implementation of the method, what attributes they contain, and what operations they perform. This makes it possible to formalize the internal logic of the software-algorithmic complex (SAC) and to unambiguously interpret the structure of the program

code. A class diagram provides logical order to the components, which helps to structure, for example, a Java project in such a way that each module performs a clearly defined function: storage and processing of production parameters; implementation of a mathematical model; access to a database (DB); logic for selecting the optimal option; integration with other components of the enterprise IS.

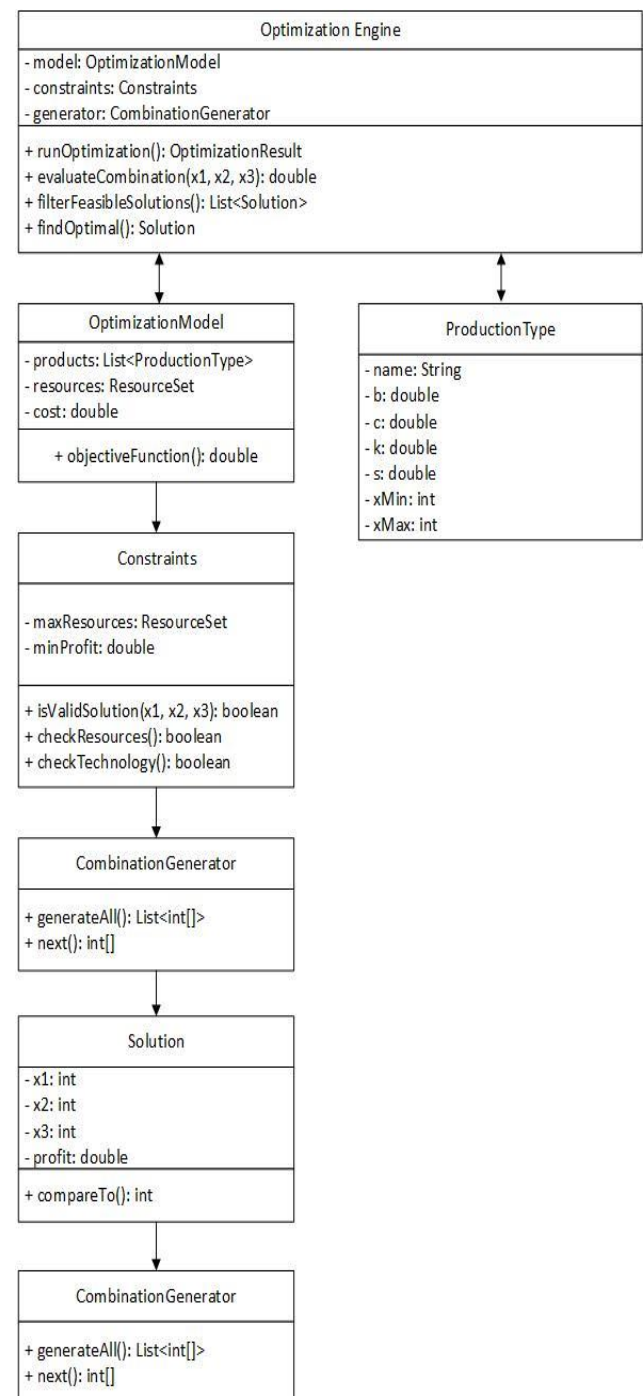


Fig. 1. Class diagram of the resource allocation optimization problem

The class diagram performs a number of important functions: it structures the program logic; formalizes the implementation of the method; defines the architecture of the modules; documents architectural decisions; serves as the basis for further implementation, maintenance, and development of the PAC; ensures transparency and reproducibility of the resource allocation optimization process.

This structure shows: dependencies between the model, constraints, and optimizer; separation of combination generation from evaluation logic; presence of an information access layer (*Data Access Object*, *DAO*) for storing parameters and results.

The proposed method of optimizing resource allocation at a confectionery enterprise is being implemented in practice through the integration of a mathematical model, algorithmic procedures, and software. The developed model and method for optimizing resource allocation take into account a combination of factors: the cost of production of individual types of products, the level of sales prices, the efficiency of resource use (coefficients and), the availability of raw materials and warehouse stocks, the throughput capacity of equipment, and planned production restrictions. This creates a universal tool that can be integrated into any enterprise management system, from traditional ERP to highly specialized MES solutions.

Fig. 2 shows a diagram of the application of the resource allocation optimization method in the production processes of a confectionery enterprise. The diagram demonstrates the step-by-step implementation of the method, which begins with the collection of actual production indicators.

These include: raw material consumption rates per unit of a particular product type; price parameters – purchase prices and projected sales prices; production constraints in the form of available equipment and labor; seasonal fluctuations in demand and production line utilization schedules. The information collection and validation stage ensures the accuracy of the initial parameters, which is a key prerequisite for the implementation of the model. After that, a resource matrix is formed – a centralized structured data set that contains available resources and possible options for their distribution among product types.

Table 1 presents a comparative description of the existing and improved methods of resource allocation.

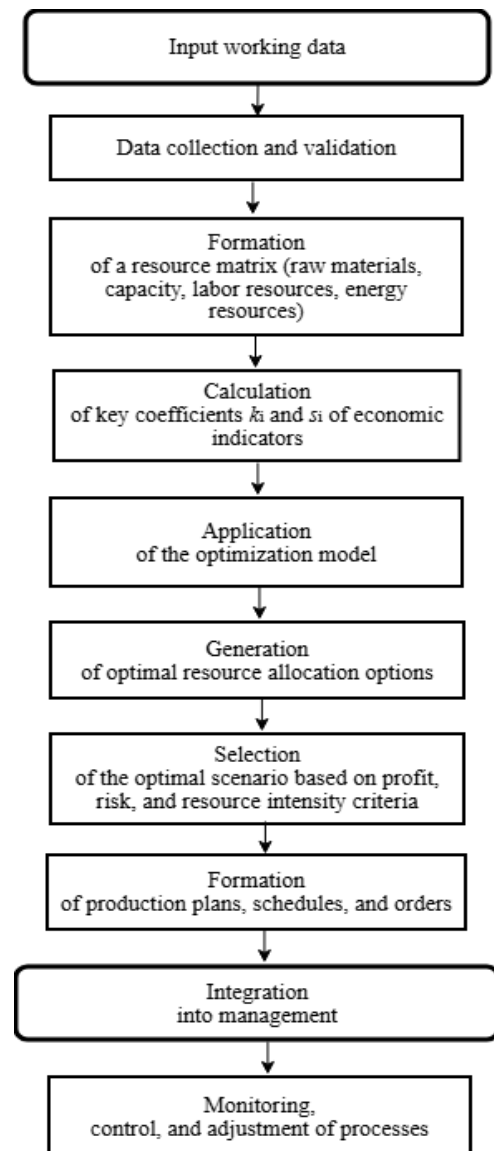


Fig. 2. Scheme for implementing the resource allocation method

A comparative analysis shows that the existing method is effective under stable production and economic conditions, but has limited flexibility and does not take market dynamics into account. The improved method eliminates these shortcomings through the use of adaptive parameters, expansion of the mathematical model, and improvement of the optimization algorithm, which increases the accuracy and practical applicability of the solutions obtained. To implement the improved method, a software module has been developed using the *Java* programming language, which ensures reliability, platform independence, an object-oriented approach, and the possibility of integration with existing modules of the enterprise information system. Listing 1 shows a fragment of *Java* code that implements the functional basis of the IS optimization module.

Table 1. Comparative characteristics of existing and improved methods of resource allocation

Comparison criterion	Existing (basic) method	Improved method
Optimization goal	profit maximization under fixed resource constraints	profit maximization taking into account dynamic economic and technological factors
Type of mathematical model	linear profit model	improved economic-mathematical model of profit with adaptive parameters
Type of input data	static, deterministic	partially dynamic taking into account seasonal and market fluctuations
Demand	fixed or average	variable, with the possibility of adjusting the production program
Resource constraints	rigidly defined, unchanging	flexible, with the possibility of adaptation in case of changing conditions
Technological coefficients	constant normative values	Dynamic coefficients
Seasonality	not taken into account or taken into account indirectly	and, reflecting real production conditions
Model scalability	limited with an increase in the number of variables	taken into account directly in the model parameters
Planning flexibility	low, requires a complete recalculation	ensures operation with a large number of interrelated parameters
Optimization result	optimal production volumes under given conditions	high, with the possibility of operational adjustment of plans
Practical application	limited by stable production conditions	optimal production program with adaptation to changes in the environment
Economic effect	moderate	effective in a changing market

Listing 1. Java code of the optimization module (*OptimizationModel.java*)

```
public class OptimizationModel {
    private List<ProductionType> products;
    private ResourceSet resources;

    public OptimizationModel(List<ProductionType> products, ResourceSet resources) {
        this.products = products;
        this.resources = resources;
    }

    public double calculateProfit(int x1, int x2, int x3) {
        return products.get(0).profit(x1) +
            products.get(1).profit(x2) +
            products.get(2).profit(x3);
    }

    public List<ProductionType> getProducts() {
        return products;
    }
}
```

The proposed *Java* code ensures: correct execution of the mathematical model; verification of technological and resource constraints; complete enumeration of permissible combinations; determination of the optimal resource allocation option. The proposed method for optimizing the distribution of production resources at a confectionery enterprise is implemented by combining a mathematical model, algorithmic procedures, and software implementation as part

of a production management information system. The improved method is integrated into a software and algorithmic complex that provides automated collection and processing of input information, formation of acceptable production alternatives, calculation of the target function, and selection of the optimal solution.

At the practical level, the method is implemented in the form of a separate optimization module that interacts with modules for collecting production indicators,

forming input parameters, and supporting management decision-making.

The input information for the optimization module is data on available resources (raw materials, production capacity, labor resources), economic indicators (cost price, sales prices, marginal income), as well as demand parameters taking into account seasonal and market coefficients.

This information is processed automatically, which minimizes the human factor and increases the accuracy of planning. To demonstrate the feasibility and practical significance of the method, we consider an example of its implementation for a confectionery enterprise that manufactures three types of products. Within the example, a space of possible production programs is

$$b_1 - c_1 = 40; k_1 = 1,10; s_1 = 0,90 \Rightarrow \text{multiplier } f_1 = 40 \cdot 1,10 \cdot 0,90 = 39,6 (\text{UAH/kg});$$

$$b_2 - c_2 = 40; k_1 = 0,95; s_1 = 0,85 \Rightarrow \text{multiplier } f_2 = 40 \cdot 0,95 \cdot 0,85 = 32,3 (\text{UAH/kg});$$

$$b_3 - c_3 = 50; k_3 = 1,20; s_3 = 1,00 \Rightarrow \text{multiplier } f_3 = 50 \cdot 1,20 \cdot 1,00 = 60,0 (\text{UAH/kg}).$$

Discretization step: $\Delta x = 500 \text{ kg}$ (technologically justified minimum batch). Target function for specific numbers:

$$P = 39,6 \cdot x_1 + 32,3 \cdot x_2 + 60 \cdot x_3, \quad (8)$$

Table 2. Examples of combinations x_1, x_2, x_3 and profit P calculations

№	(x_1) (кг)	(x_2) (кг)	(x_3) (кг)	P calculations (UAH)	P (UAH)
1	1000	800	600	$39,6 \cdot 1000 + 32,3 \cdot 800 + 60 \cdot 600$	101440
2	1500	800	600	$39,6 \cdot 1500 + 32,3 \cdot 800 + 60 \cdot 600$	121240
3	2000	600	400	$39,6 \cdot 2000 + 32,3 \cdot 600 + 60 \cdot 400$	122580
4	2500	0	500	$39,6 \cdot 2000 + 32,3 \cdot 600 + 60 \cdot 400$	129000
5	0	1800	1500	$39,6 \cdot 0 + 32,3 \cdot 1800 + 60 \cdot 1500$	148140
6	500	500	500	$39,6 \cdot 500 + 32,3 \cdot 500 + 60 \cdot 500$	65950
7	2500	1800	1500	$39,6 \cdot 500 + 32,3 \cdot 500 + 60 \cdot 500$	247140
8	2000	1000	1000	$39,6 \cdot 2500 + 32,3 \cdot 1800 + 60 \cdot 1500$	171500
9	1500	1800	0	$39,6 \cdot 2000 + 32,3 \cdot 1000 + 60 \cdot 1000$	117540
10	1000	0	1500	$39,6 \cdot 1000 + 32,3 \cdot 0 + 60 \cdot 1500$	129600

The right column shows the total profit P for each combination (rounded to UAH). Combination No. 7 (maximum values for each type) gives the maximum P in this sample (247,140 UAH), but before the final selection, it must be checked for compliance with all resource constraints (raw materials, total equipment operating time, logistics, etc.). If the combination violates one of the constraints, it must be filtered out.

Thus, the results achieved confirm the implementation of the developed complex as an element of a modern information system for managing confectionery production. Practical results show that

formed by discretely enumerating the production volumes of each type of product. All combinations are checked for compliance with resource, technological, and economic constraints, after which the acceptable options are evaluated according to the criterion of profit maximization. The result of the implementation is an optimal production program that determines the rational distribution of resources between product types, the level of equipment utilization, and the expected economic effect. The results can be used to quickly adjust production plans, adapt the product range to changes in demand, and make economically sound management decisions.

Units: x_i – kg (in the table).

Accepted coefficients:

where P – profit, UAH; x_i – volumes in kg.

Table 2 shows examples of combinations x_1, x_2, x_3 and profit P calculations.

the use of the improved method makes it possible to increase the accuracy of determining the rational distribution of resources, reduce the share of inefficiently used resources, and increase the value of the objective function compared to basic planning methods. On average, the increase in profit as a result of optimization ranged from 5 to 12%, depending on the scenario and product range structure. A particularly significant effect was achieved in conditions of resource scarcity or unstable demand, which confirms the practical feasibility of implementing the improved approach.

The results achieved confirm that the improved method is an effective tool for supporting management decisions in the field of production planning. Its application contributes to increasing the economic stability of the enterprise and reducing the risks associated with the irrational use of resources.

6. Discussion of research results

The results achieved demonstrate the feasibility and effectiveness of the improved method and mathematical model for optimizing the distribution of production resources. Unlike most previous studies in this area, which are dominated by static linear programming models or normative-planning approaches, this work implements a more flexible discrete-analytical approach focused on the real conditions of production systems.

A distinctive feature of the results achieved is the comprehensive consideration of both economic and technological factors within a single objective function. Previous studies have tended to focus either on economic indicators (profit, cost) or on technological constraints, which has led to a simplification of the model and a reduction in the accuracy of the results. The improved model proposed in this work makes it possible to combine these aspects and ensure a more adequate reproduction of the real production process. Compared to existing planning methods, the results of implementing the improved method demonstrate a number of significant advantages. First, it ensures greater accuracy in determining the rational allocation of resources by reviewing acceptable alternatives and rejecting ineffective options at the early stages of the method. Second, the use of dynamic coefficients makes it possible to take into account seasonal and market fluctuations, which is practically not implemented in traditional approaches. Third, the method provides high planning flexibility and the ability to quickly adapt the production program in case of changes in constraints or available resources.

An important advantage of the results achieved is their practical orientation. Unlike purely theoretical models, the improved method is implemented in the form of a software and algorithmic complex, which creates the prerequisites for its practical implementation in production management information systems. At the same time, the results achieved have certain limitations that must be taken into account in practical application. In particular, the discrete-analytical approach to

generating alternatives can lead to an increase in the computational complexity of the task if there is a significant increase in the number of product types or model parameters. This limits the application of the method to very large production systems without additional procedures for reducing the search space or parallel computations.

In addition, the accuracy of the results largely depends on the quality of the input information. Inaccurate or outdated information on cost, equipment productivity, or raw material availability can negatively affect the adequacy of the decisions made. Therefore, the effective application of the method requires a well-established system for collecting and updating production and economic information.

7. Conclusions and directions for further research

The study analyzed modern approaches to the allocation of production resources in the confectionery area and identified their limitations related to the use of static models and insufficient attention to the dynamics of production conditions. Key factors affecting the efficiency of resource use were identified, namely: the structure of production processes, equipment limitations, availability of raw materials, and economic properties of products.

The mathematical model of enterprise profit was improved by detailing the relationships between technological parameters and economic indicators. The objective function of optimization has been formulated taking into account production, resource, and time constraints. The method of solving the optimization problem has been improved, ensuring the search for a rational distribution of resources based on the generation and evaluation of acceptable alternatives.

During the testing of the improved method on the example of a typical enterprise in the confectionery area, its effectiveness and economic feasibility were confirmed. In general, the set research goal has been achieved, and its results can be used to support management decision-making in production information systems.

Prospects for further research on this issue are related to expanding the functionality of the proposed method and model. In particular, it is advisable to include stochastic factors, which will allow for the uncertainty of demand and raw material prices to be taken into

account. Another promising direction is the integration of the improved method with ERP and MES systems to ensure automated planning in real time.

In further research, it is advisable to pay special attention to the application of intelligent methods, in particular elements of machine learning, for forecasting the input parameters of the model.

The implementation of these areas will contribute to increasing the practical value of the results achieved and expanding their scope of application in the context of the digital transformation of food industry enterprises.

Conflict of interest

The authors declare that they have no conflict of interest with respect to this research, including

References

1. Beamon, B. (2021), "Supply chain design and analysis: Models and methods", *International Journal of Production Economics*, Vol. 240, P. 108–118. DOI: <https://doi.org/10.1016/j.ijpe.2021.108118>
2. Heizer, J., Render, B., Munson, C. (2022), *Operations management: Sustainability and supply chain management*, 14th ed., Pearson Education, Harlow, 912 p.
3. Slack, N., Brandon-Jones, A. (2023), *Operations and process management: Principles and practice for strategic impact*, 6th ed., Pearson Education, Harlow, 816 p.
4. Nahmias, S., Olsen, T. (2022), *Production and operations analysis*, 8th ed., Waveland Press, Long Grove, 854 p.
5. Beamon, B., Chen, V. (2022), "Hybrid optimization approaches in production planning", *Production Planning & Control*, Vol. 33, No. 4, P. 310–324. DOI: <https://doi.org/10.1080/09537287.2021.1883684>
6. Tarver, M. (2023), *Process optimization in food manufacturing*, Elsevier, Oxford, 472 p.
7. Silver, E., Pyke, D., Thomas, D. J. (2021), *Inventory and production management in supply chains*, 4th ed., CRC Press, Boca Raton, 794 p.
8. Arteaga-Cabrera, E., Ramírez-Márquez, C., Segovia-Hernández, J., Gómez-Salazar, J. (2025), "Advancing optimization strategies in the food industry: From traditional approaches to multi-objective solutions", *Applied Sciences*, Vol. 15, No. 7, Art. 3846. DOI: <https://doi.org/10.3390/app15073846>
9. Stadler, H. (2015), "Supply chain management: An overview", in: *Supply Chain Management and Advanced Planning*, Springer, Berlin, Heidelberg, P. 3–28. DOI: https://doi.org/10.1007/978-3-642-55309-7_1
10. Dolgui, A., Proth, J.-M. (2016), *Supply chain engineering: Useful methods and techniques*, Springer International Publishing, 540 p. DOI: <https://doi.org/10.1007/978-3-319-27798-0>
11. Stevenson, M., Spring, M. (2017), "Flexibility from a supply chain perspective: Definition and review", *International Journal of Operations & Production Management*, Vol. 37, No. 6, P. 682–713. DOI: <https://doi.org/10.1108/IJOPM-01-2016-0020>
12. Govindan, K., Fattahi, M., Keyvanshokoo, E. (2017), "Supply chain network design under uncertainty: A comprehensive review and future research directions", *European Journal of Operational Research*, Vol. 263, No. 1, P. 108–141. DOI: <https://doi.org/10.1016/j.ejor.2017.04.009>
13. Ivanov, D., Dolgui, A., Sokolov, B. (2019), "The impact of digital technology and Industry 4.0 on the ripple effect and supply chain risk analytics", *International Journal of Production Research*, Vol. 57, No. 3, P. 829–846. DOI: <https://doi.org/10.1080/00207543.2018.1488086>
14. Ghiani, G., Laporte, G., Musmanno, R. (2019), *Introduction to logistics systems management*, 2nd ed., Wiley, 480 p. DOI: <https://doi.org/10.1002/9781119384960>
15. Fandel, G., Giese, A., Raub, M. (2018), "Cost-oriented aggregation planning in food production", *OR Spectrum*, Vol. 40, No. 4, P. 927–955. DOI: <https://doi.org/10.1007/s00291-018-0517-5>

financial, personal, authorship, or other conflicts that could influence the research and its results presented in this article.

Funding

The study was conducted without financial support.

Data availability

The manuscript has no associated data.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence technology in the creation of this work.

16. Akkerman, R., Farahani, P., Grunow, M. (2016), "Quality, safety and sustainability in food distribution: A review of quantitative operations management approaches", *International Journal of Production Economics*, Vol. 152, P. 1–15. DOI: <https://doi.org/10.1016/j.ijpe.2016.07.009>
17. Zhang, Y., Wang, L., Li, X. (2020), "Optimization of production planning in food manufacturing considering multiple resource constraints", *Journal of Cleaner Production*, Vol. 256, Art. 120345. DOI: <https://doi.org/10.1016/j.jclepro.2020.120345>
18. Ivanov, D. (2021), "Viable supply chain model: Integrating agility, resilience and sustainability perspectives – lessons from and thinking beyond the COVID-19 pandemic", *Annals of Operations Research*, Vol. 303, P. 1–21. DOI: <https://doi.org/10.1007/s10479-020-03640-6>
19. Beamon, B., Balcik, B. (2018), "Performance measurement in humanitarian relief chains", *International Journal of Public Sector Management*, Vol. 31, No. 2, P. 159–174. DOI: <https://doi.org/10.1108/IJPSM-05-2017-0132>
20. Pinedo, M. (2022), *Scheduling: Theory, algorithms, and systems*, 6th ed., Springer, 720 p. DOI: <https://doi.org/10.1007/978-3-030-85408-4>
21. Kühnle, H., Bauernhansl, T. (2021), *Digital transformation in manufacturing*, Springer Nature, 410 p. DOI: <https://doi.org/10.1007/978-3-030-43616-7>
22. Trienekens, J., Wognum, P., Beulens, A., Van der Vorst, J. (2018), "Transparency in complex dynamic food supply chains", *Advanced Engineering Informatics*, Vol. 35, P. 65–76. DOI: <https://doi.org/10.1016/j.aei.2018.01.007>
23. Дмитревський, Д.В., Левикін, В.М. (2025), "Розробка інформаційної системи управління підприємством з виробництва кондитерських виробів", 29-й Міжнародний молодіжний форум «Радіоелектроніка та молодь у XXI столітті». Зб. матеріалів форуму. Т. 6., Ч. I. Харків: ХНУРЕ. С. 125.
24. Hill, A., Hill, T. (2017), *Essential operations management*, 2nd ed., Red Globe Press, London, 422 p.
25. Starr, M. K., Gupta, S. K. (eds.) (2017), *The Routledge Companion to Production and Operations Management*, Routledge, London, 676 p.

Received (Надійшла) 06.01.2026

Accepted for publication (Прийнята до друку) 28.01.2026

Publication date (Дата публікації) 12.03.2026

Відомості про авторів / About the Authors

Левикін Віктор Макарович – доктор технічних наук, професор, Харківський національний університет радіоелектроніки, професор кафедри інформаційних управляючих систем; Харків, Україна;
Viktor Levykin – Doctor of Technical Sciences, Professor, Kharkiv National University of Radio Electronics, Professor at the Department of Information and Control Systems; Kharkiv, Ukraine;
 e-mail: viktor.levykin@nure.ua
 ORCID ID: <https://orcid.org/0000-0002-7929-515X>

Дмитревський Дмитро В'ячеславович – Харківський національний університет радіоелектроніки, здобувач вищої освіти; Харків, Україна;
Dmytro Dmytrevskiy – Kharkiv National University of Radio Electronics, Master's Student; Kharkiv, Ukraine;
 e-mail: dmytro.dmytrevskiy@nure.ua
 ORCID ID: <https://orcid.org/0000-0003-1330-7514>

МЕТОД РОЗПОДІЛУ ВИРОБНИЧИХ РЕСУРСІВ У КОНДИТЕРСЬКІЙ ПРОМИСЛОВОСТІ

Об'єктом дослідження є процеси планування виробництва й використання промислових ресурсів у кондитерській діяльності в умовах обмежених виробничо-технологічних можливостей і варіативності асортименту продукції.
Предмет вивчення – методи й математичні моделі оптимізації розподілу матеріальних, енергетичних і виробничих ресурсів підприємства з метою підвищення ефективності його функціонування. **Метою дослідження** є вдосконалення методу й математичної моделі оптимізації розподілу виробничих ресурсів підприємства, що забезпечує максимізацію прибутку відповідно до заданих виробничо-технологічних умов, обмежень на ресурси й структури асортименту продукції,

а також підвищує ефективність управлінських рішень у процесі планування виробництва. Для досягнення окресленої мети в роботі визначено такі **завдання**: аналіз сучасних методів і підходів до розподілу ресурсів у кондитерській галузі; визначення ключових факторів, що впливають на ефективність використання матеріальних, енергетичних і виробничих ресурсів; удосконалення математичної моделі формування прибутку підприємства; побудова цільової функції оптимізації з огляду на економічні й технологічні обмеження; удосконалення методу розв'язання задачі оптимізації для знаходження раціональних параметрів розподілу ресурсів. У дослідженні застосовано такі **методи**: системний аналіз і узагальнення наукових джерел; лінійного й нелінійного програмування; економіко-математичне моделювання; дискретно-аналітичні методи оптимізації; алгоритмізації та програмної реалізації. Основні **результати** дослідження полягають в удосконаленні математичної моделі оптимізації прибутку способом урахування взаємозв'язків між технологічними параметрами виробництва, ресурсними обмеженнями й структурою асортименту продукції. Сформульовано цільову функцію оптимізації та вдосконалено метод її розв'язання, який забезпечує формування, оцінювання й вибір оптимального варіанта розподілу ресурсів. Запропонований підхід реалізовано у вигляді програмно-алгоритмічного інструментарію для застосування в інформаційних системах управління виробництвом. **Висновки**: впровадження вдосконаленого методу сприяє підвищенню ефективності планування виробництва, раціоналізації використання ресурсів і зростанню прибутковості кондитерського підприємства, що підтверджує його практичну доцільність та ефективність.

Ключові слова: оптимізація; ресурси; виробництво; прибуток; моделювання; алгоритм; планування; автоматизація; система; підприємство.

Бібліографічні описи / Bibliographic descriptions

Левикін В. М., Дмитревський Д. В. Метод розподілу виробничих ресурсів у кондитерській промисловості. *Автоматизовані системи управління та прилади автоматики*. 2026. № 1 (188). С. 163–176. DOI: <https://doi.org/10.30837/0135-1710.2026.188.163>

Levykin, V., Dmytrevskiy, D. (2026), "Improving the method of optimizing the distribution of production resources in confectionery production", *Management Information System and Devices*, No. 1 (188), P. 163–176. DOI: <https://doi.org/10.30837/0135-1710.2026.188.163>

Пащенко Р. Е.

АВТОМАТИЗОВАНЕ ВИЯВЛЕННЯ АНОМАЛІЙ НА КОСМІЧНИХ ЗНІМКАХ СІЛЬСЬКОГОСПОДАРСЬКИХ ЗЕМЕЛЬ ІЗ ВИКОРИСТАННЯМ ФРАКТАЛЬНОГО АНАЛІЗУ

Сучасні методи космічного моніторингу сільськогосподарських земель із використанням вегетаційних індексів не дають змоги оцінювати аномалії на космічних знімках і автоматизувати цей процес. **Предметом дослідження** є оцінювання можливості виявлення аномалій на космічних світлинах із застосуванням "пірамідального" фрактального аналізу. **Об'єктом вивчення** є космічні знімки супутника Sentinel-2 сільськогосподарських земель з аномаліями й без них. **Мета роботи** – дослідження методу автоматизованого виявлення аномалій на космічних знімках сільськогосподарських земель із використанням фрактального аналізу. **Досягнуті результати.** Установлено, що візуалізація поля фрактальних розмірностей космічного знімка стає неефективною за умови використання "вікна" великих розмірів. Крім того, застосування "стрибаючого вікна" не дає змоги визначити аномальні ділянки на космічних зображеннях. Запропоновано метод виявлення аномалій на космічних знімках сільськогосподарських земель з упровадженням "пірамідального" фрактального аналізу, що сприяє автоматизації цього процесу. "Піраміди" космічних знімків створено за допомогою розподілу вихідного космічного зображення на знімки менших розмірів, які надалі індексуються. На кожному рівні "піраміди" розраховано фрактальні розмірності космічних знімків і побудовано "піраміду" фрактальних розмірностей, що еквівалентно застосуванню "стрибаючого вікна" під час побудови поля фрактальних розмірностей. **Висновки.** Дослідження розробленого методу виявлення аномалій продемонстрували, що якщо фрактальні розмірності космічних знімків усіх рівнів "піраміди" перевищують фрактальну розмірність $D = 2,9$ або фрактальні розмірності на нижньому рівні "піраміди" більші, ніж на верхньому рівні, то аномалії на космічному знімку відсутні. Якщо фрактальні розмірності на окремих ділянках нижнього рівня "піраміди" менші за фрактальні розмірності на верхньому рівні й на вихідному космічному знімку, то на цих ділянках знімка спостерігаються аномалії.

Ключові слова: моніторинг сільськогосподарських земель; аномалія; космічний знімок; фрактальний аналіз.

Вступ

Сучасні засоби дистанційного зондування Землі (ДЗЗ) дають змогу отримувати інформацію про стан і структуру сільськогосподарських угідь, а також виявляти ділянки, що мають аномалії, зокрема ерозійні порушення та інші види деградації землі [1]. Вплив природних, техногенних і антропогенних факторів на посівні площі також оцінюється за допомогою методів агроекологічного моніторингу [2]. Сучасні супутники ДЗЗ допомагають регулярно отримувати космічні знімки на заданій території (один знімок за 5–8 діб). Такі зображення відбивають реальний стан земної поверхні з великою просторовою роздільною здатністю (10–30 м), і отримана інформація надходить від відкаліброваних датчиків у різних діапазонах хвиль. Наприклад, такими супутниками ДЗЗ є Sentinel-2 [3], космічні знімки яких надаються користувачам у вільному доступі через мережу Інтернет. Світлина Sentinel-2 мають високу просторову роздільну здатність (в одному пікселі 10 м) у 13 спектральних діапазонах, тобто є мультиспектральними. Космічні знімки

супутника Sentinel-2 можна отримувати двічі на тиждень на заданій території.

Регулярний космічний моніторинг земель сільськогосподарського призначення найчастіше проводиться з упровадженням методів, основаних на використанні вегетаційних індексів [4]. Різноманітні вегетаційні індекси беруть до уваги різну відбивну здатність рослин, а також фізичні властивості світла та інші властивості. Найбільш часто застосовують нормалізований різницевий вегетаційний індекс (NDVI). Вегетаційні індекси визначаються емпірично, і їх застосування не дає змоги оцінювати аномалії на космічних знімках, а також автоматизувати цей процес.

Аналіз останніх досліджень і публікацій

У роботі [5] зазначено, що для розрахунку індексу NDVI використовуються показники двох спектральних каналів – ближнього інфрачервоного й червоного. Тобто для отримання інформації цих каналів необхідна апаратура багатоспектральної зйомки каналів, що ускладнює отримання вихідного

зображення. Крім того, у роботі наголошено, що доцільність методів моніторингу сільськогосподарських земель, оснований на використанні індексу NDVI, значно залежить від кількості активної біомаси, і ці методи не ефективні, коли рослинність має низький рівень вегетації. У дослідженні [6] наведено результати оцінювання можливості моніторингу сільськогосподарських земель з використанням космічних знімків із супутника *Sentinel-2* і розрахунку індексів NDVI. У роботі зазначено, що за допомогою індексів NDVI можна зробити висновок про перехід сільськогосподарської культури з однієї фази розвитку в іншу, але не досліджено доцільність виявлення аномалій на космічних знімках.

У студіях [7, 8] подано результати оцінювання врожайності й стану ґрунту та підстильної поверхні з використанням широкосмугових показників ДЗЗ і продемонстровано, що для оброблення інформації гіперспектральної зйомки необхідні спеціальні алгоритми аналізу зображень, отриманих за допомогою безпілотних літальних апаратів і штучних супутників Землі. Крім цього, широко застосовуються нейронні мережі. Але впровадження сучасних алгоритмів отримання й оброблення гіперспектральних знімків призводить до збільшення складності та вартості апаратури, а також потребують тривалого часу для їх реалізації та ускладнення автоматизації.

Методи фрактального аналізу цифрових зображень [9, 10] дають змогу кількісно оцінити складність поверхонь і рослинності на космічних знімках, отриманих тільки в одному діапазоні хвиль. У роботі [11] продемонстровано можливість розрізнення різних типів ґрунтів із використанням фрактальних розмірностей, розрахованих для показників дистанційного зондування, але не розглянуто автоматизацію цього процесу. Стан різних сільськогосподарських культур посівів кукурудзи на різних фазах вегетації із застосуванням фрактального аналізу космічних знімків *Sentinel-2* досліджено в праці [12]. Крім того, продемонстровано, що за величиною середніх фрактальних розмірностей можна розподілити поля, засіяні гречкою та кукурудзою, і поля соняшнику, пшениці та ячменю, але в роботі не проаналізовано можливість виявлення на космічних знімках аномалій, викликаних впливом різних чинників.

Для визначення аномалій на космічних світлинах сільськогосподарських земель доцільно

розробити метод їх фрактального аналізу, що дасть змогу автоматизувати цей процес.

Мета статті – дослідити метод автоматизованого виявлення аномалій на космічних знімках сільськогосподарських земель за допомогою фрактального аналізу.

Матеріали й методи виявлення аномалій з використанням фрактального аналізу

Оцінити складність структури космічного знімка (заповнення тривимірного подання зображення) можна за допомогою розрахунку фрактальної розмірності, яка є дробовою величиною. Вищому ступеню заповнення тривимірного подання знімка відповідає більша величина фрактальної розмірності. Порядок обчислення фрактальних розмірностей з упровадженням методів покриття й призми, які найчастіше застосовуються на практиці, та побудови поля фрактальних розмірностей цифрових зображень подано в роботах [10, 13].

Розрахунок фрактальної розмірності методом покриття [14] забезпечує геометричне обчислення розмірності Гаусдорфа–Безіковича. У цьому разі тривимірне подання космічного знімка покривається кубами із стороною заданої довжини ε , а потім визначається сумарна кількість кубів $N(\varepsilon)$, використана для покриття зображення. Така процедура повторюється для декількох розмірів сторони куба ε і будується графік логарифмічної залежності $\log N(\varepsilon)$ від $\log 1/\varepsilon$. Для побудови апроксимованого графіка впроваджується метод найменших квадратів і після побудови аналізуються кути їх нахилу. Тангенс кута нахилу отриманої прямої є числовим значенням фрактальної розмірності D .

Розрахунок фрактальної розмірності методом призми [15] забезпечує визначення розмірності за допомогою обчислення площини верхньої грані призми. У цьому разі за показниками яскравості зображення обирається довжина основи призми розміром ε і формується верхня грань призми, а потім обчислюється її площа $P(\varepsilon)$. Така процедура повторюється для декількох розмірів довжини основи призми ε і будується графік логарифмічної залежності $\log P(\varepsilon)$ від $\log 1/\varepsilon$. Надалі фрактальна

розмірність D визначається так само, як і в методі покриття.

Під час фрактального аналізу космічних (цифрових) знімків доцільно застосовувати поле фрактальних розмірностей (ПФР), процес побудови якого детально розглянуто в роботі [10]. Основними параметрами, які використовуються під час побудови ПФР, є розмір "вікна", яким сканується зображення, і крок переміщення "вікна" s . "Вікно" називають ковзним, якщо $s=1$, і стрибаючим, якщо $s>1$. Розмір ПФР буде меншим за розмір космічного знімка на величину "вікна". Побудоване ПФР космічного зображення можна розподілити на декілька рівнів за величиною фрактальної розмірності, що дає змогу його візуалізувати. Кожен рівень (діапазон змін фрактальних розмірностей) відтворюється відповідним кольором. "Вікна" з близькими значеннями фрактальних розмірностей можуть бути об'єднані в окремі ділянки.

У роботі [13] визначено, що розрахунок фрактальних розмірностей і візуалізація ПФР можуть застосовуватися для сегментації сільськогосподарських земель на космічних знімках супутника *Sentinel-2*. Продемонстровано, що візуалізація ПФР космічних знімків дає змогу наочно показати зміну стану сільськогосподарських угідь і межі аномалій на знімках. Але в цьому дослідженні не проаналізовано, як розмір "вікна" й величина кроку його переміщення впливають на якість побудови ПФР і його візуалізацію. Проведемо такий аналіз з використанням космічних знімків каналу b8 супутника *Sentinel-2* розміром 56×56 пікселів.

На рис. 1 подано космічні знімки супутника *Sentinel-2* поля з пшеницею без аномалій розміром 56×56 пікселів (а), з аномаліями розмірами 56×56 (б) і 28×28 пікселів (в). Космічні знімки розміром 28×28 пікселів (див. рис. 1, в) є чотирма частинами космічного знімка з аномаліями розміром 56×56 пікселів (див. рис. 1, б).

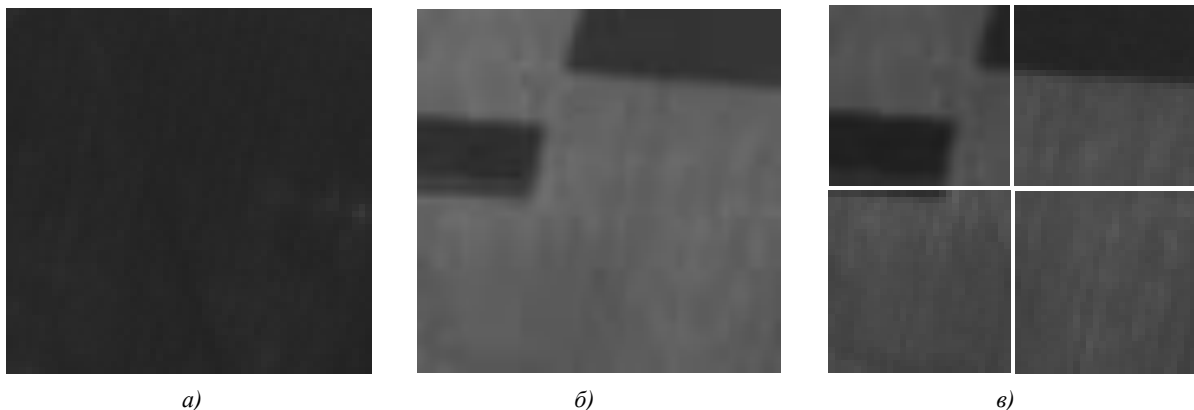


Рис. 1. Космічні знімки супутника *Sentinel-2* (канал b8) поля з пшеницею без аномалій розміром 56×56 пікселів (а), з аномаліями розмірами 56×56 (б) і 28×28 пікселів (в)

Побудуємо ПФР космічних знімків розмірами 56×56 і 28×28 пікселів з використанням методу призми й "ковзного вікна", а також виконаємо їх візуалізацію.

На рис. 2 наведено візуалізацію ПФР космічних знімків (поданих на рис. 1, б, в) з використанням "вікна" розміром 5×5 пікселів з кроком переміщення 1 піксель; розміри ПФР становлять 51×51 (а) і 23×23 пікселі (б). Кількість діапазонів зміни фрактальних розмірностей, на які розділялося ПФР (кількість рівнів візуалізації ПФР), дорівнювала п'яти.

Як видно на рис. 2, за допомогою візуалізації ПФР можна виділити частини космічного знімка

з різною структурою та визначити на них межі аномальних ділянок, але неможливо автоматизувати процес визначення аномалій.

Порівняння візуального подання ПФР на рис. 2, а і б демонструє, що розміри космічних знімків за однакових параметрів побудови ПФР не впливають на якість візуалізації ПФР. На ПФР космічних знімків різних розмірів одного й того самого поля добре визначаються межі аномалій. Отже, візуалізація ПФР космічного знімка дає змогу оцінити ділянки поля з різними значеннями фрактальних розмірностей, які визначають різний тип підстильної поверхні.

Розглянемо, як розмір "вікна" впливає на якість побудови ПФР і його візуалізацію.

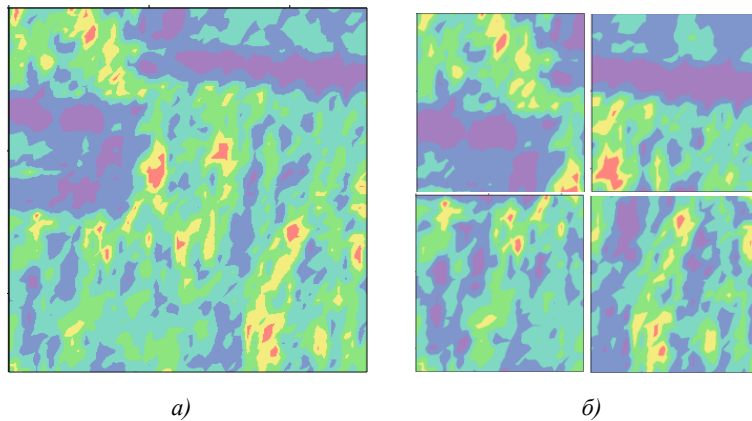


Рис. 2. Візуалізація ПФР під час використання "вікна" розміром 5×5 пікселів з кроком переміщення 1 піксель розміром 51×51 піксель (а) і чотири розміром 23×23 пікселі (б)

На рис. 3 наведено візуалізацію ПФР космічного знімка (рис. 1, б), побудованого з використанням "вікна" розміром 5×5 пікселів (а), "вікна" розміром

9×9 (б) і "вікна" розміром 17×17 пікселів (в) з кроком переміщення 1 піксель. Кількість рівнів візуалізації ПФР дорівнювала п'яти.

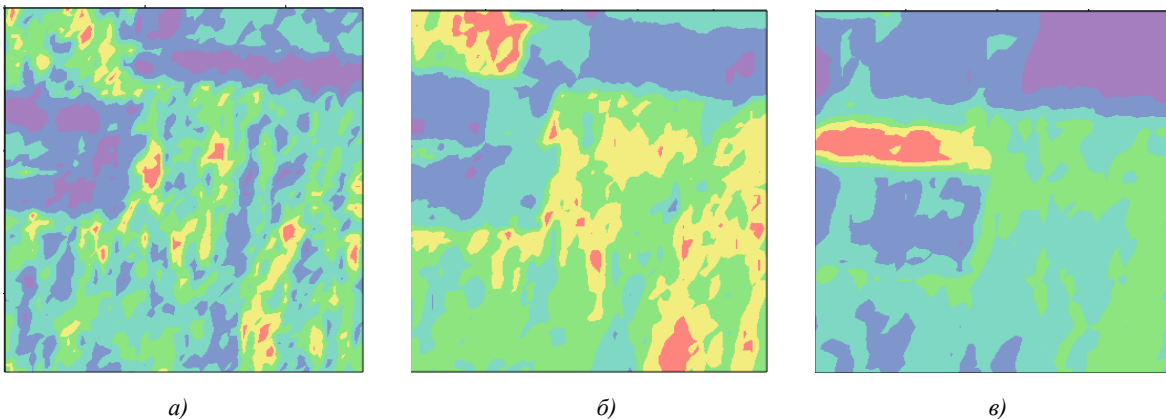


Рис. 3. Візуалізація ПФР під час використання "вікна" розміром 5×5 пікселів (а), "вікна" розміром 9×9 (б), "вікна" розміром 17×17 пікселів (в) з кроком переміщення 1 піксель

Як видно на рис. 3, в, збільшення розмірів "вікна" спричиняє згладжування візуального подання ПФР, що погіршує визначення меж аномальних ділянок на космічному знімку. Поряд з цим за умови незначних розмірів "вікна" (див. рис. 3, а) візуальне подання ПФР має велику кількість дрібних елементів, що ускладнює його аналіз і унеможливорює прийняття рішення щодо того, яка ділянка є аномальною. Застосування "вікна" середніх розмірів (див. рис. 3, б) дає змогу виокремити аномальні ділянки на космічній світлинні (синій колір), але на частині знімка, на якому немає аномалій, також виокремлюються особливості (червоний колір), що можна прийняти за аномальні.

Розглянемо як величина кроку переміщення "вікна" впливає на якість побудови ПФР та його візуалізацію.

На рис. 4 зображено візуалізацію ПФР космічного знімка (рис. 1, б), побудованого з використанням "вікна" розміром 5×5 пікселів із кроком переміщення 5 пікселів (а), "вікна" розміром 9×9 пікселів із кроком переміщення 9 пікселів (б) і "вікна" розміром 17×17 пікселів із кроком переміщення 17 пікселів (в). Кількість рівнів візуалізації ПФР дорівнювала п'яти.

Порівняння візуального подання ПФР на рис. 3 і 4 демонструє, що величина кроку переміщення "вікна" за інших однакових параметрів побудови ПФР значно впливає на якість візуалізації ПФР. Так, під час використання "стрибаючого вікна" візуальне подання ПФР стає неефективним для визначення аномалій на космічному знімку. У разі незначного розміру "вікна" (5×5 пікселів, рис. 4, а)

спостерігається згладжування ПФР, але дві аномальні ділянки позначаються синім кольором. За умови збільшення розмірів "вікна" (9×9 пікселів, рис. 4, б)

до розмірів 17×17 пікселів (рис. 4, в) узагалі неможливо проводити аналіз під час визначення аномальних ділянок на космічному знімку.

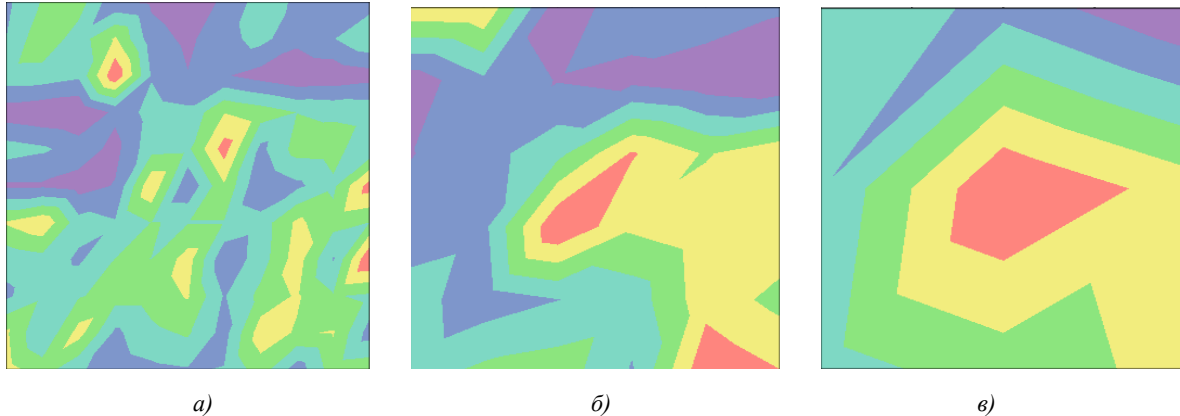


Рис. 4. Візуалізація ПФР за умови використання "вікна" розміром 5×5 пікселів із кроком переміщення 5 пікселів (а), "вікна" розміром 9×9 пікселів із кроком переміщення 9 пікселів (б), "вікна" розміром 17×17 пікселів із кроком переміщення 17 пікселів (в)

Отже, аналіз продемонстрував, що сучасний стан виявлення аномалій на космічних знімках за допомогою побудови ПФР дає змогу визначити особливості підстильної поверхні. Але візуалізація ПФР космічного знімка стає неефективною в умовах використання "вікна" великих розмірів, а також використання "стрибаючого вікна" не дає змоги виявити аномальні ділянки на космічних знімках. Тобто під час візуалізації ПФР необхідно обирати параметри побудови ПФР залежно від типу поверхні, зображеної на космічних світлинах, і автоматизувати процес виявлення аномалій стає неможливим.

Мета статті – розглянути можливість автоматизованого визначення аномалій на космічних знімках сільськогосподарських земель із застосуванням фрактального аналізу.

Метод виявлення аномалій з використанням "пірамідального" фрактального аналізу

Дослідження дають змогу запропонувати метод виявлення аномалій на космічних знімках сільськогосподарських земель із застосуванням "пірамідального" фрактального аналізу (структурну схему методу зображено на рис. 5).

На рис. 5 видно, що на першому етапі можна отримати космічні знімки з мережі Інтернет і визначити об'єкти дослідження – сільськогосподарські землі.

На другому етапі здійснюється вибір полів, які необхідно дослідити і які визначені замовником. На космічному знімку вони встановлюються візуально. На цьому етапі також можна визначити розмір зображення, наприклад, 56×56 пікселів, і після цього вирізати зображення полів заданого розміру.

На третьому етапі створюється "піраміда" космічних знімків та індексуються отримані зображення. Тобто, наприклад, із зображення розміром 56×56 пікселів (індекс 1) на першому рівні будується чотири зображення розміром 28×28 пікселів (індекси 111, 112, 113, 114). На другому рівні "піраміди" будується 16 зображень розміром 14×14 пікселів (індекси 11111, 11112, 11121, 11122, 11211, 11212, 11211, 11222, 11311, 11312, 11321, 11322, 11411, 11412, 11421, 11422), а на третьому рівні – 64 зображення розміром 7×7 пікселів із відповідною індексацією. Створення "піраміди" космічних знімків рівнозначно застосуванню "стрибаючого вікна" під час побудови ПФР. У цьому разі величина кроку переміщення дорівнює розміру "вікна" на кожному рівні "піраміди".

На четвертому етапі розраховуються фрактальні розмірності "піраміди" космічних знімків. Величини фрактальних розмірностей можна визначити за методом покриття [14] або за методом призми [15]. У такий спосіб створюється індексована "піраміда" фрактальних розмірностей, яка зберігається в пам'яті комп'ютера.



Рис. 5. Структурна схема методу виявлення аномалій на космічних знімках сільськогосподарських земель із використанням "пірамідального" фрактального аналізу

На п'ятому етапі порівнюються величини фрактальних розмірностей на різних рівнях "піраміди" космічних знімків. Якщо фрактальні розмірності на нижчих рівнях "піраміди" не відрізняються за першим знаком після коми від фрактальних розмірностей на вищому рівні, тоді можна вважати, що аномалій на космічному знімку немає. Різниця фрактальних розмірностей за першим знаком після коми на різних рівнях "піраміди" дає змогу визначити аномальні ділянки на вихідному космічному зображенні поля за індексом "піраміди".

На шостому етапі визначаються мінімальні фрактальні розмірності на різних рівнях "піраміди" космічних світлин, які відповідають найбільш аномальним ділянкам.

На сьомому етапі узагальнюються результати виявлення аномалій на космічних знімках сільськогосподарських земель. У цьому разі за результатами фрактального аналізу визначається кількість аномальних ділянок та їх положення, а за величинами мінімальних фрактальних розмірностей виявляються найбільш аномальні ділянки на космічному зображенні.

Результати дослідження та їх обговорення

Розглянемо застосування розробленого методу для автоматизованого виявлення аномалій на космічних знімках сільськогосподарських земель.

Космічні світлини каналу b8 супутника Sentinel-2 отримано з мережі Інтернет. На знімках візуально обрано одне поле з пшеницею без аномалій і те саме поле з пшеницею з аномаліями в іншу дату. Визначено розміри поля на космічній світлині й вирізано зображення полів розміром 56×56 пікселів, які наведено на рис. 1, а і б. Далі створено "піраміду" космічних знімків та індексовано отримані зображення. За цими знімками розраховано фрактальні розмірності, визначені методом покриття. На рис. 6–9 подано космічні світлини поля з пшеницею без аномалій, що створюють "піраміду", а поряд з ними величини їх фрактальних розмірностей. Індикацію зображень наведено зверху й ліворуч від зображень, а також проіндексовано й фрактальні розмірності D .

На рис. 6 подано вихідний космічний знімок поля з пшеницею без аномалій розміром 56×56 пікселів і його фрактальну розмірність для "вікна" 56×56 пікселів.


D	1
1	 $D_1 = 2,912$

Рис. 6. Космічний знімок поля з пшеницею без аномалій розміром 56×56 пікселів, фрактальна розмірність для "вікна" 56×56 пікселів

Відповідно до рис. 6 на полі з пшеницею аномалії візуально не спостерігаються, а фрактальна розмірність дорівнює $D_1=2,912$. Тобто можна вважати, що фрактальна розмірність понад 2,9 ($D>2,9$) визначає стан сільськогосподарських земель на космічних знімках без аномалій.

На рис. 7 подано чотири космічні зображення першого рівня "піраміди" розміром 28×28 пікселів та їх фрактальні розмірності.

На рис. 7 видно, що фрактальні розмірності трьох знімків ($D_{111}=2,954$; $D_{112}=2,969$; $D_{121}=2,957$) перевищують фрактальну розмірність вихідного зображення $D_1=2,912$ і на цих ділянках знімка аномалії не спостерігаються. Фрактальна розмірність четвертого знімка мінімальна ($D_{122}=2,889$) і менша ніж $D_1=2,912$, але різниця не значна – $\Delta D=0,023$, і можна вважати, що аномалія на ньому також відсутня.





1	1	2
1	 $D_{111}=2,954$	 $D_{112}=2,969$
2	 $D_{121}=2,957$	 $D_{122}=2,889$

Рис. 7. Космічні знімки поля з пшеницею без аномалій розміром 28×28 пікселів, фрактальні розмірності для "вікна" 28×28 пікселів

На рис. 8 наведено 16 космічних знімків другого рівня "піраміди" розміром 14×14 пікселів та їх фрактальні розмірності.

Як видно на рис. 8, фрактальні розмірності всіх космічних знімків другого рівня "піраміди" перевищують $D=2,9$. Отже, можна зробити висновок, що аномалій на них немає. Необхідно зауважити, що фрактальні розмірності другого рівня "піраміди", які відповідають четвертому знімку першого рівня "піраміди", також перевищують $D=2,9$ ($D_{12211}=2,947$; $D_{12212}=2,974$; $D_{12221}=2,961$; $D_{12222}=2,954$). Це підтверджує, що на вихідному космічному знімку аномалій немає.

На рис. 9 наведено частку космічних знімків третього рівня "піраміди" розміром 7×7 пікселів (16 знімків) та їх фрактальні розмірності. Ці космічні зображення відповідають першому верхньому лівому квадранту вихідного космічного знімка, поданого на рис. 6.

На рис. 8 видно, що фрактальні розмірності всіх космічних знімків третього рівня "піраміди" (як і другого рівня) більші за фрактальну розмірність $D=2,9$, що підтверджує висновок про відсутність аномалій на космічному зображенні.

Для спрощення узагальнення результатів виявлення аномалій на космічних знімках сільськогосподарських земель побудуємо "піраміду" фрактальних розмірностей космічної світлин поля з пшеницею (див. табл. 1).

Аналіз показників, поданих у табл. 1, демонструє, що фрактальні розмірності всіх космічних знімків всіх рівнів "піраміди" перевищують $D=2,9$, окрім фрактальної розмірності на першому рівні "піраміди" для нижнього правого квадранту вихідного космічного зображення.

Узагальнення фрактального аналізу космічного знімка поля з пшеницею дає змогу зробити висновок, що на ньому аномалій немає.








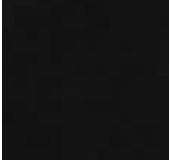



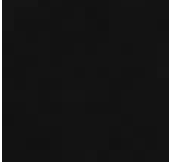
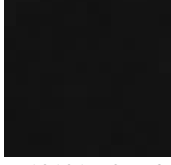

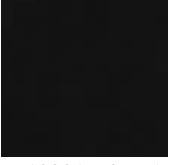
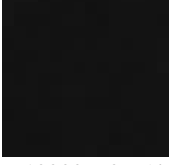
11	1	2	12	1	2
1	 $D11111 = 2,956$	 $D11112 = 2,957$	1	 $D11211 = 2,968$	 $D11212 = 2,960$
2	 $D11121 = 2,968$	 $D11122 = 2,964$	2	 $D11221 = 2,960$	 $D11222 = 2,974$
21	1	2	22	1	2
1	 $D12111 = 2,955$	 $D12112 = 2,963$	1	 $D12211 = 2,947$	 $D12212 = 2,974$
2	 $D12121 = 2,968$	 $D12122 = 2,970$	2	 $D12221 = 2,961$	 $D12222 = 2,954$

Рис. 8. Космічні знімки поля з пшеницею без аномалій розміром 14×14 пікселів, фрактальні розмірності для "вікна" 14×14 пікселів


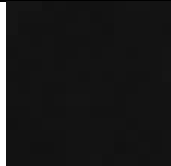

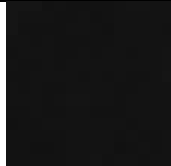












111	1	2	112	1	2
1	 $D1111111 = 2,956$	 $D1111112 = 2,971$	1	 $D1111211 = 2,997$	 $D1111212 = 2,950$
2	 $D1111121 = 2,941$	 $D1111122 = 2,976$	2	 $D1111221 = 2,955$	 $D1111222 = 2,973$
121	1	2	122	1	2
1	 $D11121 = 2,939$	 $D11121 = 2,963$	1	 $D1112211 = 2,976$	 $D1112212 = 2,966$
2	 $D11121 = 2,958$	 $D11121 = 2,950$	2	 $D1112221 = 2,970$	 $D1112222 = 2,973$

Рис. 9. Космічні знімки поля з пшеницею без аномалій розміром 7×7 пікселів, фрактальні розмірності для "вікна" 7×7 пікселів

Таблиця 1. "Піраміда" фрактальних розмірностей космічного знімка поля з пшеницею без аномалій для різних розмірів "вікна"

D	1				
1	2,912				
1	1	2			
1	2,954	2,969			
2	2,957	2,889			
11	1	2	12	1	2
1	2,956	2,957	1	2,968	2,960
2	2,968	2,964	2	2,960	2,974
21	1	2	22	1	2
1	2,955	2,963	1	2,947	2,974
2	2,968	2,970	2	2,961	2,954
11	1	2			
1	2,956	2,957			
2	2,968	2,964			
111	1	2	112	1	2
1	2,956	2,971	1	2,997	2,950
2	2,941	2,976	2	2,955	2,973
121	1	2	122	1	2
1	2,939	2,963	1	2,976	2,966
2	2,958	2,950	2	2,970	2,973

На рис. 10–13 наведено космічні світлини поля з пшеницею з аномаліями, що створюють "піраміду", а поряд з ними подано величини їх фрактальних розмірностей.

На рис. 10 запропоновано вихідний космічний знімок розміром 56×56 пікселів і його фрактальна розмірність для "вікна" 56×56 пікселів.

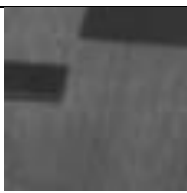
D	1
1	 $D_1 = 2,937$

Рис. 10. Космічний знімок поля з пшеницею з аномаліями розміром 56×56 пікселів, фрактальна розмірність для "вікна" 56×56 пікселів

Як видно на рис. 10, на полі з пшеницею візуально спостерігаються дві аномалії, але фрактальна розмірність становить $D_1 = 2,937$, тобто перевищує 2,9. Отже, величина фрактальної розмірності всього вихідного космічного знімка не дає змоги виявити аномалії на ньому. Вище зазначалося, що $D > 2,9$ визначає стан сільськогосподарських земель на космічних світлинах без аномалій.

На рис. 11 подано чотири космічні знімки першого рівня "піраміди" розміром 28×28 пікселів та їх фрактальні розмірності.

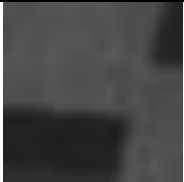
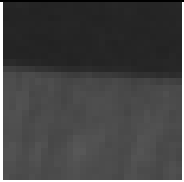
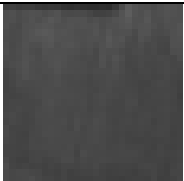
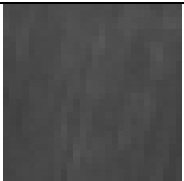
1	1	2
1	 $D_{111} = 2,901$	 $D_{112} = 2,885$
2	 $D_{121} = 2,926$	 $D_{122} = 2,948$

Рис. 11. Космічні знімки поля з пшеницею з аномаліями розміром 28×28 пікселів, фрактальні розмірності для "вікна" 28×28 пікселів

На рис. 11 видно, що фрактальні розмірності трьох зображень ($D_{111} = 2,901$; $D_{112} = 2,885$; $D_{121} = 2,926$) менші, ніж фрактальна розмірність вихідного космічного знімка $D_1 = 2,937$, і на цих ділянках спостерігаються аномалії. Більш аномальні ділянки визначаються меншими фрактальними розмірностями. Мінімальна фрактальна розмірність $D_{112} = 2,885$ відповідає другому космічному знімку, на якому аномалія займає майже половину зображення. Фрактальна розмірність четвертої світлини $D_{122} = 2,948$ перевищує $D_1 = 2,937$, і можна вважати, що аномалії на ній немає.

На рис. 12 подано 16 космічних знімків другого рівня "піраміди" розміром 14×14 пікселів та їх фрактальні розмірності.

Як видно на рис. 12, три фрактальні розмірності світлин другого рівня "піраміди" ($D_{11112} = 2,881$; $D_{11121} = 2,680$; $D_{11122} = 2,866$), що відповідають першому знімку першого рівня "піраміди", менші за фрактальну розмірність $D_{111} = 2,901$. Тому можна зробити висновок, що на них є аномалії. Мінімальна фрактальна розмірність $D_{11121} = 2,680$ відповідає третьому космічному знімку, на якому аномалія поширена майже на всьому зображенні. Фрактальна розмірність першого космічного знімка другого рівня "піраміди" дорівнює $D_{11111} = 2,966$, що

перевищує $D_{111}=2,901$, і $D=2,9$, тобто на ньому аномалії відсутні.

Крім того, на рис. 12 видно, що дві фрактальні розмірності космічних знімків другого рівня "піраміди" ($D_{11211}=2,735$; $D_{11212}=2,786$), що відповідають другому знімку першого рівня "піраміди", менші за фрактальну розмірність $D_{112}=2,885$ і на них є аномалії. Мінімальна

фрактальна розмірність $D_{11211}=2,735$ відповідає першому космічному знімку, й аномалія на ньому найбільша. Також необхідно зазначити, що фрактальні розмірності третього й четвертого космічного знімка другого рівня "піраміди" дорівнюють $D_{11221}=2,941$ і $D_{11222}=2,957$, які більші за $D_{112}=2,885$ і $D=2,9$. Це відповідає відсутності аномалій на них.





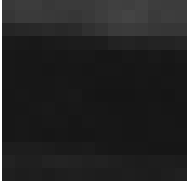

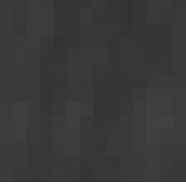
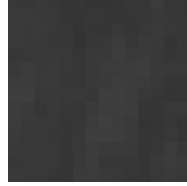
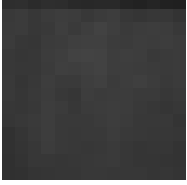
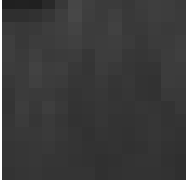
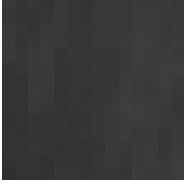
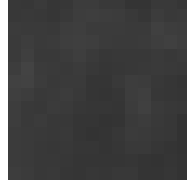
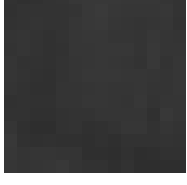

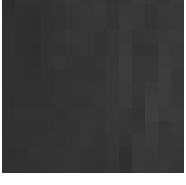
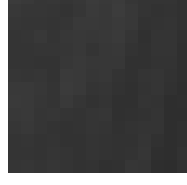
11	1	2	12	1	2
1	 $D_{11111} = 2,966$	 $D_{11112} = 2,881$	1	 $D_{11211} = 2,735$	 $D_{11212} = 2,786$
2	 $D_{11121} = 2,680$	 $D_{11122} = 2,866$	2	 $D_{11221} = 2,941$	 $D_{11222} = 2,957$
21	1	2	22	1	2
1	 $D_{12111} = 2,938$	 $D_{12112} = 2,904$	1	 $D_{21211} = 2,956$	 $D_{12212} = 2,937$
2	 $D_{12121} = 2,955$	 $D_{21122} = 2,961$	2	 $D_{12221} = 2,929$	 $D_{12222} = 2,949$

Рис. 12. Космічні знімки поля з пшеницею з аномаліями розміром 14×14 пікселів, фрактальні розмірності для "вікна" 14×14 пікселів

Також три фрактальні розмірності космічних знімків другого рівня "піраміди" (див. рис. 12) ($D_{12111}=2,938$; $D_{12121}=2,955$; $D_{21122}=2,961$), що відповідають третьому знімку першого рівня "піраміди", перевищують фрактальну розмірність $D_{121}=2,926$. Можна зробити висновок, що на них аномалій немає. Фрактальна розмірність другого космічного знімка другого рівня "піраміди" дорівнює $D_{12112}=2,904$, що незначно менше, ніж

$D_{121}=2,926$, але перевищує $D=2,9$. І також можна дійти висновку, що на ньому аномалій немає.

Фрактальні розмірності всіх космічних світлин другого рівня "піраміди" (див. рис. 12), які відповідають четвертому знімку першого рівня "піраміди" більші за $D=2,9$, що дає змогу зробити висновок, що на них аномалії відсутні.

На рис. 13 наведено частку космічних знімків третього рівня "піраміди" розміром 7×7 пікселів

(16 світлин) та їх фрактальні розмірності. Ці зображення відповідають першому верхньому лівому квадранту вихідного космічного знімку, поданого на рис. 10.

Як видно на рис. 13, фрактальні розмірності всіх космічних зображень третього рівня "піраміди", які відповідають першому знімку другого рівня "піраміди", перевищують $D=2,9$, тобто на них аномалій немає.

Крім того, на рис. 13 видно, що дві фрактальні розмірності космічних знімків третього рівня "піраміди" ($D_{1111212}=2,792$; $D_{1111222}=2,812$), які відповідають другій світлинці другого рівня "піраміди", менші за фрактальну розмірність $D_{11112}=2,881$

і на них є аномалії. Найбільша аномалія спостерігається на другому космічному знімку, і йому відповідає мінімальна фрактальна розмірність $D_{1111212}=2,792$. Також необхідно зазначити, що фрактальні розмірності першого й третього космічного зображення третього рівня "піраміди" становлять $D_{1111211}=2,944$ і $D_{1111221}=2,885$, тобто перевищують $D_{11112}=2,881$. Але фрактальна розмірність першого космічного знімка більша й за $D=2,9$, що відповідає відсутності аномалій на ньому, а фрактальна розмірність третього зображення наближена до $D_{11112}=2,881$, і на ньому є незначна аномалія.




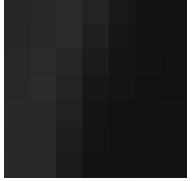



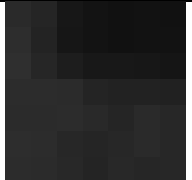
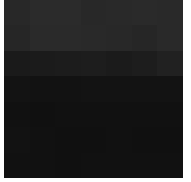
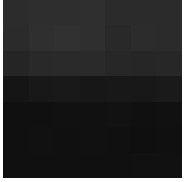





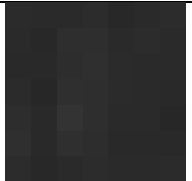
111	1	2	112	1	2
1	 $D_{1111111} = 2,943$	 $D_{1111112} = 2,950$	1	 $D_{1111211} = 2,944$	 $D_{1111212} = 2,792$
2	 $D_{1111121} = 2,964$	 $D_{1111122} = 2,946$	2	 $D_{1111221} = 2,885$	 $D_{1111222} = 2,812$
121	1	2	122	1	2
1	 $D_{1112111} = 2,705$	 $D_{1112112} = 2,721$	1	 $D_{1112211} = 2,791$	 $D_{1112212} = 2,947$
2	 $D_{1112121} = 2,873$	 $D_{1112122} = 2,903$	2	 $D_{1112221} = 2,646$	 $D_{1112222} = 2,902$

Рис. 13. Космічні знімки поля з пшеницею з аномаліями розміром 7×7 пікселів, фрактальні розмірності для "вікна" 7×7 пікселів

Дві фрактальні розмірності космічних знімків третього рівня "піраміди" (див. рис. 13) ($D_{1112111}=2,705$; $D_{1112112}=2,721$), що відповідають третій світлинці другого рівня "піраміди", близькі до фрактальної розмірності $D_{11121}=2,680$,

і це свідчить про наявність на них аномалій. Мінімальна фрактальна розмірність $D_{11121}=2,680$ відповідає третьому космічному знімку, на якому аномалія поширена майже на всій світлинці. Фрактальні розмірності третього й четвертого

космічних зображень третього рівня "піраміди" становлять $D_{1112121}=2,873$ і $D_{1112122}=2,903$, що перевищує $D_{11121}=2,680$ і $D_{1112211}=2,791$. Тобто можна вважати, що на них аномалій немає. Але зауважимо, що на четвертому космічному знімку аномалія, виявлена на вищих рівнях "піраміди", займає все зображення розміром 7×7 і для цієї світлини не є відхиленням ($D_{1112122}=2,903$ більше за $D=2,9$).

Дві космічні світлини третього рівня "піраміди", що відповідають четвертому знімку другого рівня "піраміди" (див. рис. 13) мають фрактальні розмірності ($D_{1112211}=2,791$; $D_{1112221}=2,646$) менші, ніж $D_{11122}=2,866$, тобто на цих знімках є аномалії. Мінімальна фрактальна розмірність $D_{1112221}=2,646$ свідчить, що найбільша аномалія є на третьому космічному зображенні. Фрактальні розмірності другого й четвертого космічних світлин третього рівня "піраміди" дорівнюють $D_{1112212}=2,947$ і $D_{1112222}=2,902$, і вони є більшими за $D_{11122}=2,866$ і $D=2,9$, тобто на цих знімках аномалій немає.

Результати виявлення аномалій на зображеннях сільськогосподарських земель узагальнимо за допомогою "піраміди" фрактальних розмірностей космічного знімка поля з пшеницею з аномаліями (див. табл. 2).

Аналіз показників, поданих у табл. 2, демонструє, що на першому рівні "піраміди" фрактальні розмірності першого, другого й третього космічних знімків менші за фрактальну розмірність усього вихідного космічного зображення. Це свідчить про наявність на них аномалій, а на четвертому зображенні аномалії відсутні (його фрактальна розмірність перевищує фрактальну розмірність усього вихідного космічного знімка).

Порівняння фрактальних розмірностей другого рівня "піраміди" (див. табл. 2) з фрактальними розмірностями першого рівня дає змогу локалізувати аномалії. Так, на першому космічному знімку першого рівня "піраміди" аномалія займає тільки три четверті зображення, на другому – половину, на третьому – четверть світлини, а на четвертому зображенні аномалій немає.

Третій рівень "піраміди" фрактальних розмірностей дає змогу продовжити аналіз аномалій (більш детально завдяки меншим розмірам знімків). У табл. 2 для прикладу наведено тільки четверту частину

фрактальних розмірностей третього рівня "піраміди", яка відповідає першому верхньому лівому квадранту вихідного космічного знімка. Перші чотири фрактальні розмірності третього рівня "піраміди" більші, ніж $D=2,9$, тобто на світлинах, які вони характеризують, аномалій немає. Аналіз других чотирьох фрактальних розмірностей третього рівня "піраміди" демонструє, що аномалія займає три четверті знімка вищого рівня. Треті чотири фрактальні розмірності показують, що аномалія поширена на половині світлини вищого рівня й розташована горизонтально, а четверті чотири фрактальні розмірності демонструють, що аномалія займає половину зображення вищого рівня вертикально. Але, як зазначалося раніше, четверта фрактальна розмірність четвертих чотирьох фрактальних розмірностей дорівнює $D_{1112122}=2,903$, що перевищує $D=2,9$, але цей космічний знімок розміром 7×7 повністю заповнений аномалією, виявленою на вищих рівнях "піраміди".

Таблиця 2. "Піраміда" фрактальних розмірностей космічного знімка поля з пшеницею з аномаліями для різних розмірів "вікна"

D	1				
1	2,937				
11	1	2			
1	2,901	2,885			
2	2,926	2,948			
111	1	2	12	1	2
1	2,966	2,881	1	2,735	2,786
2	2,680	2,866	2	2,941	2,957
21	1	2	22	1	2
1	2,938	2,904	1	2,956	2,937
2	2,955	2,961	2	2,929	2,949
1111	1	2			
1	2,966	2,881			
2	2,680	2,866			
11111	1	2	112	1	2
1	2,943	2,950	1	2,944	2,792
2	2,964	2,946	2	2,885	2,812
121	1	2	122	1	2
1	2,705	2,721	1	2,791	2,947
2	2,873	2,903	2	2,646	2,902

Висновки й перспективи подальшого розвитку

Регулярний космічний моніторинг земель сільськогосподарського призначення найчастіше проводиться із застосуванням вегетаційних індексів, але їх використання не дає змоги оцінити аномалії

на космічних знімках, а також автоматизувати цей процес.

Оцінити складність структури космічної світлини (заповнення тривимірного подання знімка) можна за допомогою розрахунку фрактальної розмірності. Але візуалізація поля фрактальних розмірностей космічного зображення стає неефективною за умови використання "вікна" великих розмірів, а також застосування "стрибаючого вікна" не дає змоги виявити аномальні ділянки на космічних світлинах. Крім того, під час візуалізації ПФР необхідно обирати параметри побудови ПФР залежно від типу поверхні, й автоматизувати виявлення аномалій стає неможливим.

Запропоновано метод виявлення аномалій на космічних знімках сільськогосподарських земель з використанням "пірамідального" фрактального аналізу, який дає змогу автоматизувати цей процес. Під час створення "піраміди" космічних світлин вихідний знімок розподіляється на зображення менших розмірів, і вони індексуються. На кожному рівні "піраміди" розраховуються фрактальні розмірності й будується їх "піраміда", що рівнозначно застосуванню "стрибаючого вікна" під час побудови ПФР.

Продемонстровано, що, якщо фрактальні розмірності космічних знімків всіх рівнів "піраміди" перевищують $D=2,9$ або фрактальні розмірності на нижньому рівні "піраміди" більші, ніж на верхньому рівні, тоді аномалій на космічному зображенні немає. Водночас, якщо фрактальні розмірності окремих ділянок нижнього рівня "піраміди" менші за фрактальні

розмірності на верхньому рівні й на вихідному космічному знімку, тоді на цих ділянках зображення спостерігаються аномалії.

У подальших дослідженнях доцільно розглянути можливість виявлення аномалій на космічних світлинах із використанням "ковзного вікна" під час розрахунку фрактальних розмірностей на кожному рівні "піраміди" космічних знімків

Конфлікт інтересів

Автор декларує, що не має конфлікту інтересів, зокрема фінансового, особистого, авторського чи будь-якого іншого характеру, який міг би вплинути на дослідження, а також на результати, опубліковані в цій статті.

Фінансування

Дослідження проводилося без фінансової підтримки.

Доступність даних

Рукопис не має пов'язаних матеріалів у сховищі даних.

Використання засобів штучного інтелекту

Автор підтверджує, що не застосовував технології штучного інтелекту для написання роботи.

References

1. Yang, L., Lu, B., Schmidt, M., Natesan, S. et al. (2025), "Applications of remote sensing for crop residue cover mapping", *Smart Agricultural Technology*, No. 11, P. 1–14. DOI: <https://doi.org/10.1016/j.atech.2025.1008080>
2. Tararik, O.G., Syrotenko, O.V., Iliencko, T.V., Kuchma, T.L. (2019), "Agroecological Satellite Monitoring", *Kyiv: Agrarian Science*, 204 p.
3. "Copernicus Europe's eyes on Earth, Sentinel-2. Copernicus Europe's eyes on Earth" [Electronic resource]. – Access mode: <https://www.copernicus.eu/en/about-copernicus/infrastructure/discover-our-satellites> – 26.12.2025.
4. Garajeh, M. K., Hassangholizadeh, K., Lomer, A. R. B., Ranjbari, A., et al. (2023), "Monitoring the impacts of crop residue cover on agricultural productivity and soil chemical and physical characteristics", *Scientific Reports*, No. 13-15054, P. 1–13. DOI: <https://doi.org/10.1038/s41598-023-42367-9>
5. Slobodjnik, M. P. (2014) "Prognostication productivity of agricultural cultures on materials of RSE and vegetations indexes", *Announcer of geodesy and cartography*, No. 6 (93), P. 16–20.
6. Maryushko, M. V., Pashchenko, R. E., Koblyuk, N. S. (2019), "Monitoring of agricultural crops using SENTINEL-2 satellite images", *Radioelectronic and Computer Systems*, No. 1 (89), P. 99–108. DOI: <https://doi.org/10.32620/reks.2019.1.11>
7. Yue, J., Tian, Q. (2020), "Estimating fractional cover of crop, crop residue, and soil in cropland using broadband remote sensing data and machine learning", *Int J Appl Earth Obs Geoinformation*, No. 89, P. 1–15. DOI: <https://doi.org/10.1016/j.jag.2020.102089>
8. Sudheer, K. P., Gowda, P., Chaubey, I., Howell, T. (2010), "Artificial Neural Network Approach for Mapping Contrasting Tillage Practices", *Remote Sens*, No. 2, P. 579–590. DOI: <https://doi.org/10.3390/rs2020579>

9. Feder, J. (1988), "Fractals", *New York: Springer US*, 263 p. DOI: <https://doi.org/10.1007/978-1-4899-2124-6>
10. Dolya, G. N., Ivanov, V. K., Kuchuk, G. A., Pashchenko, R. E. et al. (2006), "Fractal analysis of processes, structures, and signals" / Ed. by R. E. Pashchenko, *Kharkiv: NEO Ekoperspektiva*, 348 p.
11. Pashchenko, R., Mariushko, M. (2020), "Analysis of the Earth's Surface Type in Remote Sensing Using Fractal Dimension", *IEEE Ukrainian Microwave Week (UkrMW), MRRS-2020*, P. 1–4. DOI: <https://doi.org/10.1109/ukrmw49653.2020.9252811>
12. Pashchenko, R. E., Maryushko, M. V. (2023), "Estimation the state of different agricultural cultures with use fractal analysis", *Advanced Information Systems*, Vol. 7, Iss. 3, P. 81–88. DOI: <https://doi.org/10.20998/2522-9052.2023.3.14>
13. Pashchenko, R. E., Maryushko, M. V. (2021), "Monitoring changes of agricultural earths state is from data of space imageries fractal analysis", *Control, navigation and communication systems*, Vol. 65, Iss. 3, P. 8–17. DOI: <https://doi.org/10.26906/SUNZ.2021.3.008>
14. Crownover, R. M. (1995), "Introduction to Fractals and Chaos", *London: Jones and Bartlett Publishers, Inc.*, 352 p.
15. Clarke, K. C. (1986), "Computation of the Fractal Dimension of Topographic Surface Using the Triangular Prism Surface Area Method", *Computers & Geosciences*, Vol. 12, No. 5, P. 713–722.

Received (Надійшла) 08.02.2026

Accepted for publication (Прийнята до друку) 14.02.2026

Publication date (Дата публікації) 12.03.2026

Відомості про авторів / About the Authors

Пашченко Руслан Едуардович – доктор технічних наук, професор, Інститут радіофізики та електроніки ім. О. Я. Усикова НАН України, старший науковий співробітник відділу дистанційного зондування Землі; Харків, Україна;
Ruslan Pashchenko – Doctor of Technical Sciences, Professor, O. Ya. Usikov Institute for Radiophysics and Electronics of the National Academy of Sciences of Ukraine, Senior research scientist at the Department remote sensing of the Earth; Kharkiv, Ukraine;
 e-mail: r.pashchenko@khai.edu, r.paschenko@i.ua
 ORCID ID: <https://orcid.org/0000-0001-6218-0324>

HYBRID MODELS OF ARTIFICIAL INTELLIGENCE IN AUTOMATED ASSESSMENT TASKS FOR THE RELIABILITY OF INFORMATION SYSTEM VALIDATION RESULTS

Abstract. *The modern methods of the agricultural earths space monitoring with use vegetation indexes do not allow to carry out estimation anomalies on spaces pictures and to automatize this process. **The subject** of the study is to assess possibility discovery anomalies on spaces pictures with use "pyramidal" fractal analysis. **The object** of the study is satellite images of agricultural earths with anomalies and without anomalies taken by the Sentinel-2 satellite. **The aim** is to consideration method of automated discovery anomalies on the agricultural earths spaces pictures with use fractal analysis. The following **results** were obtained. It is set that visualization the field of fractals dimensions space picture becomes not effective during use of large nesses "window" and also use of "jumping window" does not allow to define anomalous areas on spaces pictures. The method of discovery anomalies on the agricultural earths spaces pictures with use "pyramidal" fractal analysis which allows to automatize this process is offered. Creation "pyramid" of spaces pictures is carried out due to the division the base space picture on the less sizes pictures which in subsequent are indexed. At every level "pyramid" settle accounts fractal dimensions of spaces pictures and the "pyramid" fractals dimensions is built, that equivalently to application "jumping window" during construction the field of fractals dimensions. **Conclusions.** The conducted researches application the developed method of discovery anomalies showed that if fractals dimensions of all levels "pyramid" spaces pictures more fractal dimension of $D = 2,9$, or fractals dimensions at lower level "pyramid" more than at top level, anomalies on a space picture are not present. If fractals dimensions of separate areas lower level "pyramid" less fractals dimensions at top level and on an initial space picture, there are anomalies on these areas are present.*

Keywords: *monitoring of agricultural earths; anomaly; space picture; fractal analysis.*

Бібліографічні описи / Bibliographic descriptions

Пашченко Р. Е. Автоматизоване виявлення аномалій на космічних знімках сільськогосподарських земель із використанням фрактального аналізу. *Автоматизовані системи управління та прилади автоматики*. 2026. № 1 (188). С. 177–190. DOI: <https://doi.org/10.30837/0135-1710.2026.188.177>

Pashchenko, R. (2026), "Automated discovery anomalies on the agricultural earths spaces pictures with use fractal analysis", *Management Information System and Devices*, No. 1 (188), P. 177–190. DOI: <https://doi.org/10.30837/0135-1710.2026.188.177>

Yuriy Polupan, Olga Malyyeva, Maksym Kikot

MODELING AND ANALYSIS OF LOGISTICS SUPPLY AND DISTRIBUTION ROUTES FOR HIGH-TECH PRODUCTS CONSIDERING DIFFERENT MODES OF TRANSPORT

The **subject** of the article is the processes of organizing and optimizing the logistics of supply and distribution of high-tech defense products under wartime conditions, in particular, the development of a network-based logistics structure taking into account reverse flows, risks, and limited resources. The **purpose** of the study is to develop a model of supply and distribution routes that optimizes delivery time and costs while simultaneously reducing the risks of cargo loss and increasing the transport utilization rate. **Tasks** to be solved: to form a generalized combined structure of the logistics chain; to identify and analyze the main parameters of supply and distribution routes across different modes of transport; to develop a computer experiment plan; to conduct simulation modeling of supply and distribution routes; and to develop practical recommendations for implementing logistics projects under wartime conditions. **Research methods** and tools: a systems approach, simulation modeling, graphical and comparative analysis, and the AnyLogic platform. **Results** obtained: A combined structure of the logistics chain for a military drone manufacturing enterprise with multiple route options is schematically presented. A model is proposed that formalizes the logistics chain as a network of nodes and arcs within a transport system and includes a set of input parameters describing vehicles, routes, and order characteristics. The essence of the computer experiment lies in reproducing the properties of infrastructure and logistics projects using a discrete-event simulation model, which makes it possible to analyze critical parameters separately for supply and distribution, as well as to model cases of their interdependence. Particular attention is paid to the separation of routes into supply and distribution, as well as to the use of different modes of transport. **Conclusions**: The obtained results confirm that a decisive factor in logistics decision-making is a balanced ratio between economic costs, risk level, and expected effect. The proposed approach enables a quantitative assessment of management decision options within supply and distribution logistics chains of manufacturing enterprises, thereby increasing the validity of choosing an optimal development strategy.

Keywords: logistics; supply chain; risk management; optimization; combined structure; simulation modeling; alternative routes; supply and distribution; manufacturing enterprise; AnyLogic.

Introduction

Efficient management of logistics flows is a key factor in enhancing the competitiveness of manufacturing enterprises. In the current context of rising demand and constrained transportation infrastructure under martial law, the optimization of delivery routes for raw materials, components, and finished products plays a crucial role.

Modern logistics systems of manufacturing enterprises are characterized by a high level of complexity and a multi-level structure. They encompass a network of raw material and component suppliers, manufacturing plants, warehouses, and end consumers, among which material, informational, and financial flows are organized. The effective functioning of such a system depends on the coordination of all elements of the chain, the optimization of transportation routes.

The study of logistics systems under martial law conditions is further complicated by the fact that the response of the transport and logistics network to changes in management parameters may manifest over months

or even years. Therefore, conducting a full-scale experiment becomes economically and organizationally impractical. This is especially relevant for logistics and infrastructure projects in the field of security and defense, which are characterized by high costs and increased risks.

Analysis of publications

An analysis of publications on the topic reveals a significant interest in the issue of combined logistics routes. In [1], various types of reverse flows, as well as the factors, barriers, and costs influencing managerial decisions in this field, are systematized. Article [2] examines cooperation between the civilian and military sectors in reverse logistics to support the EU's circular economy, proposing a model for integrating a sustainable supply chain. In [3], an information system for managing the military supply chain based on incident analysis is proposed.

Article [4] investigates the use of military logistics drones for rapid resource delivery in challenging

environments, emphasizing their autonomy and flexibility. Study [5] presents mathematical models and geographic information systems for optimizing logistics routes under wartime conditions.

The authors of [6] analyze strategies for managing supply chains in conflict conditions to enhance business resilience. Paper [7] proposes a concept for integrating forward and reverse logistics within the framework of a circular economy, focusing on overcoming barriers to closing material cycles.

Studies [8] and [9] present algorithms and systems designed to improve the efficiency of military logistics support, automate flow management, and forecast resource needs. In [10], key nodes of reverse logistics networks with remanufacturing are examined, highlighting the role of emerging technologies.

It can be concluded that traditional methods for planning and optimizing logistics processes fail to account for the dynamics and uncertainty inherent in wartime conditions. Moreover, existing approaches typically consider the supply chain as a single entity, without distinguishing between supply routes and distribution routes. This reduces flexibility and prevents the consideration of situations in which supply and distribution have different priorities, risks, or time constraints.

Therefore, for conducting research on supply and distribution logistics, the use of the following methods and technologies is considered effective:

- operations research methods – for the formalization and forecasting of the technical parameters of transport infrastructure in order to design optimal routes [11];
- risk theory – for developing risk maps and performing quantitative and qualitative analyses of potential threats, taking into account uncertain or incomplete data [12];
- computer modeling methods – for reproducing complex scenarios of vehicle movement and the interaction of agents within supply logistics systems through agent-based models [13];
- artificial intelligence technologies, in particular multi-agent systems, for modeling and optimizing the independent or interdependent operation of supply and distribution routes [14].

Problem Statement

The purpose of this study is to develop a model of supply and distribution routes that enables the optimization of delivery time and costs while

simultaneously reducing the risk of cargo loss and increasing transport utilization. Particular attention is given to the distinction between supply and distribution routes, as well as to the use of different modes of transportation: trucks, trains, and aircraft.

To achieve this goal, the following objectives must be addressed:

- to formulate a generalized combined structure of the supply chain encompassing both supply and distribution routes;
- to identify and analyze the key parameters of supply and distribution routes across different modes of transport;
- to design a computer experiment plan including several modeling scenarios;
- to perform modeling of supply and distribution paths using an illustrative example, taking into account the following criteria: time, risks, cost, and reliability;
- to develop practical recommendations for the implementation of logistics projects during wartime, considering both separate and joint optimization of supply and distribution processes.

Model Description

The logistics supply – distribution chains of a manufacturing enterprise can be divided into two components: the supply routes for material resources and the distribution routes for finished products. Each of these components may include various types of participants in the supply chain and, accordingly, different structural configurations.

In supply logistics, the main participants are:

- suppliers,
- intermediate warehouses (both wholesale and retail),
- intermediaries (carriers or logistics operators).

In distribution logistics, the main participants include:

- intermediaries (both wholesale and retail),
- end consumers.

It should be noted that when considering the production of dual-use goods during wartime, where the end users are military units, the storage of finished products should not be employed in distribution logistics in order to ensure the fastest possible delivery. Moreover, volunteer organizations may act as intermediaries in such cases. Figure 1 schematically illustrates a combined structure of the supply chain for an enterprise engaged in the production of military-purpose drones. It should

also be emphasized that within such a supply chain, multiple route options exist between the specified elements. For the analysis of logistics networks and the selection of optimal solutions, it is effective to apply simulation modeling, particularly discrete-event models. This method makes it possible to account for the temporal dynamics of processes, random events, and variability in the operating conditions of logistics channels, which cannot be achieved using a classical analytical approach.

The model formalizes the supply chain as a network of nodes and arcs of the transport system represented as a graph $G = (V, E)$, where V – is the set of nodes. These represent individual elements of the logistics network: suppliers of raw materials and components, manufacturing enterprises, warehouses, and end consumers. Each node is characterized by its own storage parameters, throughput capacity, and order processing capabilities; E is the set of routes between nodes, i.e., the possible transportation paths. Each arc of the graph is described by parameters such as route length, expected travel time, type of transport, throughput capacity, and risk factors.

Thus, the model includes a set of input parameters describing transport vehicles, routes, and order

characteristics. The transport parameters include the type of vehicle $T = \{road, rail, air\}$, movement speed v , carrying capacity Q , and unit cost per kilometer c . A route is characterized by the distance between nodes d , base travel time $t = d/v$, and risk coefficient r , which reflects the probability of delays or additional costs. Order parameters include cargo volume W , delivery priority p , and delivery deadline D .

The input parameters can dynamically change during the simulation, reflecting fluctuations in supply and demand as well as the state of transport infrastructure.

As output parameters, the model provides a comprehensive assessment of the efficiency of logistics solutions using the following indicators: delivery time T_{del} , which accounts for possible delays along the route; transportation cost C_{del} , which includes additional expenses associated with risks and the choice of transport mode; and route reliability S_{rel} , inversely proportional to the risk coefficient. To obtain an integrated evaluation of efficiency, an aggregate criterion is applied, combining normalized indicators of time, cost, and reliability with corresponding weighting coefficients.

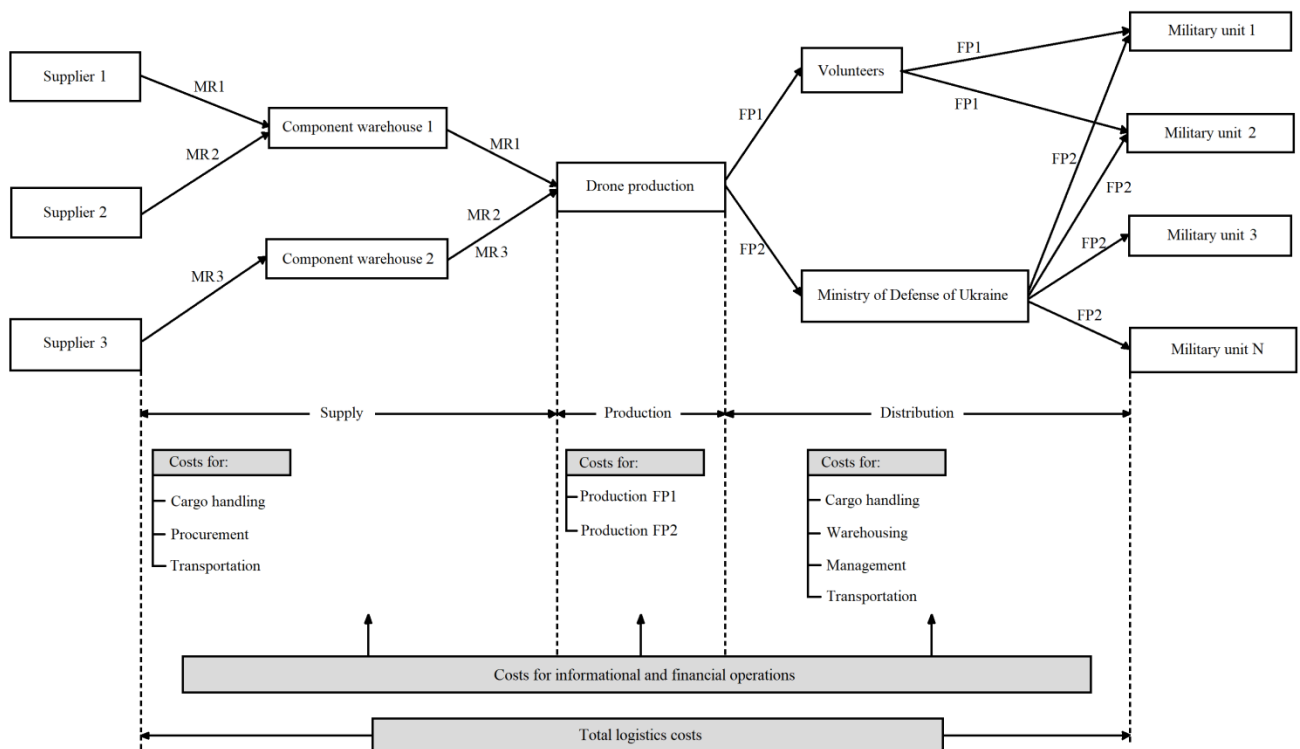


Fig. 1. The supply chain of supply, production, and distribution of unmanned systems

The logic of the model's operation can be described as follows. The model begins by initializing a network with predefined nodes and routes characterized by their parameters. Then, an order flow is generated to represent the demand for raw materials and finished products from consumers. For each order, the system determines feasible routes and available transport modes, calculates the expected delivery time, transportation cost, and probability of successful delivery. During the simulation process, random events such as delays, breakdowns, or other force majeure circumstances are considered, which modify the base parameters. An integrated performance score is calculated for all alternatives, allowing the identification of optimal routes and the formulation of recommendations for logistics flow management.

Among the key application features, the model enables a comprehensive analysis of both supply channels (from suppliers to production enterprises) and distribution channels (from manufacturers to end users), ensuring the evaluation of logistics flow efficiency and reliability. It also allows for extension and customization by incorporating additional criteria such as transportation environmental impact, infrastructure load, and warehouse turnover, thereby improving the flexibility and accuracy of assessments. Under martial law or emergency conditions, the model can be adapted by adjusting risk coefficients for specific routes or restricting their availability, which enables forecasting and managing potential delays and risks. Thus, the proposed discrete-event model provides a systematic approach to logistics

network analysis, accounts for temporal and stochastic factors, and supports scientifically grounded decision-making in optimizing supply and distribution routes.

Description of the Computer Experiment

For the evaluation of the efficiency and reliability of logistics channels in a manufacturing enterprise, a computer experiment was conducted using a transport network model of Ukraine.

In this study, the enterprise under consideration has a single plant located in Lviv and a network of suppliers and customers distributed across cities throughout the territory of Ukraine. The input parameters of the experiment were formed based on open data on transport infrastructure and transportation statistics, taking into account the specific conditions of operation under martial law. The main focus was placed on two types of logistics flows: supply channels for delivering resources from supplier cities to Lviv, and distribution channels for delivering products from Lviv to frontline and key regional cities.

The input parameters for each route included the type of transport, the distance between nodes, the expected travel time (both without delays and considering risk factors), the cargo capacity, and the qualitative risk characteristics. During the simulation, potential delays caused by technical or security-related restrictions were also taken into account. Tables 1 and 2 present the main parameters of local supply and distribution routes with the production center located in Lviv.

Table 1. Main parameters of local supply routes to Lviv (fragment)

№	City (Supplier)	Transport	Distance (km)	Time (h) (without delays / considering risks)	Load capacity (approx.)	Qualitative advantages / risks
1	2	3	4	5	6	7
1	Lutsk	Road	170.7	1.71 / 1.71–4.28	~20 t (truck)	Close to Lviv – fast delivery; moderate risk of local traffic jams
2	Lutsk	Rail	177.6	2.28 / 2.28–6.84	~75 t (wagon)	Higher throughput; requires coordination at stations
3	Lutsk	Air	136.6	0.23 / 0.23–0.58	~10–20 t (depending on aircraft)	Very fast for urgent batches; expensive, limited availability
4	Rivne	Road	225.4	2.25 / 2.25–6.75	~20 t	Short regional route; possible local restrictions
5	Rivne	Rail	234.4	3.00 / 3.00–12.00	~75 t	Suitable for larger batches; risks at junctions
6	Rivne	Air	180.3	0.30 / 0.30–0.75	~10–20 t	High speed but costly
7	Ternopil	Road	146.2	1.46 / 1.46–3.65	~20 t	Short routes – quick delivery; min delays in peacetime
8	Ternopil	Rail	152.1	1.95 / 1.95–5.85	~75 t	Suitable for medium-sized batches
9	Ternopil	Air	117.0	0.19 / 0.19–0.47	~10–20 t	Minimal delivery time; high cost
10	Vinnitsia	Road	409.1	4.09 / 4.09–12.27	~20 t	Medium distance; possible delays on highways
11	Vinnitsia	Rail	425.4	5.45 / 5.45–21.80	~75 t	Good capacity but high risk of node delays

Continuation of the table 1

1	2	3	4	5	6	7
12	Vinnitsia	Air	327.3	0.55 / 0.55–1.38	~10–20 t	Fast but costly and dependent on airports
13	Zhytomyr	Road	417.1	4.17 / 4.17–12.51	~20 t	Medium distance; risks in the central corridor
14	Zhytomyr	Rail	433.8	5.56 / 5.56–22.24	~75 t	Large shipments possible; complex coordination
15	Zhytomyr	Air	333.7	0.56 / 0.56–1.40	~10–20 t	Fast for small urgent cargo
16	Kyiv	Road	584.4	5.84 / 5.84–17.52	~20 t	Major supplier; high traffic intensity; increased road risks
17	Kyiv	Rail	607.8	7.79 / 7.79–38.95	~75 t	High capacity, but junctions are vulnerable points
18	Kyiv	Air	467.5	0.78 / 0.78–1.95	~10–20 t	Fast delivery of critical materials
...
25	Kropyvnytskyi	Road	770.2	7.70 / 7.70–26.95	~20 t	Medium distance; moderate risk
26	Kropyvnytskyi	Rail	801.0	10.27 / 10.27–51.35	~75 t	Effective for large batches
27	Kropyvnytskyi	Air	616.2	1.03 / 1.03–2.58	~10–20 t	Fast but more expensive

Table 2. Main parameters of distribution routes from Lviv

Direction	Transport type	Approx. distance (km)	Travel time (without delays / considering risks)	Load capacity	Qualitative advantages / risks
Lviv → Zaporizhzhia	Road	~900	16–20 h / 20–36 h	~20 t	Door-to-door delivery possible; high attack risks; long distance
Lviv → Zaporizhzhia	Rail	~930	18–24 h / 24–72 h	~75 t (×50–70 wagons)	Large-scale deliveries, but dependence on railway junctions
Lviv → Zaporizhzhia	Air	~800	1.5–2 h / 3–5 h	~10–20 t	Very fast delivery; high risk of air defense engagement
Lviv → Dnipro	Road	~820	14–18 h / 18–30 h	~20 t	High mobility; road attack risks
Lviv → Dnipro	Rail	~850	16–22 h / 24–60 h	~75 t	Large volumes; vulnerable infrastructure nodes
Lviv → Dnipro	Air	~720	1.5 h / 3–4 h	~10–20 t	Fastest option; increased operational risks
Lviv → Kharkiv	Road	~1000	18–22 h / 24–40 h	~20 t	Long route; risk of shelling
Lviv → Kharkiv	Rail	~1030	20–26 h / 30–72 h	~75 t	Large-scale transport; dependence on junctions
Lviv → Kharkiv	Air	~950	2 h / 3–5 h	~10–20 t	Very fast; vulnerable to attacks
Lviv → Poltava	Road	~720	12–16 h / 16–28 h	~20 t	Direct delivery; risks in central Ukraine
Lviv → Poltava	Rail	~740	14–20 h / 24–60 h	~75 t	Suitable for bulk cargo; rail junctions under threat
Lviv → Poltava	Air	~650	1.5 h / 3 h	~10–20 t	Fastest route; high cost and risk
Lviv → Sumy	Road	~965	9.5 h / 34 h	~20 t	Long route; elevated risk
Lviv → Sumy	Rail	~1004	13 h / 64.5 h	~75 t	Suitable for large batches if infrastructure remains stable
Lviv → Sumy	Air	~772	1.5 h / 3.5 h	~10–20 t	Fast delivery; high cost

Figures 2 and 3 graphically illustrate the inverse relationship between delivery time and cost parameters across different types of transport. Thus, even considering only two parameters, the task of selecting a mode of transport is a compromise problem. Therefore, solving the task of forming supply and distribution routes based on five criteria (some of which are fuzzy and some qualitative) by algorithmic means is highly complex and requires the use of simulation modeling.

Table 3 presents the comparative characteristics of the three main types of transport used in the logistics routes for the supply and distribution of products: road, rail, and air transport. To generalize the indicators, average values from the distribution routes from Lviv were used, which makes it possible to obtain a representative picture of the efficiency and limitations of each transport mode.

The main parameters include route length, expected travel time, average speed, maximum load capacity, transportation cost, route throughput, route flexibility, critical points, qualitative advantages and risks, as well as infrastructure availability and reliability.

In the logistic network model, each mode of transport is assigned indicators of risk, reliability, and weight coefficients for the integrated efficiency assessment, reflecting the specifics of transport channel functioning under conditions of increased uncertainty.

For making strategic decisions regarding cargo transportation, it is necessary to take into account possible changes in transportation conditions and to provide, in addition to the main mode of transport, reserve and critical alternatives. Moreover, combined routes based on different modes of transport can be

developed. By summarizing the data presented in Table 3 and identifying the primary decision-making criterion, it is possible to formulate recommendations, which are presented in Table 4. However, it should be noted

that more substantiated decisions can be made by considering the specific characteristics of the territories along the planned route.

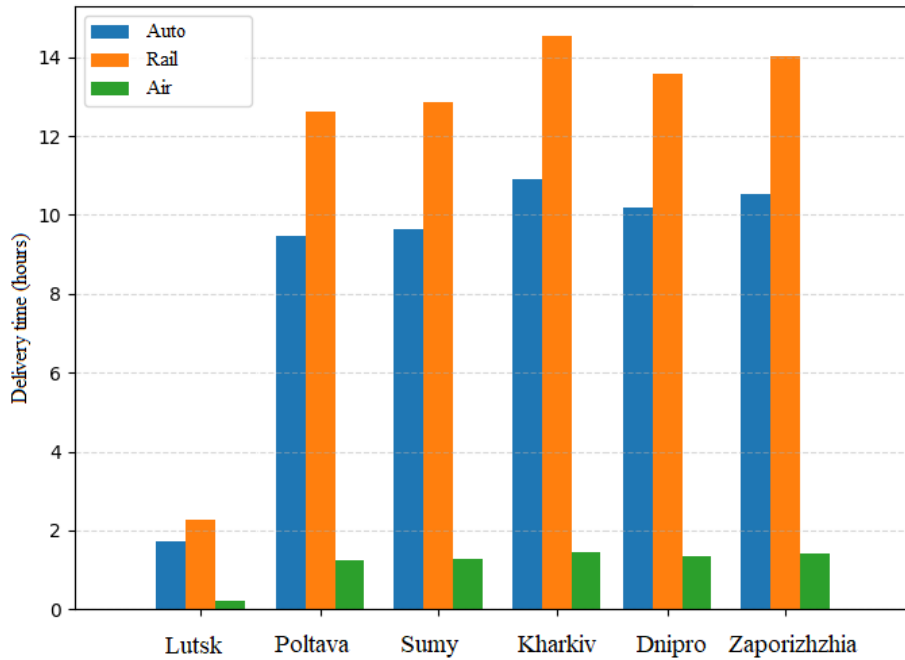


Fig. 2. Delivery time from Lviv by transport type

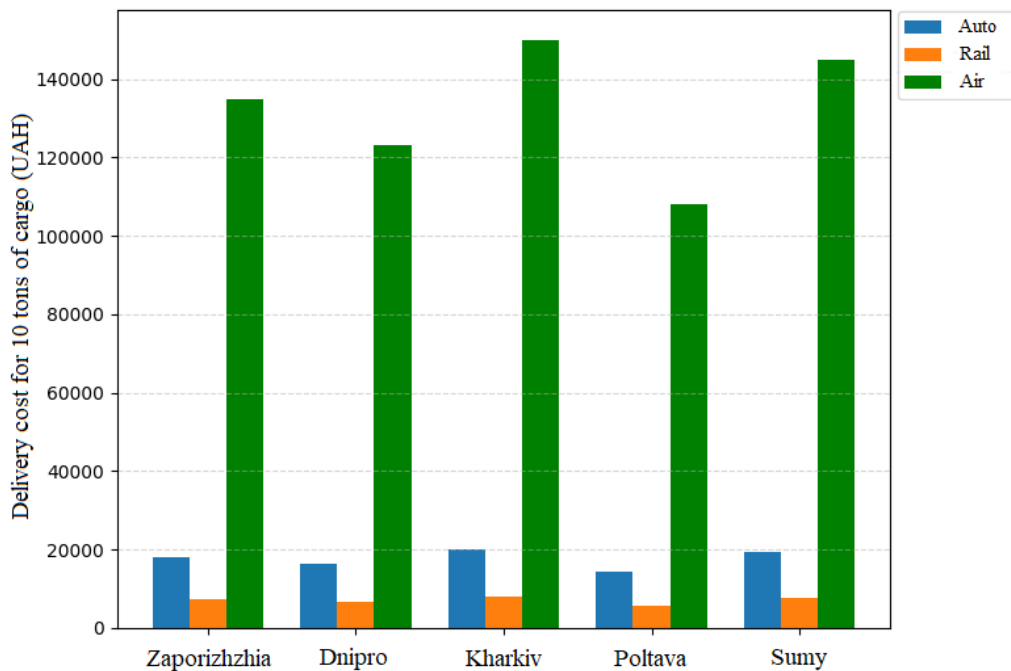


Fig. 3. Delivery cost from Lviv by transport type

Table 3. Distribution of transport modes depending on the primary delivery criterion

Parameter	Truck	Train	Air
Route length, km	820	840	650
Expected travel time, h (without delays / considering risks)	8 / 18	10 / 35	1 / 2
Average speed, km/h	~80	~60	~600
Maximum load capacity, t	20	75	15
Transportation cost per km, UAH	20	8	150
Transportation cost per tonne, UAH	820	90	6500
Route throughput, t/day	~400	~3,000–5,000	~200–300
Route flexibility	High	Medium	Low
Critical route points	Roadblocks, road repairs, shelling zones	Railway junctions, track damage, transshipment stations	Airport accessibility, air defense, weather conditions
Qualitative advantages	Door-to-door delivery, flexibility, local logistics	High capacity, cost efficiency, stability	Speed, urgency, contactless delivery
Main risks	Shelling, sabotage groups, traffic jams, road repairs	Track damage, infrastructure shelling, delays	Air defense strikes, weather limitations, accident losses
Infrastructure availability	High	High	Low
Reliability	Medium	High	Low
Cargo destruction risk, %	10	4	35

The risk indicator characterizes the probability of delays or additional costs during cargo transportation. Its values were established based on expert evaluations and analysis of Ukraine's transport infrastructure.

Table 4. Comparative characteristics of transport modes on the main supply route

Primary criterion	Main transport mode	Reserve transport mode	Critical transport mode
Minimization of delivery time	Air	Auto	Rail
Minimization of delivery cost	Rail	Auto	Air
Minimization of delivery risk	Rail	Auto	Air

For road transport, the risk is 0.2, reflecting the average influence of road conditions, checkpoints, and wartime threats. For rail transport, the risk is 0.25, accounting for the concentration at junction stations, possible track damage, and delays at transshipment points. The highest risk is associated with air transport (0.8) due to limited airport accessibility, air defense threats, weather conditions, and high maintenance costs.

To assess risk, factors determined by the hazards of wartime conditions can be used, namely the possibility of attacks on transport infrastructure. The consequences may include damage to railway tracks, highways, or bridges. Such events cannot be predicted in advance; however, based on historical data, it is possible to estimate the overall level of danger in a specific territory by considering the frequency of air raid alerts as well as artillery and missile strikes. Such studies were

conducted in [15], resulting in the determination of territorial hazard indices for both railway and road transport.

Table 5. Distribution of cities by hazard indices

Hazard index	Number of hazardous events	City
0.1	1	Ivano-Frankivsk
0.1	2	Lutsk
0.1	0	Uzhhorod
0.1	0	Chernivtsi
0.5	64	Vynnytsia
0.5	46	Zhytomyr
0.5	47	Kropyvnytskyi
0.5	60	Lviv
0.5	94	Poltava
0.5	18	Rivne
0.5	18	Ternopil
0.5	48	Cherkasy
0.5	69	Chernihiv
0.7	971	Zaporizhzhia
0.7	290	Kyiv
0.7	362	Odesa
0.7	105	Khmelnyskyi
0.9	9301	Dnipro
0.9	1187	Mykolaiv
0.9	3623	Sumy
0.9	3286	Kharkiv
0.9	3769	Kherson
1.0	Temporarily occupied territories	Donetsk
1.0	Temporarily occupied territories	Luhansk
1.0	Temporarily occupied territories	Simferopol

Reliability reflects the stability and predictability of delivery. Although this indicator is not directly used in the integrated efficiency formula, it serves for qualitative comparison of transport modes. For road

transport, reliability is 0.85, for rail transport – 0.90, and for air transport – 0.50. This correlation aligns with real-world characteristics: road transport is the most flexible, rail transport is the most stable, and air transport is the fastest but least predictable.

The integral assessment of route efficiency (*Score*), taking into account time, cost, and risk indicators, is calculated using an additive aggregation formula:

$$Score = w_1 T_{norm} + w_2 C_{norm} + w_3 R_{norm},$$

where $T_{norm}, C_{norm}, R_{norm}$ – are the normalized indicators of time, cost, and risk, respectively, and w_1, w_2, w_3 are the corresponding weighting coefficients.

The weight coefficients of the integrated evaluation (w_1 – time, w_2 – cost, w_3 – risk) in this experiment are set as $w_1 = 0.2, w_2 = 0.2, w_3 = 0.6$. The significant weight of the risk factor reflects the strategic priority

of transportation safety during wartime, when even short-term disruptions or infrastructure damage can lead to critical consequences. It should be noted that when the weighting coefficients change – for example, when transportation time becomes significantly more important than risk (which may be the case when delivering cargo for immediate needs at the front line) – preference will be given to air transport, such as helicopters or transport unmanned aerial vehicles.

The logistic network model, implemented in the AnyLogic environment, is shown in Figure 4. On the schematic diagram, blue squares represent suppliers, and red squares denote end customers. The model simulates the dynamic movement of road and rail transport, allowing analysis of supply routes, transportation duration, and the collection of statistical data for further efficiency calculations.

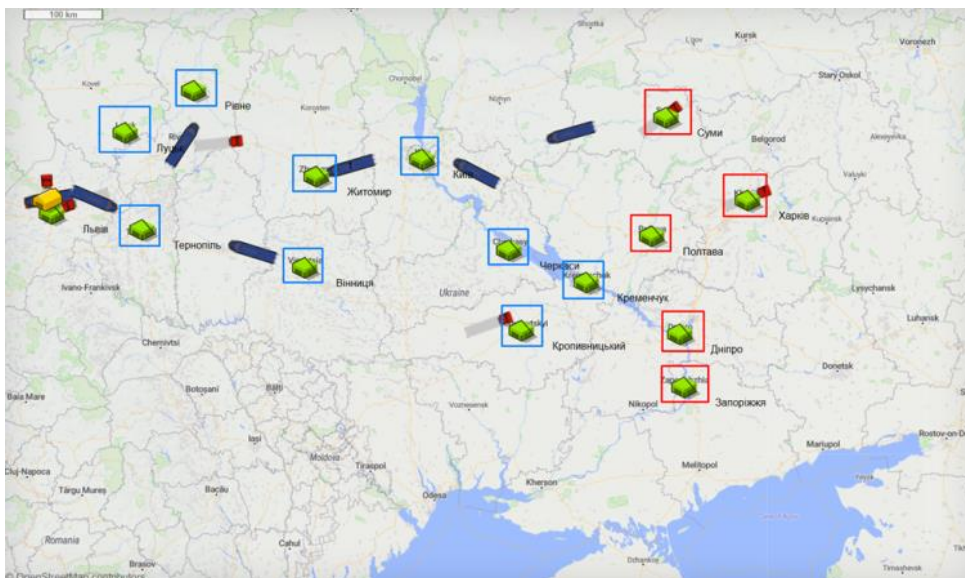


Fig. 4. Scheme of the supply and distribution logistics network implemented in the AnyLogic environment

For the experiment, a discrete-event model of the logistics system was developed, covering supply and distribution routes between major industrial and regional centers. Three baseline scenarios were formulated:

- Scenario 1 – Dependent supply and distribution using in-house transport. Characterized by stable operation of road transport over short and medium distances. The main advantage is route flexibility and the possibility of direct delivery; however, sensitivity to wartime risks and road congestion increases. The average reliability is 0.85.

- Scenario 2 – Dependent supply and distribution with cost, time, and risk optimization. This scenario

implements an algorithmic selection of the most efficient transport type according to an integral indicator (*Score*) calculated using established weighting coefficients. The model minimizes the risks of delays and cost overruns, prioritizing railway transport for large shipments and road transport for short routes. The average reliability is 0.87, indicating a balanced system.

- Scenario 3 – Independent supply and distribution. The channels operate autonomously, involving external logistics operators. The average reliability is 0.90; however, the system has lower flexibility and higher costs.

For each scenario, a series of computational experiments was conducted, resulting in the

determination of integral efficiency indicators for the routes. The evaluation, in addition to the three main criteria, also included an additional reliability parameter. The obtained results are presented in graphical form (Figures 5–7), which illustrate the advantages and limitations of each scenario.



Fig. 5. Integral efficiency of the experimental scenarios

The analysis of the results showed that the integral efficiency (Fig. 5), which accounts for costs, risks, and delivery time, is nearly identical in Scenarios 1 and 2 (≈ 0.83), whereas independent supply (Scenario 3) demonstrates lower efficiency (0.8) due to a higher risk of flow concentration (risk coefficient ≈ 0.32).

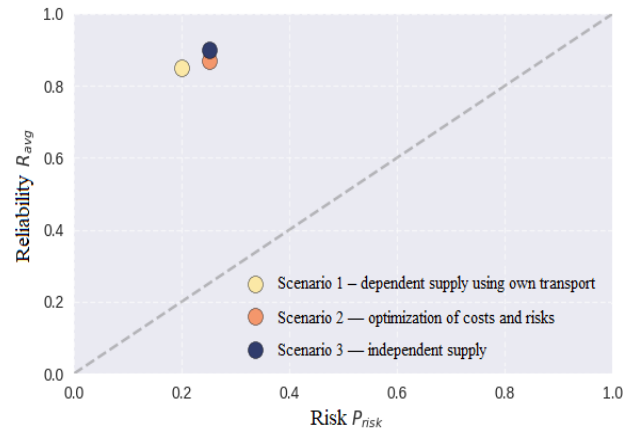


Fig. 6. Ratio of reliability to risk across scenarios

Figure 6 illustrates the relationship between transport system reliability and risk, where reliability is interpreted as an availability-oriented measure of successful delivery, while risk reflects aggregated probabilities of disruptive events under wartime conditions based on historical data and expert assessments. This relationship supports strategic decision-making by justifying rail transport as a stable primary channel, road transport as a flexible backup, and air transport for critical time-sensitive deliveries despite higher risk.

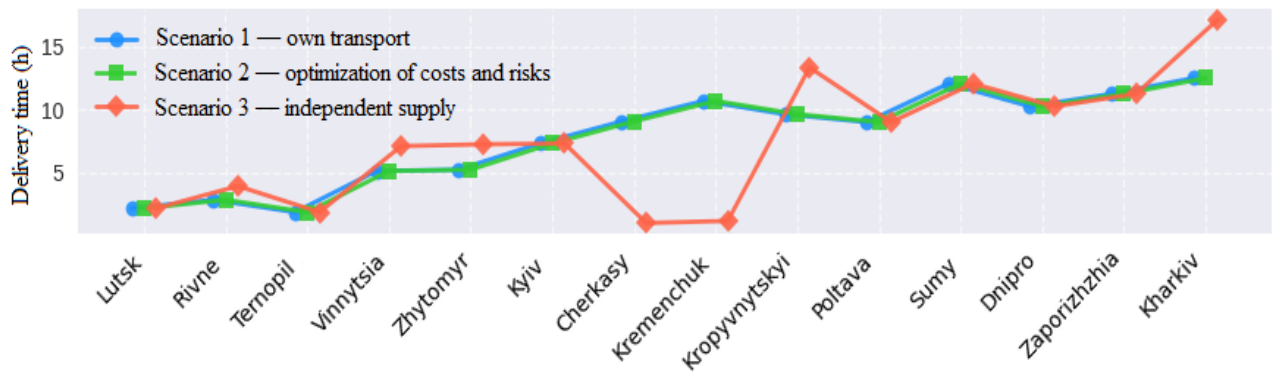


Fig. 7. Delivery time across cities in different scenarios

The best performance was achieved by rail transport in the optimization scenario, combining economic feasibility and stability. Road transport remained effective for short routes within the western and central regions.

Air transport, despite its minimal delivery time, proved economically inefficient and the least reliable (efficiency coefficient ≈ 0.3). The stability analysis based on the reliability-to-risk ratio further confirms the trade-off between efficiency and robustness across the

considered scenarios. The upper part of the chart (high reliability) represents the area of high or satisfactory stability, where the system operates efficiently with low risk. The lower part (low reliability) corresponds to the area of low stability, where risks increase significantly. Thus, the results of computer modeling confirm the feasibility of a hybrid logistics strategy, in which road transport ensures responsiveness on short routes, while rail transport provides stability and scalability over long distances.

Conclusions

Within the scope of the conducted study, the following results were obtained:

- a generalized combined structure of the logistics chain with supply and distribution routes was formed for an enterprise engaged in the production of military-purpose drones;
- the key parameters of supply and distribution routes for different modes of transport – such as delivery time, cost, risk, and reliability – were identified and analyzed; an integral indicator of route efficiency was proposed;
- a computer experiment plan was developed, including three simulation scenarios that take into account the interdependence of supply and distribution processes;
- a discrete-event model of supply and distribution routes was developed, enabling the analysis of alternative logistics scenarios with the determination of optimal delivery time and costs while simultaneously reducing the risk of cargo loss;
- supply and distribution routes were simulated using an illustrative case study within the territory of Ukraine, considering the criteria of distance, time, risk, cost, and reliability;
- recommendations were developed to support strategic decision-making in the implementation of logistics projects under wartime conditions.

The obtained results confirm that increased costs do not always correlate with improved

performance indicators, as the decisive factor is a balanced ratio between economic expenses, risk level, and expected outcomes. The proposed approach enables quantitative evaluation of management decision alternatives within the supply and distribution chains of manufacturing enterprises, thereby enhancing the justification for selecting an optimal development strategy.

Conflict of interest

The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, authorship or otherwise, that could affect the research and its results presented in this paper.

Funding

This research was conducted without financial support.

Data availability

The manuscript has no associated data.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence methods while creating the presented work.

References

1. Savchenko, L., Hryhorak, M. (2018), "Conceptual foundations for the development of reverse logistics in the circular economy", *Pryazovskiy Economic Bulletin*, P. 14–32. URL: http://www.pev.kpu.zp.ua/journals/2018/5_10_uk/5_10_2018.pdf
 2. Vasiliauskas, A. V., Ivanauskas, S., Čižiūnienė, K. (2024), "Challenges of ensuring reverse logistics in a military organization using outsourced services", *Sustainability*, Vol. 16, No. 11, Article 4569. DOI: <https://doi.org/10.3390/su16114569>
 3. Cantelmi, R., Di Gravio, G., Patriarca, R. (2020), "Learning from incidents: A supply chain management perspective in military environments", *Sustainability*, Vol. 12, No. 14, Article 5750. DOI: <https://doi.org/10.3390/su12145750>
 4. Minculete, G. (2025), "Military logistics drones: The innovative solution for transportation challenges on the battlefield", *Review of the Air Force Academy*, P. 342–354. DOI: <https://doi.org/10.2478/raft-2025-0033>
 5. Zahoryanska, O., Ziabriev, V., Kyshchuk, D. (2025), "Logistics support of enterprise activity under martial law: problems and optimization ways", *Scientific Bulletin of Kremenchuk Mykhailo Ostrohradskiy National University*, No. 3 (2), P. 342–354. DOI: [https://doi.org/10.31891/2307-5740-2025-342-3\(2\)-6](https://doi.org/10.31891/2307-5740-2025-342-3(2)-6)
 6. Tereshchenko, S. I., Yevtushenko, A. M. (2023), "Supply chain: management and optimization", *Journal of Strategic Economic Studies*, No. 6 (17), P. 207–214. DOI: <https://doi.org/10.30857/2786-5398.2023.6.21>
 7. Ding, L., Wang, T., Chan, P. (2023), "Forward and reverse logistics for circular economy in construction: a systematic literature review", *Journal of Cleaner Production*, Vol. 388, Article 135981. DOI: <https://doi.org/10.1016/j.jclepro.2023.135981>
-

8. Dachkovskiy, V. O., Sampir, O. M. (2019), "Algorithm of the functioning of the logistics support system", *Modern Information Technologies in the Sphere of Security and Defence*, No. 2 (35), P. 87–92. DOI: <https://doi.org/10.33099/2311-7249/2019-35-2-87-92>
9. Havryliuk, I. Yu., Matsko, O. Y., Dachkovskiy, V. O. (2019), "Conceptual foundations of flow management in the logistics support system of the Armed Forces of Ukraine", *Modern Information Technologies in the Sphere of Security and Defence*, Vol. 34, No. 1, P. 37–44. DOI: <https://doi.org/10.33099/2311-7249/2019-34-1-37-44>
10. Zhang, X., Zou, B., Feng, Z., Wang, Y., Yan, W. (2022), "A review on remanufacturing reverse logistics network design and model optimization", *Processes*, Vol. 10, No. 1, Article 84. DOI: <https://doi.org/10.3390/pr10010084>
11. Abba Dabo, A.-A., Hosseinian-Far, A. (2023), "An integrated methodology for enhancing reverse logistics flows and networks in Industry 5.0", *Logistics*, Vol. 7, No. 4, Article 97. DOI: <https://doi.org/10.3390/logistics7040097>
12. Omosa, G. B., Numfor, S. A., Kosacka-Olejnik, M. (2023), "Modeling a reverse logistics supply chain for end-of-life vehicle recycling risk management: a fuzzy risk analysis approach", *Sustainability*, Vol. 15, No. 3, Article 2142. DOI: <https://doi.org/10.3390/su15032142>
13. Fedorovych, O., Kritskiy, D., Malieiev, L., Rybka, K., Rybka, A. (2024), "Military logistics planning models for enemy targets attack by a swarm of combat drones", *Radioelectronic and Computer Systems*, No. 1, P. 207–216. DOI: <https://doi.org/10.32620/reks.2024.1.16>
14. Burkivskiy, O. S. (2025), "Optimization of military logistics under combat conditions using the A* algorithm", *Natural and Technical Sciences*, Vol. 1, No. 17, P. 168–172. DOI: <https://jvestnik-sss.donnu.edu.ua/article/view/17323>
15. Nabokin, D., Malyeyeva, O., Malieieva, Yu. (2024), "Selection of the most effective transportation route considering territory safety indexes in wartime conditions", *Control, Navigation and Communication Systems*, No. 3, P. 130–134. DOI: <https://doi.org/10.26906/SUNZ.2024.3.141>

Received (Надійшла) 19.02.2026

Accepted for publication (Прийнята до друку) 26.02.2026

Publication date (Дата публікації) 12.03.2026

Відомості про авторів / About the Authors

Полупан Юрій Володимирович – Національний аерокосмічний університет "Харківський авіаційний інститут", аспірант кафедри комп'ютерних наук та інформаційних технологій; Харків, Україна;

Yuriy Polupan – National Aerospace University "Kharkiv Aviation Institute", Postgraduate Student at the Department of Computer Science and Information Technologies; Kharkiv, Ukraine;

e-mail: yuriypolupan6@gmail.com

ORCID ID: <https://orcid.org/0009-0009-3030-8448>

Малєєва Ольга Володимирівна – доктор технічних наук, професор, Національний аерокосмічний університет "Харківський авіаційний інститут", професор кафедри комп'ютерних наук та інформаційних технологій; Харків, Україна;

Olga Malyeyeva – Doctor of Technical Sciences, Professor, National Aerospace University "Kharkiv Aviation Institute", Professor at the Department of Computer Science and Information Technologies; Kharkiv, Ukraine;

e-mail: o.maleyeva@khai.edu

ORCID ID: <https://orcid.org/0000-0002-9336-4182>

Кікоть Максим Сергійович – Національний аерокосмічний університет "Харківський авіаційний інститут", аспірант кафедри комп'ютерних наук та інформаційних технологій; Харків, Україна;

Maksym Kikot – National Aerospace University "Kharkiv Aviation Institute", Postgraduate Student at the Department of Computer Science and Information Technologies; Kharkiv, Ukraine;

e-mail: maxym111111@gmail.com

ORCID ID: <https://orcid.org/0009-0005-7554-9246>

МОДЕЛЮВАННЯ Й АНАЛІЗ ЛОГІСТИЧНИХ МАРШРУТІВ ПОСТАЧАННЯ ТА ЗБУТУ ВИСОКОТЕХНОЛОГІЧНОЇ ПРОДУКЦІЇ З ОГЛЯДУ НА РІЗНІ ВИДИ ТРАНСПОРТУ

Предметом дослідження в статті є процеси організації та оптимізації логістики постачання й збуту високотехнологічної оборонної продукції в умовах воєнного часу, зокрема розроблення мережевої логістичної структури з огляду на зворотні потоки, ризики й обмежені ресурси. **Мета дослідження** – розроблення моделі маршрутів постачання та розподілу, яка оптимізує час і витрати доставки, одночасно знижуючи ризики втрати вантажів і підвищуючи коефіцієнт використання транспорту. **Завдання:** сформулювати узагальнену комбіновану структуру логістичного ланцюга; визначити й проаналізувати основні параметри маршрутів постачання та збуту за різними видами транспорту; розробити план комп'ютерного експерименту; змодельовати шляхи постачання й збуту; запропонувати практичні рекомендації для реалізації логістичних проєктів у воєнний час. **Методи й засоби дослідження:** системний підхід, імітаційне моделювання, графічно-порівняльний аналіз, платформа AnyLogic. **Досягнуті результати.** Схематично подано комбіновану структуру логістичного ланцюга для підприємства з виробництва дронів військового призначення з множиною варіантів шляхів. Запропоновано модель, що формалізує логістичний ланцюг як мережу вузлів і дуг транспортної системи та містить набір вхідних параметрів, що описують транспортні засоби, маршрути й характеристики замовлень. Сутність комп'ютерного експерименту полягає у відтворенні властивостей інфраструктурних і логістичних проєктів за допомогою дискретно-подієвої імітаційної моделі, що дає змогу аналізувати критичні параметри окремо для постачання та розподілу, а також моделювати випадки їх взаємозалежності. Особливу увагу приділено розділенню маршрутів на постачання й розподіл, а також використанню різних видів транспорту. **Висновки.** Досягнуті результати підтверджують, що вирішальним чинником логістичних рішень є збалансоване співвідношення між економічними витратами, рівнем ризику й очікуваним ефектом. Запропонований підхід сприяє кількісному оцінюванню варіантів управлінських рішень у логістичних ланцюгах постачання й збуту виробничих підприємств, що підвищує обґрунтованість вибору оптимальної стратегії розвитку.

Ключові слова: логістика; ланцюг постачання; управління ризиками; оптимізація; комбінована структура; моделювання; альтернативні маршрути; постачання та збут; виробниче підприємство; AnyLogic.

Бібліографічні описи / Bibliographic descriptions

Полупан Ю. В., Малєєва О. В., Кікоть М. С. Моделювання й аналіз логістичних маршрутів постачання та збуту високотехнологічної продукції з огляду на різні види транспорту. *Автоматизовані системи управління та прилади автоматики*. 2026. № 1 (188). С. 191–202. DOI: <https://doi.org/10.30837/0135-1710.2026.188.191>

Polupan, Yu., Mal'yeyeva, O., Kikot, M. (2026), "Modeling and analysis of logistics supply and distribution routes for high-tech products considering different modes of transport", *Management Information System and Devices*, No. 1 (188), P. 191–202. DOI: <https://doi.org/10.30837/0135-1710.2026.188.191>

АЛФАВІТНИЙ ПОКАЖЧИК

	5
Ахундов Р. Г.	54
Божко О. Ю.	5
Гашимов Е. Г.	17
Гребеннік І. В.	163
Дмитревський Д. В.	33
Калита Н. І.	191
Кікоть М. С.	17
Коваленко О. А.	108
Корнієнко Є. Д.	44
Кучук Г. А.	163
Левикін В. М.	98
Лещинська І. О.	108
Ляшенко О. С.	191
Малєєва О. В.	33
Неліпа О. Д.	44
Панченко В. І.	177
Пащенко Р. Е.	54
Петров К. Е.	69
Пилипенко А. Г.	191
Полупан Ю. В.	129
Прокопенко О. С.	145
Репіхов В. М.	129
Смеляков С. В.	78
Ткачов В. М.	121
Усик В. В.	121
Федорчак Б. П.	98
Чалий С. Ф.	145
Чуприна А. С.	5

ALPHABETICAL INDEX

Akhundov Ramil	5
Bozhko Oleksandr	54
Hashimov Elshan	5
Grebennik Igor	17
Dmytrevskyi Dmytro	163
Kalyta Nadiia	33
Kikot Maksym	191
Kovalenko Oleksii	17
Korniienko Yehor	108
Kuchuk Heorhii	44
Levykin Viktor	163
Leshchynska Iryna	98
Liashenko Oleksii	108
Malyeyeva Olga	191
Nelipa Oleksandr	33
Panchenko Volodymyr	44
Pashchenko Ruslan	177
Petrov Konstantin	54
Pylypenko Andrii	69
Polupan Yuriy	191
Prokopenko Oleksandr	129
Repikhov Vadym	145
Smelyakov Sergii	129
Tkachov Vitalii	78
Usik Victoriya	121
Fedorchak Bohdan	121
Chalyi Serhii	98
Chupryna Anastasiya	145

НАУКОВЕ ВИДАННЯ

SCIENTIFIC PUBLICATION

Автоматизовані системи управління та прилади автоматики

Management Information System and Devices

*Всеукраїнський міжвідомчий
науково-технічний збірник*

Випуск 188, 2026

*All-Ukrainian Interdepartmental
Scientific and Technical Collection*

Issue 188, 2026

Відповідальний секретар журналу

О.М. Цимбал

Відповідальний за ліцензування *В.В. Косенко*

Редактор *Л.В. Кузьміна*

Комп'ютерна верстка *Л.В. Кузьміна*

Л.Ю. Светайло

Responsible secretary of journal

O. Tsymbal

Responsible for licensing *V. Kosenko*

Editor *L. Kuzmina*

Computer layout *L. Kuzmina*

L. Svietailo

Формат 60×84/8. Умов. друк. арк. 23,7.
Тираж 150 прим.

Format 60×84/8. Conventional printed sheets 23,7.
Edition of 150 copies.

Віддруковано з готових оригінал-макетів
в типографії ФОП Андреев К.В.
Єдиний державний реєстр юридичних осіб
та фізичних осіб-підприємців.
Запис №24800170000045020 від 30.05.2003.

Printed from ready-made original layouts
in the printing house
of Individual Entrepreneur Andreev K.V.
Unified State Register of Legal Entities
and Individual Entrepreneurs.
Entry No. 24800170000045020 of 30.05.2003.

61157, Харків, вул. Акад. Богомольця, 9, кв. 50,
тел. +38 (063) 993-62-73
e-mail: ep.zakaz@gmail.com

fl. 50, 9, Acad. Bogomolets Str., Kharkiv, 61157,
тел. +38 (063) 993-62-73
e-mail: ep.zakaz@gmail.com