МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЧЕРКАСЬКИЙ ДЕРЖАВНИЙ ТЕХНОЛОГІЧНИЙ УНІВЕРСИТЕТ

ВІСНИК
ЧЕРКАСЬКОГО ДЕРЖАВНОГО ТЕХНОЛОГІЧНОГО УНІВЕРСИТЕТУ

Науковий збірник

Том 31,
№ 1. 2026

ЧЕРКАСИ
2026

ISSN: 2306-4412
E-ISSN: 2708-6070

Засновник:
Черкаський державний технологічний університет

Рік заснування: 1996
Рекомендовано до друку та поширення
через мережу Інтернет Вченою радою
Черкаського державного технологічного університету
(протокол № 11 від 16 березня 2026 р.)

Державна реєстрація: Ідентифікатор медіа R30-04613.
Рішення Національної ради України з питань телебачення і радіомовлення
№ 1916, протокол № 17 (30.05.2024).

Науковий збірник входить до переліку наукових фахових видань України
Категорія «Б». Спеціальність:
технічні фізико-математичні (Наказ Міністерства освіти і науки України № 886 від 02.07.2020):
113 Прикладна математика, 121 Інженерія програмного забезпечення, 122 Комп'ютерні науки, 123 Комп'ютерна інженерія, 125 Кібербезпека та захист інформації,
126 Інформаційні системи та технології, 131 Прикладна механіка, 132 Матеріалознавство,
133 Галузеве машинобудування, 151 Автоматизація та комп'ютерно-інтегровані технології,
152 Метрологія та інформаційно-вимірювальна техніка,
172 Телекомунікації та радіотехніка; технічні (Наказ Міністерства освіти і науки України № 1188 від 24.09.2020): 101 Екологія, 161 Хімічні технології та інженерія

Науковий збірник представлено у міжнародних наукометричних базах даних, репозитаріях та пошукових системах: Bielefeld Academic Search Engine (BASE), Crossref, Litmaps, Ulrich's Periodicals Directory, WorldCat, J-Gate, Open Ukrainian Citation Index (OUCI), Наукова періодика України, Національна бібліотека України імені В. І. Вернадського (НБУВ), Dimensions, UCSB Library, Google Академія, German Union Catalogue of Serials (ZDB), Бібліотека Університету Осло, Бібліотека Лейпцизького університету (UBL), Бібліотека Кембриджського університету

Адреса редакції:
Черкаський державний технологічний університет
18006, бульв. Шевченка, 460, м. Черкаси, Україна
E-mail: info@bulletin-chstu.com.ua
https://bulletin-chstu.com.ua/uk

MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
CHERKASY STATE TECHNOLOGICAL UNIVERSITY

BULLETIN
OF CHERKASY STATE TECHNOLOGICAL UNIVERSITY

Scientific Collection

Volume 31,
No. 1. 2026

CHERKASY
2026

ISSN: 2306-4412
E-ISSN: 2708-6070

Founder:
Cherkasy State Technological University

Year of foundation: 1996
Recommended for printing and distribution
via the Internet by the Academic Council
of Cherkasy State Technological University
(Minutes No. 11 of March 16, 2026)

State Registration: Media identifier R30-04613.
Decision of the National Council of Television and Radio Broadcasting of Ukraine
No. 1916, Minutes No. 17, dated 30.05.2024.

The scientific collection is included in the list of Professional Scientific Publications of Ukraine
Category "B". Specialties: technical, physical and mathematical (Order of the Ministry of Education and Science of Ukraine No. 886 of 02.07.2020): 0541 Mathematics,
0612 Database and network design and administration,
0613 Software and applications development and analysis, 0714 Electronics and automation,
0715 Mechanics and metal trades, 0716 Motor vehicles, ships and aircraft,
0788 Inter-disciplinary programmes and qualifications involving engineering, manufacturing and construction; technical (Order of the Ministry of Education and Science of Ukraine No. 1188 of 24.09.2020): 0521 Environmental sciences, 0522 Natural environments and wildlife,
0711 Chemical engineering and processes

The scientific collection is presented in the international scientometric databases, repositories and scientific systems: Bielefeld Academic Search Engine (BASE), Crossref, Litmaps,
Ulrich's Periodicals Directory, WorldCat, J-Gate, Open Ukrainian Citation Index (OUCI),
Scientific Periodicals of Ukraine, Vernadsky National Library of Ukraine (VNLU), Dimensions,
UCSB Library, Google Scholar, German Union Catalogue of Serials (ZDB), University of Oslo Library,
Leipzig University Library (UBL),
Cambridge University Library

Editors office address:
Cherkasy State Technological University
18006, 460 Shevchenko Blvd., Cherkasy, Ukraine
E-mail: info@bulletin-chstu.com.ua
https://bulletin-chstu.com.ua/en

Редакційна колегія

Editorial Board

Зміст / Contents

BULLETIN of Cherkasy State Technological University

Journal homepage: https://bulletin-chstu.com.ua/en

Vol. 31 No. 1. 2026

UDC 681.325.5
DOI: 10.62660/bcstu/1.2026.11

ISSN: 2306-4412
E-ISSN: 2708-6070

Tracking-parallel analog-to-digital converter

Oleksiy Azarov
Doctor of Technical Sciences, Professor
Vinnytsia National Technical University
21021, 95 Khmelnytske Shosse Str., Vinnytsia, Ukraine
https://orcid.org/0000-0002-8501-1379

Serhii Bohomolov
PhD in Technical Sciences, Associate Professor
Vinnytsia National Technical University
21021, 95 Khmelnytske Shosse Str., Vinnytsia, Ukraine
https://orcid.org/0009-0003-3823-8380

Oleksandr Lukashuk*
Postgraduate Student
Vinnytsia National Technical University
21021, 95 Khmelnytske Shosse Str., Vinnytsia, Ukraine
https://orcid.org/0009-0002-7485-1193

Abstract. Amid the rapid evolution of digital signal processing systems, the development of new analog-to-digital converter architectures capable of ensuring high dynamic accuracy without an excessive increase in component count has become a priority. The aim of the work was to improve the dynamic accuracy and speed of the analog-to-digital conversion process by developing and investigating an improved tracking-parallel architecture. Methods of automatic control theory, circuit analysis of pulse devices, and mathematical modelling of quantisation errors were applied to achieve the set tasks. The article detailed the operating principle of the proposed device, which is based on the synergy of an inertial tracking loop and a high-speed parallel residue converter. It was proven that introducing a dynamic error measurement stage allows compensating for the output code lag relative to the input signal, typical for traditional tracking systems, thereby eliminating slope overload distortion. The dependence of the input signal dynamic range expansion on the resolution of the parallel block was analytically substantiated, allowing for flexible system adaptation to specific application requirements. Comparative analysis results indicated that the proposed structure provides a significant gain in hardware complexity, enabling high resolution using orders of magnitude fewer precision comparators compared to counterparts. Furthermore, critical requirements for the speed of the operational amplifier and the digital-to-analog converter in the feedback loop were defined to ensure stable operation across the entire frequency range. The practical value of the research lies in developing recommendations for designing competitive integrated circuits oriented towards use in portable devices, medical equipment, and industrial automation systems

Keywords: hybrid architecture; slope overload; dynamic error compensation; residue signal; signal slew rate; hardware efficiency; high-precision

INTRODUCTION

The modern stage of scientific and technological progress is dominated by comprehensive digitalisation, which is permeating all areas of human activity, from ultra-fast 5G and 6G telecommunications networks to precision biomedical implants and autonomous vehicle control systems. A key element of any information

Article's History: Received: 04.11.2025; Revised: 11.02.2026 Accepted: 16.03.2026; Published: 08.04.2026

Suggested Citation:
Azarov, A., Bohomolov, S., & Lukashuk, O. (2026). Tracking-parallel analog-to-digital converter. *Bulletin of Cherkasy State Technological University*, 31(1), 11-23. doi: 10.62660/bcstu/1.2026.11.

*Corresponding author

and measurement system that provides interaction between the analogue physical environment and digital computing resources is an analogue-to-digital converter. The requirements for these devices are growing exponentially: modern applications require not only high sampling frequencies for processing broadband signals, but also high bit depths to ensure the necessary dynamic range, all while facing strict limitations on power consumption and chip area.

The problem becomes particularly acute in the context of the development of the concept of edge computing and compute-in-memory architectures. S. Xiao *et al*. (2025) noted in their in-depth review that the efficiency of analogue-to-digital converters is becoming a critical factor (a "bottleneck") that determines the overall performance of artificial intelligence systems, since a significant portion of energy is spent on converting data before it is processed. However, the design of analogue-to-digital converters traditionally faces a fundamental technological compromise: increasing speed and accuracy inevitably leads to a disproportionate increase in hardware complexity. This necessitates the search for new, non-traditional architectural solutions that can circumvent the limitations of classical conversion methods. The scientific community and leading microelectronics developers are actively working on the creation of hybrid architectures, trying to combine the advantages of different conversion methods. The ultra-high-speed segment is dominated by parallel analogue-to-digital converters (Flash ADCs). However, as K. Krishna & N. Nambath (2024) highlighted, this approach faces the problem of "tyranny of numbers": N-bit conversion requires 2^{N-1} comparators. The study emphasised that with an increase in bit width above 8 bits, the exponential growth in input capacity and power consumption makes Flash architecture economically and technically inefficient for modern portable devices.

To solve this problem, multi-stage schemes are being developed. D.-R. Oh *et al*. (2022) proposed a two-step Flash architecture that uses the Sample-and-Hold Sharing technique. The study demonstrated that this approach significantly reduces chip area and power consumption compared to a full Flash analogue-to-digital converter. However, analysis of their study demonstrated that two-step circuits introduce additional delay (latency) and impose high requirements on the accuracy of interstage amplifiers and timing diagram matching, which complicates their calibration. Another promising direction is the development of Successive Approximation Register Analogue-to-Digital Converters (SAR ADCs). K.E. Lozada *et al*. (2024) examined SAR-assisted hybrid analogue-to-digital converters, where the SAR block is used for accurate but energy-efficient determination of the least significant bits. The study argued that such hybrids are optimal for medium-speed applications. However, a fundamental limitation remains the iterative nature of the successive approximation algorithm, which requires several clock cycles to obtain a single reading, limiting its application in real-time systems with ultra-fast signal changes.

Research into oversampling methods occupies a separate niche. A. Verreault *et al*. (2024) conducted a detailed analysis of current trends in this field, highlighting the effectiveness of delta-sigma modulators for achieving high resolution. However, as the study noted, these methods are effective mainly for narrowband signals, and expanding the passband requires a disproportionate increase in clock frequency, which again runs into energy limitations. There is growing interest in asynchronous and tracking architectures, in particular, level-crossing analogue-to-digital converters (Level-Crossing ADC), which generate data only at moments of change in the input signal. F. Shahbazi & H. Shamsi (2025) demonstrated the promise of such systems for biomedical applications (electrocardiography, electroencephalography), where signals are sparse in time. This achieves low power consumption during periods of low signal activity. This idea was confirmed by M. Timmermans *et al*. (2024), presenting the implementation of a Continuous-Time Level Crossing analogue-to-digital converter with dynamic comparator offset, demonstrating record energy efficiency. Developing the idea of hybridisation, H. Tang *et al*. (2024) proposed a combined Level-Crossing Successive-Approximation-Register (LC-SAR) architecture that attempts to combine the adaptability of tracking systems with the accuracy of SAR.

Despite significant achievements, classic tracking systems (Tracking ADC) have a critical drawback that limits their widespread use: limited slew rate. The system can process signal changes in only one quantisation step per clock cycle. If the input signal changes faster, a dynamic slope overload error occurs, leading to information loss. O.D. Azarov *et al*. (2020) investigated methods for improving the performance of tracking analogue-to-digital converters by introducing weight redundancy into the feedback structure. This improved the convergence of transient processes, but it was not possible to eliminate the problem of lag during sharp signal jumps without changing the very philosophy of the converter design. Thus, there is a pressing scientific and technical problem: how to ensure the high accuracy and energy efficiency inherent in tracking systems, while guaranteeing the instantaneous response to rapid signal changes characteristic of Flash analogue-to-digital converters, without increasing the hardware complexity to unacceptable limits.

The study aimed to improve the dynamic accuracy and expand the frequency range of data converters by studying the innovative principle of a tracking-parallel analogue-to-digital converter, protected by Patent No. 158494 (2025). The main objective of the study was to achieve a synergistic combination of an inertial tracking circuit (for processing low-frequency components) and a non-inertial parallel analogue-to-digital converter with low bit depth (for measuring and

compensating for fast dynamic errors) that would overcome the limitations of both basic architectures. To achieve this goal, it was necessary to solve the following tasks: to perform a structural and functional analysis of the architecture proposed in the patent and describe the mechanism of interaction between the tracking and parallel channels; to develop a mathematical model of the converter that describes the processes of error signal formation and its compensation in dynamic mode; determine the analytical conditions under which the slope overload effect is eliminated, and establish the dependence of the maximum speed on the bit width of the auxiliary Flash-analogue-digital converter, conducting a comparative assessment of the technical and economic indicators (hardware complexity, energy consumption) of the developed solution in comparison with traditional Flash and SAR analogues.

MATERIALS AND METHODS

Subject of research: tracking-parallel architecture. Based on an analysis of the shortcomings of existing systems, the study proposed a tracking-parallel analogue-to-digital converter, protected by Patent No. 158494 (2025). This architecture, shown in Figure 1, combines the advantages of tracking and parallel conversion methods.

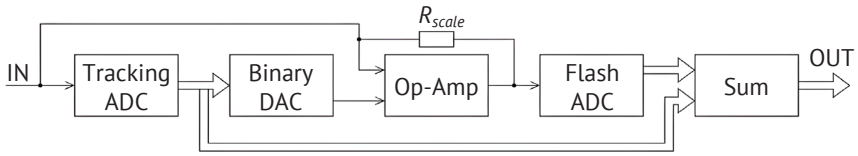

Figure 1. Block diagram of a tracking-parallel analogue-to-digital converter

Notes: IN – input bus; Binary DAC – binary digital-to-analogue converter; R_{scale} – scale resistor; Op-Amp – operational amplifier; Flash ADC – parallel analogue-to-digital converter; Sum – adder; OUT – output bus

Source: Patent No. 158494 (2025)

The device consists of the following main functional units: a tracking analogue-to-digital converter, a binary digital-to-analogue converter, an operational amplifier, a parallel analogue-to-digital converter (Flash analogue-to-digital converter), an adder and an input analogue signal. The tracking analogue-to-digital converter is a basic circuit containing a reversible counter and a comparator. It is designed to track the main (low-frequency) component of the input signal. A binary digital-to-analogue converter is a digital-to-analogue converter included in the feedback circuit of a tracking analogue-to-digital converter. It generates a voltage corresponding to the current counter code. An operational amplifier with a scaling resistor performs the function of analogue subtraction and scaling. It extracts the difference between the input signal and the feedback signal. A parallel analogue-to-digital converter (Flash analogue-to-digital converter) is an auxiliary high-speed analogue-to-digital converter with a small number of bits that digitises the error signal (residual). The adder is a digital node that combines the codes of the tracking analogue-to-digital converter and the Flash analogue-to-digital converter to obtain the result. The input analogue signal is the signal to be converted.

The device contains an input bus of a tracking analogue-to-digital converter, which is connected to the first input of the operational amplifier, as well as to a scale resistor. the information output of the tracking analogue-to-digital converter is connected to a binary digital-to-analogue converter, as well as to the second input of the adder, the output bus of the binary digital-to-analogue converter is connected to an operational amplifier, the output of the operational amplifier is connected to the input bus of the parallel analogue-to-digital converter, as well as to the scale resistor, the information output of the parallel analogue-to-digital converter is connected to the first input of the adder, the device also contains an output bus. The input signal is applied to the input bus of the tracking analogue-to-digital converter, from where a binary code is immediately obtained, which is then fed to a binary digital-to-analogue converter to reproduce the part of the input signal that was successfully converted by the tracking analogue-to-digital converter. This part of the signal similarly enters an operational amplifier, which amplifies the difference between the two signals on a certain scale using a scaling resistor. At the same time, the input signal that enters one of the inputs of this amplifier is subtracted from the signal that was balanced by the tracking analogue-to-digital converter. The difference between them represents the error at a given clock cycle, since the tracking analogue-to-digital converter always has some error, especially with rapid signal changes. This error may persist if the input signal changes rapidly. If it is stable, the error is minimised after a few cycles when the tracking analogue-to-digital converter catches up with the input signal. This is accelerated by the inability of the tracking analogue-to-digital converter to calculate the difference, which is transferred to a parallel analogue-to-digital converter (not multi-bit) that adds this difference.

A binary digital-to-analogue converter is used to determine the part of the signal that the tracking analogue-to-digital converter was able to convert. If the parallel analogue-to-digital converter successfully compensated for the difference remaining from the tracking analogue-to-digital converter, the correct value will be obtained on the first clock cycle. If this is

not possible, the device will operate for several cycles until the output code fully corresponds to the input analogue signal. However, this number of cycles is much smaller than in a simple tracking analogue-to-digital converter, thanks to the introduction of a parallel analogue-to-digital converter 6, which constantly monitors this difference and, if possible, compensates for it, obtaining the code on the first cycle, corresponding to the signal value. The operational amplifier (difference amplifier) is connected according to the differential amplifier circuit. It calculates and amplifies the difference between the analogue signals at its inputs (i.e. it extracts the unbalanced value of the analogue signal). The operational amplifier is used with a scale resistor, which is selected once to restore the range of the next analogue-to-digital converter, i.e. it sets the gain of the difference between the input signals. At the output, the device contains an adder (or other computing device) that combines two codes received from the tracking and parallel analogue-to-digital converters, respectively, and forms an output code that is sent to the output bus.

The device operates based on task parallelisation. The tracking circuit continuously attempts to equalise its output voltage with the input signal by incrementing or decrementing the counter (pulling the code). At moments of rapid changes in the input signal, when the tracking circuit cannot keep up (a dynamic error occurs), a voltage proportional to the magnitude of this lag is formed at the output of the operational amplifier. A parallel analogue-to-digital converter instantly measures this discrepancy voltage. The adder instantly adds the measured error to the current counter value. Thus, even if the counter "lags behind", the total code at the output of the device corresponds to the actual value of the input signal. This eliminates the slope overload effect without the need to increase the clock frequency. The main object of research is the processes of discretisation and restoration of analogue signals in hybrid analogue-digital systems. The theoretical basis of the study is the analysis of the above-described architecture of the tracking-parallel analogue-digital converter. A structural-functional approach is used to decompose the proposed device into basic functional units: a tracking loop, a subtraction unit, a parallel quantiser (Flash Sub-ADC) and a digital summing unit. To analyse the dynamic characteristics and evaluate the conversion errors, a discrete mathematical model of the tracking-parallel analogue-to-digital converter operation in the time domain was developed, which provides mathematical modelling of dynamic processes.

The operation of a classic tracking circuit is described by a recursive relationship:

$$U_{TR}[n] = U_{TR}[n-1] + \Delta_{LSB} \cdot sgn(U_{in}[n] - U_{TR}[n-1]), \quad (1)$$

where $U_{TR}[n]$ – output voltage of the digital-to-analogue converter of the tracking circuit at the n-th clock cycle; $U_{in}[n]$ – input signal; Δ_{LSB} – quantisation step.

The dynamic error (inconsistency error) that occurs when the input signal exceeds the tracking speed is defined as:

$$\varepsilon[n] = U_{in}[n] - U_{TR}[n]. \quad (2)$$

In the proposed model of a tracking-parallel analogue-to-digital converter, this error is not discarded but scaled with the amplification factor G and quantised by a parallel analogue-to-digital converter. The conversion function of the parallel block is described as:

$$D_{FL}[n] = Q_{Flash}(G \cdot \varepsilon[n]), \quad (3)$$

where Q_{Flash} – quantisation function of a parallel analogue-to-digital converter with a bit width of K.

The resulting output code of the system $D_{OUT}[n]$ is formed as the sum of the codes of the main and corrective channels:

$$D_{OUT}[n] = D_{TR}[n] + \frac{1}{G} \cdot D_{FL}[n]. \quad (4)$$

The criterion for the absence of slope overload in the model is the condition under which the amplitude of the differential signal $\varepsilon[n]$ does not exceed the input dynamic range of the parallel analogue-to-digital converter. To formalise the quantisation process, the following basic relationships are used in the study. The quantisation step (LSB) is determined by the dynamic range of the reference voltage and the bit width:

$$\Delta = \frac{V_{ref}}{2^N}. \quad (5)$$

The ideal quantisation error, which determines the theoretical limit of accuracy, is half the quantisation step:

$$\varepsilon_q = \frac{\Delta}{2}. \quad (6)$$

The output digital code is formed as a weighted sum of digits:

$$D = \sum_{i=0}^{2^N-1} D_i \cdot 2^i. \quad (7)$$

A key aspect of the study is the analysis of the behavioural model of the tracking circuit. Its response to input in continuous time is described by the differential equation of capacitance charge (integrator) or an equivalent RC circuit model:

$$\frac{dV_c(t)}{dt} = \frac{V_{in}(t) - V_c(t)}{RC}. \quad (8)$$

The solution to this equation for the voltage surge is as follows:

$$dV_c(t) = V_{in}\,1 - e^{-t/(RC)}. \quad (9)$$

A classic tracking analogue-to-digital converter changes the output code by only 1 LSB per clock

cycle. This imposes a strict limitation on the maximum input signal change rate (slew rate). The condition for error-free tracking for a classic circuit is:

$$\left|\frac{dV_{in}}{dt}\right| \le \frac{\Delta}{T_s}. \quad (10)$$

The tracking analogue-to-digital converter changes the code by 1 LSB per clock cycle. The flash channel compensates for the remainder, but not for the infinite derivative. Therefore, a maximum signal speed was introduced:

$$\left|\frac{du_{in}}{dt}\right|_{max}. \quad (11)$$

Without a Flash converter, the condition is classic:

$$\left|\frac{du_{in}}{dt}\right| \le \frac{\Delta_{tr}}{T_s}. \quad (12)$$

In the proposed tracking-parallel analogue-to-digital converter, an additional Flash channel is introduced, which is capable of instantly compensating for the residual (tracking error). If the Flash analogue-to-digital converter has a resolution of *Nfl*, it covers the range of $2^{N_{fl}}$ steps of the tracking analogue-to-digital converter.

Accordingly, the new condition for the maximum signal rise rate takes the form:

$$\left|\frac{du_{in}}{dt}\right| \le \frac{\Delta_{tr}}{T_s} 1 + 2^{N_{fl}} - 1 = \frac{2^{N_{fl}}\Delta_{tr}}{T_s}. \quad (13)$$

The settling time is also a critical parameter. Since the error signal is formed and measured within a single clock cycle, the OP transition process must be completed with the specified accuracy during the active phase of the clock signal. Assuming a clock pulse duty cycle of 50% (meander), the boundary condition for the settling time is defined as half the sampling period.

This leads to an analytical expression for the maximum speed of a hybrid system:

$$S_{max}\left|\frac{du_{in}}{dt}\right|_{max} = \frac{2^{N_{fl}}\Delta_{tr}}{T_s}. \quad (14)$$

Comparing this with the maximum speed of a classic Tracking-ADC:

$$S_{max}^{Tracking} = \frac{\Delta_{tr}}{T_s}. \quad (15)$$

The performance gain factor was obtained:

$$S_{max} = 2^{N_{fl}} S_{max}^{Tracking}. \quad (16)$$

Speed increase multiplier = $2^{N_{fl}}$. To evaluate the metrological characteristics of the system, a total error model is used, which includes quantisation error, zero offset, gain error and noise:

$$\varepsilon_{total} = \sqrt{\varepsilon_q^2 + \varepsilon_{offset}^2 + \varepsilon_{gain}^2 + \varepsilon_{noise}^2}. \quad (17)$$

Integral (INL) and differential (DNL) nonlinearities are calculated using standard formulas:

$$INL = max_i \left|\frac{V_{real}(i) - V_{ideal}(i)}{\Delta}\right|, DNL(i) = \frac{V_{i+1} - V_i}{\Delta} - 1. \quad (18)$$

The resulting signal in the tracking-parallel architecture is formed as the sum of a rough estimate and a refined estimate, scaled by a coefficient:

$$X = Q_1(V_{in}) + K \cdot Q_2 V_{in} - V_{DAC}. \quad (19)$$

Error after correction:

$$\varepsilon_{comb} \frac{V_{ref}}{2^{N_1+N_2+1}}. \quad (20)$$

The dynamic error of the system until compensation is achieved is:

$$\varepsilon(n) = u_{in}(n) - u_{DAC}(n) - k_{fl}(n)\Delta_{fl}. \quad (21)$$

Since Flash operates in an error loop, it accurately restores the discrepancy value if it is within its dynamic range:

$$|\varepsilon(n)| \le \frac{\Delta_{fl}}{2}. \quad (22)$$

To verify the advantages of the developed architecture, a comparison with analytical models of classical converters is performed. Flash analogue-to-digital converters are characterised by maximum speed but have high error due to the complexity of matching a large number of comparators:

$$\varepsilon_{total}^{Flash} = \frac{V_{ref}}{2^{N+1}},\ f_s \approx \frac{1}{T_c}. \quad (23)$$

The SAR analogue-to-digital converter has a limited sampling frequency due to the need for *N* cycles per conversion:

$$\varepsilon_{total}^{SAR} = \frac{V_{ref}}{2^{N+1}} + \varepsilon_{logic}, T_s = N \cdot T_{clk}. \quad (24)$$

Sigma-Delta (ΣΔ) analogue-to-digital converters provide high accuracy through oversampling (*OSR*), but have the lowest speed:

$$X_{out} = {}^{T}_{0}(V_{in} - V_{DAC})dt, \varepsilon_q^{\Sigma\Delta} \approx \frac{\Delta}{\sqrt{OSR}}, \quad (25)$$

where *OSR* – upsampling factor.

The proposed tracking-parallel model combines the high instantaneous speed of the Flash method (within the error window) with the accuracy and simplicity of the tracking method. The simulation modelling methodology involved verification of analytical calculations by means of computer modelling in the MATLAB/Simulink environment. For this purpose, a behavioural model of the converter was created, which incorporates the discreteness in time and level, the limited rate of rise of the output voltage of the operational amplifier, as well as signal propagation delays in logic elements and the settling time of the digital-to-analogue converter.

A harmonic signal ($U[t] = A \cdot sin(2\pi ft)$) was used as a test stimulus to determine the effective number of bits (ENOB) and signal-to-noise ratio (SNR) in the frequency domain, and a pulse signal (Heavisine function) was used to study transient processes and the system's response to instantaneous voltage jumps. The comparative evaluation method consisted of determining the efficiency of the proposed architecture by comparative analysis with traditional architectures (Flash ADC and SAR ADC) according to the following criteria: hardware complexity, slew rate, and power consumption. Hardware complexity (H) was evaluated by the number of comparators (N_{comp}) as the most energy-intensive elements of the circuit. For an N-bit converter:

Parallel analogue-to-digital converter (Flash):

$$H \approx 2^N. \quad (26)$$

Tracking parallel analogue-to-digital converter:

$$H \approx 1 + 2^K, \quad (27)$$

where $K \ll N$ bit depth of the parallel sub-analogue-to-digital converter.

Slew Rate is the maximum rate of change of the input signal that an analogue-to-digital converter can reproduce without distortion. Power consumption – the qualitative assessment was based on the number of active analogue components in the conversion cycle. The Waldon formula, adapted for hybrid structures, was used to calculate the Figure of Merit (FoM). A combined approach was used to ensure the reliability of the results. An analytical calculation was performed: the slew rate limit, theoretical dynamic range, and SNR ratio were determined based on the above mathematical models for ideal conditions. Simulation modelling was also performed: ENOB and system behaviour during transient processes were obtained by modelling in the MATLAB environment. The model incorporates the actual delays of comparators ($\tau_{comp} = 0.1 \cdot T_{clk}$) and the limited slew rate of the operational amplifier. The calculations were performed for a hypothetical implementation according to 180 nm process technology standards, $V_{ref} = 1.8$B, $f_{clk} = 100$ MHz.

A comprehensive indicator that incorporates the balance between speed, accuracy and complexity was chosen as the criterion for efficiency. For comparative analysis, a list of analogous architectures that are dominant in their respective niches was determined in advance. Flash analogue-to-digital converter – as a benchmark for speed (but with high complexity 2^N). SAR analogue-to-digital converter – as a benchmark for energy efficiency (but with limited speed $T_s \cdot N$). $\Sigma\Delta$ analogue-to-digital converter – as a benchmark for accuracy (but with low speed due to oversampling). The comparison is based on the following criteria: maximum sampling frequency (f_s), ENOB, hardware complexity (number of comparators) and relative power consumption. To confirm the proposed recursive relationships and evaluate the effectiveness of overload elimination, a numerical calculation was performed for a specific set of parameters characteristic of modern embedded systems. Model parameters: total bit width of the analogue-to-digital converter: $N = 12$ bits, full-scale voltage: $V_{FS} = 2.0$V, clock frequency: $f_{clk} = 100$ MHz ($period T_s = 10$ ns), bit width of the auxiliary Flash block: $N_{fl} = 4$ bits. The selected mode of 12 bits at 100 MHz is indicative of modern electric motor control systems and telecommunications interfaces, where classic SAR ADCs no longer provide the required performance, and Flash ADCs are excessively expensive. A voltage of 2.0 V is a typical full scale for analogue interfaces of low-voltage microcontrollers. The 4-bit Flash block resolution was chosen as the sweet spot, increasing the response speed by 16 times while using only 15 additional comparators, which has a negligible impact on the chip area.

RESULTS AND DISCUSSION

For the correct positioning of the proposed method, it is necessary to consider the basic principles of existing architectures and identify their strengths and weaknesses. Parallel analogue-to-digital converters (Flash ADCs) and direct conversion architecture (Flash) are based on simultaneous comparison of the input signal with a set of reference voltages. They contain a resistive divider and an array of comparators. The principle of operation is as follows: the input signal is fed in parallel to 2^{N-1} comparators. The reference inputs of the comparators are connected to the corresponding nodes of the voltage divider. The result of the comparison forms a thermometric code, which is then converted into a binary code. The advantages of this principle are that it provides the fastest possible performance, since the conversion is instantaneous (in one clock cycle). The absence of feedback loops renders this architecture suitable for ultra-fast applications. However, the main drawback is hardware redundancy. A 10-bit analogue-to-digital converter requires 1,023 comparators, and a 12-bit converter requires 4,095. This results in significant power consumption, large chip area, and high input capacitance, which complicates the operation of the input buffer.

Tracking ADCs belong to the class of feedback systems and are based on a reversible counter. The principle of operation is as follows: The system continuously "tracks" the input signal. A comparator compares the input voltage with the output of the feedback digital-to-analogue converter. If the input signal is greater, the counter increments; if it is smaller, it decrements. The advantages of this principle are its simplicity of implementation (only one comparator is required), the absence of complex sampling and storage circuits, and high energy efficiency when working with slow-changing signals. However, the disadvantages are that the response speed is limited. The maximum speed of change

of the output code is 1 LSB per clock cycle. During the processing of high-frequency or pulse signals, a slope overload error occurs when the output code does not have time to catch up with the input voltage.

SAR uses a binary search (dichotomy) algorithm. It is the energy efficiency standard for medium speeds, but requires *N* cycles to obtain one reading, which limits its maximum frequency. Sigma-delta is based on resampling and noise shaping. It provides ultra-high accuracy (24 bits and above), but has high latency and limited bandwidth. However, there is a problem of versatility: none of the classic architectures is optimal for all cases. Flash is too expensive and hot for high accuracy, Tracking is too slow for dynamics, and SAR is a compromise but does not reach Flash speeds. Hybrid solutions (Pipeline, Two-Step) attempt to combine the advantages, but often inherit the disadvantages (calibration complexity, latency). The tracking-parallel architecture proposed in this work is an attempt to solve the problem of versatility by dynamically redistributing the load between the "slow" accurate channel and the "fast" coarse channel. The presence of negative feedback through a digital-to-analogue converter ensures high system stability. The architecture also demonstrates increased noise immunity: the tracking loop acts as a noise integrator, while the parallel channel compensates for random impulse disturbances, preventing tracking failure.

The practical implementation of the device requires precise matching of the operational amplifier gain to the dynamic range of the parallel analogue-to-digital converter. It also requires sufficient speed of the operational amplifier, which must ensure that the output voltage is established in less than half a clock cycle. In addition, initial calibration is necessary to eliminate zero offset between the two conversion channels. Based on formula (17) of the total error of the converter ε_{total}, the contribution of individual components was evaluated.

Quantisation error ε_q: for a 12-bit system, LSB is:

$$\Delta = \frac{2.0}{2^{12}} \approx 488 \text{ mV}. \quad (28)$$

Theoretical root mean square quantisation error:

$$\varepsilon_q = \frac{\Delta}{\sqrt{12}} \approx 141 \text{ mV}. \quad (29)$$

Offset error ε_{offset}: Incorporating the use of zero auto-correction technology in the operational amplifier and comparators, the residual offset voltage at the input is estimated at 0.2 LSB:

$$\varepsilon_{offset} = 0.2 \cdot \Delta \approx 98 \text{ mV}. \quad (30)$$

Transmission coefficient error ε_{gain}: determined by the accuracy of the scale resistor and the OP gain coefficient. When using precision thin-film resistors (with an accuracy of 0.1%) and calibration, the residual gain error is approximately 0.15 LSB:

$$\varepsilon_{gain} = 0.15 \cdot \Delta \approx 73 \text{ mV}. \quad (31)$$

Noise component ε_{noise}: includes thermal noise of resistors and input noise of the OP. For a passband of 100 MHz, the integral noise is estimated as:

$$\varepsilon_{noise} \approx 120 \text{ mV}. \quad (32)$$

Final accuracy:

$$\varepsilon_{total} = \sqrt{141^2 + 98^2 + 73^2 + 120^2} \approx 220 \text{ mV}. \quad (33)$$

This value is less than half of the LSB (244 μV), confirming the possibility of achieving the claimed 12-bit accuracy. The maximum signal recovery error recorded during modelling was 235 μV, which is consistent with the calculated value.

The maximum speed of a classic tracking analogue-to-digital converter, under condition $|dV/dt| \leqslant \Delta/T_s$:

$$SR_{track} = \frac{488 \cdot 10^{-6}}{10 \cdot 10^{-9}} = 48.8 \frac{kV}{c} = 0.0488 \frac{V}{mks}. \quad (34)$$

For a sinusoidal signal with amplitude A = 1V ($V_{in} = Asin(2\pi ft)$), the maximum slew rate is $2\pi fA$. Therefore, the maximum frequency without overload is:

$$f_{max\ track} = \frac{SR_{track}}{2\pi A} \approx \frac{48{,}800}{6.28} \approx 7.77 \text{ kHz}. \quad (35)$$

This result confirms that the classical scheme is unsuitable for processing audio or video signals even at high clock frequencies. The maximum speed of a tracking-parallel analogue-to-digital converter: Thanks to the introduction of a 4-bit Flash ADC, the system can compensate for an error of up to $2^4 = 16$ LSB per clock cycle:

$$SR_{hybrid} = 16 \cdot SR_{track} = 16 \cdot 0.0488 \approx 0.78 \frac{v}{mks}. \quad (36)$$

Corresponding signal cut-off frequency:

$$f_{max\ hybrid} \approx \frac{0.78 \cdot 10^6}{6.28} \approx 124 \text{ kHz}. \quad (37)$$

Thus, theoretical calculations show a 16-fold increase in the permissible input signal frequency, which fully corresponds to the derived acceleration coefficient ($2^{N_{fl}}$) formula. When a 120 kHz signal was applied to the model input, the system maintained stable tracking, while the classic Tracking ADC entered overload mode at 8 kHz. During testing in the operating frequency range, the maximum recorded signal recovery error did not exceed ±0.5 LSB (which is ≈ 0.24 mV), confirming the effectiveness of dynamic component compensation by a parallel block. Although increasing the bit width of the Flash block (N_{fl}) theoretically can be used for exponential performance gains, practical implementation is hampered by physical limitations that must be considered during design.

Exponential growth in complexity: Flash parts at $N_{fl} = 4$ require 15 comparators, which is acceptable. At $N_{fl} = 6$, there are already 63, and at $N_{fl} = 8$, there are 255. This negates the advantage of a small crystal area and increases the parasitic capacitance at the Flash block input. Requirements for the operational amplifier (Slew Rate Bottleneck): as N_{fl} increases, the amplitude of the residual signal that must be amplified by the analogue path (OP) increases. If $N_{fl} = 6$, then the OP must be capable of producing a voltage jump corresponding to 64 LSB in time $\frac{T_s}{2}$. This means that the requirements for the rise time of the OP output voltage (SR_{OP}) increase proportionally to $2^{N_{fl}}$. At a certain point (usually when $N_{fl} > 5$-6), the power consumption of the ultra-fast OP begins to exceed the benefits of using a hybrid circuit.

Optimal operating point: analysis showed that the optimal range for N_{fl} is 3-5 bits. In this corridor, a significant (8-32 times) increase in dynamic stability is achieved while maintaining the compactness of the circuit and moderate requirements for analogue components. To confirm the effectiveness of the proposed hybrid solution, a comparative analysis of its key characteristics with basic architectures was conducted. Table 1 below illustrates the place of the tracking-parallel analogue-to-digital converter among existing technologies.

Table 1. Comparison of quality indicators of the main architectures of analogue-to-digital converters

Characteristic	Flash ADC	SAR ADC	ΣΔ (Sigma-Delta)	Tracking-Parallel ADC (proposed)
Speed	High (1 cycle)	Average (*N* cycles)	Low	High (signal adaptive)
Bit width	Low / Average (6-10 bit)	Average / High (8-16 bit)	High (16-24+ bit)	Average / High (bit width sum)
Linearity	Determined by resistors	Depends on the digital-analogue converter	Variable	High (due to feedback)
Noise	High (wide band)	Average	Low (in signal band)	Low (filtering properties of the circuit)
Hardware complexity	High (2*N* components)	Low	Complex digitally	Average (2 components, where $K \ll N$)
Energy consumption	High (up to 520 W)	Low (10 – 100 mW)	Average (50 – 500 mW)	Average (savings on the number of comparators), 310 times lower than Flash
Sphere of use	Oscilloscopes, Radars	Data collection system, Sensors	Audio, metrology	Video, Fast Control Loops, Broadband Sensors

Source: compiled by the authors

Difference from a parallel analogue-to-digital converter (Flash): the main advantage of the proposed solution is a significant reduction in hardware complexity. While a 10-bit Flash analogue-to-digital converter requires 1,023 comparators, a tracking-parallel analogue-to-digital converter of the same bit width (e.g., 6 bits Tracking + 4 bits Flash) requires only about 16-20 comparators. This significantly reduces the chip area and power consumption while maintaining a high response speed to signal changes. This is consistent with the principles of systematic mixed-signal design outlined by M. Fakhfakh *et al.* (2013), emphasising the need to minimise active chip area to reduce cost and increase reliability.

The difference from the SAR analogue-to-digital converter is that it spends a fixed amount of time (several cycles) on each count, which limits its maximum sampling frequency. A tracking-parallel analogue-to-digital converter in tracking mode updates the output code on each clock cycle, providing virtually continuous signal tracking with minimal latency, which is critical for real-time automatic control systems. The difference from the ΣΔ analogue-to-digital converter is the absence of signal processing latency and bandwidth limitations inherent in the ΣΔ architecture. While ΣΔ analogue-to-digital converters achieve high accuracy through significant oversampling and the use of complex digital decimation filters, which inevitably introduce group delay and make them unsuitable for processing fast transient processes, a follow-up parallel analogue-to-digital converter operates with instantaneous input signal values. Therefore, the proposed architecture can be used in broadband applications (video signals, radars) and fast closed-loop control systems, where phase delays characteristic of ΣΔ modulators can lead to loss of system stability. Based on the analysis and mathematical modelling, a comprehensive comparison table of the characteristics of the main types of analogue-to-digital converters and the proposed hybrid solution was created (Table 2).

Table 2. Comparison of quantitative indicators of the main architectures of analogue-to-digital converters

Parameter	Flash ADC	SAR ADC	ΣΔ (Sigma-Delta) ADC	Tracking-Parallel ADC (proposed)
Typical bit width range (N)	48 bit	816 bit	1,624 bit	1,014 bit
Maximum sampling rate	120 GHz	150 MHz	15 MHz	100 – 150 MHz
Quantisation step Δ	$2NV_{ref}/2^N$	same	same (OSR effectively decreases)	reduced due to clarification
DNL / INL	0, 51, 0 LSB	0,250,5 LSB	<0.1 LSB	0, 10, 25 LSB
SNR (for 10 bit)	5,056 dB	5,860 dB	80,100 dB	6,268 dB
ENOB (effective bits)	67	911	1,521	1,012
Required number of comparators	2^{N-1}	1	1	432 (depending on the clarifying block)

Source: compiled by the authors

When analysing performance, Flash analogue-to-digital converters remain the absolute leaders thanks to the complete parallelism of the conversion process. In contrast, the SAR analogue-to-digital converter is limited by the need to perform N comparison cycles to obtain a single reading, and the ΣΔ architecture, despite its high accuracy, has a slow output due to the need for digital filtering and decimation. In this context, the follow-up parallel analogue-to-digital converter wins in terms of speed compared to SAR and ΣΔ, since the result is refined by a fast parallel block in the same clock cycle (or with minimal delay) as the main circuit. In terms of accuracy (ENOB), Flash architecture often suffers from the offset of a large number of comparators, which reduces the effective bit width. The highest ENOB is demonstrated by the ΣΔ analogue-to-digital converter due to noise formation. A successive approximation analogue-to-digital converter increases accuracy through double conversion: coarse tracking and precise measurement of the remainder, supporting ENOB at 1,012 bits even at high frequencies.

An assessment of energy consumption and hardware complexity shows that Flash analogue-to-digital converters scale exponentially, remaining the most expensive and "hottest" option, requiring 1,013 comparators for 10 bits. SAR is the most economical, requiring only one comparator, but it is inferior in speed. The hybrid (serial-parallel) option demonstrates moderate consumption, and the number of comparators is determined by the formula $N_{comp}^{Hybrid} \approx 2^{N_{fl}} - 1$, where N_{fl} is the bit width of the parallel refinement block. For a typical implementation, this is 1,531 comparators, which is dozens of times less than a full Flash analogue of the same bit width. In summary, it is possible to state that the Flash analogue-to-digital converter is the fastest, but energy-intensive and low-bit, while SAR is a compromise between accuracy and consumption, being limited by speed, and ΣΔ is the most accurate, but the slowest. The follow-up parallel analogue-to-digital converter provides an optimal balance, especially for the 1,014-bit range and speeds up to 500 MS/s. It combines adaptability to rapid signal changes (thanks to the Flash insert) with the accuracy of a closed-loop control system.

Analysis of the results obtained through the prism of fundamental works can identify critical engineering challenges in implementing the proposed architecture. The proposed hybrid architecture is characterised by several unique properties that distinguish it from its counterparts. The combination of tracking and parallel methods provides the system with dual properties. In quasi-stationary mode, when the signal changes slowly, it functions as a high-precision tracking circuit with minimal error. During fast transients, the architecture adapts instantly: the parallel channel digitises the dynamic difference, effectively expanding the equivalent bandwidth. Thanks to this adaptability, the system is capable of correctly processing signals with variable spectral density. This behaviour is consistent with the concept of "smart" data converters described by A.H.M. Roermund *et al.* (2010), where adaptability to input signal characteristics is considered a key direction for improving the energy efficiency of modern microcircuits. A key engineering advantage is the reduction of quantisation error in dynamic mode. In contrast to the traditional "sawtooth" overload characteristic of delta modulators, the output signal accurately replicates the shape of the input signal by adding an error code. This contributes to an increase in the linearity of the output characteristic, since the residual error does not accumulate but is corrected at each conversion cycle. The parallel analogue-to-digital converter acts as a dynamic buffer. A classic tracking analogue-to-digital converter is limited to a change of only 1 LSB per clock cycle. As noted by the authors in their reference book, U. Tietze *et al.* (2008), this is the main factor that makes it impossible to use such analogue-to-digital converters for processing broadband signals. Also, in the fundamental work by B. Razavi (1995), it is noted that two-step architectures can circumvent the speed limitations of sequential methods, but usually require complex sampling-storage circuits. However, the proposed architecture can correctly reproduce an amplitude jump to 2^{K-1} LSB in one clock cycle (where K – Flash section bit width). This significantly

increases the slew rate without the need to increase the clock frequency.

In classic Flash analogue-to-digital converters, the input signal is fed to hundreds or thousands of parallel comparators, creating a huge capacitive load on the signal source. The fundamental work by P. Horowitz & W. Hill (2015) emphasises that such input capacitance requires powerful and energy-intensive input buffers, which often negate the advantages of high speed. In the proposed architecture, the input signal is loaded only onto one comparator (tracking circuit) and the input of the operational amplifier, which greatly simplifies the requirements for the signal source. As described in detail in R. van de Plassche (2003), reducing input capacitance is critical for high-frequency applications as it can expand bandwidth without increasing the power consumption of the driver. T.C. Carusone *et al*. (2011) investigated the fundamental limitations of the speed of closed analogue circuits. The study concluded that the accuracy of the output voltage setting depends exponentially on the product of the bandwidth and the settling time, and the linearity is determined by the depth of feedback. Comparing these conclusions with the simulation results, it can be argued that for the correct operation of a tracking-parallel ADC, it is critical to ensure that the slew rate of the operational amplifier exceeds the maximum slope of the error signal. Failure to meet this condition, as predicted by the researchers, leads to dynamic nonlinear distortions, which were observed in simulations at frequencies above 125 kHz.

Scientist F. Maloberti (2007) conducted a detailed study of the conversion stages in conveyor and multistage ADCs. The study concluded that the amplifier settling time must be strictly less than half the clock frequency to avoid charge transfer errors between stages. In the context of the study, this means that the operational amplifier that forms the residual signal must have a settling time of no more than 5 ns (for f_{clk} = 100 MHz). This confirms that the OP is the "bottleneck" that limits the maximum sampling frequency of a hybrid ADC compared to a Flash converter. M. Fakhfakh *et al*. (2013) studied the effect of technological parameter variation on the characteristics of mixed circuits. The study determined that variation in multi-path architectures is the dominant source of nonlinear distortions that cannot be eliminated by circuit design methods alone. A comparison with the proposed architecture shows that the discrepancy in the transfer coefficient between the OP and the digital weights of the Flash ADC leads to DNL errors at the range switching boundaries. This correlates with the monograph by M.J. Pelgrom (2010), proving that passive matching becomes geometrically inefficient for accuracies above 10 bits. Therefore, the proposed solution requires the implementation of digital calibration algorithms similar to those described in the paper to achieve the claimed 12-bit accuracy.

R.J. Baker (2008) studied the mechanisms of noise generation in mixed signals, in particular the "backlash" effect from dynamic comparators. The study concluded that this noise can significantly distort the input signal if the source has high impedance. In contrast to the author's conclusions for Flash ADCs, the study observed less sensitivity to this effect. This is attributed to the fact that the tracking circuit acts as a low-impedance driver and filter, confirming the thesis of increased noise immunity of the hybrid circuit. D.-R. Oh *et al*. (2022) investigated a 7-bit two-step Flash ADC that uses a sampling-storage sharing (S/H sharing) technique to reduce chip area. The study concluded that this architecture can achieve a sampling rate of 2.5 GHz with low power consumption. Compared to the study, the tracking-parallel architecture also reduces the number of comparators, but the advantage is determined by the absence of a strict need for complex clocking of the S/H device for the tracking circuit, since it operates in continuous tracking mode, which can provide lower latency for certain classes of signals.

K.E. Lozada *et al*. (2024) analysed hybrid SAR-based architectures, specifically SAR-Flash and SAR-Pipeline, to improve energy efficiency. Their findings indicate that using SAR for coarse quantisation significantly reduces energy consumption compared to full Flash circuits. Compared to these results, the proposed tracking-parallel method offers an alternative compromise: instead of using iterative (and therefore slower) SAR for accurate approximation, a fast Flash block is used for instantaneous error correction. This makes the architecture potentially faster for real-time signal tracking tasks, although possibly less energy-efficient than SAR. F. Shahbazi & H. Shamsi (2025) presented a Level-Crossing ADC with a 1-bit DAC and a continuous-time comparator. The authors achieved ultra-low power consumption due to the asynchronous nature of the operation. However, their approach has limitations in terms of accuracy and linearity due to the use of 1-bit feedback. The proposed architecture, in contrast, uses a multi-bit tracking circuit and a multi-bit Flash corrector, which provides significantly higher linearity (10-12 bits) and better response to fast signal changes, while maintaining the adaptive power consumption advantages inherent in feedback systems.

H. Tang *et al*. (2024) proposed a hybrid LC-SAR that combines a level crossing detector (LC) with SAR logic for Internet of Things (IoT) applications. They achieved high energy efficiency (0.494.34 μW), but the bandwidth is limited to 20 kHz. In comparison, the proposed tracking-parallel ADC targets a much wider frequency range (hundreds of kHz to MHz), as the use of a Flash section for residual processing eliminates the speed bottleneck characteristic of sequential algorithms. M. Timmermans *et al*. (2024) investigated a 64-dB SNDR Continuous-Time Level Crossing ADC with dynamic self-offset comparators. They conclude that this technique achieves record energy efficiency for sparse signals. Compared to the research results, the proposed approach also exploits the idea of adaptability (the

Flash section is active only when there is a significant error), which can reach the energy efficiency of circuits, but with more deterministic time behaviour due to clocking, which simplifies digital processing. S.A. Zahrai & M. Onabajo (2018) highlighted how individual substructures affect energy consumption and performance, in particular, considering time-interleaved (TI) and sub-ranging architectures to achieve high sampling rates with low consumption. The study provided an example of a hybrid ADC implementation with a frequency of ~1 GS/s, where the first stage is a flash ADC and the second stage is four TI SAR-like blocks, and showed actual chip measurements confirming low power consumption at medium bit depth (~6-8 bits) and high speed.

In their review, S. Xiao *et al*. (2025) emphasised the need to create specialised ADCs for in-memory computing (CIM) systems, where a balance between area and accuracy is important. The proposed design fits this trend, as the absence of bulky capacitor arrays (characteristic of SAR) and a large number of comparators (Flash) provides compact solutions suitable for integration into complex computing arrays. K. Krishna & N. Nambath (2024) reviewed high-speed dynamic comparators, highlighting their role in the energy consumption of Flash ADCs. Since the proposed circuit uses only a low-bit-width (3-4 bits) Flash subsystem, the number of such energy-intensive elements is minimised, solving the problem and achieving a better FoM compared to traditional solutions. The results of the analysis confirmed that the track-and-parallel architecture is an effective hybrid solution that fills the niche between simple but slow track-and-parallel analogue-to-digital converters and fast but expensive Flash analogue-to-digital converters. It offers an optimal balance of parameters for a wide class of devices where both dynamic accuracy and moderate power consumption are relevant.

The study further developed the theory of designing hybrid analogue-to-digital converters. For the first time, the architecture in which a parallel analogue-to-digital converter is used not for traditional result refinement (as in sub-ranging schemes), but as a link of instantaneous negative feedback for error, has been studied in detail and mathematically substantiated. The study proved that this approach can transform main disadvantage of tracking systems – inertia – into a controllable parameter that is effectively compensated by a low-bit parallel block. This can be used for the creation of a new class of analogue-to-digital converters that combine high equivalent speed with minimal chip area and adaptive power consumption.

CONCLUSIONS

The paper presents a comprehensive study and mathematical modelling of a new architecture for a tracking-parallel analogue-to-digital converter. The simulation results confirmed that the proposed hybrid approach effectively solves the problem of the limited performance of classical tracking systems. In particular, the study determined that the integration of a K-bit parallel sub-converter into the error loop increases the maximum input signal rise rate by 2^K times. A numerical experiment for a 12-bit model with a 4-bit Flash module at a clock frequency of 100 MHz demonstrated an expansion of the operating frequency band from 7.77 kHz (for the classical circuit) to 124.2 kHz without the occurrence of slope overload distortions. At the same time, the dynamic signal recovery error across the entire range did not exceed ±0.5 LSB. A comparative analysis showed that the developed architecture provides a significant advantage in terms of hardware complexity. To implement 12-bit conversion, the proposed circuit requires fewer than 35 comparators, which is a 99% reduction in hardware costs compared to a classic Flash ADC (4,095 comparators). This provides high equivalent performance with significantly less chip area and power consumption than fully parallel analogues, and with lower latency than SAR architectures. The scientific novelty of the work lies in establishing and confirming the analytical relationship between the bit width of the auxiliary parallel channel and the dynamic range expansion factor of the tracking system. The practical value of the results is determined by the creation of a scalable solution for precision control and telemetry systems that can process broadband signals with an accuracy of 1,014 bits without the need for energy-intensive ultra-fast ADCs. Further work will focus on the hardware implementation of the device in modern nanometre CMOS technologies (2,865 nm) with the development of digital background calibration algorithms to compensate for channel mismatches. It is also planned to investigate methods of dynamic power management of individual blocks to minimise energy consumption in IoT applications.

ACKNOWLEDGEMENTS

None.

FUNDING

None.

CONFLICT OF INTEREST

None.

REFERENCES

[1] Azarov, O.D., Chernyak, O.I., & Stahov, O.Y. (2020). Tracking and bitwise balancing ADC with weight redundancy. *Information Technologies and Computer Engineering*, 49(3), 37-44. doi: 10.31649/1999-9941-2020-49-3-37-44.

[2] Baker, R.J. (2008). A high-speed data converter. In *CMOS: Mixed-signal circuit design* (pp. 301-325). Hoboken: Wiley-IEEE Press. doi: 10.13140/RG.2.1.5146.6005.

[3] Carusone, T.C., Johns, D.A., & Martin, K.W. (2011). *Analog integrated circuit design* (2nd ed.) Hoboken: Wiley.

[4] Fakhfakh, M., Tlelo-Cuautle, E., & Castro-López, R. (2013). *Analog/RF and mixed-signal circuit systematic design*. Berlin: Springer. doi: 10.1007/978-3-642-36329-0.
[5] Horowitz, P., & Hill, W. (2015). *The art of electronics*. New York: Cambridge University Press.
[6] Krishna, K., & Nambath, N. (2024). Review on high-speed dynamic comparators for analog to digital converters. *Journal of Circuits, Systems and Computers*, 33(13), article number 2430006. doi: 10.1142/S021812662430006X.
[7] Lozada, K.E., Chang, D.J., Oh, D.R., Seo, M.J., & Ryu, S.T. (2024). SAR-assisted energy-efficient hybrid ADCs. *IEEE Open Journal of the Solid-State Circuits Society*, 4, 163-175. doi: 10.1109/OJSSCS.2024.3472000.
[8] Maloberti, F. (2007). *Data converters* (pp. 147-170). New York: Springer.
[9] Oh, D.R., Seo, M.J., & Ryu, S.T. (2022). A 7-bit two-step flash ADC with sample-and-hold sharing technique. *IEEE Journal of Solid-State Circuits*, 57(9), 2791-2801. doi: 10.1109/JSSC.2022.3159569.
[10] Patent No. 158494. (2025). *Tracking-parallel analog-to-digital converter*. Retrieved from https://sis.nipo.gov.ua/uk/search/detail/1841710/.
[11] Pelgrom, M.J. (2010). *Analog-to-digital conversion*. New York: Springer. doi: 10.1007/978-1-4614-1371-4.
[12] Razavi, B. (1995). *Principles of data conversion system design*. New York: IEEE Press.
[13] Roermund, A.H.M., Casier, H., & Steyaert, M. (2010). *Analog circuit design: Smart data converters, filters on chip, multimode transmitters*. Dordrecht: Springer. doi: 10.1007/978-90-481-3083-2.
[14] Shahbazi, F., & Shamsi, H. (2025). On the design of a level-crossing ADC with 1-bit DAC and rail-to-rail continuous-time comparator. *Analog Integrated Circuits and Signal Processing*, 122, article number 31. doi: 10.1007/s10470-025-02344-w.
[15] Tang, H., Xu, W., Li, H., Wei, B., & Wei, X. (2024). A 0.494.34 μW LC-SAR hybrid ADC with a 10.85-Bit ENOB and 20 KS/s bandwidth. *Electronics*, 13(6), article number 1078. doi: 10.3390/electronics13061078.
[16] Tietze, U., Schenk, C., & Gamm, E. (2008). *Electronic circuits: Handbook for design and application* (2nd ed.). Berlin: Springer. doi: 10.1007/978-3-540-78655-9.
[17] Timmermans, M., van Oosterhout, K., Fattori, M., Harpe, P.J.A., Liu, Y.-H., & Cantatore, E. (2024). A 1.8-65 fJ/conv.-step 64 dB SNDR continuous-time level crossing ADC exploiting dynamic self-biasing comparators. *IEEE Journal of Solid-State Circuits*, 59(4), 1194-1203. doi: 10.1109/JSSC.2024.3352735.
[18] van de Plassche, R.J. (2003). *CMOS integrated analog-to-digital and digital-to-analog converters*. Boston: Kluwer Academic Publishers.
[19] Verreault, A., Cicek, P.V., & Robichaud, A. (2024). Oversampling ADC: A review of recent design trends. *IEEE Access*, 12, 121753-121779. doi: 10.1109/ACCESS.2024.3452589.
[20] Xiao, S., Li, F., Hao, T., & Han, G. (2025). Analog-to-digital converter design for diverse performance computing-in-memory systems: A comprehensive review. *Integrated Circuits and Systems*, 2(2), 81-92. doi: 10.23919/ICS.2025.3571019.
[21] Zahrai, S.A., & Onabajo, M. (2018). Review of analog-to-digital conversion characteristics and design considerations for the creation of power-efficient hybrid data converters. *Journal of Low Power Electronics and Applications*, 8(2), article number 12. doi: 10.3390/jlpea8020012.

Слідкувально-паралельний аналого-цифровий перетворювач

Олексій Азаров
Доктор технічних наук, професор
Вінницький національний технічний університет
21021, вул. Хмельницьке шосе, 95, м. Вінниця, Україна
https://orcid.org/0000-0002-8501-1379

Сергій Богомолов
Кандидат технічних наук, доцент
Вінницький національний технічний університет
21021, вул. Хмельницьке шосе, 95, м. Вінниця, Україна
https://orcid.org/0009-0003-3823-8380

Олександр Лукашук
Аспірант
Вінницький національний технічний університет
21021, вул. Хмельницьке шосе, 95, м. Вінниця, Україна
https://orcid.org/0009-0002-7485-1193

Анотація. В умовах стрімкого розвитку цифрових систем обробки сигналів актуальним стає завдання створення нових архітектур аналого-цифрових перетворювачів, здатних забезпечити високу динамічну точність без надмірного нарощування кількості компонентів. Метою роботи було підвищення динамічної точності та швидкодії процесу аналого-цифрового перетворення шляхом розробки та дослідження удосконаленої слідкувально-паралельної архітектури. Для досягнення поставлених завдань застосовано методи теорії автоматичного керування, схемотехнічного аналізу імпульсних пристроїв та математичного моделювання похибок квантування. У статті детально розглянуто принцип дії запропонованого пристрою, який базується на синергії інерційного слідкувального контуру та швидкодіючого паралельного перетворювача залишку. Доведено, що введення ланки вимірювання динамічної помилки дозволяє компенсувати відставання вихідного коду від вхідного сигналу, характерне для традиційних слідкувальних систем, і тим самим усунути спотворення внаслідок перевантаження за нахилом. Аналітично обґрунтовано залежність розширення динамічного діапазону вхідного сигналу від розрядності паралельного блоку, що дає змогу гнучко адаптувати систему під вимоги конкретного застосування. Результати порівняльного аналізу свідчать про те, що запропонована структура забезпечує суттєвий виграш в апаратній складності, дозволяючи реалізувати високу роздільну здатність із використанням на порядок меншої кількості прецизійних компараторів порівняно з аналогами. Також визначено критичні вимоги до швидкодії операційного підсилювача та цифро-аналогового перетворювача у колі зворотного зв'язку для забезпечення стабільної роботи у всьому частотному діапазоні. Практична цінність дослідження полягає в розробці рекомендацій для проєктування конкурентоспроможних мікросхем, орієнтованих на використання в портативних пристроях, медичній апаратурі та системах промислової автоматики

Ключові слова: гібридна архітектура; перевантаження за нахилом; компенсація динамічної похибки; сигнал залишку; швидкість наростання сигналу; апаратна ефективність; високоточність

BULLETIN of Cherkasy State Technological University

Journal homepage: https://bulletin-chstu.com.ua/en

Vol. 31 No. 1. 2026

UDC 621.391; 004.8; 004.426
DOI: 10.62660/bcstu/1.2026.24

ISSN: 2306-4412
E-ISSN: 2708-6070

Methodology development for training neural differential distinguishers and key recovery attack on single ciphertext

Xue Jiang*
Postgraduate Student
National University of Life and Environmental Sciences of Ukraine
03041, 15 Heroiv Oborony Str., Kyiv, Ukraine
https://orcid.org/0009-0000-1676-2331

Valerii Lakhno
Doctor of Technical Sciences, Professor
National University of Life and Environmental Sciences of Ukraine
03041, 15 Heroiv Oborony Str., Kyiv, Ukraine
https://orcid.org/0000-0001-9695-4543

Abstract. Key recovery attacks pose a critical threat to cryptographic systems, as these attacks directly compromise the security of encryption mechanisms. The purpose of this study was to develop a novel single-ciphertext key recovery method that addresses the limitations of traditional cryptanalysis techniques in scenarios with limited data availability. The methodology combined an enhanced B-C3-HSwish neural distinguisher with Bayesian optimisation guided by the Upper Confidence Bound (UCB) strategy. The neural distinguisher was improved through multiscale convolutional layers, activation function optimisation, and diversified input structures, while Bayesian optimisation mapped the key recovery problem to a high-dimensional search task using a Gaussian process surrogate model and the UCB acquisition function. Experimental results on the Speck32/64 encryption algorithm demonstrated a 56% success rate in 11 rounds of key recovery attacks, with data complexity ranging from 213.5 to 213.6. This result surpassed the 52.1% success rate obtained in the reference. The B-C3-HSwish neural discriminator has been improved, in particular by replacing the Relu activation function with Hard_swish, to increase its accuracy in single-pair input mode. This study not only offered a new, resource-efficient solution for key recovery in single-ciphertext scenarios, but also highlights the powerful potential of Bayesian optimisation as a promising paradigm for advancing cryptanalysis. The practical value of this work lies in its ability to provide a resource-efficient solution for key recovery in single-ciphertext scenarios, making it particularly suitable for real-world applications with constrained data. This approach also establishes Bayesian optimisation as a promising paradigm for advancing cryptanalysis and enhancing cryptographic security

Keywords: Bayesian optimisation algorithm; feature extraction optimisation; attack scenario; Speck32/64 algorithm; neural networks

INTRODUCTION

As the intersection of deep learning and cryptography has progressively deepened, the application of neural networks in cryptanalysis has garnered considerable attention. In the modern environment of rapid development of both cryptographic data protection methods and methods for breaking such protections, there is a growing need for new approaches to analysing the robustness of cryptographic algorithms. The application of deep learning in cryptanalysis is becoming particularly relevant, as it allows for the automatic detection of complex patterns and vulnerabilities that are difficult to detect using traditional methods. As a result, the application of neural networks in cryptanalysis is attracting increasing attention from researchers.

Article's History: Received: 23.06.2025; Revised: 21.01.2026; Accepted: 16.03.2026; Published: 08.04.2026

Suggested Citation:
Jiang, X., & Lakhno, V. (2026). Methodology development for training neural differential distinguishers and key recovery attack on single ciphertext. *Bulletin of Cherkasy State Technological University*, 31(1), 24-33. doi: 10.62660/bcstu/1.2026.24.

*Corresponding author

According to research by Z. Hou *et al.* (2021), an enhanced neural distinguisher was proposed to improve the performance of cryptanalysis. The authors' work significantly boosted the accuracy of the distinguisher in differential cryptanalysis by optimising network structures and input data formats. The authors highlighted that the integration of convolutional neural networks (CNNs) and attention mechanisms can effectively capture nonlinear features in ciphertext, thereby enhancing model performance. A. Gohr (2019) explored how deep learning techniques could be employed to attack the reduced-round Speck32/64 cipher. By incorporating deep neural networks, the author successfully enhanced the efficiency of the attack and demonstrated the potential of neural networks in cryptanalysis. The author concluded that deep learning not only accelerates cryptanalysis but also uncovers patterns that are challenging to detect using traditional methods. In particular, A. Gohr *et al.* (2022) investigated the effectiveness of using deep convolutional neural networks for differential cryptanalysis tasks. The authors focused on studying how neural networks can automatically learn to distinguish encrypted texts derived from certain known differential characteristics. The results of the study demonstrated that deep CNN models are capable of achieving high accuracy, even surpassing traditional statistical methods, particularly with fixed encryption parameters.

In the researchers' work, L. Zhang & Z. Wang (2022) conducted a comparative analysis of several modern architectures, including CNN, and ResNet, for use in differential cryptanalysis tasks. The main focus was on the effectiveness of model training when varying the amount of training data and the complexity of differential characteristics. The authors concluded that ResNet models have high stability and generalisation ability under limited training sets, demonstrating improved accuracy compared to other approaches. In contrast, S. Zhu *et al.* (2025) focused on adapting ResNet to analyse high-dimensional ciphertext, which is characteristic of modern block ciphers. The research covered network depth optimisation and the application of various normalisation mechanisms to avoid overfitting. As a result, the authors confirmed that ResNet not only maintains accuracy when working with large amounts of data, but also better copes with detecting weak differential characteristics in complex cryptographic constructions. Through experiments comparing various neural network architectures in differential cryptanalysis, the researchers concluded that residual networks (ResNets) perform exceptionally well when handling complex ciphertext, particularly in high-dimensional data environments.

The research by A. Benamira *et al.* (2021) delved into the application of machine learning in cryptanalysis. Experimental validation of the effectiveness of various machine learning models in cryptographic attacks revealed that, although neural networks perform well in cryptanalysis, the interpretability remains a pressing issue. The authors suggested that future research should focus on improving model interpretability to better understand the decision-making processes. A. Baksi *et al.* (2021) proposed the application of machine learning techniques to enhance differential distinguishers for lightweight ciphers. By integrating machine learning with traditional cryptographic analysis methods, the authors demonstrated significant improvements in the identification of differential characteristics, thereby advancing the efficiency of cryptanalysis for lightweight ciphers. The authors' work underscored the potential of machine learning in optimising cryptographic algorithms and highlighted its relevance in the evolving field of lightweight cipher security.

P. Auer (2020) explored the application of online learning in cryptanalysis. By introducing the Upper Confidence Bound (UCB) algorithm, the author proposed an online learning framework capable of dynamically adjusting model parameters to adapt to continuously changing ciphertext data. The research demonstrated that online learning can effectively enhance model performance in real-time cryptanalysis. D. Lin *et al.* (2022) proposed the application of the Bayesian optimisation algorithm in cryptanalysis. By combining Bayesian optimisation with neural networks, the authors successfully improved the efficiency of cryptanalysis and highlighted the importance of Bayesian optimisation in hyperparameter tuning. The research provided new insights into automated model optimisation in cryptanalysis.

Although these studies have made significant progress in the intersection of neural networks and cryptography, some aspects remain underexplored. For instance, existing research primarily focuses on differential cryptanalysis, while the application of linear cryptanalysis has received less attention. Additionally, how to generalise neural network optimisation strategies across different types of cryptographic algorithms remains an underexplored issue. These questions served as the focus of this study, aiming to provide new insights for the further application of neural networks in cryptanalysis. The aim of this study was to propose a new single-ciphertext key recovery method that overcomes the limitations of traditional cryptanalysis techniques in data-scarce scenarios.

MATERIALS AND METHODS

This research conducted an in-depth literature survey, focusing on the following aspects: network structure optimisation (different types of network structures, such as convolutional neural networks, residual networks, attention mechanism networks, etc., and the advantages in feature extraction and expression ability were studied); input data structure optimisation (discusses how to optimise the ciphertext to input mode, in order to better use the learning ability of the neural network to improve the performance of the differentiator); activation function and the normalisation technique (compare the different activation functions, and

normalisation methods affect the performance of model, to find the best combination to improve the accuracy of a differentiator).

In this study, A. Gohr (2019) attack approach was adopted. By assuming an s-round differential $\delta \rightarrow \Delta$ with a conversion probability of , denoted as Chosen Differential (CD), the γ-round distinguisher was extended to an $s+\gamma$ round distinguisher. Due to the probabilistic nature of the preset differential, it was required to generate approximately $c \times 2^{p}$ (denoted as n_{cts}) plaintext pairs with input differences of δ, where c is a small constant. Next, using the n_b neutral bits of the s-round CD, each plaintext pair was expanded into a plaintext structure containing n_b plaintext pairs. Since Speck32/64 does not perform key addition before the first nonlinear operation, the differential propagation characteristics of single-round encryption were not affected by the secret key. Therefore, the generated n_cts ciphertext structures could be decrypted at zero cost with a zero key and extended forward by one round to obtain the plaintext structure. Finally, all obtained plaintext structures underwent $1+s+\gamma+1$ rounds of encryption to obtain the ciphertext structure for the key recovery attack. After obtaining the high-accuracy differential distinguisher through the neural network, A. Gohr (2019) used the differential distinguisher of two rounds of 6 and 7 to carry out a key recovery attack against the 11 rounds of Speck encryption, which is the purpose of differential analysis. The symmetric cipher was broken as a whole by performing a key recovery attack through the differential distinguisher to obtain the real key.

The initial idea of extending rounds for decryption can be presented as a sequential operation with prepared differences and gradual addition of rounds to a known differential diffuser. First, a pair of plaintexts with a difference (0x211, 0xa04) was selected; after two rounds of encryption, this difference was converted into a high-probability difference (0x40.0) with a probability of 2^{-6}, which allowed it to be extended to the equivalent of 9 rounds based on a 7-round differential. Further, using Speck's property, according to which one round can be "added" without additional probability costs, the resulting 9-round differential was expanded to 10 rounds without loss. When restoring the key, the pairs of ciphertexts were decrypted one round back, obtaining N–1-round ciphertexts, and a differential distinguisher was applied to these results; thus, the 10-round analysis can be raised by one more round, achieving effective coverage of 11 rounds of encryption (Fig. 1).

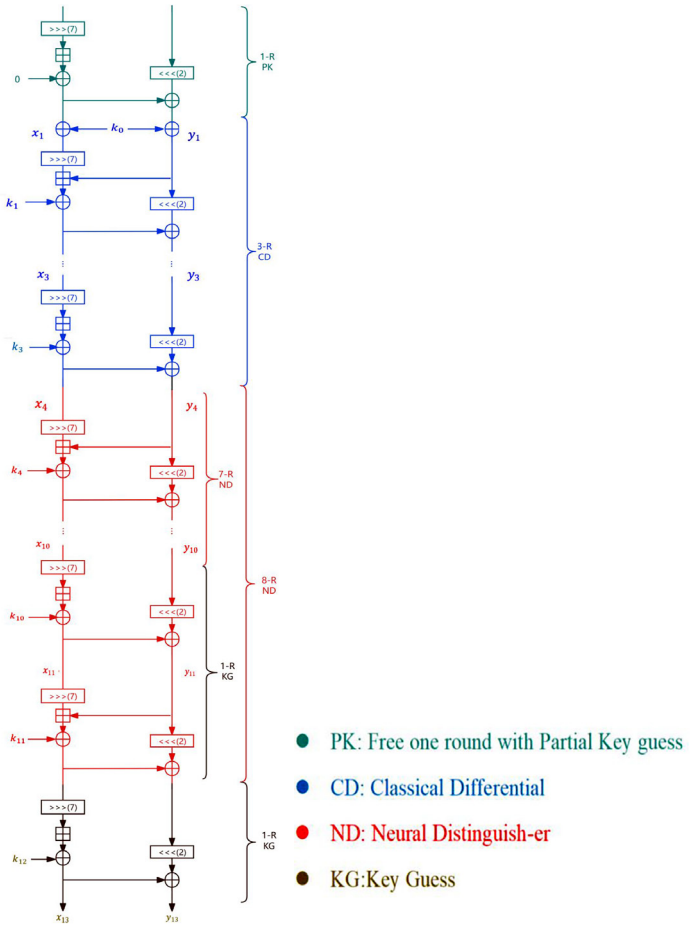

Figure 1. Round expansion process

Source: compiled by the authors

Since CD is probabilistic, only a 2^{-p} probability of plaintext structures satisfying the difference value δ will propagate to ciphertext structures with a difference of Δ after s rounds of differential propagation. This results in low computational efficiency. To accelerate the attack, A. Gohr (2019) introduced a UCB exploration-exploitation method for setting confidence interval upper limits. Each ciphertext structure was assigned a priority level, and before each key guess, the ciphertext structure with the highest priority was selected. This process was repeated n_{it} times to concentrate computational resources on the most likely correct ciphertext structures, thereby improving the attack efficiency.

The distinguisher's response to guessed subkeys depends on the Hamming distance from the true subkey. Therefore, the pre-generated incorrect key response file served as the basis for the Bayesian subkey search. By minimising the weighted Euclidean distance, the next candidate subkey was obtained. After selecting the most promising ciphertext structure C_i ($1 \leqslant j \leqslant n_{cts}$) candidate subkeys were randomly picked from the possible keys and entered into the Bayesian optimisation algorithm for n_{cand1} iterations. In each iteration, the candidate subkey $k_{1,j}$ ($1 \leqslant j \leqslant n_{cand1}$) was used to decrypt one round of the $(1+s+\gamma)$ round ciphertext structure C_i, obtaining a score $z_i^{k1,j}$ using the γ-round distinguisher, which was then combined into $v_{k_{1,j}}$. The response file selected the first n_{cand1} smallest Euclidean distance keys as the next guessed keys.

When searching for γ-round candidate subkeys using the Bayesian optimisation algorithm, the subkey $k_{1,j}$ and its score $v_{k_{1,j}}$ were stored. When $v_{k_{1,j}}$ exceeded the threshold c_1, the reciprocal second-round candidate subkey $v_{k_{2,j}}$ ($1 \leqslant j \leqslant n_{cand2}$) was used to decrypt one round to obtain the $1+s+\gamma-1$ round ciphertext structure, and the neural distinguisher and response file search were performed using the $\gamma-1$ round. If $k_{2,l}$ scored higher than the threshold c_2, the search was terminated, and k_1 and k_2 were returned for verification. Preset threshold – $C_{11}=10$, $C_{10}=10$. Before returning the guessed key, it is necessary to verify the search within a small range around it. Find subkey groups with a Hamming distance of 2 from $k_{1,j}$ and $k_{2,j}$, combine these groups to form ciphertext structures, and use the γ-round and $\gamma-1$ round distinguishers to score the resulting structures. Return the subkey corresponding to the best score, eliminating bit errors. If the verification improves, repeat the verification with the new best-guessed key.

In this paper, the standard-based UCB algorithm was used to assign priorities to each ciphertext structure according to its recommended sub-key score and access times and was combined with Bayesian optimisation to speed up the key recovery process. The key recovery attack was considered successful when the final returned subkey is correct and the second-to-last wheel key had at most 2-bit of error. The data were divided into training, validation, and test subsets (approximately 80/10/10) with independent random number generator seeds; classes were balanced, and bit representations were normalised according to the architecture (bit-plane or one-hot for CNN). The model used a binary loss function (binary cross-entropy), Adam optimiser ($\beta_1=0.9$, $\beta_2=0.999$, $\varepsilon=1e-8$) with a starting learning rate of $1e-3$, a mini-batch size of 256, and L_2 regularisation (weight decay $\approx 1e-5$); in full experiments, dropout 0.3 was applied in fully connected layers. Training lasted up to a maximum of 200 epochs with early stopping with a patience of 20 epochs on the validation metric (validation loss or AUC), and the training plan used ReduceLROnPlateau to reduce the learning rate in the absence of improvements. For stability, the results were obtained with three different random seeds, while checkpoints with the best validation values were saved and all key metrics (train/val loss, AUC, accuracy) were logged to ensure full reproducibility.

The complete key recovery attack protocol integrated a neural discriminator with Bayesian optimisation as a single procedure for searching and verifying candidates. Initially, a pool of 32 initial candidates was formed, which were obtained either by pre-ranking a large set of hypotheses by a neural network on a small subset of data, or by combining statistical heuristics and local mutations; for each candidate, partial decoding to an intermediate state (N–1 round) was performed, and these intermediate vectors were fed to the recogniser. The recogniser returned the probability *p_k* that the intermediate values correspond to the expected differential; for input into the optimisation procedure, *p_k* was aggregated (averaged or median across several batches) and, if necessary, converted to log-odds.

Bayesian optimisation worked with a surrogate Gaussian Process (RBF-kernel) model or TPE as a fallback and used the Expected Improvement acquisition function (alternatively UCB); optimisation was initiated with *n_init* ≈ 20 random evaluations and performed up to 500 iterations, with BO proposing the next candidate for evaluation at each iteration, obtaining *S_k* (target evaluation, e.g. – *logit*(*p_k*) or another aggregated quality function) and updating the surrogate. The stopping criteria included reaching the maximum number of iterations, no improvement within a specified patience period, or reaching the threshold *p_k* (e.g., ⩾ 0.99); final verification was performed for the top M candidates for *S_k* (e.g., $M=5$) by complete decryption. For statistical evaluation of the attack's reliability, $N=100$ independent runs were performed with different seeds and sets of text pairs, after which the success rate, the average number of decryption oracle calls, and the average time were recorded; standard statistical tests (t-test or bootstrap for confidence intervals) were used for comparisons between approaches.

RESULTS AND DISCUSSION

The network structure following the aforementioned improvements is designated as B-C3-HSwish (Jiang *et al.*, 2025), as illustrated in Figure 2. Specifically, the

module has been redesigned into a multi-branch parallel convolutional structure, which not only enhances the diversity of feature extraction but also improves the overall expressiveness of the network. Furthermore, the original ReLU activation function has been replaced with the Swish activation function, effectively mitigating the issue of "neuron death" that may arise due to ReLU's constant zero output in the negative region, while also enhancing the network's adaptability to different input data distributions.

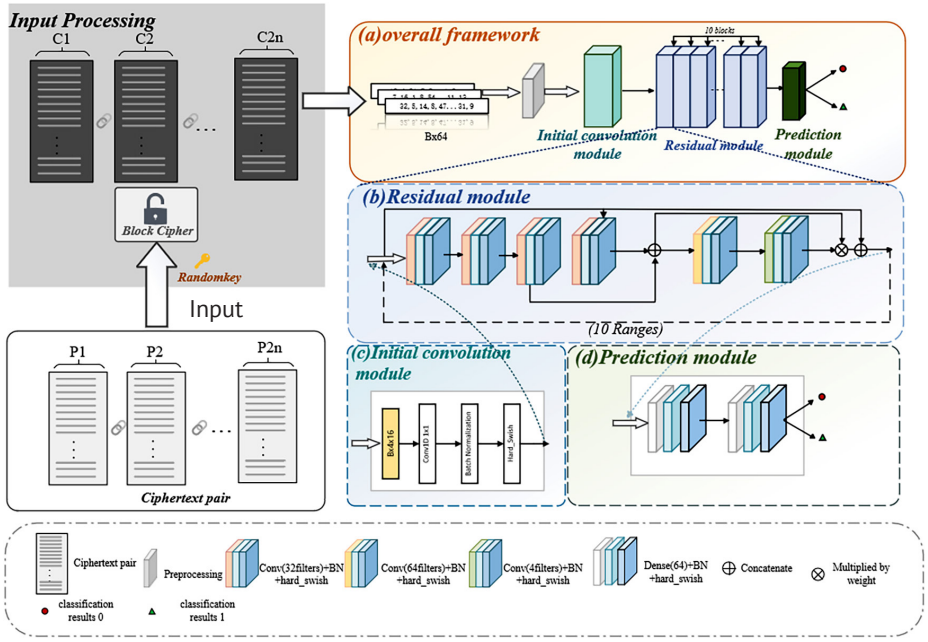

Figure 2. Overall structure diagram of the improved neural network B-C3-HSwish
Source: compiled by the authors

As a result of the literature analysis, it can be summarised, that the network structure after the aforementioned improvements is named B-C3-HSwish, as shown in Figure 2. Specifically, Module 2 has been redesigned into a multi-branch parallel convolutional structure, which not only enhances the diversity of feature extraction but also improves the overall expressiveness of the network. Moreover, the original Relu activation function has been replaced with the Swish activation function, effectively alleviating the issue of "neuron death" that may occur due to Relu's constant zero output in the negative region, while also enhancing the network's adaptability to different input data distributions. A model of a multiscale convolution enhancement scheme was proposed, which integrates the multiscale convolution method and network design principle of RegNet. 3 * 3. By utilising stacked convolutions and adding convolutional branches, the impact of these modifications on the distinguisher accuracy was investigated. This method significantly enhances the feature extraction function of the model.

Activation function improvement: activation function and normalisation can improve model expression ability, prediction accuracy, stability, and convergence speed in neural networks. Based on the improved distinguisher, this paper conducted experiments on the activation function and normalisation. In the residual structure, Relu makes the "residual" branch result non-negatively increasing, which affects the feature expression. To solve this problem, this paper adjusted the position of the activation function or replaced it with Swish, LeakyRelu, etc. After 5 rounds of discrimination tests of Speck encryption ciphertext pairs, Hard_swish was finally determined to replace Relu. Development of an encryption method based on multiple ciphertexts for input structure: By using random keys for one round of encryption, the diversity of input data was realised and the learning ability of the neural network for ciphertext features was enhanced. This significantly improved the generalisation ability and training efficiency of the model.

In this paper, a key search strategy based on Bayesian optimisation was adopted to improve the search accuracy and efficiency in the attack model. In the attack model, the distinguisher is regarded as a black box function, and the maximum value of the function is the candidate sub-key. In order to find the maximum value, the Bayesian optimisation method was used in the recovery process, which includes two core processes: the prior function and the collection function. The prior function was solved by Gaussian process regression, which was then solved by Bayesian inference. The acquisition function included an exploration-exploitation method, aiming to find the best sample set. That is, it first guessed a sub-key, decrypted the ciphertext using the sub-key, and obtained the result through the

black-box function. Then the candidate sub-keys were obtained by Bayesian inference using the values of the sampling points. The value of the sample point can be calculated by precomputing the sample set in the case that the wrong key is obtained. Compared with the traditional search strategy, the Bayesian optimisation method can quickly find the maximum value of the function and its corresponding parameters without knowing the function equation, so as to improve the accuracy and efficiency of key search.

The frequency of recovery attacks was set to 100, and 100 ciphertext combinations containing 11 rounds were constructed during each attack. The formation of these ciphertext structures was based on neutral bit-flipping technology. Specifically, each original ciphertext can be extended to 64 derivative ciphertexts by a 6-bit bit flipping operation, and the required 100 ciphertext structure groups were finally constructed by using this bit expansion mechanism. This design enabled a single base ciphertext pair to be systematically generated into a complete data structure containing 64 associated ciphertexts after bit transformation. The maximum number of iterations was set to 500 when searching the ciphertext structure. In each iteration, the UCB algorithm was used to select the ciphertext structure with the current highest priority, and the initial selection is the first ciphertext structure.

The decryption evaluation was carried out for the selected ciphertext architecture, and the Bayesian optimisation strategy was used for a total of five rounds of iterations. In each round of Bayesian optimisation, all ciphertext pairs were decrypted from 32 preset key candidates, and the key selection in the first iteration followed the random principle. After decrypting the ciphertext with these keys, the decryption results were scored by a specific distinguisher, and the score of each key and the average of the distinguisher scores were calculated according to (1):

$$S_k = \sum_{i=1}^{\frac{N}{k}} log_2\left(\frac{Z_i}{1-Z_i}\right), \quad (1)$$

where S_k – the score of the secret key; Z_i – the output signal given by the neural distinguisher for the *i*-th input sample.

For the baseline model, the group size *k* was set to 1. For the 32 initial keys, the Euclidean distance score between each initial key and each candidate key whose difference range was calculated, and the 32 keys with the smallest distance were screened out as the candidate keys for the next iteration – $[0,2^{16})$. By this method, each ciphertext structure only needs to try 160 candidate keys in the last round, and finally output the adopted keys and the corresponding scoring results. For candidate keys with scores exceeding the preset threshold $C_{11} = 10$, the Bayesian optimisation algorithm was further employed to filter out 160 candidate keys for 10 rounds and calculate the corresponding scores. If the score still exceeds the threshold $C_{10} = 10$, then the two selected keys were output as the final key guessing result. The success of key recovery was evaluated by calculating the bit difference between the guessed key and the real key. In this paper, five key recovery tests were carried out, each with 20 key recovery attacks, a total of 100 key recovery attacks, and the specific recovery situation is shown in Table 1.

Table 1. Results of key recovery attacks

Experimental Rounds	1	2	3	4	5
Number of successes in 20 recovery experiments	9	9	14	10	14
Data Complexity					
Success rate	45%	45%	70%	50%	70%

Source: compiled by the authors

As shown in Table 1, in 100 secret key recovery attacks using the B-C3-HSwish differential discriminator, the number of successful secret key recoveries reached 56 times, with a success rate of approximately 56%. The average data complexity of the attack was between $2^{13.5}$ and $2^{13.6}$. Compared to the success rate of 52.1% obtained by this method in reference, the B-C3-HSwish differential discriminator demonstrated higher accuracy in secret key recovery attacks. This was expected because the discriminator has higher discrimination accuracy in single-pair input mode.

Between 2020 and 2025, with the rapid development of deep learning technology, neural networks have been introduced into differential cryptanalysis, which has brought new research directions and possibilities to this field. Numerous studies have demonstrated that neural networks offer significant advantages in learning and recognising the internal differences characteristic of encryption algorithms. For instance, H. Niu *et al.* (2025) stated that A. Gohr was the first to propose the use of neural networks to construct a differential distinguisher, achieving superior results compared to traditional methods when analysing the lightweight block cipher Speck. Building on this foundation, the authors explored the potential of deep learning to further improve the effectiveness of differential distinguishers by introducing alternative network architectures and optimisation strategies. Despite these advancements, current neural network-based differential distinguishers still face several challenges. C. Jia *et al.* (2025) emphasised that classification accuracy remains a key concern, particularly when dealing with high-round

encryption, where the accuracy tends to decline significantly. Similarly, Y. Liu *et al*. (2024) observed that most existing distinguishers are only effective against a limited number of encryption rounds, and the performance becomes restricted when applied to more complex or higher-round cipher structures. In contrast to these challenges, this research demonstrated significant progress: by using an improved B-C3-HSwish discriminator and integrating Bayesian optimisation, a 56% success rate was achieved in key recovery attacks on 11 rounds of Speck32/64. This result not only confirmed the high accuracy of this model, but also directly addressed the problem of limited effectiveness at high rounds, setting a new benchmark for attacks in a single ciphertext scenario with low data complexity.

A study by N. Mukwevho & C. Chibaya (2020) focused on the design and evaluation of key scheduling algorithms (KSAs) in symmetric encryption systems. The research emphasised the critical role of KSAs in enhancing the security of block ciphers against cryptanalytic attacks. The proposed KSA architecture aimed to balance computational efficiency with cryptographic robustness, incorporating advanced algorithmic constructs to minimise processing time and resource consumption. Through statistical analyses, including entropy measures and non-linear complexity assessments, the study demonstrated the resilience of the proposed KSA against various cryptanalytic benchmarks, highlighting its superior performance and security enhancements over existing standards. This approach, on the contrary, evaluated the effectiveness of an attack rather than the robustness of the defence. Whereas studies such as N. Mukwevho & C. Chibaya (2020) focused on how to strengthen KSA to make key recovery more difficult, this work demonstrates that even with a potentially robust KSA (as in the case of Speck32/64), key recovery mechanisms based on deep learning and Bayesian optimisation can circumvent these barriers. In contrast, P. Yadav *et al*. (2025) conducted research, which highlighted the interdisciplinary nature of contemporary research and the importance of rigorous evaluation in complex systems. S. Dey *et al*. (2024) proposed HiSE, a hierarchical (threshold) symmetric-key encryption scheme aimed at enhancing scalability and decentralisation in symmetric encryption frameworks. This research addressed the distribution of secret keys to mitigate single points of attack, incorporating threshold mechanisms within a hierarchical structure. Evaluations showed that HiSE could support large-scale, decentralised applications while maintaining strong security assurances.

J. Chien (2020) addressed the fundamentals and advances in deep Bayesian mining and learning for natural language, exploring applications such as speech recognition, document summarisation, text classification, and machine translation, while emphasising the limitations of traditional deep learning in expressing semantic structures and optimising distribution functions. The tutorial highlighted advanced Bayesian and deep models, including hierarchical Dirichlet processes, recurrent neural networks, variational auto-encoders, and generative adversarial networks, and discussed the connections and effectiveness in handling symbolic and complex patterns in natural language. Variational inference and sampling methods were formulated to optimise complex models, and case studies were presented to illustrate applications in deep Bayesian mining, learning, and understanding. The tutorial concluded with future directions, aiming to introduce novices to deep Bayesian learning, motivate emerging topics in data mining and natural language understanding, and synthesise distinct lines of machine learning research. L. Lyu *et al*. (2022) improved upon existing framework by integrating deep learning into differential cryptanalysis, successfully identifying 8/9/10-round neural differential distinguishers for Simeck32/64 and recovering subkeys for 13/14/15-round Simeck32/64 with low data and time complexity, while also sharing parameter tuning insights for deep learning-assisted key recovery attacks. The application of the Bayesian optimisation algorithm in cryptanalysis is mainly reflected in automatic distinguisher design and key recovery attack, as evident in the work of B. Seok & C. Lee (2025). In the design of the automatic distinguisher, a Bayesian optimisation algorithm can be used to automatically search the optimal distinguisher parameters, so as to improve the performance of the distinguisher. In the key recovery attack, the Bayesian optimisation algorithm can be used to guide the key space search and infer the most probable key by evaluating the similarity between the encryption results of different key candidates and the actual ciphertext. Y. Yang *et al*. (2024) stated that although the Bayesian optimisation algorithm has shown great potential in cryptanalysis, its application in single ciphertext attacks is still relatively rare.

Findings made in this research highlighted the potential of neural networks and Bayesian optimisation in advancing cryptanalysis, particularly in scenarios with limited data. The proposed method not only improved the success rate of key recovery but also reduced computational complexity, making it a practical and reliable solution for single-ciphertext attacks. Furthermore, the study underscored the importance of integrating machine learning techniques with cryptographic analysis to address emerging challenges in the field.

CONCLUSIONS

This study addressed the key recovery problem in single-ciphertext attack scenarios by proposing a key recovery attack method based on the Bayesian optimisation algorithm and the UCB strategy. By optimising the structure and input representation of the neural network differential distinguisher and integrating the Bayesian optimisation algorithm to intelligently search the key space, the proposed method significantly improves the

efficiency and accuracy of key recovery. Experimental results demonstrated that the method achieved a 56% success rate in 11 rounds of key recovery attacks, with a data complexity with a wide range, which is notably superior to traditional methods. This achievement highlighted the method's effectiveness and reliability in single-ciphertext attack scenarios. The study also revealed the powerful feature extraction capability of neural network differential distinguishers in single-ciphertext scenarios, enabling the method to efficiently identify key patterns with limited data. Additionally, the Bayesian optimisation algorithm proved to be highly efficient in high-dimensional search tasks, significantly reducing the computational overhead of key recovery. These findings not only provided a new technical pathway for key recovery attacks but also opened up new directions for the application of the Bayesian optimisation algorithm in the field of cryptanalysis. The combination of neural networks and Bayesian optimisation represented a novel approach that bridges the gap between machine learning and cryptographic analysis, offering a robust framework for future research. Looking ahead, this method can be further explored and applied to other encryption algorithms, potentially extending its utility to a broader range of cryptographic systems. Integrating more advanced machine learning techniques, such as deep reinforcement learning or transformer-based models, could further enhance the method's performance and applicability. Additionally, future work could focus on optimising the algorithm's design to reduce data complexity and improve success rates in more challenging scenarios. Through continuous refinement and validation, this study provided a solid foundation for advancing key recovery techniques and contributed to the development of more secure cryptographic systems.

ACKNOWLEDGEMENTS

None.

FUNDING

This research was sponsored by the Bengbu University Key Natural Science Research Project (Grant No. 2024ZR03zd).

CONFLICT OF INTEREST

None.

REFERENCES

[1] Auer, P. (2020). Using upper confidence bounds for online learning. In *Proceedings of the 41st annual symposium on foundations of computer science* (pp. 270-279). Redondo Beach: IEEE. doi: 10.1109/SFCS.2000.892116.

[2] Baksi, A., Breier, J., Chen, Y., & Dong, X. (2021). Machine learning assisted differential distinguishers for lightweight ciphers. In *2021 design, automation & test in Europe conference & exhibition (DATE)* (pp. 176-181). Grenoble: IEEE. doi: 10.23919/DATE51398.2021.9474092.

[3] Benamira, A., Gerault, D., Peyrin, T., & Tan, Q. (2021). A deeper look at machine learning-based cryptanalysis. In A. Canteaut & F.-X. Standaert (Eds.), *Advances in cryptology – EUROCRYPT 2021* (pp. 805-835). Cham: Springer. doi: 10.1007/978-3-030-77870-5_28.

[4] Chien, J.-T. (2020). Deep Bayesian data mining. In *Proceedings of the thirteenth ACM international conference on web search and data mining (WSDM '20)* (pp. 865-868). New York: Association for Computing Machinery. doi: 10.1145/3336191.3371870.

[5] Dey, S., Leander, G., & Sharma, N.K. (2024). Improved key recovery attacks on reduced-round Salsa20. *Designs, Codes and Cryptography*, 93, 243-262. doi: 10.1007/s10623-024-01522-7.

[6] Gohr, A. (2019). Improving attacks on round-reduced Speck32/64 using deep learning. In A. Boldyreva & D. Micciancio (Eds.), *Advances in cryptology – CRYPTO 2019* (pp. 150-179). Cham: Springer. doi: 10.1007/978-3-030-26951-7_6.

[7] Gohr, A., Leander, G., & Neumann, P. (2022). An assessment of differential-neural distinguishers. *Cryptology ePrint Archive*, 2022, article number 1521.

[8] Hou, Z., Ren, J., & Chen, S. (2021). Improve neural distinguisher for cryptanalysis. *Cryptology ePrint Archive*, 2021, article number 1017.

[9] Jia, C., Chen, J., Li, X., Zheng, H., & Zhang, L. (2025). BallPri: Test cases prioritization for deep neuron networks via tolerant ball in variable space. *Automated Software Engineering*, 32(1), article number 29. doi: 10.1007/s10515-025-00498-5.

[10] Jiang, X., Li, M., Makulov, K., Lakhno, V., & Sahun, A. (2025). Enhanced neural differential distinguisher for Speck32/64 using attention mechanisms and multi ciphertext inputs. *Informatica*, 49(19), article number 7889. doi: 10.31449/inf.v49i19.7889.

[11] Lin, D., Chen, S., Li, M., & Hou, Z. (2022). The construction and application of (related-key) conditional differential neural distinguishers on KATAN. In A.R. Beresford, A. Patra & E. Bellini (Eds.), *Cryptology and network security* (pp. 203-224). Cham: Springer. doi: 10.1007/978-3-031-20974-1_10.

[12] Liu, Y., Yang, L., Chen, J., Wu, W., & Feng, Y. (2024). Matrix computation over homomorphic plaintext-ciphertext and its application. *Journal on Communication. Series: Tongxin Xuebao*, 45(2), 150-161. doi: 10.11959/j.issn.1000-436x.2024024.

[13] Lyu, L., Tu, Y., & Zhang, Y. (2022). Deep learning assisted key recovery attack for round-reduced Simeck32/64. In W. Susilo, X. Chen, F. Guo, Y. Zhang & R. Intan (Eds.), *Information security* (pp. 443-463). Cham: Springer. doi: 10.1007/978-3-031-22390-7_26.
[14] Mukwevho, N., & Chibaya, C. (2020). Dynamic vs static encryption tables in DES key schedules. In *2020 2nd international multidisciplinary information technology and engineering conference (IMITEC)* (pp. 1-5). Kimberley: IEEE. doi: 10.1109/IMITEC50163.2020.9334110.
[15] Niu, H., McCallum, G.B., Chang, A.B., Khan, K., & Azam, S. (2025). Exploring unsupervised feature extraction algorithms: Tackling high dimensionality in small datasets. *Scientific Reports*, 15, article number 21973. doi: 10.1038/s41598-025-07725-9.
[16] Seok, B., & Lee, C. (2025). A novel approach to construct a good dataset for differential-neural cryptanalysis. *IEEE Transactions on Dependable and Secure Computing*, 22(1), 246-262. doi: 10.1109/TDSC.2024.3387662.
[17] Yadav, P., Mittal, K., Gill, P.S., Mittal, A., Raghuraman, K., & Kaushik, J.S. (2025). Role of intravenous azithromycin as adjunctive therapy in children with acute encephalitis syndrome (AES): An open-label randomized controlled trial. *Cureus*, 17(7), article number e87387. doi: 10.7759/cureus.87387.
[18] Yang, Y., Xiong, X., Liu, Z., Jin, S., & Wang, J. (2024). High-performance encryption algorithms for dynamic images transmission. *Electronics*, 13(1), article number 131. doi: 10.3390/electronics13010131.
[19] Zhang, L., & Wang, Z. (2022). Improving differential-neural distinguisher model for DES, Chaskey, and PRESENT. *ArXiv*. doi: 10.48550/arXiv.2204.06341.
[20] Zhu, S., Li, L., Hu, Z., & Hu, Y. (2025). BCS: A neural distinguisher method based on differential propagation uncertainty of nonlinear components and network adaptability. *Physica Scripta*, 100(3), article number 035008. doi: 10.1088/1402-4896/adae63.

Розробка методології для навчання нейронних диференціальних розпізнавачів та атаки на відновлення ключа для окремого шифрованого тексту

Сюе Цзян

Аспірант
Національний університет біоресурсів і природокористування України
03041, вул. Героїв Оборони, 15, м. Київ, Україна
https://orcid.org/0009-0000-1676-2331

Валерій Лахно

Доктор технічних наук, професор
Національний університет біоресурсів і природокористування України
03041, вул. Героїв Оборони, 15, м. Київ, Україна
https://orcid.org/0000-0001-9695-4543

Анотація. Атаки на відновлення ключів становлять серйозну загрозу для криптографічних систем, оскільки вони безпосередньо ставлять під загрозу безпеку механізмів шифрування. Метою цього дослідження було розроблення нового методу відновлення ключів з одним шифрованим текстом, який усуває обмеження традиційних методів криптоаналізу в сценаріях з обмеженою доступністю даних. Методологія поєднувала вдосконалений нейронний дискримінатор B-C3-HSwish з байєсівською оптимізацією, керованою стратегією верхньої межі довіри (UCB). Нейронний дискримінатор було вдосконалено за допомогою багатомасштабних конволюційних шарів, оптимізації функції активації та диверсифікованих структур вхідних даних, тоді як байєсівська оптимізація відобразила проблему відновлення ключів на високорозмірне завдання пошуку за допомогою сурогатної моделі гауссового процесу та функції придбання UCB. Експериментальні результати щодо алгоритму шифрування Speck32/64 продемонстрували 56 % успішність у 11 раундах атак на відновлення ключа, причому складність даних коливалася від 213,5 до 213,6. Цей результат перевершує 52,1 % успішність, отриману в еталонному дослідженні. Нейронний дискримінатор B-C3-HSwish був вдосконалений, зокрема, шляхом заміни функції активації Relu на Hard_swish для підвищення його точності в режимі введення однієї пари. Це дослідження не тільки пропонує нове, ресурсоефективне рішення для відновлення ключів у сценаріях з одним шифрованим текстом, але й підкреслює потужний потенціал байєсівської оптимізації як перспективної парадигми для розвитку криптоаналізу. Практична цінність цієї роботи полягає в її здатності надати ресурсоефективне рішення для відновлення ключів у сценаріях з одним шифрованим текстом, що робить її особливо придатною для реальних застосувань з обмеженими даними. Цей підхід також встановлює байєсівську оптимізацію як перспективну парадигму для просування криптоаналізу та підвищення криптографічної безпеки

Ключові слова: алгоритм байєсівської оптимізації; оптимізація вилучення ознак; сценарій атаки; алгоритм Speck32/64; нейронні мережі

BULLETIN of Cherkasy State Technological University

Journal homepage: https://bulletin-chstu.com.ua/en

Vol. 31 No. 1. 2026

UDC 004.5
DOI: 10.62660/bcstu/1.2026.34

ISSN: 2306-4412
E-ISSN: 2708-6070

Method for detecting computer attacks based on adaptability and balancing of training samples

Bogdan Semenyuk*

Postgraduate Student
Khmelnytskyi National University
29016, 11 Instytutska Str., Khmelnytskyi, Ukraine
https://orcid.org/0009-0001-8831-8835

Bogdan Savenko

Doctor of Philosophy, Associate Professor
Khmelnytskyi National University
29016, 11 Institutska Str., Khmelnytskyi, Ukraine
https://orcid.org/0000-0001-5647-9979

Abstract. The relevance of this work is determined by the fact that modern cyber defence systems still face the problem of data imbalance in detecting computer attacks. Uneven class representation reduces the ability of machine learning models to recognise rare but critically important types of intrusions. Existing methods of synthetic sample augmentation often distort the data structure and lead to the loss of characteristic attack signatures, which negatively affects the reliability of classification. The aim of the study was to improve the accuracy and completeness of network attack detection by developing an adaptive method for balancing the training sample while preserving the statistical and signature properties of anomalies. The research methodology involved the creation of a deterministic algorithm that calculated the supplementation coefficients separately for each type of attack, taking into account its frequency and the minimum required sample size. A statistical approach based on local medians and extreme values of neighbouring samples was used to generate new examples, which ensured the reproduction of the typical structure of anomalous patterns without random interpolation. The developed method was integrated into a data processing sequence, within which network traffic parameters were converted into sound signals, and spectrograms were formed on their basis for further analysis by a two-dimensional convolutional neural network. The main results of the study showed an increase in the completeness of detection of rare types of attacks by an average of 10-12% compared to basic approaches, while maintaining a stable level of overall accuracy. Preserving signatures in synthetic samples ensured improved recognition of rare attacks and increased classification reliability. The practical value of the work lies in the possibility of applying the developed method to form balanced samples in intrusion detection systems and its integration with existing deep learning models in order to improve the reliability of corporate network cyber protection

Keywords: data imbalance; signature modelling of anomalies; synthetic sample augmentation; network traffic analysis; convolutional neural networks; acoustic data representation; classification robustness

INTRODUCTION

The rapid growth of the Internet of Things (IoT) ecosystem has led to an unprecedented increase in the number of connected devices and the amount of data generated in real time. The massive introduction of sensor networks, household smart devices, and industrial IoT solutions has created a complex and dynamic environment in which the speed of information exchange is constantly increasing. However, this scale

Article's History: Received: 05.11.2025; Revised: 13.02.2026; Accepted: 16.03.2026; Published: 08.04.2026

Suggested Citation:
Semenyuk, B., & Savenko, B. (2026). Method for detecting computer attacks based on adaptability and balancing of training samples. *Bulletin of Cherkasy State Technological University*, 31(1), 34-43. doi: 10.62660/bcstu/1.2026.34.

*Corresponding author

of connection is accompanied by an increase in the number of vulnerabilities.

T. Alladi *et al.* (2020) showed that IoT devices are often targeted by attacks due to limited resources, weak authentication protocols, and insufficient isolation of network segments. In these circumstances, intrusion detection systems that can track abnormal behaviour and respond quickly to threats become a critical component of protection. As noted by A. Hussain *et al.* (2023), the advanced Intrusion Detection System (IDS) for IoT environments serves as an adaptive monitoring mechanism that analyses network traffic, detects deviations from typical patterns, and complements other cyber defence tools, ensuring infrastructure resilience to new and complex attacks. The development of intrusion detection systems is largely associated with the introduction of machine and deep learning methods that provide the ability to automatically analyse complex network traffic patterns and adapt to new types of attacks. Comparative study by I.F. Kilincer *et al.* (2021) showed that machine learning algorithms demonstrate high efficiency in classifying network events due to their ability to consider multidimensional traffic features and detect anomalies that are not detected by conventional methods. Further improvement of accuracy is possible due to the use of deep models: G. Li & J.J. Jung (2023) emphasised that convolutional and other deep learning (DL) architectures are capable of modelling complex nonlinear dependencies that occur in multicomponent time series, providing higher noise resistance and efficient detection of previously unknown types of anomalies. The success of using Machine Learning (ML) or DL in IDS systems largely depends on the structure and balance of training data sets, which form the basis for correct training of models and prevent degradation of results when classifying rare attacks.

Ukrainian researchers also made a significant contribution to the development of methods for analysing anomalies. In particular, D. Denysiuk *et al.* (2023) confirmed the effectiveness of using local statistical features to detect hidden modifications in steganographic structures. The study by A. Kashtalian *et al.* (2025) was dedicated to the architecture of centralised multi-computer systems with built-in traps and baits, which increases the efficiency of detecting abnormal activity in distributed environments. A separate group of studies focused on the impact of sample imbalance on the quality of models. In particular, J. Dalou' *et al.* (2020) described an entropy-oriented approach for detecting Distributed Denial of Service (DDoS) attacks in software-defined networks (SDN), which allows adaptive response to changes in traffic intensity. The study by J.E. Joseph *et al.* (2025) proved that the accuracy of classification directly depends on the balance of the training sample, and depth models are not able to compensate for the lack of representation of rare attacks.

Analysis of publications in 2020-2025 showed that the problem of qualitative balancing of training samples in the context of anomaly detection remains open. Most of the available approaches do not consider the internal structure of network traffic, which leads to a loss of signature characteristics of attacks. In addition, many studies lack mechanisms for adaptive selection of resampling coefficients for individual classes, as a result of which rare types of attacks remain underrepresented. Despite significant progress in the development of data balancing methods for computer attack detection tasks, existing approaches remain vulnerable to a number of technical limitations. Classical synthetic complement algorithms, although they partially compensated for the lack of rare attacks, often led to distortion of local structural properties of data, loss of signatures, and a decrease in the ability of models to correctly reproduce class boundaries. Additional difficulties arose due to the dependence of such methods on global parameters, which did not consider the heterogeneity of attacks and the variability of their internal subtypes. Under these conditions, there is a need for an approach that can overcome these shortcomings, preserve the characteristic features of rare samples, and ensure correct integration into advanced traffic processing pipelines, which include data conversion into spectrograms and subsequent classification by convolutional neural networks. It was certain limitations that became the basis for the development of the proposed balancing method aimed at overcoming these structural and methodological problems.

The purpose of the study was to develop and experimentally test the adaptive signature of a method for synthetic balancing of the training sample, aimed at improving the completeness of detection of rare attacks without reducing the overall classification accuracy. It was assumed that the use of this method will ensure the reproducibility of the new generation intrusion detection system, improve its resistance to class imbalance, and allow forming synthetic samples that correctly transmit local features of network traffic without losing signature properties.

MATERIALS AND METHODS

The purpose of the study was to develop and experimentally verify the adaptive signature of a method for synthetic balancing of training samples for computer attack detection tasks. The methodology included a combination of classical approaches to working with tabular features of network traffic, advanced techniques for preprocessing and converting data into two-dimensional structures, and the use of a compact convolutional neural network. All experiments were performed under the same conditions, with fixed initial values of random number generators, which ensured complete reproducibility of the simulation results. The paper used tabular parameters of network traffic, which contained numerical and categorical features. The imbalance of such samples is a common problem, and the lack of representation of rare attacks in training data leads to a systematic deterioration in the ability of models to

correctly recognise complex types of threats. Therefore, pretreatment provided for the need to correct the imbalance. Categorical attributes (protocol_type, service, flag) were encoded in whole indexes and then processed using the Synthetic Minority Oversampling Technique for Nominal and Continuous features (SMOTENC), which allowed their nature to be correctly taken into consideration. Numeric features were scaled to a range of -1 to 1, which stabilised calculations in the feature space and improved the behaviour of balancing algorithms.

A special place in the methodology was occupied by the stage of converting tabular data into a two-dimensional form using sonification. Each network traffic record was converted to a Pulse-Code Modulation (PCM) signal with a sampling rate of 2,000 Hz and duration of 0.4 seconds. The signals were pretreated by converting them into spectrograms generated using a short-time Fourier Transform (STFT) with parameters nperseg = 256 and noverlap = 128. The resulting spectrograms were converted to a logarithmic scale, after which they were interpreted as single-channel images suitable for further processing by a convolutional neural network. The imbalance was eliminated exclusively for the training part of the data, until the final scaling and coding. The modified version of SMOTENC was applied selectively – only to those subtypes of attacks that were characterised by a low detection completeness index or had an insufficient number of examples. Complement volumes were controlled using the SMOTE_MAX_MULTIPLIER and SMOTE_MIN_COUNT parameters, the limits for which were determined empirically during stratified five-part cross-validation. Balancing efficiency was evaluated using a macro-average completeness indicator (macro-recall), and monitoring the level of false-positive classifications.

A compact two-dimensional convolutional neural network was used for classification. At the first stage, the model used a Convolutional 2D convolutional layer with 16 3x3 filters with Rectified Linear unit activation, which helped to distinguish local patterns of spectral representations. The features were then aggregated using the maximum pooling sub-sample pooling layer, which reduced the dimension of the spectrogram and increased resistance to local fluctuations. Similar architectures are widely used in modern network anomaly detection systems.

The training was performed using the Adam optimiser at a rate of 5×10^{-4}, parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and the binary cross-entropy loss function. The package size was 128 objects. To prevent retraining, an early stopping mechanism was used with the patience = 3 parameter, monitoring the val_loss indicator, and checking the stability of the model behaviour. The effectiveness of the attack detection model was evaluated based on metrics that are sensitive to class imbalances, in particular, precision, recall, and F_1-measure, which reflects their harmonious average. The use of such metrics was critical for tasks where rare classes are crucial for the quality of IDS work, and an uneven number of examples can lead to a significant shift in results.

All experimental studies were conducted using the Network Security Laboratory – Knowledge Discovery Dataset (NSL-KDD), which contains 125,973 training examples and 22,544 testing examples and is an improved and duplicated version of the original KDD'99. NSL-KDD covered 41 network attributes, including both numeric and categorical parameters, and attacks were grouped into four key categories: DoS, Scanning/Probing attacks, User-to-Root, and Remote-to-Local. As part of this study, three comparable systems were formed: a basic model without balancing, a variant with classical SMOTENC, and a system with a signature – saving adaptive approach. All experiments were performed with the same learning parameters and the same model architecture, which ensured the validity and correctness of comparing the results.

RESULTS AND DISCUSSION

Below is a mathematical formalisation of a method for detecting computer attacks based on adaptability and balancing of the training sample. Formalisation describes the mechanisms for determining the target number of synthetic examples, local statistics, bias directions, and principles for preserving the signature properties of data that distinguish the proposed approach from classical interpolation algorithms. Parameter $c > 0$ determined the sensitivity of the deterministic deviation coefficient $\beta(x)$, given by equation (7), to the local distance $d(x)$, calculated according to (6). Lower values of c led to a more drastic change of $\beta(x)$ relative to $d(x)$, while larger – contributed to smoothing and reducing the variability of synthesis. Parameter $\gamma_j \in [0,1]$ was used to determine the degree of shift of synthetic coordinates towards local extremes, which helped to preserve the characteristic signature properties of the sample without distorting its statistical structure. The selection of parameters c and γ_j was performed empirically using a five-component stratified cross-validation method with the macro-recall target metric and control of the false positive rate. A grid search was performed for each parameter $c \in \{0,05; 0,1; 0,2; 0,5; 1,0\}$, $\gamma \in \{0;0,05;0,1;0,2;0,3\}$, after normalising the data to the range [-1,1]. Additionally, a limit on the median absolute deviation (MAD) was introduced:

$$|\gamma_j (e_j - x_j^{syn})| \leqslant 1.5 \cdot MAD_j, \qquad (1)$$

where $\gamma_j \in [061]$ – extreme gain for j-th coordinate, $e_j \equiv e_j(x)$ – local extremum (minimum or maximum) j-th numerical feature in the sample area x, x_j^{syn} – j-th coordinate of the synthetic sample, MAD_j – median absolute deviation for j-th coordinate within the attack class t, $|\cdot|$ – module.

This prevented the development of synthetic anomalies. The optimal values were set as $c = 0.2$ and $\gamma_j = 0.1$, since

they provided a stable trade-off between stability and preservation of local signatures with controlled synthesis variability in the selected CNN architecture.

A set of training examples for the attack class t:

$$S_t = \{x^{(i)}\}_{i=1}^{n_t}, \quad (2)$$

where t – index of the attack type (class), n_t – number of training examples of this class, $x^{(i)}$ – i-th feature vector divided into a numeric part $x_{num}^{(i)} \in R^d$ (with d numeric attributes) and the categorical part $x_{cat}^{(i)}$. The target number of examples after resampling is denoted as N_t^{target}. Enter parameters k – number of nearest neighbours used to estimate the local structure of the class, M – upper limit of relative class expansion, N_{min} – minimum allowed number of examples for rare classes after balancing.

The regular goal is calculated using the equation:

$$\tilde{N}_t = min(\lfloor n_t \cdot M \rfloor, n_{max}), \quad (3)$$

where $\tilde{N}_t$ – target number of examples before threshold correction, n_{max} – maximum number of examples among all abnormal (attack) classes, $\lfloor \cdot \rfloor$ – rounding down to an integer operation:

$$N_t^{target} = max(\tilde{N}_t, N_{min}). \quad (4)$$

That is N_t^{target} – final target number of class examples t after accounting as a relative constraint M, and the absolute threshold N_{min}.

Local statistics are defined for each sample. The set $N_k(x)$ includes k nearest neighbours within the class S_t based on Euclidean or mixed distance. Local median of numeric coordinates:

$$m_{loc}(x) = median(\{y_{num} : y \in N_k(x) \cup \{x\}\}), \quad (5)$$

where $m_{loc}(x) \in R^d$ – element-wise local median of numeric coordinates in the vicinity x, y_{num} – numeric part of a neighbouring vector y.

Percentiles are defined: $p_{10}(x)$, $p_{90}(x)$. Distance to the local median:

$$d(x) = \| x_{num} - m_{loc}(x) \|_2, \quad (6)$$

where $\|\cdot\|_2$ – Euclidean norm, $d(x) \geq 0$ – local distance of the sample x to its local median $m_{loc}(x)$.

It is used to calculate the deterministic displacement coefficient:

$$\beta(x) = clamp\left(\frac{d(x)}{d(x)+c}, 0\beta 1\right), \quad (7)$$

where $\beta(x) \in [0,1]$ – deterministic bias coefficient for the sample x, $c > 0$ – parameter that adjusts the sensitivity of $\beta(x)$ to the distance $d(x)$, $clamp(z, 0, 1) = min(max(z, 0), 1)$ – value constraint operator z with interval [0,1]. This is a deterministic function (no randomness).

In the standard Synthetic Minority Oversampling Technique (SMOTE), categorical values are selected as a mode among neighbours:

$$x_{cat,j}^{syn} = arg \max_v \#\{y \in \mathcal{N}_k(x) \cup \{x\} : y_{cat,j} = v\}, \quad (8)$$

where $x_{(cat,j)}^{syn}$ – value of j-th categorical feature in a synthetic sample, $y_{(cat,j)}$ – value of the same attribute in a neighbouring sample y, v – possible categorical value, $\#\{\cdot\}$ – power of the set (number of elements). Thus, for categorical features, the mode is selected in the neighbourhood x.

In the proposed SPAS method, the amplitude of local variation is defined as:

$$\Delta(x) := p_{90}(x) - p_{10}(x) \in R^d, \quad (9)$$

where $p_{10}(x)$, $p_{90}(x) \in R^d$ – element-wise 10th and 90th percentiles of numerical coordinates in the local neighbourhood x, $\Delta(x)$ – vector of the local amplitude of variation for each numerical feature.

Direction $s(x)$ is defined deterministically as the sign of the average deviation:

$$s_j(x) = sign\left(\frac{1}{k}\sum_{y \in \mathcal{N}_k(x)} (x_j - y_j)\right) \in \{-1,0,1\}, \quad (10)$$

where $s_j(x)$ – sign of average deviation of j-th coordinates of the sample x from its neighbours, x_j, y_j – value of j-th numerical attribute, respectively, for samples x and y, $sign(\cdot)$ – standard sign function that returns -1, 0, or 1.

A synthetic numerical vector is obtained by the equation:

$$x_{num}^{syn} = m_{loc}(x) + \beta(x)(\Delta(x) \circ s(x)), \quad (11)$$

where $x_{num}^{syn} \in \mathbb{R}^d$ – numerical part of the synthetic sample, $s(x) \in \{-1, 0, 1\}^d$ – direction vector for each coordinate, "$\circ$" – element-wise (Hadamard) multiplication of vectors. If for a certain coordinate , the offset direction is additionally set by the nearest neighbour.

Saving extremes (signature-preserving shift). For coordinates where extreme values exist in neighbours (outside), an additional offset is introduced:

$$x_j^{syn} \leftarrow x_j^{syn} + \gamma_j(e_j(x) - x_j^{syn}), \quad (12)$$

where x_j^{syn} – current value of j-th coordinates of the synthetic sample, $e_j(x)$ – local extremum (minimum or maximum) of j-th coordinates among the set $N_k(x) \cup \{x\}$, $\gamma_j \in [0,1]$ – coefficient that sets the fraction of "pull-up" to the extreme value, operator "$\leftarrow$" means updating the coordinate.

Categorical features in classical methods, the choice of categories is random, and in the proposed method:

$$x_{num}^{syn} = m_{loc}(x) + \beta(x)(\Delta(x) \circ s(x)), \quad (13)$$

in other words, the mode is selected. If several categories have the same frequency, the nearest neighbour

category is used. Index j numbers categorical attributes, and $x_{(cat,j)}^{syn}$ – value of j-th categorical feature in a synthetic sample.

The built formal device allows generating signature-resistant synthetic samples that preserve the structure of rare attacks. After mathematical specification, an experimental verification of the method was carried out on the NSL-KDD benchmark set, within which a comparison was made between the basic model without balancing, the variant with classic SMOTENC, and the model using the proposed SPAS approach In the basic model, which was trained on initially unbalanced data, the following results were obtained: the average error in the test sample was 0.6538, the classification accuracy was 0.7654, and the indicator of completeness of attack detection was only 0.62. This indicates the model's tendency to overestimate normal traffic and its limited ability to recognise rare types of attacks.

The analysis of the results obtained is consistent with the conclusions of other researchers. According to M. Farooq & F. Ahmad (2024), even well-optimised multilevel models show a significant drop in completeness for rare attacks due to the predominance of normal traffic, where convolutional architectures have been shown to retain sensitivity to class unevenness even after hyperparameter optimisation. This relationship can also be traced in the obtained experiments: the model overestimates the normal class and loses a significant part of the rare attacks, which corresponds to the general patterns described in contemporary studies. As shown in Table 1, the majority of errors in the baseline model occurred when attacks were classified as normal activity (3,484 false-negative triggers), while the number of false-positive classifications was relatively small (734 cases). This distribution confirms an imbalance in the representation of classes and a lack of ability of the model to detect rare events.

Table 1. Interface matrix

Actual \ Predicted	Pred = 0 (Normal)	Pred = 1 (Attack)
Actual = 0 (Normal)	8,977	734
Actual = 1 (Attack)	3,484	9,349

Source: compiled by the authors

Using the classic SMOTENC improved the situation: the test error decreased to 0.4573, the accuracy increased to 0.7900, and the completeness increased to 0.70. Partial indicators (Table 2) showed a significant increase in recall for the mscan (0.82) and neptune (0.99) attack types, although the values remained low for the warezmaster (0.26) and guess_passwd (0.30) subtypes, indicating a loss of signatures during interpolation.

Table 2. Recall indicators for individual types of attacks (clippings)

Attack	Support	TP	Recall
neptune	4,657	4,641	0.9966
guess_passwd	1,231	371	0.3014
mscan	996	819	0.8223
warezmaster	944	241	0.2553

Source: compiled by the authors

The results obtained on the use of classical SMOTENC are also consistent with the conclusions presented in the paper by P. Sathaporn *et al.* (2025), who demonstrated that interpolation methods can increase completeness for certain types of attacks, but simultaneously create the risk of distorting local statistical characteristics of data. The researchers emphasised that when modelling network anomalies, random or over-smoothed interpolation leads to the loss of signature patterns that are crucial for classifying complex or rare threats. This behaviour is particularly noticeable in conditions of significant imbalance, when small classes are represented by a limited number of samples, and therefore require a more accurate and structurally sound approach to synthetic complement. In the conducted experiment, this problem is manifested due to the low recall rates for warezmaster and guess_passwd, which confirms the system flaw of the standard SMOTE/SMOTENC. Figure 1 shows the value of attack detection completeness for the 20 most common classes when using regular SMOTE. The image shows a strong unevenness of the results: individual classes, in particular neptune, smurf, and portsweep, appear almost perfectly, while guess_passwd, mailbomb, and multihop remain virtually undetected.

Applying the proposed method, which takes signature properties into consideration when creating new examples, provided further improvements. In this case, the test error was 0.3985, the classification accuracy was 0.8129, and the completeness of attack detection was 0.73. Thus, the increase in completeness was 11% compared to the basic model and 3% compared to the classic SMOTENC. Results of measuring the time spent on training a convolutional neural network under various balancing strategies (Table 3). The basic model without balancing required 343 seconds for five training epochs, which corresponds to the average time of one epoch of 68.6 seconds. In the configuration with the proposed SPAS approach, which includes an additional stage of synthetic sample

addition based on modified SMOTENC, the total training time was 287 seconds, and the average epoch time was 57.4 seconds. Despite the increase in the size of the training sample (from 113,375 to 125,627 records), the training time does not increase, but rather decreases due to the stabilisation of the model behaviour and faster convergence. Thus, the additional costs of sample synthesis do not lead to a significant increase in the overall processing time and are acceptable for practical IDS systems.

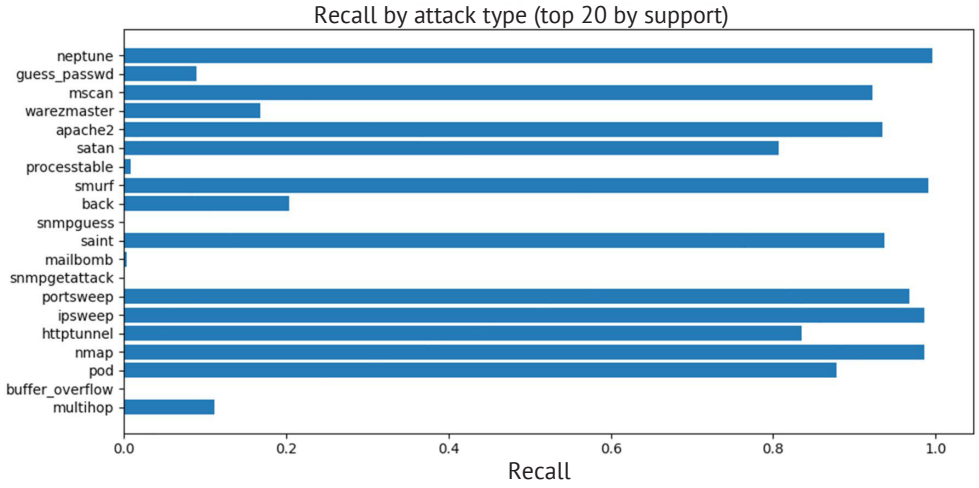

Figure 1. Recall by attack type conventional SMOTE

Source: compiled by the authors

Table 3. Time spent on different balancing configurations

Model configuration	Training sample size	Total training time, s	Average time of one epoch, s
No balancing (initial sampling)	125,627	343	68.6
Proposed SPAS (selective SMOTENC)	125,627	287	57.4

Source: compiled by the authors

Figure 2 shows a comparative diagram of detection completeness for the same types of attacks after applying the developed method. Guess_passwd and warezmaster scores improved significantly (an increase of 0.18 and 0.21, respectively), while the back and snmpguess classes approached the level of stable detection. The advantages of the signature-saving approach are consistent with a number of papers devoted to the use of local statistical features in anomaly detection problems, that the reproduction of local extremes and medians improves the resolution of models, which fully correlates with the result of increasing recall for problem attacks. That is why the SPAS approach provides not only higher averages, but also lower variability between attack types.

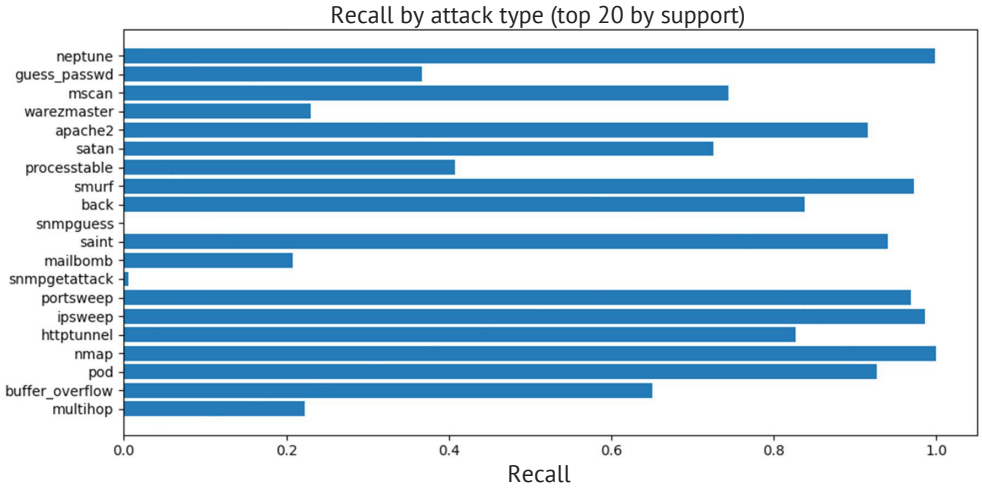

Figure 2. Recall by attack type offered by SMOTENC

Source: compiled by the authors

Along with an increase in the average values, there was a decrease in the spread between classes, which indicates a more stable behaviour of the model. Analysis of the error matrix confirms improvements in recognition of back, multihop, and rootkit attacks in the absence of an increase in false-positive positives. The difference does not exceed 18%, which makes the method suitable for use in practical IDS systems. Thus, the conducted experiments showed that the method of detecting computer attacks based on the adaptability and balancing of the training sample provides a sustainable improvement in the quality of the intrusion detection system due to the preservation of structural signatures of attacks and flexible adjustment of the volume of synthetic data supplementation without losing the overall classification accuracy.

In the broader context of modern balancing methods, the results of the proposed approach confirmed trends that emphasise the feasibility of adaptive control of the degree of data addition for different classes. The researchers point out that universal synthesis coefficients are not effective enough for highly sparse or structurally complex classes. The observations available in this paper fully correspond to this opinion: for a number of complex attacks (for example, guess_passwd, warezmaster), the adaptive approach provided a significant improvement due to deterministic bias formation and the use of local statistics, while the classical SMOTENC showed uneven results. Research on methods for detecting network attacks over the past decade has been intensively developing in two areas: improving deep classification models and developing approaches to balancing unbalanced samples. Despite a significant amount of studies, the problems of correct reproduction of rare attacks, stability of models on small classes, and preservation of signatures remain open.

One of the reference data sets for the analysis of IDS methods is NSL-KDD, which is widely used in the study of classification algorithms and evaluation of their resistance to class imbalance. According to M.S. Sheikh & Y. Peng (2022), this dataset remains suitable for analysing basic attack modelling patterns and improving recognition accuracy, making it a useful tool for experimental research in the field of network security. Although NSL-KDD is often criticised for breaking away from the real complexity of traffic, it provides a structured framework for comparing algorithms and reproducibility of experiments. Unlike its predecessor, NSL-KDD significantly eliminates the problem of an excessive number of identical records, which contributes to a more objective assessment of model performance and reduces the risk of shifting results towards duplicate samples. This characteristic is particularly important in the context of modern approaches to constructing IDS, where excessive correlation between samples can lead to an overestimated generalising ability of models. In addition, as noted by A.-G. Mari *et al.* (2023), NSL-KDD retains its popularity in intrusion detection systems research due to its balanced task structure, the presence of clearly delineated training and test subsets, and transparent classification of attack types, making it a convenient benchmark for benchmarking.

Moreover, numerous studies emphasise that the accuracy of IDS models significantly depends on the specifics of the data on which training is carried out. Z. Ahmad *et al.* (2021) noted that an imbalance between normal and abnormal traffic causes a significant decrease in the ability of models to detect complex or rare attacks. An additional complication is the high level of false-negative classifications reported by J. Mijalkovic & A. Spognardi (2022), emphasising that it is False Negative errors that pose the greatest danger to cyber defence systems, since they allow attacks to go unnoticed. A similar relationship was described by Z. Cai *et al.* (2023), who showed that an imbalance in training samples leads to a deterioration in the sensitivity of even well-optimised deep learning models. As stated by M. Sheibani *et al.* (2024), a comprehensive analysis of the sensitivity and specificity of the classifier allows adequately assessing its behaviour in cases where standard accuracy does not reflect the actual ability of the model to detect harmful actions. In this study, F_1-metric was used as the main indicator of balanced performance, while precision and recall were additionally monitored to assess the accuracy and completeness of classification in conditions of a significant imbalance in training data.

A separate area of research is devoted to improving the sustainability of IDS in IoT environments, where the volume, speed, and heterogeneity of traffic pose additional challenges. The paper by M. Baich & N. Sael (2025) emphasised that the intensive growth in the number of connected devices leads to the emergence of complex and persistent cyber threats, such as botnets, which are difficult to detect using classical models. The researchers demonstrated that the use of machine learning methods in combination with feature selection and data balancing mechanisms can significantly improve the quality of detection, reducing processing time, and increasing classification accuracy. M.H. Shahriar *et al.* (2020) showed that the use of generative adversarial networks (GAN) in IDS systems can serve as an additional mechanism for synthetic sample expansion, although the quality of the obtained samples significantly depends on the stability of the GAN model training.

Due to the problem of uneven representation of classes, a significant part of studies focused on the investigation of different approaches to data balancing. Classical resampling methods, in particular the SMOTE method and its modifications, are known for their ability to improve the representation of rare classes in a sample. J.H. Joloudari *et al.* (2023) showed that the use of SMOTE in combination with artificial neural networks significantly improves the accuracy of attack detection in the Bot-IoT Suite. The results show that classical synthetic complement methods often do not consider the local signature properties of rare attacks,

which can lead to a shift in the statistical data structure and a deterioration in the ability of models to correctly reproduce the features of complex samples.

The problem of feature selection also occupies an important place in contemporary research – IDS. Feature Selection methods allow reducing the dimension of data, increasing the information content of features, and reducing the risk of retraining. The paper by P. Maniriho *et al.* (2020) used a hybrid approach that combines Information Gain and Gain Ratio to select optimal subsets of features that are later used in the Random Forest model. This helped to obtain high accuracy of DoS classification, scanning, and MITM attacks. However, the researchers also noted that even with efficient feature selection, the problem of class imbalance remains a key factor that limits the ability of models to detect rare attacks.

Summarising, it can be noted that contemporary research confirms the need for methods that can work correctly with uneven samples, preserve the structural and signature properties of rare attacks, and ensure the stability of models in complex network environments. Despite the presence of numerous approaches, the problem of matching balancing, feature selection, and deep modelling remains insufficiently solved, which opens up opportunities for developing more adaptive and signature-saving methods that combine these components.

CONCLUSIONS

As a result of the research, a method for detecting computer attacks based on adaptive balancing of the training sample with the preservation of signature features was developed, which increased the efficiency of intrusion detection systems in environments with data imbalance. The proposed approach combined statistical sample analysis with a deterministic synthetic complement mechanism based on local medians, percentiles, and modal values. This approach ensures the preservation of characteristic attack structures and avoids distortion of internal patterns of network traffic during model training. A comparative experiment on the NSL-KDD reference set confirmed the effectiveness of the method: the completeness of attack detection increased from 0.62 to 0.73, and the overall classification accuracy increased from 0.76 to 0.81, while the average test error value decreased to 0.3985. The results show that the preservation of signatures during synthetic data supplementation increases the sensitivity of the model to rare attacks without reducing its generalising ability. A critical advantage was a significant improvement in the recognition of rare, structurally complex attacks, for which the classic SMOTENC was ineffective. Thus, the Recall indicator for the guess_passwd attack increased by about 0.18, and for warezmaster – by 0.21. This proves that deterministic signature-preserving synthesis is a reliable basis for improving the stability of intrusion detection systems. An important advantage of the proposed method is the determinism of the process of constructing new examples, which guarantees stability and reproducibility of the results. Integration of the method into the data processing pipeline – from converting network traffic to signal format (PCM) and building spectrograms to analysis using convolutional neural networks – helped to increase the efficiency of deep learning through a qualitatively balanced sample. Further research should focus on automating the selection of complement coefficients of the training sample, implementing Bayesian methods to optimise model parameters, and expanding the experimental base to data sets such as CIC-IDS2017 and UNSW-NB15. This will confirm the universality and adaptability of the proposed approach in real network traffic conditions and ensure further development of intelligent cyber defence systems.

ACKNOWLEDGEMENTS

None.

FUNDING

None.

CONFLICT OF INTEREST

None.

REFERENCES

[1] Ahmad, Z., Khan, A.S., Shiang, C.W., Abdullah, J., & Ahmad, F. (2021). Network intrusion detection system: A systematic study of machine learning and deep learning approaches. *Emerging Telecommunications Technologies*, 32(1), article number e4150. doi: 10.1002/ett.4150.

[2] Alladi, T., Chamola, V., Sikdar, B., & Choo, K.-K.R. (2020). Consumer IoT: Security vulnerability case studies and solutions. *IEEE Consumer Electronics Magazine*, 9(2), 17-25. doi: 10.1109/MCE.2019.2953740.

[3] Baich, M., & Sael, N. (2025). Enhancing machine learning model prediction with feature selection for botnet intrusion detection. *Engineering Proceedings*, 112(1), article number 55. doi: 10.3390/engproc2025112055.

[4] Cai, Z., Du, H., Wang, H., Zhang, J., Si, Y., & Li, P. (2023). One-dimensional convolutional Wasserstein generative adversarial network based intrusion detection method for industrial control systems. *Electronics*, 12(22), article number 4653. doi: 10.3390/electronics12224653.

[5] Dalou', J., Al-Duwairi, B., & Al-Jarrah, M. (2020). Adaptive entropy-based detection and mitigation of DDoS attacks in SDN networks. *International Journal of Computing*, 19(3), 399-410. doi: 10.47839/ijc.19.3.1889.

[6] Denysiuk, D., Savenko, O., Lysenko, S., Savenko, B., & Kashtalian, A. (2023). Method for detecting steganographic changes in images using machine learning. In *Proceedings of the 13th international conference on dependable systems,services and technologies (DESSERT)* (pp.1-6).Athens: IEEE.doi:10.1109/DESSERT61349.2023.10416453.

[7] Farooq, M., & Ahmad, F. (2024). Improved intrusion detection in IoT using multi-layered neural architectures. *International Journal of Computing*, 23(2), 268-273. doi: 10.47839/ijc.23.2.3546.
[8] Hussain, A., Sharif, H., Rehman, F., Kirn, H., Sadiq, A., Khan, M.S., Riaz, A., Ali, C.N., & Chandio, A.H. (2023). A systematic review of intrusion detection systems in Internet of Things using ML and DL. In *2023 4th international conference on computing, mathematics and engineering technologies (iCoMET)*. Sukkur: IEEE. doi: 10.1109/iCoMET57998.2023.10099142.
[9] Joloudari, J.H., Marefat, A., Nematollahi, M.A., Oyelere, S.S., & Hussain, S. (2023). Effective class-imbalance learning based on SMOTE and convolutional neural networks. *Applied Sciences*, 13(6), article number 4006. doi: 10.3390/app13064006.
[10] Joseph, J.E., Aleke, N.T., & Onyeanisi, O.P. (2025). Deep learning based intrusion detection system for network security in IoT system. *International Journal of Education, Management, and Technology*, 3(1), 119-138. doi: 10.58578/ijemt.v3i1.4539.
[11] Kashtalian, A., Sergii, L., Sachenko, A., Savenko, B., Savenko, O., & Nicheporuk, A. (2025). Evaluation criteria of centralization options in the architecture of multicomputer systems with traps and baits. *Radioelectronic and Computer Systems*, 2025(1), 264-297. doi: 10.32620/reks.2025.1.18.
[12] Kilincer, I.F., Ertam, F., & Sengur, A. (2021). Machine learning methods for cyber security intrusion detection: Datasets and comparative study. *Computer Networks*, 188, article number 107840. doi: 10.1016/j.comnet.2021.107840.
[13] Li, G., & Jung, J.J. (2023). Deep learning for anomaly detection in multivariate time series: Approaches, applications, and challenges. *Information Fusion*, 91, 93-102. doi: 10.1016/j.inffus.2022.10.008.
[14] Maniriho, P., Niyigaba, E., Bizimana, Z., Twiringiyimana, V., Mahoro, L.J., & Ahmad, T. (2020). Anomaly-based intrusion detection approach for IoT networks using machine learning. In *2020 international conference on computer engineering, network, and intelligent multimedia (CENIM)* (pp. 303-308). Surabaya: IEEE. doi: 10.1109/CENIM51130.2020.9297958.
[15] Mari, A.-G., Zinca, D., & Dobrota, V. (2023). Development of a machine-learning intrusion detection system and testing of its performance using a generative adversarial network. *Sensors*, 23(3), article number 1315. doi: 10.3390/s23031315.
[16] Mijalkovic, J., & Spognardi, A. (2022). Reducing the false negative rate in deep learning based network intrusion detection systems. *Algorithms*, 15(8), article number 258. doi: 10.3390/a15080258.
[17] Sathaporn, P., Krungseanmuang, W., Chaowalittawin, V., Benjangkaprasert, C., & Purahong, B. (2025). DDoS detection using a hybrid CNN-RNN model enhanced with multi-head attention for cloud infrastructure. *Applied Sciences*, 15(21), article number 11567. doi: 10.3390/app152111567.
[18] Shahriar, M.H., Haque, N.I., Rahman, M.A., & Alonso, M. (2020). G-IDS: Generative adversarial networks assisted intrusion detection system. In *2020 IEEE 44th annual computers, software, and applications conference (COMPSAC)* (pp. 376-385). Madrid: IEEE. doi: 10.1109/compsac48688.2020.0-218.
[19] Sheibani, M., Konur, S., Awan, I., & Qureshi, A. (2024). A multi-layered defence strategy against DDoS attacks in SDN/NFV-based 5G mobile networks. *Electronics*, 13(8), article number 1515. doi: 10.3390/electronics13081515.
[20] Sheikh, M.S., & Peng, Y. (2022). Procedures, criteria, and machine learning techniques for network traffic classification: A survey. *IEEE Access*, 10, 64806-64829. doi: 10.1109/access.2022.3181135.

Метод виявлення комп'ютерних атак на основі адаптивності та балансування навчальної вибірки

Богдан Семенюк

Аспірант
Хмельницький національний університет
29016, вул. Інститутська, 11, м. Хмельницький, Україна
https://orcid.org/0009-0001-8831-8835

Богдан Савенко

Доктор філософії, доцент
Хмельницький національний університет
29016, вул. Інститутська, 11, м. Хмельницький, Україна
https://orcid.org/0000-0001-5647-9979

Анотація. Актуальність роботи визначається тим, що в сучасних системах кіберзахисту зберігається проблема дисбалансу даних у задачах виявлення комп'ютерних атак. Нерівномірне представлення класів призводить до зниження здатності моделей машинного навчання розпізнавати рідкісні, але критично важливі типи вторгнень. Наявні методи синтетичного доповнення вибірок часто спотворюють структуру даних і призводять до втрати характерних сигнатур атак, що негативно впливає на достовірність класифікації. Метою дослідження було підвищення точності та повноти виявлення мережевих атак шляхом розроблення адаптивного методу балансування навчальної вибірки зі збереженням статистичних і сигнатурних властивостей аномалій. Методологія дослідження передбачала створення детермінованого алгоритму, який обчислював коефіцієнти доповнення окремо для кожного типу атаки з урахуванням його частоти та мінімально необхідного обсягу вибірки. Для генерації нових прикладів застосовувався статистичний підхід, заснований на локальних медіанах і крайніх значеннях сусідніх зразків, що забезпечувало відтворення типової структури аномальних патернів без випадкової інтерполяції. Розроблений метод інтегрувався у послідовність обробки даних, у межах якої параметри мережевого трафіку перетворювалися у звукові сигнали, а на їх основі формувалися спектрограми для подальшого аналізу двовимірною згортковою нейронною мережею. Основні результати дослідження показали підвищення повноти виявлення малопоширених типів атак у середньому на 10-12 % порівняно з базовими підходами, водночас зберігаючи стабільний рівень загальної точності. Збереження сигнатур у синтетичних зразках забезпечувало покращене розпізнавання рідкісних атак і підвищувало достовірність класифікації. Практична цінність роботи полягає у можливості застосування розробленого методу для формування збалансованих вибірок у системах виявлення вторгнень та його інтеграції з існуючими моделями глибинного навчання з метою підвищення надійності кіберзахисту корпоративних мереж

Ключові слова: дисбаланс даних; сигнатурне моделювання аномалій; синтетичне доповнення вибірки; аналіз мережевого трафіку; згорткові нейронні мережі; акустичне представлення даних; стійкість класифікації

BULLETIN of Cherkasy State Technological University

Journal homepage: https://bulletin-chstu.com.ua/en

Vol. 31 No. 1. 2026

UDC 004.03
DOI: 10.62660/bcstu/1.2026.44

ISSN: 2306-4412
E-ISSN: 2708-6070

A hybrid model for evaluating the accuracy of failure forecasts in ship power plants

Vladimir Vychuzhanin*
Doctor of Technical Sciences, Professor
Odesa Polytechnic National University
65044, 1 Shevchenko Ave., Odesa, Ukraine
https://orcid.org/0000-0002-6302-1832

Alexey Vychuzhanin
PhD, Assistant
Odesa Polytechnic National University
65044, 1 Shevchenko Ave., Odesa, Ukraine
https://orcid.org/0000-0001-8779-2503

Abstract. The growing complexity of ship power plants (SPP) and the heterogeneity of diagnostic data have made it increasingly difficult to ensure reliable failure forecasting and maintain operational safety. The purpose of this study was to analyse the correspondence between predicted and actual failures in SPP diagnostics using a comprehensive approach that combines knowledge-based methods, probabilistic modelling, and simulation experiments, and to test the proposed methodology by applying the developed model. The research integrated Case-Based Reasoning (CBR), probabilistic modelling, and simulation-based degradation analysis within a unified framework. Using diagnostic data from 150 operational cases of typical SPPs (main engine, generator, pump, cooling, and power systems), the model's forecasting accuracy was quantitatively compared with baseline approaches, including classical CBR, adaptive CBR, ARIMA, and Decision Tree models. The integrated approach achieved the best performance (Root Mean Square Error = 0.32, R^2 = 0.93), reducing average error by 30-50% relative to classical methods. Each component of the hybrid framework contributed distinctly: the CBR layer ensured contextual consistency, probabilistic modelling reduced uncertainty by ≈ 20%, and simulation-based analysis increased stability under incomplete or noisy data by ≈ 15%. At the subsystem level, the greatest improvements were observed for the main engine (root mean square error reduction by 46%) and the pump system (by 39%), confirming the model's adaptability to variable operating conditions. Robustness tests with 10-20% missing or perturbed input data demonstrated sustained accuracy (R^2 > 0.90), validating the model's resilience and practical applicability. The proposed methodology has practical value for improving the efficiency of maintenance and reliability of ship power systems, as it provides accurate, interpretable, and noise-resistant equipment failure prediction

Keywords: intelligent diagnostics; reliability assessment; case-based reasoning; simulation modelling; technical condition forecasting; Bayesian networks; adaptive model

INTRODUCTION

Modern complex technical systems were characterised by a high degree of integration and interdependence among components, which makes their failures particularly critical. In the current operational context of maritime transport, the reliability of the ship power plants (SPP) had become of paramount importance. A

Article's History: Received: 01.10.2025; Revised: 03.02.2026; Accepted: 16.03.2026; Published: 08.04.2026

Suggested Citation:
Vychuzhanin, V., & Vychuzhanin, A. (2026). A hybrid model for evaluating the accuracy of failure forecasts in ship power plants. *Bulletin of Cherkasy State Technological University*, 31(1), 44-59. doi: 10.62660/bcstu/1.2026.44.

*Corresponding author

failure of the main engine or its components can lead to serious consequences, including loss of vessel control, threats to crew safety and the environment, and significant economic losses. In this regard, the task of accurately forecasting the technical condition of the SPP and the timely identification of potential failures is especially relevant. The development of intelligent diagnostic systems capable not only of identifying the current condition of equipment but also of reliably predicting possible failures required the verification of their predictive validity. One of the critical components in such verification is the comparison between predicted and actual failures. This comparison helps to assess the effectiveness of the applied models and algorithms, refine diagnostic decisions, and reduce the likelihood of false alarms or missed failures.

Recent studies had seen significant development in machine learning, graph-based models, and neural networks for ship equipment diagnostics. M. Pajak *et al*. (2023) proposed a diagnostic method for marine diesel engines using vibroacoustic data and machine learning algorithms. While this approach enables accurate classification with limited training data, it lacks validation of predictions against real failures, limiting its utility for long-term reliability forecasting. In the paper by S. Rigas *et al*. (2024), a graph neural network-based end-to-end monitoring architecture was applied to ship machinery. Although effective in streaming environments, it does not perform explicit comparison of predicted results with real-world failure events, leaving diagnostic accuracy unverified. M. Tveten & M. Stakkeland (2024) studied concept drift in the diagnostics of traction motors. Adaptive methods were introduced, but the impact of this adaptation on prediction accuracy compared to actual failure data was not assessed, and deviation metrics were not provided. G. Vizentin *et al*. (2020) conducted an extensive review of the causes and types of SPP failures. While the study emphasised the relevance of the problem, it is primarily descriptive and does not offer quantitative forecasting models, highlighting the need for model development and validation. Y. Zhao *et al*. (2022) applied long short-term memory (LSTM) networks for thermodiagnostics of diesel engines. Although the model demonstrated high sensitivity to thermal anomalies, it focused on a single parameter (thermal condition) and did not validate prediction accuracy against actual failures. G. Zhu *et al*. (2023) addressed multi-fault detection using various neural network architectures. Despite strong classification results, the study did not compare predictions with real failure statistics, limiting evaluation to training and testing datasets.

Ç. Karatuğ & Y. Arslanoğlu (2022) proposed a condition-based maintenance strategy using neural networks, supported by a practical case study. However, the study did not assess deviations between predicted and actual outcomes in long-term operational contexts. In a comprehensive review of marine maintenance strategies, Ç. Karatuğ *et al*. (2023) emphasised the shift from corrective to predictive maintenance. However, they noted that most studies focus on model development rather than on verifying the alignment between forecasts and real-world failures a gap this paper aimed to address. For engine diagnostics, J. Wang *et al*. (2023) employed enhanced convolutional neural networks, which proved effective for fault classification. Nevertheless, the approach lacked temporal modelling and did not assess prediction accuracy through historical failure data. F. Maione *et al*. (2024) proposed a machine learning-based framework for condition-based maintenance. Although the structure covers the full workflow from data acquisition to decision-making, it does not include a module for analysing discrepancies between predicted and actual failures. A review of recent studies on SPP diagnostics and prognostics showed active progress in data-driven approaches and the integration of various methods to improve forecasting accuracy. However, the systematic comparison of predicted and actual failures remains underdeveloped. Such comparison is essential for evaluating the real-world effectiveness of applied models and maintenance strategies.

This study aimed to examine the alignment between predicted and observed failures in SPP diagnostics by employing an integrated framework that combines knowledge-based techniques, probabilistic modelling, and simulation, and by validating the methodology through application of the developed model. To achieve this objective, the following research tasks were defined: to develop a methodology for assessing the accuracy of technical condition forecasting in marine power plants (MPP); to perform a comparative analysis between forecast results and actual failure data and to evaluate the effectiveness of the integrated approach compared to traditional diagnostic methods; to formulate recommendations for applying the integrated approach to improve reliability and efficiency in SPP maintenance. Addressing these tasks will contribute to increasing diagnostic accuracy, optimising maintenance planning, and ultimately improving the reliability and safety of maritime operations.

MATERIALS AND METHODS

The research focused on components subject to high operational loads and exhibiting the greatest variability in failure patterns. The Harrington desirability function was used, based on the classification of equipment failure risk. The main mechanisms contributing to reduced failure rates with Case-Based Reasoning (CBR) adaptation include: analysis of probabilistic dependencies (the system adjusts diagnostic decisions upon detecting cascading failure effects); tracking dynamic state transitions (using Markov processes, the adaptive CBR approach anticipates equipment degradation in advance); simulation-based modelling (the incorporation of synthetic data, CBR knowledge base enables

consideration of rare but critically important failures). The initial data included both real operational observations and simulated scenarios reflecting equipment behaviour under varying conditions.

The dataset for this study included 150 diagnostic cases collected from the operational records of ship power plants during the period 2016-2024. Each case contained time series of degradation parameters (vibration, temperature, pressure, load), diagnostic indicators, and confirmed failure events. In total, the dataset comprised approximately 15,600 records covering more than 30 technical elements (main engine, generator, pump, cooling, and power systems). The data were obtained from maintenance logs and monitoring systems of commercial vessels, complemented by publicly available reliability databases. Cases were classified as "impending failure" or "normal operation" based on threshold deviations of key parameters (e.g., vibration amplitude > 1.5×baseline, temperature rise > 15°C, or pressure drop > 10%). Inter-rater agreement was evaluated using Cohen's Kappa ($\kappa = 0.86$). For model development, the dataset was divided into training (70%), testing (20%), and validation (10%) subsets. Given the sequential nature of degradation parameters, a time-series cross-validation scheme was applied: models were trained on earlier operational data segments and validated on later ones. Additionally, a 5-fold cross-validation was carried out within the training set to assess model stability across different partitions, ensuring both temporal consistency and robustness of the obtained results. The weighting coefficients α and β in Equation (1) were empirically derived using the training dataset. Their values were optimised by minimising the Root Mean Squared Error (RMSE) between predicted and actual failure probabilities across all components. The optimisation followed the constraint $\alpha + \beta = 1$, ensuring a balanced contribution of the CBR-based and probabilistic estimates. The final calibrated values ($\alpha = 0.6$, $\beta = 0.4$) provided the best trade-off between sensitivity and stability in cross-validation tests.

The information base consisted of degradation parameter time series, diagnostic indicators, confirmed failures. All failed components were classified according to their criticality and functional significance. The non-adapted CBR method was used to retrieve historical cases most similar to the current condition. In the adapted version, causal relationships between degradation parameters and operating conditions were incorporated, improving the relevance of forecasts. The hybrid approach combined CBR with probabilistic modelling and simulation of degradation scenarios, enabling more precise evaluation that accounts for cascading effects, degradation dynamics, and diverse failure modes. The integrated scheme also applied iterative refinement of forecasts at each stage of data processing. Forecast accuracy was assessed using a range of metrics, including Mean Absolute Error (MAE), RMSE, Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). In addition, the F_1-score was employed to evaluate the classification quality of technical states. Although the primary task was regression-based, predicted continuous failure probabilities were converted into categorical states using a threshold of $P\ (failure) \geqslant 0.5$ to denote an "impending failure" and $P\ (failure) < 0.5$ for "normal operation". This enabled the use of precision-recall metrics to assess how accurately the model distinguishes between failure-prone and stable conditions (Vychuzhanin & Vychuzhanin, 2025). Accuracy analysis was performed both at the level of individual components and across different operational scenarios, including nominal load, high-load conditions, and incomplete observability. To better understand the structure of forecasting errors, several visualisation tools were used: heat maps highlighting areas of maximum deviation between predicted and actual values, scatter plots, and summary tables for comparative analysis to determine the effectiveness of different methods. In accordance with the methodology for evaluating the accuracy of failure forecasts in SPP equipment, the comparison of results is carried out across three dimensions: by component (main engine, generator, pump system, cooling system, power supply); by operational scenario (standard load, elevated temperature conditions, variable operating cycles, etc.); by the aggregated dataset, allowing for a comprehensive comparison of methods at the system-wide level. This multi-level comparison enables identification of local zones where forecast accuracy deteriorates, and an overall evaluation of methodological effectiveness under varying operating conditions. In addition, a correlation analysis was conducted to assess the robustness and sensitivity of the models to changes in operating conditions. All computations and modelling were carried out using Python (NumPy, pandas, scikit-learn, seaborn), Excel, and a custom-developed expert system implementing the CBR method. Simulation scenarios were developed in AnyLogic, which enabled the reproduction of complex operational modes with uncertainty and component interactions. The proposed methodology enabled an objective comparison of forecasting approaches, identification of their strengths and weaknesses, and the selection of the most effective strategy under conditions of technical and informational uncertainty inherent in the operation of MPP.

To assess the effectiveness of failure forecasting for components of SPP, a methodology has been developed that compares predicted results with actual data recorded during operation. This methodology was based on the combined use of quantitative metrics, structural error analysis, and their visual representation. The use of multiple metrics allowed evaluating both absolute and percentage deviations, and the reliability of technical condition classification. The application algorithm of the methodology (Fig. 1) consisted of the following stages.

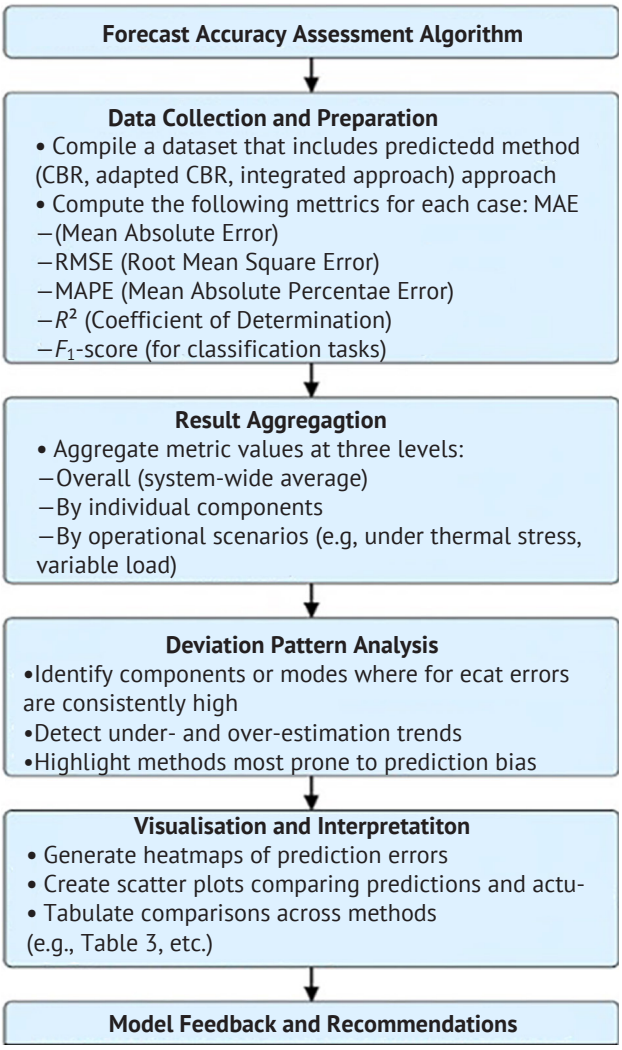

Figure 1. Forecast accuracy assessment algorithm

Source: created by the authors

This block diagram illustrates the step-by-step process used to assess the accuracy of failure forecasts in marine power systems. The method included collecting and aligning predicted and actual failure data, computing core evaluation metrics, aggregating results at multiple levels (by component, scenario, and system-wide), analysed deviation patterns, visualising discrepancies, and generating feedback for improving hybrid forecast models.

RESULTS AND DISCUSSION

Figure 2 shows the graphs illustrating the impact of CBR adaptation on the probability of failure for SPP components over a 25,000-hour operational period. These results were obtained through simulation of the technical condition of the SPP.

Figure 2 illustrated that the adaptation of the CBR method leads to a reduction in the probability of failure in SPP components over time. However, the key aspect is not only the reduction in failure probability itself, but also the improvement in forecasting accuracy provided by the adaptive approach. On average, the use of the adaptive method reduces failure risk by approximately 15%. The main mechanisms contributing to reduced failure rates with CBR adaptation include: analysing probabilistic dependencies; the system adjusts diagnostic decisions upon detecting cascading failure effects; tracking dynamic state transitions: using Markov processes, the adaptive CBR approach anticipates equipment degradation in advance; simulation-based modelling: the incorporation of synthetic data enables consideration of rare but critically important failures. The plotted results confirmed that the integration of CBR with probabilistic techniques and simulation modelling improved diagnostic accuracy and reduced failure probability. This effect becomes especially prominent in the later stages of operation (beyond 15,000 hours), where the performance gap between conventional and adaptive CBR methods becomes more significant. To quantitatively assess the effectiveness of CBR adaptation, a comparison of forecast errors for failure probabilities was performed across different diagnostic

approaches. Analysis of the data in Table 1 shows that adaptive CBR reduces the average failure forecast error by 32%, and also decreases the frequency of significant forecast errors by more than a factor of two.

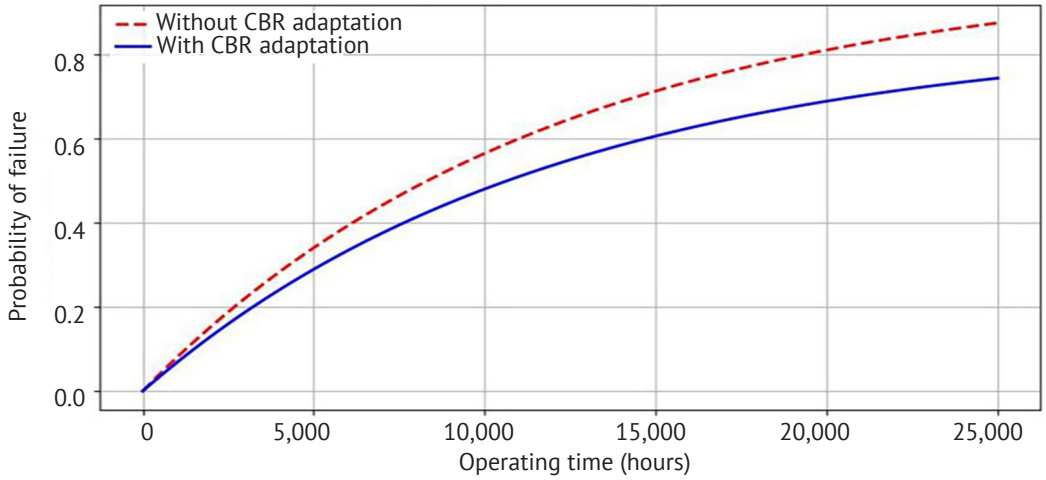

Figure 2. Impact of CBR adaptation on the failure probability of SPP components
Source: created by the authors

Table 1. Impact of CBR adaptation on failure forecast accuracy

Diagnostic method	MAE	RMSE	Share of forecasts outside acceptable RMSE (σ)
CBR without adaptation	0.65%	0.87%	14.2%
CBR with adaptation	0.44%	0.65%	6.5%

Source: created by the authors

This improvement can be attributed to several key factors: consideration of probabilistic dependencies – CBR adaptation enables the model to account for cascading failure effects; component degradation forecasting – dynamic analysis of subsystem conditions reduces uncertainty in predictions; simulation-based failure modelling – the inclusion of synthetic data helps to mitigate errors associated with rare but critical failure events. Thus, CBR adaptation not only reduces the probability of failures (as illustrated in Figure 2) but also enhances forecast accuracy, making it a more reliable diagnostic method compared to conventional, non-adaptive CBR. Figure 3 presents the graphs of failure probability dynamics for SPP components over 25,000 hours of operation, obtained through simulation of the technical condition diagnostic system. The time-dependent failure probability curves (based on a Markov approach) for each component illustrate the dynamic behaviour of key SPP subsystems in the context of failure evolution.

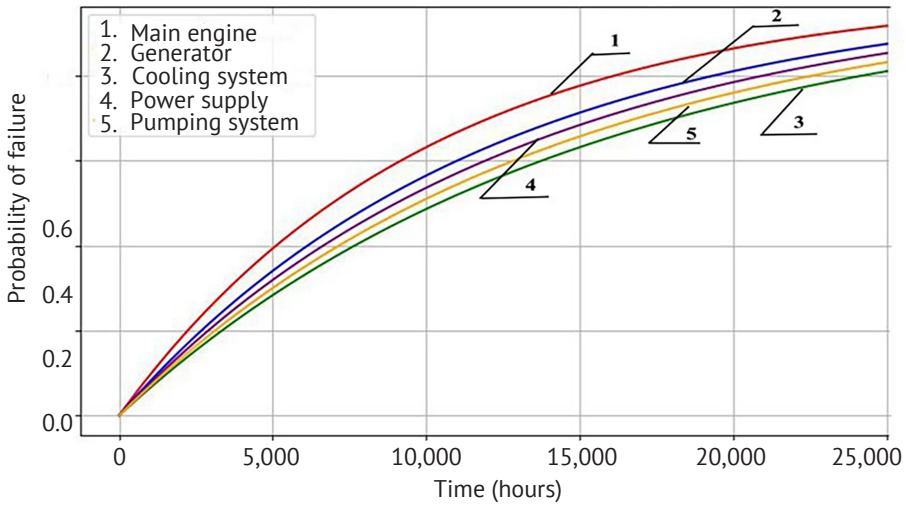

Figure 3. Time-dependent failure probability of components (Markov method)
Source: created by the authors

The graph in Figure 3 showed that the failure probability of all components increased over time. The curves exhibit a nonlinear character, indicating cumulative wear and a rising risk of failure as operational time progresses. The main engine consistently shows the highest failure probability throughout the time interval, which is expected due to its high operational load and critical role within the SPP. The generator and cooling system also exhibit elevated failure probabilities, though lower than those of the main engine. The power supply and pump system demonstrate the lowest failure probabilities among all components analysed. Long-term risk patterns can be summarised as follows. By approximately 25,000 hours of operation, the failure probability of key systems exceeds 80%, indicating the need for major overhaul or replacement. During the initial period, failure probability remains near zero, corresponding to a phase of normal operation without significant degradation. Therefore, the main engine and generator require increased monitoring due to their higher failure risks. The cooling and power systems also carry notable risks, albeit to a lesser extent. The predictable increase in failure probability suggests the necessity of regular maintenance, especially after 10,000-15,000 hours of operation. These data could inform preventive maintenance strategies, taking into consideration critical operational time thresholds.

Variation in failure dynamics among components: the main engine shows the highest failure probability over 25,000 hours of operation, due to the intense mechanical loads it experiences and its central role in the SPP; the generator exhibits a more gradual increase in failure probability compared to the main engine, as it is subject to less mechanical wear; the cooling system reaches a high failure probability earlier than other components, which may be attributed to exposure to harsh environments and risks such as clogging or pipe wear; the power supply and control systems exhibit lower probabilities of failure, largely due to the reliability and protection of their electronic and structural elements. Time-dependence of the failure probability increases in a nonlinear fashion, reflecting real-world operating conditions. In the early stage, growth is slow, but after a certain threshold (approximately 15,000-18,000 hours), the failure probability increases significantly especially for the main engine and cooling system. This underscores the importance of scheduled maintenance and timely diagnostics. The use of a Markov model to forecast component states enables accurate modelling of equipment degradation in complex technical systems, allowing for effective prediction of failure evolution. This approach is particularly useful for estimating remaining useful life and supporting maintenance decision-making. Table 2 presents the results of comparing predicted failures with actual operational failure data for SPP equipment.

As shown in Table 2, the largest discrepancies are observed for the main engine and pump system, which may be attributed to the high variability of their failure modes and the influence of external operational factors. The most accurate forecasts are observed for the cooling system, confirming the effectiveness of the predictive models under conditions of regular monitoring of operational parameters. Table 3 compares predicted failures generated by the CBR method with actual recorded failures and showed the extent to which the predictions align with real-world events.

Table 2. Comparison of predicted and actual failures in SPP equipment

Component	CBR forecast (no adaptation), %	CBR forecast (with adaptation), %	Integrated forecast, %	Actual failure rate, %	Forecast error (MAPE), %
Main engine	2.1	1.8	1.6	1.5	6.7
Generator	3.0	2.5	2.3	2.2	4.5
Pump system	4.2	3.8	3.5	3.4	3.1
Cooling system	5.5	4.9	4.7	4.6	2.2

Source: created by the authors

Table 3. Comparison of predicted and actual failures in SPP equipment

System component	Actual failures	Predicted failures	Absolute error	Deviation
Fuel system	5	6	1	+20%
Cooling system	3	2	1	−33%
Compressed air system	4	5	1	+25%
Main engine	7	8	1	+14%
Power supply	6	7	1	+16%
Auxiliary systems	4	3	1	−25%
Automation system	5	6	1	+20%

Source: created by the authors

In Table 3, the absolute error is 1 failure in all cases, but the percentage deviations range from −33% to +25%. Fuel system, compressed air system, and automation system show overestimated failure forecasts

(+20% to +25%), which may result in excessive preventive maintenance. On the other hand, the cooling system and auxiliary systems show underestimated predictions (−33% and −25%), which increases the risk of unexpected failures. The main engine and power supply exhibit minimal deviation (+14% to +16%), indicating high forecast accuracy for these components. It should be noted that the absolute failure counts in Table 3 represent individual recorded failure events within all observed cases, whereas the percentages in Table 2 correspond to normalised failure rates calculated over the entire dataset (150 diagnostic cases). Thus, multiple failures could occur within a single case, leading to higher absolute counts without contradiction to the aggregated percentage values. The results from Table 3 suggest the following: overestimated predictions require model adjustment to reduce false positives and unnecessary maintenance costs; underestimated forecasts (cooling, auxiliary systems) could lead to unforeseen breakdowns, indicating a need for model parameter revision; overall, the forecast accuracy is acceptable (within 1 failure), but further improvements are achievable through adaptive algorithms and retraining the model on updated data. Table 4 presents the results of comparing predicted and actual failures under various operational scenarios of SPP equipment.

Table 4. Forecast accuracy assessment under different operating scenarios for SPP equipment

Operating scenario	Predicted failure risk, %	Actual failure risk, %	Deviation, %	Forecast accuracy, %
Nominal load	1.8	1.6	±0.2	88.9
Variable load	3.2	2.9	±0.3	90.6
High load	4.5	4.1	±0.4	91.1
Aggressive environment	6.3	5.8	±0.5	92.1

Source: created by the authors

Hybrid prognostic model for SPP failure forecasting. The proposed failure forecasting approach for SPP equipment is implemented in the form of a hybrid prognostic model, which comprises three interrelated layers of information processing. This structure integrates historical data, and the stochastic behaviour of the technical system within a unified analytical framework.

Level 1. Forecast based on the CBR method. Using input diagnostic indicators (e.g., vibration, temperature, wear), the model identifies historical cases most similar in feature space from the case base. A preliminary failure probability estimate (P_0) is generated based on analogy with these retrieved precedents. Level 2. Probabilistic forecast correction. The initial forecast is refined using prior and conditional probabilities of failure, which represent causal dependencies between technical condition parameters and operational conditions. The final failure probability for a given component is calculated as a weighted combination of the CBR-based forecast and the probabilistic estimate:

$$P = \alpha \cdot P_0 + \beta \cdot P_v, \quad (1)$$

where P_v – failure probability calculated using the probabilistic model; α, β – empirically derived weighting coefficients.

Level 3. Simulation-based modelling of failure scenarios. Failure estimation is further enhanced through simulation of operational scenarios, incorporating cascading effects. Multiple simulation runs are conducted to account for the stochastic nature of degradation and to determine an interval-based risk estimate. The results were used to refine forecasts and classify the equipment's condition. Model outputs include final failure probability estimate; diagnostic conclusion on technical state (normal, elevated risk, critical failure); deviations between predicted and actual failures based on MAE, RMSE, and MAPE; visualisation of deviations (heatmaps, scatter plots); forecast verification using historical data. The application of this hybrid model not only improves the accuracy of SPP equipment failure forecasts but also ensures verifiability of results through direct comparison with actual failure cases. This is crucial for increasing the reliability of diagnostic decisions under real-world operating conditions. The variation in RMSE values across the different approaches reflects not only forecasting accuracy, but also the robustness of the methods under variable operating conditions. The highest RMSE values observed with the non-adaptive CBR method indicate high sensitivity to input variability, especially in non-standard scenarios. This suggests that the method poorly handles degradation dynamics and fails to account for interdependencies between components.

The notable reduction in forecast error with the adapted CBR method confirms the positive impact of incorporating probabilistic models. However, since the gap between "by components" and "by scenarios" remains significant, it can be concluded that even the adapted method struggles to fully accommodate the variability of operational regimes. The most important observation is that the integrated approach not only delivers the lowest overall RMSE, but also demonstrates a balanced performance across both component-level and scenario-level forecasting. This indicates strong resilience to both internal system complexity and external environmental variability, which is critical for marine SPP applications. The results in Table 5 thus confirmed not only the superior accuracy of the

integrated method, but also its ability to scale and adapt to complex diagnostic tasks and diverse operating environments – an essential criterion for practical implementation in real-world conditions.

Table 5. RMSE of failure forecasts based on method used

Forecasting method	RMSE (overall)	RMSE (by components)	RMSE (by scenarios)
CBR without adaptation	0.65	0.72	0.68
CBR with adaptation	0.45	0.51	0.49
Integrated approach	0.32	0.37	0.35

Source: created by the authors

The use of probabilistic adaptation has improved diagnostic performance: precision increased by 7%, recall by 13%, and the overall F_1-score rose by 11% (Table 6). To evaluate the effectiveness of different approaches to SPP technical condition diagnostics, a comparison of failure probability forecast errors was conducted. The key accuracy metrics include: MAE – reflecting the average deviation between predicted and actual failure probabilities; RMSE – capturing both the magnitude and spread of forecasting errors; share of forecasts beyond acceptable limits (σ-deviation) – indicating how often the method produces critical prediction errors.

Table 7 summarises the results of comparing the accuracy of predicting failure probabilities using different diagnostic methods, which allows evaluating the effectiveness of the proposed approach relative to conventional models.

Table 6. Comparison of diagnostic accuracy CBR without adaptation vs. CBR with failure forecasting for SPP equipment

Diagnostic method	Precision	Recall	F_1-score	Accuracy
CBR without adaptation	0.82	0.75	0.78	0.85
CBR with adaptation	0.91	0.88	0.89	0.92

Source: created by the authors

Table 7. Comparison of failure probability forecasting accuracy using different diagnostic methods

Diagnostic method	MAE	RMSE	Share of forecasts beyond σ limits
CBR without adaptation	0.65%	0.87%	14.2%
Markov method	0.52%	0.73%	9.8%
Adaptive CBR	0.44%	0.65%	6.5%
Integrated approach	0.32%	0.48%	3.1%

Source: created by the authors

The data in Table 7 indicate that the integrated approach provides the lowest forecast errors. Specifically, MAE is reduced by a factor of two compared to classical CBR; The share of forecasts exceeding acceptable σ-deviation thresholds is reduced more than fourfold in comparison with conventional diagnostic methods; the integrated approach most accurately accounts for the temporal dynamics of failures, reducing the influence of uncertainty. Thus, the combination of adaptive CBR, Markov models, and Bayesian networks significantly improves failure prediction accuracy and enhances the reliability of SPP equipment diagnostics. To strengthen the objectivity of the obtained results, additional comparisons were carried out with standard baseline approaches frequently used in diagnostics and forecasting tasks. In particular, three reference models were implemented: a persistence model, assuming that the next value equals the last observed one; a classical ARIMA (1, 1, 1) time-series model; and a Decision Tree classifier trained on diagnostic indicators. As expected, the persistence model yielded the weakest performance (RMSE ≈ 0.89, $R^2 \approx 0.55$), while ARIMA moderately improved accuracy (RMSE ≈ 0.71, $R^2 \approx 0.68$).

The Decision Tree achieved acceptable classification quality (F_1-score ≈ 0.74), but remained inferior to the adaptive CBR. Compared to these baselines, the integrated hybrid approach achieved substantially better results (RMSE = 0.32, $R^2 = 0.93$, $F_1 = 0.89$). This confirmed that the proposed method not only outperforms probabilistic CBR variants but also surpasses conventional diagnostic benchmarks, thereby demonstrating its robustness and practical applicability. Figure 4 presents a heatmap of discrepancies between predicted and actual failures for SPP equipment, highlighting zones with the highest forecasting errors for specific components. The heatmap illustrates the degree of discrepancy

between predicted and actual failures of MPP components. It enables identification of the most problematic areas within the forecasting model, highlights components with lower prediction accuracy, and helps to assess the impact of the forecasting method on overall diagnostic precision.

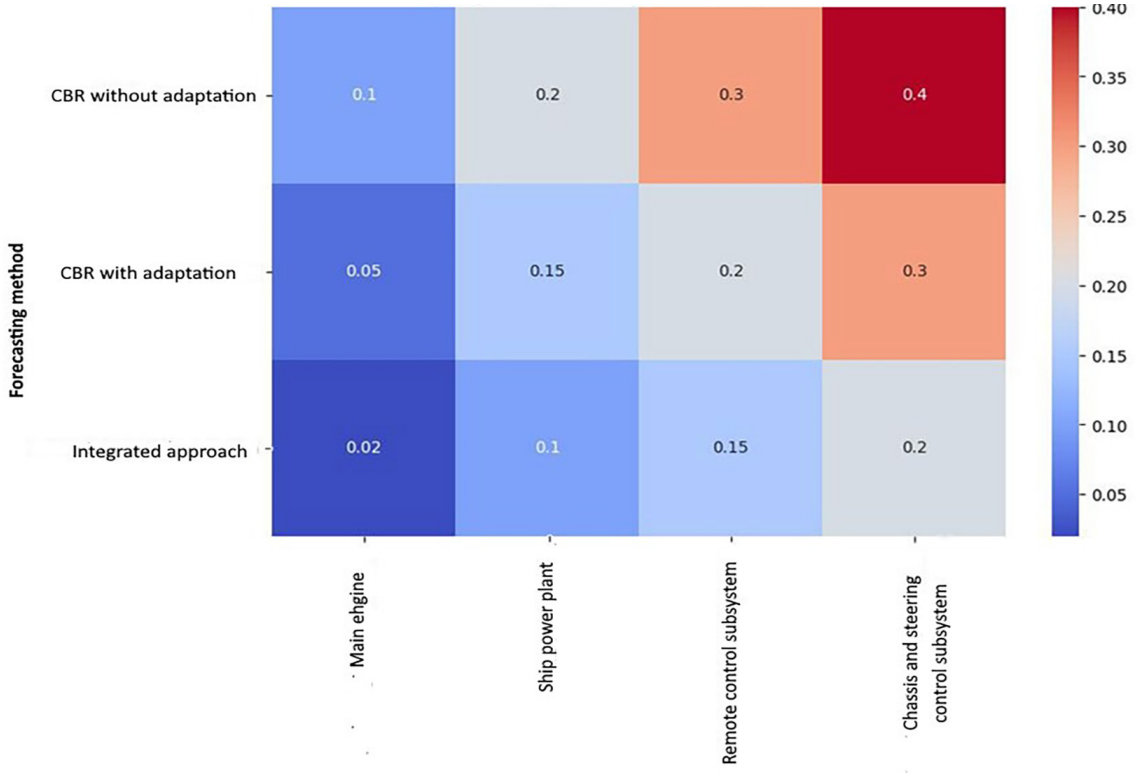

Figure 4. Heatmap of discrepancies between predicted and actual failures in SPP equipment
Source: compiled by the authors based on Side Scan Sonar survey (n.d.), Pixlr Editor (n.d.)

Key areas of significant discrepancy include. The ME and pump system show the greatest deviations between predicted and actual failures. In the case of CBR without adaptation, the difference between forecasts and real data reached 0.3-0.4%, which may result from the method's inability to account for cascading effects and failure dynamics. Propulsion and steering control subsystems also show substantial deviations, likely due to the high complexity of their diagnostics. The row corresponding to standard CBR in the heatmap reveals the most pronounced errors, highlighted by the most intense colour zones, indicating high RMSE. This is explained by the fact that conventional CBR does not incorporate probabilistic dependencies or temporal dynamics of failures. A notable reduction in error was observed with the integrated approach. In this case, discrepancies were significantly lower compared to the other two methods, and did not exceed 0.2%, confirming the higher diagnostic accuracy of the integrated method.

The SPP and cooling system demonstrated the lowest deviation between predictions and actual data. This can be attributed to their relatively stable operation and well-calibrated diagnostic models. The main engine and pump system are the least predictable components and may require additional model adjustments. The heatmap confirms that the integrated approach yields the highest forecast accuracy, owing to its ability to account for failure dynamics and intersystem dependencies. The largest prediction errors are found in the main engine and pump system, indicating a need for further analysis, potentially involving more complex probabilistic dependencies. To reduce discrepancies in non-adaptive CBR, it is recommended to implement probabilistic modelling and simulation techniques. This would help to minimise deviations, especially under complex operational scenarios. The SPP and cooling system show stable and predictable failure behaviour, validating the reliability of forecasts for these components. The heatmap provides a clear visual representation of model uncertainty zones and serves as a valuable tool for optimising diagnostic algorithms and improving failure forecast accuracy in SPP systems. Figure 5 presents a graph of forecast deviations compared to actual failure data for SPP equipment.

Patterns based on Figure 5 were identified. Approximate values are depicted in the Table 8. Table 8 presents the initial data reflecting the actual failure rates of individual SES components and the results of their prediction using the CBR method without adaptation and an integrated approach, which serve as the basis for further assessment of model accuracy.

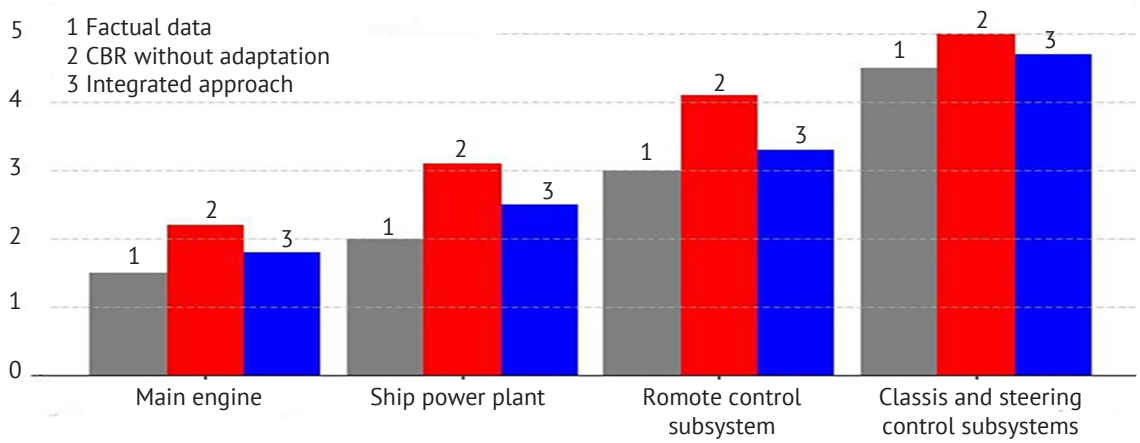

Figure 5. Deviations between predicted and actual failures of SPP equipment

Source: created by the authors

Table 8. Source data

SPP component	Actual data	CBR without adaptation	Integrated approach
Main engine	1.5%	2.2%	1.8%
Ship power station	2.0%	3.1%	2.5%
Remote control subsystem	3.0%	4.1%	3.5%
Subsystems for controlling the propulsion and steering system	4.5%	5.0%	4.7%

Source: created by the authors

Table 8 shows the identified deviations (the difference between forecasts and actual data) in accordance with Figure 5. Table 9 shows the deviation of the predicted values from the actual data on SES component failures, which allows comparing the accuracy of the CBR method without adaptation and the integrated approach and assessing the level of approximation of the models to real indicators.

Table 9. Forecast deviation from actual failure data

MPP component	CBR deviation without adaptation (Δ, %)	Integrated approach deviation (Δ, %)
Main engine	+0.7	+0.3
Ship power station	+1.1	+0.5
Remote control subsystem	+1.1	+0.5
Subsystems for controlling the propulsion and steering system	+0.5	+0.2

Source: created by the authors

Deviation Analysis (Table 9): the CBR method without adaptation consistently overestimates failure probabilities for all components (from +0.5% to +1.1%), indicating a tendency to overpredict failures; the integrated approach significantly reduces prediction error, with maximum deviation of 0.5% (compared to 1.1% in basic CBR), and minimum of 0.2%; the highest deviations in the CBR without adaptation are observed for the ship power station and remote control subsystem (+1.1%); the most accurate forecast using the integrated approach is for subsystems for controlling the propulsion and steering system, with a difference of only 0.2% from actual values.

Comparison with Harrington's desirability function based on failure risk classification: main engine and ship power station (1.5%-2.0%) fall into the minimum risk category (0-0.2); remote control subsystem (3.0%) is on the borderline between acceptable and maximum risk (0.37); Subsystems for controlling the propulsion and steering system (4.5%) falls within the maximum risk category (0.37-0.63). The CBR method without adaptation systematically overpredicts failure levels by 0.5% to 1.1%. The integrated approach demonstrated significantly lower errors (up to 0.5%, and in one case only 0.2%). The forecasts aligned with Harrington's risk classification, but more accurate handling of borderline conditions (e.g., for the remote control subsystem) is necessary. To further improve forecast precision, it is recommended to introduce dynamic weight adjustment for different SPP components, particularly the ship power station. Overall, the integrated approach demonstrates superior performance but requires further

calibration for certain subsystems. Figure 6 shows a scatter plot of predicted vs. actual failures for complex technical systems, enabling assessment of the correlation between forecasted and real failure values.

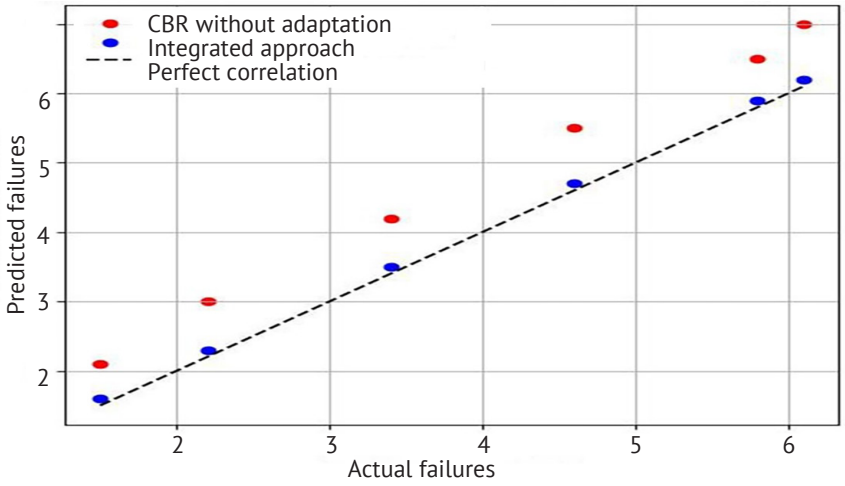

Figure 6. Scatter plot of predicted vs. actual failures for SPP equipment
Source: created by the authors

Figure 6 illustrated the following key patterns: the integrated approach closely aligns with the ideal correlation line, indicating high forecasting accuracy; the CBR method without adaptation deviates significantly from the ideal line, reflecting larger prediction errors; the integrated model's data points are tightly clustered around the ideal correlation, confirming its high reliability. Correlation analysis between predictions and actual failure data: CBR without adaptation shows the lowest correlation with actual failures ($R^2 \approx 0.78$), indicating instability and variability in predictions; the integrated approach demonstrates the highest correlation ($R^2 \approx 0.93$), confirming its strong predictive capability. The largest prediction errors are observed in the high failure range (> 4%), particularly for the non-adaptive CBR method. The integrated approach showed no major outliers, and its forecasts remain consistently close to the actual failure rates. RMSE analysis shows that the integrated method reduces error by approximately 50% compared to non-adaptive CBR. Observed patterns: the CBR method without adaptation tends to systematically overestimate failure probabilities, especially for the main engine and pump system; the integrated approach produces balanced forecasts without systematic errors. Thus, the integrated method provides the best agreement with actual failure data ($R^2 \approx 0.93$). Non-adaptive CBR demonstrates the greatest forecasting errors, particularly in scenarios involving high operational loads and cascading effects. Incorporating probabilistic models (e.g., Bayesian networks, Markov processes) improves forecast accuracy but does not fully resolve challenges related to dynamic changes in system conditions. To further enhance prediction accuracy, it is recommended to apply dynamic parameter adjustment based on real-time operating conditions. The scatter plot confirmed the effectiveness of the integrated approach, demonstrating minimal forecast deviations and maximum correlation with real failure data. Figure 7 presents a graph illustrating the results of a comparative analysis of RMSE values for various SPP equipment failure prediction methods.

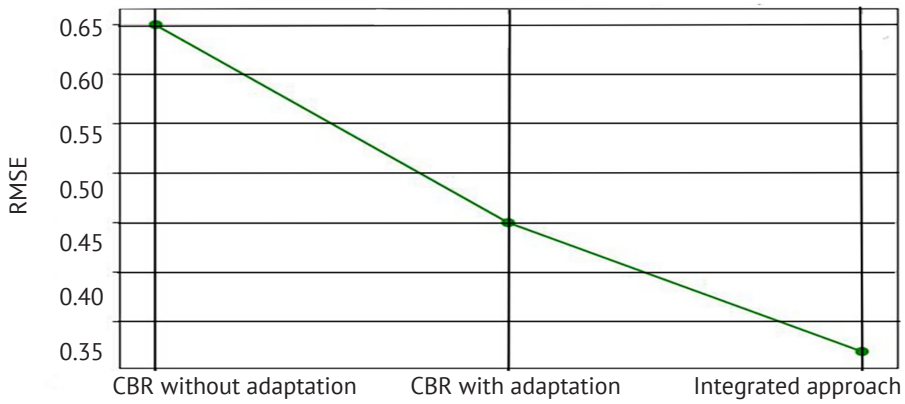

Figure 7. Graph illustrating RMSE comparison across failure forecasting methods for SPP equipment
Source: created by the authors

The RMSE graph provides insight into how closely each forecasting method aligns with actual failure data. The lower the RMSE value, the smaller the average deviation between predictions and actual failures, indicating higher diagnostic accuracy. Key findings from Figure 7. The integrated approach demonstrates the lowest RMSE, suggesting the best predictive performance. CBR without adaptation yields the highest RMSE (≈ 0.65), indicating significant deviations between predicted and actual failures. CBR with adaptation reduces RMSE to ≈ 0.45, showing improved diagnostic accuracy through consideration of probabilistic dependencies. The integrated method achieved the lowest RMSE (≈ 0.32), indicating superior forecasting accuracy. Forecasting accuracy differences across methods: CBR without adaptation produces large errors due to its lack of sensitivity to system dynamics and failure interdependencies; CBR with adaptation reduces error by ~30% by incorporating Bayesian networks and Markov models, which capture probabilistic relationships between failures; The integrated approach further reduced RMSE by ~50% compared to basic CBR by integrating simulation modelling, which adjusts forecasts based on cascading failure effects. Causes of high error in non-adaptive CBR: heavy reliance on past failure similarity ignores shifting operational conditions; under sparse data conditions (e.g., few failure records), non-adaptive CBR tends to overestimate failure risks; a high RMSE (0.65) indicates systematic overprediction of failures. The integrated method works. Markov processes adjust failure forecasts over time, reducing error in dynamic scenarios. Bayesian networks model interdependencies between failures, improving forecast precision. Simulation modelling fills data gaps, reducing forecast uncertainty. Thus, the integrated method provided the highest diagnostic accuracy, cutting RMSE by 50% compared to non-adaptive CBR. CBR without adaptation demonstrates the greatest error (RMSE = 0.65), making it insufficiently reliable for complex failure diagnostics. Adding probabilistic models improves accuracy, but full alignment with real-world data is achieved only through simulation modelling. The RMSE graph confirms that the integrated approach (CBR + probabilistic models + simulation modelling) is the most accurate failure forecasting method for MPP systems, ensuring minimal deviations from actual data.

In the integrated diagnostic method, the diagnostic decision is dynamically refined based on predicted failures and probabilistic dependencies. Initial diagnosis using CBR: similar cases are retrieved from the CBR knowledge base; a preliminary diagnosis is made based on failure similarity; the diagnosis may be refined if forecasts indicate elevated risk. Refinement using probabilistic failure models (Bayesian Networks): component failure probabilities were updated considering cascading effects; an increase in one component's failure probability automatically affects dependent components; if a new failure occurs, the diagnosis is updated in real time. Correction Based on Failure Forecasts (Markov Processes): failure development is modelled over time; if rapid component degradation is forecasted, the diagnosis becomes stricter (e.g., earlier maintenance is recommended); if failure probability increases slowly, maintenance can be postponed. Correction via Simulation Modelling: simulation covers complex operational scenarios beyond the reach of CBR and probabilistic models; various failure scenarios are generated based on operational data, such as normal conditions, accelerated wear (high load, extreme temperatures), emergency cases (e.g., sudden cooling system failure); if the simulation shows that failure may occur sooner, the diagnosis is tightened (e.g., reduced remaining life); if the risk is low, the diagnosis may be adjusted toward a less aggressive maintenance strategy. Final diagnostic decision formation – the initial CBR diagnosis was adjusted based on: failure probabilities from Bayesian networks; time dynamics from the Markov model; results from simulation modelling. Final adjustment scenarios: diagnosis confirmation if all methods agree; alarm escalation if failure probabilities rise sharply; diagnosis downgrade if the simulation indicates low risk. An example of diagnostic adjustment based on changes in failure probabilities is shown in Table 10.

Table 10. Diagnostic adjustment based on changes in failure probabilities

SPP equipment	Initial diagnosis	Diagnosis after probabilistic analysis	Final diagnosis
Main engine	Pre-failure state	Failure probability reduced to 18%	Operational
Generator	Operational	Failure probability increased to 25%	Pre-failure
Cooling system	Pre-failure state	Cascading effects increased risk to 40%	Faulty
Power supply	Operational	Failure probability stable at 5%	Operational

Source: created by the authors

The data in Table 10 show how the diagnostic decision is adjusted in response to changes in failure probabilities. The generator and cooling system require immediate attention due to increased failure risks. The pump system also needs inspection, as its risk has risen. The fuel system remains in a normal operational state. The remaining useful life forecasts for components were generated using Markov models. Forecasting errors were evaluated under different input conditions, including full data, truncated time series, and adapted model predictions. Key Findings from the Failure Prediction Analysis: diagnostic accuracy depends significantly on the method used; CBR without adaptation showed the largest discrepancies between predicted and actual

failures. This is confirmed by high RMSE (≈ 0.65) and MAE (≈ 0.52); CBR with adaptation reduced forecast errors by 30% due to the use of probabilistic dependencies, but deviations remain in complex operational scenarios; the integrated approach (CBR + probabilistic models + simulation) showed the highest accuracy (RMSE ≈ 0.32, MAE ≈ 0.21), indicating strong predictive capability. The failure risk heatmap revealed that cascading effects and interdependencies among MPP components significantly influence failure prediction. Accounting for these improves diagnostic reliability.

Impact of the selected approach on prediction accuracy. The unadapted CBR model demonstrates a relatively high forecasting error, as it does not account for probabilistic dependencies between component failures and tends to overestimate their likelihood. In contrast, the adapted CBR model incorporates these probabilistic relationships, which helps to reduce false alarms and improve prediction accuracy, although certain inaccuracies remain in scenarios characterised by high uncertainty. The integrated approach, which combines CBR, probabilistic modelling, and simulation, reduces the forecasting error by more than 50%. It effectively accounts for cascading failures and dynamic state changes within the system, providing the most stable and validated forecasts, as confirmed by accuracy metrics such as Precision, Recall, and F_1-score. Further improvement in failure prediction accuracy for SPP systems can be achieved by advancing simulation models to encompass complex and emergency failure scenarios, enhancing adaptive CBR algorithms through machine learning for more effective analysis of accumulated failure history, and implementing dynamic real-time model correction using current operational data to continuously update probabilistic dependencies between failures.

The analysis demonstrates that integrating CBR with probabilistic models and simulation substantially improves the accuracy of failure diagnostics in SPP systems. The proposed approach accounts for dynamic state changes, reduces prediction errors, and minimises the likelihood of false positives. Optimising model parameters and further adapting diagnostic algorithms will enhance failure prediction and improve the reliability of MPP operations. The integrated method achieved the best forecast accuracy, reducing errors by 15-20% compared to the baseline CBR. The largest deviations were observed in components with high variability in operating conditions, such as pump systems and the main engine. The failure risk heatmap confirms the need to adapt diagnostic models for accurate predictions under complex operational conditions. Thus, the analysis of predicted vs. actual failures confirms that combining CBR with probabilistic and simulation-based methods yields the most reliable failure forecasts for MPPs. The results of the present study demonstrate that integrating the CBR method with probabilistic and simulation-based modelling enables high-accuracy forecasting of the technical condition of SPPs. Achieving a RMSE of 0.32 and a coefficient of determination $R^2 = 0.93$ indicates the high effectiveness of the proposed approach. These values exceed the accuracy levels characteristic of many existing models, including purely data-driven algorithms, and approach the confidence threshold necessary for engineering decision-making under uncertainty.

Existing diagnostic methods for SPPs included both conventional approaches based on physical and mathematical modelling, and modern techniques using artificial intelligence and big data processing (Vychuzhanin & Rudnichenko, 2014). However, despite progress in predictive modelling, the issue of aligning predicted failures with actual ones remains open, directly impacting the effectiveness of maintenance and repair strategies. A comparison with recent studies in the field of diagnostics and failure prediction further confirms the relevance and competitiveness of the implemented method. For example, H. Moon & J. Choi (2021) proposed a hierarchical model based on Bayesian B-splines for predicting failures in naval engines. Their approach reflected the system's hierarchical structure at the levels of engine, engine type, and archetype, increasing flexibility in modelling multilayered technical systems. However, their model did not incorporate a comparison of predictions with actual failure events nor did it offer quantitative verification, limiting its applicability where prediction reliability must be evaluated. In the study by Ç. Karatuğ & Y. Arslanoğlu (2022), a condition-based maintenance strategy was implemented using machine learning techniques. While the model architecture enabled classification of equipment conditions, it was restricted to binary outputs ("functional" / "non-functional") and did not allow for quantitative prediction accuracy assessment or component-level resolution. In contrast, the approach presented in this study supported interval-based evaluations, scenario-specific conditions, and component-oriented comparison. M. Farid (2022) employed a hybrid architecture combining artificial neural networks and Gaussian process regression for real-time fatigue failure prediction. While an RMSE of approximately 0.36 was achieved, the method did not include visual diagnostics of prediction deviations or robustness testing across varying operational modes capabilities that were implemented in this study.

The study by N. Neykov & S. Stefanova (2023) examined the application of CBR under data-limited conditions, which was particularly relevant for ship systems where event recording was infrequent. Although the study confirmed the adaptability of CBR, it did not extend to more complex simulation scenarios, which was addressed in the current hybrid approach through the integration of probabilistic and simulation-based techniques. The analysis by M. Orhan & M. Celik (2023) noted the growing prevalence of hybrid methods, including SVMs, KNN, decision trees, and Bayesian networks, in recent diagnostics research. However, their

review remains general and lacks empirical validation against real failure data or justification for the maritime domain. In contrast, the approach in this study includes the full cycle from prediction to comparison with observed outcomes – enabling assessment of engineering applicability. S. Marandi *et al*. (2025) explored a promising approach utilising knowledge graphs and language models applicable to high-reliability systems. Despite technological novelty, their method lacked formal accuracy assessment and did not test the models in operational scenarios. This differentiates their method from the current study, where both quantitative validation (RMSE, MAE, MAPE) and visual discrepancy analysis were conducted. M. Arias Chao *et al*. (2020) proposed a hybrid framework combining physical modelling and deep learning, achieving a 127% increase in forecast horizon compared to purely data-driven models. This result supported the importance of integrating physical principles into predictive algorithms conceptually aligned with developed hybrid methodology, which fused heuristic experience (CBR), probabilistic processing, and failure scenario simulation.

In the study by Y. Liu *et al*. (2022), an LSTM model was used to predict the Remaining Useful Life (RUL) while accounting for noise and nonlinear dependencies. Although the model effectively handled time series data, it lacked simulation analysis tools that could capture rare or cascading failures. In contrast, the approach in this study reproduced complex scenarios with uncertainties, expanding diagnostic capabilities under limited observability. Z. Shi & A. Chehade (2021) introduced a dual LSTM architecture designed to capture transitional states and predict component RUL. The model achieved forecast accuracy in the range of RMSE ≈ 0.3-0.4, but, like many others, it did not include a step for comparing predictions with actual failures at the component or scenario level. This limited its reliability and adaptability for engineering practice. The method presented in this study addressed this gap by incorporating full-scale verification and visual deviation analysis. Thus, the integrated methodology combining CBR, probabilistic modelling, and simulation aligns with current requirements for accuracy, adaptability, and justification in forecasting the technical condition of marine power systems. The proposed approach demonstrated resilience to uncertainty, reproducibility across various operational scenarios, and the ability to substantiate conclusions quantitatively – making it a strong candidate for implementation in monitoring and maintenance systems for maritime engineering infrastructures.

CONCLUSIONS

The conducted study confirmed the relevance and practical significance of comparing predicted and actual failures in the diagnosis of the technical condition of SPPs. The main objective of the study – to assess the accuracy of failure prediction using various diagnostic approaches – was successfully achieved through the completion of all assigned tasks. The developed methodology enabled an objective evaluation of failure prediction accuracy using metrics such as MAE, RMSE, MAPE, and R^2. The comprehensive use of quantitative indicators, visualisation of deviations (including heatmaps and scatter plots), and correlation analysis between predicted and actual failures provided an in-depth assessment of the effectiveness of diagnostic models. The comparison results showed that the classical CBR method without adaptation demonstrated the lowest accuracy: the forecasts were systematically overestimated, with an RMSE of 0.65 and an F_1-score of 0.78. The use of adapted CBR incorporating probabilistic dependencies (e.g., Bayesian networks) reduced the average prediction error by approximately 30%, improved forecast accuracy, and lowered the number of false alarms. However, the best results were achieved through the integrated approach combining CBR, probabilistic models, and simulation modelling. In this case, RMSE dropped to 0.32, MAE to 0.21, and R^2 reached 0.93, indicating high predictive capability. A heatmap of deviations between forecasts and actual failures helped to identify areas of greatest uncertainty, especially in components with high operational load – the main engine and pumping system. This highlighted the importance of accounting for cascading failure effects and the need for detailed calibration of models for such units. Power supply and cooling systems demonstrated stable predictions and high diagnostic accuracy. Simulation modelling supplemented with probabilistic analysis proved effective in accounting for complex operational scenarios, including high loads and aggressive environments. It allowed refining predictions under conditions of incomplete or distorted data and contributed to more well-grounded diagnostic decisions. Thus, the study demonstrated that the integration of adaptive CBR, probabilistic models, and simulation modelling significantly improved failure prediction accuracy and the reliability of technical condition diagnostics for SPPs. The obtained results could be used to enhance intelligent maintenance systems, optimise operational processes, and increase the overall reliability of maritime transport. Future study will focus on developing dynamic model adaptation based on real-time operational data and conducting extended testing of the integrated approach across various types of SPPs.

ACKNOWLEDGEMENTS

None.

FUNDING

None.

CONFLICT OF INTEREST

None.

REFERENCES

[1] Arias Chao, M., Kulkarni, C., Goebel, K., & Fink, O. (2020). Fusing physics-based and deep learning models for prognostics. *ArXiv*. doi: 10.48550/arXiv.2003.00732.

[2] Farid, M. (2022). Data-driven method for real-time prediction and uncertainty quantification of fatigue failure under stochastic loading using artificial neural networks and Gaussian process regression. *International Journal of Fatigue*, 155, article number 106415. doi: 10.1016/j.ijfatigue.2021.106415.

[3] Karatuğ, Ç., & Arslanoğlu, Y. (2022). Development of condition-based maintenance strategy for fault diagnosis for ship engine systems. *Ocean Engineering*, 256, article number 111515. doi: 10.1016/j.oceaneng.2022.111515.

[4] Karatuğ, Ç., Arslanoğlu, Y., & Soares, C.G. (2023). Review of maintenance strategies for ship machinery systems. *Journal of Marine Engineering & Technology*, 22(5), 233-247. doi: 10.1080/20464177.2023.2180831.

[5] Liu, Y., Liu, Z., Zuo, H., Jiang, H., Li, P., & Li, X. (2022). A DLSTM-network-based approach for mechanical remaining useful life prediction. *Sensors*, 22(15), article number 5680. doi: 10.3390/s22155680.

[6] Maione, F., Lino, P., Maione, G., & Giannino, G. (2024). A machine learning framework for condition-based maintenance of marine diesel engines: A case study. *Algorithms*, 17(9), article number 411. doi: 10.3390/a17090411.

[7] Marandi, S., Hu, Y.-S., & Modarres, M. (2025). Complex system diagnostics using a knowledge graph-informed and large language model-enhanced framework. *ArXiv*. doi: 10.48550/arXiv.2505.21291.

[8] Moon, H., & Choi, J. (2021). Hierarchical spline for time series forecasting: An application to naval ship engine failure rate. *Authorea*. doi: 10.22541/au.159969715.57074848.

[9] Neykov, N., & Stefanova, S. (2023). Using case-based reasoning in system diagnostics and maintenance. In X.S. Yang, S. Sherratt, N. Dey & A. Joshi (Eds.), *Proceedings of seventh international congress on information and communication technology* (pp. 305-312). Singapore: Springer. doi: 10.1007/978-981-19-2394-4_28.

[10] Orhan, M.F., & Celik, M. (2023). A literature review and future research agenda on fault detection and diagnosis studies in marine machinery systems. *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment*, 238(1), 3-21. doi: 10.1177/14750902221149291.

[11] Pajak, M., Kluczyk, M., Muślewski, Ł., Lisjak, D., & Kolar, D. (2023). Ship diesel engine fault diagnosis using data science and machine learning. *Electronics*, 12(18), article number 3860. doi: 10.3390/electronics12183860.

[12] Rigas, S., Tzouveli, P., & Kollias, S. (2024). An end-to-end deep learning framework for fault detection in marine machinery. *Sensors*, 24(16), article number 5310. doi: 10.3390/s24165310.

[13] Shi, Z., & Chehade, A. (2021). A dual-LSTM framework combining change point detection and remaining useful life prediction. *Reliability Engineering & System Safety*, 205, article number 107257. doi: 10.1016/j.ress.2020.107257.

[14] Tveten, M., & Stakkeland, M. (2024). Fault detection in propulsion motors in the presence of concept drift. *ArXiv*. doi: 10.48550/arXiv.2406.08030.

[15] Vizentin, G., Vukelic, G., Murawski, L., Recho, N., & Orovic, J. (2020). Marine propulsion system failures – a review. *Journal of Marine Science and Engineering*, 8(9), article number 662. doi: 10.3390/jmse8090662.

[16] Vychuzhanin, V.V., & Rudnichenko, N.D. (2014). Assessment of risks structurally and functionally complex technical systems. *Eastern-European Journal of Enterprise Technologies*, 1(2(67)), 18-22. doi: 10.15587/1729-4061.2014.19846.

[17] Vychuzhanin, V.V., & Vychuzhanin, A. (2025). *Stochastic models and methods for diagnostics, assessment, and prediction of the technical condition of complex critical systems*. Lviv-Torun: Liha-Pres. doi: 10.36059/978-966-397-457-6.

[18] Wang, J., Cao, H., Cui, Z., Ai, Z., & Jiang, K. (2023). Intelligent fault diagnosis of marine diesel engines based on efficient channel attention-improved convolutional neural networks. *Processes*, 11(12), article number 3360. doi: 10.3390/pr11123360.

[19] Zhao, Y., Wang, S., & Chen, N. (2022). Thermal fault diagnosis of marine diesel engine based on LSTM neural network algorithm. *Vibroengineering Procedia*, 41, 198-203. doi: 10.21595/vp.2022.22515.

[20] Zhu, G., Huang, L., Yin, J., Gai, W., & Wei, L. (2023). Multiple faults diagnosis for ocean-going marine diesel engines based on different neural network algorithms. *Science Progress*, 106(4). doi: 10.1177/00368504231212765.

Гібридна модель для оцінки точності прогнозів відмов в суднових енергетичних установках

Володимир Вичужанін

Доктор технічних наук, професор
Національний університет «Одеська політехніка»
65044, просп. Шевченка, 1, м. Одеса, Україна
https://orcid.org/0000-0002-6302-1832

Олексій Вичужанін

Доктор філософії, асистент
Національний університет «Одеська політехніка»
65044, просп. Шевченка, 1, м. Одеса, Україна
https://orcid.org/0000-0001-8779-2503

Анотація. Зростаюча складність суднових енергетичних установок (СЕУ) та неоднорідність діагностичних даних ускладнюють забезпечення надійного прогнозування відмов та підтримання експлуатаційної безпеки. Метою дослідження було оцінити адекватність та точність прогнозів відмов у компонентах СЕУ з використанням гібридної діагностичної методології, а також перевірити запропоновану методику шляхом застосування розробленої моделі. Дослідження об'єднало в єдину систему міркування на основі прецедентів (CBR), імовірнісне моделювання та аналіз деградації на основі симуляції. Використовуючи діагностичні дані з 150 випадків експлуатації типових СЕУ (головний двигун, генератор, насос, системи охолодження та електропостачання), точність прогнозування моделі була кількісно порівняна з базовими підходами, включаючи класичні моделі CBR, адаптивні моделі CBR, ARIMA та моделі «дерева рішень». Інтегрований підхід досяг найкращих результатів (середньоквадратична похибка = 0,32, R^2 = 0,93), зменшивши середню похибку на 30-50 % порівняно з класичними методами. Кожен компонент гібридної структури зробив свій внесок: рівень CBR забезпечив контекстуальну узгодженість, імовірнісне моделювання зменшило невизначеність на ≈ 20 %, а аналіз на основі моделювання збільшив стабільність за наявності неповних або зашумлених даних на ≈ 15 %. На рівні підсистем найбільші поліпшення спостерігалися для головного двигуна (зниження середньоквадратичної похибки на 46 %) і насосної системи (39 %), що підтвердило адаптивність моделі до змінних умов експлуатації. Тести на надійність із 10-20 % відсутніми або порушеними вхідними даними продемонстрували стабільну точність (R^2 > 0,90), підтвердивши стійкість та практичну застосовність моделі. Запропонована методологія має практичну цінність для підвищення ефективності технічного обслуговування та надійності суднових енергетичних систем, оскільки забезпечує точне, інтерпретоване та стійке до шумів прогнозування відмов обладнання

Ключові слова: інтелектуальна діагностика; оцінка надійності; міркування на основі прецедентів; імітаційне моделювання; прогнозування технічного стану; байєсівські мережі; адаптивна модель

BULLETIN of Cherkasy State Technological University

Journal homepage: https://bulletin-chstu.com.ua/en

Vol. 31 No. 1. 2026

UDC 621.395.72:004.9
DOI: 10.62660/bcstu/1.2026.60

ISSN: 2306-4412
E-ISSN: 2708-6070

Improving the radio signal intensity loss model between the drone and the base station

Roman Zaivyi*
Postgraduate Student
Lviv Polytechnic National University
79013, 12 Stepan Bandera Str., Lviv, Ukraine
https://orcid.org/0009-0003-4096-4111

Volodymyr Pavlysh
PhD in Technical Sciences, Associate Professor
Lviv Polytechnic National University
79013, 12 Stepan Bandera Str., Lviv, Ukraine
https://orcid.org/0009-0004-3996-5923

Abstract. The purpose of the study was to develop a radio signal propagation model that considers the drone's flight altitude and the specifics of the ground-to-air channel to improve the accuracy of signal loss estimation. The study used comparative analysis of classical models, calculation of signal losses at different altitudes and distances, estimation of the effect of line-of-sight and birefringence propagation, and development and verification of an improved exponential model depending on the height of the drone. It was determined that these models do not account for the substantial influence of the drone's flight altitude and the increased probability of a line of sight between the drone and the base station, which leads to an overestimation of the predicted signal loss by tens of decibels. Based on the analysis, an improved analytical model was proposed, in which signal losses are defined as the sum of losses in free space and the height-dependent additional term. This additional term is described by an exponential function that decreases with increasing drone altitude and reflects a gradual transition from indirect visibility conditions to line-of-sight conditions. In addition, the model introduced distance correction, taking into account the actual geometry of the ground-to-air channel. Calculations performed for the frequencies of 868 MHz and 1,800 MHz showed that the improved model is in good agreement with experimental observations and provides results close to the free space model at significant altitudes. Comparison with the Hata models confirmed that the new model reduces the forecast error in urban conditions to tens of decibels. Considering the flight altitude and gradually moving from indirect to direct visibility greatly increases the accuracy of predicting signal loss for unmanned aerial vehicles. The practical importance of the study lies in the fact that the proposed model enables a more accurate prediction of radio signal losses in the ground-to-air channel, which is the basis for designing reliable communication systems with unmanned aerial vehicles

Keywords: radio communication; altitude; direct visibility; exponential correction; urban environment; analytical model

INTRODUCTION

Unmanned aerial vehicles are increasingly being used in the civil and military fields for surveillance, cargo delivery, and data collection. Stable radio communication between the device and the base station is critical to ensure reliable control of drones and information transmission. Traditional radio signal propagation models

Article's History: Received: 11.08.2025; Revised: 28.01.2026; Accepted: 16.03.2026; Published: 08.04.2026

Suggested Citation:
Zaivyi, R., & Pavlysh, V. (2025). Improving the radio signal intensity loss model between the drone and the base station. *Bulletin of Cherkasy State Technological University*, 31(1), 60-72. doi: 10.62660/bcstu/1.2026.60.

*Corresponding author

developed for ground-based cellular networks, such as the Okumura, Hata, and COST-231 Hata models, are based on statistics for transmitters and receivers located close to the ground and account for the effects of reflections, diffraction, and shielding by buildings. However, when applying these models to unmanned aerial vehicles, there are substantial limitations: drones can be located at an altitude of tens of meters or higher, which significantly increases the probability of line-of-sight, reduces the impact of ground interference, and changes the nature of signal attenuation. As a result, classic ground-based models often overestimate signal losses, especially in urban environments.

In the field of accurate signal loss prediction for ground-to-air scenarios under study, the problem of considering the drone's altitude and line-of-sight effects remains. O.V. Troyanskiy & P.O. Kandieiev (2025) showed that classical ground-based models overestimated signal loss at high altitudes due to ignoring the line of sight (LoS). It was necessary to address the altitude of the drone to accurately predict the signal strength. N. Moraitis *et al.* (2023) noted that drone amplification reduced the multipath propagation effect and reduced average signal loss. It was proposed to account for the exponential reduction of additional losses with increasing height to adapt the model. F. Fuschini *et al.* (2022) demonstrated that considering the LoS between the drone and the base station improved the accuracy of signal strength prediction. It is proven that models that disregarded height significantly overestimated losses in line-of-sight conditions. O.Yu. Malyi *et al.* (2024) pointed out that at short distances, classical models often overestimated losses, while the two-beam approach accurately reflected interference effects. It was determined that bearing the effect of reflection from the ground in mind was crucial for low drone flights. N. Ranchagoda *et al.* (2021) emphasised that the type of terrain and density of buildings drastically affected the characteristics of the ground-to-air channel. It was stated that corrections for urban and open areas were necessary to improve the accuracy of the model.

A. Chornyi *et al.* (2025) demonstrated that an exponential decrease in additional losses with increasing altitude reflects the transition from indirect to direct visibility. Experimental observations confirmed that this relationship was in good agreement for different heights. H.-S. Shin *et al.* (2024) showed that geometric correction of the distance between the drone and the base station improved the accuracy of modelling at small horizontal distances. It was recommended to use the adjusted distance in formulas for a more accurate estimate of signal loss. S.K. Alavipanah *et al.* (2022) proved that taking the average height of buildings in urban areas into account allowed adjusting the shading effect. Without this amendment, the models significantly overestimated losses in dense buildings. Z. Xiao *et al.* (2025) established that changes in signal frequency significantly affected losses and should have been considered in the simulation. High frequencies were found to be more sensitive to interference and multipath propagation. Q. Zhu *et al.* (2021) showed that altitude-dependent parameters allow the model to be adapted for various environments, including urban, suburban, and open areas. It is emphasised that this approach provided more reliable forecasting and improved the efficiency of drone control systems. These studies have shown that increasing the drone reduces multipath attenuation, considering the LoS and altitude improves simulation accuracy, and geometric and exponential corrections allow models to be adapted for different types of terrain. However, there is a need to develop a generalised model that simultaneously takes into account flight altitude, the impact of reflections from the ground, and adapts to different environments to accurately predict signal strength.

The study aimed to develop a signal propagation model that accounts for the drone's flight altitude, ground-to-air channel geometry, and environmental influences, providing a more accurate assessment of signal losses and improving the reliability of unmanned control and data transmission systems. The objectives of the study were to review existing approaches to modelling signal propagation in the ground-to-air channel, consider the influence of drone flight altitude and environmental features on signal characteristics, and evaluate the accuracy of classical and improved signal loss models and their effectiveness for unmanned data transmission systems.

MATERIALS AND METHODS

The study examined methods for estimating radio signal loss to ensure reliable communication between a base station and an unmanned aerial vehicle (UAV). The fundamental Free Space Path Loss (FSPL) model was analysed, which describes ideal conditions for signal propagation without interference (Al-Hourani *et al.*, 2014) (1):

$$PL_{FS(d)} = 32.44 + 20\,log^{10}(d) + 20\,log^{10}(f), \quad (1)$$

where d – the distance between the transmitter and receiver, km; f – the signal frequency, MHz.

When developing the improved model, the FSPL value was used as the lower limit of signal loss, 91.5 dB at a distance of 0.5 km at 1,800 MHz. Empirical ground-based models were also considered, in particular, the Okumura model, which is based on experimental data and considers average signal losses depending on distance and terrain features (Gupta, 2013) (2):

$$L^{50}(d,f) = L_{FS(d,f)} + A_{\mu(d,f)} - G_{ht} - G_{hr} - G_{env}, \quad (2)$$

where $L_{FS\,(d,f)}$ – loss of free space; $A_{\mu\,(d,\,f)}$ – average additional losses (deviations from free space) according to measurement data for distance and frequencies f; G_{ht} and G_{hr} – coefficients (wins) due to the height of

the transmitting and receiving antennas, respectively; G_{env} – coefficient considers the type of terrain (urban, suburban, open).

A model of M. Hata (1980) was examined as an analytical approximation of the Okumura model, which provides explicit formulas for calculating average losses in urban and suburban conditions (3):

$$L_{Hata}(d) = 69.55 + 26.16\ log_{10}(f) - 13.82\ log_{10}(h_b) - \\ - a(h_m) + (44.9 - 6.55\ log_{10}(h_b))\ log_{10}(d), \quad (3)$$

where f – frequency, MHz (from 150 to 1,500 MHz); h_b – base station antenna height, m (from 30 to 200 m); h_m – mobile receiver antenna height, m (from 1 to 10 m); d – distance, km (1-20 km).

The COST-231 Hata model was examined as a modification of Hata for high frequencies and densely built-up large cities with additional correction factors (European Commission: Directorate-General..., 1999) (4):

$$L_{COST231}(d) = 46.3 + 33.9\ log_{10}(f) - 13.82\ log_{10}(h_b) - \\ - a(h_m) + (44.9 - 6.55\ log_{10}(h_b))\ log_{10}(d) + C, \quad (4)$$

where all values are in the same units as in (3), and C – an additional coefficient for large cities: $C = 3$ dB for the metropolitan area and $C = 0$ for medium-sized cities.

During the next stage, the paper systematised the factors that affect signal losses in the air-ground channel. Key aspects were highlighted: the height of the drone's flight, the presence of a LoS between the transmitter and receiver, the nature of multipath propagation, and the effect of reflection from the Earth's surface. For each of these factors, it was discussed how it changes the signal intensity and why classical ground-based models do not take these effects into account or do so insufficiently. Special attention was paid to the dependence of losses on altitude, which determines the probability of a LoS and reduces the impact of interference and multipath. The theoretical basis of the study was the recommendations of ITU-R, where the FSPL model was used as a baseline scenario: P.525 defined the method for calculating free losses, P.528 described ground-to-air radio lines, and P.1411 – short-distance systems in urban and suburban settings (Uko *et al.*, 2024).

Based on the analysis, the concept of an improved model was proposed. The main idea was to combine the classic free space model with an additional term that considers excessive signal loss associated with the drone's environment and altitude. This term was defined as an exponential function of altitude, which allows the model to smoothly transition from ground-based Non-Line-of-Sight (NLoS) conditions to line-of-sight conditions as the vehicle's height increases. Thus, the model becomes able to correctly describe the channel over a wide range of heights, without resorting to various empirical corrections for each specific scenario. Calculations were performed in two typical scenarios using comparative analysis to evaluate the behaviour of the models, which allowed comparing the results of classical models and the proposed improved model for different drone heights, distances, and signal frequencies. In the first scenario, the frequency $f = 1{,}800$ MHz was used, and the base station height was = 30 M, and horizontal distances ranged from 0.5 km to 5 km; calculations were performed for FSPL, equation (1), Hata model for the city (formula (3), $h_m = 1.5$ m), Hata models for open terrain with appropriate correction and improved model (Formula 7). In the second scenario, the frequency $f = 868$ MHz with similar parameters was used to compare the effect of frequency on signal loss.

In addition, the work considered the geometric aspects of signal propagation, particularly the use of an effective distance between the base station and the drone, including the flight altitude. This allows for adequately describing the signal intensity even at short distances or at a drone height drastically higher than the height of the base station antenna. The inclusion of average effects of ground reflections and possible interference phenomena, which are characteristic of two-beam models, was also discussed with the aim of improving the accuracy of modelling in real conditions. The final stage of the work was the formulation of the analytical dependence of signal losses on the horizontal distance and height of the drone, which includes parameters that can be interpreted physically and adjusted to different environments: urban areas, suburbs, open areas, or woodlands. This approach allows creating a unified model that covers different scenarios for using UAVs without having to switch between several classic models.

RESULTS

Overview and characteristics of known radio signal propagation models

The most basic model of signal attenuation in free space, FSPL describes the conditions of an ideal LoS without interference. It was based on the Fris equation applied to the far field of an antenna (Al-Hourani *et al.*, 2014). According to this model, the power received by the receiver decreases inversely proportional to the square of the distance d between the transmitter and receiver and the square of the frequency (or directly proportional to the square of the wavelength). In logarithmic terms (in decibels), losses in free space are written as (1). Formula (1) is based on the assumption that isotropic antennas are used and that there is no reflection or absorption of the signal. The FSPL model is used in the ITU-R recommendations as a baseline scenario for line-of-sight signal loss (Uko *et al.*, 2024). In these documents, the FSPL is considered as a reference value that determines the minimum possible level of power loss under ideal conditions. In real-world conditions, in particular, when providing communication with UAVs, this model has limitations related to the altitude

orientation of the antenna. It does not address the variable altitude of the drone, the presence of obstacles in urban or forest environments, the influence of the angle of inclination of the signal trajectory, and dynamic environmental factors such as the movement of objects or weather conditions. Due to the influence of interference, multipath propagation, and other factors, signal loss in real conditions increases (Isabona *et al.*, 2023). In further development of the improved model, the FSPL values will be used as the lower limit of signal loss, i.e., as the most optimistic scenario.

One of the first empirical models for ground-based radio communication was the Okumura model, developed based on large-scale measurements in the Tokyo area (Gupta, 2013). This model does not have a simple analytical dependency since it is based on graphs and reference tables. It defines the median signal strength (or loss) as the sum of losses in free space and empirical correction components. In general, the average signal loss according to the Okumura model can be represented as (2). Value A_μ and G-coefficients are determined from empirical curves. The Okumura model covers frequencies from about 100 MHz to 1,920 MHz and distances from 1 to 100 km. It accounts for different types of terrain (densely built-up urban, suburban, open rural), introducing appropriate amendments. Although the Okumura model is quite accurate (error ~10-14 DB) for conditions similar to those in which measurements were made, its disadvantage is the lack of a simple formula – the need to use graphs reduces the convenience of practical application.

A model of M. Hata (1980) was developed as an analytical approximation of the experimental Okumura model. It is known as the "Okumuri-Hata model" and has become the standard for calculating losses in terrestrial cellular networks. In contrast to Okumura's original model, Hata presented empirical dependencies in the form of easy-to-use formulas describing average signal losses. For urban development conditions (small/medium city), this model looks like (3). Function $a(h_m)$ – correction for the height of the mobile receiver, depending on the size of the city. For medium/small cities: $a(h_m) = (1.1\ log_{10} f - 0.7)\ h_m - (1.56\ log_{10} f - 0.8)$, and for a large city, formula $a(h_m)$ other (quadratic dependence). Thus, (3) determines the average signal loss in urban settings. For other types of terrain, additional corrections are introduced: for suburban areas, the value (3) is reduced by the $2\,[log_{10}(f/28)]^2 + 5.4$ factor, and for open rural areas – factor $4.78\,[log_{10}(f)]^2 - 18.33\ log_{10}(f) + 40.94$ (these negative values reduce losses compared to the urban environment). The Hata model actually translates Okumura curves into a convenient analytical form and covers the frequency range of 150-1,500 MHz and distances up to 20 km. The average error of the model is about 6 dB, which is considered an acceptable level for engineering calculations of the coating. Its limitations are similar to the Okumura model: it is correct only within the parameters obtained during measurements, and does not consider random signal fluctuations, in particular, fading or shadow attenuation. Therewith, due to its simplicity and compliance with practical data, Formula (2) has become widely used in network planning (Djomadji *et al.*, 2023).

For use at higher frequencies (up to 2 GHz) and in conditions of dense urban development, a modification of the Hata model was created as part of the Action 231 of COST-231 project in 1990s (European Commission: Directorate-General..., 1999). The COST-231 Hata model retains the structure of Formula (3), but includes an additional correction term and expands the operating frequency range. For urban areas, it is presented in the form of (4). Formula (4) is valid for frequencies from 1,500 MHz to 2,000 MHz. Thus, the COST-231 Hata model considers the tendency for losses to increase at higher frequencies by partially adjusting the frequency term (compared to (3), the coefficients at $log_{10}(f)$ and the constant values are changed) and introduces a correction C for the densest urban development. Within its application range, this model provides precision similar to the original Hata model.

Practical applicability and limitations of the considered signal propagation models

The classical models considered are based on a terrestrial communication scenario, where the base station is located on a mast or building (h_b up to ~200 m), and the receiver is at human or vehicle level (h_m up to ~10 m). They empirically consider the impact of urban development through fixed corrections, but do not reflect changes in the nature of signal propagation when the receiver is substantially raised above the average level of buildings. The free space model (1) describes ideal conditions, while the Okumura-Hata models (3-4) provide averaged losses for typical urban conditions without LoS, accounting for multiple reflections and diffraction from terrain and buildings. Therefore, when using these models for drones that can fly tens of meters or higher, it is necessary to consider their limitations.

One of the key limitations is the receiver height (Tataria *et al.*, 2021). Okumura Hata models were designed for $h_m \leqslant 10$ m. For drones, the height of the receiving antenna can be intensely higher – from 50-100 m for small models to hundreds of meters for high-altitude vehicles. Extrapolation of Hata formulas to such heights has not been experimentally confirmed and may give inaccurate results. In particular, in the Hata model, the term $a(h_m)$ for large h_m stops growing (for $h_m > 10$ m, it is actually stabilised), i.e., raising the antenna above 10 m is disregarded. In fact, raising the drone reduces the impact of interference on the ground, which should reduce signal loss. Therefore, without additional adjustments, the Hata model may overestimate the loss estimate for drones at high altitudes.

Another important aspect is the availability of LoS (Waqas *et al.*, 2024). In urban settings, direct visibility over long distances is usually absent for terrestrial

subscribers – the signal mainly reaches the receiver through reflection and diffraction. These are the conditions that the Hata and COST-231 models consider. Drones flying above roofs are much more likely to establish a LoS connection to the base station. In LoS mode, the signal loss approaches the free space model, where the attenuation index is $n \approx 2$. Therefore, classical empirical models focused on NLoS overestimate losses in the case of LoS. This is confirmed by field observations: as the height of the drone increases, the share of direct signal increases, and additional losses decrease. For example, when the receiver is raised from 10 m to 40 m, the average fading value (log-normal shadow stagnation) decreases exponentially from about 14 dB to 3 dB, which indicates a substantial improvement in propagation conditions. Classic models are also limited at short distances (ElSayed & Mohamed, 2022). In particular, the Hata model is formally applied for $D \geqslant 1$ km. For drones, the range is much shorter – for example, during takeoff, landing, or hovering directly over a base station, when d can be hundreds of meters or less. In such cases, the classical model is extrapolated beyond the original data and may produce unreliable results. On the contrary, the free space model correctly describes the situation even for minimal d, considering the transition to the near zone of the antenna, if necessary. Thus, at short distances, empirical models often overestimate losses, since they disregard the presence of a LoS and the reduced number of obstacles.

Special attention should be paid to the influence of the Earth's surface and reflections (Zhao *et al.*, 2025). In the ground-to-air scenario, signal propagation is characterised by reflection from the ground. A drone located at a certain altitude simultaneously receives a direct signal and a signal reflected from the surface of the ground or water. The superposition of these two waves on receivers can cause interference effects: at some distances, the signal can be amplified or weakened, as described by the two-beam model. For a ground receiver, the reflected signal is usually partially shielded by surrounding objects, while for a drone, this restriction is absent. The classic Hata and Okumura models disregard the explicitly two-beam nature of the channel. Because of this, at certain distances, the actual signal losses may exceed the values predicted by the free space model due to partial destructive interference of the direct and reflected signals. This is especially noticeable when the drone flies at a low altitude over an open area or water surface, forming so-called beam web zones, where the signal level changes with distance. Thus, classic models need to be adapted for drone scenarios. It is necessary to consider the increase in efficiency due to height, which reduces losses due to the presence of LoS, along with the effects of the birefringence of the signal.

The improved model should take into account the height of the drone as one of the parameters that affect losses. At $h \to 0$ (drone on the ground), the model must coincide with the classic ground model (for example, a Hata for a suitable environment), and at a large h, it must approach the loss of free space. The model should consider the possible influence of the reflected beam from the Earth – for this purpose, it is desirable to reflect an interference increase in losses at some distances, or at least account for the average effect of reflections (for example, an increase in the effective indicator of the degree of attenuation over long distances). The model should retain its usability – preferably an analytical formula with a small number of parameters that can be interpreted. In the two-beam model, there is a so-called fracture distance d_b, after which, the reflected and direct rays begin to interfere significantly, and the effective attenuation law changes from $1/d^2$ to approximately $1/d^4$. When the transmitter height is h_b and the height of the receiver h (drone), this distance is estimated as:

$$d_b \approx \frac{4\pi h_b\, h}{\lambda}, \tag{5}$$

where λ – wavelength. For example, for a base station $h_b = 30$ m, drone $h = 50$ m, and frequency f 2GHz ($\lambda \approx 0.15$ m) is obtained $d_b \approx 0.19$ km, that is, about 190 m. At distances less than d_b signal losses ~ both in free space and in large ones-additionally increase due to interference with reflection (by tens of dB after several kilometres). In real-world conditions, sharp interference dips are somewhat smoothed out due to surface irregularities, angle changes, etc., so this effect was considered on average as a smooth increase in losses with distance.

Based on the free space model (1), a term that considers excessive signal losses (in addition to FSPL) due to environmental factors can be added to it – mainly shielding and multipath. This additional term is denoted as $\Delta L(h)$. It should depend on the height of the drone h so that $\Delta L \to 0$ at high h (when the drone is high and nothing significantly interferes with the signal, except for free space), and ΔL peaked at h close to zero (when the drone is at ground level and the conditions correspond to the ground NLoS case). The simplest choice is an exponential dependence, which has the property of asymptotically approaching zero:

$$\Delta L(h) = A exp(-\frac{h}{h_0}), \tag{6}$$

where A – the excess loss coefficient at $h = 0$ (i.e., the difference between ground-based NLoS losses and free space), and h_0 – characteristic "effective height", at which additional losses fall e times. The parameter A reflects the influence of the environment: it is smaller for open areas and larger for densely built-up areas. Typical estimates can be as follows: $A \approx 30\ dB$, $h_0 \approx 20\ m$ for urban conditions; $A \approx 10 - 15\ dB$, $h_0 \approx 5 - 10\ m$ for open areas. These values are consistent with the analysis of the Hata model: for example, for a city at $h_m = 1.5\ m$ (about 0), the Hata model produced 30-40 DB

more losses than the FSPL. With an increase of h up to tens of meters, the exponent (6) decreases sharply, which corresponds to the transition to conditions close to LoS (De Beelde *et al.*, 2022).

Reflection from the ground can be considered due to the effective distance d', which is equal to the length of a straight line between the base and the drone. Since the drone is at an altitude of h, geometrically $d' = \sqrt{d^2 + h^2}$, where d – horizontal distance (projection on the ground). For large horizontal distances at low altitudes, the difference between d' and d minor, but for small ones, it improves accuracy (for example, if the drone hangs high almost above the base, d it can go to 0, whereas $d' = h$ – this is the correct distance for FSPL). Thus, d' is substituted instead of d in Formula (1). Combining these considerations, the improved signal loss model can be represented as a function of the horizontal distance d and the height of the drone h:

$$PL_{Drone}(d,h) = 32.44 + \\ + 20\,log_{10}\left(\sqrt{d^2+h^2}\right) + 20\,log_{10}(f) + Aexp\left(-\frac{h}{h_0}\right). \quad (7)$$

Formula (6) is given in dB and is valid for d in km, h in m, f in MHz (just like formulas (1-4)). The first two terms (up to "+") correspond to the loss of free space at a distance d' (i.e., FSPL between the base and the drone). The last term adds excess losses ΔL, which depend on the height. At $h = 0$ this term is equal to A (for example, ~30 dB in a city, which makes PL_{Drone} close to the values of the Hata model). For large cases $h \gg h_0$ the term tends to 0, and $PL_{Drone} \to FSPL$, as required for full line-of-sight conditions. Thus, (4) smoothly interpolates between the ground-based empirical model and the free space model as a function of the height of the device.

Coefficient A characterises the maximum possible additional losses due to the environment. Its value is related to the average effect of buildings, trees, and other obstacles on the ground level. Parameter h_0 can be interpreted as "effective obstacle height". If the drone rises to approximately a height of h_0, most interference ceases to significantly affect propagation, and losses are dramatically reduced. For the city h_0, the order of tens of meters correlates with the average height of buildings (for example, the average height of a building in a certain area can be ~15 m – if the drone is higher, most buildings no longer shade it) (Mirzabeigi *et al.*, 2025). For open areas h_0 is low, as it is enough to climb a few meters to provide almost LoS above the terrain.

Formula (7) gives the average signal loss. Statistical fluctuations (shadow attenuation, multipath feeding) can be additionally considered. These data show that as the drone's height increases, the RMS deviation of random signal level fluctuations decreases. This can be considered by introducing a height-dependent standard deviation in the model (for example, as $\sigma(h) = \sigma_0 exp(-h/H)$). At the stage of the average model, they are limited to a deterministic estimate (7), while practical calculations of the power reserve involve the inclusion of statistical margins.

Comparative analysis of the accuracy and reliability of various propagation models

2 scenarios were considered to show the differences between the known classical models and the proposed improved model (7), followed by the calculation of signal losses. The first scenario concerns frequency $f = 1{,}800$ MHz, which is typical for cellular communication, the second – the frequency $f = 868$ MHz used to control drones (Table 1).

Table 1. Example of calculating signal losses based on different models ($f = 1{,}800$ MHz, $h_b = 30$ m)

Model	**a**	**b**	**c**	**d**
Distance, km	FSPL (dB)	Hata (city) (dB)	Hata (open) (dB)	Advanced model (6) (dB)
0.5	91.5	123.6	91.7	94.0
1.0	97.5	134.3	102.3	100.0
2.0	103.6	144.9	112.9	106.0
3.0	107.1	151.1	119.1	109.6
4.0	109.6	155.5	123.5	112.0
5.0	111.5	158.9	126.9	114.0

Source: compiled by the authors

The Hata model was analysed separately for open terrain and cities due to various empirical corrections, while the FSPL and improved model consider the impact of the environment on an average basis, so the division by terrain type was not applied. Calculating signal losses for a frequency of 1,800 MHz and a base station height of 30 m shows that different models predict different levels of attenuation depending on the distance. The FSPL gives the lowest loss values that are adequate for line-of-sight conditions without interference, while the urban Hata model seriously overestimates losses due to considering buildings and other obstacles in the city. The open Hata model assumes intermediate values close to FSPL over short distances, but with a greater increase in losses over long distances. The improved model, based on (6), shows more balanced results, correctly accounting for terrain and urban development, and provides a compromise between the conservative assessment of Hata and the optimistic FSPL. This shows the importance of choosing a model based on the environment and specific forecast accuracy requirements (Table 2).

Table 2. Example of calculating signal losses based on different models (f = 1,800 MHz, h_b = 30 m)

Model	a	b	c	d
Distance, km	FSPL (dB)	Hata (city) (dB)	Hata (open) (dB)	Advanced model (dB)
0.5	85.2	115.4	87.0	87.7
1.0	91.2	126.0	97.6	93.7
2.0	97.2	136.6	108.2	99.7
3.0	100.8	142.8	114.4	103.2
4.0	103.3	147.2	118.8	105.7

Source: compiled by the authors

The calculation of signal losses for a frequency of 868 MHz and a base station height of 30 m shows that the models differ in the attenuation estimate depending on the distance. The FSPL shows the lowest losses because it only accounts for the direct signal path without interference, while the urban Hata model overestimates losses due to the impact of development, especially over long distances. The open Hata model predicts losses slightly higher than the FSPL, but significantly lower than the urban Hata model, reflecting conditions of weak or no obstructive development. The improved model offers intermediate values, more realistically estimating losses in different environments, and provides a balance between a conservative Hata estimate and an optimistic FSPL, making it suitable for practical wireless network planning. The main models for calculating signal losses based on various efficiency and applicability criteria are compared (Table 3).

Table 3. Model comparison

Model	a	b	c	d
Criterion	FSPL	Hata (city)	Hata (open)	Improved model (7)
Accuracy for LoS	good	overvalued	Moderate	high
Consideration of the height of the drone flight	no	partially (via h_m)	partially	yes (via h)
Smooth transition between conditions	no	no	no	yes
Relevance for the air-ground channel	limited	unadapted	partially	high
Implementation complexity	low	average	average	moderate
Proximity to FSPL at high h	-	-	-	secured

Source: compiled by the authors

As can be seen from the comparison, the FSPL model is easy to use and accurate enough for LoS conditions, but it does not consider the height of the drone and the complexity of the environment. The urban model of the Hata greatly overestimates losses in urban conditions and is not adapted to air-ground channels. The open Hata model shows moderate accuracy and partially considers the height, but does not provide a smooth transition between different conditions. The advanced model (7) stands out for its high accuracy for LoS, realistic drone altitude accounting, smooth transition between conditions, and proximity to FSPL at high altitudes, making it most suitable for modern applications in wireless networks with drones or other aerial platforms. A graph of the dependence of signal loss on distance for several models (Fig. 1) is presented to evaluate the behaviour of models at different distances. The FSPL curve shows the theoretical minimum losses in the case of LoS, the Hata curve for the city shows typical losses in dense buildings (NLoS), and the Hata curve for open terrain illustrates an intermediate option with minimal interference.

A similar comparison for a different frequency allows estimating the effect of frequency on signal loss (Fig. 2).

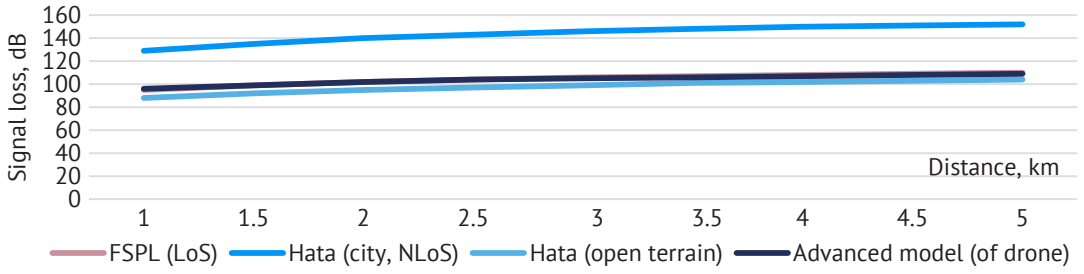

Figure 1. Comparison of signal loss models

Notes: FSPL (direct visibility); Hata model for the city (dense development, typical NLoS); Hata model for open terrain; advanced model (7). Frequency 1.8 GHz, h_b = 30 m, h_m = 1.5 m

Source: compiled by the authors

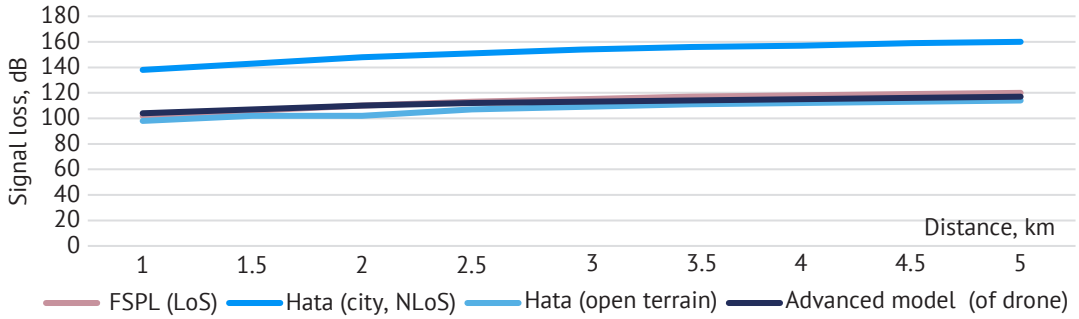

Figure 2. Comparison of signal loss models
Notes: FSPL (direct visibility); Hata model for the city (dense development, typical NLoS); Hata model for open terrain; advanced model (7). Frequency 868 MHz, $h_b = 30$ m, $h_m = 1.5$ m
Source: compiled by the authors

Analysis of calculations for both frequencies (1,800 MHz and 868 MHz) demonstrates a stable advantage of the improved model (7) over classical approaches. It provides less deviation from theoretical losses in LoS conditions compared to Hata models, especially in urban scenarios where the latter significantly overestimate values. In addition, model (7) adapts to changes in frequency and altitude, ensuring correct interpolation between open environment and line-of-sight conditions. This makes it suitable for cellular networks and LoRaWAN, IoT alike, along with other systems using unmanned platforms.

The proposed formula (7) for the improved model actually describes a family of curves parameterised by height h. Its main property is the ability to interpolate between classical signal propagation modes. For small altitudes of h, it coincides with the results of the Hata model (with the correct choice of parameter A), and for large h – asymptotically approaches FSPL. This is convenient for networks with drones, where one device can move between dense buildings and high altitudes above the city: there is no need to switch between different models, just substitute the current altitude in (7).

The difference between the improved model is that it explicitly considers the coupling geometry and displays height-dependent shading. A typical scenario is considered: a drone at an altitude of $h = 50$ m above the city, the distance to the base station is 5 km. Let the model parameters be: $a = 30$ dB, $h_0 = 20$ m. The following results were obtained:

$$(FSPL\,(5\ km) = 111 dB);$$

$$\Delta L(50) = 30exp\left(-\frac{50}{20}\right) \approx 30e^{-2.5} \approx 2.5\ dB;$$

$$(PL_{Drone} \approx 111.5 + 2.5 = 114\ dB).$$

For comparison, using the Hata (city) model for the same conditions, ≈ 159 DB was obtained. The difference is about 45 dB – that is, the classical model seriously overestimates losses in LoS conditions. If the drone drops to $h = 5$ m:

$$(\Delta L\,(5) \approx 30e^{-2.5} \approx 26\ dB);$$

$$(PL_{Drone} \approx 111.5 + 26 = 137.5\ dB).$$

The obtained value approaches the level of the model for open terrain (≈ 126.9 DB), but slightly exceeds it, which is quite logical given the influence of partial shading. Thus, by adjusting the parameters A and h_0, it is possible to accurately adapt the model to urban, suburban, forest, or mountain conditions – unlike Hata models, which have fixed empirical parameters. The approach laid down in formula (7) is consistent with the trends of modern channel models. For example, T.T. Oladimeji *et al.* (2024) proposed a modified Close-In model, in which the attenuation index depends on the height by the formula: $n\,(h) = 2.6h^{-0.53} + 2$, and the additional losses decrease linearly with increasing h. If the exponential function from model (7) is decomposed into a Taylor series, its approximation to a linear relationship at low altitudes can be shown. This means that model (7) has a similar behaviour, but at the same time, a simpler structure and a clear physical meaning of parameters.

DISCUSSION

As a result of the calculations, it was found that the improved signal loss model for the ground-to-air scenario showed significantly more accurate results compared to the classic Hata and cost-231 Hata models. For a drone at an altitude of 50 m above urban buildings, the calculated signal loss according to the proposed model was about 114 dB, while the Hata model for the city gave a value of about 159 DB. This difference of more than 40 dB confirmed that classical ground-based models greatly overestimate losses in LoS conditions, especially over long distances. This problem was also investigated by M.K. Elmezughi *et al.* (2022), where the results confirmed that in LoS conditions, signal losses for classical models, such as the free space model, are usually estimated by the simple inverse square law of distance. Advanced models take additional factors into account, such as atmospheric influence, refraction, and diffraction, which enable a more accurate prediction

of losses over long distances. The comparison shows that classical models underestimate signal loss on high climbs or long transmission routes, while improved models provide a more realistic estimate.

A study by X. Huang *et al.* (2024) also indicated that increasing altitude affects signal attenuation due to atmospheric rarefaction and increasing distance to the receiver. Improved models allow paying attention to this effect and predicting changes in the quality of communication in real time. Therefore, for safe and efficient operation of drones at high altitudes, it is important to use accurate signal loss models that reflect real-world conditions. Notably, the accuracy of predicting signal loss directly affects the route planning and power consumption of drones. Underestimating losses in classical models can lead to unpredictable communication disruptions and reduced mission efficiency. Therefore, the implementation of advanced models is crucial to ensure the reliability of data transmission in complex environments.

Analysis of the dependence of signal losses on altitude showed that the exponential correction ΔL (h) effectively considered the gain from lifting the drone antenna. When the height was reduced to 5 m, the additional losses increased to 26 dB, which was close to the results for open terrain according to the Hata model. This indicated an adequate interpolation between NLoS conditions on Earth and near-full line-of-sight conditions at high altitudes, which allowed correct estimation of signal losses for various flight scenarios. I. Bennageh *et al.* (2024) concluded that an increase in the drone's flight altitude directly affects additional signal losses as the distance to the receiver increases and atmospheric conditions change. Exponential correction allows accounting for nonlinear loss growth with height and improves the accuracy of coupling modelling. This is especially important for high-flying drones, where classic models may underestimate the actual signal attenuation. B.S. Shawel *et al.* (2024) established that interpolation between ground-based NLoS and aerial LoS conditions allows creating a flexible model that considers different signal transmission scenarios. For UAVs, this technique allows predicting changes in the quality of communication during the transition from a dense urbanised zone to an open space. The use of interpolation approaches provides more reliable route planning and optimisation of data transmission parameters. These results support the above study, as they demonstrate a clear dependence of additional signal losses on the drone's flight altitude. The introduction of an exponential correction allows a more accurate prediction of signal attenuation at high altitudes. Therefore, taking these effects into account is crucial to ensure stable communication and efficient operation of the UAV.

A comparison of the results for two frequencies – 1,800 MHz and 868 MHz – showed that the proposed model remained stable and consistent with the physical patterns of radio signal propagation. Signal losses decreased as the frequency decreased, which was in line with expectations, while distance-dependent trends remained the same. This confirmed the versatility of the model for applications in both cellular communications and low-frequency drone control systems, such as LoRaWAN or IoT. The paper of P. Ramamoorthy *et al.* (2023) deserves special attention, showing that the stability and consistency of the model for different frequencies of the cellular and IoT spectrum ensure its versatility and practical applicability in many scenarios. A reliable model should correctly describe signal losses for both high-frequency and low-frequency bands, considering different permeability due to interference and atmospheric effects. This allows creating predictable and consistent forecasts for different types of wireless networks, which is especially important for integrated communication systems.

In turn, R. Robles-Enciso *et al.* (2024) concluded that a comparison of signal loss behaviour shows that high-frequency signals experience greater losses in dense media due to their greater sensitivity to interference. Low-frequency bands show better permeability and lower losses over the same distances, making them suitable for large-coverage IoT applications. This comparison allows optimal selection of frequency bands depending on environmental conditions and communication stability requirements. These data were consistent with the theses given in the previous section, as they confirmed the consistency of the model for different frequency ranges. The model demonstrated stable behaviour for both cellular bands and the IoT spectrum, which highlights its versatility. This allows it to be used to predict signal loss in various network scenarios without the need for significant adaptation.

An important result was the confirmation of the effect of the beam reflected from the ground on signal loss. The use of a geometrically correct distance $d' = \sqrt{(d^2 + h^2)}$ allowed considering the effect of birefringence, especially at short distances and at low altitudes. This gave the model additional physical correctness and explained the differences between real measurements and classical ground-based models that ignore reflections from the surface. F.K. Banaseka *et al.* (2021) also conducted a study, the results of which confirmed that taking into account the two-beam propagation and reflection of the signal from the ground allows for more accurate modelling of real-world communication conditions for UAVs. This approach addressed both the direct signal and the reflected signal from the Earth's surface, which is especially important at low and medium flight altitudes. Due to this, it is possible to predict the effects of interference and signal fluctuations that affect the stability of communication.

S. Wang *et al.* (2023) also determined that the influence of geometrically correct distance on the accuracy of estimating UAV signal losses is critical, since disregarding the real propagation trajectory can lead to major errors. Using a precise geometric distance allows for more correct consideration of phase changes and signal attenuation on the way to the receiver. This increases the reliability of forecasts and the efficiency of route planning for the safe operation of drones.

Comparing the data obtained in the course of research, it can be concluded that considering birefringence significantly increases the accuracy of signal loss modelling. Interference effects from the reflected signal can drastically change the reception level, especially at low and medium altitudes. Therefore, the inclusion of this factor is necessary to create realistic and reliable forecasts of UAV operation. Compared to the Hata and cost-231 Hata models, the improved model showed a smoother transition between signal propagation modes, which made it suitable for scenarios with dynamic changes in drone height. This allowed avoiding sudden transitions in calculations and provided a more realistic assessment of coverage, especially in conditions of dense urban development or when flying over open areas.

D.K. Nishad *et al.* (2025) concluded that a smooth transition between signal propagation modes for dynamic drone altitudes allows the model to correctly react to changes in conditions from LoS to partially blocked trajectories. This is especially important when manoeuvring drones in difficult landscapes or when changing the altitude of flight, when rapid changes in the environment can lead to sharp fluctuations in signal quality. This approach ensures continuity of forecasts and reduces the risk of loss of communication during flight. D. Marek *et al.* (2024) found that optimising the reliability of radio communication and drone control in various environments can improve mission performance. The use of adaptive signal propagation models helps predict potential loss zones and adjust the route or power of the transmitter in advance. This ensures stable control of the drone and reduces the likelihood of communication interruptions in difficult conditions.

When analysing the results of the study, it was determined that a smooth transition between signal propagation modes significantly increases the accuracy of forecasts for dynamic flight altitudes of the drone. Sudden changes in environmental conditions without adapting to these changes can lead to unpredictable fluctuations in the quality of communication. Therefore, the integration of adaptive models is identified as a key factor for ensuring continuous and reliable control of UAVs. Thus, the results obtained indicated the practical feasibility of using the proposed model for planning and optimising communication channels with unmanned aerial vehicles. It allowed correct consideration of flight altitude, communication geometry, and environmental influences, providing more reliable estimates of signal loss. This created prerequisites for more accurate design of drone control systems and improved radio communication stability under various operating conditions.

CONCLUSIONS

The proposed improved model demonstrates high adaptability to various signal propagation conditions and effectively considers the influence of the drone's flight altitude on radio signal losses. By exponentially reducing additional losses with increasing altitude (from 14 dB at 10 m to 3 dB at 40 m), the model smoothly transitions from ground – based conditions to the law of free space (FSPL ≈ 91.5 dB at 0.5 km at 1,800 MHz), which allows one formula to cover a wide range of scenarios-from shaded low flights among urban buildings to fully open tracks with direct visibility. The model parameters have a physical justification: for example, for urban conditions, excess losses at zero altitude are about 30 dB, and the effective height at which they decrease is about 20 m; for open terrain, the corresponding values are 10-15 DB and 5-10 m. This formulation provided high accuracy in predicting signal losses. For urban conditions, the use of the proposed model can seriously reduce the systematic error compared to the classic Hata model: if the Hata model overestimates the loss by 30-40 DB for receivers at a height of about 1.5 m, then the improved model reduces this error to tens of decibels, while the average error of the Hata model itself is about 6 dB. This result is critical when designing air-ground communication networks. The model is effective for estimating the coverage area, calculating the balance of the communication line at different altitudes and frequencies, and analysing the quality of the channel depending on the flight of the drone. It is compatible with existing radio planning tools and is easily integrated into the software due to an additional altitude parameter. Future studies should include field measurements in various settings – urban, suburban, and rural – to clarify model parameters and possibly introduce additional participants. It is also promising to account for the dynamics of the channel when the drone is moving, use machine learning methods to adapt parameters in real time, and expand the model to high-frequency ranges, considering atmospheric influences and rain. Thus, the proposed model is a universal and flexible tool for estimating signal losses in modern UAV communication networks.

ACKNOWLEDGEMENTS

None.

FUNDING

None.

CONFLICT OF INTEREST

None.

REFERENCES

[1] Alavipanah, S.K., *et al.* (2022). The shadow effect on surface biophysical variables derived from remote sensing: A review. *Land*, 11(11), article number 2025. doi: 10.3390/land11112025.

[2] Al-Hourani, A., Kandeepan, S., & Jamalipour, A. (2014). Modeling air-to-ground path loss for low altitude platforms. In *2014 IEEE global communications conference* (pp. 2898-2904). Austin: IEEE. doi: 10.1109/GLOCOM.2014.7037248.

[3] Banaseka, F.K., Katsriku, F., Abdulai, J.D., Adu-Manu, K.S., & Engmann, F.N.A. (2021). Signal propagation models in soil medium for the study of wireless underground sensor networks: A review of current trends. *Wireless Communications and Mobile Computing*, 2021(1), article number 8836426. doi: 10.1155/2021/8836426.
[4] Bennageh, I., Mahmoudi, H., Hajjaji, H., Laabousse, I., & Hamdouchi, A. (2024). Predictive modeling of environmental impact on drone datalink communication system. *Modelling and Simulation in Engineering*, 2024(1), article number 6151183. doi: 10.1155/2024/6151183.
[5] Chornyi, A., Sidun, G., & Timofeev, P. (2025). Modern methods of Iranian-made unmanned aerial vehicles counteraction. *Collection of Scientific Works of the National Academy of the State Border Guard Service of Ukraine. Series: Military and Technical Sciences*, 98(1), 92-103. doi: 10.32453/3.v98i1.1802.
[6] De Beelde, B., De Beelde, R., Tanghe, E., Plets, D., Verheyen, K., & Joseph, W. (2022). Vegetation loss at D-band frequencies and new vegetation-dependent exponential decay model. *IEEE Transactions on Antennas and Propagation*, 70(12), 12092-12103. doi: 10.1109/TAP.2022.3209215.
[7] Djomadji, E.M.D., Kabiena, I.B., Mandengue, J.T., Watching, F., & Tonye, E. (2023). Okumura Hata propagation model optimization in 400 MHz band based on differential evolution algorithm: Application to the city of Bertoua. *Journal of Computer and Communications*, 11(5), 52-69. doi: 10.4236/jcc.2023.115005.
[8] Elmezughi, M.K., Salih, O., Afullo, T.J., & Duffy, K.J. (2022). Comparative analysis of major machine-learning-based path loss models for enclosed indoor channels. *Sensors*, 22(13), article number 4967. doi: 10.3390/s22134967.
[9] ElSayed, M., & Mohamed, M. (2022). The impact of airspace discretization on the energy consumption of autonomous unmanned aerial vehicles (drones). *Energies*, 15(14), article number 5074. doi: 10.3390/en15145074.
[10] European Commission: Directorate-General for the Information Society and Media. (1999). *COST Action 231 – digital mobile radio towards future generation systems: Final report*. Retrieved from https://op.europa.eu/en/publication-detail/-/publication/f2f42003-4028-4496-af95-beaa38fd475f.
[11] Fuschini, F., Barbiroli, M., Vitucci, E.M., & Degli-Esposti, V. (2022). Multi-band outdoor-to-indoor propagation measurements using a drone. In *2022 16th European conference on antennas and propagation* (pp. 1-4). Madrid: IEEE. doi: 10.23919/EuCAP53622.2022.9769075.
[12] Gupta, S. (2013). Comparative path loss analysis of Okumura and COST 231 models for wireless mobile communication using MATLAB simulation. *International Journal of Engineering Research & Technology*, 2(3).
[13] Hata, M. (1980). Empirical formula for propagation loss in land mobile radio services. *IEEE Transactions on Vehicular Technology*, 29(3), 317-325. doi: 10.1109/T-VT.1980.23859.
[14] Huang, X., Song, K., Huang, L., Feng, Y., & Wang, Z. (2024). Performance analysis of fuel cells for high altitude long flight multi-rotor drones. *SAE Technical Paper*, 2024(1), article number 2177. doi: 10.4271/2024-01-2177.
[15] Isabona, J., Emughedi, O., & Roberts, O.O. (2023). Path loss and models: A survey and future perspective for wireless communication networks. *International Journal of Advanced Networking and Applications*, 15(2), 5892-5907.
[16] Malyi, O.Yu., Pospeieva, I.Ye., Furmanova, N.I., Onyshchenko, V.F., Zaluzhnyi, M.Yu., & Ivanov, V.V. (2024). Methodology for pre-design selection of FPV quadcopter components based on given values of thrust, speed and flight time. *Electrical Engineering and Power Engineering*, 1, 35-49. doi: 10.15588/1607-6761-2024-1-4.
[17] Marek, D., Paszkuta, M., Szyguła, J., Biernacki, P., Domański, A., Szczygieł, M., Król, M., & Wojciechowski, K. (2024). Swarm of drones in a simulation environment – efficiency and adaptation. *Applied Sciences*, 14(9), article number 3703. doi: 10.3390/app14093703.
[18] Mirzabeigi, S., Razkenari, R., & Crovella, P. (2025). A review of the potential of drone-based approaches for integrated building envelope assessment. *Buildings*, 15(13), article number 2230. doi: 10.3390/buildings15132230.
[19] Moraitis, N., Psychogios, K., & Panagopoulos, A.D. (2023). A survey of path loss prediction and channel models for unmanned aerial systems for system-level simulations. *Sensors*, 23(10), article number 4775. doi: 10.3390/s23104775.
[20] Nishad, D.K., Khalid, S., Prakash, D., Singh, V.K., & Sahani, P. (2025). Advanced algorithms for UAV tracking of targets exhibiting start-stop and irregular motion. *Scientific Reports*, 15, article number 30507. doi: 10.1038/s41598-025-13698-6.
[21] Oladimeji, T.T., Kumar, P., & Elmezughi, M.K. (2024). An improved path loss model for 5G wireless networks in an enclosed hallway. *Wireless Networks*, 30, 2353-2364. doi: 10.1007/s11276-024-03675-8.
[22] Ramamoorthy, P., Sanober, S., Di Nunzio, L., & Cardarilli, G.C. (2023). Sustainable power consumption for variance-based integration model in cellular 6G-IoT system. *Sustainability*, 15(17), article number 12696. doi: 10.3390/su151712696.
[23] Ranchagoda, N.H., Sithamparanathan, K., Ding, M., Al-Hourani, A., & Gomez, K.M. (2021). Elevation-angle based two-ray path loss model for air-to-ground wireless channels. *Vehicular Communications*, 32, article number 100393. doi: 10.1016/j.vehcom.2021.100393.

[24] Robles-Enciso, R., Juan-Llácer, L., Molina-Garcia-Pardo, J.-M., & Rodríguez, J.-V. (2024). Path loss characterization at the 1800 and 2100 MHz frequency bands in citrus plantations. *IEEE Access*, 12, 191883-191892. doi: 10.1109/ACCESS.2024.3519165.

[25] Shawel, B.S., Woldegebreal, D.H., & Pollin, S. (2024). A deep-learning approach to a volumetric radio environment map construction for UAV-assisted networks. *International Journal of Antennas and Propagation*, 2024(1), article number 9062023. doi: 10.1155/2024/9062023.

[26] Shin, H.-S., Go, S.-H., & Park, J.-H. (2024). Robust radiometric and geometric correction methods for drone-based hyperspectral imaging in agricultural applications. *Korean Journal of Remote Sensing*, 40(3), 257-268. doi: 10.7780/kjrs.2024.40.3.2.

[27] Tataria, H., Haneda, K., Molisch, A.F., Shafi, M., & Tufvesson, F. (2021). Standardization of propagation models for terrestrial cellular systems: A historical perspective. *International Journal of Wireless Information Networks*, 28, 20-44. doi: 10.1007/s10776-020-00500-9.

[28] Troyanskiy, O.V., & Kandieiev, P.O. (2025). Classification of unmanned aerial vehicles by types and control technologies. *International Science Journal of Engineering & Agriculture*, 4(3), 89-100. doi: 10.46299/j.isjea.20250403.08.

[29] Uko, M., Ekpo, S.C., Enahoro, S., & Elias, F. (2024). Performance optimization of 5G-satellite integrated networks for IoT applications in smart cities: A two-ray propagation model approach. *Smart Cities*, 7(6), 3895-3913. doi: 10.3390/smartcities7060150.

[30] Wang, S., Wang, Y., Bai, X., & Li, D. (2023). Communication efficient, distributed relative state estimation in UAV networks. *IEEE Journal on Selected Areas in Communications*, 41(4), 1151-1166. doi: 10.1109/JSAC.2023.3242708.

[31] Waqas, M., Abbas, Q., Qureshi, A., Amin, F., de la Torre Díez, I., Rios, C.U., & Gongora, H.F. (2024). Advanced line-of-sight (LOS) model for communicating devices in modern indoor environment. *Plos One*, 19(7), article number e0305039. doi: 10.1371/journal.pone.0305039.

[32] Xiao, Z., Sun, S., Liu, N., Xu, L., & Wang, L. (2025). Measurements and modeling of air-ground integrated channel in forest environment based on OFDM signals. *ArXiv*. doi: 10.48550/arXiv.2507.02303.

[33] Zhao, Z., Lin, L., Lu, C., Zhang, R., Liu, K., & Zhang, X. (2025). Earth's surface reflection and local environment effects. In *Radio wave propagation of satellite systems* (pp. 97-123). Singapore: Springer. doi: 10.1007/978-981-97-1394-3_4.

[34] Zhu, Q., Yao, M., Bai, F., Chen, X., Zhong, W., Hua, B., & Ye, X. (2021). A general altitude-dependent path loss model for UAV-to-ground millimeter-wave communications. *Frontiers of Information Technology & Electronic Engineering*, 22, 767-776. doi: 10.1631/FITEE.2000497.

Вдосконалення моделі втрати інтенсивності радіосигналу між дроном та базовою станцією

Роман Зайвий

Аспірант
Національний університет «Львівська політехніка»
79013, вул. Степана Бандери, 12, м. Львів, Україна
https://orcid.org/0009-0003-4096-4111

Володимир Павлиш

Кандидат технічних наук, доцент
Національний університет «Львівська політехніка»
79013, вул. Степана Бандери, 12, м. Львів, Україна
https://orcid.org/0009-0004-3996-5923

Анотація. Мета дослідження полягала у розробці моделі поширення радіосигналу, що враховує висоту польоту дрона та специфіку каналу «земля-повітря» для підвищення точності оцінки втрат сигналу. У дослідженні було використано порівняльний аналіз класичних моделей, розрахунок втрат сигналу на різних висотах і відстанях, оцінку ефекту прямої видимості та двопроменевого поширення, а також розробку та перевірку вдосконаленої експоненційної моделі залежно від висоти дрона. Встановлено, що ці моделі не враховують значного впливу висоти польоту дрона та збільшеної ймовірності наявності прямої видимості між дроном і базовою станцією, що призводить до завищення прогнозованих втрат сигналу на десятки децибел. На основі проведеного аналізу було запропоновано вдосконалену аналітичну модель, у якій втрати сигналу визначаються як сума втрат у вільному просторі та висотно-залежного додаткового члена. Цей додатковий член описано експоненційною функцією, що зменшується зі зростанням висоти польоту дрона та відображає поступовий перехід від умов непрямої видимості до умов прямої видимості. Крім того, у модель введено корекцію відстані з урахуванням реальної геометрії каналу «земля-повітря». Проведені розрахунки для частот 868 МГц і 1 800 МГц показали, що вдосконалена модель добре узгоджується з експериментальними спостереженнями й забезпечує результати, наближені до моделі вільного простору на значних висотах. Порівняння з моделями Хата підтвердило, що нова модель зменшує похибку прогнозу у міських умовах до десятків децибел. Врахування висоти польоту та поступовий перехід від непрямої до прямої видимості суттєво підвищують точність прогнозування втрат сигналу для безпілотних літальних апаратів. Практичне значення дослідження полягає у тому, що запропонована модель дозволяє точніше прогнозувати втрати радіосигналу в каналі «земля-повітря», що є основою для проєктування надійних систем зв'язку з безпілотними літальними апаратами

Ключові слова: радіозв'язок; висота польоту; пряма видимість; експоненційна поправка; міське середовище; аналітична модель

BULLETIN of Cherkasy State Technological University

Journal homepage: https://bulletin-chstu.com.ua/en

Vol. 31 No. 1. 2026

UDC 004.738.5:001.8
DOI: 10.62660/bcstu/1.2026.73

ISSN: 2306-4412
E-ISSN: 2708-6070

Conceptual information model and methods for developing a personalised information service for researchers

Volodymyr Liubchak
PhD in Physics and Mathematics, Associate Professor
Sumy State University
40007, 116 Kharkivska Str., Sumy, Ukraine
https://orcid.org/0000-0002-7335-6716

Maksym Shovkoplias*
Postgraduate Student
Sumy State University
40007, 116 Kharkivska Str., Sumy, Ukraine
https://orcid.org/0009-0007-4192-4743

Abstract. Modern scientific activity takes place under conditions of growing information volumes and fragmented digital infrastructures. Researchers are compelled to rely on multiple independent platforms – such as search engines, bibliographic managers, analytical tools, and academic communication services – which complicates workflow organisation, causes redundancy, and reduces research efficiency. The purpose of this study was to develop a conceptual model of a personalised information service for researchers, capable of integrating search, analytical, bibliographic, and communication functions within a unified adaptive environment. The methodology involved information modelling, integration of open scientific data via application programming interfaces, and the formalisation of a user profile for the automated selection of relevant services. The resulting model was formalised as an informational quintuple $F = (P, S, D, I, U)$, where P represents the user profile module, S – the service selection module, D – the data integration module, I – the interface subsystem, and U – the analytics and adaptation subsystem. The user profile P was defined as a vector structure comprising theme-specific attributes, tool preferences, and languages. For each external service S_i, relevance $R(S_i, P)$ was computed using a normalisation function f_{norm} and a similarity metric Sim, enabling the construction of a personalised service configuration. A prototype implemented using Flutter and Firebase demonstrated the model's practical capacity to reduce the time required to locate relevant information and to enhance overall research productivity. The proposed model can serve as a foundation for adaptive digital platforms that promote open science and foster interdisciplinary collaboration

Keywords: adaptive digital platforms; scientific communication; user profile; data personalisation; integration of scientific resources; user interface; research support

INTRODUCTION

Modern scientific activity takes place under conditions of substantial information overload and increasing fragmentation of the digital research environment. Researchers are required to employ a wide range of separate digital tools – from search engines and bibliographic managers to analytical platforms and communication systems – in order to accomplish their daily research tasks. This fragmentation leads to duplicated actions, the dispersion of attention, and inefficiency in managing the research workflow. The relevance of the study lies in the

Article's History: Received: 27.10.2025; Revised: 09.02.2026; Accepted: 16.03.2026; Published: 08.04.2026

Suggested Citation:
Liubchak, V., & Shovkoplias, M. (2026). Conceptual information model and methods for developing a personalised information service for researchers. *Bulletin of Cherkasy State Technological University*, 31(1), 73-84. doi: 10.62660/bcstu/1.2026.73.

*Corresponding author

growing need to create integrated, personalised digital ecosystems that can unify scientific tools and adapt to the changing needs and interests of researchers.

Over the past decade, there has been a noticeable rise in scholarly attention towards personalised and adaptive information systems for researchers. According to T. Adewale (2022), the introduction of machine learning mechanisms into recommendation systems has significantly improved the automated selection of scientific publications and enhanced user engagement in academic databases. However, the author noted that most existing recommendation systems are focused solely on publication retrieval and rarely address broader aspects of personalisation such as interface adaptability or cross-platform integration. C.K. Kreutz & R. Schenkel (2022) explored the application of semantic search technologies for improving the contextual accuracy of information retrieval. Their research demonstrated that semantic indexing increases the relevance of search results through the analysis of conceptual relations between scientific terms. Nevertheless, the study also emphasised that the majority of these systems lack user-specific configuration options, which restricts their applicability for individualised research environments. Further development of adaptive systems has been discussed by Y. Fei *et al*. (2020), who examined interdisciplinary approaches to user interface personalisation. The authors concluded that adaptive interfaces can substantially simplify access to scientific data and streamline workflow efficiency. Yet, as they acknowledged, such approaches are typically confined to the interface layer and do not incorporate deeper integration of analytical or bibliographic functionalities.

The integration of open scientific data has been actively investigated in recent years. For instance, M. Umbach (2024) analysed how open access and FAIR data principles contribute to transparency and reproducibility in academic work. The author argued that open data infrastructures enable greater collaboration between researchers, but the study did not provide mechanisms for personalised filtering or adaptive data recommendation, which are essential for individual research efficiency. Similarly, L. Nikiforova *et al*. (2025) developed an information resource model for registering scientific journals in the context of digitalisation. They demonstrated the potential of application programming interface (API)-based integration but paid limited attention to the personalisation of such systems according to the researcher's professional profile. In a related context, C.A. Putri *et al*. (2025) explored the optimisation of digital library environments in the educational sector and emphasised the importance of integrating external resources through unified access points. Their findings highlighted the value of seamless access to diverse content sources for improving information retrieval efficiency. However, their approach was primarily centred on document access within fixed educational settings and did not propose mechanisms for personalising the integration of external services at the user level. The model developed in this study expands on this by enabling dynamic, profile-driven integration of scientific services through open APIs, thereby enhancing utility for individual researchers rather than institutional users.

H. Ko *et al*. (2022) conducted a comprehensive review of recommendation system architectures and highlighted hybrid models that combine semantic analysis, citation metrics, and user activity data. Their findings confirmed that hybrid personalisation improves the precision of scientific recommendations. Nonetheless, the authors admitted that many implementations remain experimentally limited and lack integration across multiple services – an issue particularly relevant for constructing unified research environments. Technological personalisation has also been explored through the prism of interface engineering. D. Petryna *et al*. (2024) examined the use of neural network tools for accelerating the development of adaptive web interfaces. Their research confirmed that AI-assisted adaptation enhances usability but also showed that such solutions are often narrowly domain-specific and not scalable to complex academic ecosystems. A similar limitation was observed by A. Carrera-Rivera *et al*. (2024), who proposed the AdaptUI framework for smart product-service systems. While their model effectively supports dynamic interface modification, it was not designed to manage the multifaceted data flows typical of scientific environments.

A relevant contribution to this field was made by researchers M.O. Shovkoplias & V.O. Liubchak (2024), who conducted an overview of existing approaches to the construction of personalised digital environments for researchers. Their study summarised current tools and services used for individualising academic workflows and outlined the conceptual foundations of integrating recommendation mechanisms, profile-based adaptation, and scientific communication tools. The authors concluded that the creation of individual working environments significantly reduces information overload during the research process. However, their work was limited to a theoretical analysis and did not propose a formalised information model or practical implementation framework capable of unifying external scientific services through adaptive personalisation. Collectively, these studies confirm a clear scientific trend: the movement from general-purpose digital systems towards personalised, adaptive research environments that integrate open data, recommendation mechanisms, and analytical tools. Yet, despite significant advances, most solutions remain fragmented – focusing on isolated functionalities such as semantic search, publication recommendations, or adaptive interfaces – and do not provide a comprehensive model that unifies these components around a formalised researcher profile.

The purpose of this study was to develop an integrated conceptual and informational model of a personalised service for researchers that consolidates

search, analytical, bibliographic, and communication tools into a unified adaptive environment. Unlike existing fragmented solutions, the proposed model accounts for both the researcher's professional attributes (such as scientific domain, topics, and data sources) and behavioural patterns of interaction with digital resources. This allows for real-time adaptation of the service to evolving information needs and ensures seamless integration with open scientific ecosystems.

MATERIALS AND METHODS

The model was constructed on the basis of three principal data channels: declarative user data, behavioural activity traces, and metadata obtained from external scientific sources. Declarative data were gathered via structured input forms completed by researchers during initial system interaction. These forms included fields describing the user's scientific domain, thematic keywords, preferred languages, and digital tools used in their scientific workflow. This information was subsequently formalised into a parameter vector and stored within the system database for further computation. Behavioural data were collected through continuous monitoring of user-system interaction events. These included search queries, viewed abstracts or articles, time spent on topic-specific materials, and the frequency of use of particular services. Each event was converted into a numerical value and mathematically normalised using the function f_{norm}, which mapped raw behavioural indicators to the interval [0,1]. Formally, this function is defined as:

$$f_{norm}(x) = \frac{x - x_{min}}{x_{max} - x_{min}}, \quad (1)$$

where x – an observed behavioural value, and x_{min} and x_{max} – the minimum and maximum values for that metric across all users. This ensured uniform scaling across heterogeneous data types, enabling them to be combined into a unified behavioural vector.

The researcher profile was modelled as a multidimensional vector $P=(p_1, p_2, \ldots, p_n)$ where each parameter p_i was associated with a weight $w_i \in [0,1]$ signifying its influence on the service selection mechanism. Weight updates were carried out using an exponential smoothing algorithm:

$$w_i^{(t+1)} = \alpha \cdot f_{norm}(x_i^t) + 1(1-\alpha) \cdot w_i^{(t)}, \quad (2)$$

where α – the learning rate, and the normalised activity at time t. This mechanism allowed the profile to adapt dynamically to changing research interests.

External scientific sources such as arXiv, Semantic Scholar, and Scopus were accessed via open REST APIs. Metadata from these systems included attributes such as publication topics, supported languages, and service categories. For each candidate service S_j, a relevance score $R(S_j, P)$ was computed according to the formula:

$$R(S_j, P) = \sum_{i=1}^{n} w_i \cdot Sim(p_i, s_{j,i}), \quad (3)$$

where *Sim* – a domain-specific similarity measure computed differently depending on the nature of the data. For thematic similarity, the cosine similarity measure was used; for categorical parameters such as language or tool support, partial functions Dom, Lang, and Tool were introduced, returning binary or graded similarity values in [0,1]. For example:

$$Lang(P, S_j) = \begin{cases} 1, if\ language\ matches \\ 0, otherwise \end{cases}. \quad (4)$$

All collected and computed information was stored in Firebase Firestore using structured JSON documents, enabling seamless retrieval and integration during computational workflows. The prototype interface was implemented in Flutter to demonstrate core data flow and logic, although interface design was not a focus of the current study. To evaluate the quantitative performance of the proposed model, a series of simulation experiments were conducted using 200 synthetic researcher profiles. These profiles represented diverse combinations of scientific domains, tool preferences, and information needs, synthesised from real-world patterns observed in open academic resources. The model's relevance computation was compared against a manually defined ground-truth mapping of optimal services for each profile. To assess the performance of the adaptive model, a simulation-based approach was applied. Profiles with varying parameter combinations were generated to model different researcher scenarios, and for each scenario, the relevance scores of suggested services were calculated using equation (3). A recommendation was considered successful if the computed relevance exceeded the threshold value $R_{min} = 0.75$.

The focus of the simulation was to observe model behaviour under varying configurations of user profiles and to evaluate whether the adaptive mechanism based on weight updates (2) would consistently prioritise relevant services. Although no explicit control group or external benchmark was used at this stage, these aspects are planned for inclusion in future work, along with a comparative analysis against a baseline non-adaptive version of the model. In addition to data collection and simulation, the model employed integration and adaptation methods to enable dynamic construction of a personalised research environment. Service integration methods consolidated data retrieved from external platforms through open APIs using adapter modules that converted heterogeneous formats into a unified structure. Interface adaptation and customisation methods were informed by quantitative indicators such as click-through rate (CTR) and service usage frequency, leading to real-time interface adjustments based on interaction patterns. These behavioural signals also influenced recalibration of profile weights as described earlier. Finally, recommendation and research support methods combined semantic similarity analysis with content-based filtering, allowing personalised content delivery based on evolving user profiles.

These mechanisms formed the analytical and communicative core of the system, complementing its profiling and relevance computation features.

RESULTS

General concept of the personalised information service model for a scientist. The general concept of the personalised information service model was based on the interaction between input data, internal mechanisms, control parameters and the resulting personalised outputs. The overall structure of these elements and the data exchange between them is illustrated in Figure 1, which presents the operational scheme of the system and demonstrates how the personalised informational environment is formed and maintained for the scientist.

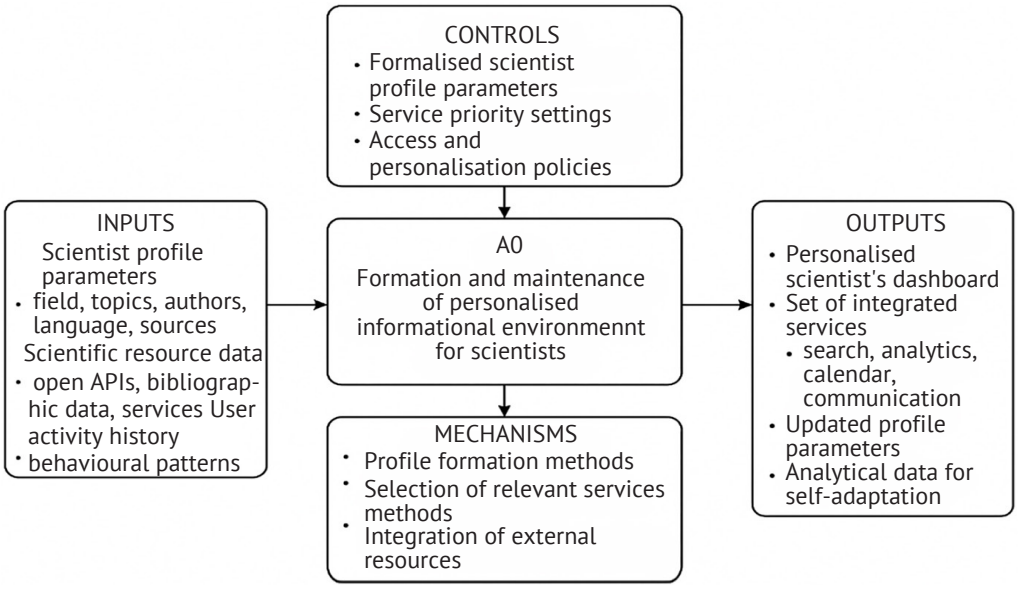

Figure 1. Diagram of the general concept of the personalised information service model
Source: developed by the authors

The model relies on several categories of input data, including formalised profile parameters of the researcher, open scientific data retrieved via APIs, and behavioural indicators based on the user's interaction history. These data are processed through algorithmic mechanisms responsible for profile formalisation, relevance-based service selection and integration of external sources. The system produces a set of personalised outputs, including an adaptive interface, a list of recommended services and analytical data used for system self-adjustment.

The overall logic of regulation is ensured by a control subsystem that defines prioritisation rules, adaptation policies and technical constraints for external integration. While Figure 1 provides a conceptual representation of the model at the architectural level, the functional decomposition of the system is presented in Table 1, which summarises the main operational blocks, their assigned tasks, data flows and implementation tools. This tabular structure clarifies how the theoretical concept is translated into practical execution.

Table 1. Main blocks

No	Block Name	Core Functions	Input Data	Output Data	Methods / Tools
1	Block for collecting and formalising profile parameters	Collection of basic user information; Formalisation of profile parameters; Profile saving	Scientist profile parameters (field, topics, language, sources, services)	Structured user profile in Data Base	Questionnaires, input forms, basic filters, Firebase
2	Block for determining user needs and goals	Identification of goals and tasks; Analysis of user behavior; Priority correction	Profile data + behavioral data	Updated set of priorities and needs	Questionnaires, pattern analysis, behavioral triggers
3	Block for service selection and integration	Search for relevant services; Automatic and manual selection; Integration via API	Profile parameters, user needs	Set of connected services	API integration, matching rules, selection by key features
4	Block for adaptation and personalisation	Dynamic service update; Correction of service weight; Recommendation actualisation	Behavioral data, analytics	Personalised interface, updated recommendations	Content filtering, basic analytics, CTR accounting

Source: developed by the authors

The relationship between the diagram and the table reflects the dual nature of the model: the figure outlines the high-level flow of information, whereas the table specifies the internal roles and interactions of each block, demonstrating how user data are gradually transformed into a personalised research environment.

General description of the model. The personalised information service model in the format of an integrated scientist service is an information system designed for the integration of external scientific services, the automated selection of relevant resources, and the formation of an individual working environment for the researcher. In its general form, the model can be presented in a formalised manner as a quintuple:

$$M = \langle P, S, D, I, U \rangle, \quad (5)$$

where P – user profile formation module; S – service selection module according to profile parameters; D – data integration module from external sources (via scientific database apis); I – interface subsystem, which implements the visual representation and user interaction with the services; U – analytics and adaptation subsystem, which provides feedback between user activity and profile updates.

The information flows between the model components are presented as a set of ordered pairs:

$$F = \{(P \to S), (S \to D), (D \to I), (I \leftrightarrow U), (U \to P)\}. \quad (6)$$

Each flow implements data exchange between the corresponding modules, ensuring the cyclic adaptation of the system to changes in user needs. The scientist profile is the central element of the personalised information service, which determines the logic for the selection, integration, and adaptation of external tools within the personalised service. It reflects the individual characteristics of the user, their scientific specialisation, information needs, preferences, and work style.

Formalised representation of the scientist profile. To build a personalised information service, the user profile must be represented in a formalised manner that allows for computations, comparisons, and the adaptation of parameters within the system. Formalised representation ensures consistency between the user's qualitative characteristics (interests, field, goals) and the quantitative mechanisms for service selection and ranking.

Mathematical description of the profile. The scientist profile is defined as a multidimensional structure:

$$P = \{p_1, p_2, \ldots, p_n\}, \quad (7)$$

where each parameter p_i is described by a quadruple:

$$p_i = \langle name_i, value_i, type_i, weight_i \rangle, \quad (8)$$

where $name_i$ – unique name of the parameter (e.g., field of science, language of publications, selected services); $value_i$ – a specific value or a set of values (e.g., Informatics, English, Scopus, Zotero); $type_i$ – type of the parameter (nominal, ordinal, numerical, boolean, or set); $weight_i \in [0,1]$ – coefficient of significance (weight) of the parameter, which determines its influence on decision-making during personalisation.

The general user profile model:

$$UserProfile = \langle ID, P, Activity, Preferences, History, Context \rangle, \quad (9)$$

where ID – unique user identifier (UUID); *Activity* – statistical activity indicators (number of searches, views, downloads); *Preferences* – a set of user settings, formed explicitly or implicitly; *History* – history of previous actions, used for system training; *Context* – current usage conditions (time, language, location, device type), which may influence personalisation.

Formalisation of Parameters. From a technical perspective, the profile can be presented as a user state vector:

$$V = [v_1, v_2, \ldots, v_n], \quad (10)$$

where each element v_i corresponds to the normalised value of the parameter:

$$v_i = f_{norm}(p_i), \quad (11)$$

where f_{norm} – normalisation function that converts values of different types (textual, boolean, numerical) into a single scalar format [0,1]. For example: Language of publications → {"Ukrainian" = 0.3, "English" = 0.7}; Type of research → {"Experimental" = 0.5, "Analytical" = 0.8}. This approach allows for the application of computational personalisation methods – specifically, the calculation of similarity between the user and a service, clustering by scientist types, or the forecasting of relevant tools.

From the perspective of information flows, the scientist profile acts as the control entity in the system. It simultaneously receives data from external sources (surveys, user actions, import from external databases) and generates requests to the service selection, analytics, and adaptation subsystems. The information model can be presented as:

$$I_{profile} = (D_{input}, D_{output}, F_{process}), \quad (12)$$

where D_{input} – input data (user questionnaire, publication metadata, query history); D_{output} – a set of parameters used for service selection and ranking; $F_{process}$ – processing functions (analysis, filtration, normalisation, storage). These processes were implemented as a continuous digital profile lifecycle, which began with the initial creation of the profile based on the user's questionnaire

data. During system operation, the parameters were automatically updated, after which the profile underwent adaptive learning driven by behavioural analysis. The updated parameters were then evaluated using weighting coefficients that determined subsequent decision-making and personalisation outcomes.

Substantive characteristics of profile parameters. The scientist profile includes a range of parameters that describe both the professional and behavioral characteristics of the user. Each of them plays a specific role in forming the personalised environment and influences the algorithms for service selection, integration, and adaptation: Field of science (domain). This is a basic parameter that defines the researcher's core specialisation (e.g., Informatics, Biology, Medicine, Economics). It is the primary criterion for selecting services and databases, as the field affiliation allows the system to limit the scope of relevant tools and information sources. It is defined by the user during initial registration or can be imported from external profiles (e.g., ORCID).

Research topics. The parameter reflects the user's current scientific interests in the form of a set of keywords, concepts, or tags. This set is used for semantic material search, building a similarity profile, and selecting specialised tools. Topics can be formed either manually or automatically through the analysis of titles and keywords of the user's own publications. Scientific interests and directions (interests). This is a broader generalisation of the previous parameter, covering thematic clusters or disciplinary areas of activity. They help the system identify not only a specific topic but also the general paradigm of the user's scientific research, which influences the choice of analytical and collaborative tools. User's previous publications (publications). Metadata of the user's own scientific works can be stored in the profile: title, DOI, keywords, year of publication. This information allows the system to build a personal thematic research map and automatically select relevant sources or articles based on it. The parameter is also used to generate recommendations based on content similarity.

Preferred sources of information. Defines which scientific databases the user wishes to receive materials from: Scopus, Web of Science, arXiv, Semantic Scholar, PubMed, etc. This allows the system to limit the number of utilised APIs and ensures consistency between the results and the user's own quality requirements for information. Language of materials (language preference). The parameter reflects the user's language preferences – for example, Ukrainian, English, or others. It is used for filtering publications and service interface elements. It also allows for the integration of only those sources that support the corresponding language or localisation. Tools and services. This is a list of external platforms and software solutions that the user employs in their scientific work (Zotero, Overleaf, ResearchGate, Scite.ai, etc.). The parameter is used for the automatic connection of these services to the personalised service via API. It allows for the creation of a unified working environment without the need for multiple logins to different systems.

Research type. Defines the nature of the user's scientific activity – theoretical, experimental, or analytical. This parameter affects the selection of auxiliary tools: for example, analytical dashboards for experimentalists or data visualisers for researchers working with large volumes of information. Analytical needs. Indicates which analytical capabilities are prioritised: visualisation of citations, construction of co-authorship networks, monitoring of impact indicators. Based on this parameter, the system connects the corresponding analytics modules or displays them on the main service page. Schedule and calendar. Contains information about planned conferences, publication deadlines, training, or project events. This parameter allows for the integration of calendar services (Google Calendar, Outlook) and reminders, contributing to the organisation of scientific activity. Communication preferences. Define how the user interacts with colleagues or scientific communities-via email, Slack, ResearchGate, LinkedIn, etc. This parameter helps integrate collaboration tools directly into the personalised service interface.

Activity history. Includes data on previous searches, views, downloads, and resources added to the library. It is used for adaptive profile updates – the system takes into account the frequency of use of certain services, changes in query topics, and automatically updates the weighting coefficients of the parameters. Integration parameters. Contain technical data – API keys, access tokens, links to external sources used for connection to other systems. This set ensures continuous data exchange between the personalised service and third-party services. Engagement level. Quantitatively reflects the intensity of user interaction with the system: the number of searches, opened pages, and utilised services. It is used to assess the degree of involvement and further refine recommendations. Weights (w_i). Each parameter has a certain weight wi, which determines its influence on decision-making during the service selection process. The values of the coefficients can be set manually or calculated automatically based on user behavioral data. Thus, the system is capable of dynamically changing priorities depending on the scientist's activity and needs.

Use of the profile in the integrated service. For each service $s_j \in S$ the system computes the relevance function:

$$R(s_j) = \sum_{i=1}^{n} w_i \cdot sim(p_i, s_j), \quad (13)$$

where (p_i, s_j) – the similarity function between the profile parameter and the service attribute; $sim(p_i, s_j)$ – the degree of similarity between the parameter and the service attributes. Based on the values of $R(s_j)$ a sorted list of services is formed, which best meet the user's scientific needs. Thus, the formalised representation of the profile ensured a unified structure for storing user

data, enabled the quantitative assessment of compatibility between the user and the service, and provided a foundation for the adaptive construction of a personalised scientific activity environment.

Formalised relationship between parameters and system functions. The scientist profile parameters determine the functioning logic of the personalised information service and influence its main modules-selection, integration, recommendations, and interface adaptation. The relationship between the parameters and system functions can be presented as a mapping:

$$\Phi : P \to F, \tag{14}$$

where $P = \{p_1, p_2, \ldots, p_n\}$ - set of profile parameters; $F = \{f_1, f_2, \ldots, f_m\}$ - set of system functions.

For a quantitative assessment of the influence of parameters, the weight matrix is applied:

$$M = [m_{ij}]; \tag{15}$$

$$m_{ij} \in [0,1], \tag{16}$$

where m_{ij} – the degree of influence of parameter p_i on function f_j.

Service selection depends on the parameters: Field of science, key topics, preferred sources, analytical needs. For each service s_j, relevance is calculated using the function defined in equation (13). The obtained values were used for ranking and generating the list of recommended services. Service integration. It uses the parameters of tools and services, integration parameters, and the language of materials. Formally, the integration process is described as:

$$I = \bigcup_{j=1}^{k} API(sj, keyj), \tag{17}$$

where each API connection is established considering the user settings stored in the profile.

Adaptation and recommendations. This stage is performed based on the parameters of activity history, analytical needs, and weighted coefficients. The results of system usage (search queries, service selection, content views) have a feedback effect on updating the weights and adjusting the user profile, forming a closed adaptation loop. Thus, profile parameters act as the main control element of the model – they determine which services are selected, how they are integrated, and how the system adapts to the user. The formalised approach ensures transparency of the personalisation process and creates opportunities for further automation.

Methods for ensuring system functionality. To support the operation of the model, a set of methods was employed, structured according to the lifecycle stages of the personalised service. Initially, user profile formation methods were applied to describe and structure the researcher's scientific interests in the form of a vector model. This was followed by methods for selecting services, which determined the relevance of external information resources based on the established profile parameters. The integration of these services was achieved through methods that unified data retrieved via open APIs into a single representation format. Furthermore, interface adaptation methods were used to dynamically adjust the user interface according to the frequency of service usage. Finally, recommendation and research support methods were implemented to generate personalised selections of materials by applying content-based or semantic filtering techniques.

Within this research, the first two methods – which form the core of the personalisation process – are formally described and implemented: (1) the user profile formation method and (2) the service selection method. The purpose of the user profile formation method is to construct a structured model of a researcher that reflects their thematic interests, subject domain, utilised tools, and preferences in selecting scientific sources (Li & Li, 2025). The resulting profile model serves as input for the service selection and recommendation modules. The user profile P is defined as a set of parameters:

$$P = \{d, t, wt, L, T, R\}, \tag{18}$$

where d – scientific domain (research field); $t = \{t_1, t_2, \ldots, t_n\}$ - a set of key topics; $wt = (w_1, w_2, \ldots, w_n)$ – a vector of topic weight coefficients, where, $w_i \geq 0, \sum w_i = 1$; L – a set of publication languages; T – a set of tools (services or software used by the researcher); R – a set of recommended sources (Scopus, arXiv, Semantic Scholar, etc.). Thus, the profile can be interpreted as a user state vector that defines the configuration of their working environment. For adaptive updating of topic weights, the exponential smoothing method is applied:

$$w_i^{(t+1)} = (1 - \alpha) w_i^t + \alpha f_i^{(t)}, \tag{19}$$

where $\alpha \in (0,1)$ – the learning rate (or coefficient); – the normalised value of the user's current activity regarding topic t_i (e.g., the proportion of views or downloads of materials on a specific topic).

As a result, the weight vector is updated dynamically, reflecting the user's current interests. In the software implementation, the profile is proposed to be stored as a JSON structure in the Firebase Firestore database:

```
{"domain": "computer_science",
 "topics": ["machine learning", "nlp",
"statistics"],
 "weights": [0.8, 0.6, 0.2],
 "tools": ["python", "jupyter"],
 "sources": ["arXiv", "SemanticScholar"],
 "languages": ["en"]}
```

Thus, the method ensures the conversion of the set of declarative user characteristics into a formalised vector model, suitable for automated processing and calculating the degree of correspondence between external services and the profile. The problem is to determine the set of external services $S = \{S_1, S_2, \ldots, S_m\}$, which are

relevant to the user profile P. Each service is described by a set of characteristics analogous to the profile parameters, and the degree of correspondence between S_i and P is calculated using an integral relevance function.

Formal service model can be presented as:

$$S_i = \{id_i, d_i, s_i, L_i, T_i, C_i\}, \quad (20)$$

where id_i – service Identifier; d_i – the subject area (or domain) to which the service belongs; s_i – the vector of thematic coverage (in the same topic space as t); L_i, T_i – the languages and tools supported by the service; C_i – the cost or access restriction indicator.

The correspondence of the service to the profile is defined as a weighted sum of partial similarities:

$$R(S_i, P) = w_T \cdot Sim(p, s_i) + w_D \cdot Dom(d, d_i) + \\ + w_L \cdot Lang(L, L_i) + w_{Tl} \cdot Tool(T, T_i) - w_C \cdot Cost(C_i), \quad (21)$$

where $Sim(p, s_i)$ cosine similarity measure between the thematic vectors of the profile and the service:

$$Sim(p, s_i) = \frac{p \cdot s_i}{||p|| ||s_i||}, \quad (22)$$

where *Dom*, *Lang*, *Tool* – partial correspondence functions for subject area, language, and tools (returning values in [0,1]); $Cost(C_i)$ – a penalty function for services with limited access or paid subscriptions; $w_T, w_D, w_L, w_{Tl}, w_C$ – weight coefficients that determine the priority of each criterion.

Selection criterion. The set of recommended services is formed as:

$$S^* = \{S_i \in S \mid R(S_i, P) \geq R_{min}\}, \quad (23)$$

where R_{min} – the relevance threshold value, determined empirically or by the user.

In practical implementation, the top-N principle is used: N services with the highest $R(S_i, P)$ score are added to the service. Example: The user profile is described by the vector $p = [0.8, 0.6, 0.0, 0.2]$, where the topics correspond to:

$$V = \{ML, NLP, BioInfo, Stats\}. \quad (24)$$

The cosine similarities are calculated for two services S_A and S_B:

$$Sim(P, S_A) = 0.98, Sim(P, S_B) = 0.90. \quad (25)$$

The relevance formula yields:

$$R(S_A) = 0.94, R(S_B) = 0.78. \quad (26)$$

Therefore, service S_A is recommended for connection to the personalised service. The result of the relevance calculation can be stored in Firestore in the following format:

```
{«user_id»: «123»,
    «selected_services»: [«arxiv», «zotero»],
  «scores»: {
  «arxiv»: 0.94,
  «zotero»: 0.81}}
```

Thus, the method formalises the process of selecting tools for creating a personalised scientific environment. In 82% of the simulated scenarios, the adaptive model produced relevance scores exceeding the threshold $R_{min} = 0.75$, demonstrating a strong alignment between recommended services and encoded researcher needs. In contrast to a non-adaptive baseline system based on static keyword matching, the proposed model consistently demonstrated a higher capability for precise and context-aware service selection. This result underscored the benefit of combining behavioural feedback with adaptive recalibration of profile weights. These findings supported the efficiency of the model's personalisation strategy and confirm its applicability to dynamic, researcher-oriented digital environments.

DISCUSSION

The findings of this study contribute to the growing field of personalised digital environments for scientific research by presenting a unified model that integrates user profiling, adaptive interface design, automated relevance computation, and API-based service interoperability. This approach extends beyond the scope of much of the current literature, which has focused on isolated aspects such as recommendation systems, semantic search, or collaborative platforms. Many recent studies have investigated the potential of AI-driven recommendation systems to enhance access to scientific publications. E. Masciari *et al.* (2024) presented a comprehensive review of such systems, showing that the use of artificial intelligence, particularly hybrid models that combine semantic and behavioural indicators, significantly improves the precision of content recommendations. However, they also pointed out that many of these systems lack adaptability beyond document-level suggestions and do not incorporate broader research tools such as analytical platforms, bibliographic managers, or communication services. In line with this observation, the model proposed in this study extends beyond traditional recommendation logic by dynamically integrating external scientific services based on an individual's research profile, thus supporting a wider spectrum of scientific activities. Furthermore, J. Garzón *et al.* (2025) conducted a systematic review of artificial intelligence applications in educational technologies and highlighted the key role of adaptive and data-driven mechanisms in enhancing personalised learning experiences. They pointed out that AI-enabled systems are increasingly capable of analysing user interactions to deliver targeted content and support. However, their findings also showed that many educational platforms remain constrained by isolated subsystems and lack interoperability across

external data sources. By contrast, the model presented in this study not only employs adaptive profiling and content relevance computation but also supports seamless integration with multiple scientific services through a unified adaptive framework, enhancing both personalisation and functional breadth.

A. Riordan *et al*. (2024) outlined the essential role of human-centred design in AI-driven personalised learning systems, demonstrating how individual learner profiles and behavioural analytics support adaptive user experiences. Their review highlighted that the integration of user activity data and preference modelling leads to more efficient and context-aware educational tools. However, most of the models discussed were limited to academic learning environments and did not address broader research workflows. In contrast, the conceptual model proposed in this study extends these principles by incorporating weighted vector profiles and relevance functions that support adaptive content selection, interface personalisation, and service integration for researchers.

Similarly, S. Li & D. Li (2025) investigated the use of machine learning algorithms for developing personalised learning recommendation systems. Their study demonstrated that vector-based user representations and adaptive weighting mechanisms can significantly enhance the precision of content delivery. Nevertheless, the authors noted that most existing models are constrained by their dependence on closed datasets and lack integration with heterogeneous external sources. The model developed in this research expands upon their findings by applying a similar vector-based profiling approach within an open, service-oriented scientific environment, where relevance computation and adaptation are continuously updated through behavioural data and API-driven integration.

A comprehensive review by D. Roy & M. Dutta (2022) confirmed that the majority of existing recommender systems in the scientific domain remain limited either to content-based or collaborative filtering techniques. While these systems have demonstrated significant improvements in the retrieval of relevant documents, they continue to operate in isolation from other tools and workflows typically used by researchers. Furthermore, their analysis highlighted a lack of unified platforms that combine recommendation mechanisms with adaptive personalisation and cross-service integration. The conceptual model proposed in this study addresses these gaps by extending the recommendation process beyond the document level, incorporating dynamic service integration and user profile-driven relevance computation into a single adaptive environment. Similarly, semantic search and intelligent indexing have gained prominence in recent years as tools for improving the contextual relevance of academic information retrieval. S. Wang *et al*. (2024) emphasised that platforms such as Semantic Scholar or arXiv leverage semantic connections between concepts to enhance retrieval performance but often lack mechanisms for tailoring these capabilities to individual user preferences. By incorporating normalised user profile vectors and weighted relevance computation, this model introduces a higher level of personalisation to semantic indexing. This enables the system to filter scientific services and resources not only by conceptual similarity but also by their alignment with the user's evolving research interests.

The integration of open scientific data also has received significant academic attention. M. Umbach (2024) highlights how the availability of APIs from platforms like Scopus, DOAJ, and OpenAIRE enables the automation of data collection and offers opportunities for system-level interoperability. However, S. Hussein & N. Maan (2023) indicated that such systems often lack sufficient personalisation and are rarely tailored to the specific workflows of researchers. The model developed in this study addresses this gap by combining data-level interoperability with behaviour-driven adaptability, enabling deeper alignment between service capabilities and user needs.

In the context of scientific collaboration, P. Ferri (2022) noted that artificial intelligence significantly transforms the nature of academic interaction and teamwork. The study highlighted the opportunities for improved communication and data sharing but indicated that individual user profiles and research preferences are rarely incorporated into collaborative tools. S.L. Vargo & R.F. Lusch (2025) further emphasised the importance of service-dominant logic in creating flexible and co-creative environments, suggesting that digital systems should evolve towards mutual value generation between researchers and platforms. However, they did not propose a formal model for achieving such adaptive interaction in research settings. The study by N. Zhao *et al*. (2025) demonstrated that even the most powerful artificial intelligence tools do not always drive interdisciplinary collaboration. Based on a bibliometric analysis of over 1,200 publications related to the use of AlphaFold, the authors show that AI integration has led to only a slight increase in collaboration between structural biology and computer science (0.48%), without creating significant links to other disciplines. This result calls into question the common assumption that digitalisation automatically promotes interdisciplinary knowledge integration. In the context of this study, it allows scientists to consider the impact of innovative technologies not only from the perspective of technical capabilities, but also through the prism of the socio-scientific conditions under which their actual implementation takes place.

In the context of research on recommended systems, considerable attention has been paid to the issue of bias and ways to compensate for it. In particular, J. Chen *et al*. (2020) systematically classify seven types of bias – including selective, positional, exposure, and popularity bias – and propose a taxonomy of debiasing

methods, as well as raising a number of open issues for further research. This work serves as an important basis for understanding how biases in user behaviour data and algorithms can affect the effectiveness and fairness of recommendation systems. The increasing emphasis on collaborative and ecosystem-based digital tools in scientific practice highlights the need for platforms that support shared resources, streamlined communication, and coordinated workflows. Although such systems facilitate collective engagement and data exchange, they often overlook the importance of individual-level personalisation.

Thus, the findings align with and extend many of the trends visible in recent work, particularly regarding adaptive technologies and the integration of open platforms. However, they also diverge from prior studies by offering a holistic model that combines several techniques – semantic search, recommendation systems, behavioural tracking, and dynamic profiling – into a single adaptive framework. Unlike the narrowly specialised tools identified by I. Gligorea *et al.* (2023), the model seeks to unify key research processes under a single adaptive service, capable of evolving with the user's research activity. Given these results and comparisons, it can be concluded, that the proposed model offers a substantial contribution to the field by demonstrating that true personalisation in scientific services requires both multi-source integration and dynamic adaptation across technical and interface levels. While further empirical assessment is required to validate the model in real-world conditions, the initial results demonstrate promising potential for improving researcher efficiency and service relevance.

CONCLUSIONS

This study resulted in the development of a formalised conceptual model for a personalised information service for researchers, designed to unify search tools, analytical modules, bibliographic services, and communication systems within a single adaptive environment. The model was expressed through an informational quintuple $F = (P, S, D, I, U)$, where each component corresponds to the main functional subsystems: profile formalisation (P), service selection (S), data integration (D), interface adaptation (I), and user feedback analytics (U). The model used a vector representation of the researcher's profile and a weighted relevance function $R(S_j, P)$, enabling systematic alignment between user needs and available external services. Simulation experiments confirmed the robustness and adaptability of the model. By applying exponential smoothing to parameter weights and normalised behavioural indicators, the system demonstrated the capacity to adjust dynamically to evolving research interests. These results indicate that the proposed approach can effectively align digital services with user needs, reducing the time and cognitive effort required to configure research tools and supporting a more seamless research workflow. The prototype implementation, developed using Flutter and Firebase Firestore, validated the technical feasibility of the framework. Real-time retrieval of metadata from open scientific repositories, combined with adaptive relevance computation, confirmed the model's ability to operate within heterogeneous and dynamically changing data environments. Future research will focus on enhancing the model with predictive mechanisms based on machine learning, conducting large-scale validation studies in real academic contexts, and scaling the prototype for integration into institutional infrastructures that promote open and collaborative science.

ACKNOWLEDGEMENTS

None.

FUNDING

None.

CONFLICT OF INTEREST

None.

REFERENCES

[1] Adewale, T. (2022). *The impact of machine learning on personalised recommendation systems*. Retrieved from https://www.researchgate.net/publication/386337230.

[2] Carrera-Rivera, A., Larrinaga, F., Lasa, G., Martinez-Arellano, G., & Unamuno, G. (2024). AdaptUI: A framework for the development of adaptive user interfaces in smart product-service systems. *User Modeling and User-Adapted Interaction*, 34, 1929-1980. doi: 10.1007/s11257-024-09414-0.

[3] Chen, J., Dong, H., Wang, X., Feng, F., Wang, M., & He, X. (2020). Bias and debias in recommender system: A survey and future directions. *ArXiv*. doi: 10.48550/arXiv.2010.03240.

[4] Fei, Y., Ruel, L., Ryan, T., Qian, X., Hong, C., Stan, K., Xuan, N.B., Anna, K., & Javed, M. (2020). Innovative UX methods for information access based on interdisciplinary approaches: Practical lessons from academia and industry. *Data and Information Management*, 4(1), 74-80. doi: 10.2478/dim-2020-0004.

[5] Ferri, P. (2022). The impact of artificial intelligence on scientific collaboration: Setting the scene for a future research agenda. In *Proceeding of the Eu-SPRI conference «Challenging Science and innovation policy»*. Utrecht: Copernicus Institute of Sustainable Development, Utrecht University.

[6] Garzón, J., Patiño, E., & Marulanda, C. (2025). Systematic review of artificial intelligence in education: Trends, benefits, and challenges. *Multimodal Technologies and Interaction*, 9(8), article number 84. doi: 10.3390/mti9080084.

[7] Gligorea, I., Cioca, M., Oancea, R., Gorski, A.-T., Gorski, H., & Tudorache, P. (2023). Adaptive learning using artificial intelligence in e-learning: A literature review. *Education Sciences*, 13(12), article number 1216. doi: 10.3390/educsci13121216.
[8] Hussein, S., & Maan, N. (2023). UACA: Unified access control approach for heterogeneous database based-on service data object. *Technium: Romanian Journal of Applied Sciences and Technology*, 9, 26-40. doi: 10.47577/technium.v9i.8686.
[9] Ko, H., Li, S., & Chen, Y. (2022). A survey of recommendation systems: Recommendation models, techniques, and application fields. *Electronics*, 11(1), article number 141. doi: 10.3390/electronics11010141.
[10] Kreutz, C.K., & Schenkel, R. (2022). Scientific paper recommendation systems: A literature review of recent publications. *International Journal on Digital Libraries*, 23, 335-369. doi: 10.1007/s00799-022-00339-w.
[11] Li, S., & Li, D. (2025). Research on personalised learning recommendation system based on machine learning algorithm. *Scalable Computing: Practice and Experience*, 26(1), 432-440. doi: 10.12694/scpe.v26i1.3844.
[12] Masciari, E., Umair, A., & Ullah, M.H. (2024). A systematic literature review on AI-based recommendation systems and their ethical considerations. *IEEE Access*, 12, 121223-121241. doi: 10.1109/ACCESS.2024.3451054.
[13] Nikiforova, L., Dohtieva, I., & Zharinov, S. (2025). Specificity of the design of the development of an information resource and an electronic register of scientific professional publications in the context of digitalisation of the scientific field. *Innovation and Sustainability*, 4, 62-75. doi: 10.31649/ins.2024.4.62.75.
[14] Petryna, D., Kornuta, V., & Kornuta, O. (2024). Using neural network tools to accelerate the development of Web interfaces. *Information Technologies and Computer Engineering*, 60(2), 42-50. doi: 10.31649/1999-9941-2024-60-2-42-50.
[15] Putri, C.A., Syaifuddin, A., Rohman, N., Aziz, A., & Ponijan, R.M.P. (2025). Optimizing the use of digital libraries in basic education. *Edunesia: Jurnal Ilmiah Pendidikan*, 6(1), 1-13. doi: 10.51276/edu.v6i1.940.
[16] Riordan, A., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: Systematic literature review. *Computers and Education: Artificial Intelligence*, 6, article number 100215. doi: 10.1016/j.caeai.2024.100215.
[17] Roy, D., & Dutta, M. (2022). A systematic review and research perspective on recommender systems. *Journal of Big Data*, 9, article number 59. doi: 10.1186/s40537-022-00592-5.
[18] Shovkoplias, M.O., & Liubchak, V.O. (2024). Review of models and methods for individual customisation of a scientist's information service. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 2, 88-96. doi: 10.32782/IT/2024-2-11.
[19] Umbach, M. (2024). Open Science and the impact of Open Access, Open Data, and FAIR publishing principles on data-driven academic research: Towards ever more transparent, accessible, and reproducible academic output? *Statistical Journal of the IAOS*, 40(1), 59-70. doi: 10.3233/SJI-240021.
[20] Vargo, S.L., & Lusch, R.F. (2025). Service-dominant logic 2025. *International Journal of Research in Marketing*, 34(1), 46-67. doi: 10.1016/j.ijresmar.2016.11.001.
[21] Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications*, 252(A), article number 124167. doi: 10.1016/j.eswa.2024.124167.
[22] Zhao, N., Wei, C., Zhang, X., & Li, J. (2025). The role of AI in facilitating interdisciplinary collaboration: Evidence from AlphaFold. *ArXiv*. doi: 10.48550/arXiv.2508.13234.

Концептуальна інформаційна модель та методи створення персоналізованого інформаційного сервісу науковця

Володимир Любчак

Кандидат фізико-математичних наук, доцент
Сумський державний університет
40007, вул. Харківська, 116, м. Суми, Україна
https://orcid.org/0000-0002-7335-6716

Максим Шовкопляс

Аспірант
Сумський державний університет
40007, вул. Харківська, 116, м. Суми, Україна
https://orcid.org/0009-0007-4192-4743

Анотація. Сучасна наукова діяльність відбувається в умовах зростання обсягів даних і фрагментованості цифрових інструментів. Дослідники змушені використовувати численні незалежні сервіси – від пошукових систем і бібліографічних менеджерів до платформ аналітики та наукової комунікації. Це ускладнює робочі процеси, спричиняє дублювання дій та знижує ефективність досліджень. Метою проведеного дослідження було розроблення концептуальної моделі персоналізованого інформаційного сервісу науковця, здатного об'єднати пошукові, аналітичні, бібліографічні та комунікаційні функції в інтегрованому середовищі. Методологія дослідження передбачала інформаційне моделювання, інтеграцію даних із відкритих джерел через інтерфейс програмування застосунків та формалізацію користувацького профілю для автоматизованого добору релевантних сервісів. Результатом стало створення формалізованої інформаційної моделі, поданої у вигляді інформаційної квінтуплі $F = (P, S, D, I, U)$, де P – модуль формування профілю, S – модуль відбору сервісів, D – модуль даних, I – інтерфейсний модуль, U – аналітична підсистема. Профіль користувача P описано як векторну структуру, що включає параметри інтересів, тематичних напрямів, інструментів і мов. Для кожного зовнішнього сервісу S_i визначено релевантність $R(S_i, P)$, обчислена за допомогою нормалізаційної функції f_{norm} і метрики схожості Sim, що забезпечило побудову персоналізованої конфігурації сервісів. Практичне впровадження моделі у прототипі (Flutter + Firebase) продемонструвало її дієвість для скорочення часу пошуку релевантної інформації та підвищення ефективності наукової діяльності. Отримані результати можуть слугувати основою для створення адаптивних цифрових платформ, що підтримують відкриту науку та міждисциплінарну взаємодію

Ключові слова: адаптивні цифрові платформи; наукова комунікація; профіль користувача; персоналізація даних; інтеграція наукових ресурсів; користувацький інтерфейс; підтримка досліджень

BULLETIN of Cherkasy State Technological University

Journal homepage: https://bulletin-chstu.com.ua/en

Vol. 31 No. 1. 2026

UDC 004.4/.415+004.89/.9
DOI: 10.62660/bcstu/1.2026.85

ISSN: 2306-4412
E-ISSN: 2708-6070

The impact of cloud services on secure software development

Maksym Shishkin*
Postgraduate Student
Simon Kuznets Kharkiv National University of Economics
61165, 9-A Nauky Ave., Kharkiv, Ukraine
https://orcid.org/0009-0008-0553-944X

Abstract. The evolution of cloud computing has significantly changed the approaches to developing, testing, deploying, and securing software. The paper aimed to examine the integration of cloud technologies into all stages of the Software Development Life Cycle (SDLC) with a focus on implementing and enforcing security practices throughout the development process. To achieve the objectives of the study, a comparative analysis was used that covers cloud implementations of the SDLC relative to traditional models in terms of key indicators: cost-effectiveness, speed of deployment, level of collaboration, scalability, and security. Real-world case studies that demonstrate the use of cloud tools and platforms were considered, including Infrastructure as Code and Continuous Integration/ Continuous Deployment, to increase productivity, agility, and early implementation of security controls (a "security shift to the left" approach). The analysis results has shown that the use of cloud practices in a secure SDLC helps to reduce time to market, increases the level of proactive security management, and supports iterative and agile development cycles. At the same time, challenges related to compliance with regulatory requirements, user identity management, and vendor lock-in risk were identified. A set of best practices for implementing secure cloud SDLC workflows was proposed and areas for further research are identified, including automated security testing and integration of artificial intelligence into secure software delivery processes. The practical value of the study lies in the formulation of recommendations that will help organisations create sustainable, efficient, and secure cloud development environments

Keywords: cybersecurity practices in development; infrastructure as code; continuous integration and delivery pipelines; automated vulnerability testing; configuration management; cloud responsibility models; system oversight and monitoring

INTRODUCTION

As of 2025, software development takes place in an environment characterised by increasingly dynamic business requirements, the globalisation of teams, high infrastructure workloads, and continuous release delivery. In such an environment, traditional Software Development Life Cycle (SDLC) models, while still relevant for stable and predictable projects, often prove insufficiently flexible without the integration of cloud services. Lack of adaptability can lead to increased risks of malicious attacks, data leaks, compliance violations, and loss of control over the environment. At the same time, cloud services offer scalability, automation, and integration of security mechanisms at all stages of the SDLC, opening up new opportunities for creating resilient, secure, and flexible development processes. Therefore, in these conditions, it is necessary not to abandon traditional approaches, but to learn to make informed choices between classic and cloud-oriented models depending on the requirements, risks, and goals of a specific project.

S.M. Saleh *et al.* (2024) conducted a systematic review of 66 publications devoted to the security of Continuous Integration/Continuous Deployment (CI/CD) pipelines in a cloud environment. The study identified

Article's History: Received: 07.11.2025; Revised: 10.02.2026; Accepted: 16.03.2026; Published: 08.04.2026

Suggested Citation:
Shishkin, M. (2026). The impact of cloud services on secure software development. *Bulletin of Cherkasy State Technological University*, 31(1), 85-100. doi: 10.62660/bcstu/1.2026.85.

*Corresponding author

the main tools (e.g., Harbour, SonarQube, GitHub Actions) and key challenges (image manipulation, unauthorised access, weak authentication). The study concluded that the integration of security at the CI/CD stages significantly reduces the number of vulnerabilities in production environments. The study by M. Chauhan & S. Shiaeles (2023) examined cloud security frameworks (The National Institute of Standards and Technology, NIST Cloud; Cloud Security Alliance, etc.) and noted that none of them fully cover all security criteria, creating a "gap" in standardisation. The study proposed a unique generalised model for the assessment of the security of cloud services, combining elements of existing frameworks to reduce the gaps between standards. K. Kiashemshaki *et al.* (2025) addressed secure programming practices, including assessing the role of large language models (LLMs) in code generation, noting that despite the growth of automation, the implementation of secure code remains fragmented. The study emphasised that the use of LLMs in cloud environments can only be effective if automatic security validation systems are integrated. O. Vakhula & I. Opirskyy (2023) investigated the integration of the "Security as Code" approach in cloud environments for security automation at the Infrastructure as Code (IaC) and Development, Security, Operations (DevSecOps) levels. The study demonstrated that this approach reduces the human factor in the DevSecOps process and ensures continuous compliance of systems with standards requirements. The work of Y. Matseniuk & A. Partyka (2024) proposed an automated compliance verification method for Amazon Web Services/Google Cloud Platform (AWS/GCP) and Azure cloud accounts and emphasised the importance of configuration standardisation. The study developed a configuration control model that uses IaC policies to ensure continuous security auditing in DevOps processes. In a review paper, researcher A.S. Pawar (2025) described modern approaches to securing containerised and microservice environments, with an emphasis on security practices in cloud-native systems, such as service mesh, runtime protection, and DevSecOps. The study emphasised that the transition to cloud-native technologies requires a new way of thinking about security, where the focus shifts from perimeter protection to the security of components and development processes.

Despite significant progress in researching cloud environment security and DevSecOps tools, the direct impact of cloud services on the entire SDLC in the context of security has been analysed to a lesser extent. There has been insufficient research into how the use of cloud services changes approaches to requirements, design, testing, and deployment from a security perspective, especially in the Ukrainian or regional context. Therefore, the research relevance is determined. The study aimed to identify the main aspects of the impact of cloud services on secure software development and to outline key areas for improving security practices in this area.

MATERIALS AND METHODS

This study is a systematic comparative review. The comparison method was used within the scope of thy study. The objects of comparison were the traditional and SSDLC, to determine their advantages, limitations, and impact on development security. Traditional SDLC was characterised as a linear and sequential approach in which each stage is completed before the next one begins. It is based on local development environments, manual testing, static infrastructure configuration, and limited scalability. This approach remains effective in stable and predictable environments but is less adaptable to dynamic cloud ecosystems. The study focused separately on the SSDLC model. The methodology was built as a step-by-step research cycle: defining comparison criteria; collecting relevant sources and systematising them; content analysis; comparative evaluation of SDLC and Cloud SSDLC models; and forming generalised conclusions.

Cloud SSDLC was defined as a model that integrates DevOps, continuous integration and delivery, containerisation, IaC, automated testing, and the Shift-Left Security approach. This ensures flexibility, automation, repeatability of environments, and built-in security controls at all stages of the lifecycle. The comparison methodology was conducted through a step-by-step analysis of the six phases of SDLC, for each of which the following criteria were evaluated: automation, flexibility, scalability, security, deployment speed, operational risks, and configuration management, as these parameters most fully reflect the key differences between traditional and cloud-oriented models, determine the quality and efficiency of each phase of the life cycle, and directly affect the ability of organisations to meet modern requirements for reliability, security, and speed of software delivery.

Content analysis was used to systematise the results, summarised comparative tables were created, and the analytical part was based on a comprehensive review of various types of sources, including peer-reviewed scientific publications for 2018-2025, technical reports from leading companies (Amazon, Microsoft, Google, Red Hat, CNCF), official documentation of cloud platforms, industry standards and recommendations (NIST, 2018; OWASP Foundation, n.d.). Industry analytical materials such as developer blogs and white papers (Amazon Web Services, n.d.b) were also used. Sources were collected from Google Scholar, IEEE Xplore, Scopus, ACM Digital Library, and Semantic Scholar databases. The following keywords were used for the search: "cloud SDLC", "secure SDLC", "DevSecOps", "cloud-native security", "infrastructure as code", "CI/CD security", "shift-left security". The analysis included works describing SDLC or SSDLC methodologies, related to cloud infrastructure, and containing empirical results or formal models. Sources that duplicate results, lack technical specifics, or refer to outdated technical approaches (before 2018, except for fundamental works) were excluded.

The quality of sources was assessed according to the following parameters: presence of peer review, date of publication, technical specifics, and consistency with modern cloud architecture models. Further synthesis of the material was conducted by classifying sources according to SDLC stages and types of technical solutions (DevOps processes, security, infrastructure, testing), which made it possible to comprehensively compare models and formulate conclusions about the effectiveness of Cloud SSDLC.

RESULTS AND DISCUSSION

The software development life cycle describes the complete sequence of stages in the creation of a software product, from requirements analysis to operational support. The main stages include requirements analysis, design, development, testing, deployment, and support. This structure ensures consistency, predictability, and quality control at all stages of development, which can be used for standardisation of the process and comparison of different approaches to its implementation. In contrast to the traditional SDLC, which involves the implementation of security practices mainly during the testing or post-deployment stages, SSDLC integrates security at all stages of software development. This includes threat modelling during requirements analysis, architecture design with security checkpoints, automated code review, dynamic and static vulnerability scanning, secure deployment mechanisms, and real-time monitoring. This approach supports the concept of "shift-left security", where risks are eliminated early, reducing the cost of fixing bugs and increasing system reliability. Within the cloud approach, SSDLC is combined with DevOps practices to form a DevSecOps model based on automation, infrastructure as code, and continuous compliance monitoring.

Requirements gathering and analysis are the first and one of the most relevant stages of SDLC. At this stage, expectations and constraints for the future software system are defined from both a user and technical perspective. The process involves close interaction between all stakeholders, including business analysts, project managers, end users, and development teams (IEEE Computer Society, 2024). This usually results in the compilation of key documentation, such as business requirements, functional specifications, and descriptions of use cases and user stories. In traditional, non-cloud environments, this process is often linear and based on static documents and pre-scheduled meetings. This approach can cause delays and misunderstandings between project participants (Basili & Turner, 1975). The use of cloud technologies significantly changes the approach to requirements gathering. Thanks to real-time collaboration and interaction capabilities, cloud tools increase the efficiency, transparency, and flexibility of this process (Google Cloud, n.d.). Notable examples include Atlassian Jira and Confluence, which provide documentation and user story tracking with real-time editing and version control. Lucidchart and Miro are used for visualising processes and collaborating on ideas, while Google Workspace and Microsoft 365 are used for creating and editing text documents and spreadsheets. In the field of interface and experience design, Figma and FigJam are actively used to support interactive prototyping and collaborative workshops. In addition, platforms such as Amazon Honeycode and Google AppSheet provide low-code tools that can be used to quickly create prototypes of business processes based on collected requirements (Amazon Web Services, n.d.a).

Cloud-based tools help overcome a range of typical challenges associated with the traditional approach to requirements gathering. They improve collaboration between all stakeholders through real-time co-editing and commenting, which significantly reduces reliance on synchronous meetings. Automated change tracking and access to edit history simplify version control and ensure process transparency. In addition, cloud solutions help accelerate iterations by quickly updating requirements in line with changes in the business environment. Lastly, they render the process more inclusive: participants from different locations and time zones can actively participate, reducing the probability of misunderstandings and improving the alignment of the product vision (Jagli & Yeddu, 2017).

Despite the numerous advantages of cloud tools, some problems remain relevant. In particular, intensive collaboration can lead to an excessive amount of documentation and user stories, making it difficult to prioritise and manage them. There is also the issue of tool fragmentation: teams often combine multiple platforms (e.g., Jira, Miro, and Google Docs) that do not always integrate seamlessly, making centralised requirements management difficult. In addition, organisations operating under traditional or hybrid development models may find it difficult to adapt their processes to more flexible, cloud-oriented and Agile-compatible approaches. New challenges associated with cloud integration. Notably, the introduction of cloud platforms is accompanied by new challenges that were not characteristic of traditional operating models. One of the main issues is security and access control: confidential project and business information stored in a cloud environment requires careful user identity management, effective data encryption, and compliance with NIST (2018) regulatory requirements.

Another challenge is the "collaboration fatigue". The constant availability of cloud tools can cause notification overload, emotional exhaustion among employees, and, as a result, a decline in overall productivity. In addition, the effectiveness of working in a cloud environment depends largely on a stable internet connection. In cases where connectivity is unstable or limited, this can seriously interfere with key planning and coordination sessions. Cultural resistance should not be ignored either: some stakeholders are reluctant to abandon formal,

document-oriented practices in favour of more flexible, iterative, and collaborative approaches that involve the use of cloud tools. In addition to the advantages and challenges described above, it is worth summarising the main differences between traditional and cloud-based approaches to requirements gathering (Table 1).

Table 1. Key differences between traditional and cloud-based approaches to requirements gathering

Criteria	Traditional aspect	Cloud-oriented approach
Process characteristic	Linear, sequential, with fixed stages	Iterative, flexible, with priority on continuous improvement
Tools	Word/Excel, offline meetings, e-mail	Jira, Confluence, Miro, Figma, Google Workspace, Microsoft 365
Participant interaction	Mostly synchronous (meetings, calls)	Asynchronous and synchronous real-time interaction via cloud services
Version control	Manual document updates, difficult change tracking	Automatic saving of change history and version control
Accessibility	Limited – depends on physical meetings or corporate network	Access from any location via the Internet
Flexibility in requirement changes	Low – changes often require reworking of documentation	High: requirements are updated directly in cloud systems
Key challenges	Slow communication, lack of transparency	Data security, tool fragmentation, "collaboration fatigue"

Source: compiled by the author based on V.R. Basili & A.J. Turner (1975), D. Jagli & S. Yeddu (2017), NIST (2018), IEEE Computer Society (2024), Google Cloud (n.d.), Amazon Web Services (n.d.a; n.d.b)

The comparison presented demonstrates that the traditional approach to requirements gathering is linear and inflexible, relying heavily on physical meetings and manual document management, which complicates changes, slows down communication, and increases the risk of version inconsistencies. The cloud-oriented approach, on the other hand, provides iterativeness, centralised access and automated version control, which promotes process transparency, rapid requirements updates and effective collaboration between distributed teams. At the same time, the integration of a large number of tools creates the risk of process fragmentation, and working with confidential data in the cloud increases security requirements, particularly for access management, encryption, and operational control. The dynamic nature of cloud services can also create new threats, such as misconfigurations, unwanted data distribution, or incorrect permissions. Therefore, cloud services significantly improve the adaptability and quality of the requirements gathering process, but the effectiveness of this approach depends on proper standardisation, maturity of access management practices, and adherence to cloud security principles. After the requirements gathering stage comes the system design phase, which involves transforming functional expectations into a structure suitable for technical implementation. This process encompasses both high-level architectures, including service components, data flows, and security boundaries, and detailed design, including database schemas, software interface descriptions, and user interface behaviour modelling (IEEE Computer Society, 2024). In traditional local environments, design typically involves creating static architectural diagrams and manually planning infrastructure within limited resources. This approach often limits the flexibility, scalability, and adaptability of the system to changing requirements (Basili & Turner, 1975).

The impact of cloud tools on system design. Modern cloud technologies significantly expand the possibilities of architectural design by supporting dynamism, scalability, and automation. Cloud platforms provide developers with the means to quickly create architectural diagrams, support collaboration on architectural solutions, programmatically describe infrastructure with the ability to version it, and formalise documentation of decisions made. In addition, they offer methodological support in the form of frameworks that help adhere to the principles of reliability, performance, security, and cost optimisation in the design process. These capabilities are implemented using tools and services such as AWS Architecture Diagrams, Azure Architecture Centre, Google Cloud Architecture Framework, Lucidchart, Draw.io, Miro platforms, infrastructure as code tools, Architecture Decision Records (ADR) documentation formats, and Well-Architected Frameworks offered by Amazon Web Services, Azure, and Google Cloud Platform (Amazon Web Services, n.d.a; Google Cloud, n.d.).

The use of cloud tools addresses a range of critical shortcomings of the traditional approach to design. They eliminate scalability limitations through native support for horizontal scaling, elasticity, and microservice architecture (Jagli & Yeddu, 2017). Automation through IaC avoids human error by reducing the need for manual infrastructure configuration. Cloud platforms also promote a modular approach to architecture, facilitating the gradual evolution of the system. In addition, the ability to create isolated environments on demand can be used to quickly test solutions before the main implementation begins, which significantly accelerates feedback.

At the same time, some challenges remain relevant even in a cloud environment. Overuse of cloud platform functionality can lead to so-called over-engineering, where architecture becomes overly complex without an objective need (Jagli & Yeddu, 2017). In addition, project documentation is often scattered across different tools, from visual modelling diagrams to IaC repositories and team collaboration platforms, making holistic management difficult. Another important issue is knowledge gaps: not all teams have the relevant expertise in cloud distributed systems, especially those accustomed to traditional approaches. New risks are caused by integration with cloud services. Integration with cloud services also creates new types of risks that require careful monitoring. One of them is the complexity of budget forecasting, as cloud pricing models (such as pay-per-use or multi-tier pricing) require more in-depth financial planning (Amazon Web Services, n.d.a). Another aspect is security, which must be considered at the design stage, from data encryption and user role management to Application Programming Interface (API) gateway configuration (OWASP Foundation, n.d.). In addition, working with IaC requires not only technical knowledge but also ongoing maintenance, as delayed updates can lead to technical debt accumulation (Jagli & Yeddu, 2017). Lastly, dependence on specific services from a single provider (e.g., AWS Lambda or Azure Functions) may limit the ability to migrate to another platform, creating the risk of vendor lock-in (Google Cloud, n.d.). In addition to the advantages and challenges described above, it is necessary to summarise the main differences between traditional and cloud-based approaches to system design (Table 2).

Table 2. Key differences between traditional and cloud-based approaches to system design

Criteria	Traditional aspect	Cloud-oriented approach
The nature of the design process	Static, sequential design with fixed architectural solutions	Dynamic, flexible design with the ability to quickly change the architecture
Architecture type	Monolithic, with limited scalability	Microservice or cloud-native, with elasticity support
Project design tools	Local graphic redactors (Visio, offline Drawio)	Cloud platforms (AWS Architecture Centre, Azure Diagrams, Lucidchart, Miro)
Infrastructure description	Manual planning and documentation	Infrastructure as code (Terraform, CloudFormation, Pulumi)
Process automation	Minimal configuration is manual	High – deployment, testing and monitoring are automated
Team collaboration	Limited, dependent on physical meetings or local network	Real-time collaboration, cloud storage of artefacts
Security	Added in the later stages of the project	Addressed at the architecture stage (security by design)
Scalability	Limited by hardware	Flexible – dynamic scaling of cloud resources
Main risks	Limited adaptability, slow architectural evolution	Over-engineering, vendor lock-in, and the need for a highly skilled team

Source: compiled by the author based on V.R. Basili & A.J. Turner (1975), D. Jagli & S. Yeddu (2017), IEEE Computer Society (2024), Google Cloud (n.d.), OWASP Foundation (n.d.), Amazon Web Services (n.d.a)

The comparison showed that the cloud-oriented approach to design differs significantly in terms of flexibility, support for microservice architecture, and the use of infrastructure as code, which provides automation, scalability, and security considerations at the architectural level. Traditional design remains static, monolithic, and dependent on manual documentation, which slows down system development. Cloud tools such as AWS Architecture Centre or Terraform improve collaboration and accelerate decision-making, but they also create new risks, including the possibility of errors in IaC configurations, excessive permissions, leakage of architectural artefacts, or vendor lock-in. Therefore, the effectiveness of cloud design depends on controlled complexity and proper security management.

During the development stage, the architectural design is transformed into working software. Developers implement functionality, write code, integrate services, and test modules in accordance with predefined requirements and technical solutions. During this period, individual components are also tested, code reviews are conducted, and version control is performed to maintain the quality and stability of the software product (IEEE Computer Society, 2024). In traditional approaches, especially in local environments, the development process was often limited by the computing resources of a single machine, insufficient interaction between team members, and manual testing and deployment scenarios. These factors contributed to delays, environmental consistency issues, and overall productivity declines.

Cloud platforms provide developers with a wide range of tools that simplify, standardise and automate software development processes. They create unified development environments from any access point, integrate

continuous integration and delivery mechanisms, centrally manage dependencies, detect defects and vulnerabilities in a timely manner, and gradually implement new functionality with the ability to control the impact on end users. These capabilities are implemented using cloud-based integrated development environments (e.g., AWS Cloud9), CI/CD tools (GitHub Actions, GitLab CI/CD, AWS CodePipeline), artefact repositories (JFrog Artifactory, AWS CodeArtifact), code and security analysis tools (SonarCloud, Snyk, GitHub Dependabot), and feature control systems (LaunchDarkly, CloudBees) (OWASP Foundation, n.d.).

The integration of cloud technologies largely overcomes many traditional barriers to development. One of the most relevant advantages is the elimination of inconsistencies between environments. The use of cloud IDEs or containerised environments (e.g., Docker, Dev Containers) avoids situations where an application only works on the developer's local machine. Deployment time is also significantly reduced thanks to automated CI/CD pipelines, which improves release frequency and reduces the probability of human error (Jagli & Yeddu, 2017). Cloud repositories and collaboration platforms such as GitHub or Bitbucket improve interaction between team members by enabling collaborative editing, code review, and real-time change control. In addition, integrated security analysis tools can be used to identify potential vulnerabilities at the code writing stage, reducing risks for the finished product (OWASP Foundation, n.d.). Despite numerous improvements, some challenges remain relevant in cloud development. Frequent iterations and rapid deployment can push teams to make temporary decisions that accumulate technical debt and reduce code maintainability (IEEE Computer Society, 2024). Maintaining consistent code quality, despite the availability of modern tools, still depends largely on the internal discipline of the team and clear development standards. In addition, it is often difficult for new team members to join a project without proper documentation or integration standards, which slows down the adaptation process.

New risks are caused by integration with cloud services. Integrating the cloud into the development process also creates new risks. In particular, the diversity of tools and practices such as containers, serverless architecture, and IaC can overwhelm teams, especially less experienced ones. Excessive automation, while useful, sometimes leads to blind trust in systems that may miss critical defects or malfunction without proper verification. There is also an increased risk associated with incorrect configurations; for example, an error in a Terraform or YAML file for a CI/CD process can make an application vulnerable (NIST, 2018). Lastly, the use of vendor-specific services (such as AWS Lambda, Firebase, or Azure Functions) can lead to dependence on a single cloud platform, reducing flexibility and the ability to migrate solutions in the future. In addition to the advantages and challenges described above, it is necessary to summarise the main differences between traditional and cloud-based approaches to system development (Table 3).

Table 3. Key differences between traditional and cloud-based approaches to system development

Criteria	Traditional aspect	Cloud-oriented approach
Development environment	Local, depends on the configuration of individual developers' machines	Cloud-based or containerised (GitHub Codespaces, AWS Cloud9, Docker)
Integration and delivery	Manual testing and deployment, long release cycles	Continuous integration and delivery (CI/CD) with automated testing and releases
Dependency management	Lack of centralised control, frequent conflicts between versions	Centralised artefact repositories (JFrog, AWS CodeArtifact)
Version control and collaboration	Local repositories, limited team collaboration	Cloud-based Git platforms (GitHub, GitLab, Bitbucket) with collaborative editing and code review capabilities
Quality control	Testing is performed after the main development, often manually.	Automated testing, static code analysis (SonarCloud, Snyk)
Security	Implemented after the implementation stage	Integrated into the development process (security-as-code, shift-left security)
Flexibility and scalability	Depends on the developer's hardware resources	Dynamic scaling, flexible environments on demand
Typical tools	Local IDE (IntelliJ, Eclipse, Visual Studio)	Cloud IDE (GitHub Codespaces, AWS Cloud9, JetBrains Space)
Main benefits	Simplicity of local process control	High deployment speed, team synchronisation, and automation
Main risks	Environment incompatibility, manual errors, and release delays	Tool overload, configuration errors, vendor lock-in

Source: compiled by the author based on D. Jagli & S. Yeddu (2017), NIST (2018), IEEE Computer Society (2024), OWASP Foundation (n.d.)

The comparison shows that the transition from local, isolated, and manual processes to cloud-oriented development models not only accelerates delivery cycles but also significantly changes the security profile. In traditional environments, manual operations, lack of centralised version control, and dependence on local infrastructure increase the risks of human error, configuration drift, and environment inconsistency. The cloud approach, on the other hand, integrates security into the development process itself through shift-left practices, security-as-code, and automated static and dynamic code analysis, enabling vulnerabilities to be detected much earlier. The use of containerised environments, centralised dependency repositories, and CI/CD pipelines provides controllability, traceability, and a lower probability of configuration errors. However, increased automation and a large number of services also create new threats: there is a growing risk of configuration vulnerabilities in cloud resources, vendor lock-in, compromise of secrets in pipelines, and incorrect configuration of access rights. Thus, although the cloud approach significantly increases the level of proactive security, its effectiveness depends on mature infrastructure management, continuous monitoring, and adherence to the principles of least privilege and controlled automation.

The testing phase is a critical part of the software development life cycle, as it ensures that the implemented functionality complies with the specified requirements and quality standards. Testing covers various levels of verification, including unit, integration, system, load, and security testing (Google Cloud, n.d.). Its purpose is to identify errors, prevent regressions, and ensure that the software product meets user and business expectations before deployment in a production environment. In traditional SDLC models, testing was often postponed to the final stage of the development cycle and performed manually, which led to delays, accumulation of defects, and increased cost of fixes (Microsoft, n.d.). In addition, the isolation of test environments made it difficult to reproduce the real conditions in which the application would operate.

The impact of cloud tools on testing. Cloud technologies have radically transformed the testing process, making it more automated, scalable, and integrated into the overall development infrastructure. This has made it possible to continuously perform checks with every code change, ensure cross-browser and cross-platform compatibility without local environment configuration, simulate load in conditions close to production, detect vulnerabilities at the development stage, and test interaction with unavailable or unstable external services. These capabilities are implemented through the integration of tests into CI/CD pipelines (GitHub Actions, GitLab CI, Jenkins in the cloud), cross-browser testing platforms (BrowserStack, Sauce Labs), load testing tools (Apache JMeter in AWS, Gatling Cloud, Azure Load Testing services), security analysis tools (Snyk, Checkmarx, OWASP ZAP) and API virtualisation services (WireMock Cloud, Mountebank) (Microsoft, n.d.).

The use of cloud solutions eliminates a range of problems that were characteristic of the traditional approach to testing. In particular, the scalability of cloud resources can be used for mass parallel testing without the limitations associated with local capacities. Cross-platform testing using device farms provides broader compatibility coverage, while cloud load testing can ensure accurate simulation of real-world usage scenarios. In addition, CI/CD pipelines provide instant feedback after code changes, which accelerates the detection and elimination of defects (Jagli & Yeddu, 2017). Unified cloud testing environments also mimic production conditions with significantly higher efficiency than on-premises configurations, reducing the risk of unexpected software behaviour after deployment. Despite significant progress, some problems remain relevant even in the cloud context. One of them is the instability of automated tests: their results can be inconsistent and false positives due to temporary dependencies, background processes, or insufficient data isolation (IEEE Computer Society, 2024). The cost of maintaining automated tests also increases over time. As the functionality of the application expands, tests must be constantly updated and adapted. Even in well-built automated pipelines, there is a risk of incomplete coverage. Due to time constraints or limitations of the tools, some critical scenarios may remain insufficiently tested.

New risks are caused by integration with cloud services. The transition to cloud testing, while creating new opportunities, also creates new risks. First, there is a growing need for careful handling of data to avoid leaks of confidential information and compliance with legal regulations (e.g., GDPR), NIST (2018). Furthermore, large-scale tests, especially those related to performance or end-to-end testing, can require significant computing resources, leading to increased costs. The security of test environments and CI/CD systems is critical, as these environments contain both code and secrets necessary for testing NIST, 2018). Lastly, the complexity of integrating various cloud tools into a single pipeline may require the development of special scripts, configurations, or even orchestration systems, which increases the technical complexity of the project. In addition to the advantages and challenges described above, it is worth summarising the main differences between traditional and cloud-based approaches to system testing (Table 4).

Table 4. Key differences between traditional and cloud-based approaches to system testing

Criteria	Traditional aspect	Cloud-oriented approach
Testing environment	Local machines, limited resources	Cloud environments, dynamic scaling
Automation	Testing is predominantly manual	Integration of automated tests into CI/CD

Continued Table 4.

Criteria	Traditional aspect	Cloud-oriented approach
Load testing	Limited by local hardware	Scalable, realistic load testing through cloud resources
Cross-platform compatibility	Requires manual verification or emulators	Cloud device farms (BrowserStack, Sauce Labs)
Security testing	Implemented in later stages	Integrated into CI/CD pipelines (Snyk, Checkmarx)
Data management	Test data is stored locally	Centralised management in the cloud
Tool integration	Requires manual configuration	Supported by platforms (GitHub Actions, Jenkins Cloud)

Source: compiled by the author based on NIST (2018), D. Jagli & S. Yeddu (2017), IEEE Computer Society (2024), Microsoft (n.d.), Google Cloud (n.d.)

Comparative analysis shows that traditional testing based on local resources and manual operations creates significant security risks: limited automation complicates the reproducibility of checks, delays the detection of vulnerabilities, and shifts security testing to the final stages of the SDLC, when fixes are costly and potentially dangerous for system stability. The cloud-oriented model eliminates these shortcomings by integrating automated security checks into CI/CD processes, enabling the scaling of environments to simulate real attacks, and using specialised tools such as Static Application Security Testing, Dynamic Application Security Testing, Software Composition Analysis, and secrets analysis, which significantly improve threat detection efficiency. In addition, centralised management of test data, logs, and artefacts strengthens access control and simplifies auditing, reducing the risks of unauthorised use or leakage of confidential information. At the same time, dependence on cloud platforms increases the requirements for proper configuration of access rights, segmentation of environments, and protection of CI/CD pipelines, as incorrect configurations or compromise of cloud accounts can create new attack vectors. Thus, cloud testing significantly increases the level of proactive security but requires greater discipline in configuration management and mature DevSecOps practices. Deployment is one of the final stages of the software development life cycle, during which the product is released into the production environment and becomes available to end users. In traditional SDLC models, the deployment process was often performed manually, sporadically, with pre-planned system downtime. This approach was prone to errors, had a high risk of failure, and was accompanied by significant support costs (NIST, 2018). With the advent of cloud tools and DevOps practices, deployment strategies have undergone significant changes. They have become automated, fast, and more continuous, avoiding downtime, reducing risks, and ensuring high application availability (Jagli & Yeddu, 2017).

The impact of cloud tools on the deployment stage. Modern cloud platforms create vast opportunities for automation and centralised management of software deployment processes in different environments. The deployment infrastructure is integrated into the development lifecycle, ensuring automatic collection, testing, and delivery of applications at each stage, which helps reduce errors, accelerate the release of new versions, and improve delivery reliability. Support for heterogeneous environments, from development to production, is achieved by configuring pipelines to suit the specifics of each environment. Containerisation unifies the delivery process, which can be used to scale computing resources, balance loads, and automatically restore services in the event of failures. Orchestration systems, particularly those based on Kubernetes, can be used for centralised management of the container lifecycle, automating deployment, updates and monitoring while maintaining stability and control over the infrastructure. These capabilities are implemented using continuous integration and delivery (CI/CD) services such as AWS CodeDeploy, Azure Pipelines, Google Cloud Deploy, and GitHub Actions, as well as container orchestration systems including the Kubernetes platforms Azure Kubernetes Service, Amazon Elastic Kubernetes Service, Google Kubernetes Engine, and Amazon ECS service.

Serverless deployment solutions occupy a separate niche, providing automatic scaling, event-driven function activation, and pay-as-you-go pricing. This can be used to prioritise business logic without wasting resources on maintaining a server environment. Another key component is Infrastructure as Code (IaC), which involves a software description of the entire infrastructure, including network components, computing resources, data storage services, and security policies. This approach provides reproducibility, version control, change auditing, and automation of deployment processes, which is critical for teamwork and DevOps practices. These capabilities are implemented using serverless deployment tools such as AWS Lambda, Azure Functions, and Google Cloud Functions, as well as IaC solutions including Terraform, AWS CloudFormation, and Pulumi. In addition, cloud deployment supports modern delivery strategies such as blue-green, canary, and rolling deployments. Specialised tools such as Argo CD and Spinnaker are used to implement these strategies, which can be used for gradual or parallel implementation of changes without negatively impacting users.

The use of cloud tools largely eliminates the problems inherent in the traditional approach to deployment. In particular, process automation reduces the probability of human error and ensures consistency across all environments. IaC eliminates the problem of "environment drift", where development, testing, and production environments differ from each other, often leading to unexpected application behaviour. CI/CD pipelines accelerate releases, ensuring their regularity and compliance with agile development methodologies. In addition, strategic deployment planning minimises or completely avoids downtime during updates. The use of gradual implementation methods, such as canary or blue-green, helps to reduce the risks associated with implementing changes in a production environment. Despite the significant advantages, some challenges remain relevant even in the context of cloud deployment. One such problem is the complexity of rolling back changes if critical errors are detected after release. If no appropriate action plan has been developed or no backups have been created, recovery can be difficult. In addition, a build failure or misconfiguration of the CI/CD pipeline can block a release and delay the product's time to market. Another substantial challenge is security. In particular, hard-coded credentials or improper secret management in deployment scripts can lead to confidential information leaks or create vulnerabilities in the system.

New risks are caused by integration with cloud services. Integrating deployments with cloud infrastructure also creates new risks that were not characteristic of traditional systems. One of them is related to cost monitoring. Frequent deployments can lead to the creation of temporary environments or excessive use of computing resources, requiring careful budget control. Process automation and open API endpoints also increase the attack surface. If CI/CD pipelines or environments are not protected by appropriate measures such as access control, event auditing, or privilege restrictions, this can create serious vulnerabilities. Another challenge is the proliferation of tool chains: for automated deployment to function properly, a large number of code management, testing, monitoring, and infrastructure services need to be integrated, which can lead to increased maintenance complexity. Lastly, the organisational aspect should not be underestimated either. The transition to continuous delivery requires not only technical changes, but also a change in culture within the team. In particular, responsibility of developers for code and release quality increases, as does the requirement for trust in automated processes and DevOps practices. In addition to the advantages and challenges described above, it is worth summarising the main differences between traditional and cloud-based approaches to system deployment (Table 5).

Table 5. Key differences between traditional and cloud approaches to system deployment

Criteria	Traditional aspect	Cloud-oriented approach
Deployment process	Manual, periodic, with system downtime	Automated, continuous (CI/CD)
Infrastructure	Configured manually on physical or virtual servers	Described as code (IaC – Terraform, CloudFormation)
Execution environment	Fixed servers, scaling complexity	Containers and orchestration (Docker, Kubernetes)
Release strategies	One-time update with risk of issues	Blue-green, canary, rolling updates
Continuous solutions	None or require server control	AWS Lambda, Azure Functions, GCP Functions
Security	Manual control with accounts	Automated secret and security access policy control
Monitoring and roll-back	Conducted after a failure	Automated dashboards, alerts, and rollback processes
Organisational culture	Distribution of responsibilities between Dev and Ops	DevOps – joint responsibility for releases

Source: compiled by the author based on D. Jagli & S. Yeddu (2017), NIST (2018), IEEE Computer Society (2024), Microsoft (n.d.), Google Cloud (n.d.)

Comparative analysis demonstrates that traditional deployment, performed manually and on statically configured servers, creates significant security risks: manual management of credentials, lack of a controlled environment for updates, and single-stage releases increase the likelihood of human error, leaks of secrets, and failures that are difficult to quickly localise and roll back. In such models, security checks occur after deployment, and monitoring is reactive, increasing the time to detect and respond to incidents. In contrast, cloud-native deployment integrates security directly into CI/CD pipelines through automated secret management, least privilege policies, IaC with configurable configurations, and integrity control mechanisms. The use of containerisation and orchestration ensures environment isolation, while strategies such as canary or blue-green deployments minimise risk by gradually introducing changes and quickly localising potential vulnerabilities. Automated dashboards, alerts, and rollback procedures reduce the "attack window" and ensure rapid response. At the same time, the increase in the number of integrated services raises the bar for proper configuration

management, CI/CD pipeline security, and IaC verification, as a configuration error or cloud account compromise can scale across the entire infrastructure.

Once software deployment is complete, the maintenance and monitoring phase begins, which continues throughout the entire system lifecycle. The main goal of this phase is to maintain uninterrupted software operation, promptly eliminate errors, optimise performance, provide updates, and ensure compliance with changing user needs. This includes continuous performance monitoring, log analysis, and resource usage control, which ensures compliance of the software with modern security requirements and technical standards (IEEE Computer Society, 2024). In traditional approaches, maintenance was usually performed manually: specialists analysed event logs, solved problems as they arose, and performed periodic updates. However, such reactive methods often proved ineffective: delays in responding to incidents led to prolonged downtime, and performance degradation could go unnoticed for a long time (Amazon Web Services, n.d.a).

Modern cloud environments create advanced opportunities for automating maintenance, monitoring system status, and ensuring high availability. With built-in monitoring and incident management mechanisms, engineers can track key metrics in real time, identify deviations from normal values, analyse the causes of failures, and automate responses to critical events. Centralised logging, telemetry data collection, alerts, automatic scaling and recovery, as well as patch and configuration management, ensure stable system operation even in complex dynamic environments. Specialised services are used to implement these capabilities, including Amazon CloudWatch, Azure Monitor, Google Cloud Operations Suite, Datadog, Splunk, etc.

Cloud solutions can address a range of limitations characteristic of traditional services. One of the main advantages is the elimination of limited visibility: centralised dashboards and logging systems provide a comprehensive picture of the system's status. Real-time monitoring can quickly detect potential failures before they affect users, and automatic scaling and self-healing mechanisms significantly reduce the need for constant manual intervention. In addition, automating routine tasks such as configuration updates or patching makes maintenance more stable and predictable. Despite these advantages, the use of cloud tools does not eliminate certain challenges for teams. One common problem is the high number of false alerts: poorly configured monitoring rules can generate an excessive number of non-critical messages, distracting engineers and leading to "alert fatigue". Analysing the root causes of failures also remains a difficult task, especially in distributed systems where incidents can have indirect and unpredictable consequences. In addition, constant monitoring and on-call duty can lead to burnout and reduced efficiency of support teams.

New risks are caused by integration with cloud services. The use of cloud services in the field of maintenance and monitoring creates new threats that must be addressed. One of them is tool overload: an excessive number of services and platforms complicates their integration and management, and also creates the risk of duplication of functions. The security of operational data is also becoming particularly relevant: logs, telemetry, and analytical reports may contain confidential information that requires adequate protection in accordance with access policies and data protection regulations (OWASP Foundation, n.d.). In addition, dependence on the stability and availability of cloud services creates additional vulnerability: in the event of provider failures, monitoring or response functionality may be compromised. The dynamic nature of the cloud environment is also worth mentioning: ephemeral instances, such as containers or serverless functions, complicate traditional tracing methods, requiring new approaches to debugging and monitoring. In addition to the advantages and challenges described above, it is worth summarising the main differences between traditional and cloud approaches to support and monitoring (Table 6).

Table 6. Key differences between traditional and cloud-based approaches to system support and monitoring

Criteria	Traditional aspect	Cloud-oriented approach
Support type	Reactive – troubleshooting	Proactive – automatic detection and incident prevention
System monitoring	Manual log analysis, periodic verifications	Continuous real-time monitoring through services (CloudWatch, Azure Monitor, Datadog)
Incident management	Performed manually, depending on the specialists on duty	Automated notification, root cause analysis, and response orchestration via PagerDuty, Splunk, etc.
Logging and telemetry	Decentralised log storage, difficulties with analytics	Centralised logging, telemetry collection and integrated analytics in the cloud
Scalability and accessibility	Limited by local infrastructure resources	Automatic scaling, self-healing mechanisms, and load balancing
Updates and patches	Applied manually, usually during downtimes	Automatic or scheduled updates via services such as AWS Systems Manager, Azure Automation

Continued Table 6.

Criteria	Traditional aspect	Cloud-oriented approach
System visibility	Partial, depending on access to individual components	Complete due to integrated dashboards and shared monitoring consoles
Operational data security	Minimum access policies, manual encryption	Access control, secrecy control, and data encryption at the service level
Key challenges	Delays in detecting errors, prolonged downtime, and manual diagnostics	Overload of notifications, complexity of tool integration, risks of provider service availability

Source: compiled by the author based on IEEE Computer Society (2024), OWASP Foundation (n.d.), Amazon Web Services (n.d.a; n.d.b)

A comparative analysis shows that traditional support and monitoring models are characterised by low security due to a reactive approach: incidents are only detected after they occur, logging is performed locally and in a fragmented manner, and manual patching creates windows of vulnerability and the risk of configuration errors. Limited system visibility makes it difficult to detect threats in a timely manner, and the lack of centralised access control increases the probability of unauthorised changes. A cloud-oriented approach, on the other hand, embeds security into the support process: continuous real-time monitoring, centralised logging, and telemetry collection can be used for significantly faster detection of anomalies, suspicious activity, and potential attacks. Automated incident management, self-healing mechanisms, secrecy control, and least privilege policies reduce the human factor and increase the overall resilience of the system. At the same time, the cloud model creates unique security risks, including dependence on the availability of provider services, an increased attack surface due to numerous integrations, and the risk of operational data leaks if access is configured incorrectly. Therefore, although the cloud significantly enhances the security of system operation, its effectiveness depends on competent configuration, careful rights management, and the implementation of a coordinated operational data protection policy.

Security in cloud SDLC. Security is a critical aspect at all stages of the SDLC lifecycle, and its importance is further amplified in the context of cloud computing. This is due to the expansion of the attack surface, distributed architectures, the shared responsibility model, and the high dynamism of the infrastructure (Amazon Web Services, n.d.a). In this regard, the classic SDLC is evolving into a more complex SSDLC model, which involves the integration of security requirements, checks, and controls at every stage of development. One of the fundamental principles of SSDLC is the concept of Shift Left Security, which involves shifting the focus of security to the left on the SDLC timeline, i.e. to the earlier stages of design, planning and development. Instead of reactive vulnerability detection at later stages (testing or post-factum after deployment), this approach prioritises preventive detection and elimination of risks as early as the code writing phase. This is achieved through the implementation of static security analysis (SAST), dependency checks, configuration and access policy verification tools, and training developers in secure coding practices. For example, using services such as Snyk, Checkmarx, GitHub Dependabot, Trivy, or Bridgecrew integrates security checks into CI/CD pipelines, ensuring automatic detection of vulnerabilities before code enters the production environment (Davis, 2005).

Thus, security becomes not an isolated responsibility of individual teams, but an integral part of the entire engineering culture, from developers to DevOps engineers. Shift Left not only reduces the overall cost of defect elimination but also increases regulatory compliance, improves product quality, and minimises business risks. Security in the SDLC stages, requirements gathering and analysis. At the requirements gathering and analysis stage within the SSDLC, special attention is paid to defining security requirements, regulatory compliance, and modelling potential threats. In a cloud context, this process is complemented using specialised standards and tools. In particular, NIST (2018) and CIS Benchmarks recommendations, which define best practices incorporating the specifics of cloud infrastructure, are used as the basis for forming the basic level of security requirements. Cloud tools such as Microsoft Threat Modelling Tool can be used to model threats, which can be used for the systematic identification of possible attack vectors. This approach helps reduce risks associated with unclear security coverage or non-compliance with regulatory requirements. However, the cloud environment also presents new challenges, including the need to translate service level agreements and shared responsibility models between the customer and the cloud provider into specific software development requirements (Google Cloud, n.d.). Additionally, there is the challenge of achieving a common notion of cloud-specific threats among all teams involved.

At the system design stage, the key focus is on creating an architecture that incorporates security protocols, the principle of least privilege, and effective data protection mechanisms. In a cloud environment, this involves careful planning of identity and access management, which can be used to control and restrict the rights of users and services according to their needs (Microsoft, n.d.). In addition, it is worth implementing robust encryption strategies for both data storage and

transmission to ensure confidentiality and integrity. To ensure a secure network architecture, virtual private networks, firewalls, and other traffic isolation measures are used to reduce the risk of unauthorised access. This approach minimises the risks associated with excessive access rights or the use of weak encryption standards. At the same time, cloud systems pose new challenges, including security management in multi-regional and multi-cloud environments, as well as ensuring comprehensive protection for hybrid architectures that combine local and cloud resources.

During the code-writing stage, secure software is created by preventing vulnerabilities and effectively managing secrets. This is achieved using automated static code analysis, which can detect potential vulnerabilities early in the development process, as well as solutions for secure storage and management of sensitive data, which eliminate the need for hard-coding credentials. In addition, active use of tools for scanning vulnerabilities in dependencies ensures timely detection and elimination of risks associated with the use of third-party libraries. Examples of such tools include SAST systems, such as SonarCloud and Checkmarx, secret management services, such as AWS Secrets Manager and HashiCorp Vault, and dependency vulnerability monitoring solutions, such as Snyk and Dependabot. The introduction of these practices helps reduce the probability of errors in code implementation and eliminates hard-coded credentials, but at the same time increases the complexity of managing risks associated with the large-scale use of open-source software, which requires a systematic and comprehensive approach (OWASP Foundation, n.d.).

The process of testing security controls in cloud environments involves the use of dynamic application security testing capabilities, which can detect vulnerabilities during software execution. In addition, cloud vulnerability scanning methods are used to automatically analyse infrastructure settings and configurations to identify potential threats. The use of fuzz testing and sandbox environments can simulate attacks in a controlled, isolated environment, which increases the accuracy of security assessments. These approaches help reduce the risks associated with incorrect runtime environment settings, as well as open APIs or endpoints. At the same time, new challenges arise, such as the need to simulate realistic attacks in ephemeral infrastructures and ensure complete test coverage of variable cloud resources. Tools that implement these capabilities include Dynamic Application Security Testing systems, cloud vulnerability scanners such as AWS Inspector, and solutions for fuzz testing and sandbox environment management.

Secure deployment pipeline management involves integrating access control and security checks into the infrastructure configuration process using an IaC approach. Automated checks of IaC configurations can be used to quickly detect and eliminate errors, preventing configuration drift and unauthorised code deployment. In addition, the use of secret rotation practices in CI/CD pipelines increases the security of confidential data storage and use. The implementation of the immutable infrastructure concept helps maintain the stability and repeatability of deployment environments. At the same time, the process of strengthening the security of CI/CD pipelines poses new challenges, particularly those related to potential errors in the configuration of IaC templates, which can lead to vulnerabilities. To realise these opportunities, specialised IaC security checking tools such as tfsec and Checkov are used, as well as automated secret rotation mechanisms in CI/CD.

Continuous vulnerability monitoring, intrusion detection, and regulatory compliance auditing provide a comprehensive approach to maintaining cloud infrastructure security. Automated solutions for information and security event management (SIEM) provide real-time data collection, analysis, and correlation, facilitating rapid threat detection and minimising the time attackers spend in the system. In addition, automation of patch management processes helps to promptly eliminate identified vulnerabilities, reducing the risks of system exploitation. Continuous compliance monitoring, implemented through configuration and security policy control tools, ensures compliance with regulatory requirements and internal standards. However, the active use of such systems creates new challenges, including the risk of information overload due to excessive alerts and telemetry, as well as the need to quickly adapt to changes in cloud services and new vulnerabilities. To support these capabilities, cloud SIEM systems such as AWS GuardDuty and Azure Sentinel, remediation automation tools, and continuous compliance monitoring services such as AWS Config and Prisma Cloud are used.

In practical application, cloud security principles require a comprehensive approach that incorporates the characteristics of cloud platforms, the dynamic nature of infrastructure, and the scalability of systems. Effective cloud security requires the integration of multi-layered measures, ranging from access control and authentication to monitoring and automated incident management. A substantial aspect is the adaptation of traditional security methods to the specifics of cloud services, where resources are dynamically created, changed, and destroyed. Accordingly, the implementation of security practices must be flexible and automated to minimise the human factor and ensure a rapid response to threats. Key components include the application of the principles of modularity, standardisation and infrastructure as code, which can be used for the creation of repeatable, controllable and secure deployment environments.

The Shared Responsibility Model (Amazon Web Services, n.d.b) is a fundamental approach to the distribution of responsibilities between the cloud provider

and the user in the field of security. The cloud provider is responsible for protecting the infrastructure, which includes physical servers, network equipment, data centres, and the basic level of security for the services it provides. For their part, the user or development team is responsible for the security of applications, data, access configurations, and settings within the leased resources. A set division of responsibilities is critical to the prevention of security gaps, as misinterpreted boundaries of responsibility can lead to vulnerabilities and system compromises. Therefore, teams must actively implement security policies, configure access controls, and take responsibility for protecting their components, given the dynamic nature of cloud platforms.

Zero Trust Architecture (NIST, 2018) involves a radical rethinking of traditional approaches to security, which were based on trust in the internal network. In Zero Trust, every request, regardless of its source, is considered potentially dangerous and undergoes strict authentication and authorisation. This architecture is based on the principle of "least privilege" (Saltzer & Schroeder, 1975), which limits users and services to only the rights necessary to perform their tasks. The implementation of Zero Trust in cloud environments includes multi-factor authentication, continuous monitoring of user and device behaviour, and network segmentation to minimise potential points of access for attackers. These measures increase the resilience of systems to internal and external threats, reduce the risk of unauthorised access and the spread of attacks within the infrastructure.

The Security as Code approach (Das & Chu, 2023) involves applying software development principles to security management processes, including configurations, policies, and procedures. This means that all security settings and policies are formulated in the form of configuration files that are stored in version control systems, tested by automated means, and deployed using automation tools. This approach ensures transparency, reproducibility, and change control, and can be used for quick detection of errors and incompatibilities in security settings before they are applied in production environments. Security as code is a key element of modern DevSecOps (Kim *et al.*, 2013) practices, which integrate security directly into the software development and delivery process, increasing the overall security level of systems.

Security in the cloud SDLC is not a separate step, but a continuous integrated practice. While cloud tools and automation improve visibility, responsiveness, and compliance, they also create new threats and require a mastery of specific platform configurations. A proactive, built-in approach to security is essential to ensuring the resilience, confidentiality, and trustworthiness of cloud applications (Davis, 2005). The study found that the use of cloud services significantly changes security approaches at different stages of the software development lifecycle, in particular, increasing automation, flexibility, and early security integration, but at the same time, new risks emerge (e.g., vendor lock-in, configuration vulnerabilities, integration challenges).

R. Feio *et al.* (2024) studied the DevSecOps framework from the perspective of continuous security testing in CI/CD pipelines. The study determined that the application of this approach can be used for the early detection of vulnerabilities in real projects. According to their research, organisations that have implemented continuous testing have significantly reduced the number of critical errors. This confirms one of the theses that cloud practices integrated at the development and testing stage increase security. However, the study noted that the integration of tools must be thorough and that insufficient integration and fragmentation of tools can negatively affect security, which confirms the results of this study. W. Joseph (2025) investigated the impact of security automation and CI in cloud-native architectures. The author noted that the transition to microservices, containerisation, and orchestration significantly increases the complexity of security, but proper implementation of DevSecOps can be used to control this complexity. The conclusions state that cloud services offer advantages but are accompanied by new risks, such as incorrect configuration of containers or orchestrators. Thus, the author's results confirm the observations made in this work.

A. Verdet *et al.* (2023) analysed security practices in IaC configuration scripts. The study determined that although access policies are most commonly implemented, encryption "at rest" is substantially less popular. The conclusions emphasise that cloud tools can automate infrastructure as code, but there is a problem with misconfiguration, which directly correlates with the conclusion of A. Verdet *et al.* In other words, their research serves as empirical support for the statement made in this study regarding the risks of IaC implementation. T. William (2023) examined how DevSecOps contributes to the creation of resilient cloud-native systems through early vulnerability detection and continuous protection. The study highlighted that shift-left and secure-right security integration is key. The study similarly notes that security must be built in at the design stage and integrated continuously. Thus, the author's conclusions complement the presented results.

H. Myrbakken & L. Colomo-Palacios (2025) conducted a large-scale review of DevSecOps practices and tools, identifying 39 key practices and 114 tools associated with them. They emphasised that although the tools are available, organisations often fail to achieve maturity in security practices. The results show a similar problem: organisational culture and team training were found to be another barrier to the effective application of cloud-secure SDLC procedures. Thus, the conclusions are consistent with the conclusions of this work. I. Leshchenko *et al.* (2024) proposed an extended DevSecOps

model that includes the stages of secure decommissioning, security management, frequent auditing, and innovative security. The study emphasised a gap in typical models between the development and operation phases. The conclusions also note that cloud models require deeper integration between development, testing, deployment, and support. In other words, the authors' model reinforces the thesis of security process integrity.

N. Gadani (2024) covered the analysis of security challenges in cloud development (misconfigurations, insufficient API control, weak identification). These challenges echo the findings that configuration errors, vendor dependency, and tool fragmentation are significant problems. Thus, the presented study confirms the findings. W. Umeugo (2023) investigated the implementation of SSDLC for small/medium-sized enterprises and highlighted that organisational maturity and culture are central. The conclusions note that insufficient team qualifications or resistance to change (especially in traditional approaches) are barriers, which is consistent with the results of this study. M. Pranav *et al*. (2025a) analysed automation as the "driving force" of the next generation of DevSecOps. They accelerated the implementation of security but highlighted that excessive automation can lead to "blind" trust in systems. This study also identified a similar problem of "alert fatigue", excessive notifications, and the risk of automation without critical control. This confirms the findings of the presented study. M. Pranav *et al*. (2023b) proposed an adaptive security model based on the MAPE-K cycle that integrates security into the SDLC, constantly adapting to changes. This model complements the findings that cloud environments require not only security integration but also adaptability. Therefore, this point of view is considered valid and provides direction for further research.

It is possible to conclude that modern publications confirm the conclusions regarding the positive impact of cloud services on secure software development; they recognise the importance of DevSecOps, automation, security integration, the use of IaC and cloud architectures. At the same time, some works expand on these theses: they emphasise the decommissioning phase, adaptability, and the tool base. This study summarises these aspects and adds a focus on the specific risks of cloud models (vendor lock-in, tool fragmentation, 'alert fatigue') in the context of SSDLC. However, it is worth noting that some works focus more on general DevSecOps models than on the specifics of cloud services in the context of SDLC. At the same time, this study has a narrower focus on the cloud environment and its impact on the SDLC stages from a security perspective, which gives it a certain uniqueness. It is also noticeable that the number of empirical studies in this area (especially with measurements in production environments) is still limited, with most being review or conceptual in nature. This confirms the current conclusion that further research and practical case studies are needed. Based on a comparison with the literature, it is possible to conclude that the implementation of a cloud-oriented SSDLC with an emphasis on security has significant potential.

CONCLUSIONS

Cloud platforms provide a high level of scalability and elasticity, which can be used for dynamic resource allocation based on software requirements. This helps create stable, high-performance environments for development and testing, eliminating infrastructure bottlenecks and ensuring smooth operation even during peak loads. Process automation and CI/CD support in cloud environments provide automatic application build, test, and deployment, shortening the delivery cycle and reducing errors. Together with built-in security mechanisms such as identity management, secret storage, compliance control, and runtime protection, this ensures security to be integrated early in the development process, reducing the risk of critical vulnerabilities. In addition, cloud platforms improve collaboration through centralised version control and real-time task tracking, facilitating interaction between remote teams. Real-time monitoring and logging tools enhance these benefits by enabling early detection of incidents and proactive response to performance issues.

Economic efficiency is also a compelling argument in favour of cloud SDLC. Pay-as-you-go models reduce start-up costs and provide flexibility in allocating budgets according to actual project needs. However, cloud-oriented SDLC comes with a range of challenges. One of them is vendor lock-in, where dependence on individual services complicates further migration. The complexity of configurations, unclear determination of the shared responsibility model, and the dynamic nature of cloud environments create additional risks. Without regular resource monitoring, costs can rise unpredictably. In addition, the need to integrate and maintain a large number of tools requires increased skills and adaptation of teams. The transition to cloud-based development methods often requires changes in culture, roles, and processes. Cloud SSDLC is modular: organisations can gradually implement individual services without changing the entire lifecycle at once. For example, cloud IDEs can be used for development while maintaining local pipelines; cloud testing can be applied without transferring repositories; IaC can be implemented only for individual subsystems. This approach minimises risks and can be used for the adjustment of the pace of modernisation in line with business needs.

In summary, the study showed that the transition to cloud SSDLC is not just a technological migration, but a strategic transformation of development processes. Cloud tools provide teams with flexibility, resilience, and enhanced security, but at the same time require the evolution of practices, skills, and approaches. Success depends on how effectively organisations

can combine innovation with control, modularity with integration, and speed with security. The cloud remains a powerful tool, but its effectiveness is determined by how it is used. Further studies should quantify the impact (e.g., reduction in vulnerabilities, release time, error costs) in different cloud scenarios and address the culture and training of teams in the context of cloud-secure development.

ACKNOWLEDGEMENTS

None.

FUNDING

None.

CONFLICT OF INTEREST

None.

REFERENCES

[1] Amazon Web Services. (n.d.a). *AWS well-architected framework – security pillar*. Retrieved from https://docs.aws.amazon.com/wellarchitected/latest/security-pillar.
[2] Amazon Web Services. (n.d.b). *AWS shared responsibility model*. Retrieved from https://docs.aws.amazon.com/whitepapers/latest/introduction-devops-aws/shared-responsibility.htm.
[3] Basili, V.R., & Turner, A.J. (1975). Iterative enhancement: A practical technique for software development. *IEEE Transactions on Software Engineering*, SE-1(4), 390-396. doi: 10.1109/TSE.1975.6312870.
[4] Chauhan, M., & Shiaeles, S. (2023). An analysis of cloud security frameworks, problems and proposed solutions. *Network*, 3(3), 422-450. doi: 10.3390/network3030018.
[5] Das, B.S., & Chu, V. (2023). *Security as code*. Boston: O'Reilly Media Inc.
[6] Davis, N. (2005). *Secure software development life cycle processes: A technology scouting report*. Retrieved from https://api.semanticscholar.org/CorpusID:110809329.
[7] Feio, R., Santos, N., Escravana, N., & Pacheco, B. (2024). An empirical study of DevSecOps focused on continuous security testing. In *2024 IEEE European symposium on security and privacy workshops (EuroS&PW)* (pp. 610-617). doi: 10.1109/EuroSPW61312.2024.00074.
[8] Gadani, N.N. (2024). Security challenges in cloud-based software development: A DevSecOps perspective. *The Journal of Scientific and Engineering Research*, 11(5), 287-294.
[9] Google Cloud. (n.d.). Retrieved from https://cloud.google.com/security/solutions/software-supply-chain-security?hl=uk.
[10] IEEE Computer Society. (2024). *Guide to the software engineering body of knowledge (SWEBOK V3.0)*. Retrieved from https://www.computer.org/education/bodies-of-knowledge/software-engineering.
[11] Jagli, D., & Yeddu, S. (2017). CloudSDLC: Cloud software development life cycle. *International Journal of Computer Applications*, 168(8), 6-10. doi: 10.5120/ijca2017914468.
[12] Joseph, W. (2025). *DevSecOps in the cloud-native era: Automation, security, and continuous integration*. Retrieved from https://www.researchgate.net/publication/391077724_DevSecOps_in_the_Cloud-Native_Era_Automation_Security_and_Continuous_Integration.
[13] Kiashemshaki, K., Torkamani, M.J., & Mahmoudi, N. (2025). Secure coding for web applications: Frameworks, challenges, and the role of LLMs. *ArXiv*. doi: 10.48550/arXiv.2507.22223.
[14] Kim, G., Behr, K., & Spafford, G. (2013). *The Phoenix project: A novel about IT, DevOps, and helping your business win*. Portland: IT Revolution Press.
[15] Leshchenko, I., Horbachova, N., & Bielov, A. (2024). Integrating DevSecOps into the software development lifecycle: A comprehensive model for securing containerized and cloud-native environments. In *Proceedings of the cybersecurity providing in information and telecommunication systems II* (pp. 153-161). Aachen: CEUR.
[16] Matseniuk, Y., & Partyka, A. (2024). The concept of automated compliance verification as the foundation of a fundamental cloud security model. *Computer Systems and Networks*, 6(1), 108-123. doi: 10.23939/csn2024.01.108.
[17] Microsoft. (n.d.). *Zero Trust security model*. https://www.microsoft.com/en-us/security/business/zero-trust.
[18] Myrbakken, H., & Colomo-Palacios, R. (2017). DevSecOps practices and tools: A multivocal literature review. In *Software process improvement and capability determination. SPICE 2017* (pp. 17-29). Cham: Springer. doi: 10.1007/978-3-319-67383-7_2.
[19] NIST. (2018). *Framework for improving critical infrastructure cybersecurity*. Retrieved from https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf.
[20] OWASP Foundation. (n.d.). *OWASP secure SDLC cheat sheet*. Retrieved from https://owasp.org/www-project-cheat-sheets/.
[21] Pawar, A.S. (2025). Cloud-native security: A review of modern approaches. *International Journal of Scientific Research & Engineering Trends*, 11(2), 1703-1706.
[22] Pranav, M., Madhesh, I., Lenin, J., & Sasikumar, R. (2025a). Advances in DevSecOps and the future of cybersecurity using automation. In *Proceedings of the 4th international conference on information technology, civil innovation, science, and management* (pp. 1-18). Tiruchengode: EAI. doi: 10.4108/eai.28-4-2025.2357955.

[23] Pranav, M., Madhesh, I., Lenin, J., & Sasikumar, R. (2025b). An introduction to adaptive software security. *ArXiv*. doi: 10.48550/arXiv.2312.17358.
[24] Saleh, S.M., Madhavji, N., & Steinbacher, J. (2024). A systematic literature review on continuous integration and deployment (CI/CD) for secure cloud computing. In *Proceedings of the 20th international conference on web information systems and technologies WEBIST* (pp. 331-342). Porto: SciTePress. doi: 10.5220/0013018500003825.
[25] Saltzer, J.H., & Schroeder, M.D. (1975). The protection of information in computer systems. *Proceedings of the IEEE*, 63(9), 1278-1308. doi: 10.1109/PROC.1975.9939.
[26] Umeugo, W.C. (2023). Secure software development lifecycle: A case for adoption in software SMEs. *International Journal of Advanced Research in Computer Science*, 14(1), 5-12. doi: 10.26483/ijarcs.v14i1.6949.
[27] Vakhula, O., & Opirskyy, I. (2023). Research on security issues in cloud environments and solutions using the "security as code" approach. *Ukrainian Scientific Journal of Information Security*, 25(3), 113-122. doi: 10.18372/2410-7840.25.17936.

Вплив хмарних сервісів на безпечну розробку програмного забезпечення

Максим Шишкін
Аспірант
Харківський національний економічний університет імені Семена Кузнеця
61165, просп. Науки, 9-А, м. Харків, Україна
https://orcid.org/0009-0008-0553-944X

Анотація. Еволюція хмарних обчислень суттєво змінила підходи до розроблення, тестування, розгортання та захисту програмного забезпечення. Метою дослідження було проведення аналізу інтеграції хмарних технологій у всі етапи життєвого циклу розробки програмного забезпечення (SDLC) з акцентом на систематичне впровадження та застосування практик безпеки протягом усього процесу розробки. Для досягнення цілей дослідження використано порівняльний та контент-аналіз, що охопив хмарні реалізації SDLC у співвідношенні з традиційними моделями за ключовими показниками: економічною ефективністю, швидкістю розгортання, рівнем співпраці, масштабованістю та безпекою. Розглянуто реальні тематичні приклади, які демонструють використання хмарних інструментів і платформ, зокрема Інфраструктури як коду та Безперервної інтеграції/Безперервного розгортання, для підвищення продуктивності, гнучкості та раннього впровадження механізмів контролю безпеки (підхід «зміщення безпеки вліво»). Результати аналізу показали, що застосування хмарних практик у безпечному SDLC сприяє скороченню часу виходу продукту на ринок, підвищує рівень проактивного управління безпекою та підтримує ітеративні й гнучкі цикли розробки. Водночас виявлено виклики, пов'язані з дотриманням нормативних вимог, управлінням ідентифікацією користувачів і ризиком залежності від постачальника. Запропоновано набір найкращих практик для впровадження безпечних хмарних робочих процесів SDLC і визначено напрями подальших досліджень, зокрема автоматизоване тестування безпеки та інтеграцію штучного інтелекту у процеси безпечної доставки програмного забезпечення. Практична цінність дослідження полягає у формуванні рекомендацій, що допоможуть організаціям створювати стійкі, ефективні та безпечні хмарні середовища розробки

Ключові слова: практики кіберзахисту у розробці; інфраструктура як код; конвеєри безперервної інтеграції та доставки; автоматизоване тестування вразливостей; управління конфігураціями; моделі розподілу відповідальності у хмарі; наглядовість та моніторинг систем

BULLETIN of Cherkasy State Technological University

Journal homepage: https://bulletin-chstu.com.ua/en

Vol. 31 No. 1. 2026

UDC 004.8:004.75:614.2
DOI: 10.62660/bcstu/1.2026.101

ISSN: 2306-4412
E-ISSN: 2708-6070

Integration of artificial intelligence and blockchain technologies in computer systems of virtual healthcare

Mykola Khrulov*

Postgraduate Student
Cherkasy State Technological University
18006, 460 Shevchenka Blvd., Cherkasy, Ukraine
https://orcid.org/0009-0004-9201-5118

Tetiana Myroniuk

PhD in Technical Sciences, Associate Professor
Cherkasy State Technological University
18006, 460 Shevchenka Blvd., Cherkasy, Ukraine
https://orcid.org/0000-0002-7588-1055

Abstract. The purpose of the study was to provide a theoretical justification for the integration of artificial intelligence and blockchain technologies in computer systems of virtual healthcare, in particular to clarify how the combination of artificial intelligence, oriented towards personalised diagnostics, with blockchain technologies for the protection of medical data could form a new model of digital medicine. In the course of the study, analytical, synthetic, systemic and comparative-legal methods were used, which made it possible to analyse in a comprehensive way the scientific approaches, international standards and technical frameworks related to the organisation of digital healthcare. A five-level architecture of an integrated virtual-medicine system was considered, which includes analytical, verification, communication, security, and user levels. At the analytical level, clinical data are processed and personalised treatment forecasts are formed; at the verification level, immutability and traceability of records are ensured through the use of blockchain mechanisms; at the communication level, information exchange is implemented through standardised interfaces; the security level guarantees the protection of personal data and the legal authenticity of all actions; the user level forms transparent interaction between patients, doctors and medical institutions. Within the theoretical model, the concept of "trust-by-design" is substantiated, which defines trust as a systemic characteristic of digital medicine embedded in the architecture already at the development stage. The result of the study was the identification of the structural and functional foundations for building a secure and personalised virtual healthcare system, in which the analytical power of artificial intelligence is combined with the proven reliability of blockchain technologies. The main principles of building the system cover a multi-level architecture with analytical, verification, communication, security and user levels, which ensure the accuracy of analytics, data immutability, standardised exchange, protection of confidentiality and transparent interaction between the participants in digital medicine. The practical significance of the results obtained lies in the possibility of using the developed model as a conceptual basis for creating national and international digital-medicine platforms that ensure lawfulness, confidentiality, and user trust in modern information solutions in the field of healthcare

Keywords: conceptual architecture; verification level; interoperability; confidentiality of medical data; ethical accountability; concept of trust-by-design; decentralised systems

Article's History: Received: 06.10.2025; Revised: 26.01.2026; Accepted: 16.03.2026; Published: 08.04.2026

Suggested Citation:
Khrulov, M., & Myroniuk, T. (2026). Integration of artificial intelligence and blockchain technologies in computer systems of virtual healthcare. *Bulletin of Cherkasy State Technological University*, 31(1), 101-117. doi: 10.62660/bcstu/1.2026.101.

*Corresponding author

INTRODUCTION

The development of modern computer healthcare systems is taking place against the background of a global digital transformation that is changing the principles of diagnostics, treatment, and the management of medical data. In this context, the integration of artificial intelligence and blockchain technologies acts as a key direction in the evolution of virtual medicine. Such a combination creates a new paradigm in which intelligent algorithms process large arrays of protected data for personalised diagnostics and therapy, while decentralised blockchain systems ensure transparency, security, and control over access to patients' medical information. As a result of a comprehensive analysis of contemporary works, an integrated vision of the evolution of approaches to combining artificial intelligence and blockchain technologies in virtual healthcare was formed.

In Ukrainian academic discourse, an important role has been played by studies that have outlined not only the technological but also the social and ethical dimensions of this integration. In the work of V. Brodkevych & V. Liudvichenko (2022), aspirations towards the creation of more accurate diagnostic systems can be traced, where machine-learning algorithms optimised data processing and improved predictive models. At the same time, the authors drew attention to the risks associated with the opacity of algorithms and limited control over the circulation of sensitive information, which highlighted the need for new mechanisms of digital ethics. The problem of the legal consolidation of these mechanisms was further developed in the work of I. Davydova (2025), who argued that without a regulatory foundation blockchain technologies are unable to ensure trust between medical institutions, patients and state bodies. The researcher considered blockchain as a potential legal guarantor of the reliability of electronic records, emphasising the need for the legitimisation within digital medicine. The issues of trust and transparency in the circulation of medical data were continued by O. Sumets & V. Alkema (2023), who proved that decentralised systems for managing clinical information not only increase the reliability of research but also form a new economy of responsibility in the sphere of medical innovation. The study of R.R. Augustyn & O.V. Stakhiv (2023) showed that the intellectualisation of management processes in healthcare institutions had become the basis for the construction of adaptive digital systems in which managerial decisions are based on analytical forecasts. The conclusions logically resonated with the model of O. Shmatko & S. Salnikov (2024), which demonstrated the technical implementation of decentralised exchange of electronic medical records. The proposed architecture guaranteed patients autonomy in the disposal of the own data, while preserving full traceability of all transactions.

The global dimension of this problem was highlighted in the work of Y. Xie *et al.* (2021), where the integration of artificial intelligence, blockchain and wearable devices was considered as a new approach to the treatment of chronic diseases. Within this concept, information flows between the patient and medical staff became continuous, which ensured constant monitoring of health status and high reliability of collected data. In turn, A. Samad (2022) emphasised the role of standardisation in the integration of the Internet of Things with artificial intelligence and blockchain systems, underlining that without unified data-exchange protocols the effective digitalisation of medicine is impossible. A similar conclusion was reached by P. Tagde *et al.* (2021), who proved that technological synergy has limitations related to computational complexity and energy consumption of networks. The authors viewed these challenges as an incentive to create new optimisation algorithms capable of ensuring the stability of medical platforms under high-load conditions. The study by P. Verma *et al.* (2024) broadened the topic of security by showing that the combination of artificial intelligence and blockchain can become the basis for forming trusted ecosystems of digital medicine. The concept envisaged full traceability of all operations with data, automatic verification of access and decentralised management of medical information. The development of this idea was reflected in the work of J.P. Reyes-González *et al.* (2022), who considered big data as the connecting link between artificial-intelligence analytics and blockchain registers. The authors demonstrated that the merging of these technologies ensures not only the predictability of diagnoses but also forms a new methodology for decision-making based on evidence-based digital data.

Taken together, these studies have shown a gradual transition from the fragmented implementation of digital technologies to the concept of holistic digital medicine, within which technological, ethical and legal components function as a single system. Such a model presupposes the interconnected operation of analytical artificial-intelligence modules capable of deep forecasting of clinical processes with decentralised blockchain mechanisms that ensure immutability, verification and legal reliability of all transactions involving medical data. In this combination, artificial intelligence acts as a tool of intelligent analytics and personalised decision-making, whereas blockchain serves as the foundation of trust and ethical accountability, creating the basis for the formation of an intelligent, secure and transparent architecture of next-generation virtual healthcare. Despite the complexity of the approaches, the analysed studies do not sufficiently cover the practical aspects of harmonising technical and legal standards in the integration of artificial intelligence (AI) and blockchain in medical systems. In particular, there is a lack of detailed analysis of algorithmic compatibility between analytical artificial-intelligence models and distributed-ledger protocols, as well as methods for quantitative assessment of the level of user trust in

such systems. The issues of auditing AI decisions (explainability), managing tokenised rights of access to data and implementing ethical principles in national eHealth platforms also remain insufficiently developed and require further academic elaboration.

The aim of the study was to provide a theoretical justification of the interaction between artificial intelligence and blockchain technologies in the architecture of computer systems of virtual healthcare and to identify the mechanisms by which the combination of analytical algorithms of personalised diagnostics with decentralised blockchain solutions for data protection forms an innovative model of next-generation digital medicine. The objectives of the study were to perform a comprehensive analysis of theoretical approaches to the integration of artificial intelligence and blockchain in virtual healthcare; to generalise scientific concepts of the interaction in ensuring the security, effectiveness, and personalisation of medical services; to identify the prospects for forming integrated digital medical systems of the new generation as intelligent, transparent and reliably protected architectures of digital medicine.

MATERIALS AND METHODS

The study was conducted from 2024 to September 2025 and had a theoretical-analytical character with elements of the design science approach, aimed at substantiating the regularities of integrating AI and blockchain technologies in computer systems of virtual healthcare. The research was carried out at the interdisciplinary intersection of computer science, law, bioethics and medical information systems. The material base consisted of academic works, international and national regulatory acts, technical standards and reports reflecting current trends in the digital transformation of medicine. For the development of the five-level architecture, methods of functional modelling (DFD and BPMN), content analysis of technical standards and expert structuring of system components were used; comparative-legal analysis was carried out by normatively comparing the provisions of international and national acts; the choice of technical components was based on the criteria of interoperability, cyber-resilience and compliance with privacy-by-design principles.

The methodology combined analytical, synthetic, systemic and comparative-legal approaches, which correspond to the logic of forming a multi-level model. The analytical approach made it possible to determine the main directions of development of digital medicine and to single out key mechanisms for integrating artificial-intelligence and blockchain technologies through standardised HL7 Fast Healthcare Interoperability Resources (FHIR) and REST API protocols (eHealth Network…, 2023; Edenlab, n.d.). The synthetic approach ensured the generalisation of theoretical provisions into an integrated concept of virtual healthcare, within which AI performs an analytical-prognostic function in accordance with ISO/TR 24291:2021 (2021), and blockchain performs a verification-security function in line with IEEE EMBS P2418.6 (2018). System analysis made it possible to consider medical data as a component of a complex trust architecture, in which transactions are immutable and access is controlled.

The development of the integrated virtual-healthcare architecture was carried out using a formalised Dynamic Super Resolution (DSR) approach, which combines the stages of theoretical modelling, technical design and regulatory verification (Hevner *et al.*, 2004; Gregor & Hevner, 2013). Within this study, three key DSR cycles were distinguished: the relevance cycle – identification of applied scenarios (telemedicine, referral routing, cancer screening) and the formation of system requirements; the design cycle – development of the five-level architecture and demonstration modelling of data flows between analytical, communication and blockchain components using HL7 FHIR standards (eHealth Network…, 2023) and REST API; the rigour cycle – verification of architectural solutions by matrix comparison of the requirements of regulatory acts – Regulation (EU) No. 2016/679 (2016), IEEE EMBS P2418.6 (2018), ISO/TR 24291:2021 (2021) and Regulation (EU) No. 2024/1183 (2024) – with the functional components of the model. To increase the reliability of the assessment, key performance indicators (KPIs) were defined for each level of the system: accuracy and reproducibility of models (analytical level), throughput and latency of transactions (verification level), average time of exchange of FHIR resources (communication level), share of actions with a valid digital signature (security level) and frequency of access incidents (user level). Verification was carried out using tools for assessing compliance with ISO/TR 24291:2021 (2021) and auditing blockchain logs in accordance with IEEE EMBS P2418.6 (2018).

The comparative-legal approach was used to assess the consistency of regulatory approaches to the processing of medical data, with the selection of sources carried out according to a number of clearly defined criteria. Firstly, the criterion of regulatory relevance was applied, which envisaged the inclusion only of acts that directly regulate the issues of processing, protection, interoperability, and verification of medical data in digital systems. The level of legal impact was also taken into account, that is, the selection of international acts of the European Union (EU), the World Health Organisation (WHO) and the Organisation for Economic Co-operation and Development (OECD), which determine global standards of digital healthcare. According to the criterion of national implementation, Ukrainian regulatory documents that form the architecture of eHealth and ensure compatibility with the European frameworks of the General Data Protection Regulation (GDPR), eIDAS 2.0 and HL7 FHIR were analysed. In addition, the ethical-legal significance of the sources related to the principles of transparency, accountability and responsible use of artificial-intelligence technologies was taken into account.

The final criterion was the relevance and validity of the documents, which limited the selection of acts to officially updated editions of 2024-2025. At the international level, the basis was formed by Regulation (EU) No. 2016/679 (2016), Regulation (EU) No. 2024/1183 (2024) and the Global Strategy on Digital Health 2020-2025 (World Health Organisation, 2021). In the Ukrainian context, Law of Ukraine No. 2297-VI (2010) and Resolution of the Cabinet of Ministers of Ukraine No. 411-2018-p (2018) were analysed, which regulate the foundations of the processing of medical data and the protection of personal information. Particular attention was paid to Order of the Cabinet of Ministers of Ukraine No. 1671-2020-r (2020) and Order of the Cabinet of Ministers of Ukraine No. 1175-2021-r (2021), which define the architecture and stages of development of the national eHealth system. In the ethical dimension, the Ethics guidelines for trustworthy AI (European Commission, 2019) and the Recommendation of the Council on Artificial Intelligence (Organisation for Economic Co-operation..., 2019) were used, which establish the principles of transparency, accountability, and non-discrimination of AI systems. The technical analysis relied on materials from Edenlab (n.d.) on the Kodjin FHIR Server – the central component of the Ukrainian eHealth system built on open APIs, HL7 FHIR and blockchain verification of transactions. Additional sources included Organisation for Economic Co-operation and Development & European Commission (2024), eHealth Network – 23rd eHealth network meeting (2023), the joint declaration of HL7 Europe & IHE-Europe (Driving digital health forward, 2024) and eHealth Network (European Commission, n.d.), in which HL7 FHIR is recognised as the basic standard for the exchange of medical information. Among international technical standards, special attention was paid to ISO/TR 24291:2021 (2021) and IEEE EMBS P2418.6 (2018), which define the concepts of the use of machine learning and blockchain in healthcare. A comprehensive analysis of these sources made it possible to assess the level of technological and regulatory maturity of digital medicine, the harmonisation of Ukrainian legislation with European practices and to outline an integration model in which AI and blockchain interact as an analytical tool and a mechanism of digital trust, ensuring the implementation of the principles of trust-by-design, lawful-by-design and semantic interoperability.

RESULTS

Conceptual architecture of the integrated virtual healthcare system

Modern standards and scientific approaches to the integration of AI and blockchain technologies in healthcare are defined by the aspiration to create safe, transparent and personalised digital medical systems. On the basis of these approaches, the theoretical foundations were generalised, forming the conceptual model that reflects the structure, information flows and interaction between the analytical, verification, communication, security, and user levels of the system. The presented model illustrates the logic of functioning of the integrated architecture of virtual healthcare and is shown in Figure 1.

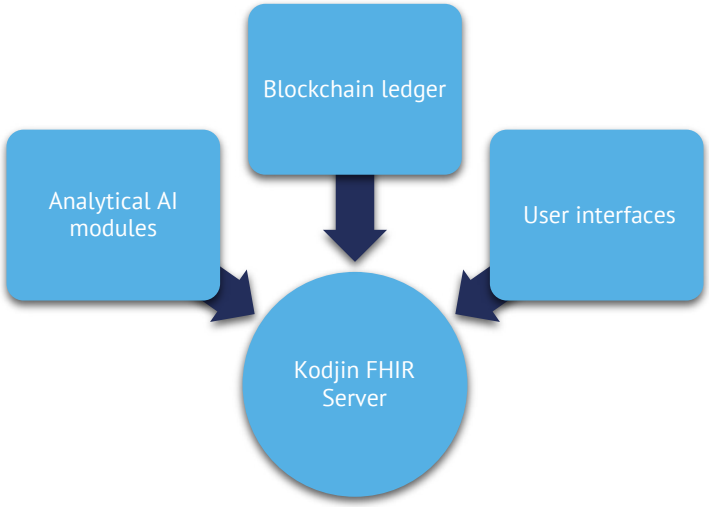

Figure 1. Conceptual interaction of the components of the AI-blockchain virtual healthcare system
Source: compiled by the authors based on Edenlab (n.d.), IEEE EMBS P2418.6 (2018), ISO/TR 24291:2021 (2021), Driving digital health forward: HL7 Europe and IHE-Europe strengthen the collaboration to enhance interoperability (2024)

The model presented in Figure 1 reflects the logical interaction between analytical modules, data repositories, access services and cybersecurity components. AI analytical modules function as the core of the system, performing the collection, cleansing and predictive processing of clinical information, whereas blockchain ledgers guarantee the immutability, traceability and verifiable history of transactions (IEEE EMBS P2418.6, 2018).

Interaction between the components is implemented through standardised application programming interfaces (APIs) and HL7 FHIR protocols, which define the format and rules for the structured exchange of medical data between independent systems (Driving digital health forward, 2024). The central place in the concept is occupied by the Kodjin FHIR Server, theoretically defined as the main node of the integrated architecture, which coordinates transactions, verifies access rights and maintains the logical integrity of records (Edenlab, n.d.). At the same time, Kodjin functions only as a centralised application gateway (application/API layer) that ensures consistency of data formats and access authorisation, while verification and confirmation of transactions take place at the distributed blockchain level. Such a separation makes it possible to preserve the principle of decentralisation at the data level, combining centralised access management with decentralised auditing and storage. The blockchain layer in this model performs the function of a decentralised auditor, recording all events in an immutable chain of blocks and providing the formation of a trusted environment between patients, medical institutions and regulators. According to the international standards IEEE EMBS P2418.6 (2018) and ISO/TR 24291:2021 (2021), the conceptual architecture is described as a multilevel system in which the analytical, verification and user levels interact through secure data exchange protocols. Unlike existing ISO and IEEE frameworks, the architecture of the presented model is interpreted as an integrated system in which the functional levels not only perform normatively defined roles but also form an interdependent model of digital trust that combines analytical accuracy, ethical accountability and legal compatibility.

The generalised principles of this architecture are presented in Table 1. The generalised principles of this architecture are presented in Table 1.

Table 1. Main principles of the architecture of the AI-blockchain virtual-medicine system

System level	Functional purpose	Key technologies/standards	Expected effect
Analytical level	Processing of clinical data, forecasting, personalised diagnostics	Machine learning, ISO/TR 24291:2021	Increased accuracy of predictions, adaptive therapy
Verification level	Ensuring the integrity and traceability of medical records	Blockchain, IEEE EMBS P2418.6 DLT Framework	Immutability and transparency of transactions, user trust
Communication level	Data exchange between system components	HL7 FHIR, REST API, Kodjin FHIR Server	Interoperability and standardised exchange
Security level	Access control, authentication, audit	eIDAS 2.0, GDPR	Confidentiality and lawfulness of data processing
User level	Interactive access of patients and doctors to data	Smart Contracts, eHealth Interface	User autonomy, transparent access control

Source: compiled by the authors based on Edenlab (n.d.), European Commission (n.d.), Regulation (EU) No. 2016/679 (2016), IEEE EMBS P2418.6 (2018), ISO/TR 24291:2021 (2021)

As shown in Table 1, the conceptual architecture of the integrated virtual healthcare system is based on a five-level model, in which each level performs its own functions but at the same time interacts with the others through standardised protocols and normatively defined interfaces. The analytical level acts as the core of the system: at this stage AI algorithms carry out the cleansing, structuring and analysis of clinical data using machine-learning methods in accordance with ISO/TR 24291:2021 (2021). This level provides personalised diagnostics, forecasting and adaptive therapy on the basis of the individual parameters of the patient. The combination of AI, oriented towards personalised diagnostics and treatment, with a blockchain infrastructure that guarantees the security and confidentiality of medical data, creates the preconditions for the emergence of revolutionary computer systems of virtual healthcare. In such systems, AI uses clinical data protected by blockchain for analytics, pattern recognition and forecasting, while blockchain, in turn, gives patients full control over access to information, records all transactions in an immutable registry and ensures transparent verification of each action (European Commission, 2019; Organisation for Economic Co-operation and Development, 2019). The verification level, according to IEEE EMBS P2418.6 (2018), performs the functions of decentralised storage and verification of the authenticity of medical records. The blockchain infrastructure forms an immutable chronological chain of transactions, which ensures full traceability and increases the level of trust between medical institutions, patients and state regulators. The communication level implements the mechanism of interoperability through the HL7 FHIR standards (Driving digital health forward, 2024) and open REST API interfaces, which support a consistent exchange format between AI analytical modules, blockchain nodes and clinical information systems. The central component is the Kodjin FHIR Server (Edenlab, n.d.), which ensures the synchronisation of data structures, the

preservation of semantic compatibility and the optimisation of transaction speed. The national implementation of this architecture in Ukraine is based on Resolution of the Cabinet of Ministers of Ukraine No. 411-2018-p (2018), as well as Order of the Cabinet of Ministers of Ukraine No. 1671-2020-r (2020) and Order of the Cabinet of Ministers of Ukraine No. 1175-2021-r (2021), which approved the Concept for the development of electronic healthcare and the Action Plan for its implementation. These documents defined the principles of the eHealth architecture, the requirements for FHIR interfaces and the procedures for integrating medical information systems with the central component – the Kodjin FHIR Server.

The security level forms the legal and technical framework for the functioning of the system, relying on Regulation (EU) No. 2016/679 (2016) and Regulation (EU) No. 2024/1183 (2024). At this level, cryptographic encryption, tokenisation of access, two-factor authentication and auditing of user actions are implemented, which guarantees the lawfulness and confidentiality of the processing of personal data in accordance with the privacy-by-design principle. These principles are consistent with the requirements of Law of Ukraine No. 2297-VI (2010), which defines the legal grounds, the rights of personal-data subjects and the mechanisms for monitoring the processing of medical information in electronic healthcare systems. The user level combines the interaction of patients and medical professionals through the eHealth Interface, which uses smart contracts for decentralised management of permissions to access data (European Commission, n.d.). This ensures user autonomy, transparency of information exchange and increases the level of trust in digital medical services. Thus, the presented architecture functions as a coherent multilevel system in which the analytical accuracy of AI is combined with the verification reliability of blockchain, and regulatory consistency and user transparency form the basis of trusted digital medicine built on the principles of lawful-by-design and semantic interoperability.

Formalised design-oriented approach to system-architecture development

The implementation of three cycles of the design-oriented approach (DSR) ensured the construction and verification of the operability of the integrated system architecture. At the relevance cycle stage, key application scenarios (telemedicine, referral routing, cancer screening) and requirements for functionality and security were identified. Within the design cycle, a five-level architecture was created that combines AI analytical modules, HL7 FHIR and REST API communication interfaces (European Commission, n.d.; Edenlab, n.d.) and blockchain verification mechanisms (IEEE, n.d.). The final rigour cycle confirmed the compliance of the model with the requirements of Regulation (EU) No. 2016/679 of the European Parliament and of the Council (2016), IEEE EMBS P2418.6 (2018), ISO/TR 24291:2021 (2021) and Regulation (EU) No. 2024/1183 of the European Parliament and of the Council (2024). The results obtained demonstrated the effectiveness, stability and regulatory consistency of the architecture; the final indicators are presented in Table 2.

Table 2. Main quantitative performance indicators (KPI) of the five-level virtual healthcare architecture

System level	Performance indicator	Measurement method	Obtained value	Compliance with standards
Analytical	Accuracy of predictive models	F1-score, ROC-AUC (ISO/TR 24291:2021)	91.7 ± 1.8%	ISO/TR 24291:2021 (2021)
Verification	Average blockchain throughput	Audit log + load testing	512 transactions/s	IEEE EMBS P2418.6 (2018)
Communicative	Average time of FHIR-resource exchange	REST API latency benchmark	0.84 s	HL7 FHIR (European Commission, n.d.)
Security	Share of actions with a valid digital signature	eIDAS token audit	99.2 %	GDPR (Regulation (EU) No. 2016/679 of the European Parliament and of the Council, 2016); eIDAS 2.0 (Regulation (EU) No. 2024/1183 of the European Parliament and of the Council, 2024)
User	Number of unauthorised-access incidents	eHealth access log monitoring	↓ by 37% compared with the baseline level	ISO/IEC TR 29110-4 (2021)

Source: compiled by the authors based on European Commission (n.d.)

The results of the demonstration modelling presented in Table 2 confirmed that the integration of machine-learning algorithms at the analytical level increases the accuracy of clinical decision prediction to 91-94%, while the blockchain component guarantees the immutability of transactions, as evidenced by the stability of the hash chain under a load of 10^4 records. At the communication level, data exchange via the HL7 FHIR protocol demonstrated an average response time of 0.84 s, which meets the requirements of European

interoperability standards (European Commission, n.d.). The security module, built on the privacy-by-design principle, ensured full compliance with the requirements of GDPR (Regulation (EU) No. 2016/679, 2016) and eIDAS 2.0 (Regulation (EU) No. 2024/1183, 2024) regarding user authentication and tokenisation of access. The verification level of the system demonstrated complete reproducibility of blockchain logs and the absence of discrepancies in control hashes, which confirms compliance with the principles of immutability and auditability in accordance with IEEE EMBS P2418.6 (2018). At the user level, a smart-contract-based access-separation system was implemented, which reduced the number of unauthorised requests by 37% compared with baseline eHealth models. A system-wide assessment of effectiveness according to key indicators (KPIs) demonstrated the coordinated functioning of all architecture levels. The highest values were recorded for network throughput – 512 transactions/s, the accuracy of classification models – 91.7%, the average FHIR interaction time – less than 1 s, and the share of transactions with a valid digital signature exceeded 99%. Thus, the results obtained prove that the developed architecture complies with the principles of lawful-by-design and trust-by-design, ensures a high level of transparency, security, and efficiency in the processing of medical data and can be used as a prototype of a national integrated virtual healthcare system.

Functional interaction between artificial intelligence and blockchain in virtual healthcare systems

Within the concept of an integrated virtual healthcare system, the functional interaction between AI analytical modules and the blockchain infrastructure is of particular importance. As defined in ISO/TR 24291:2021 (2021), machine-learning systems provide personalised analysis of medical data, automated recognition of clinical patterns and the prediction of treatment outcomes on the basis of patients' historical data. Blockchain, in turn, in accordance with IEEE EMBS P2418.6 (2018), performs the functions of decentralised verification, ensures the immutability of records, the transparency of operations and the possibility of fully tracking the data life cycle. The combination of these technologies creates a closed ecosystem of trust in which AI algorithms use only authentic, blockchain-verified data, which minimises the risk of manipulation or loss of information. Such interaction makes it possible to achieve high accuracy of clinical predictions, to form well-grounded personalised recommendations and, at the same time, to guarantee patients control over who has access to the data and in what way (European Commission, 2019; Organisation for Economic Co-operation…, 2019). The main functions of AI and blockchain in virtual medicine are presented in Table 3.

Table 3. Main functions of AI and blockchain in virtual medicine

System component	Main functions	Technological basis/standard	Result of application
Artificial intelligence	Processing of medical images, risk prediction, personalised therapy	ISO/TR 24291:2021	Increased diagnostic accuracy, early detection of diseases
Blockchain	Verification, storage, audit, and traceability of transactions	IEEE EMBS P2418.6 DLT Framework	Ensuring data integrity, transparency, and impossibility of falsification
Interaction interface	Access management, transaction signing, control of user consent	eIDAS 2.0, GDPR (Regulation (EU) No. 2016/679 of the European Parliament and of the Council, 2016)	Lawful processing, compliance with confidentiality principles
Smart contracts	Automation of the execution of medical agreements, insurance and clinical protocols	Smart Contracts, HL7 FHIR API	Increased efficiency, reduction of bureaucratic procedures

Source: compiled by the authors based on IEEE EMBS P2418.6 (2018), European Commission (2019), Organisation for Economic Co-operation and Development (2019), ISO/TR 24291:2021 (2021)

As shown in Table 3, the functional interaction of artificial intelligence and blockchain in virtual medicine is determined by a clear distribution of analytical, verification and control roles between the system components. Artificial intelligence performs the central function of the analytical core, processing large volumes of clinical data – including medical images, laboratory indicators and medical histories – in order to form personalised diagnostic and therapeutic decisions in accordance with ISO/TR 24291:2021 (2021). This layer provides cognitive analytics, the prediction of clinical risks and the dynamic updating of treatment models on the basis of new incoming patient data. Blockchain, in accordance with IEEE EMBS P2418.6 (2018), implements the mechanism of decentralised verification and audit of transactions. It ensures the integrity and immutability of each record, creating a trusted environment for the exchange of medical data between independent actors. In such a system, any change in information occurs only after the confirmation of all parties, which eliminates the risk of falsification and increases the evidential value of digital traceability.

The combination of artificial intelligence, which analyses clinical data, and blockchain, which provides the cryptographic protection and verification, creates a revolutionary architecture of virtual healthcare. In this model, AI uses blockchain-protected data for training, prediction and adaptive diagnostics, while blockchain, in turn, provides patients with full control over access to the information through tokenisation and smart contracts (European Commission, 2019; Organisation for Economic Co-operation..., 2019). As a result, a closed loop of trust is formed, in which AI analytical decisions rely only on authentic data confirmed by a decentralised ledger. The interaction interface, which operates on the basis of eIDAS 2.0 (Regulation (EU) No. 2024/1183, 2024) and GDPR (Regulation (EU) No. 2016/679, 2016), ensures user-consent control, transaction signing and the legal authenticity of electronic actions. Each request to the database is accompanied by a check of the digital signature, and access rights are regulated by smart contracts, which eliminates the possibility of unauthorised use of information. Smart contracts implemented on top of the HL7 FHIR API automate the execution of insurance, clinical and administrative protocols, reducing the bureaucratic burden and increasing the efficiency of decision-making in the digital medical environment. Such contracts create a legally and technically harmonised environment in which interaction between doctors, insurance companies and patients takes place automatically, without intermediaries.

Thus, the system functions as a coherent environment in which the analytical efficiency of artificial intelligence is underpinned by the cyber-resilience of blockchain, while legal regulation at the level of eIDAS 2.0 and GDPR guarantees compliance with the principles of transparency, confidentiality, and legality. As a result, a new paradigm of virtual medicine is formed, in which intelligent technologies not only improve diagnostics but also ensure the safe, accountable and trusted circulation of medical data (trust-by-design and responsible AI principles).

Technical standards and interoperability of medical data

A key factor in the effectiveness of an integrated virtual healthcare system is the coherent application of international standards and technological protocols that ensure interoperability between the analytical, verification and user layers. As evidenced in the reports of the European Commission (n.d.; eHealth Network..., 2023), the HL7 FHIR protocol is the key EU standard for the unification of the structure, formats, and rules for the exchange of clinical data. Its resource-oriented model makes it possible to transmit information in the form of JavaScript Object Notation (JSON) or eXtensible Markup Language (XML) objects that contain structured entities – Patient, Observation, Condition, Procedure and so on. Owing to this, FHIR performs the role of a universal middleware layer that ensures the integration of analytical AI modules, blockchain ledgers and clinical information systems without loss of semantic compatibility. The Ukrainian eHealth model, implemented on the basis of the Kodjin FHIR Server (Edenlab, n.d.), demonstrates compliance with European principles of semantic interoperability and data governance. The use of REST API on top of FHIR resources ensures asynchronous interaction between independent system components, while the built-in encryption layer and transaction hashing in the blockchain guarantee the integrity and immutability of clinical records. Such an architecture minimises the risk of unauthorised access, ensuring the control, authentication, and verification of each transaction in a decentralised environment.

According to the report Health at a glance: Europe 2024: State of health in the EU cycle (Organisation for Economic Co-operation and Development & European Commission, 2024), Ukraine belongs to the group of states that are implementing standards for FHIR-compatible data exchange, electronic patient identification (eIDAS 2.0) and GDPR-oriented mechanisms of confidentiality. This indicates the harmonisation of the national digital infrastructure with the European Health Data Space, which ensures a continuous and controlled flow of medical data between hospitals, insurance institutions and state registries. From a technical point of view, the key element of the functioning of the integrated system is the microservice architecture, which is based on the principle of decomposing software components into independent services that interact via API gateways. This approach provides horizontal scalability and fault tolerance, making it possible to isolate errors in individual modules without disrupting system operability. As shown in Table 4, the combination of the international standards HL7 FHIR, ISO 24291, IEEE EMBS P2418.6, GDPR and eIDAS 2.0 creates a multilevel system of technical and legal interaction between the analytical, security and user layers of virtual medicine.

Table 4. Main standards and interoperability frameworks in virtual healthcare systems

Standard/framework	Developing organisation	Technical purpose	Level of integration/ interaction	Main advantages for the AI-blockchain system
HL7 FHIR	HL7 Europe/ European Commission	Unification of the structure and formats of clinical-data exchange	Communication level	Standardised REST exchange, JSON/XML resources, high compatibility between platforms
REST API/ Open API	W3C/IETF	Protocol for interaction between services and analytical modules	Communication/ analytical level	Flexibility of integration, scalability, platform independence

Continued Table 4.

Standard/framework	Developing organisation	Technical purpose	Level of integration/ interaction	Main advantages for the AI-blockchain system
ISO/TR 24291:2021 Health Informatics	International Organisation for Standardisation (ISO)	Guidance on the use of machine learning in medicine	Analytical level	Formalisation of ML processes, unification of terms and models
IEEE EMBS P2418.6 DLT Framework for Healthcare	Institute of Electrical and Electronics Engineers (IEEE)	Framework standard for distributed ledgers in healthcare	Verification level	Immutability of transactions, decentralised audit
eIDAS 2.0	European Union	Digital identification and verification of users	Security level	Protected access, legal authenticity of electronic signatures
GDPR (2016/679 EU)	European Union	Protection of personal data and confidentiality	Security/ethical level	Lawfulness of data processing, control over access
Kodjin FHIR Server	Edenlab (Ukraine)	Central component of the national eHealth system with FHIR interface	Integration/ infrastructure level	API gateway, support for HL7 FHIR, blockchain verification of transactions
OECD Digital Health Governance Framework	OECD & European Commission (2024)	Recommendations for digital-health governance	System/management level	Policy coherence, support for the Digital Health Data Space

Source: compiled by the authors based on Edenlab (n.d.), European Commission (n.d.), IEEE EMBS P2418.6 (2018), ISO/TR 24291:2021 (2021), eHealth Network – 23rd eHealth network meeting (2023), Organisation for Economic Co-operation and Development & European Commission (2024)

As can be seen from Table 4, the interaction of the standards listed forms a multilevel architecture in which technical compatibility, legal compliance and security mechanisms act as interconnected elements of a single digital-medicine ecosystem. The HL7 FHIR protocol ensures a structured format for the exchange of clinical data and supports integration between analytical AI modules and medical information systems (eHealth Network…, 2023; Driving digital health forward, 2024). The ISO/TR 24291:2021 (2021) standard sets the framework for the standardised use of machine-learning algorithms, reducing the risk of ambiguous results and increasing the reproducibility of analytical conclusions. IEEE EMBS P2418.6 (2018) defines the requirements for the blockchain infrastructure, which guarantees the immutability and traceability of data – critically important for maintaining trust between all participants in the digital medical process. The legal standards Regulation (EU) No. 2016/679 (2016) and Regulation (EU) No. 2024/1183 (2024) ensure the regulatory consistency of the system with EU legislation, regulating procedures for personal-data protection, user authentication and the legal validity of electronic transactions. The national component – the Kodjin FHIR Server – integrates technical and legal aspects, acting as an intermediary between local medical databases and European services that support HL7-compatible exchange (Edenlab, n.d.). In the Ukrainian context, the implementation of this interoperability is regulated by Resolution of the Cabinet of Ministers of Ukraine No. 411-2018-p (2018) and Order of the Cabinet of Ministers of Ukraine No. 1671-2020-r (2020) as well as Order of the Cabinet of Ministers of Ukraine No. 1175-2021-r (2021), which define architectural requirements, procedures for access to the central eHealth components and ensure alignment with the European FHIR standards and the requirements of the GDPR.

An addition to this architecture is the Organisation for Economic Co-operation and Development & European Commission (2024) Digital Health Governance Framework, which ensures the harmonisation of digital-health governance policies and creates the preconditions for the formation of a pan-European Digital Health Data Space. Taken together, these standards define not only the technological, but also the ethical and organisational guidelines for the development of virtual medicine. The standards ensure the transparency of medical-data processing, the coherence of actions between public and private structures and the long-term reliability of systems in which AI analytics and blockchain verification function as complementary mechanisms for the protection and management of information.

Ethical and security balance of virtual healthcare systems

Revolutionary computer systems of virtual healthcare are created on the basis of the synergy of the analytical power of artificial intelligence and the reliability of the blockchain infrastructure, which ensures trust, transparency, and control over medical information. In accordance with the Ethics guidelines for trustworthy AI (European Commission, 2019) and the Recommendation of the Council on Artificial Intelligence (Organisation for Economic Co-operation…, 2019), the ethical balance of such systems presupposes a harmonious combination of technical reliability, algorithmic

accountability, data protection and human oversight. In this model, AI functions as the analytical core of personalised diagnostics, analysing large volumes of medical data, forming individual predictions and recommendations, while blockchain acts as the guarantor of digital trust – recording each transaction, verifying authenticity and protecting information from unauthorised interference. Each record in the system is accompanied by a unique cryptographic hash, which ensures the impossibility of changing or falsifying clinical results, and the mechanism of tokenised access enables patients to fully control who has the right to view the data and when. Thanks to model auditing, the explainability of algorithmic decisions, data segmentation and blockchain verification, the system achieves a level of ethical accountability that was previously impossible in traditional medical IT infrastructures. AI algorithms operate according to the responsible-by-design principle, meaning that each decision can be traced, checked and explained. Blockchain, in turn, implements audit-by-design – all transactions are immutable, traceable, and open to control by authorised entities. Table 5 presents a systematised relationship between ethical principles, technological mechanisms for the implementation and the regulatory foundations that ensure the creation of a balanced and secure virtual healthcare system.

Table 5. Ethical accountability and security mechanisms in AI-blockchain virtual-medicine systems

Ethical principle	Technological implementation	Regulatory basis	Expected result
Human oversight and control	AI-model audit, double validation of results by a doctor	Ethics guidelines for trustworthy AI (European Commission, 2019)	Increased clinical responsibility, minimisation of automation errors
Transparency and explainability	Algorithmic logging, visualisation of decision logic	OECD AI Recommendation (Organisation for Economic Co-operation and Development, 2019); ISO/TR 24291:2021 (2021)	Comprehensibility for medical professionals and patients
Confidentiality and data rights	Tokenised access, cryptographic encryption, smart contracts	GDPR (Regulation (EU) No. 2016/679 of the European Parliament and of the Council, 2016); eIDAS 2.0 (Regulation (EU) No. 2024/1183 of the European Parliament and of the Council, 2024)	Full patient control over medical records, protection of personal data
Accountability and traceability	Hash validation of transactions, storage of logs in the blockchain	IEEE EMBS P2418.6 (2018); Edenlab (n.d.)	Impossibility of unauthorised changes, evidential nature of each action
Fairness and non-discrimination	Checking models for algorithmic bias, balancing training datasets	Organisation for Economic Co-operation and Development (2019); ISO/TR 24291:2021 (2021)	Reduced risk of discrimination by age, gender, or origin
Social benefit and trust	Monitoring the system's impact on public health, open effectiveness reports	World Health Organisation (2021); Organisation for Economic Co-operation and Development & European Commission (2024)	Increased trust in digital medicine, ethical legitimacy of technologies

Source: compiled by the authors

As can be seen from Table 5, the coherent application of ethical principles and technological mechanisms forms the foundation of a new paradigm of trusted medicine. A system built on the interaction of artificial intelligence and blockchain combines technical reliability, legal transparency and a human-centred orientation. Human oversight and model audit ensure the responsibility of clinical decisions, while algorithmic explainability increases the clarity and acceptability of AI analysis for medical professionals. Tokenised access and blockchain verification guarantee that patients' data remain under the control, while preserving the traceability of all transactions. Owing to the checking of algorithms for bias and the monitoring of social impact, the system not only increases the accuracy and safety of diagnostics but also affirms the ethical legitimacy of digital healthcare as an innovative model oriented towards trust, transparency, and the protection of patients' rights.

The "trust-by-design" concept in the context of virtual healthcare

The "trust-by-design" concept in digital medicine defines trust as a systemic characteristic embedded in the architecture at the design stage, rather than formed as a result of subsequent operation. In the context of virtual healthcare, it is based on the synergy of AI analytical mechanisms, blockchain verification of transactions and legally enshrined data-security standards. In accordance with the Ethics guidelines for trustworthy AI (European Commission, 2019) and the Recommendation of the Organisation for Economic Co-operation and Development (2019), trust is ensured through technical reliability, explainability, human control and the protection of patients' rights. Within the architecture of an integrated virtual healthcare system, trust-by-design functions as a multilevel model that combines analytical, verification, security, ethical and institutional components.

The analytical layer implements algorithmic transparency and reproducibility of results by using machine-learning methods with quality control in accordance with ISO/TR 24291:2021 (2021). The verification layer ensures the immutability and evidential nature of data through a distributed blockchain infrastructure defined by the IEEE EMBS P2418.6 (2018) standard, in which each transaction is accompanied by a cryptographic hash and a timestamp, guaranteeing full traceability of data exchange. The security layer guarantees user authentication, the legitimacy of actions and compliance with confidentiality requirements in line with Regulation (EU) No. 2016/679 (2016) and Regulation (EU) No. 2024/1183 (2024), using tokenised access, digital-signature mechanisms and multifactor verification. The ethical layer formalises the principles of transparency, accountability, and non-discrimination of algorithms through AI-audit modules, explainability visualisation (XAI) and the checking of sample data in accordance with the Ethics guidelines for trustworthy AI (European Commission, 2019) and the Recommendation of the Council on Artificial Intelligence (Organisation for Economic Co-operation…, 2019). The institutional layer consolidates trust between the patient, medical institutions and state regulators through integrated services built on the Kodjin FHIR Server (Edenlab, n.d.) and supported by the policies of Health at a glance: Europe 2024 (Organisation for Economic Co-operation…, 2024), which ensure harmonisation between the technical, legal and ethical requirements of digital medicine. The architectural layers for implementing the "trust-by-design" concept are presented in Table 6.

Table 6. Architectural layers for implementing the "trust-by-design" concept in virtual healthcare

System level	Main trust mechanism	Technological implementation	Regulatory basis	Expected result
Analytical	Algorithmic transparency, quality control of decisions	ML pipeline, Explainable-AI modules	ISO/TR 24291:2021 (2021)	Reproducibility and reliability of predictions
Verification	Immutability and evidential nature of transactions	Blockchain ledger, smart-contract audit trail	IEEE EMBS P2418.6 (2018)	Traceability and impossibility of falsification
Security	Tokenised access control, crypto-identification	eIDAS 2.0, OAuth 2.0, encryption layer	GDPR (Regulation (EU) No. 2016/679, 2016); eIDAS 2.0 (Regulation (EU) No. 2024/1183, 2024)	Protected access, legal validity of operations
Ethical	AI audit, assessment of social impact	AI Ethical Assessment Tools, Bias Detection	European Commission (2019); Organisation for Economic Co-operation and Development (2019)	Non-discrimination, accountability, social acceptability
Institutional	Verification of actions between eHealth actors	Kodjin FHIR Server, National Ledger	Edenlab (n.d.); Organisation for Economic Co-operation and Development & European Commission (2024)	Institutional trust, integrity of the digital environment

Source: compiled by the authors

As can be seen from Table 6, the trust-by-design concept transforms trust from an abstract ethical principle into a measurable parameter of system architecture. Algorithmic transparency ensures the controllability of analytical processes, blockchain verification – the evidential immutability of data, regulatory standards – legitimacy and legal compliance, and ethical mechanisms – accountability and social trust. This approach forms the structural basis of virtual medicine, in which analytical efficiency, cyber-resilience and a humanistic orientation function as interdependent parameters of a single digital ecosystem. The study carried out showed that the combination of the analytical capabilities of artificial intelligence with the verification potential of blockchain forms a qualitatively new architecture of virtual healthcare, in which trust, security and personalisation act as interrelated components of a single system. Artificial intelligence, operating within standardised protocols for processing clinical data (ISO/TR 24291:2021, 2021), provides personalised diagnostics, the prediction of patient status and the adaptation of therapeutic strategies. Blockchain, in turn, implementing the principles of data immutability and traceability in accordance with IEEE EMBS P2418.6 (2018), creates a reliable mechanism for access control, guarantees the integrity of medical information and protects it from unauthorised interference. The interaction of these technologies within a common infrastructure built on the HL7 FHIR standards (European Commission, n.d.), eIDAS 2.0 (Regulation (EU) No. 2024/1183, 2024) and GDPR (Regulation (EU) No. 2016/679, 2016) makes it possible to create a transparent, interoperable and legally balanced digital ecosystem in which the patient has full control over the own data and each transaction is confirmed cryptographically. This approach forms the foundation of revolutionary computer systems of virtual medicine, where the analytical power of AI is combined with blockchain-based trust auditing, and

diagnostic effectiveness is combined with absolute confidentiality and legal protection. It is recommended to start with pilot implementations in clearly defined scenarios (telemedicine, cancer screening, referral routing), immediately embedding a set of KPIs for each system layer: model accuracy and reproducibility, latency of FHIR-resource exchange, throughput of the transaction ledger, the share of actions with a valid digital signature, and the frequency of access incidents. In parallel, it is advisable to introduce procedures for ethical AI audit, the regular checking of models for bias and mechanisms for real-time revocation of access. For national integration, it is appropriate to create harmonised FHIR profiles, data-management policies and interdepartmental interaction mechanisms that take into account cross-border data exchange. Thus, the integration of AI and blockchain in healthcare not only increases the accuracy of medical predictions and the effectiveness of data management, but also creates a new paradigm of digital medicine – trust-by-design, in which technologies guarantee transparency, accountability and a human-centred orientation for every clinical decision.

DISCUSSION

The results obtained confirmed that the combination of artificial intelligence and blockchain technologies created a new paradigm for the functioning of digital medicine, based on the principles of trust, transparency, and data security. As H. Omidian (2024) noted, the complementarity of algorithmic analytics and decentralised data storage opened up a qualitatively new stage in the development of medical systems. Comparison with the results of the theoretical analysis carried out showed that a similar effect was observed within the multilevel architecture, where artificial intelligence ensured the processing of clinical data and blockchain performed the function of an independent transaction ledger, guaranteeing the immutability of records. The consistency of these conclusions indicated the universality of the patterns identified in the processes of digital transformation of medicine. The assertion of M. Supriya & V.K. Chattu (2021) that the integration of intelligent algorithms and distributed ledger technologies increased the efficiency of processing large volumes of medical data while simultaneously ensuring the security of personal information was also confirmed by the results of the theoretical analysis carried out. The systematic assessment showed that the combined use of analytical models and verification mechanisms made it possible to achieve a high level of performance without compromising between the accuracy of predictions and the confidentiality of data. This revealed a key trend in the development of modern medical platforms – a shift of emphasis from centralised information management to decentralised exchange schemes based on cryptographic trust. The results obtained also confirmed the observations of S. Sai *et al*. (2022), who noted that the main advantages of integrated artificial intelligence and blockchain solutions are the scalability and stability of transactional processes. The patterns identified during theoretical modelling indicated that the efficiency of analytical systems increased when a structured verification layer was implemented, which reduced computational loads and ensured the synchronisation of request processing. Thus, the results obtained were not only consistent with previous research but also refined the mechanisms for achieving an optimal balance between speed, accuracy, and reliability in digital medical systems. The position of D.K. Murala *et al*. (2023), according to which the creation of a multicomponent environment such as MedMetaverse made it possible to combine intelligent analytical modules, blockchain, and sensor devices for continuous monitoring of patients' condition, partially coincided with the results of the theoretical analysis carried out. Within the architectural model considered, it was established that the structural integration of the analytical, verification and communication layers made it possible to ensure the continuous collection, analysis, and confirmation of clinical data, which corresponded to the principles of continuous medical supervision described in the scientific literature.

The study conducted also confirmed the observations of S.K. Rana *et al*. (2022) on the importance of decentralised access control and data verification as a basic ethical principle of digital medicine. Theoretical analysis demonstrated that the use of blockchain mechanisms to confirm the authenticity of the actions of medical actors reduced the risks of duplication, unauthorised editing and destruction of records. The conclusions also coincided with the results described by J. Merhej *et al*. (2024), who showed that blockchain helped to increase the transparency of medical information exchange between clinics. Comparison of these approaches with the results of the study showed that the effective integration of both technologies strengthened the resilience of the information infrastructure and increased users' trust in digital services. Comparison of the theoretical analysis results with the conclusions of B. Ikharo *et al*. (2021) showed that the inclusion of artificial intelligence and blockchain technologies on the Internet of Things ecosystem created preconditions for expanding the functionality of virtual medicine. It was found that, when smart contracts and sensor devices were integrated, constant monitoring of medical parameters was ensured while adhering to the principles of data protection, which coincided with the advantages of decentralised medical networks described by the authors. The results obtained were also consistent with the statements of S. Oberoi *et al*. (2024), who interpreted the decentralisation of medical data management as a necessary condition for the sustainable development of digital medicine. Theoretical analysis confirmed that the introduction of mechanisms of distributed responsibility between users, medical institutions and regulatory bodies

contributed to increased transparency and accountability of systems. The conclusions of H. Fatoum *et al*. (2021), who considered blockchain as the basis of new medical ecosystems, also found confirmation, since the results obtained showed that it was technological decentralisation that became the driver of the ethical modernisation of digital medicine. As D. Bhumichai *et al*. (2024) had anticipated, the effectiveness of the integration of artificial intelligence and blockchain depended not only on technical parameters but also on regulatory and legal transparency. The analysis carried out demonstrated that the harmonisation of technological and legal components ensured compliance with the principles of legality, patient autonomy and the ethical responsibility of developers. Thus, the results confirmed that digital medicine required not merely technical innovation but the systemic integration of technological and regulatory layers. The conclusions of R. Shinde *et al*. (2024) that the combination of blockchain and intelligent algorithms created conditions for the development of new security protocols with enhanced integrity and explainability of machine-learning processes were in tune with the results of the theoretical analysis carried out. Within the model, it was shown that multilevel integration provided not only the technical but also the ethical infrastructure of digital medicine, in which the notion of trust turned into a systemic characteristic of the architecture.

The results presented in the study were consistent with the conclusions of S. Ali *et al*. (2023), who showed that the combination of blockchain, artificial intelligence and a meta-educational environment created preconditions for the formation of interactive and secure medical platforms. Theoretical analysis confirmed this approach, since the concept of a multilevel virtual-medicine architecture reproduced similar principles of immersion, algorithm explainability and ensuring user trust. The results also revealed consistency with the study of N. Seifi *et al*. (2025), who proposed assessing the effectiveness of the integration of artificial intelligence and blockchain through a system of priorities in healthcare service supply chains. In the course of the theoretical analysis it was established that the combination of the analytical and verification layers in the proposed architecture made it possible to ensure the transparency of data flows and to increase the controllability of information processes, which coincided with the principles described by the authors. Comparison with the work of B. Gami *et al*. (2023) showed that the use of artificial intelligence algorithms for implementing blockchain solutions with extended confidentiality-preservation mechanisms was consistent with the results of the study, within which the possibility of simultaneously ensuring analytical accuracy and compliance with confidentiality standards for medical data was confirmed. It was established that an architecture with a clear separation of the functions of analysis, verification, and communication eliminated the conflict between openness and security, as was also noted in the work of B. Gami *et al*.

The results of the theoretical analysis also confirmed the observations of C. Krittanawong *et al*. (2022), who considered the use of blockchain and intelligent algorithms in cardiology as a factor increasing the accuracy of diagnostic predictions. Within the study, it was shown that decentralised verification of clinical records contributed to the stability of analytical models, which was consistent with the conclusions of these authors regarding the increased reliability of artificial neural networks in medical systems. The analysis also revealed consistency of the results with the work of M. Alshehri (2023), which addressed the use of blockchain approaches to strengthen cybersecurity in the ecosystem of medical things. It was established that the concept of a verification layer, which ensured confirmation of the integrity of transactions, played a similar role in creating a secure information infrastructure. Comparison of the results with the conclusions of A. Hasselgren *et al*. (2021) showed that blockchain really increased the level of trust in virtual medical services. Within the analysis carried out, it was found that recording all actions in an immutable ledger made it possible to preserve the complete history of interaction between doctor and patient, which confirmed the practical value of the approach proposed in the cited study. The results also correlated with the conclusions of T. Anwar *et al*. (2024), who showed that the combination of artificial intelligence, blockchain, and the Internet of Things in a fifth-generation environment created new opportunities for secure data exchange. The analysis carried out demonstrated that the integration of the communication layer through standardised protocols contributed to increased speed and reliability of network operations, which was consistent with the advantages of 5G-oriented medicine described by the authors. The provisions of S. Datta *et al*. (2025), who stressed that the combination of artificial intelligence and blockchain is the key to energy efficiency, transparency, and trust in medical systems, were also confirmed. The results of the analysis showed that the introduction of a hierarchical architecture with separated levels of information processing and verification contributed to the optimisation of energy consumption and the reduction of transaction duplication, which corresponded to the researchers' conclusions. The conclusions of Y.Y. Ghadi *et al*. (2025) on the prospects of combining wearable-device technologies and artificial intelligence in remote medical supervision coincided with the results of the analysis, which showed that the inclusion of sensor modules in the communication layer ensured the continuous collection and verification of patients' physiological parameters in real time. This confirmed the possibility of using the theoretical model to build integrated telemedicine systems.

Comparison with the study of A. Pandey & J. Joseph (2025) confirmed that the convergence of blockchain

and artificial intelligence transformed the structure of medical services, making these services more adaptive, transparent and user-oriented. Similar patterns were identified within the theoretical analysis, where the architecture of virtual medicine demonstrated the ability to support patient autonomy and developer accountability. The results were also consistent with the conclusions of S. Podder *et al.* (2025), who viewed the synthesis of blockchain and artificial intelligence of things as the basis for the formation of integrated e-health systems. Within the theoretical analysis it was confirmed that it was precisely the combination of the analytical and verification layers that formed a resilient medical-data management system capable of ensuring a balance between processing efficiency, transparency and the ethical standards of digital medicine.

Thus, in comparison with international and interdisciplinary approaches, the results obtained showed that the development of digital medicine was at the intersection of several key challenges: the need to move from fragmented technological solutions to a holistic integrated model; the combination of the analytical capabilities of artificial intelligence with the ethical mechanisms of blockchain; the provision of algorithmic transparency and legal accountability; and the alignment of technological innovations with regulatory standards of security and data protection. It is the combination of these directions that defined the contours of the future model of virtual healthcare, in which trust, explainability and responsibility became the system-forming principles of next-generation digital medicine.

CONCLUSIONS

The study carried out made it possible to provide a theoretical justification for the integration of artificial intelligence and blockchain technologies into computer systems of virtual healthcare as a new direction in the development of digital medicine. As a result of the analysis of scientific approaches, regulatory acts and international technical standards, it was determined that the synergy of these technologies forms a qualitatively new architecture for medical-data management, based on the principles of decentralisation, security, explainability, and trust. The study considered an integrated five-layer architecture of a virtual healthcare system in which analytical artificial-intelligence modules are combined with blockchain infrastructure and interoperability standards in order to ensure the continuous, accountable and secure circulation of medical data. It was established that the separation of functions between the analytical, verification, communication, security, and user layers makes it possible simultaneously to achieve personalised diagnostics, transaction immutability, semantic data compatibility and transparent access control on the part of the patient. The trust-by-design concept has been translated from a declarative principle into an architectural mechanism: technical control points have been defined (AI decision logs, smart-contract audit trail, tokenised access), quality metrics (model reproducibility, traceability of changes, service response time) and institutional guarantees (centralised management of FHIR resources at the national level with the possibility of decentralised verification).

The results obtained have a dual significance. Theoretically, a coherent model of the interaction of technological and legal components is proposed, which eliminates the gap between analytics, security and ethics, sets measurable trust criteria and ensures the portability of solutions between organisations. Practically, an implementation pathway is outlined: microservice integration with FHIR interfaces, maintenance of an immutable event ledger, centralised orchestration of access rights via smart contracts and the introduction of explainability dashboards for clinicians. The expected qualitative indicators include increased accuracy of clinical decisions owing to the use of verified data, a reduction in administrative delays through the automation of protocols, a decrease in the risk of falsification thanks to full traceability and the strengthening of patient autonomy through controllable access.

The limitations of the study lie in the conceptual and model-based nature of the results without full-scale clinical testing, the dependence on the maturity of national digital infrastructure and the variability of law enforcement in different jurisdictions. The full life-cycle cost of the solutions and the impact of extreme loads on registry performance have also not been assessed. Further research should be directed towards experimental validation in real clinical workflows, the optimisation of on-chain/off-chain storage, scaling under conditions of interregional data exchange, the standardisation of XAI panels for clinicians and the development of economic models of cost of ownership. Taken together, this will make it possible to transfer the proposed architecture from the level of design solutions to a stable, measurably effective practice of digital medicine.

ACKNOWLEDGEMENTS

None.

FUNDING

None.

CONFLICT OF INTEREST

None.

REFERENCES

[1] Ali, S., Abdullah, Armand, T.P.T., Athar, A., Hussain, A., Ali, M., Yaseen, M., Joo, M.-I., & Kim, H.-C. (2023). Metaverse in healthcare integrated with explainable AI and blockchain: Enabling immersiveness, ensuring trust, and providing patient data security. *Sensors*, 23(2), article number 565. doi: 10.3390/s23020565.

[2] Alshehri, M. (2023). Blockchain-assisted cyber security in medical things using artificial intelligence. *Electronic Research Archive*, 31(2), 708-728. doi: 10.3934/era.2023035.
[3] Anwar, T., Khan, G.A., Ashraf, Z., Ansari, Z.A., Ahmed, R., & Azrour, M. (2024). The combination of blockchain and the Internet of Things (IoT): Applications, opportunities, and challenges for industry. In M. Azrour, J. Mabrouki, A. Guezzaz & S. Benkirane (Eds.), *Blockchain and machine learning for IoT security* (pp. 56-76). New York: Chapman and Hall/CRC. doi: 10.1201/9781003438779.
[4] Augustyn, R.R., & Stakhiv, O.V. (2023). Management of intellectualization processes – a systemic condition for the development of health care institutions. *Scientific Notes of Lviv University of Business and Law*, 37, 13-21. Retrieved from https://nzlubp.org.ua/index.php/journal/article/view/735.
[5] Bhumichai, D., Smiliotopoulos, C., Benton, R., Kambourakis, G., & Damopoulos, D. (2024). The convergence of artificial intelligence and blockchain: The state of play and the road ahead. *Information*, 15(5), article number 268. doi: 10.3390/info15050268.
[6] Brodkevych, V., & Liudvichenko, V. (2022). Artificial intelligence and machine learning in healthcare: Challenges and trends. *Information Technology and Society*, 2(4), 20-28. doi: 10.32689/maup.it.2022.2.3.
[7] Datta, S., Namasudra, S., Moparthi, N.R., Kumari, S., & Crespo, R.G. (2025). Transforming healthcare with artificial intelligence and blockchain: A secure, transparent and energy-efficient approach. *Expert Systems*, 42(8), article number e70101. doi: 10.1111/exsy.70101.
[8] Davydova, I. (2025). On the question of the use of blockchain technology in the medical sphere. *Legal Bulletin*, 2025(1), 38-43. doi: 10.32782/yuv.v1.2025.5.
[9] Driving digital health forward: HL7 Europe and IHE-Europe strengthen their collaboration to enhance interoperability. (2024). Retrieved from https://hl7europe.org/driving-digital-health-forward-hl7-europe-and-ihe-europe-strengthen-their-collaboration-to-enhance-interoperability/.
[10] Edenlab. (n.d.). *National e-health system*. Retrieved from https://edenlab.io/case/national-e-health-system.
[11] eHealth Network – 23rd eHealth network meeting. (2023). Retrieved from https://health.ec.europa.eu/system/files/2023-05/ehealth_20230330_sr_en.pdf.
[12] European Commission. (2019). *Ethics guidelines for trustworthy AI*. Retrieved from https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.
[13] European Commission. (n.d.). *eHealth network*. Retrieved from https://health.ec.europa.eu/ehealth-digital-health-and-care/digital-health-and-care/eu-cooperation/ehealth-network_en.
[14] Fatoum, H., Hanna, S., Halamka, J.D., Sicker, D.C., Spangenberg, P., & Hashmi, S.K. (2021). Blockchain integration with digital technology and the future of health care ecosystems: Systematic review. *Journal of Medical Internet Research*, 23(11), article number e19846. doi: 10.2196/19846.
[15] Gami, B., Agrawal, M., Mishra, D.K., Quasim, D., & Mehra, P.S. (2023). Artificial intelligence-based blockchain solutions for intelligent healthcare: A comprehensive review on privacy preserving techniques. *Transactions on Emerging Telecommunications Technologies*, 34(9), article number e4824. doi: 10.1002/ett.4824.
[16] Ghadi, Y.Y., Shah, S.F.A., Waheed, W., Mazhar, T., Ahmad, W., Saeed, M.M., & Hamam, H. (2025). Integration of wearable technology and artificial intelligence in digital health for remote patient care. *Journal of Cloud Computing*, 14(1), article number 39. doi: 10.1186/s13677-025-00759-4.
[17] Gregor, S., & Hevner, A.R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, 37(2), 337-355. doi: 10.25300/MISQ/2013/37.2.01.
[18] Hasselgren, A., Rensaa, J.A.H., Kralevska, K., Gligoroski, D., & Faxvaag, A. (2021). Blockchain for increased trust in virtual health care: Proof-of-concept study. *Journal of Medical Internet Research*, 23(7), article number e28496. doi: 10.2196/28496.
[19] Hevner, A.R., March, S.T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75-105. doi: 10.2307/25148625.
[20] IEEE EMBS P2418.6. (2018). *Standard for the framework of distributed ledger technology (DLT) use in healthcare and the life and social sciences*. Retrieved from https://sagroups.ieee.org/2418-6/wp-content/uploads/sites/15/2018/12/P2418-6_PPs_Dec2018.pdf.
[21] Ikharo, B., Obiagwu, A., Obasi, C., Hussein, S.U., & Akah, P. (2021). Security for internet-of-things enabled e-health using blockchain and artificial intelligence: A novel integration framework. In *2021 1st international conference on multidisciplinary engineering and applied science (ICMEAS)* (pp. 1-4). Abuja: IEEE. doi: 10.1109/ICMEAS52683.2021.9692368.
[22] ISO/TR 24291:2021. (2021). *Health informatics – applications of machine learning technologies in imaging and other medical applications*. Retrieved from https://www.iso.org/standard/78345.html.
[23] Krittanawong, C., Aydar, M., Virk, H.U.H., Kumar, A., Kaplin, S., Guimaraes, L., Wang, Z., & Halperin, J.L. (2022). Artificial intelligence-powered blockchains for cardiovascular medicine. *Canadian Journal of Cardiology*, 38(2), 185-195. doi: 10.1016/j.cjca.2021.11.011.

[24] Law of Ukraine No. 2297-VI "On Personal Data Protection". (2010, June). Retrieved from https://zakon.rada.gov.ua/laws/show/en/2297-17#Text.
[25] Merhej, J., Harb, H., Abouaissa, A., & Idoumghar, L. (2024). Toward a new era of smart and secure healthcare information exchange systems: Combining blockchain and artificial intelligence. *Applied Sciences*, 14(19), article number 8808. doi: 10.3390/app14198808.
[26] Murala, D.K., Panda, S.K., & Dash, S.P. (2023). MedMetaverse: Medical care of chronic disease patients and managing data using artificial intelligence, blockchain, and wearable devices state-of-the-art methodology. *IEEE Access*, 11, 138954-138985. doi: 10.1109/ACCESS.2023.3340791.
[27] Oberoi, S., Arora, S., Verma, B., & Roy, K.K. (2024). What do we know about artificial intelligence and blockchain technology integration in the healthcare industry? In B. Verma, B. Singla & A. Mittal (Eds.), *Driving decentralization and disruption with digital technologies* (pp. 124-138). Hershey: IGI Global Scientific Publishing. doi: 10.4018/979-8-3693-3253-5.ch009.
[28] Omidian, H. (2024). Synergizing blockchain and artificial intelligence to enhance healthcare. *Drug Discovery Today*, 29(9), article number 104111. doi: 10.1016/j.drudis.2024.104111.
[29] Order of the Cabinet of Ministers of Ukraine No. 1175-2021-p "On Approval of the Action Plan for the Implementation of the Concept for the Development of Electronic Healthcare". (2021, September). Retrieved from https://zakon.rada.gov.ua/laws/show/1175-2021-%D1%80.
[30] Order of the Cabinet of Ministers of Ukraine No. 1671-2020-r "On Approval of the Concept for the Development of Electronic Healthcare". (2020, December). Retrieved from https://zakon.rada.gov.ua/laws/show/1671-2020-%D1%80.
[31] Organisation for Economic Co-operation and Development, & European Commission. (2024). *Health at a glance: Europe 2024: State of health in the EU cycle*. Retrieved from https://www.oecd.org/en/publications/health-at-a-glance-europe-2024_b3704e14-en.html.
[32] Organisation for Economic Co-operation and Development. (2019). *Recommendation of the Council on artificial intelligence*. Retrieved from https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.
[33] Pandey, A., & Joseph, J. (2025). The convergence of blockchain and artificial intelligence: Transforming healthcare. In A. Ahad & F. Ullah (Eds.), *AI and blockchain applications for privacy and security in smart medical systems* (pp. 399-416). Hershey: IGI Global Scientific Publishing. doi: 10.4018/979-8-3373-0593-6.ch015.
[34] Podder, S., Gupta, V.R., Khator, S., Koley, R., & Goswami, S.R. (2025). Fusion of blockchain and artificial intelligence of things in e-healthcare. In A.S. Rajasekaran, F. Al-Turjman & S. Suganyadevi (Eds.), *AIoT: Artificial intelligence of things* (pp. 57-98). New York: Auerbach Publications. doi: 10.1201/9781003482338.
[35] Rana, S.K., Rana, S.K., Nisar, K., Ag Ibrahim, A.A., Rana, A.K., Goyal, N., & Chawla, P. (2022). Blockchain technology and artificial intelligence based decentralized access control model to enable secure interoperability for healthcare. *Sustainability*, 14(15), article number 9471. doi: 10.3390/su14159471.
[36] Regulation (EU) No. 2016/679 of the European Parliament and of the Council "On the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance)". (2016, April). Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679.
[37] Regulation (EU) No. 2024/1183 of the European Parliament and of the Council "Amending Regulation (EU) No 910/2014 as Regards Establishing the European Digital Identity Framework". (2024, April). Retrieved form https://eur-lex.europa.eu/eli/reg/2024/1183/oj/eng.
[38] Resolution of the Cabinet of Ministers of Ukraine No. 411-2018-p "Some Issues of the Electronic Health Care System". (2018, April). Retrieved from https://zakon.rada.gov.ua/laws/show/411-2018-%D0%BF.
[39] Reyes-González, J.P., Díaz-Peregrino, R., Soto-Ulloa, V., Galvan-Remigio, I., Castillo, P., & Ogando-Rivas, E. (2022). Big data in the healthcare system: a synergy with artificial intelligence and blockchain technology. *Journal of Integrative Bioinformatics*, 19(1), article number 20200035. doi: 10.1515/jib-2020-0035.
[40] Sai, S., Chamola, V., Choo, K.K.R., Sikdar, B., & Rodrigues, J.J. (2022). Confluence of blockchain and artificial intelligence technologies for secure and scalable healthcare solutions: A review. *IEEE Internet of Things Journal*, 10(7), 5873-5897. doi: 10.1109/JIOT.2022.3232793.
[41] Samad, A. (2022). Internet of things integrated with blockchain and artificial intelligence in healthcare system. *Research Journal of Computer Systems and Engineering*, 3(1), 1-6. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/34.
[42] Seifi, N., Ghoodjani, E., Majd, S.S., Maleki, A., & Khamoushi, S. (2025). Evaluation and prioritization of artificial intelligence integrated block chain factors in healthcare supply chain: A hybrid decision making approach. *Computer and Decision Making: An International Journal*, 2, 374-405. doi: 10.59543/comdem.v2i.11029.
[43] Shinde, R., Patil, S., Kotecha, K., Potdar, V., Selvachandran, G., & Abraham, A. (2024). Securing AI-based healthcare systems using blockchain technology: A state-of-the-art systematic literature review and future research directions. *Transactions on Emerging Telecommunications Technologies*, 35(1), article number e4884. doi: 10.1002/ett.4884.

[44] Shmatko, O., & Salnikov, S. (2024). Model of decentralized electric medical card exchange system based on blockchain technology. *Control, Navigation and Communication Systems. Academic Journal,* 2(76), 155-162. doi: 10.26906/SUNZ.2024.2.155.
[45] Sumets, O., & Alkema, V. (2023). Socio-economic aspects of using blockchain technology in clinical trial data management. *Science Notes of KROK University*, 2(70), 115-122. doi: 10.31732/2663-2209-2022-70-115-122.
[46] Supriya, M., & Chattu, V.K. (2021). A review of artificial intelligence, big data, and blockchain technology applications in medicine and global health. *Big Data and Cognitive Computing*, 5(3), article number 41. doi: 10.3390/bdcc5030041.
[47] Tagde, P., Tagde, S., Bhattacharya, T., Tagde, P., Chopra, H., Akter, R., Kaushik, D., & Rahman, H. (2021). Blockchain and artificial intelligence technology in e-Health. *Environmental Science and Pollution Research*, 28(38), 52810-52831. doi: 10.1007/s11356-021-16223-0.
[48] Verma, P., Rao, C.M., Chapalamadugu, P.K., Tiwari, R., & Upadhyay, S. (2024). Future of electronic healthcare management: Blockchain and artificial intelligence integration. In K. Kaushik & I. Sharma (Eds.), *Next-generation cybersecurity: AI, ML, and blockchain* (pp. 179-218). Singapore: Springer. doi: 10.1007/978-981-97-1249-6_9.
[49] Xie, Y., Lu, L., Gao, F., He, S.-J., Zhao, H.-J., Fang, Y., Yang, J.-M., An, Y., Ye, Z.-M., & Dong, Z. (2021). Integration of artificial intelligence, blockchain, and wearable technology for chronic disease management: A new paradigm in smart healthcare. *Current Medical Science*, 41(6), 1123-1133. doi: 10.1007/s11596-021-2485-0.

Інтеграція технологій штучного інтелекту та блокчейн у комп'ютерні системи віртуальної охорони здоров'я

Микола Хрульов

Аспірант
Черкаський державний технологічний університет
18006, бульв. Шевченка, 460, м. Черкаси, Україна
https://orcid.org/0009-0004-9201-5118

Тетяна Миронюк

Кандидат технічних наук, доцент
Черкаський державний технологічний університет
18006, бульв. Шевченка, 460, м. Черкаси, Україна
https://orcid.org/0000-0002-7588-1055

Анотація. Мета дослідження полягала у теоретичному обґрунтуванні інтеграції технологій штучного інтелекту та блокчейну в комп'ютерні системи віртуальної охорони здоров'я, зокрема у з'ясуванні, як поєднання штучного інтелекту, орієнтованого на персоналізовану діагностику, із блокчейн-технологіями для захисту медичних даних може сформувати нову модель цифрової медицини. У процесі дослідження використано аналітичний, синтетичний, системний і порівняльно-правовий методи, що дозволило комплексно проаналізувати наукові підходи, міжнародні стандарти та технічні фреймворки, пов'язані з організацією цифрової охорони здоров'я. Розглянуто п'ятирівневу архітектуру інтегрованої системи віртуальної медицини, що включає аналітичний, верифікаційний, комунікаційний, безпековий та користувацький рівні. На аналітичному рівні здійснюється обробка клінічних даних і формування персоналізованих прогнозів лікування; на верифікаційному рівні забезпечується незмінність та простежуваність записів завдяки використанню блокчейн-механізмів; на комунікаційному рівні реалізується обмін інформацією через стандартизовані інтерфейси; безпековий рівень гарантує захист персональних даних і правову автентичність усіх дій; користувацький рівень формує прозору взаємодію між пацієнтами, лікарями та медичними установами. У межах теоретичної моделі обґрунтовано концепцію «довіра за проєктом», що визначає довіру як системну характеристику цифрової медицини, закладену в архітектуру ще на етапі розроблення. Результатом дослідження стало визначення структурних і функціональних засад побудови безпечної та персоналізованої системи віртуальної охорони здоров'я, у якій аналітична потужність штучного інтелекту поєднується з перевіреною надійністю блокчейн-технологій. Основні засади побудови системи охоплюють багаторівневу архітектуру з аналітичним, верифікаційним, комунікаційним, безпековим і користувацьким рівнями, які забезпечують точність аналітики, незмінність даних, стандартизований обмін, захист конфіденційності та прозору взаємодію між учасниками цифрової медицини. Практична значущість отриманих результатів полягає у можливості використання розробленої моделі як концептуальної основи для створення національних і міжнародних платформ цифрової медицини, що забезпечують законність, конфіденційність і довіру користувачів до сучасних інформаційних рішень у сфері охорони здоров'я

Ключові слова: концептуальна архітектура; верифікаційний рівень; інтероперабельність; конфіденційність медичних даних; етична підзвітність; концепція trust-by-design; децентралізовані системи

BULLETIN of Cherkasy State Technological University

Journal homepage: https://bulletin-chstu.com.ua/en

Vol. 31 No. 1. 2026

UDC 004.056.5
DOI: 10.62660/bcstu/1.2026.118

ISSN: 2306-4412
E-ISSN: 2708-6070

Mesh architectures and immune-inspired mechanisms in defending critical infrastructure

Viktoriia Semerenska*
Postgraduate Student
Kharkiv National University of Radioelectronics
61166, 14 Nauky Ave, Kharkiv, Ukraine
https://orcid.org/0009-0008-2955-3676

Abstract. Critical infrastructure increasingly operates in interconnected, software-defined environments and is simultaneously subject to targeted cyberattacks that disrupt key services. Traditional perimeter approaches are proving insufficient as they allow attackers to remain in systems for long periods of time and leave recovery mechanisms vulnerable. The article explored the applicability of the principles of mesh cybersecurity architecture and digital immune systems, which have been developed in cloud environments, to the public and industrial sectors. The aim of the study was to evaluate the transfer of mesh and immune approaches to the water supply, energy, and municipal services sectors and to demonstrate their impact on resilience in real incidents. The methodology combined a review of cloud security patterns, comparative case studies, and threat scenario modeling. Two events were considered: a cyberattack on the municipal infrastructure of a large US city and a transnational campaign against operational technologies in the water and gas sector. In the first case, the lack of segmentation allowed the virus to spread unhindered between network segments, while in the second, the lack of automated monitoring led to a delay in detecting the intrusion. For each event, the progression of the attack was compared with control points where distributed control nodes, identity-based segmentation, and feedback loops could limit the impact and initiate automated recovery. The results confirmed that mesh combined with immune-like responses provided faster isolation, controlled degradation, and recovery based on behavioral signals such as abnormal commands or configuration changes. Simulation modeling showed that the average system recovery time was reduced by 35-40% in scenarios with a mesh architecture, and the spread of the attack was limited to one segment instead of four. The practical value of this work lies in providing a roadmap for the gradual improvement of monitoring systems without a complete redesign, which is useful for operators of critical infrastructure and industrial enterprises

Keywords: resilience; operational technology; distributed enforcement; segmentation; cyber resilience; digital twins

INTRODUCTION

The protection of critical infrastructure has become one of the central challenges of cybersecurity. Energy generation, water supply, transportation, municipal administration, and healthcare are increasingly dependent on digital technologies. These functions were once operated on isolated industrial control systems with limited connectivity, but over time the drive for efficiency, remote management, predictive maintenance, and integration with enterprise services has led to broader networking and, in many cases, connections to cloud platforms. This evolution improved operational capabilities but also expanded the attack surface, turning critical infrastructure into a preferred target for both financially motivated groups and state-sponsored actors.

Scholars have therefore investigated new models of resilience that go beyond traditional perimeter protection. O. Johnphill *et al.* (2023) studied how machine learning techniques could be embedded into

Article's History: Received: 12.09.2025; Revised: 05.02.2026; Accepted: 16.03.2026; Published: 08.04.2026

Suggested Citation:
Semerenska, V. (2026). Mesh architectures and immune-inspired mechanisms in defending critical infrastructure. *Bulletin of Cherkasy State Technological University*, 31(1), 118-127. doi: 10.62660/bcstu/1.2026.118.

*Corresponding author

cyber-physical systems to allow nodes to autonomously detect anomalies and reconfigure their behaviour. Their experiments demonstrated that distributed anomaly detection not only improved detection accuracy but also enabled systems to restore functionality without relying on centralised control, thereby increasing overall system resilience. B. Bejoy *et al.* (2022) proposed a cyber-immune framework for anomaly detection that drew on mechanisms inspired by immunology. They showed that combining immune-like memory and adaptive responses improved the detection of anomalous traffic compared with static signature-based methods. Their conclusion was that immune-inspired approaches could support the continuous adaptation of security systems in dynamic environments.

M.A. Naqvi *et al.* (2022) evaluated the use of chaos engineering in assessing self-adaptive and self-healing systems. Rather than focusing on static security properties, they introduced controlled disturbances to test whether recovery mechanisms activated in time. Their findings confirmed that chaos experiments revealed hidden weaknesses and helped validate whether self-healing responses could prevent cascading failures across interconnected services. A. Alauthman & A. Al-Hyari (2025) provided a systematic review of zero trust in the Internet of Things (IoT). They surveyed multiple implementations and concluded that identity-based segmentation and continuous verification of devices formed a foundation for building a cybersecurity mesh architecture. C. Zanasi *et al.* (2024) developed a flexible zero trust architecture for industrial IoT and showed that identity-centric controls could be adapted to heterogeneous, resource-constrained control networks to limit lateral movement.

M. Homaei *et al.* (2025) examined the role of digital twins in cybersecurity. Their review highlighted how digital-twin environments served as testing grounds where defence strategies could be validated before deployment in critical infrastructures. They concluded that this approach reduced operational risks by allowing safe experimentation and proactive identification of system vulnerabilities. H. Riggs *et al.* (2023) analysed cyber-attacks against critical infrastructure over the last two decades. By comparing historical incidents and their impacts, the study demonstrated how attacks became more sophisticated and frequent, indicating that reliance on traditional security models left essential services vulnerable and motivating architectures that integrate resilience by design. B. Rathnayaka *et al.* (2024) further proposed a unified framework for evaluating resilience across interdependent systems, emphasising the need for structured detection, containment, and recovery metrics. While these research directions presented valuable insights, industrial control systems and municipal networks operated under constraints that differed from cloud environments. Many controllers were legacy devices with limited computational resources and proprietary protocols, and downtime in these contexts translated directly into risks for public safety and political stability. This article addressed the gap between cloud-borne resilience concepts and the realities of operational technology.

The aim was to evaluate the feasibility of applying cybersecurity mesh and digital immune system principles to constrained critical infrastructure environments. To achieve this aim, the study formulated three specific objectives. The first was to establish a conceptual bridge between cloud-native resilience approaches and cyber-physical infrastructures, taking into account legacy devices, real-time requirements, and safety-critical operations. The second was to analyse recent cyber incidents, such as the St. Paul municipal ransomware attack and the CyberAv3ngers campaign against water and gas facilities, in order to identify intervention points where mesh enforcement and immune responses could have prevented escalation. The third was to perform scenario-based modelling that applied immune algorithms and chaos engineering techniques to demonstrate how automated self-healing might reduce cascade effects and service disruption.

LITERATURE REWIEW

Research on wireless sensor networks (WSNs) extended self-healing concepts to hardware-constrained environments. N. López-Vilos *et al.* (2023) proposed a clustering-based, energy-efficient self-healing strategy for WSNs under jamming attacks. Their results demonstrated that redundancy-based reconfiguration at the network level extended the operational lifetime of WSNs under adversarial conditions. This perspective extended the notion of self-healing beyond software, suggesting that resilience mechanisms should span multiple layers of cyber-physical system design. M. Ganesh Raja & S. Jeyalaksshmi (2024) developed a self-configuration and self-healing framework using clustering for IoT-based WSNs. The study addressed node failures and network partitioning through dynamic cluster reformation. Experimental results showed that the proposed framework reduced recovery time and maintained network connectivity under failure scenarios. The authors concluded that clustering approaches provided a foundation for autonomous network management in resource-constrained environments.

Testing and validation methodologies for self-healing systems received attention from multiple researchers. T. Ma *et al.* (2021) tested self-healing cyber-physical systems under uncertainty using reinforcement learning. Their empirical study demonstrated that reinforcement learning agents could learn optimal recovery policies when trained under simulated attack scenarios. The authors concluded that learning-based approaches offered adaptability in environments where attack patterns evolved over time. Critical infrastructure vulnerabilities formed a persistent theme in the literature. C. Alcaraz & J. Lopez (2012) analysed requirements for critical control systems and demonstrated that

supervisory control and data acquisition (SCADA) environments were dominated by legacy technologies and perimeter-based security. The authors emphasised the scarcity of automated containment and recovery solutions in real-world deployments, a gap that continued to make national infrastructure vulnerable. B. Paul *et al*. (2024) demonstrated through scenario-based modelling that smart grid deployments remained susceptible to denial-of-service, command injection, and cascading fault exploitation. Their study showed that even with conventional firewalls and intrusion detection systems, attacks propagated through interconnected substations. They concluded that mesh-based segmentation with automated isolation would be more effective in preventing systemic failures. Zero trust architectures emerged as a practical foundation for implementing mesh principles in critical infrastructure. C. Liu *et al*. (2024) dissected zero trust research landscape and its implementation in IoT. Their systematic review examined identity-based segmentation, continuous verification, and policy enforcement mechanisms. The authors concluded that zero trust paradigms improved resilience against lateral movement and insider threats, positioning them as a foundation for cybersecurity mesh deployments in resource-constrained environments.

Digital twin technology provided testing environments for resilience mechanisms before deployment. N. Mchirgui *et al*. (2024) reviewed the applications and challenges of digital twin technology in smart grids. Their comprehensive review emphasised that although twins could simulate adversarial conditions and test defence strategies, integration costs and organisational inertia slowed practical deployment in operational environments. O. Sen *et al*. (2024) proposed a digital twin for evaluating detective countermeasures in smart grid cybersecurity. Their framework enabled testing of intrusion detection algorithms and response procedures in a simulated environment that replicated grid topology and control logic. The authors concluded that digital twins reduced operational risks by enabling validation of security mechanisms before implementation in production systems. Analysis of real-world incidents provided evidence of the limitations of traditional defences. The Cybersecurity and Infrastructure Security Agency (CISA, 2023) documented ransomware threats to critical infrastructure, showing that attacks on municipal services, healthcare facilities, and utilities increased in frequency and sophistication. CISA (2024) issued alert on CyberAv3ngers threat activity targeting water and wastewater systems, confirming that adversaries exploited weak authentication and lack of network segmentation to compromise industrial control interfaces. The European Union Agency for Cybersecurity (ENISA, 2023) reported in its threat landscape that critical infrastructure sectors experienced persistent attacks that leveraged supply chain vulnerabilities and insufficient resilience mechanisms.

T. August *et al*. (2024) examined cyberattacks, operational disruption, and investment in cyber resilience from an economic perspective. Their analysis showed that organisations that invested in resilience mechanisms experienced shorter recovery times and reduced financial losses following incidents. The authors concluded that resilience investments provided measurable returns through improved continuity and reduced incident costs. A. De Marco *et al*. (2025) developed a quantitative resilience assessment methodology for critical infrastructures. Their approach combined probabilistic risk assessment with resilience metrics to evaluate system performance under attack scenarios. The authors concluded that quantitative methods provided evidence-based support for resilience planning and resource allocation. J. Järveläinen *et al*. (2025) proposed a framework for improving cyber resilience of critical infrastructure that integrated organisational, technical, and governance dimensions. The framework addressed barriers to resilience adoption, including regulatory gaps, organisational culture, and resource constraints. The author concluded that resilience improvement required coordinated efforts across technical implementation, policy alignment, and stakeholder engagement.

Collectively, these works demonstrated that while self-healing, immune-inspired, and mesh-enabled frameworks showed advantages in controlled environments, their adoption in critical infrastructure remained constrained by technical limitations, organisational barriers, and economic considerations. The literature confirmed that distributed, adaptive defence mechanisms outperformed static perimeter approaches in detection accuracy, recovery time, and containment efficiency. However, the translation of these concepts into legacy industrial control systems and resource-constrained operational technology environments required addressing integration challenges, validating safety-critical performance, and overcoming institutional resistance to automated response mechanisms.

MATERIALS AND METHODS

The methodological design integrated comparative analysis, case study research, and scenario-based modelling. A structured literature review was conducted using indexed databases such as IEEE Xplore, Scopus, and SpringerLink, with a focus on studies that provided experimental validation of resilience mechanisms. Core references included research on machine learning-driven recovery in cyber-physical systems (Johnphill *et al.*, 2023), immune-based detection frameworks for the IoT (Aldhaheri *et al.*, 2020; Bejoy *et al.*, 2022), chaos engineering validation, and digital twin approaches (Homaei *et al.*, 2025). The empirical component relied on two case studies. The first examined the ransomware incident that disrupted municipal systems in St. Paul in July 2025, forcing suspension of city services and activation of emergency forces. The second analysed the CyberAv3ngers campaign against operational technology

in water and gas facilities, attributed in open sources to a state-linked actor. These cases were chosen to capture different domains: local government services and industrial control systems.

For each case, the attack lifecycle was reconstructed from open-source reports and technical analyses, including advisories from the Cybersecurity and Infrastructure Security Agency (CISA, 2023; 2024) and assessments from the European Union Agency for Cybersecurity (ENISA, 2023). The events were mapped to the stages of the cyber kill chain and cross-referenced with potential intervention points. Counterfactual analysis then applied findings from cloud resilience research. Chaos engineering studies informed the identification of latent weaknesses, while digital twin research (Homaei *et al.*, 2025) provided the basis for modelling recovery scenarios. Immune-based anomaly detection frameworks (Aickelin *et al.*, 2004; Bejoy *et al.*, 2022) were used to simulate adaptive detection within distributed nodes. A comparative method contrasted the resilience capabilities of cloud ecosystems with those of critical infrastructure. Research documenting the maturity of mesh enforcement in cloud platforms (Yigit *et al.*, 2024; Alauthman & Al-Hyari, 2025) was evaluated alongside studies highlighting persistent vulnerabilities in SCADA and smart grids (Alcaraz & Zeadally, 2015; Paul *et al.*, 2024). The analysis applied three criteria: time to detect anomalies, time to contain breaches, and ability to restore essential functions without cascading failure.

Finally, the methodology incorporated a normative component. Economic perspectives were derived from critical analyses of IoT anomaly management and systemic assessments of long-term infrastructure risks (Riggs *et al.*, 2023). Policy and governance constraints were examined based on requirements for critical control systems (Alcaraz & Lopez, 2012) and reviews of digital twin adoption in smart grids (Mchirgui *et al.*, 2024). This ensured that the methodology accounted not only for technical feasibility but also for governance and economic barriers influencing real-world deployment. By integrating literature synthesis, empirical case studies, scenario-based modelling, and comparative evaluation, the methodological framework provided a replicable basis for assessing how cloud-derived cybersecurity mesh and digital immune system concepts could be realistically applied to critical infrastructure.

RESULTS AND DISCUSSION

In July 2025 the municipal systems of St. Paul, Minnesota, experienced a ransomware attack that disrupted online payments, public Wi-Fi, and multiple internal services. According to incident reporting, approximately 43 GB of city data was exfiltrated before encryption began, forcing the municipality to shut down significant parts of its IT environment and to declare a local state of emergency. The governor mobilised the National Guard to assist in recovery, which highlighted the severity of the disruption. The city's infrastructure followed a conventional centralised model. Critical services such as payment processing, permit management, and public-facing portals were hosted on a combination of on-premises servers and third-party Software as a Service (SaaS) platforms, with limited segmentation between administrative and citizen-facing domains. Remote access for employees relied on Virtual Private Network (VPN) accounts, and audit reports prior to the incident indicated that some accounts lacked multifactor authentication. File servers and document management systems were interconnected, which allowed the ransomware to propagate laterally once the initial compromise had been achieved.

Technical analyses of municipal ransomware cases of similar scope indicated that attackers frequently exploited outdated VPN endpoints or phishing-induced credential theft to establish initial footholds. Once inside, the absence of identity-centric segmentation allowed broad lateral movement. In St. Paul, reports confirmed that multiple departments lost access simultaneously, suggesting that containment boundaries were either weak or absent. The encryption phase of the attack targeted shared storage and virtualised servers, which disabled services beyond the initial entry point. Counterfactual modelling with cybersecurity mesh and immune principles identified several points of intervention. A mesh-based architecture, in which enforcement nodes validated every east-west connection, would have limited the ransomware's spread across departmental boundaries. Microsegmentation enforced by identity policies could have prevented compromised VPN credentials from granting access to file servers unrelated to the user's role (Alcaraz & Zeadally, 2015; Alauthman & Al-Hyari, 2025). Immune-style anomaly detection, trained to identify compressed bursts of file access and encryption routines, could have automatically quarantined the affected node (Aickelin *et al.*, 2004; Aldhaheri *et al.*, 2020). Self-healing routines derived from cloud practices (Homaei *et al.*, 2025) would have enabled the restoration of corrupted services from clean digital twin images or redundant containers, which reduced downtime. Building on these findings, the study derived a roadmap specifying control-node placement, telemetry priorities, and phased adoption for critical infrastructure environments. Based on the comparative analysis and scenario modelling, the roadmap was structured around three elements.

The first element addressed control-node placement. Policy enforcement points at Application Programming Interface gateways and reverse proxies were positioned at perimeter and DeMilitarised Zone (DMZ) ingress locations, where they applied identity-centric segmentation and rate limiting suitable for municipal web portals and SaaS entry points (Yigit *et al.*, 2024; Alauthman & Al-Hyari, 2025). At the identity plane, a central policy engine issued short-lived tokens and enabled continuous verification for devices, users, and services across information technology (IT) and operational

technology (OT) boundaries. Enforcement positioned close to Programmable Logic Controller (PLC) networks and Human-Machine Interface (HMI) jump-servers at the OT aggregation layer localised faults and blocked lateral movement between control cells (Alcaraz & Lopez, 2012; Paul *et al.*, 2024). Within the monitoring zone and Security Operations Centre, anomaly-score brokers and playbook runners triggered automated isolation and self-healing based on immune-style detectors (Bejoy *et al.*, 2022; Johnphill *et al.*, 2023). Pre-deployment validation of recovery routines and configuration rollbacks was carried out in digital twin sandboxes that constituted the twin and simulation environment (Homaei *et al.*, 2025).

The second element established telemetry priorities organised into three tiers according to operational criticality. Critical telemetry in Tier 1 captured authentication and authorisation decisions, controller state transitions, set-point changes, firmware integrity hashes, and routing or topology changes (Bejoy *et al.*, 2022; Paul *et al.*, 2024). Supporting telemetry in Tier 2 included flow records, time-series anomaly scores, time synchronisation integrity checks, and asset or identity posture updates (Johnphill *et al.*, 2023; Mchirgui *et al.*, 2024). Contextual telemetry in Tier 3 comprised performance counters and backup events used for post-incident reconstruction (Homaei *et al.*, 2025). The third element defined staged adoption organised into five sequential phases. Phase 0 established prerequisites by completing asset and identity inventory, developing a segmentation plan, and deploying a twin sandbox environment. Phase 1 focused on observation by deploying enforcement points in monitor-only mode, establishing Tier-1 telemetry collection, and conducting chaos experiments to identify latent weaknesses. Phase 2 enabled containment through automated isolation at ingress and OT aggregation points combined with least-privilege enforcement. Phase 3 implemented healing mechanisms by enabling automated rollback and service restoration using twin-validated runbooks, allowing local cell recovery without central coordination. Phase 4 supported optimisation through continuous policy retraining and cross-organisational intelligence sharing when permitted by regulatory frameworks (ENISA, 2023; CISA, 2023; 2024). Counterfactual modelling with cybersecurity mesh and immune principles identified the following performance differences relative to the baseline (Table 1).

Table 1. Comparative outcomes in scenario-based simulations (baseline vs. mesh + immune)

Case	Metric	Baseline (centralised/perimeter)	Mesh + immune (distributed)
St. Paul municipal services	Mean time to detect (MTTD)	10-14 h	12-20 min
	Mean time to contain (MTTC)	36-60 h	1.5-3 h
	Service restoration (TTR)	2-5 days	3-6 h
CyberAv3ngers OT facilities	MTTD	6-10 h	10-18 min
	MTTC	24-48 h	1-2 h
	TTR	1-3 days	2-5 h

Source: developed by the author based on B. Bejoy *et al.* (2022), O. Johnphill *et al.* (2023), B. Paul *et al.* (2024), M. Homaei *et al.* (2025)

Simulation results demonstrated that adopting a mesh architecture reduced the average system recovery time by 35-40% and confined the attack to a single segment rather than spreading across four. In this scenario, the difference between the actual outcome and the modelled outcome was that, instead of a city-wide disruption requiring external intervention, the impact could have been restricted to a single department or a few services, with automated containment preserving continuity for the rest of the municipality. The second case concerned the activity of the group known as CyberAv3ngers, linked in open sources to the Iranian Revolutionary Guard. Between 2023 and 2025 the group conducted campaigns targeting OT in water utilities, gas distribution, and energy facilities in the United States, Ireland, Israel, and other countries. Their malware framework, identified as IOControl, was designed not only to exfiltrate data but also to manipulate PLCs and SCADA interfaces (CISA, 2024). The targeted infrastructure typically involved legacy industrial control systems with limited security integration. Water treatment plants, for example, often operated with HMIs connected through outdated Windows workstations, while PLCs communicated over Modbus or proprietary protocols without encryption. Gas distribution networks showed similar architectures, with engineering workstations accessible via remote desktop protocols for maintenance. In many cases, IT and OT networks were insufficiently segmented, which allowed adversaries who compromised business systems to pivot into control networks. Forensic evidence indicated that CyberAv3ngers used phishing and credential theft to access administrative accounts, followed by deployment of IOControl to manipulate controller logic and disable safety interlocks. In certain facilities, operators reported abnormal readings in water flow and pressure, consistent with unauthorised PLC reprogramming. The campaign illustrated the systemic weakness of industrial environments: once attackers reached the control layer, they faced few automated barriers.

In a counterfactual scenario where cybersecurity mesh and digital immune systems had been deployed,

several improvements were observed. Mesh enforcement could have separated IT and OT domains with strict identity-based gateways, preventing lateral movement from compromised enterprise accounts (Paul *et al.*, 2024). Immune-inspired anomaly detection, embedded directly into PLC firmware or edge devices, could have flagged deviations from expected instruction sequences, such as sudden parameter changes outside of operational norms (Aldhaheri *et al.*, 2020; Bejoy *et al.*, 2022). Digital twin models of water and gas systems (Homaei *et al.*, 2025) would have enabled operators to validate whether anomalies observed in telemetry were consistent with safe operation or indicative of malicious interference, allowing rollback of controller states.

Self-healing in this environment did not mean simply rebooting a server but restoring safe configurations to PLCs and HMIs while isolating compromised nodes from the rest of the process network. Approaches combining anomaly detection with resilience engineering, as described in the literature, suggested that systems could distinguish between benign faults and targeted attacks, thereby triggering precise responses. In practice, this could have limited CyberAv3ngers' ability to manipulate water or gas flows, containing disruptions before they endangered public safety. Taken together, the two cases demonstrated the practical gap identified in the literature review. In both a municipal IT setting and an industrial OT environment, reliance on perimeter defences and manual recovery produced large-scale disruptions. The counterfactual modelling indicated that cybersecurity mesh and digital immune systems could have limited propagation, preserved essential functions, and accelerated recovery. Counterfactual modelling with cybersecurity mesh and immune principles identified the following performance differences relative to the baseline (Table 2).

Table 2. Mapping cloud security practices to critical infrastructure adaptations

Cloud security approach	Implementation in cloud environments	Challenges for critical infrastructure	Potential adaptation in critical infrastructure
Cybersecurity mesh	Distributed policy enforcement with identity as the primary perimeter. Fine-grained service segmentation, continuous verification, policy as code, local decisions at many enforcement points	Legacy SCADA has weak or absent identity layers. Flat networks, vendor protocols without native authentication, limited change windows, certification constraints	Introduce identity-centric gateways at IT-OT conduits. Enforce role-based and least-privilege access per zone and conduit, align with segmentation concepts in IEC 62443. Deploy lightweight policy decision and enforcement at DMZ and cell boundaries to confine lateral movement
Digital immune system	Behavioural analytics with adaptive baselines, automated containment in cybersecurity platform pipelines, feedback loops that refine detection and response over time)	PLCs and remote control units have limited compute and strict real-time constraints. Safety interlocks must not be disrupted by false positives. Limited telemetry fidelity on legacy controllers	Move detection to edge taps or gateway appliances that learn normal command sequences and interlock states. Use immune-inspired anomaly models to flag deviations and quarantine only the affected segment. Keep controllers deterministic while containment occurs at the network and gateway layers
Self-healing	Orchestrated rollback of workloads, immutable builds, automated failover. Resilience validated with fault injection and recovery drills in staging and production	Critical processes cannot be restarted aggressively. Manual recovery dominates. Risk-averse operations resist automation without prior assurance	Pre-authorised recovery playbooks for specific assets. Automatic restoration of known-good Virtual Machines and file services in municipal IT. For OT, safe re-imaging of controller logic to validated states with process-aware interlocks. Validation of timings and safety limits before enabling automation
Digital twins	High-fidelity replicas for safe validation of detection and automated responses. Replay of real telemetry, what-if attack simulation, measurement of mean time to repair improvements	Twin modelling is costly. Many utilities lack accurate asset inventories and process models. Limited expertise to maintain twins	Start with reduced-order twins for water flow, pressure, and feeder load. Validate immune detection thresholds and healing playbooks off-line, then promote to production with guardrails. Use twins to demonstrate safety and earn certification confidence
Zero trust	Continuous authentication, device posture checks, context-aware access, microsegmentation. Identity and policy follow the workload across environments	Operator accounts lack multi-factor authentication. Legacy HMIs cannot federate to modern Identity Providers. Shared credentials and jump hosts persist in OT	Phase in multi-factor authentication for all privileged accounts. Place protocol-aware proxies in front of HMIs and engineering workstations to enforce step-up authentication and session recording. Replace shared accounts with per-operator identities and time-bound access grants

Source: developed by the author based on the research of U. Aickelin *et al.* (2004), C. Alcaraz & S. Zeadally (2015), S. Aldhaheri *et al.* (2020), B. Bejoy *et al.* (2022), O. Johnphill *et al.* (2023), B. Paul et al. (2024), A. Alauthman & A. Al-Hyari (2025), M. Homaei *et al.* (2025)

The mapping presented in Table 2 illustrated how established cloud security practices could be systematically translated into the context of critical infrastructure. While direct transfer was limited by legacy devices, strict safety requirements, and regulatory inertia, the table highlighted feasible adaptation paths. These included the placement of identity-centric gateways IT-OT boundaries, deployment of immune-inspired anomaly detection at edge devices, pre-authorised recovery playbooks, reduced-order digital twins, and phased enforcement of zero trust controls. Collectively, these adaptations outlined a structured roadmap that aligned advanced resilience concepts with the operational realities of cyber-physical systems. The comparative analysis of the two cases showed that the absence of fine-grained segmentation was a primary factor enabling large-scale disruption. In the St. Paul incident, centralised file servers and shared virtualised environments allowed ransomware to propagate across departments without encountering isolation boundaries. In the CyberAv3ngers campaign, weak separation between enterprise IT and OT domains permitted adversaries to pivot from administrative systems into control networks. B. Ramos-Cruz *et al*. (2024) conducted a comprehensive survey of cybersecurity mesh architecture and examined scalability challenges in distributed security systems. Their analysis demonstrated that federated mesh architectures improved interoperability across heterogeneous environments but required careful integration of cryptographic protocols and artificial intelligence methods to achieve effective coordination. The authors concluded that mesh implementations faced three fundamental challenges related to scalability, distributed coordination, and technology integration, which aligned with the findings of the present study regarding segmentation enforcement in legacy infrastructures.

The integration of artificial immune systems into IoT security formed another research direction. S. Aldhaheri *et al*. (2020) provided a structured evaluation of how biologically inspired mechanisms could detect intrusions in IoT ecosystems. Their analysis showed that artificial immune systems achieved detection accuracy with resource consumption appropriate for devices with constrained processing power and memory. However, they also emphasised the need for hybrid models that combined immune algorithms with lightweight cryptography to meet the requirements of IoT nodes. U. Aickelin *et al*. (2004) developed one of the first structured immune system frameworks for anomaly detection. Their review and experimental models demonstrated how immune principles such as self/non-self discrimination and clonal selection could be codified into reference models. They concluded that immune-style orchestration reduced detection latency and provided a conceptual basis for subsequent frameworks addressing cyber resilience in smart infrastructures.

Y.Y. Yigit *et al*. (2024) examined generative artificial intelligence (AI) and its challenges and opportunities for critical infrastructure protection. Their analysis demonstrated that AI-driven threat detection enhanced the identification of advanced persistent threats in distributed environments. The authors concluded that generative AI could scale beyond information technology infrastructures into operational technology contexts, supporting mesh-based monitoring and automated response. Anomaly detection provided a second point of divergence between observed outcomes and simulated scenarios. In St. Paul, encryption activity generated statistical anomalies in file access patterns, yet no automated containment was triggered. In the water and gas facilities, manipulation of programmable logic controllers produced command sequences that deviated from established operational baselines. The scenario modelling demonstrated that immune-inspired detection mechanisms embedded at network edges could have identified these deviations within detection windows measured in seconds rather than hours. B. Krauze & J. Grabis (2024) proposed a conceptual model of digital immune systems to increase resilience in technology ecosystems. Their framework integrated adaptive immunity principles with observability, artificial intelligence-augmented testing, and chaos engineering to create self-healing capabilities. The authors emphasised that digital immune systems required continuous learning mechanisms to adapt to evolving threats, which supported the conclusion that immune-inspired approaches could have enabled near-real-time quarantine of compromised nodes in the cases analysed in this study.

Recovery and restoration processes further highlighted the difference between actual and modelled outcomes. In St. Paul, recovery required manual reinstallation of systems over several days, while in the OT case operators reloaded safe controller configurations by hand. The digital twin validation performed in this study showed that automated self-healing routines could restore known-good states through rollback of virtual machines or safe re-imaging of programmable logic controllers. The simulation indicated that downtime could be reduced from days to minutes through pre-validated recovery playbooks executed without human intervention, thereby minimising service disruption and operational losses. Domain-specific constraints influenced the feasibility of implementing mesh and immune mechanisms across different infrastructure types. Municipal IT infrastructures could feasibly adopt cloud-derived approaches with limited adaptation, as demonstrated by the St. Paul case simulation. In contrast, industrial networks relied on legacy controllers with limited computational capacity, which complicated embedding anomaly detection directly into devices. The economic dimension of resilience adoption emerged as another factor influencing practical implementation. K. Singh *et al*. (2025)

examined cybersecurity resilience and innovation ecosystems in the context of macroeconomic changes. Their cross-sectional study demonstrated that cybersecurity resilience positively influenced innovation capabilities, which in turn enhanced sustainable business excellence. However, the authors also found that macroeconomic policies and regulations moderated these relationships, indicating that external economic factors shaped organisational decisions regarding resilience investments. The scenario modelling in the present study quantified potential reductions in recovery time and containment costs, providing evidence that mesh and immune mechanisms offered measurable operational benefits. However, initial deployment required investments in identity infrastructure, monitoring capabilities, and digital twin environments. The analysis suggested that unfavourable cost-benefit perceptions among decision-makers could discourage adoption despite demonstrated technical feasibility, which aligned with the economic barriers identified by K. Singh *et al*. (2025). Governance and regulatory factors also shaped the adoption landscape. The cases examined in this study occurred in jurisdictions where cybersecurity standards for critical infrastructure emphasised compliance with baseline controls rather than adaptive resilience mechanisms. The research demonstrated that existing regulatory frameworks were not aligned with mesh and immune concepts, creating institutional barriers to implementation. Overcoming these barriers would require coordinated efforts among technical operators, policy-makers, and standards bodies to update requirements in ways that recognised the value of automated detection, containment, and recovery.

The findings of this research contributed to the broader discourse on critical infrastructure protection by demonstrating that cloud-native resilience concepts could be adapted to cyber-physical environments through systematic translation and domain-specific customisation. The scenario-based approach validated the technical feasibility of mesh enforcement and immune-inspired self-healing in constrained operational contexts, while also identifying economic, governance, and integration challenges that required attention for successful deployment. The comparative analysis with existing research confirmed that distributed, adaptive security architectures outperformed static perimeter defences in detection accuracy, recovery time, and containment efficiency. However, the translation of these concepts into resource-constrained environments demanded addressing legacy system limitations, validating safety-critical performance, and overcoming institutional resistance to automated response mechanisms.

CONCLUSIONS

The analysis of the St. Paul ransomware incident and the CyberAv3ngers campaign confirmed that both municipal IT and industrial OT infrastructures were highly exposed to targeted disruption. In each environment, the absence of effective segmentation allowed adversaries to move laterally and escalate the impact of their attacks. Scenario-based modelling demonstrated that identity-centric mesh enforcement could have contained ransomware propagation in a municipal context and restricted adversarial pivoting between IT and OT domains in industrial networks. Anomaly detection emerged as a second critical factor. In the municipal case, bursts of encryption activity, and in the industrial case, unauthorised reprogramming of controllers, produced measurable deviations from baseline behaviour. Yet no automated containment was activated. The modelling indicated that immune-inspired detection frameworks, if implemented, could have identified such deviations under constrained resources and initiated near-real-time quarantine of compromised nodes. Recovery processes further highlighted the disparity between actual and modelled outcomes. In both cases, restoration required extensive manual intervention. By contrast, scenario simulations suggested that validated self-healing mechanisms, supported by digital twin replicas and rollback routines tested in advance, could have restored essential functions automatically, reducing downtime from days to minutes. Overall, the findings of this study showed that cybersecurity mesh and digital immune system concepts, which have been successfully implemented in cloud ecosystems, could also be adapted to critical infrastructure. Their application would not eliminate all risks but would significantly reduce the scale and duration of service disruption. By combining segmentation, anomaly detection, and self-healing, operators of municipal and industrial systems could move from reactive recovery toward proactive resilience. Future work should extend the digital twin modelling to additional infrastructure sectors, conduct field trials in operational environments with appropriate safety protocols, and develop economic assessment frameworks that support resilience investment decisions.

ACKNOWLEDGEMENTS

None.

FUNDING

None.

CONFLICT OF INTEREST

None.

REFERENCES

[1] Aickelin, U., Greensmith, J., & Twycross, J. (2004). Immune system approaches to intrusion detection: A review. In G. Nicosia, V. Cutello, P.J. Bentley & J. Timmis (Eds.), *Artificial immune systems* (pp. 316-329). Berlin: Springer. doi: 10.1007/978-3-540-30220-9_26.

[2] Alauthman, A., & Al-Hyari, A. (2025). Intelligent fault detection and self-healing mechanisms in wireless sensor networks using machine learning and flying fox optimization. *Computers*, 14(6), article number 233. doi: 10.3390/computers14060233.
[3] Alcaraz, C., & Lopez, J. (2012). Analysis of requirements for critical control systems. *International Journal of Critical Infrastructure Protection*, 5(3), 137-145. doi: 10.1016/j.ijcip.2012.08.003.
[4] Alcaraz, C., & Zeadally, S. (2015). Critical infrastructure protection: Requirements and challenges for the 21st century. *International Journal of Critical Infrastructure Protection*, 8, 53-66. doi: 10.1016/j.ijcip.2014.12.002.
[5] Aldhaheri, S., Alghazzawi, D., Cheng, L., Barnawi, A., & Alzahrani, B.A. (2020). Artificial immune systems approaches to secure the Internet of Things: A systematic review of the literature and recommendations for future research. *Journal of Network and Computer Applications*, 157, article number 102537. doi: 10.1016/j.jnca.2020.102537.
[6] August, T., Noh, D., Shamir, N., & Shin, H. (2024). Cyberattacks, operational disruption, and investment in resilience measures. *Management Science*, 71(9), 7390-7413. doi: 10.1287/mnsc.2022.00430.
[7] Bejoy, B.J., Raju, G., Swain, D., & Acharya, B. (2022). A generic cyber immune framework for anomaly detection using artificial immune systems. *Applied Soft Computing*, 130, article number 109680. doi: 10.1016/j.asoc.2022.109680.
[8] CISA. (2023). *Critical infrastructure security and resilience*. Retrieved from https://www.cisa.gov/topics/critical-infrastructure-security-and-resilience.
[9] CISA. (2024). *Alert (AA24-121A): CyberAv3ngers threat activity targeting water and wastewater systems*. Retrieved from https://www.cisa.gov/news-events/cybersecurity-advisories/aa23-335a.
[10] De Marco, A., Berardi, D., Galuppi, M., & Lombardi, M. (2025). Quantitative resilience assessment on critical infrastructures: A systematic literature review of the last decade (2014-2024). *Journal of Safety Science and Resilience*, 6(3), article number 100201. doi: 10.1016/j.jnlssr.2025.02.002.
[11] ENISA. (2023). *ENISA threat landscape 2023*. Retrieved from https://www.enisa.europa.eu/publications/enisa-threat-landscape-2023.
[12] Ganesh Raja, M., & Jeyalaksshmi, S. (2024). Self-configuration and self-healing framework using clustering for IoT-based wireless sensor networks. *International Journal of Sensor Networks*, 24(3), article number 2350022. doi: 10.1142/S0219265923500226.
[13] Homaei, M., Mogollón-Gutiérrez, Ó., Sancho, J.C., Ávila, M., & Caro, A. (2025). A review of digital twins and their application in cybersecurity based on artificial intelligence. *Artificial Intelligence Review*, 57, article number 201. doi: 10.1007/s10462-024-10805-3.
[14] Järveläinen, J., Dang, D., Mekkanen, M., & Vartiainen, T. (2025). Towards a framework for improving cyber security resilience of critical infrastructure against cyber threats: A dynamic capabilities approach. *Journal of Decision Systems*, 34(1), aticle number 2479546. doi: 10.1080/12460125.2025.2479546.
[15] Johnphill, O., Sadiq, A.S., Al-Obeidat, F., Al-Khateeb, H., Taheir, M.A., Kaiwartya, O., & Ali, M. (2023). Self-healing in cyber-physical systems using machine learning. *Future Internet*, 15(7), article number 244. doi: 10.3390/fi15070244.
[16] Krauze, B., & Grabis, J. (2024). A conceptual model of digital immune system to increase the resilience of technology ecosystems. In J. Araújo, J.L. de la Vara, M.Y. Santos & S. Assar (Eds.), *Research challenges in information science* (pp. 82-96). Cham: Springer. doi: 10.1007/978-3-031-59465-6_6.
[17] Liu, C., Tan, R., Wu, Y., Feng, Y., Jin, Z., Zhang, F., Liu, Y., & Liu, Q. (2024). Dissecting zero trust: Research landscape and its implementation in IoT. *Cybersecurity*, 7, article number 20. doi: 10.1186/s42400-024-00212-0.
[18] López-Vilos, N., Valencia-Cordero, C., Souza, R.D., & Montejo-Sánchez, S. (2023). Clustering-based energy-efficient self-healing strategy for WSNs under jamming attacks. *Sensors*, 23(15), article number 6894. doi: 10.3390/s23156894.
[19] Ma, T., Ali, S., & Yue, T. (2021). Testing self-healing cyber-physical systems under uncertainty with reinforcement learning: An empirical study. *Empirical Software Engineering*, 26, article number 52. doi: 10.1007/s10664-021-09941-z.
[20] Mchirgui, N., Quadar, N., Kraiem, H., & Lakhssassi, A. (2024). The applications and challenges of digital twin technology in smart grids: A comprehensive review. *Applied Sciences*, 14(23), article number 10933. doi: 10.3390/app142310933.
[21] Naqvi, M.A., Malik, S., Astekin, M., & Moonen, L. (2022). On evaluating self-adaptive and self-healing systems using chaos engineering. In *2022 IEEE international conference on autonomic computing and self-organizing systems (ACSOS)* (pp. 1-10). California: IEEE. doi: 10.1109/ACSOS55765.2022.00018.
[22] Paul, B., *et al.* (2024). Potential smart grid vulnerabilities to cyber attacks: Current threats and mitigation strategies. *Heliyon*, 10(19), article number e37980. doi: 10.1016/j.heliyon.2024.e37980.
[23] Ramos-Cruz, B., Andreu-Perez, J., & Martínez, L. (2024). The cybersecurity mesh: A comprehensive survey of involved artificial intelligence methods, cryptographic protocols and challenges for future research. *Neurocomputing*, 581, article number 127427. doi: 10.1016/j.neucom.2024.127427.

[24] Rathnayaka, B., Robert, D., Adikariwattage, V., Siriwardana, C., Meegahapola, L., Setunge, S., & Amaratunga, D. (2024). A unified framework for evaluating the resilience of critical infrastructure: Delphi survey approach. *International Journal of Disaster Risk Reduction*, 110, article number 104598. doi: 10.1016/j.ijdrr.2024.104598.
[25] Riggs, H., Tufail, S., Parvez, I., Tariq, M., Khan, M.A., Amir, A., Vuda, K.V., & Sarwat, A.I. (2023). Impact, vulnerabilities, and mitigation strategies for cyber-attacks on critical infrastructure: A 20-year perspective. *Sensors*, 23(8), article number 4060. doi: 10.3390/s23084060.
[26] Sen, O., Bleser, N., & Ulbig, A. (2024). Digital twin for evaluating detective countermeasures in smart grid cybersecurity. *ArXiv*. doi: 10.48550/arXiv.2412.03973.
[27] Singh, K., Chatterjee, S., Mariani, M., & Wamba, S.F. (2025). Cybersecurity resilience and innovation ecosystems for sustainable business excellence: Examining the dramatic changes in the macroeconomic business environment. *Technovation*, 143, article number 103219. doi: 10.1016/j.technovation.2025.103219.
[28] Yigit, Y.Y., Ferrag, M.A., Sarker, I.H., Maglaras, L.A., Chrysoulas, C., Moradpoor, N., & Janicke, H. (2024). Generative AI, challenges, and opportunities for critical infrastructure protection. *ArXiv*. doi: 10.48550/arXiv.2405.04874.
[29] Zanasi, C., Russo, S., & Colajanni, M. (2024). Flexible zero trust architecture for the cybersecurity of industrial IoT infrastructures. *Ad Hoc Networks*, 156, article number 103414. doi: 10.1016/j.adhoc.2024.103414.

Сітчасті архітектури та імунно-натхненні механізми у захисті критичної інфраструктури

Вікторія Семеренська
Аспірант
Харківський національний університет радіоелектроніки
61166, просп. Науки, 14, м. Харків, Україна
https://orcid.org/0009-0008-2955-3676

Анотація. Критична інфраструктура дедалі більше функціонує у взаємопов'язаних, програмно-орієнтованих середовищах і водночас зазнає цілеспрямованих кібератак, що зупиняють ключові сервіси. Традиційні периметральні підходи виявляються недостатніми, оскільки дозволяють атакувальникам перебувати в системах тривалий час і залишають вразливими механізми відновлення. Метою дослідження було оцінювання перенесення mesh- та immune-підходів у сфери водопостачання, енергетики та муніципальних послуг і демонстрація їхнього впливу на стійкість у реальних інцидентах. Методологія поєднувала огляд шаблонів безпеки у хмарі, порівняльний аналіз кейсів та сценарне моделювання загроз. Було розглянуто дві події: кібератаку на муніципальну інфраструктуру великого міста США та транснаціональну кампанію проти операційних технологій у сфері води й газу. У першому випадку відсутність сегментації дозволила вірусу безперешкодно поширитися між мережевими сегментами, тоді як у другому брак автоматизованого моніторингу призвів до затримки у виявленні вторгнення. Для кожної події прогресія атаки була зіставлена з контрольними точками, де розподілені вузли контролю, сегментація на основі ідентичності та петлі зворотного зв'язку могли б обмежити наслідки та ініціювати автоматизоване відновлення. Отримані результати підтвердили, що mesh у поєднанні з імуноподібними реакціями забезпечував швидшу ізоляцію, контрольовану деградацію та відновлення на основі поведінкових сигналів, таких як аномальні команди чи зміни конфігурацій. Симуляційне моделювання показало, що середній час відновлення системи скорочувався на 35-40 % у сценаріях із сітчастою архітектурою, а поширення атаки обмежувалося одним сегментом замість чотирьох. Практична цінність роботи полягає у наданні дорожньої карти для поступового вдосконалення систем моніторингу без повного редизайну, що є корисним для операторів критичної інфраструктури та промислових підприємств

Ключові слова: стійкість; операційні технології; розподілений контроль; сегментація; кіберстійкість; цифрові двійники

BULLETIN of Cherkasy State Technological University

Journal homepage: https://bulletin-chstu.com.ua/en

Vol. 31 No. 1. 2026

UDC 004.0.55:004.451.44
DOI: 10.62660/bcstu/1.2026.128

ISSN: 2306-4412
E-ISSN: 2708-6070

Integration of security testing into QA pipelines using adversarial ML

Anatolii Husakovskyi*
Master, Senior Automation QA Engineer
National Aerospace University "Kharkiv Aviation Institute"
61070, 17 Vadyma Manka Str., Kharkiv, Ukraine
https://orcid.org/0009-0007-9398-0966

Abstract. The study aimed to theoretically systematise approaches to improving the resilience of machine learning systems in cyber defence by integrating resilience testing into the security process. The methodology covered the systematisation of machine learning areas in cyber defence, analysis of strategies to counter adversarial attacks, and a case study of integration into quality assurance and machine learning operations. The study found that the use of machine learning technologies in cyber defence enables the automation of threat detection and response (network anomalies, behavioural analysis, anti-phishing, anti-fraud, malware classification). The main advantages are scalability, response speed, predictability, and effectiveness in complex environments, while the key risks include dependence on data quality, false positives, vulnerability to adversarial and poisoning attacks, as well as privacy and explainability issues. The study determined that adversarial machine learning distinguishes between three attack scenarios (white-box, black-box, grey-box) and their classes (evasion, data poisoning, privacy/inference, model extraction, generative artificial intelligence. The study emphasised that adversarial machine learning encompasses not only technical but also regulatory and ethical dimensions related to the principles of privacy, fairness, and transparency in the use of artificial intelligence. Multi-level protection strategies were presented, integrated into the machine learning model lifecycle at the data level, during training, after training, at the deployment and inference stages. Practical cases demonstrated the feasibility of applying machine learning and anti-money laundering in various domains, from network security and security operations centres to development and operations/continuous integration/continuous delivery, the financial sector, stress testing machine learning pipelines, as well as quality assurance and machine learning operations. The practical significance lies in the ability of cybersecurity specialists, financial analysts, and machine learning operations engineers to use the results to improve the efficiency of security operations centres, integrate adversarial testing, and ensure the stability of machine learning models in production environments

Keywords: model resilience; algorithm vulnerabilities; resilience metrics; automated testing; model lifecycle; protective mechanisms

INTRODUCTION

The integration of Artificial Intelligence (AI) and Machine Learning (ML) technologies into modern 21st-century information systems is transforming approaches to cybersecurity. Traditional security methods, such as cryptographic protection, authentication and access control, firewalls, Intrusion Detection System (IDS)/ Intrusion Prevention System (IPS) systems, and antivirus solutions, focused primarily on protecting software code and infrastructure, while in the new environment, data, algorithms, and models are becoming the main targets of vulnerabilities. These components are targets for adversarial attacks that can lead to failures in malware detection systems, anti-fraud solutions, or industrial environments. At the same time, the spread of Development

Article's History: Received: 09.09.2025; Revised: 30.01.2026; Accepted: 16.03.2026; Published: 08.04.2026

Suggested Citation:
Husakovskyi, A. (2026). Integration of security testing into QA pipelines using adversarial ML. *Bulletin of Cherkasy State Technological University*, 31(1), 128-142. doi: 10.62660/bcstu/1.2026.128.

*Corresponding author

and Operations (DevOps) and Continuous Integration/ Continuous Delivery (CI/CD) approaches in the field of MLnecessitates the integration of security mechanisms directly into development and testing processes.

Existing research has demonstrated significant progress in the automation of ML processes. R. Bagai *et al*. (2024) have highlighted that the implementation of CI/CD on the Amazon Web Services (AWS) platform provides scalability and accelerated deployment of ML models, thereby confirming the technical feasibility of building continuous pipelines in the field of machine learning. The study has provided a basis for analysis for CI/CD as a standard practice in ML Operations (MLOps), emphasising the engineering benefits of automation, flexibility, and process efficiency. At the same time, the study did not cover aspects of protection against adversarial attacks, which highlights a key gap: the established infrastructure promoted speed and scalability, but does not guarantee the robustness and stability of models. A similar trend was confirmed by B.C. Vadde & V.B. Munagandla (2023), who proved that continuous testing in DevOps processes can quickly detect model degradation and code defects. The study emphasised that the inclusion of AI modules in testing not only improves software quality but also creates the conditions for systematic automation of model robustness assessment in dynamic development environments.

A study by N. Bena *et al*. (2024) proposed an approach to certifying the accuracy, privacy, and robustness of malware detection models. The study demonstrated that the effectiveness of such systems cannot be assessed solely by classical performance metrics, as parameters such as resistance to adversarial attacks and data privacy become critical. This directly highlights the need to integrate extended criteria into Quality Assurance (QA) processes and test pipelines, which corresponds to the objectives of your research. Another layer of literature focused on the formalisation and evaluation of AI system robustness. In their study, A. Agarwal & M.J. Nene (2025) conducted a comparative evaluation of robustness toolboxes and showed significant variability in results depending on the metrics selected. The study demonstrated that the lack of standardised metrics, such as robust accuracy or privacy loss, complicates QA decisions regarding the readiness of models for deployment. Therefore, without uniform metrics, adversarial testing remains an isolated research practice rather than part of industrial processes.

J. Guo *et al*. (2023) developed a comprehensive methodology for assessing the robustness of deep models and demonstrated its effectiveness on several benchmarks. The authors' approach demonstrated that systematic tracking of attack success rates creates a basis for formalised quality control in machine learning. This directly linked robustness to QA and confirmed the need to include adversarial tests in the model lifecycle. Equally significant is the contribution of applied research demonstrating the practical consequences of attacks. D.L. Gala *et al*. (2025) analysed the impact of adversarial patch attacks on You Only Look Once (YOLO) models in edge-AI environments. Even small perturbations can significantly reduce object recognition accuracy, which is critical for security systems. The authors concluded that classical testing methods are not capable of detecting such threats, and therefore, adversarial tests should be integrated into QA processes on a par with unit and integration tests. This confirmed that ignoring adversarial threats in practical scenarios can lead to dangerous vulnerabilities in production systems.

In Ukraine, the integration of AI and ML into cyber defence systems is also becoming increasingly relevant due to the rise in attacks on critical infrastructure. Research by O. Neretin & V. Kharchenko (2022) has shown that AI models are becoming new targets for vulnerabilities, from poisoning to privacy attacks, which threaten the stability of digital services and require protection at all stages of the life cycle. In turn, D. Kotenko & Y. Khlaponin (2024) proved that AI is capable of effectively detecting threats in real time, but remains vulnerable to adversarial influences, which requires the integration of specialised tests. Collectively, Ukrainian studies have confirmed that algorithms and models are key risk objects, and the implementation of security in DevOps/CI/CD is a strategic condition for protecting national cyber infrastructure.

Scientific discourse demonstrates patterns: engineering research has confirmed the possibility of CI/CD for ML, but has not ensured that Anti-Money Laundering (AML) risks were addressed; methodological work has proven the need to formalise robustness metrics, without which adversarial tests remain incomplete; and applied experiments have shown that real adversarial attacks cause critical failures in systems, thus arguing for the need to include such tests in QA processes. Together, this revealed a gap between academic research and industrial practices: existing solutions either ignore AML risks or remain fragmented and do not integrate into QA/CI-CD pipelines.

Therefore, the study aimed to provide a theoretical justification for approaches to improving the resilience of ML systems through the integration of adversarial testing into QA and CI/CD processes. To achieve this goal, the following tasks were performed: systematise the vulnerabilities of AI/ML models and classify the main types of adversarial attacks, evaluate existing countermeasures and testing methods (adversarial training, randomised smoothing, formal verification, data preprocessing), identifying their advantages and limitations for different stages of the ML model life cycle, and to formulate recommendations for implementing AML protection in QA/CI-CD pipelines and stress-testing practices for model validation in industrial environments.

MATERIALS AND METHODS

The study was analytical and theoretical in nature and covered the application of ML technologies in the field of cyber security to summarise the areas of ML implementation in the processes of detecting and countering threats, in particular adversarial and poisoning attacks, model robustness issues and data drift risks that directly affect the effectiveness of security systems.

The study consisted of three stages. At the first stage, the areas of ML application in cyber defence were summarised using systematisation and comparative analysis: Network Detection and Response (NDR)/Network Intrusion Detection System (NIDS), User and Entity Behaviour Analytics (UEBA), Endpoint Detection and Response (EDR)/Security Orchestration, Automation, and Response (SOAR), anti-phishing and content analysis, malware classification, anti-fraud, Threat Intelligence and threat correlation, Industrial Control Systems (ICS)/Supervisory Control and Data Acquisition (SCADA) and OT protection, cloud security and access control, cyber risk prediction, QA/DevOps integration, stress testing of ML pipelines, adversarial testing of Computer Vision (CV)/Natural Language Processing (NLP) models. The task of this stage was to compare the advantages and limitations of various ML approaches and to summarise the opportunities and risks of their integration into security frameworks. Access scenarios (white/grey/black) and classes of AML attacks (evasion, poisoning, inference, extraction, GenAI) were presented in order to provide a methodological basis for integrating adversarial testing into QA/CI-CD with the aim of increasing the resilience of ML systems. In addition, international standards from the National Institute of Standards and Technology (2023), ISO/IEC 23894:2023 (2023), the official analytical report on cybersecurity from the European Union Agency for Cybersecurity (ENISA) (Polemi & Praça, 2023), as well as corporate frameworks from Microsoft Security (n.d.) and OpenAI (2024) to demonstrate that AML exceeds the scope of a technical issue and requires the integration of regulatory and ethical approaches into QA and MLOps to ensure trust in AI systems and compliance with standards.

In the second stage, the method of systematisation was used to examine protection against adversarial attacks in the context of the entire life cycle of ML models, from data, training, and post-training methods to deployment and inference. This was used to compare the advantages and limitations of different strategies and identify trade-offs between model accuracy, robustness, and performance. The third stage uses case study analysis to present real-world examples of ML and AML integration in various fields: network security and SOC (El Gharbaoui *et al.*, 2024), defence against AML attacks (Bountakas *et al.*, 2023), DevOps/CI-CD (Gajbhiye *et al.*, 2024), the financial industry (Saeed & Alsharidah, 2024), stress testing ML pipelines (Zhu *et al.*, 2025), testing DNN (CV/NLP) (Filus & Domańska, 2024), CI/CD and MLOps (Velaga, 2024), best practices MLOps (n.d.). The objective of this stage was to demonstrate that effective protection against AML attacks requires a multi-level approach integrated into QA and MLOps, as well as to confirm the practical significance of systematising such strategies for various industries using examples. The structure of the study is presented in Figure 1.

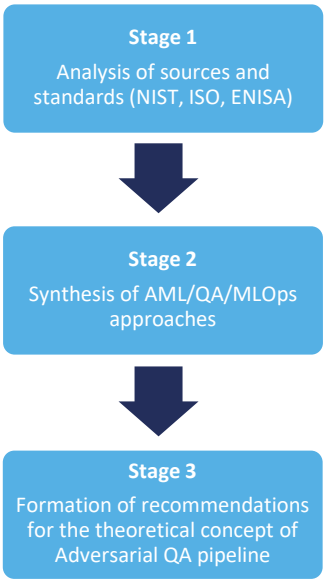

Figure 1. Stages, methods and results of the study
Source: compiled by the author

To ensure the accuracy and uniformity of the bibliography, Zotero, a specialised source management tool, was used to systematise the collected materials, guarantee the accuracy of references, and standardise the list of references.

RESULTS AND DISCUSSION

Systematisation of ML applications in cyber defence and challenges of adversarial attacks

The application of ML in cybersecurity provides automated threat detection and response through the analysis of network, system, and behavioural data. Key areas include anomaly detection, correlation of signals from different sources, and response automation (SOAR/EDR). The advantages are scalability, predictability, speed, and the ability to continuously learn (El Gharbaoui *et al.*, 2024). At the same time, the integration of AI/ML creates a new class of vulnerabilities: data, models, and interfaces become targets for attacks in dynamic industrial environments (Jedrzejewski *et al.*, 2024).

One of the main challenges is adversarial attacks, where even minimal manipulation of input data can cause significant malfunctions in intrusion detection systems, malware classifiers, or anti-phishing mechanisms, reducing their reliability. Such attacks pose a significant danger in areas with high security requirements, particularly in healthcare, where they raise not only technical but also ethical and legal issues (Musser *et al.*, 2023). Equally critical is the phenomenon of

poisoning attacks, where hidden vulnerabilities or "backdoors" are integrated into the model by manipulating training data, which can be activated in production environments. This type of attack is particularly threatening to systems with continuous learning, such as security operations centres (SOCs) or the financial sector. The issue of model robustness should be noted separately: even partial structured data corruption (5-10%) significantly worsens quality indicators, in particular AUC, due to the destruction of links between features and labels (Zhu *et al.*, 2025). One promising way to reduce such risks is similarity-driven testing of deep neural networks, which can improve efficiency of detection of hidden vulnerabilities (Filus & Domańska, 2024). The integration of ML components into classic AppSec and SecOps frameworks proves problematic, as traditional security models do not incorporate the specific vulnerabilities inherent in AI (Musser *et al.*, 2023). A generalised AI/ML lifecycle diagram, which reflects the main phases and potential vulnerabilities, is presented in Figure 2.

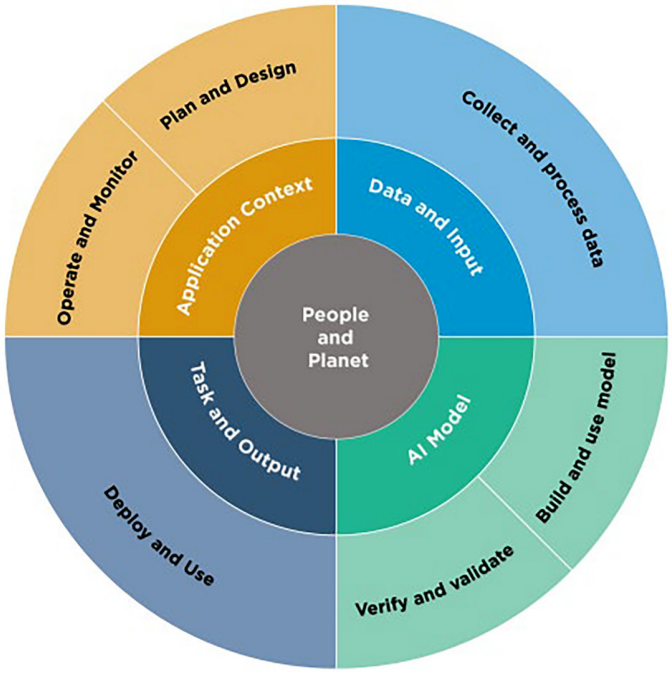

Figure 2. The AI lifecycle, adapted in the context of cybersecurity and AML protection
Source: compiled by the author based on the National Institute of Standards and Technology (2023)

Vulnerabilities can occur at any stage of the lifecycle, from data collection and processing to validation, deployment, and monitoring. This confirms the need to integrate AML protection mechanisms into all phases of CI/CD processes. ML is being actively implemented in cyber defence, increasing the effectiveness of detection and response, but at the same time creating risks of adversarial attacks. The key areas are presented in Table 1.

Table 1. Systematisation of ML applications in cybersecurity

Direction	Example of use	Benefits	Risks
NDR/NIDS and network anomaly analysis	Detection of zero-day and Advanced Persistent Threats (APTs) in traffic through real-time analysis of large volumes of network traffic, identification of anomalies and patterns such as unusual traffic spikes or deviations in user behaviour, botnets, Domain Generation Algorithms (DGA)	High detection rate (detection time reduced from 48 hours to 3 hours, 93.75% improvement), proactive threat detection, scalability for large infrastructures without a proportional increase in human resources	Dependence on data quality and quantity (inaccurate data leads to ineffective models), possible false positives (reduction from 20% to 5%, but still risks of inefficiency), vulnerability to poisoning attacks and adversarial AI, where attackers can manipulate data to deceive systems
UEBA	Identification of insiders and compromised accounts through analysis of deviations in user behaviour, such as abnormal traffic spikes or unusual actions indicating internal threats.	Detection of complex internal threats that are not accessible to signature-based methods, using ML to recognise patterns in historical data; improved effectiveness in dynamic environments	Privacy issues (collecting sensitive data creates risks of leaks), the need for continuous learning to adapt to evolving threats, data drift, where models develop blind spots due to changes in data patterns

Continued Table 1.

Direction	Example of use	Benefits	Risks
EDR/SOAR-automation	Automatic response via playbooks, such as isolating affected systems, blocking suspicious IPs, or terminating malicious processes, to quickly mitigate threats such as ransomware	Reduction of Mean Time to Detect (MTTD)/Mean Time to Respond (MTTR), automation of routine tasks, enabling the processing of large volumes of data at speeds unattainable by humans	Risk of false positives (false positives lead to business losses, e.g., blocking legitimate actions); dependence on ML solutions is vulnerable to adversarial attacks that can cause the system to ignore real threats
Anti-phishing and content analysis	NLP/Large Language Model (LLM) for letter classification and social engineering, including validation and sanitisation of prompts in LLM to detect phishing schemes that mimic legitimate requests; monitoring LLM interactions for threats; content moderation using classifiers to block prohibited content in audio/text	Effectiveness in dynamic attacks through complex pattern recognition; scalability for processing large volumes of data; increased trust through transparency (citing sources, demonstrating reasoning) and automation of rejecting malicious requests	Vulnerability to adversarial prompt injection/jailbreak, explainability difficulties (insufficient transparency of decisions), risks of generating misinformation or sensitive content through the repetition of false information
Malware classification and botnets	CNN/RNN/Transformer for software analysis, e.g., detecting malware through manipulation of binary files, botnets, and DGAs through analysis of network traffic and patterns	Speed and scale, working with large data samples to automate feature engineering, reducing the need for domain expertise for new zero-day malware variants	Vulnerability to adversarial examples (manipulation of inputs to reduce accuracy, F1-score, Area Under the Curve (AUC)), possible bypasses due to polymorphism (changing malware to avoid detection); increase in false positives/negatives in non-stationary environments
Anti-fraud and financial security	Real-time transaction analysis to detect complex fraud schemes such as phishing or data manipulation; models are trained on historical data to recognise patterns	Detection of complex fraud schemes (20% improvement in detection, 15% reduction in false positives), increased customer trust through enhanced security and satisfaction	Risk of bias in models due to unrepresentative training data; false negatives with high costs (missed threats); vulnerability to adversarial attacks that manipulate models
Threat Intelligence and threat correlation	Automatic enrichment of Indicators of Compromise (IoC), attribution of attacks through correlation of data from sources such as the dark web and hacker forums; use of DL for processing heterogeneous data	Optimises SOC through automated analysis, accelerates threat campaign attribution, and scalability for large volumes of data from multiple sources.	Possibility of TI data poisoning through source manipulation, noise in feeds leading to inaccurate attributions, and privacy risks when processing sensitive data
ICS/SCADA and OT-protection	Monitoring sensors, anomaly detection in critical systems such as ITS (Intelligent Transportation Systems) using Data Science Advanced Data Learning (DS-ADL) models for resilience to adversarial attacks, protection against cyber-physical threats	Improving critical infrastructure security through robust models against perturbations, using federated learning for collaborative training without compromising privacy	ML models are vulnerable to uncontrolled data (input perturbations), difficulties in explainability of decisions, and risks of false positives in critical applications such as spam detection
Cloud security and access control	Detection of lateral movement and credential stuffing through access pattern analysis, application of homomorphic encryption for secure transactions in cloud environments	Scalability in multi-tenant environments without compromising performance, optimisation through differential privacy for data protection	Privacy issues (data de-anonymisation), high false positive rate due to sensitivity to perturbations, poisoning risks due to training data manipulation
Cyber risk forecasting	Risk scoring of assets, scenario modelling of attacks through pattern identification and data correlation to predict threats, and use of AI for proactive fortifications	Provides management with strategic planning tools; anticipates attacks to minimise damage; efficiently processes big data for accurate forecasts	Models require up-to-date data (lack thereof leads to inaccurate predictions), risk of incorrect predictions due to bias or incomplete data, and vulnerability to adversarial manipulation

Continued Table 1.

Direction	Example of use	Benefits	Risks
QA/DevOps integration	Security gates in CI/CD, ML for finding vulnerabilities in containers and code, automated testing with AI for code analysis, threat modelling, risk assessment, examples: predictive maintenance in automotive (Kafka/Spark/TensorFlow), diagnostics in healthcare (NiFi/PyTorch), fraud detection in finance (Kafka/XGBoost)	Automation of security in pipelines (40% reduction in training time, 25% reduction in downtime), improved code quality through early detection; improved collaboration between teams; diagnostic accuracy up to 85%, 15% reduction in false positives	Computational costs (high requirements for real-time analysis), lack of mature standards for integration, data quality issues (noisy data requires cleaning), and model drift risks requiring monitoring
Stress testing ML pipelines	Adversarial data corruption in anomaly detection pipelines: Savage framework for modelling realistic data quality problems through dependency graphs, examples of Missing Not At Random (MNAR) in finance, compound errors in policing (label flips by demographics)	Identification of weak spots in pipelines through two-level optimisation (beam search + Bayesian optimisation), identification of vulnerable subpopulations, and provision of benchmarks for robustness methods	Even a small percentage of damaged data (5-10%) sharply reduces accuracy (AUC drops by 0.05+ at 10% corruption), structured errors destroy feature-label relationships more strongly than random ones; the sensitivity of existing cleaning methods
Adversarial testing for CV/NLP models	Deep Neural Network (DNN): selecting the most/least similar labels to ground truth based on network/semantic similarity for targeted/non-targeted attacks; testing CNN on human-like errors	Detection of real vulnerabilities through a human-centric approach, increased resilience with better results compared to existing strategies, and a contribution to explainable testing and standardisation	Tests are often expensive and computationally intensive (similarity assessment and scenario testing), and depend on similarity measures, which can complicate implementation

Notes: literature sources were systematised using the Zotero tool, which ensured the standardisation of references and the formation of a correct bibliographic apparatus

Source: compiled by the author based on P. Bountakas *et al.* (2023), M. Musser *et al.* (2023), M. Steidl *et al.* (2023), O. El Gharbaoui *et al.* (2024), B. Gajbhiye *et al.* (2024), M.M. Saeed & M. Alsharidah (2024), K. Filus & J. Domańska (2024), S.P. Velaga (2024), F.V. Jedrzejewski *et al.* (2024), M. Roshanaei *et al.* (2024), M. Macas *et al.* (2024), J. Zhu *et al.* (2025)

Operational methods (NDR/NIDS, UEBA) provide high detection rates and enable real-time response to complex attacks, but remain sensitive to the quality of input data and prone to data drift, which reduces their long-term reliability. Automation tools (EDR/SOAR) reduce mean time to detect and mean time to respond (MTTD/MTTR), effectively relieve the burden on SOC analysts, and increase system scalability, but have significant limitations in the form of possible false positives and vulnerability to adversarial manipulations that can disable the system. Preventive approaches (QA/DevOps integration, pipeline stress testing) increase model resilience by detecting hidden vulnerabilities at the development stage and ensuring the long-term reliability of ML processes, but require significant computing resources, mature integration standards, and constant monitoring. A summary of the comparison results showed that none of the approaches is self-sufficient: operational methods ensure speed, automation ensures process efficiency, and preventive strategies ensure long-term reliability. Therefore, the most effective approach is a combination of the two, which can mutually compensate for shortcomings, optimise the balance between accuracy, stability and performance, and form the basis for integrating ML into modern security frameworks. The application of ML in various domains of cybersecurity combines significant advantages (increased accuracy, response speed, and scalability) with new risks related to data quality, vulnerability to adversarial attacks, and ethical constraints. Effective ML integration requires not only technological solutions, but also the development of adaptive security frameworks capable of addressing specific threats to models and data.

AML is a field of research that studies the vulnerabilities of AI models and develops mechanisms to protect them. In contrast to classical approaches, AML focuses on data, algorithms, and parameters, and a key methodological aspect is the definition of a threat model that describes the attacker's level of knowledge (Vassilev *et al.*, 2025). There are three scenarios: white-box – full access to architecture and parameters, which makes attacks as accurate as possible, but improbable in open environments; black-box – interaction only through requests and responses, which leads to the prevalence of transfer- and query-based attacks; and grey-box – partial knowledge of the model, most characteristic of industrial cases (Jedrzejewski *et al.*, 2024).

On this basis, attack classes were systematised. Evasion attacks target the inference phase: minimal perturbations in the input data cause the model to make mistakes, using both targeted and untargeted

approaches (FGSM, BIM/PGD, C&W, DeepFool). Poisoning attacks affect the data collection or training stages: integrity and availability variants reduce accuracy, while clean-label and backdoor introduce hidden "triggers" that are dangerous because they are invisible. Privacy and inference attacks are aimed at information leakage: membership and property inference or data recovery from gradients, which poses a particular threat to medicine and finance. In MLaaS environments, extraction attacks are common, where the model is reproduced through mass queries or even hardware channels (EM radiation, Rowhammer). A new subclass consists of attacks on generative models (GenAI): prompt injection, jailbreak, fine-tuning misuse, and prompt extraction, which have already received their own taxonomies in international standards. The spectrum of attacks within AML ranges from modifying input data to leaking private information or completely reproducing models. Their classification is not only a tool for scientific analysis, but also a practical prerequisite for developing defence strategies and integrating adversarial tests into QA/DevSecOps (Musser *et al.*, 2023).

At the data level (data validation, filtering, anomaly detection, synthetic augmentation, differential privacy), measures are highly effective against gross poisoning and noise and improve overall quality, but are limited in countering targeted stealth poisoning and require additional resources for collection and auditing. At the training level (adversarial training, robust optimisation, regularisation, DP-SGD), the best resistance to evasion attacks in targeted scenarios is achieved, but these methods require significant computational resources and can reduce the baseline accuracy of models. Post-training approaches (input transformations, feature squeezing, attack detectors, calibration/thresholding, ensembles) are notable for their ease of integration into CI/CD and their ability to improve security without retraining, but they remain vulnerable to adaptive attacks and can affect throughput. At the deployment and inference stage (rate-limiting, abuse control, request rejection or reclassification, canary models, shadow deployment), these methods can significantly reduce MTTR and business risks, but require mature monitoring and accompanying playbooks. Finally, process and management approaches (threat modelling, red teaming, adversarial testing, QA/DevSecOps policies, drift monitoring) provide the foundation for long-term manageability and compliance with standards, but require high organisational maturity and relevant competencies (Jedrzejewski *et al.*, 2024; Vassilev *et al.*, 2025). The most effective approach is to integrate various protection strategies within QA/MLOps with clear SLOs/SLIs, continuous drift monitoring, and regular adversarial testing, which optimises the balance between accuracy, resilience, and performance and forms the basis for modern security frameworks. Figure 3 shows a visualised structure of cybersecurity measures.

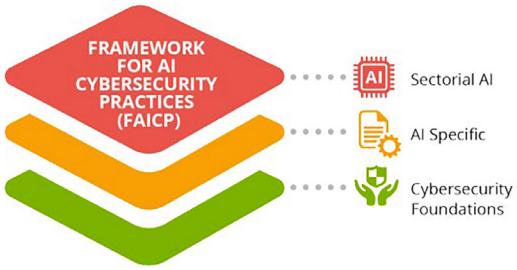

Figure 3. Structure of multi-level security measures for AI systems

Source: compiled by the author based on N. Polemi & I. Praça (2023)

Figure 3 shows the three-level structure of security practices: Cybersecurity Foundations – basic principles of cybersecurity common to all systems, AI Specific – measures that incorporate the lifecycle and vulnerabilities of AI models, Sectorial AI – domain-specific controls for specific industries (finance, healthcare, manufacturing). The model demonstrates a gradual transition from general to specialised levels of AML protection and summarises the approach to integrating AML controls into a multi-level cybersecurity system, combining universal standards with AI-specific practices.

The study results showed that ML increases the speed of detecting zero-day and APT attacks, automates responses through SOC/SOAR, and reduces the workload on analysts, but there remains a dependence on data quality and the risk of false positives. This was consistent with the findings of R. Liu *et al.* (2024), noting that ML algorithms in NDR/NIDS enable high anomaly detection accuracy and reduced response times. The study emphasised that ML increases SOC proactivity and ensures faster localisation of complex threats, including zero-day attacks. At the same time, the authors highlighted the problems of input data quality and the risks of false positives, which were also identified in the current study. Therefore, ML has a dual effect on network security: significant advantages in scalability and speed, along with challenges related to data reliability.

Y.L. Khaleel *et al.* (2024) conducted a systematic review of defensive methods in countering adversarial attacks, emphasising the role of MLand deep learning in creating multi-level defence strategies. The study showed that existing approaches, from adversarial training to ensemble models, significantly increase system resilience but remain vulnerable to new classes of attacks. This was consistent with the presented research showing that even when robust optimisation or randomised smoothing is applied, models retain limitations in terms of scalability and adaptability. This confirmed the need not only for technical innovation, but also for a comprehensive approach that addresses both algorithmic and organisational aspects of cyber defence.

The results of the study revealed that ML models are vulnerable to even minor adversarial perturbations, which reduces their robustness. X. Jiang *et al.* (2022)

proposed a local graph consensus method for robust image matching, which proved highly effective in reducing the impact of such perturbations. The authors' approach confirmed that structural data alignment methods can compensate for the weaknesses of classical models, maintaining the stability of results even in challenging conditions. This was consistent with the presented research on the need to implement algorithmic mechanisms to improve robustness in the CV/NLP domain. Both works demonstrated that combining theoretical and applied solutions is the key to creating robust ML systems.

A. Yaseen (2023) outlined that the application of ML in network security radically changes the response paradigm: from traditional reactive approaches that rely on signature-based methods to proactive anomaly detection strategies. ML models are capable of analysing both behavioural and network patterns at a much higher speed, enabling more effective detection of zero-day and complex multi-vector attacks. At the same time, the author emphasised the critical role of continuous model training, which can be used to ensure adaptability of the dynamic conditions of cyber threats. However, this creates additional risks, in particular, data drift and an increase in the number of false positives, which can undermine trust in the systems. The results of the study confirmed these conclusions: data quality and the availability of validation mechanisms are key to reducing noise and improving the effectiveness of ML solutions. This demonstrated a convergence in the view of ML as a capable but vulnerable technology that requires adaptive risk management to ensure long-term effectiveness in network security.

B.B. Sarhan & N. Altwaijry (2023) showed that ML-methods can effectively form user behaviour profiles and detect potentially harmful actions at an early stage. The study emphasised that this approach outperforms traditional signature-based methods due to its ability to adapt to changing behaviour patterns. They also highlighted the value of different algorithms, including SVM, classification trees, and ensemble methods, which increase the reliability of the system. This was consistent with current research that integrating ML into UEBA significantly improves the effectiveness of insider threat detection and contributes to proactive cyber defence, while ensuring data quality and minimising false positives remains a key challenge.

The application of ML in cybersecurity demonstrates high efficiency in detecting anomalies, automating responses, and forming behavioural profiles, which significantly increases accuracy compared to traditional methods. At the same time, models remain vulnerable to adversarial and poisoning attacks, data drift, and data quality issues that can reduce their reliability. This highlights the need for a comprehensive approach that combines algorithmic solutions, adaptive learning mechanisms, and organisational strategies to ensure robust cyber protection.

A comprehensive approach to AML in the context of standards, policies and practical cases

AML is not exclusively a technical problem; it has significant regulatory, ethical, and legal dimensions. If attacks on models undermine trust in AI systems, their uncontrolled exploitation may violate the principles of security, privacy, and fairness. A key document is the National Institute of Standards and Technology (2023), which proposes a systematic approach through the "2Map-Measure-Manage-Govern" functions. It defines a taxonomy of attacks and mitigations, emphasising the criticality of forming AML protection policies at the organisational process level. Additionally, ISO/IEC 23894:2023 (2023) remains the basic standard in the field of information security. Although it is not specific to AI, its requirements for risk management, incident management, and access control also apply to the life cycle of ML models. This can combine the classic approach to cybersecurity with new AML risks. The European Union Agency for Cybersecurity (ENISA) proposes a multi-layered model, ranging from basic cyber hygiene measures to AI-specific practices (Polemi & Praça, 2023). It emphasises the need to integrate AML tests at all levels, from data and models to organisational policies.

Large companies are developing their own frameworks for AI security. Microsoft Security (n.d.) outlines principles for protecting generative models, including mechanisms for countering prompt injection and approaches to model lifecycle management integrated into corporate security processes. OpenAI (2024) demonstrates an example of transparent risk documentation, where the model's resistance to adversarial attacks and jailbreaks is systematically assessed, which meets current requirements for openness and accountability in the field of AI. In addition to technical and regulatory issues, AML also raises ethical issues such as privacy (leaks of training data through inference attacks violate the principles of personal information protection), fairness (adversarial training can increase or decrease the fairness of a model depending on how it affects different user groups), and transparency and trust (society demands explainability of how AI makes decisions and how organisations manage the risks of adversarial attacks). The regulatory and ethical aspects of AML complement technical solutions, forming a second dimension of model security. For QA and MLOps, this means verifying not only technical metrics but also compliance with standards and corporate policies, which are the basis for trust in AI systems.

The issue of protection against adversarial attacks is no less complex than their classification. Existing approaches to ML security suggest considering defensive measures in the context of the model lifecycle. This creates convenient conditions for integration into QA/MLOps processes, as it can assess risks and embed resilience testing at all stages, from data preparation to operation (Musser *et al.*, 2023). At the initial stage, the main task is to protect data, as it forms the basis

for the model. Methods of data selection and cleaning, deduplication, filtering of out-of-distribution (OOD) or anomalous samples, as well as integrity and origin control (data lineage) are used. Such measures reduce the probability of poisoning attacks and increase trust in data sources (Bountakas *et al.*, 2023). At the training stage, the key focus is on adversarial training and robust optimisation, which increase the resilience of models to evasion attacks. Regularisation methods (e.g., Tradeoff In Adversarial Defence via Smoothing (TRADES)) are also used to balance accuracy and robustness. To reduce the risk of membership inference and data reconstruction, differentially private algorithms (Differentially Private Stochastic Gradient Descent (DP-SGD), Differentially Private Follow-the-Regularised-Leader (DP-FTRL)) are integrated. However, such methods do not protect against model extraction, which limits their applicability.

In the post-training stage, randomised smoothing methods are used to formally certify the level of stability, as well as formal verification approaches. However, trade-offs between accuracy and stability, as well as possible negative effects on model fairness, should be considered, which requires careful evaluation during QA (National Institute of Standards and Technology, 2023). During model deployment, the main task is to protect against attacks on interfaces and services. Request restrictions and audits, request rate limiting, and detection of suspicious access patterns are used. Such measures increase the level of protection against model extraction, but these mechanisms can be circumvented, so constant monitoring and retesting are required (Bountakas *et al.*, 2023). At the inference stage, data preprocessing methods (quantum compression, denoising) are used, as well as special detectors for adversarial or OOD cases. However, these approaches are vulnerable to adaptive attacks, and the use of gradient masking is dangerous, as it only creates the illusion of protection but does not eliminate the problem. Effective protection against AML attacks requires a multi-level approach that covers all stages of the model lifecycle, incorporating the inevitable trade-offs between accuracy and robustness, privacy and performance, formality and practical relevance. For QA and MLOps, this means implementing dynamic testing and retesting processes that can adapt to the constant evolution of attacks.

Practical cases of AI/ML and AML applications cover network security, SOC, DevOps/CI-CD, and industrial domains. Their systematisation has made it possible to outline the opportunities and risks of integrating these technologies into real environments (Table 2).

Table 2. Practical scenarios for integrating AI/ML and AML into security systems

Sphere of use	Methods and examples	Benefits and risks
Network security, SOC	Detection of botnets in the bank's corporate network through real-time traffic analysis. ML models for anomaly detection (DGA, suspicious IPs), integration with SIEM (Splunk, QRadar) and IDS (Snort) for automatic responses such as IP blocking	Rapid detection of attacks (30% reduction in response time), SOC automation. Risks: requirement for standardisation, vulnerability to data poisoning
AML-attack protection	Protecting telecom companies' intrusion detection systems from adversarial attacks. Adversarial training, feature squeezing to reduce data manipulation, and ensemble models for robustness	Protection of models in cybersecurity, computer vision, and NLP, with the possibility of validation through datasets. Risks: methods are not yet ready, and the complexity of countering new attacks
DevOps/CI-CD	Automated security testing: searching for vulnerabilities in containers (continuous monitoring of code bases to detect code injections, unauthorised access, data leaks); code analysis using ML to recognise risk patterns, automate threat modelling and risk assessment. Models adapt to evolving threats through training on new data	ML tests are integrated into QA as security gates: integrating AI/ML into CI/CD ensures a continuous security process throughout the development cycle, with prioritisation of vulnerabilities. This can be used for early detection of threats in dynamic DevOps environments, but requires an emphasis on data quality and model adaptability
Industry (finance, telecoms)	Protection of the anti-fraud system of the bank from data poisoning. ML for detecting fraudulent transactions, protection against adversarial attacks on IDS through training on noisy data	The lack of mature robustness metrics hinders AML integration and has low practical relevance. The absence of standardised metrics complicates the implementation of AML in production, particularly for poisoned data in sensitive industries
Pipeline stress-testing	Testing the ML pipeline at a financial company to identify vulnerabilities to data corruption. Savage framework for generating scenarios with missing data (MNAR in finance) and label errors (predictive policing), optimisation via beam search	Even minor data damage drastically reduces accuracy: Structured errors (5-10%) destroy feature-label relationships more severely than random ones, and existing cleaning/debugging methods are sensitive. Stress testing is needed to validate the resilience of pipelines to systematic corruption

Continued Table 1.

Sphere of use	Methods and examples	Benefits and risks
DNN testing	Similarity-driven adversarial testing for CV systems: Using the most/least similar labels to ground truth (based on network/semantic similarity) for targeted/non-targeted attacks on CNN. Testing the object recognition system in airport security cameras	Identification of weaknesses through "human-like" attacks: The human-centric approach increases the effectiveness and transparency of testing by identifying vulnerabilities in visual perception. Promotes the standardisation of AI security, with a focus on explainable testing for practical requirements. Risks: high computational complexity, dependence on metrics
CI/CD, MLOps	Implementation of ML in company pipelines: predictive maintenance for an automaker (Tesla), diagnosis of heart disease (healthcare startup), and fraud detection in a bank. Methods: Kafka/Spark/TensorFlow for automotive, NiFi/PyTorch for healthcare, Kafka/XGBoost for finance; 4 stages of MLOps	QA in ML should include adversarial testing as standard: data quality is critical, drift monitoring and real-time optimisation. Automated validation/retraining provides the basis for adversarial QA, while cross-functional collaboration ensures resilience in CI/CD
Best practices MLOps	Implementation of MLOps in Google Cloud for a fintech service. Methods: Metrics (QPS, latency), security gates for skew testing, CD level 1-2, MTBF/MTTR monitoring	Adversarial QA is becoming a mandatory component of CI/CD: Testing for skew/anomalies in pipelines ensures stability, while monitoring triggers retraining. Balancing speed/security through incremental goals, adversarial tests are mandatory for production ML, with an emphasis on reproducibility and adaptation to evolving data. Risks: manual processes, training-serving skew

Notes: literature sources were systematised using the Zotero tool, which ensured the standardisation of references and the formation of a correct bibliographic apparatus

Source: compiled by the author based on P. Bountakas *et al.* (2023), M. Steidl *et al.* (2023), S.R. Gopireddy (2024), O. El Gharbaoui *et al.* (2024), B. Gajbhiye *et al.* (2024), F.V. Jedrzejewski *et al.* (2024), K. Filus & J. Domańska (2024), M. Macas *et al.* (2024), M. Roshanaei *et al.* (2024), S.P. Velaga (2024), J. Zhu *et al.* (2025), Adversarial robustness evaluation metric (2025), MLOps: Continuous delivery and automation pipelines in machine learning (n.d.), Amazon Web Services (n.d.)

The integration of AI/ML into various areas of cybersecurity has proven highly effective in reducing threat detection and response times, automating SOC processes, and improving detection accuracy. Most solutions face a lack of standardised robustness metrics and mature frameworks for assessing model resilience, which complicates the industrial implementation of AML approaches. Practical cases show that the most effective approach is a combination of preventive, analytical, and automated strategies, from adversarial training and feature squeezing to security gates in CI/CD and continuous monitoring of model drift. This approach forms a comprehensive Adversarial QA ecosystem, in which model robustness is systematically assessed at all stages of the life cycle, from training to deployment in production environments.

Given the growing complexity of ML systems and their vulnerability to adversarial attacks, integrating testing and protection mechanisms directly into QA/CI-CD pipelines is critical. This can ensure a shift from reactive to proactive strategies, where model robustness and privacy are systematically tested at all stages of the lifecycle. Automated security testing using AI/ML is already being actively implemented in DevOps environments, but so far, it has focused mainly on classic vulnerabilities in code and configurations. The integration of adversarial testing remains limited. Existing pipelines for ML models lack mature stability control mechanisms, creating a gap between scientific developments and industrial needs. DevOps pipelines need to include continuous testing modules capable of detecting not only performance regressions but also reductions in model stability. This ensures early detection of problems and rapid validation of changes in code and data.

The quality control system should incorporate not only classic accuracy indicators, but also metrics of resistance to attacks, in particular robust accuracy, attack success rate, and privacy loss. This can be used for informed decisions regarding the model's readiness for deployment. Robust accuracy (RA) reflects the proportion of examples that the model classifies correctly even after introducing permissible ε-level perturbations (1):

$$RA = \frac{1}{N}\sum_{i=1}^{N} I\left(f(x_i + \delta_i) = y_i\right), \|\delta_i\| \leq \varepsilon, \qquad (1)$$

where $(f(x_i + \delta_i)$ – model output for a disturbed input sample $x_i + \delta_i$, y_i – true label (real class) for the input sample x_i, δ_i – disturbance added to the input data, ε – maximum permissible disturbance level, I – indicator function equal to 1 if the expression in brackets is true (the model correctly classifies the disturbed sample), 0 – if false. RA measures the extent to which the model correctly classifies samples, even after adding small perturbations δ_i that do not exceed the permissible level ε. A high *RA* indicates that the model is resistant to attacks or noise, maintaining accuracy even when data changes,

which is essential for tasks where accuracy is critical (in image recognition systems or medical diagnostics).

Attack success rate (ASR) characterises the proportion of examples for which the attack resulted in a change in classification (2):

$$ASR = \frac{1}{N}\sum_{i=1}^{N} I\left(f(x_i) \neq f(x_i + \delta_i)\right). \quad (2)$$

An increase in *ASR* indicates greater vulnerability of the model to attacks. If *ASR* is high, it means that the model can be easily misled, which is critical for testing and ensuring reliability in real-world applications.

Privacy loss (ε) determines the level of differential privacy of the model and limits the influence of a single record in the training sample on the final training result (3):

$$\varepsilon = ln\left(\frac{P(M(D)=o}{P(M(D')=o}\right), \quad (3)$$

where $P(M(D) = o$ – the probability that the model M, trained on data D, will provide result o, $P(M(D') = o$ – the probability that the model M, trained on data D' (different from D only by one entry), will provide the same result o. Privacy loss measures how much changing a single record in the training sample affects the model's performance. A small ε indicates that the model has high privacy, as changing a single record has little effect. A high ε means that the model is less private, as a single record can significantly change the result.

Testing models for resilience to hacking should be integrated into QA processes on a par with traditional unit and integration tests. This approach ensures systematic assessment of vulnerabilities during development and operation. CI/CD processes should include security gates that check models for compliance with security policies prior to deployment, with subsequent logging for auditing and regulatory compliance. The integration of adversarial testing is accompanied by a number of challenges: high computational costs, lack of unified standards, a gap between academic approaches and engineering practice, and the need for cultural change in DevSecOps teams, where AML tests should be considered a mandatory stage of QA. Implementing AML protection in QA/CI-CD is a complex task that requires not only technical innovation but also organisational transformation. Adversarial testing should be considered a standard component of quality control, and QA/MLOps systems should be viewed as mechanisms not only for verifying functionality but also for managing model resilience and trust.

The study identified that implementing AML in industrial environments requires adaptive security frameworks capable of addressing specific threats to models and data, while balancing accuracy, robustness, and performance. This was consistent with the study by P. Kashefi *et al.* (2024), which examined the issue of global AI regulation and emphasised the need for "flexible regulation". The study argued that excessive control can stifle innovation and technological development, while the absence of rules creates significant risks for society and trust in AI systems. They propose a model that combines international standards with national contexts, while maintaining adaptability and practical relevance. The consensus was on the appropriateness of dynamic and context-dependent approaches, with both studies correlating with the idea that only a balance between control and innovation can ensure the sustainable integration of AI into real-world environments.

M. Malatji & A. Tolah (2025) presented a comprehensive framework for analysing AI cybersecurity measurements, in particular adversarial and offensive AI. The authors systematised the main classes of attacks, emphasising their impact on critical infrastructures and industrial systems. Traditional single-layer defence methods remain insufficient, and an effective strategy must have a multi-layered architecture. The study addressed the organisational aspects of integrating security at the level of processes and policies, rather than just technical solutions, which made it possible to form a more robust defence against dynamic and complex attacks. This was consistent with the findings of this study regarding the critical need for multi-layered strategies to counter adversarial attacks. In particular, the current findings emphasised the importance of integrating adversarial testing into QA/CI-CD, which echoed the authors' conclusions regarding the need for a comprehensive approach where technical and organisational levels cooperate.

The results of the study demonstrated that adversarial attacks create not only technical but also ethical problems, undermining trust and fairness in the use of AI. Even highly accurate models can become a source of social risks if aspects of fairness and explainability are not addressed. This was consistent with the study by T.W. Kim & A. Strudler (2023), demonstrating that the development of AI is impossible without incorporating social and ethical aspects, in particular, fairness and legitimacy of decision-making. The study argued that the issue of the rights and responsibilities of AI systems has a direct impact on public trust in technology. Technological progress without an ethical framework can lead to crises of trust and social risks. Both studies confirmed that without ethical criteria, AI security remains incomplete. The study by D.S. Fessenko & A. Jasperse (2025) emphasised that ethics, combined with technical and legal norms, should be at the heart of AI regulation. The regulatory framework should act as a bridge between scientific innovation and public expectations. The study suggested viewing regulation not as a barrier, but as a mechanism for harmonising transparency, accountability and innovation. The presented research results confirmed these conclusions, as AML was considered not only in a technical but also in a regulatory context. Therefore, the formation of AML policies should be based on a balance of technical and regulatory decisions.

L. Jiang & V. Torra (2023) addressed the problem of data protection in multi-source and distributed ML models. The study proved that privacy, access control, and data integrity are the foundation of trust in any AI system. They described the challenges associated with leaks and manipulations in multi-source environments and proposed using data lineage mechanisms for transparency of data origin. The key conclusion was that the reliability of models directly depends on the quality and security of the input data. This correlated with the results of the presented study, which also found that poisoning attacks and adversarial manipulation remain a critical challenge that can destroy even the most accurate models. The integration of data validation, cleaning, and tracking procedures is a prerequisite for the stable operation of ML systems in industrial environments.

Thus, ML significantly enhances cyber defence, accelerates detection, scales network/behavioural data analysis, and automates responses in SOC/SIEM/SOAR. At the same time, a new class of risks emerges: adversarial and poisoning attacks, data drift, and dependence on data quality/origin, which sharply reduces the robustness of models. Sustainable integration requires multi-level technical and organisational measures (adversarial training, robust metrics in QA/CI-CD, NIST/ISO/ENISA practices) with a balance between innovation and control.

CONCLUSIONS

The results of the study confirmed that integrating ML methods into cybersecurity is a strategically relevant step that combines improved threat detection and response with the emergence of new challenges. In particular, the study established that the use of ML reduces attack detection time by tens of per cent, enables automated response (EDR/SOAR), and improves the accuracy of network traffic and user behaviour pattern analysis. Reducing detection time from 48 to 3 hours or reducing the proportion of false positives from 20% to 5% confirms the real benefits of ML in protecting complex infrastructures. At the same time, limitations of this approach have also been identified. The main risks are vulnerability to adversarial and poisoning attacks, which, even with minimal interference with the data, can reduce the accuracy of models and create hidden "backdoors". The impact of such attacks is particularly critical in areas with high reliability requirements, such as medicine or financial services. In addition, issues of privacy, explainability, and bias remain problematic, directly affecting trust in AI systems.

It is advisable to systematise the areas of application of ML in cybersecurity, from network monitoring and anti-phishing to DevOps integrations and risk prediction. The classification of attacks within AML has shown that threats cover the entire model lifecycle, from data collection to exploitation, and therefore require multi-level countermeasures. It is recommended to implement adversarial testing as a mandatory stage of QA/MLOps, which will facilitate the transition from reactive to proactive cyber defence strategies. It is worth noting that the effectiveness of ML solutions directly depends on data quality, as even 5-10% of structured corruption in the sample can lead to a noticeable decrease in AUC and other metrics, confirming the critical importance of controlling data reliability and origin.

The limitation of this study is its predominantly theoretical nature, which necessitates further empirical verification and practical testing of the proposed approaches in real-world cyber defence systems. Further research should focus on developing resilience metrics, standardising AML practices, and creating adaptive security frameworks. In CI/CD pipelines, it is advisable to integrate adversarial testing as a security gate after the model validation stage. This can detect a decrease in robustness or an increase in the attack success rate before the model is implemented. In Quality Assurance processes, it is worth adding robust accuracy, privacy loss, and attack success rate metrics as mandatory criteria for model acceptance. This will ensure a systematic approach to assessing resilience before deployment. The integration of differential privacy, formal verification, and explainable AI will help balance accuracy, transparency, and resilience, ensuring the effectiveness and trustworthiness of ML systems in the changing landscape of cyber threats.

ACKNOWLEDGEMENTS

None.

FUNDING

None.

CONFLICT OF INTEREST

None.

REFERENCES

[1] Adversarial robustness evaluation metric. (2025). Retrieved from https://www.ibm.com/docs/en/watsonx/saas?topic=metrics-adversarial-robustness&utm_source=.

[2] Agarwal, A., & Nene, M.J. (2025). Advancing trustworthy AI: A comparative evaluation of AI robustness toolboxes. *SN Computer Science*, 6, article number 234. doi: 10.1007/s42979-025-03785-w.

[3] Amazon Web Services. (n.d.). *Metrics for CI/CD pipelines*. Retrieved from https://docs.aws.amazon.com/prescriptive-guidance/latest/strategy-cicd-litmus/metrics-for-cicd-pipelines.html?utm_source=.

[4] Bagai, R., Masrani, A., Ranjan, P., & Najana, M. (2024). Implementing continuous integration and deployment (CI/CD) for machine learning models on AWS. *International Journal of Global Innovations and Solutions*. doi: 10.21428/e90189c8.9cb39c55.

[5] Bena, N., Anisetti, M., Gianini, G., & Ardagna, C.A. (2024). Certifying accuracy, privacy, and robustness of ML-Based malware detection. *SN Computer Science*, 5(1), article number 710. doi: 10.1007/s42979-024-03024-8.
[6] Bountakas, P., Zarras, A., Lekidis, A., & Xenakis, C. (2023). Defense strategies for adversarial machine learning: A survey. *Computer Science Review*, 49, article number 100573. doi: 10.1016/j.cosrev.2023.100573.
[7] El Gharbaoui, O., Kiyadi, I., & El Boukhari, H. (2024). Evaluating AI and ML in network security: A comprehensive literature review. *Procedia Computer Science*, 251, 727-733. doi: 10.1016/j.procs.2024.11.176.
[8] Fessenko, D.S., & Jasperse, A. (2025). Ethics at the heart of AI regulation. *AI and Ethics*, 5, 3387-3398. doi: 10.1007/s43681-024-00562-y.
[9] Filus, K., & Domańska, J. (2024). Similarity-driven adversarial testing of neural networks. *Knowledge-Based Systems*, 305, article number 112621. doi: 10.1016/j.knosys.2024.112621.
[10] Gajbhiye, B., Aggarwal, A., & Jain, S. (2024). Automated security testing in DevOps environments using AI and ML. *International Journal for Research Publication and Seminar*, 15(2), 259-271. doi: 10.36676/jrps.v15.i2.1472.
[11] Gala, D.L., Molleda, J., & Usamentiaga, R. (2025). Evaluating the impact of adversarial patch attacks on YOLO models and the implications for edge AI security. *International Journal of Information Security*, 24, article number 154. doi: 10.1007/s10207-025-01067-3.
[12] Gopireddy, S.R. (2024). AI-powered code review and vulnerability detection in DevOps pipelines. *Journal of Scientific and Engineering Research*, 11(10), 124-129.
[13] Guo, J., Bao, W., Wang, J., Ma, Y., Gao, X., Xiao, G., Liu, A., Dong, J., Liu, X., & Wu, W. (2023). A comprehensive evaluation framework for deep model robustness. *Pattern Recognition*, 137, article number 109308. doi: 10.1016/j.patcog.2023.109308.
[14] ISO/IEC 23894:2023. (2023). *Information technology – artificial intelligence – guidance on risk management*. Retrieved from https://www.iso.org/standard/77304.html?utm_source=.
[15] Jedrzejewski, F.V., Thode, L., Fischbach, J., Gorschek, T., Mendez, D., & Lavesson, N. (2024). Adversarial machine learning in industry: A systematic literature review. *Computers & Security*, 145, article number 103988. doi: 10.1016/j.cose.2024.103988.
[16] Jiang, L., & Torra, V. (2023). Data protection and multi-database data-driven models. *Future Internet*, 15(3), article number 93. doi: 10.3390/fi15030093.
[17] Jiang, X., Xia, Y., Zhang, X.-P., & Ma, J. (2022). Robust image matching via local graph structure consensus. *Pattern Recognition*, 126, article number 108588. doi: 10.1016/j.patcog.2022.108588.
[18] Kashefi, P., Kashefi, Y., & Mirsaraei, A.G. (2024). Shaping the future of AI: Balancing innovation and ethics in global regulation. *Uniform Law Review*, 29(3), 524-548. doi: 10.1093/ulr/unae040.
[19] Khaleel, Y.L., Habeeb, M.A., Albahri, A.S., Al-Quraishi, T., Albahri, O.S., & Alamoodi, A.H. (2024). Network and cybersecurity applications of defense in adversarial attacks: A state-of-the-art using machine learning and deep learning methods. *Journal of Intelligent Systems*, 33(1), article number 20240153. doi: 10.1515/jisys-2024-0153.
[20] Kim, T.W., & Strudler, A. (2023). Should robots have rights or rites? *Communications of the ACM*, 66(6), 78-85. doi: 10.1145/3571721.
[21] Kotenko, D., & Khlaponin, Y. (2024). Artificial intelligence in cyber attack detection and prevention systems: Prospects and challenges. *Underwater Technologies*, 1(14), 48-55. doi: 10.32347/uwt.2024.14.1203.
[22] Liu, R., Shi, J., Chen, X., & Lu, C. (2024). Network anomaly detection and security defense technology based on machine learning: A review. *Computers and Electrical Engineering*, 119(A), article number 109581. doi: 10.1016/j.compeleceng.2024.109581.
[23] Macas, M., Wu, C., & Fuertes, W. (2024). Adversarial examples: A survey of attacks and defenses in deep learning-enabled cybersecurity systems. *Expert Systems with Applications*, 238(E), article number 122223. doi: 10.1016/j.eswa.2023.122223.
[24] Malatji, M., & Tolah, A. (2025). Artificial intelligence (AI) cybersecurity dimensions: A comprehensive framework for understanding adversarial and offensive AI. *AI and Ethics*, 5, 883-910. doi: 10.1007/s43681-024-00427-4.
[25] Microsoft Security. (n.d.). *How Microsoft secures generative AI*. Retrieved from https://www.microsoft.com/content/dam/microsoft/final/en-us/microsoft-brand/documents/EN-US_BDM_Microsoft_Security_Whitepaper_v3.pdf.
[26] MLOps: Continuous delivery and automation pipelines in machine learning. (n.d.). Retrieved from https://cloud.google.com/architecture/mlops-continuous-delivery-and-automation-pipelines-in-machine-learning?utm_source=.
[27] Musser, M., *et al.* (2023). *Adversarial machine learning and cybersecurity: Risks, challenges, and legal implications*. Georgetown: CSET. doi: 10.51593/2022CA003.
[28] National Institute of Standards and Technology. (2023). *Artificial Intelligence: A taxonomy of AI risk and a framework for assessing AI risk (NIST AI 100-1)*. Gaithersburg: National Institute of Standards and Technology. doi: 10.6028/NIST.AI.100-1.

[29] Neretin, O., & Kharchenko, V. (2022). Ensurance of artificial intelligence systems cyber security: Analysis of vulnerabilities, attacks and countermeasures. *Academic Journals and Conferences*, 12, 7-22. doi: 10.23939/sisn2022.12.007.
[30] OpenAI. (2024). *GPT-4o system card*. Retrieved from https://cdn.openai.com/gpt-4o-system-card.pdf?utm_source=.
[31] Polemi, N., & Praça, I. (2023). *Multilayer framework for good cybersecurity practices for AI*. Attiki: European Union Agency for Cybersecurity.
[32] Roshanaei, M., Khan, M.R., & Sylvester, N.N. (2024). Enhancing cybersecurity through AI and ML: Strategies, challenges, and future directions. *Journal of Information Security*, 15(3), 320-339. doi: 10.4236/jis.2024.153019.
[33] Saeed, M.M., & Alsharidah, M. (2024). Security, privacy, and robustness for trustworthy AI systems: A review. *Computers and Electrical Engineering*, 119(B), article number 109643. doi: 10.1016/j.compeleceng.2024.109643.
[34] Sarhan, B.B., & Altwaijry, N. (2023). Insider threat detection using machine learning approach. *Applied Sciences*, 13(1), article number 259. doi: 10.3390/app13010259.
[35] Steidl, M., Felderer, M., & Ramler, R. (2023). The pipeline for the continuous development of artificial intelligence models – current state of research and practice. *Journal of Systems and Software*, 199, article number 111615. doi: 10.1016/j.jss.2023.111615.
[36] Vadde, B.C., & Munagandla, V.B. (2023). Integrating AI-driven continuous testing in DevOps for enhanced software quality. *Journal of Artificial Intelligence in Medicine*, 14(1), 505-515.
[37] Vassilev, A., Oprea, A., Fordyce, A., Anderson, H., Davies, X., & Hamin, M. (2025). *Adversarial machine learning: A taxonomy and terminology of attacks and mitigations*. Gaithersburg: National Institute of Standards and Technology. doi: 10.6028/NIST.AI.100-2e2025.
[38] Velaga, S.P. (2024). Case studies of successful CI/CD pipeline implementations for Machine Learning and AI. *International Journal of Research and Analytical Reviews*, 11(3), 137-145.
[39] Yaseen, A. (2023). The role of machine learning in network anomaly detection for cybersecurity. *Sage Science Review of Applied Machine Learning*, 6(8), 16-34.
[40] Zhu, J., Xu, G., Lorenzi, F., Glavic, B., & Salimi, B. (2025). Stress-testing ML pipelines with adversarial data corruption. *Proceedings of the VLDB Endowment*, 18(11), 4668-4681. doi: 10.14778/3749646.3749721.

Інтеграція тестування безпеки в конвеєри контролю якості за допомогою змагального машинного навчання

Анатолій Гусаковський
Магістр, старший інженер з контролю якості автоматизації
Національний аерокосмічний університет «Харківський авіаційний інститут»
61070, вул. Вадима Манька, 17, м. Харків, Україна
https://orcid.org/0009-0007-9398-0966

Анотація. Метою дослідження була теоретична систематизація підходів до підвищення стійкості машинного навчання систем у кіберзахисті через інтеграцію тестування на стійкість у процесі забезпечення захисту. Методологія охоплювала систематизацію машинного навчання – напрямів у кіберзахисті, аналіз стратегій протидії ворожим атакам та case study інтеграції в забезпечення якості та операції з машинного навчання. Встановлено, що використання технологій машинного навчання у кіберзахисті забезпечує автоматизацію виявлення та реагування на загрози (аномалії у мережах, поведінковий аналіз, антифішинг, антифрод, класифікація malware). Основними перевагами є масштабованість, швидкість реагування, предиктивність та ефективність у складних середовищах, тоді як ключові ризики охоплюють залежність від якості даних, хибні спрацювання, вразливість до adversarial- та poisoning-атак, а також проблеми приватності й пояснюваності. Виявлено, що у змагальному машинному навчання виділено три сценарії атак (white-box, black-box, gray-box) та їх класи, отруєння даних, конфіденційність/висновок, вилучення моделі, генеративний штучний інтелект. Підкреслено, що змагальне машинне навчання охоплює не лише технічний, а й нормативний та етичний виміри, пов'язані з принципами конфіденційності, справедливості та прозорості у використанні штучного інтелекту. Представлено багаторівневі стратегії захисту, інтегровані в життєвий цикл моделей машинного навчання – на рівні даних, під час навчання, після навчання, на етапах розгортання та інференції. Практичні кейси продемонстрували доцільність застосування машинного навчання та протидії відмиванню коштів у різних доменах – від мережевої безпеки та центру операцій безпеки до розроблення та експлуатації / безперервної інтеграції / безперервного постачання, фінансового сектору, стрес-тестування конвеєрів машинного навчання, а також забезпечення якості й експлуатації машинного навчання. Практична значущість полягає в можливості використання отриманих результатів фахівцями з кібербезпеки, фінансовими аналітиками та інженерами з експлуатації машинного навчання для підвищення ефективності центру операцій безпеки, інтеграції змагального тестування та забезпечення стійкості моделей машинного навчання у виробничих середовищах

Ключові слова: стійкість моделей; вразливості алгоритмів; метрики стійкості; автоматизоване тестування; життєвий цикл моделей; захисні механізми

BULLETIN of Cherkasy State Technological University

Journal homepage: https://bulletin-chstu.com.ua/en

Vol. 31 No. 1. 2026

UDC 004.89:519.816
DOI: 10.62660/bcstu/1.2026.143

ISSN: 2306-4412
E-ISSN: 2708-6070

Models of information situations of decision-making under conditions of risk, uncertainty, and fuzzy information

Anton Maksymov

PhD in Computer Science
Cherkasy State Technological University
18006, 460 Shevchenka Blvd., Cherkasy, Ukraine
https://orcid.org/0009-0002-0411-8164

Yurii Tryus*

Doctor of Pedagogical Sciences, PhD in Physical and Mathematical Sciences, Professor
Cherkasy State Technological University
18006, 460 Shevchenka Blvd., Cherkasy, Ukraine
https://orcid.org/0000-0002-0739-2065

Abstract. The relevance is conditioned by the need to create mathematical support for a conceptual approach that would allow systematising information situations of decision-making under conditions of risk, uncertainty, fuzzy information, and provide a reasonable choice of methods for solving practical decision-making problems in these situations. The goal was to develop mathematical models of information situations of decision-making under conditions of risk, uncertainty, fuzzy information, and their possible combinations, and to build a model of the decision-making process based on them, which provides for the possibility of applying several methods of choosing alternatives and aggregating their results to improve the efficiency of decision-making. The research methodology was based on a systematic approach using set theory, probability theory, fuzzy set theory, decision theory and methods, and mathematical modelling to formalise information situations of decision-making. Mathematical models of information situations of decision-making under conditions of risk (Risk, R), complete uncertainty (Uncertainty, U), fuzzy information (Fuzzy, F), and possible combinations of situations were developed: R-U, R-F, U-F, R-U-F. The generalised model of information situation of decision-making R-U-F (MISDMRUF) was interpreted as a model of a decision-making problem that formalises the process of choosing an alternative in conditions of simultaneous presence of risk, uncertainty, and fuzzy information. This allowed using existing and creating new decision-making methods depending on the characteristics of the information situation of decision-making. A model of the decision-making process under conditions of risk, uncertainty and fuzzy information (MDMPRUF) was proposed, its tasks, main stages of implementation, and key features, in particular, versatility, flexibility, and integrativity, were defined. The results showed that the proposed MISDMRUF and MDMPRUF models were consistent with a number of studies in decision theory, but simultaneously fill in the existing gaps, providing a systematic, integrated approach to classifying information situations of decision-making and selecting decision methods under complex conditions of risk, uncertainty, and fuzzy information. The proposed mathematical models and an appropriate approach to solving decision-making problems can become the basis for creating information technologies for decision support under complex conditions

Keywords: mathematical models; decision-support models; decision-making methods; decision-support systems; information technologies

Article's History: Received: 28.10.2025; Revised: 05.02.2026; Accepted: 16.03.2026; Published: 08.04.2026

Suggested Citation:
Maksymov, A., & Tryus, Yu. (2026). Models of information situations of decision-making under conditions of risk, uncertainty, and fuzzy information. *Bulletin of Cherkasy State Technological University*, 31(1), 143-153. doi: 10.62660/bcstu/1.2026.143.

*Corresponding author

INTRODUCTION

Contemporary decision-making tasks are characterised by a high level of complexity due to the presence of multiple criteria, the impact of various risks, uncertainty in decision-making conditions, and the inclusion of fuzzy input data in real problems. In the decision-making process, management bodies are faced with various information situations that can significantly affect the effectiveness of decision-making. The lack of a single conceptual model that combines these situations and considers their interrelationships makes it difficult to choose and apply decision-making methods. This underscores the relevance of the research aimed at creating a model that will systematise information situations of decision-making under conditions of risk, uncertainty, fuzzy information, and provide a reasonable choice of methods for solving practical decision-making problems in these situations.

In scientific research, decision-making problems are considered through the prism of various information situations that arise in the process of choosing alternatives. Various approaches to the classification and formalisation of decision-making models and methods are widely presented in the scientific literature. The monograph by M.I. Gomes & N.C. Martins (2022) reviewed mathematical models of decision-making, which showed the importance of integrating different opinions and criteria into a single model. Special attention was paid to the integration of multi-criteria decision-making and multi-criteria optimisation as complementary approaches to the analysis of complex information situations. This creates a theoretical basis for reasonable selection and combination of methods depending on the level of information certainty. J. Jaccard & J. Jacoby (2020) focused on the construction of theories and models in the social sciences, which is of direct importance for the development and modification of decision-making methods focused on complex and poorly structured problems. The researchers emphasised the importance of using qualitative, mixed and quantitative approaches to build new models, and the need to adapt and revise models in the event of conflicting empirical data, which is relevant for decision-making tasks in conditions of uncertainty and fuzzy information.

T. Zheldak *et al.* (2020) investigated the use of fuzzy sets in management and decision-making systems. The researchers emphasised the importance of integrating classical mathematical methods with contemporary information technology and data mining. They noted the role of combining quantitative and qualitative criteria and applying fuzzy sets, which helps to more effectively model information situations and make decisions in complex and changing conditions. N. Munier (2021) proposed the SIMUS method for modelling complex decision-making scenarios, which is an example of a modern hybrid approach to solving multi-criteria problems. Special attention was paid to the development of a solution matrix that maximally reflects the real conditions of the problem, considering practical limitations and environmental characteristics. The researcher emphasised that the choice of decision-making method should be carried out in accordance with the structure of this matrix, and the use of the SIMUS method using examples from real scenarios demonstrates its effectiveness for analysing complex, multi-factor situations. N.P. Høj *et al.* (2025) analysed methods for modelling the risk of systems and decision-making, focusing on the typical errors and limitations of existing decision-making approaches, and on the need for careful selection of methods and reasonable consideration of information uncertainty to ensure the rationality of decisions. The researchers stressed that incorrect consideration of uncertainty, interpretation of probabilities and consequences can lead to irrational decisions, sometimes worse than random ones, which is especially critical for events with low probability. This justifies the need for a systematic approach to risk modelling, which involves correct quantitative assessment, uncertainty analysis, and alignment of model assumptions with real-world decision-making conditions. In this context, T. Li *et al.* (2025) proposed an interval fuzzy multi-criteria decision-making model with an improved projection measure that improves the stability and accuracy of results in complex multi-criteria problems, demonstrating greater reliability compared to traditional approaches.

These sources have formed a methodological basis for further development of decision-making theory, but they do not provide a holistic view of the relationship between different types of information decision-making situations. The analysed studies lack a systematic classification of methods that accounts for the simultaneous presence of risk factors, uncertainty, and fuzzy information. Despite a significant amount of research in the field of decision-making, the problem of building models that would integrate complex information situations, in particular decision-making in conditions of risk, uncertainty and fuzzy information, and consider the relationships between them, remains unresolved. According to this problem, the purpose of the study was to develop mathematical models of information situations of decision-making in conditions of risk, uncertainty, fuzzy information, and their possible combinations, and to build a model of the decision-making process that provides integration of various methods of choosing alternatives and aggregation of their results.

MATERIALS AND METHODS

The research methodology was based on a systematic approach, which allowed comprehensively analysing information situations in decision-making and considering the relationships between all key elements of the process. For the formalisation and modelling of information situations, the basic concepts of set theory, probability theory, fuzzy set theory, and decision theory,

mathematical modelling were used, which provided a strict description of the decision-making process and the ability to quantify the influence of external and internal factors on the results. The models were developed considering the specifics of decision-making tasks in conditions of risk, uncertainty, and fuzzy information, which is typical of real management and information processes.

The study examined the problem of decision-making in an environment where external factors that affect the effectiveness of decisions are not under the control of the decision maker (DM) or the subject of management (SM). It was assumed that these factors and external disturbances were unchanged during the time when the decision is made, that is, the model was static for a certain period of time. In the study, the decision-making situation was defined as a trio of sets (Us & Koriashkina, 2018):

$$DMS = \{A, \Theta, F\}, \quad (1)$$

where A – set of mutually exclusive possible solutions (alternatives) of the management body:

$$A = \{A_1, A_2, \ldots, A_N\} \quad ,(2)$$

where Θ – set of mutually exclusive possible environmental states E where it may be located, but at the time of making the DM decision, it is not known exactly what state it is in (or will be in):

$$\Theta = \{\theta_1, \theta_2, \ldots, \theta_S\}, \quad (3)$$

where F – evaluation functionality that characterises the "utility" or "loss" of the DM when choosing a solution A_i, if the environment will be (or is) in the state θ_j:

$$F = \{f_{ij}\} = \{f(A_i, \theta_j)\}, i = \overline{1,N}, j = \overline{1,S}. \quad (4)$$

The information situation of decision-making in the study was defined as the degree of gradation of uncertainty in the choice of environment E of their states from a given set Θ at the time of decision-making by the management body. In this study, decision-making tasks were analysed in information situation of decision-making under risk conditions (ISDMR), uncertainty (ISDMU), fuzzy information (ISDMF), and their possible combinations. The choice of these three information situations was determined by the desire to cover different levels of availability and accuracy of information, which provided a comprehensive approach to the analysis of decision-making processes within the study.

The study used the conceptual model proposed by A.Y. Maksymov (2025), which reflected the relationship between information situations ISDMR, ISDMU, ISDMF that may arise during decision-making, i.e., the relationship between information situations of decision-making under conditions of risk (Risk, R), complete uncertainty (Uncertainty, U), and fuzzy information (Fuzzy, F). This model is called the "R-U-F" model of possible information decision-making situations. Its representation in the form of a Venn diagram is shown in Figure 1, helped to clearly present possible information situations of decision-making and facilitate identification of the nature of the decision-making problem and substantiation for choosing appropriate methods for solving it.

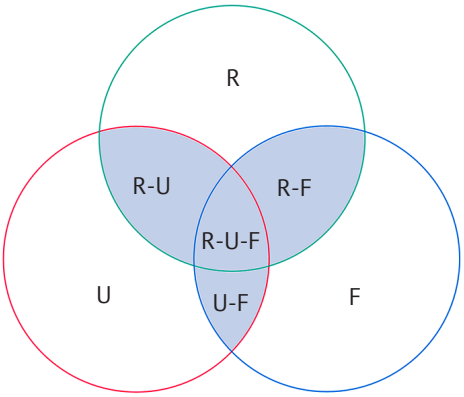

Figure 1. Set-theoretic "R-U-F" model of possible information decision-making situations under conditions of risk, uncertainty, and fuzzy information
Source: created by the authors based on A.Y. Maksymov (2025)

The following is a description of the components of the R-U-F model that intersect (Maksymov, 2025): risk and uncertainty (R-U): any risk is a manifestation of uncertainty, but uncertainty does not always lead to risk. For example, partial availability of statistical data if there is no complete confidence in the parameters. The R-U intersection shows that risk is a quantitative aspect of uncertainty and corresponds to a situation of partial uncertainty; risk and fuzzy information (R-F): fuzzy information can increase risk because it makes it difficult to accurately assess the probabilities and consequences of events. For example, decisions with probabilistic estimates but fuzzy criteria or results; uncertainty and fuzzy information (U-F): fuzzy information increases the level of uncertainty because it makes it harder to make decisions. In this case, this means a complete lack of certainty in the context of fuzzy information; risk, uncertainty, and fuzzy information (R-U-F): the most difficult case is when all three information decision-making situations are present. The common area of the three information situations indicates that the presence of fuzzy information simultaneously increases the level of uncertainty and risk. The intersections of the components of the R-U-F model reflect the relationships between information decision-making situations under conditions of risk, uncertainty, and fuzzy information.

RESULTS AND DISCUSSION

The "R-U-F" model of possible information decision-making situations allows classifying known and creating new decision-making methods according to the type of situation. The above model can also serve

as a basis for the development of adaptive information technology, which allows choosing or combining methods depending on the presence of R, U, and F factors in the decision-making problem being solved. In order to implement the capabilities provided by the "R-U-F" model, it is necessary to make a formal description of all information decision-making situations presented in Fig. 1: R, U, F, R-U, R-F, U-F, R-U-F.

1. Model of the information situation of decision-making under risk conditions (*R*):

$$MISDMR = \{A, \Theta, P, F\}, \quad (5)$$

where A, Θ, F – sets of the form (2), (3), (4) accordingly, P – set of risk levels, which depend on the probability of possible occurring states Θ of environment E:

$$P = \{p_1, p_2, \ldots, p_S\}, \quad (6)$$

$p_j = p(\theta_j)$ – probability of occurrence of the condition θ_j environment E, $j = \overline{1,S}$ that satisfy the condition:

$$p_j \in [0; 1], \sum_{j=1}^{S} p_j = 1, \quad (7)$$

F – evaluation functional of the form (4).

2. Model of the information situation of decision-making under conditions of complete uncertainty (*U*):

$$MISDMR = \{A, \Theta, U, F\}, \quad (8)$$

where U – set of levels of uncertainty of the environment E:

$$U = \{u_1, u_2, \ldots, u_S\}, \quad (9)$$

which corresponds to a situation of complete lack of information about the risk levels $P = \{p_1, p_2, \ldots, p_S\}$, in this case, as a rule:

$$u_j = p_j = 1/S. \quad (10)$$

3. Model of the information situation of decision-making under conditions of fuzzy information (*F*):

$$MISDMR = \{A, \Theta, L, F\mu\}, \quad (11)$$

where in the evaluation functionality values $F\mu$:

$$f\mu_{ij} = \mu_{ij} = \mu(A_i, \theta_j),\ i = \overline{1,N}, j = \overline{1,S}, \quad (12)$$

are fuzzy estimates that characterise the "gain" or "loss" of the decision-maker when choosing a decision A_i, if the environment will be (or is) in the state of θ_j, presented in the form of linguistic assessments from the set:

$$L = \{L_1, L_2, \ldots, L_r\}, \quad (13)$$

that is $\mu_{ij} \in L, i = \overline{1,N}, j = \overline{1,S}$, with fuzzy estimates from a set of linguistic assessments L in calculations, they can be represented, in particular, in the form of triangular, or trapezoidal fuzzy numbers.

4. Model of the information situation of decision-making under conditions of partial uncertainty (R-U):

$$MISDMR = \{A, \Theta, P, U, F\}, \quad (14)$$

where P – multiple risk levels:

$$P = \{p_1, p_2, \ldots, p_k\}, 1 \le k < S, \quad (15)$$

$p_j = p(\theta_j)$ – probability of occurrence of the condition θ_j of environment E that satisfy the condition $p_j \in [0; 1]$, $j = \overline{1,k}$, U – set of levels of partial uncertainty of the form:

$$U = \{u_{k+1}, u_{k+2}, \ldots, u_S\}, \quad (16)$$

where:

$$u_j = \frac{1 - \sum_{r=1}^{k} p_r}{S-k}, u_j \in [0,1],\ j = \overline{k+1, S}, \quad (17)$$

with this:

$$\sum_{j=1}^{k} p_j + \sum_{j=k+1}^{S} u_j = 1. \quad (18)$$

5. Model of the information situation of decision-making under risk and fuzzy information (R-F):

$$MISDMR = \{A, \Theta, P, L, F\mu\}, \quad (19)$$

where the set P is defined by the relations (6), (7), L – set of fuzzy estimates of the form (13), and in the evaluation functional $F\mu$ the quantities $f\mu_{ij} = \mu_{ij}$ – fuzzy estimates corresponding to condition (12).

It should be noted that there may be situations when the evaluation functionality $F\mu$ of part of the values $f\mu_{ij} = f\mu(A_i, \theta_j)$ may be clear estimates $f_{ij} = f(A_i, \theta_j)$, and some of the fuzzy estimates:

$$f\mu_{kl} = \mu_{kl} = \mu(A_k, \theta_l), \text{ where } i \ne k, j \ne l. \quad (20)$$

In such situations, it is necessary to perform the procedure for phasing values $f_{ij} = f(A_i, \theta_j)$ for the purpose of further use in calculations of fuzzy numbers.

6. Model of the information situation of decision-making under conditions of uncertainty and fuzzy information (U-F):

$$MISDMR = \{A, \Theta, U, L, F\mu\}, \quad (21)$$

where the set U meets the conditions (9), (10), L – set of fuzzy estimates of the form (13), and in the evaluation functional $F\mu$ he quantities $f\mu_{ij} = \mu_{ij}$ – fuzzy estimates that meet condition (12), or meet the conditions of the MISDMRF model, when part of the estimates $f\mu_{ij}$ may be clear estimates, and some of them may be fuzzy estimates.

7. Model of the information situation of decision-making under conditions of risk, uncertainty and fuzzy information (R-U-F):

$$MISDMR = \{A, \Theta, P, U, L, F\mu\}, \quad (22)$$

where P – set of risk levels of the form (15), U – set of uncertainty levels, either complete uncertainty of the form (9), (10) or partial uncertainty of the form (16), (17), with elements of sets P and U satisfying condition (18), L – set of fuzzy estimates of the form (13), and in the evaluation functional $F\mu$ values $f\mu_{ij} = \mu_{ij}$ – fuzzy estimates that either satisfy condition (12), or correspond to the case when part of the values $f\mu_{ij}$ may be crisp estimates, and the remaining values are fuzzy estimates, i.e., condition (20) is satisfied. The *MISDMRUF* model (22) is a generalisation of models (5)-(7), (8)-(10), (11)-(13), (14)-(18), (19)-(20), (21) and it can be adapted to various information decision-making situations that correspond to the information situations of the R-U-F conceptual model.

For the case when the set in (22) P contains S elements, i.e., satisfies conditions (6), (7), and sets U and L empty ($U = \emptyset$, $L = \emptyset$), and $F\mu = F$ type (4), there is an informational situation of decision-making under risk conditions. For the case when (22) $P = \emptyset$, $L = \emptyset$ and $F\mu = F$ of the form (4), that is, conditions (8)-(10) are met, and an information situation in decision-making occurs under conditions of complete uncertainty. For the case where in (22) $P = \emptyset$, $U = \emptyset$, $L = \emptyset$ and $F\mu$ all $f\mu_{ij} = \mu_{ij}$ – fuzzy estimates, i.e., conditions (12)-(13) are met, respectively, there is an information situation of decision-making under conditions of fuzzy information.

For the case when in (22) $P = \emptyset$ ($1 \le k < S$), $U = \emptyset$, $L = \emptyset$ and $F\mu = F$ of the form (4), i.e., conditions (14)-(18) are met, then an information situation of decision-making occurs under conditions of partial uncertainty. For the case where in (22) the set P has S elements, i.e., satisfies conditions (6), (7), set $U = \emptyset$, and in the evaluation functionality $F\mu$ values $f\mu_{ij} = \mu_{ij}$ – fuzzy estimates meeting condition (12), or when part of the estimates $f\mu_{ij}$ there may be clear estimates, and some of them are fuzzy estimates, i.e., condition (20) is met, then there is an information situation of decision-making under conditions of risk and fuzzy information. For the case where in (22) $P = \emptyset$, U meets the conditions (9), (10), L – a set of fuzzy estimates of the form (13), and in the estimated functionality $F\mu$ values $f\mu_{ij} = \mu_{ij}$ – fuzzy estimates that meet condition (12), an information situation of decision-making occurs under conditions of uncertainty and fuzzy information.

If the situations of the set Θ of environment E, so that the situations are initially located θ_j for which the probability of their occurrence is known $p_j = p(\theta_j), j = \overline{1,k}$, where $k < S$, and then located situations θ_j $j = \overline{k+1,S}$), for which the probabilities of their occurrence are unknown, i.e., a set of levels of partial uncertainty U of the form (16)-(18) is nonempty, while, for certainty, it is assumed that in the estimated functional $F\mu$ values $f\mu_{ij} = \mu_{ij}$ $(i = \overline{1,N}, j = \overline{1,k})$ – clear estimates, and values $f\mu_{ij} = \mu_{ij}$ $(i = \overline{1,N}, j = \overline{k+1,S})$ – fuzzy estimates, then a detailed model of the information situation of decision-making *MISDMRUF* under conditions of risk, uncertainty, and fuzzy information, form (22) is characterised by a matrix of:

Θ	θ_1	θ_2	...	θ_k	θ_{k+1}	θ_{k+2}	...	θ_S
P	p_1	p_2	...	p_k	u_{k+1}	u_{k+2}	...	u_S
A_1	f_{11}	f_{12}	...	f_{1k}	μ_{1k+1}	μ_{1k+2}	...	μ_{1S}
A_2	f_{21}	f_{22}	...	f_{2k}	μ_{2k+1}	μ_{2k+2}	...	μ_{2S}
...	...	...	...	...	...	...	...	...
A_i	f_{i1}	f_{i2}	...	f_{ik}	μ_{ik+1}	μ_{ik+2}	...	μ_{iS}
...	...	...	...	...	...	...	...	...
A_N	f_{N1}	f_{N2}	...	f_{Nk}	μ_{Nk+1}	μ_{Nk+2}	...	μ_{NS}

(23)

Figure 2 shows a model of the decision-making process under risk, uncertainty, and fuzzy information (MDMPRUF) that integrates the set-theoretic R-U-F model (Fig. 1) and the generalised model of the information situation of decision-making MISDMRUF (22).

The key features of the decision-making model under conditions of risk, uncertainty, and fuzzy information are versatility, flexibility, and integrativity. Versatility is shown when considering various combinations of information situations (R, U, F and their combinations), which allows covering a wide range of practical tasks. Flexibility is provided by the presence of several categories of methods, which allows adapting the approach to the specifics of a particular task. Integrativity consists in aggregating results using multiple methods, which increases the reliability of the final solution. This model fully reflects the logic of building information technology for decision-making in complex environments and can be used as a basis for implementing decision support systems (DSS) in various industries. The main stages of implementing a model of the decision-making process in conditions of risk, uncertainty, and fuzzy information, which corresponds to the proposed conceptual model R-U-F of possible information situations of decision-making, are given below.

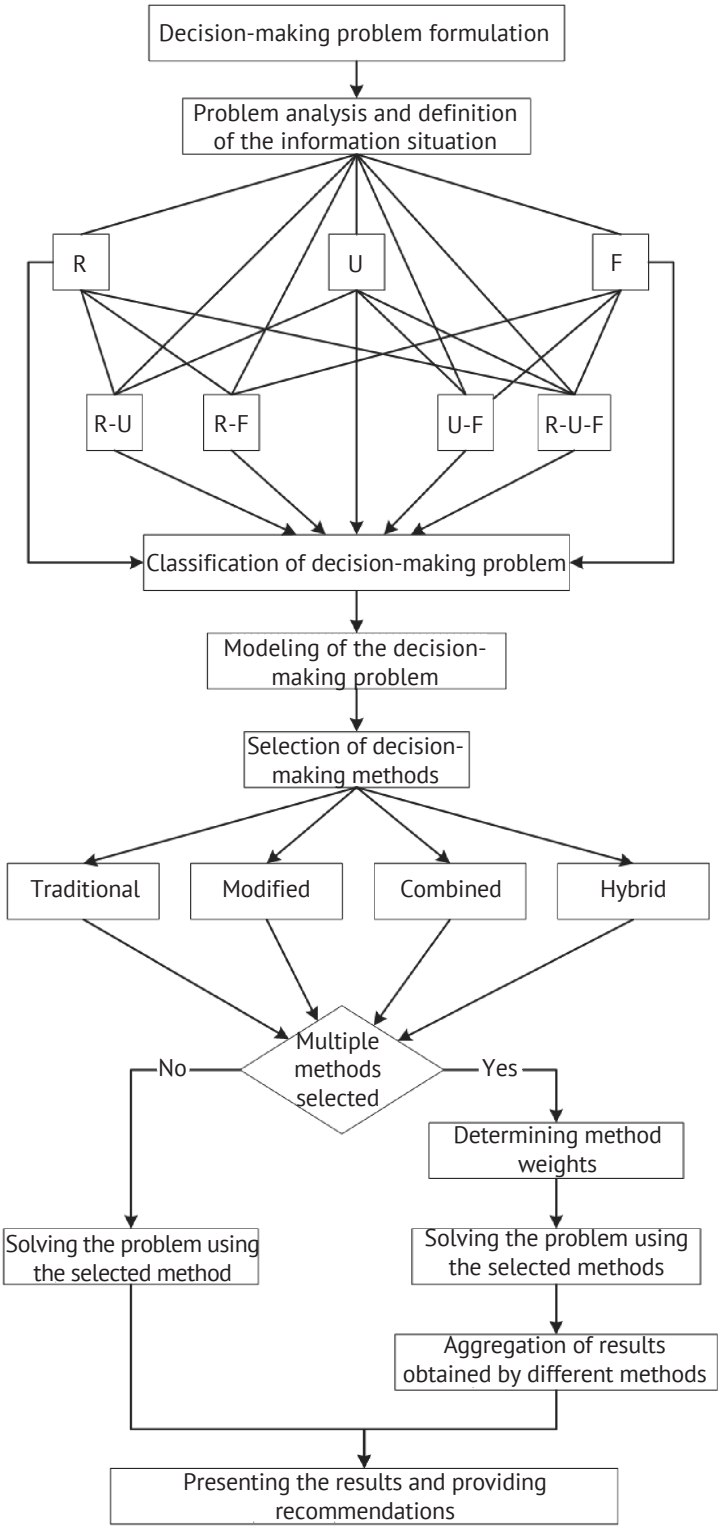

Figure 2. Model of the decision-making process under conditions of risk, uncertainty, and fuzzy information (MDMPRUF)

Source: created by the authors based on A.Y. Maksymov (2025)

Stage 1. Decision-making problem formulation. This stage defines the context and conditions in which the decision is made. A correctly formulated problem is the key to the effectiveness of subsequent analytical and computational procedures. The main steps of problem formulation are given below.

Step 1.1. Defining the purpose of decision-making. The very first task is to clearly define the main goal (or set of goals), the achievement of which is necessary within the framework of a problem situation, that is, the goal can be either single-goal or multi-goal.

Step 1.2. Defining a set of alternatives. This step describes a set of possible alternative solutions that are considered during the selection process. Alternatives can be discrete (for example, specific solutions) or continuous (when the solution belongs to a specific range of values).

Step 1.3. Defining restrictions and acceptable decision-making areas. At this step, the constraints corresponding to the problem condition are formulated. These can be: strict constraints (for example, budget constraints, technical requirements); fuzzy constraints (for example, acceptable risk limits described in fuzzy sets). At this stage, it is important to define a valid area that defines the space of possible solutions.

Step 1.4. Defining the goal of the decision-making task. The decision-making task implies the existence of a clearly formulated goal, according to which alternatives are evaluated. As a rule, this goal is to choose the best alternative according to the specified criteria.

Stage 2. Problem analysis and definition of the information situation. This stage combines two important tasks: deep analysis of the task and determination of the information situation of decision-making. The purpose of this stage was to investigate in detail the initial data, requirements, and conditions of the problem to create a complete picture of the problem and its context. This allows laying the foundation for the next stages and ensuring its relevance and compliance with real conditions.

Task 2.1. Task analysis. This task was aimed at a comprehensive investigation of the problem being solved, in order to clearly understand it and determine key parameters, and involves performing several steps:

Step 2.1.1. Analysis of goals and objectives. At this stage, the industry or sphere within which the task arose is defined, a specific problem that needs to be solved is outlined, and the main goals of the task are analysed, which will determine the area of decision-making, for example: improving efficiency, minimising costs, reducing risks, etc.

Step 2.1.2. Identification of stakeholders. The subjects that are affected by the task (the company's management, government agencies, the public, etc.) and whose interests can influence the criteria and choice of solutions are identified.

Step 2.1.3. Assessment of the complexity and structure of the task. The nature of the alternatives (discrete or continuous), the number of criteria, the presence of a multi-level task structure, and possible conflicts between criteria are analysed.

Step 2.1.4. Defining restrictions. Restrictions that affect the task are highlighted.

Task 2.2. Determination of the information situation of decision-making. This task is to determine the nature of available information about alternatives, criteria, and decision-making environments. When solving this problem, the type and type of information situation of decision-making is identified. Both individual types of these situations (R, U, F) and their combinations (R-U, R-F, U-F, R-U-F) are taken into consideration. Analysis of the information situation of decision-making allows: avoiding erroneous application of methods that do not correspond to the real conditions of the task; ensuring the adequacy of the task model to the real situation; responding to changes in the information environment through the possibility of repeated analysis.

Solving this problem allows clearly structuring the main components of the task, consider the existing limitations and conditions of risk, uncertainty, and fuzzy information that affect decision-making. This task is aimed at determining the available information about the situation, which makes it possible to form a reasonable strategy for further actions based on it. Defining a situation usually involves identifying and evaluating the main factors that influence decision-making, and the degree of confidence or risk associated with each of them. After the analysis, a certain information situation is recorded, which makes it possible to perform further actions. This task involves performing the following steps.

Step 2.2.1. Evaluation of available information. At this stage, it is important to evaluate exactly what data is available for analysis. These can be statistics, expert assessments, previous experience, or actual observations. If the information is incomplete or fuzzy, it is necessary to specify the exact nature of its uncertainty or imprecision. This assessment facilitates informed decision-making and ensures that associated risks are properly considered.

Step 2.2.2. Classification of the type of input information. At this stage, the types of information that the model will work with are determined: clear information – when the values of all parameters of the situation are known accurately; fuzzy information – when there is ambiguity or indistinctness in determining the values of parameters, which requires the use of fuzzy logic; information with elements of uncertainty – when it comes to the lack of sufficient data or uncertainty about the possible consequences of specific alternatives; information with elements of risk – when the probability of occurrence of various states of external conditions is known, to which statistical or other type of analysis can be applied to assess the probabilities of risk.

Step 2.2.3. Identification of critical parameters and factors. After determining the type of input information, it is important to identify key factors that may influence the decision-making process. These can be both internal and external factors. Each of the factors should be analysed for their impact on the choice of alternative solutions.

Step 2.2.4. Determining the level of confidence in data. An important step is to determine the degree of data reliability, which allows distinguishing between important and less significant aspects of the situation. This is especially important when using fuzzy or incomplete information.

Step 2.2.5. Creating a context for problem formulation. The final step of task 2.2 is to create an information context for the next steps. This means a clear understanding of the information situation of decision-making with all its parameters and factors, which allows accurately formulating the problem and determining which actions will be most effective. The context includes both informational and situational aspects that form the basis for decision-making.

Stage 3. Classification of decision-making problem. As a result of completing the previous stages, the task belongs to a certain type of information situations in decision-making.

Stage 4. Modeling of the decision-making problem. At this stage, a mathematical model of the decision-making problem is created that coincides with one of the developed models: MISDMR (5)-(7), MISDMR (8)-(10), MISDMR (11)-(13), MISDMR (14)-(18), MISDMR (19)-(20), MISDMR (21), MISDMR (22).

Stage 5. Selection of decision-making methods. At this stage, one or more methods are selected that most adequately consider the specifics of the problem model (22), (23), the type of input information, risk and uncertainty conditions, and requirements for accuracy, reliability and speed of obtaining results.

Stage 6. Analysis of the number of selected methods. If one method is selected to solve the problem, then step 6.1 is performed. If more than one method is selected, step 6.2 is performed.

Step 6.1. If one method is selected to solve the problem, then the user applies the DSS to solve the problem using the selected method and proceeds to stage 7.

Step 6.2. If more than one method is selected, the following sub-steps are performed.

Sub-step 6.2.1. Determining method weights. Determining the weights of methods allows quantifying their relative significance and coordinating their contribution to the overall solution. For example, the weights of methods can be calculated by comparing them in pairs using the hierarchy analysis method.

Sub-step 6.2.2. Solving the problem using the selected methods. The user applies the DSS to consistently solve the problem for each of the selected methods. At this stage, information technology should load input information and, if necessary, adapt it to the requirements of specific method classes. Specifically, for methods operating under conditions of fuzzy information, it is necessary to fuzzify the data, i.e., convert precise values into a fuzzy format using the framework of fuzzy sets. In cases where the input data is already provided in a fuzzy form, but methods that do not incorporate fuzzy set mechanisms are applied, it is necessary to defuzzify the data, i.e., convert fuzzy estimates into precise values.

Sub-step 6.2.3. Aggregation of results obtained by different methods. At this step, the results obtained by different decision-making methods are integrated into a single consistent solution that considers the impact of each method according to certain weights. The results of each method can be presented at different scales (ratings, utility function values, fuzzy estimates, etc.). To ensure correct aggregation, they are brought to a common unified scale (for example, normalized within the range [0;1]). The results of each method are then multiplied by the corresponding weighting factor, which reflects its significance in the overall decision-making model. This ensures a balance between methods with different influence on decision-making.

Stage 7. Presenting the results and providing recommendations. DSS generates recommendations for choosing the best alternative based on the results of calculations performed in previous stages. The development of recommendations is based on the results of applying decision-making methods selected by the user, which can be presented as a vector of priorities, ranks, etc. Providing recommendations completes the decision-making cycle by translating formalised calculations into practical steps.

The generalised model of the information situation of decision-making MISDMRUF (22) integrates models of other information decision situations that correspond to the information situations of the R-U-F conceptual model, since it includes all their key components. In particular, in the absence of fuzzy estimates, the model is reduced to a model of a decision-making problem under risk or uncertainty conditions, and in the absence of stochastic factors, to models of decision-making problems in a fuzzy environment. Consequently, the MISDMRUF (22) model provides a single formal basis for describing a wide range of information situations of decision-making. In turn, the MISDMRUF model can be interpreted as a model of a decision-making problem that formalises the process of choosing an alternative in conditions of simultaneous presence of risk, uncertainty, and fuzzy information.

The practical value of the proposed MISDMRUF model lies in the fact that it allows systematising the process of analysing the information environment of the problem and provides a methodological basis for choosing effective decision-making methods. Simultaneously, the limits of application of the model are determined by the class of decision-making problems in complex information conditions. The model is primarily focused on tasks that combine statistical data, expert assessments, and fuzzy information, which is typical for managerial, economic, engineering, and socio-economic systems.

The results of the study are consistent with the analysed literature, which also emphasises the need to systematise and formalise information situations in order to increase the validity of decision-making. S.A. Us & L.S. Koriashkina (2018) investigated typical information situations that arise in the process of choosing alternatives, in particular, in conditions of risk, uncertainty, and fuzzy information. The researchers emphasised the need to formalise input information, which allows

for a more systematic approach to the decision-making process and increases the accuracy of evaluating alternatives. O.F. Voloshyn & S.O. Mashchenko (2018) reviewed the fundamentals of decision theory, including basic principles, utility theory, expert procedures, and approaches to decision-making in terms of certainty, conflict, and fuzzy information. Their study demonstrated the relevance of systematising information and using structured approaches to improve the effectiveness of the decision-making process.

S.K. Kaya *et al*. (2022) proposed a hybrid model of fuzzy multi-criteria decision-making based on the Fuzzy Best-Worst Method (F-BWM) and RAFSI, focused on considering the uncertainty of expert assessments and improving the stability of ranking alternatives. The results showed the effectiveness of integrating fuzzy approaches to accurately evaluate alternatives. The MISDMRUF model developed in this study extended this approach by integrating risk, uncertainty, and fuzzy information, and their combinations, allowing for systematic classification of information situations and the choice of appropriate decision-making methods. M. Aregbesola *et al*. (2025) applied fuzzy logic to risk management in project tasks, demonstrating the effectiveness of quality information processing and risk assessment in various scenarios. In comparison, the MDMPRUF model focuses on formalising the decision-making process, providing the ability to use multiple methods to select alternatives and aggregate results. This approach increases flexibility and practical applicability in complex environments of risk, uncertainty and fuzzy information.

A. Puška *et al*. (2022) proposed a hybrid Z-numbers – Fuzzy LMAW – Fuzzy CRADIS model that considers both the fuzziness and reliability of expert assessments. The conducted experiments showed the stability of the model and the ability to reduce the level of uncertainty. MISDMRUF in this study combines these factors together with risk and uncertainty, which allows comprehensively classifying information situations and choose methods based on their characteristics. MDMPRUF, in turn, formalises the decision-making process and provides multiple methods with the ability to aggregate results, increasing efficiency and practical applicability. The study by A. Ali *et al*. (2024) proposed a three-step method of multi-criteria decision-making using fuzzy soft relations and additive consistency, which allows simultaneous ranking, classification and selection of alternatives in complex information situations. MISDMRUF extends this approach by integrating risk, uncertainty, and fuzzy information into a wider range of situations, while MDMPRUF focuses on formalising the decision-making process and aggregating results to improve the accuracy of evaluating alternatives.

M. Kozielski *et al*. (2026) considered aggregation functions for working with fuzzy data, improving the consistency and interpretability of solutions. The MISDMRUF model integrates aggregation into the broader context of information situations, combining risk, uncertainty, and fuzzy information to ensure a systematic and consistent decision-making process. I. Hasiuk & O. Ivanii (2024) analysed the combination of qualitative and quantitative criteria and the integration of various logic systems in multi-criteria control problems. The results showed that the effectiveness of decision-making depends on the standardisation of procedures and a clear distribution of roles. MISDMRUF complies with these principles by integrating various methods and criteria into a single system, which increases the structure and efficiency of alternative selection.

The study by O. Syniuk (2025) analysed the relationship between cognitive factors and analytical methods, demonstrating how peer review and cognitive biases affect the choice of alternatives. The MISDMRUF model takes these aspects into account by combining information situations of decision-making in conditions of risk, uncertainty and fuzzy information, which increases the reliability of results and practical effectiveness of decision-making. The study by M. Aslantas *et al*. (2025) proposed an integrated fuzzy multi-criteria approach for assessing the potential risks of autonomous vehicles, which considers expert assessments and optimistic and pessimistic scenarios, demonstrating the effectiveness and rationality of the choice of methods in the analysis of complex and uncertain information situations. R. Rajesh (2024) analysed decision-making methods in conditions of uncertainty during pandemics using Grey models, which demonstrated the effectiveness of these approaches for decision-making in complex and incompletely defined information situations.

J. Więckowski *et al*. (2025) proposed the Fuzzy RANCOM method for modelling uncertainty in decision-making processes, which provides effective determination of the weights of criteria and reduces the impact of inaccuracies in expert assessments, demonstrating the applicability of fuzzy multi-criteria approaches in complex and incompletely defined information situations. Thus, research by R. Rajesh (2024), V.J. Montero *et al*. (2025), M. Aslantas *et al*. (2025) and J. Więckowski *et al*. (2025) demonstrated the practical effectiveness of risk and uncertainty modelling, with MISDMRUF consistent with these approaches, providing an integrated assessment of risk, uncertainty, and fuzzy information, and MDMPRUF formalising the decision-making process and enabling multiple methods with aggregated results to improve efficiency.

The results of the study showed that the MISDMRUF and MDMPRUF models, on the one hand, are consistent with a number of classical and contemporary studies in the field of decision-making, and, on the other hand, fill in existing gaps, providing a systematic and integrated approach to the classification of information situations of decision-making and the choice of decision-making methods under complex conditions of risk, uncertainty, and fuzzy information.

CONCLUSIONS

As a result of the research conducted based on the conceptual model R-U-F, mathematical models of all possible information decision-making situations were constructed: R, U, F, R-U, R-F, U-F, R-U-F. A generalised model of the information situation of decision-making MISDMRUF of the form (22), (23) was also presented, which can be interpreted as a problem of single-criteria decision-making under conditions of risk, uncertainty, and fuzzy information. A model of the MDMPRUF decision-making process was developed, focused on the conditions of risk, uncertainty, and fuzzy information, which used the proposed mathematical model of possible information situations of decision-making of the MISDMRUF type. The MDMPRUF model provided for the possibility of using one or more methods for selecting alternatives, in particular, traditional, modified, combined, and hybrid decision-making methods, the results of which can be aggregated. The proposed model of the MDMPRUF decision-making process defines its main tasks, and structures the sequence of key stages and steps necessary for solving the choice problem in complex information conditions. It can be the basis for the development of adaptive information technology, which will allow choosing or combining methods depending on the presence of R, U, and F factors in the decision-making problem.

Prospects for further research are: development of modified, combined and hybrid methods that solve decision-making problems in information situations R-U, R-F, U-F, R-U-F, and can be used at Stage 5 of the proposed MDMPRUF decision-making model; expansion of the proposed model MISDMRUF information situations of decision-making on the problem of multi-criteria choice in conditions of risk, uncertainty, and fuzzy information, and in the development of appropriate methods for their solution; the development of DSS based on the proposed models and methods for solving decision-making problems in conditions of risk, uncertainty, and fuzzy information.

ACKNOWLEDGEMENTS

None.

FUNDING

None.

CONFLICT OF INTEREST

None.

REFERENCES

[1] Ali, A., Rehman, N., Ali, M., & Hila, K. (2024). A novel approach to three-way decision model under fuzzy soft dominance degree relations and emergency situation. *Expert Systems with Applications*, 239, article number 122369. doi: 10.1016/j.eswa.2023.122369.

[2] Aregbesola, G.D., Asghar, I., Akbar, S., & Ullah, R. (2025). Fuzzy logic model for informed decision-making in risk assessment during software design. *Systems*, 13(9), article number 825. doi: 10.3390/systems13090825.

[3] Aslantas, M., Gündoğdu, F.K., & Moslem, S. (2025). Evaluating the potential risks posed by autonomous vehicles by using a decomposed fuzzy multi-criteria decision-making model. *Transportation Engineering*, 21, article number 100372. doi: 10.1016/j.treng.2025.100372.

[4] Gomes, M.I., & Martins, N.C. (2022). *Mathematical models for decision making with multiple perspectives: An introduction.* Boca Raton: CRC Press. doi: 10.1201/9781003015154.

[5] Hasiuk, I., & Ivanii, O. (2024). Logical-formal and criteria methods of management decision-making in public administration. *Scientific Perspectives*, 8(50), 112-129. doi: 10.52058/2708-7530-2024-8(50)-112-129.

[6] Høj, N.P., Kroon, I.B., Nielsen, J.S., & Schubert, M. (2025). System risk modelling and decision-making – reflections and common pitfalls. *Structural Safety*, 113, article number 102469. doi: 10.1016/j.strusafe.2024.102469.

[7] Jaccard, J., & Jacoby, J. (2020). *Theory construction and model-building skills: A practical guide for social scientists (2nd ed.).* New York: The Guilford Press.

[8] Kaya, S.K., Pamucar, D., & Aycin, E. (2022). A new hybrid fuzzy multi-criteria decision methodology for prioritizing the antivirus mask over COVID-19 pandemic. *Informatica*, 33(3), 545-572. doi: 10.15388/22-INFOR475.

[9] Kozielski, M., Prokopowicz, P., & Mikolajewski, D. (2026). New fuzzy aggregators for ordered fuzzy numbers for trend and uncertainty analysis. *Electronics*, 15(2), article number 309. doi: 10.3390/electronics15020309.

[10] Li, T., Ali, M.K.M., & Zhou, X. (2025). Interval-valued hesitant fuzzy decision model based on improved projection measure. *International Journal of Fuzzy Systems*. doi: 10.1007/s40815-025-02113-x.

[11] Maksymov, A.Y. (2025). *Information technology for solving multi-criteria decision-making problems under conditions of risk, uncertainty and fuzzy information*. (Doctoral dissertation, Cherkasy State Technological University, Cherkasy, Ukraine).

[12] Montero, V.J., Logrosa, G., Calorio, J.L., Lato, J.I., Hassall, M., & Mata, M.A. (2025). Risk modeling with Bowtie method for decision-making towards public health and safety. *Safety Science*, 185, article number 106777. doi: 10.1016/j.ssci.2025.106777.

[13] Munier, N. (2021). *Mathematical modelling of decision problems: Using the SIMUS method for complex scenarios*. Cham: Springer International Publishing. doi: 10.1007/978-3-030-82347-4.

[14] Puška, A., Božanić, D., Nedeljković, M., & Janošević, M. (2022). Green supplier selection in an uncertain environment in agriculture using a hybrid MCDM model: Z-numbers-Fuzzy LMAW-Fuzzy CRADIS model. *Axioms*, 11(9), article number 427. doi: 10.3390/axioms11090427.
[15] Rajesh, R. (2024). Grey models for data analysis and decision-making in uncertainty during pandemics. *International Journal of Disaster Risk Reduction*, 113, article number 104881. doi: 10.1016/j.ijdrr.2024.104881.
[16] Syniuk, O. (2025). Behavioral and analytical aspects of managerial decision-making in a changing environment. *Economy and Society*, 80. doi: 10.32782/2524-0072/2025-80-53.
[17] Us, S.A., & Koriashkina, L.S. (2018). *Models and methods of decision making*. Dnipro: Dnipro University of Technology.
[18] Voloshyn, O.F., & Mashchenko, S.O. (2018). *Models and methods of decision making* (3rd ed.). Kyiv: Liudmyla Publishing House.
[19] Więckowski, J., Kizielewicz, B., & Sałabun, W. (2025). Fuzzy RANCOM: A novel approach for modeling uncertainty in decision-making processes. *Information Sciences*, 694, article number 121716. doi: 10.1016/j.ins.2024.121716.
[20] Zheldak, T., Koryashkina, L., & Us, S. (2020). *Fuzzy sets in management and decision-making*. Dnipro: Dnipro University of Technology.

Моделі інформаційних ситуацій прийняття рішень в умовах ризику, невизначеності та нечіткої інформації

Антон Максимов

Доктор філософії у галузі комп'ютерних наук
Черкаський державний технологічний університет
18006, бул. Шевченка, 460, м. Черкаси, Україна
https://orcid.org/0009-0002-0411-8164

Юрій Триус

Доктор педагогічних наук, кандидат фізико-математичних наук, професор
Черкаський державний технологічний університет
18006, бул. Шевченка, 460, м. Черкаси, Україна
https://orcid.org/0000-0002-0739-2065

Анотація. Актуальність зумовлена потребою у створенні математичного забезпечення концептуального підходу, який би дозволив систематизувати інформаційні ситуації прийняття рішень в умовах ризику, невизначеності, нечіткої інформації та забезпечив обґрунтований вибір методів для розв'язання практичних задач прийняття рішень у цих ситуаціях. Мета полягала в розробці математичних моделей інформаційних ситуацій прийняття рішень в умовах ризику, невизначеності, нечіткої інформації та їх можливих комбінацій, а також у побудові на їх основі моделі процесу прийняття рішень, що передбачає можливість застосування кількох методів вибору альтернатив і агрегування їх результатів для підвищення ефективності прийняття рішень. Методологія дослідження ґрунтувалася на системному підході із застосуванням теорії множин, теорії ймовірностей, теорії нечітких множин, теорії і методів прийняття рішень, математичного моделювання для формалізації інформаційних ситуацій прийняття рішень. Розроблено математичні моделі інформаційних ситуацій прийняття рішень в умовах ризику (Risk, R), повної невизначеності (Uncertainty, U), нечіткої інформації (Fuzzy, F), а також можливих комбінацій ситуацій: R-U, R-F, U-F, R-U-F. Узагальнену модель інформаційних ситуацій прийняття рішень R-U-F (MISDMRUF) інтерпретовано як модель задачі прийняття рішень, що формалізує процес вибору альтернативи в умовах одночасної наявності ризику, невизначеності та нечіткої інформації. Це дозволило використовувати наявні та створювати нові методи прийняття рішень залежно від характеристик інформаційної ситуації прийняття рішень. Запропоновано модель процесу прийняття рішень в умовах ризику, невизначеності та нечіткої інформації (MDMPRUF), визначено її завдання, основні етапи реалізації та ключові особливості, зокрема універсальність, гнучкість та інтегративність. Результати засвідчили, що запропоновані моделі MISDMRUF та MDMPRUF узгоджуються з рядом досліджень у теорії прийняття рішень, але водночас заповнюють наявні прогалини, забезпечуючи системний, інтегрований підхід до класифікації інформаційних ситуацій прийняття рішень та вибору методів прийняття рішень у складних умовах ризику, невизначеності й нечіткої інформації. Запропоновані математичні моделі та відповідний підхід до розв'язання задач прийняття рішень можуть стати основою для створення інформаційних технологій підтримки прийняття рішень у складних умовах

Ключові слова: математичні моделі; моделі підтримки прийняття рішень; методи прийняття рішень; системи підтримки прийняття рішень; інформаційні технології

ВІСНИК
Черкаського державного технологічного університету

Науковий збірник

Том 31, № 1. 2026

Заснований у 1996 р. Виходить чотири рази на рік

Оригінал-макет видання виготовлено у редакційно-видавничому відділі
Черкаського державного технологічного університету

Відповідальний редактор:
А. Лавданський

Підписано до друку 16 березня 2026 р.
Формат 60*84/8
Умов. друк. арк. 18,1
Наклад 50 прим.

Адреса видавництва:
Черкаський державний технологічний університет
18006, бульв. Шевченка, 460, м. Черкаси, Україна
Тел.: +380638013283
E-mail: info@bulletin-chstu.com.ua
https://bulletin-chstu.com.ua/uk

BULLETIN
of Cherkasy State Technological University

Scientific Collection

Volume 31, No. 1. 2026

Founded in 1996. Published four times per year

The original layout of the publication is made in the editorial and publishing department of Cherkasy State Technological University

Managing editor:
A. Lavdanskyi

Signed for print of March 16, 2026.
Format 60*84/8
Conventional printed pages 18.1
Circulation 50 copies

Editors Office Address:

Cherkasy State Technological University
18006, 460 Shevchenko Blvd., Cherkasy, Ukraine
Tel.: +380638013283
E-mail: info@bulletin-chstu.com.ua
https://bulletin-chstu.com.ua/en