ISSN 2786-6254 (Print)
ISSN 2786-6262 (Online)

Міністерство освіти і науки України
Запорізький національний університет

Заснований
у 1997 р.

Реєстрація суб'єкта у сфері друкованих медіа: Рішення Національної ради України з питань телебачення і радіомовлення № 2607 від 29.08.2024 року.

Computer Science and Applied Mathematics

Адреса редакції:
вул. Університетська 66, корп. 1, ауд. 21(б),
м. Запоріжжя, Україна, 69060

Телефон
для довідок:
+38 066 53 57 687

№ 2, 2025

Видавничий дім
«Гельветика»
2025

DOI Computer Science and Applied Mathematics https://doi.org/10.26661/2786-6254
DOI № 2/2025 https://doi.org/10.26661/2786-6254-2025-2

Рекомендовано до друку та поширення через мережу Internet вченою радою ЗНУ (протокол засідання № 7 від 29.12.2025 р.)

На підставі Наказу Міністерства освіти і науки України № 886 від 02.07.2020 р. (додаток 4) збірник включено до Переліку наукових фахових видань України категорії «Б» у галузі знань інформаційні технології (F1 – Прикладна математика, F2 – Інженерія програмного забезпечення, F3 – Комп'ютерні науки).

До 25 березня 2021 р. журнал виходив під назвою «Вісник Запорізького національного університету. Фізико-математичні науки».
У зв'язку зі зміною назви журналу було внесено відповідні зміни до Переліку наукових фахових видань України на підставі Наказу Міністерства освіти та науки України № 735 від 29.06.2021 р. (додаток 3).

Журнал індексується в міжнародній наукометричній базі даних Index Copernicus

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

ISSN 2786-6254 (Print)
ISSN 2786-6262 (Online)

ЗМІСТ

РОЗДІЛ І. ПРИКЛАДНА МАТЕМАТИКА

РОЗДІЛ ІІ. ІНЖЕНЕРІЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

РОЗДІЛ ІІІ. КОМП'ЮТЕРНІ НАУКИ

CONTENTS

SECTION I. APPLIED MATHEMATICS

SECTION II. SOFTWARE ENGINEERING

SECTION III. COMPUTER SCIENCES

РОЗДІЛ І. ПРИКЛАДНА МАТЕМАТИКА

UCD 004.94
DOI https://doi.org/10.26661/2786-6254-2025-2-01

МЕТАЕВРІСТИКИ НА БАЗІ ФРАГМЕНТАРНИХ МОДЕЛЕЙ ДЛЯ ЗАДАЧІ ПОКРИТТЯ МНОЖИН ТОЧОК НА ПЛОЩИНІ

Козін І. В.
доктор фізико-математичних наук, професор,
професор кафедри економічної кібернетики
Запорізький національний університет
вул. Університетська, 66, Запоріжжя, Україна
orcid.org/0000-0003-1278-8520
ainc00@gmail.com

Сардак О. В.
аспірант спеціальності прикладна математика
Запорізький національний університет
вул. Університетська, 66, Запоріжжя, Україна
orcid.org/0009-0001-2087-0287
karnelcore@gmail.com

Ключові слова: *дискретна оптимізація, задача покриття множини точок, метаевристики, орієнтована фрагментарна структура, фрагментарна модель.*

Задача покриття кінцевих множин точок на площині фігурами різних типів належить до складних задач дискретної оптимізації. Існує досить багато варіантів цієї задачі, більшість яких належать до класу NP-важких задач.
Для пошуку точних розв'язків задач покриття відсутні алгоритми поліноміальної трудомісткості. Звичайно, існують точні алгоритми для цієї задачі, але всі вони зводяться до повного перебору варіантів. Навіть більше, невідомі й наближені алгоритми розв'язку таких задач із заданими оцінками точності. Тому значний інтерес становлять розроблення та дослідження евристичних методів оптимізації. Одним із перспективних напрямів є розроблення алгоритмів, заснованих на відомих метаевристичних підходах, які з успіхом використовуються для вирішення багатьох задач дискретної оптимізації.
Одним зі слабких місць використання метаевристик є наявність обмежень у багатьох задачах точкового покриття. Зокрема, це всілякі перепони, обмеження розміщення точок, обмеження, пов'язані з можливими перетвореннями точок у координатному просторі (зрушення, повороти). Тому найчастіше доводиться модифікувати відомі алгоритми. Це призводить до великої різноманітності метаевристик, які відомі за однією назвою (генетичний алгоритм, мурашиний алгоритм), але насправді орієнтовані на чітко визначене коло задач. Це призводить до труднощів у створенні програмних продуктів, які стають індивідуальними для різних задач.
Основна мета даної роботи – розробити універсальний підхід до розв'язання складних в обчислювальному сенсі задач, який дозволяє використовувати відомі метаевристики для різних класів задач покриття без суттєвих змін алгоритмів. Для цього запропоновано метод побудови гібридних алгоритмів, що є комбінацією фрагментарного алгоритму, який є індивідуальним для кожного виду задачі покриття, і метаевристики, яка вже не прив'язана до конкретного класу задач. Такий метод може бути легко перенесений і на інші класи задач дискретної оптимізації, для яких можна побудувати фрагментарну модель.

THE SHUFFLE FROG LEAPING ALGORITHM FOR THE PRODUCTION LOCATION PROBLEM

Kozin I. V.
Doctor of Physical and Mathematical Sciences, Professor,
Professor at the Department of Economic Cybernetics
Zaporizhzhia National University
Universytetska str., 66, Zaporizhzhia, Ukraine
orcid.org/0000-0003-1278-8520
ainc00@gmail.com

Sardak O. V.
Postgraduate Student in the Specialty "Applied Mathematics"
Zaporizhzhia National University
Universytetska str., 66, Zaporizhzhia, Ukraine
orcid.org/0009-0001-2087-0287
karnelcore@gmail.com

Key words: *discrete optimization, multiply point problem, metaheuristics, oriented fragmentary structure, fragmentary model.*

The task of covering the end multiplicities of points on a plane with figures of various types can be extended to complex problems of discrete optimization. There are a lot of options for this problem, most of which fall into the class of NP-important problems. Thus, in order to find precise solutions to problems using daily polynomial complexity algorithms. It is important to find precise algorithms for this task, otherwise it will all be reduced to a complete enumeration of options. Moreover, there are unknown and close algorithms for decoupling such problems from given accuracy estimates. Therefore, it is of significant interest to develop and investigate heuristic optimization methods. One of the promising directions is the development of algorithms based on familiar meta-heuristic approaches that can be successfully used for a wide variety of discrete optimization problems.

One of the weaknesses of using metaheuristics is the presence of restrictions in many point coverage problems. In particular, these are all sorts of obstacles, restrictions on the placement of points, restrictions associated with possible transformations of points in coordinate space (shifts, rotations). Therefore, it is often necessary to modify known algorithms. This leads to a wide variety of metaheuristics, which are known by one name (genetic algorithm, ant algorithm), but are actually focused on a clearly defined range of problems. In turn, this leads to difficulties in creating software products that become individual for different problems.

The main goal of this work is to develop a universal approach to solving computationally complex problems, which allows using known metaheuristics for different classes of coverage problems without significant changes to the algorithms. For this purpose, the method of randomized hybrid algorithms is proposed, which is a combination of a fragmented algorithm, which is individual for each type of problem, and meta-heuristics, which is no longer tied to a specific class of problems. This method can be easily transferred to other classes of discrete optimization problems, for which a fragmented model may be needed.

Вступ. Задачі покриття множин точок на площині вирізняються великим різновидами постановок. Серед них існують як задачі неперервного типу, так і задачі дискретні. Більшість класів дискретних задач оптимального покриття точок відносять до *NP*-важких задач [1–3], для яких невідомі точні алгоритми пошуку оптимального розв'язку, складність яких обмежена поліномом від довжини умови задачі. Зокрема, до таких класів належать різні варіанти задачі покриття точок різноманітними фігурами на площині як без перетинів цих фігур, так і з можливими перетинами,

різні варіанти задач прямокутного розкрою [4–7], задачі покриття вершин графа типовими підграфами [8; 9] та багато інших [10; 11].

Такі задачі часто виникають у металургії, текстильній і деревообробній промисловості, землекористуванні й інших практично важливих напрямах діяльності. Відсутність точних алгоритмів розв'язання таких задач призводить до різноманітності різних наближених методів, які зазвичай не є універсальними та розраховані на використання в конкретних індивідуальних постановках. Узагальненням таких методів є метаевристики різних типів [12–14], які за розумний час дозволяють отримати наближені розв'язки задач (хоч і без теоретичного обґрунтування). Відносна простота метаевристик робить їх зручним і простим інструментом для створення комп'ютерних програм для різних завдань, які виникають на практиці.

Підтвердженням якості метаевристик є, по-перше, накопичений досвід їх використання, по-друге, перевірка на великій кількості тестових завдань. На жаль, різноманітність обмежень реальних задач призводить до того, що застосовувані метаевристики стають індивідуальними і не можуть бути просто перенесені на інші класи задач.

У роботі пропонується підхід до дискретних задач покриття множин точок із використанням орієнтованих фрагментарних структур. Використання цих об'єктів дозволяє розробити гібридний алгоритм на базі комбінації фрагментарного алгоритму й будь-якої метаевристики. До того ж усі обмеження задачі враховуються на етапі роботи фрагментарного алгоритму, а метаевристика стає універсальною і не залежить від обмежень задачі.

Огляд літератури. У літературі відомо багато різних постановок задач покриття множин точок на площині. Почнемо з найбільш простих задач покриття множин точок однаковими фігурами [15–21].

Нехай задана кінцева множина точок площини $X = \{A_1,...,A_n\} \subseteq R^2$. Кожна точка A_i визначається своїми Декартовими координатами (x_i, y_i).

Задача покриття множини точок колами заданого радіуса.

Розглянемо набір кіл заданого радіуса *R* із центрами в точках множини *X*. Задача полягає у відшуканні мінімального числа кіл із цієї множини, які містять усі точки множини *X*. Тобто потрібно вибрати підмножину індексів $I \subseteq \{1,2,...,n\}$ мінімальної потужності так, що для будь-якого $k = 1,2,...,n$ знайдеться елемент $i_0 \in I$ такий, що $\sqrt{(x_k - x_{i_0})^2 + (y_k - y_{i_0})^2} \le R$.

Задача покриття множини точок колами різних радіусів.

Більш складною є задача покриття множини точок *X* колами, радіуси яких можуть бути різними. Розглядається множина різних позитивних чисел $T = \{R_1, R_2,..., R_s\}$. Потрібно покрити множину точок колами, радіуси яких належать множині *T*. Оптимальне покриття визначається двома критеріями – кількістю елементів покриття (має бути, за можливості, мінімальним) і сумарною площею кіл покриття. Другий критерій також має бути якнайменшим.

Ускладненням даних постановок є задача кратного покриття, в якій кожна точка множини *X* повинна покриватися не менш ніж $k (k > 1)$ елементами покриття. Також можна розглядати покриття фігурами, які не є колами.

Задача покриття плоских фігур.

Тепер розглянемо випадок, коли множина точок *X* є фігурою на площині (з нескінченною множиною точок). Найбільш відомими задачами цього класу є задачі прямокутного розкрою [4–7]. Задача гільйотиного розкрою розглядається в багатьох роботах. Для даного класу задач розроблено багато точних і наближених алгоритмів. Але вдосконалення цієї задачі, коли потрібно розташувати на заданій фігурі деякий набір заданих фігур, є досить складним.

Задача покриття вершин графа.

Окремо серед задач покриття стоїть задача покриття вершин графа [1–3]. Нехай задано граф $G = (V, E)$ із множиною вершин *V* і множиною ребер *E*. Потрібно знайти набір підграфів, що містить всі вершини множини *V*. Зазвичай на підграфи накладаються обмеження щодо виду підграфа (ребро, зірка, трикутник тощо).

Результати. ***Фрагментарна модель оптимізаційної задачі.*** Орієнтованою фрагментарною структурою [13] (X, E) на кінцевій множині *X* називається сімейство впорядкованих наборів $E = \{E_1, E_2,..., E_n\}$ таких його елементів, що для будь-якої непустої послідовності $E_i = (x_1, x_2,..., x_k) \in E$ будь-яка її початкова підпослідовність $(x_1, x_2,..., x_{k'})$, $k' < k$ також належить *E*.

Елементи із множини *E* називатимемо допустимими фрагментами. Елементарним фрагментом називатимемо допустимий фрагмент, що складається з одного елемента. Максимальний фрагмент – допустимий фрагмент, який не є підмножиною іншого фрагмента. Нехай $A \in E$. Умову для елемента $x \in X$, за якої $A \cup \{x\} \in E$, називатимемо умовою приєднання елемента *x*.

Визначимо фрагментарний алгоритм як алгоритм побудови максимального фрагмента фрагментарної структури.

Цей алгоритм належить до класу «жадібних» алгоритмів і складається з таких кроків:

а) на початковому кроці елементи множини *X* лінійно впорядковуються;

б) на першому кроці алгоритму обирається порожня множина $X_0 = \varnothing$;

в) на кроці з номером $k \geq 2$ вибирається перший за обраним упорядкуванням елемент $x \in X \setminus X_{k-1}$, такий, що має властивість приєднання $X_{k-1} \bigcup \{x\} \in E$;

г) алгоритм закінчує роботу, якщо на черговому кроці не вдалося знайти елемент $x \in X \setminus X_k$, який задовольняє властивості приєднання.

Максимальний фрагмент, який є результатом роботи фрагментарного алгоритму, залежить від обраного упорядкування на множині *X*. Кожне упорядкування кінцевої множини задається перестановкою її елементів. Таким чином, будь-який максимальний фрагмент може бути описаний відповідною перестановкою елементів множини *X*.

Припишімо кожному допустимому фрагменту невід'ємну вагу, тобто визначимо функцію $\rho : E \to R^1_+$. Припускатимемо, що функція ρ монотонна зростаюча за включенням. Якщо $A, B \in E$ і $A \subseteq B$, то $\rho(A) \leq \rho(B)$. Задача оптимізації на орієнтованій фрагментарної структурі – це задача відшукання допустимого фрагмента максимальної ваги. Очевидно, що для монотонно зростаючих ваг оптимальне рішення буде максимальним фрагментом.

Задача може ставитися також іншим чином, без урахування монотонного зростання ваг, а саме: знайти максимальний фрагмент максимальної (або мінімальної) ваги.

Кожний максимальний фрагмент визначається відповідно заданим лінійним порядком перегляду елементарних фрагментів, тобто перестановкою $s \in S_n$, де *n* – кількість елементів множини *X*. Ця перестановка визначає результат роботи фрагментарного алгоритму, який побудує необхідний максимальний фрагмент. У кожній перестановці $s \in S_n$ можна зіставити єдиний максимальний фрагмент, який їй породжується. Позначимо відповідне відображення через $\varphi : S_n \to E$. Має місце природна комутативна діаграма відображень:

$$\begin{array}{ccc} & S_n & \\ \varphi \downarrow & & \searrow F \circ \varphi, \\ E & \to & R^1 \end{array}$$

яка перетворює задачу оптимізації на орієнтованій фрагментарній структурі на задачу оптимізації на множині перестановок. Будь-яка перестановка є допустимою. Усі умови, які відповідають індивідуальній задачі, сховані в умові приєднання фрагментів. Для великих значень *n* задача пошуку оптимальної перестановки зазвичай є важкою в обчислювальному сенсі. Тож для таких задач виправдано застосування метаевристик на множині перестановок.

Деякі з таких метаеврістик описано в наступному розділі.

Метаеврістики для безумовної оптимізації на множині перестановок.

Розглянемо загальну задачу безумовної оптимізації на множині перестановок S_n з *n* елементів: знайти перестановку, для якої функція $F : S_n \to R^1_+$ приймає найменше (або найбільше) значення.

Розглянемо метаеврістики, що використовують метрику Кендала на просторі перестановок S_n. Відстанню між двома перестановками називається мінімальне число транспозицій, необхідне для перетворення першої перестановки на другу. Найпростіша з таких метаеврістик – *локальний алгоритм з випадковим вибором початкової точки.*

Локальний алгоритм передбачає такі кроки:

а) на початковому етапі вибирається випадкова перестановка (наприклад, за допомогою алгоритму Фішера – Йетса) та обчислюється значення цільової функції на цій перестановці;

б) на етапі з номером *k* проглядається 1-окіл перестановки, вибирається як нове наближення точка, у якій значення цільової функції мінімальне(максимальне) в околі, що розглядається;

в) якщо отримана на черговому етапі перестановка збігається з перестановкою, отриманою на попередньому етапі, то алгоритм закінчує роботу.

Другим прикладом метаеврістики, яку можна застосовувати для пошуку субоптимальних розв'язків на множині перестановок, є еволюційний алгоритм із геометричним оператором кросовера.

Використовується стандартна схема еволюційного алгоритму на перестановках. Наведемо коротко принцип роботи такого алгоритму [13; 19]. Як базова множина рішень вибирається множина всіх перестановок з *n* елементів. На початковому етапі застосовується оператор початкової популяції, що будує першу поточну популяцію рішень Y_0. На кроці з номером $k (k \geq 0)$ алгоритму визначена поточна популяція Y_k, яка на початковому кроці збігається з Y_0. Для кожного з елементів $s \in Y_k$ обчислюється значення критерію $F(s)$.

Далі за допомогою оператора відбору в поточній популяції Y_k вибирається множина пар для кросовера. До кожної пари з вибраної множини пар застосовується оператор кросовера *Cross*, а потім до результату кросовера застосовується оператор мутації. Нехай $U = (u_1, u_2, ..., u_n)$ і $V = (v_1, v_2, ..., v_n)$ – дві довільні перестановки. Перестановка-нащадок будується як внутрішня точка відрізка, що з'єднує перестановки-батьки. У метриці Кендала таку точку можна отримати так: послідовності *U* і *V* проглядаються з початку. На черговому кроці вибирається найменший з перших елементів послідовностей і додається в нову перестанов-

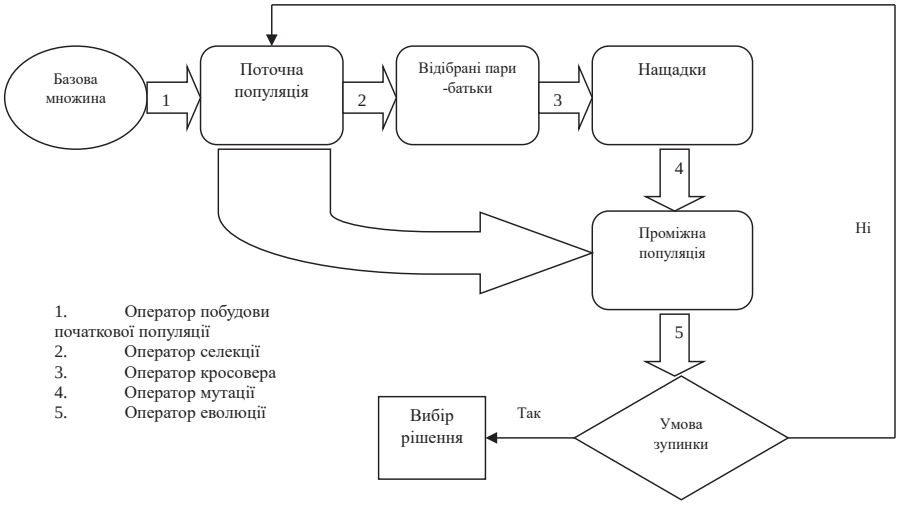


Рис. 1. Еволюційний алгоритм алгоритм

ку-нащадок. Потім цей елемент вилучається із двох послідовностей-батьків. Наприклад:

Cross ((4, 3, 1, 2, 8, 7, 6, 5), (3, 4, 2, 6, 1, 5, 8, 7)) = (3, 4, 1, 2, 8, 6, 5, 7).

Оператор мутації *Mut* із заданою імовірністю $\alpha \in (0,1)$ виконує випадкову транспозицію (заміну місцями двох елементів) у перестановці. Зауважимо, що існують також інші варіанти оператора мутації. Таким шляхом знаходиться множина $\tilde{Y}$ елементів-нащадків. До проміжної популяції $Y \cup \tilde{Y}$, що є об'єднанням поточної популяції та множини нащадків, застосовується оператор еволюції, який виділяє в цій популяції нову поточну популяцію. Процес еволюції повторюється доти, доки буде дотримано умову зупинки еволюційного алгоритму. Блок-схема еволюційного алгоритму наведена на рисунку 1.

Натепер існує безліч метаевристик у метричному просторі, які можна використовувати для пошуку субоптимальної перестановки. Серед них алгоритм перемішаних стрибаючих жаб, алгоритм імітації відпалу, алгоритм мурашиної колонії тощо.

Нехай оптимізаційна задача може бути представлена в термінах фрагментарної моделі. Якщо за цільову функцію обирати накриваючу функцію фрагментарної моделі цієї задачі, отримаємо гібридний алгоритм для пошуку субоптимальних розв'язків задачі. Для отримання значень цільової функції використовується фрагментарний алгоритм, який є індивідуальним для кожного класу дискретних оптимізаційних задач з орієнтованою фрагментарною структурою. Як приклад побудови фрагментарної моделі розглянемо один з різновидів задачі покриття множини точок на площині.

Фрагментарні моделі задачі покриття множини точок на площині.

Будемо розглядати задачу покриття множини точок у такій загальній постановці. Нехай задана кінцева множина точок площини $X = \{A_1, ..., A_n\} \subseteq R^2$ і набір типів фігур, які можуть виступати елементами покриття. У кожній такій фігурі виділена точка, яку будемо називати базовою. Розташування цієї точки на площині повністю визначає позицію конкретної фігури. Наприклад, для кіл заданого радіуса – це центри кіл. У множині X виділено підмножину $Y \subseteq X$ – точки, де можуть бути розташовані базові точки фігур – елементів покриття. Для кожної точки $y \in Y$ визначено кінцевий набір фігур $\Phi_y = \{\varphi_y^1, \varphi_y^2, ..., \varphi_y^{q_y}\}$, базова точка яких може бути розташована в точці y. Такі фігури на площині будемо називати допустимими. Звичайно, допустима фігура, окрім точки y, покриває і деякі інші точки множини X. Для кожної фігури A, базова точка якої розташована в точці y, визначена позитивна вага $\rho = \rho(A, y)$. Допустимим розв'язком задачі покриття будемо називати пару $U = (Z, A)$, що складається із множини попарно незбіжних точок $Z = \{y_1, y_2, ..., y_s\} \subseteq Y$ і набору допустимих фігур $A = \{A_1, A_2, ..., A_s\}$, базові точки яких розташовані у відповідних точках множини Z, усі точки множини X містяться в об'єднанні фігур $\bigcup_{i=1}^{s} A_i$. Вагою допустимого розв'язку U називається число $\rho(U) = \sum_{i=1}^{s} \rho(A_i, y_i)$. Задача оптимізації полягає у відшуканні допустимого покриття, вага якого мінімальна.

Побудуємо фрагментарну модель загальної задачі покриття. Визначимо розмір перестановки за формулою $N = \sum_{i=1}^{n} q_{y_i}$. Перенумеруємо послідовно всі пари $(y_i, \varphi_{y_i}^j)_{i=1,2,...,n}^{j=1,2,...,q_i}$ числами від 1 до N. Візьмемо будь-яку перестановку цих чисел і послідовність пар $(y_i, \varphi_{y_i}^j)$, що відповідає цій перестановці. Побудуємо допустимий розв'язок задачі покриття, що відповідає заданій перестановці. Для цього потрібно визначити фрагментарний алгоритм,

що по перестановці будує допустимий розв'язок задачі покриття. На початковому кроці алгоритму, на кожному кроці k фрагментарного алгоритму буде побудовано пару $U_k = (Z_k, A_k)$. Визначимо умову приєднання чергового елемента $(y_i, \varphi^j_{y_i})$ до пари U_r. Щоб побудувати наступну пару $U_{k+1} = (Z_{k+1}, A_{k+1})$ за правилом:

$$Z_{k+1} = Z_k \cup \{y_i\};\ A_{k+1} = A_k \cup \{\varphi^j_{y_i}\},$$

потрібно дотримання таких умов: 1) $y_i \notin Z_k$; 2) підмножина точок множини X, що міститься в об'єднанні фігур набору $A_k \cup \{\varphi^j_{y_i}\}$ хоча б на одну точку більше за підмножину точок множини X, що містяться в об'єднанні фігур набору A_k. Будь-який допустимий розв'язок задачі може бути досягнутий належним вибором перестановки із множини S_N.

Розглянемо як приклад задачу покриття заданої множини $X = \{A_1, ..., A_n\} \subseteq R^2$ точок площини колами радіуса R із центрами в цій же множині точок. Вага кожного кола дорівнює одиниці. Задача оптимізації полягає у знаходженні мінімального числа кіл із центрами в точках множини X, які містять усі точки множини X. Тобто базовою точкою є центр кола радіуса R.

Беремо перестановку чисел $i_1, i_2, ..., i_n$ від 1 до n. Будуємо множину кіл відповідно до умови: коло із центром в точці A_{i_k} додається до цієї множини, якщо воно містить хоча б одну точку множини X, яка не ввійшла в об'єднання вже обраних кіл. Алгоритм закінчує роботу тоді, коли всі точки множини X увійдуть в об'єднання обраних кіл. Вага розв'язку – це кількість кіл, які до нього ввійшли. Для пошуку субоптимального розв'язку можна використовувати будь-яку метаевристику на множині перестановок.

Висновки. У роботі було наведено принцип побудови гібридних алгоритмів для розв'язання задачі покриття множини точок на площині. Для пошуку субоптмальних розв'язків задачі покриття запропоновано використання гібридного алгоритму, який засновано на комбінації фрагментарного алгоритму й однієї зі стандартних метаевристик на множині перестановок. Запропонована методика побудови гібридних алгоритмів легко може бути розширена й для інших варіантів задач покриття точок.

ЛІТЕРАТУРА

1. Garey M. R., Johnson D. S. Computers and Intractability: A Guide to the Theory of NP-Completeness. San Francisco, CA : W. H. Freeman, 1979. 175 p.
2. Papadimitriou C. H., Steiglitz K. Combinatorial Optimization: Algorithms and Complexity. Prentice-Hall, 1982. 496 p.
3. Roth R. Computer solution to minimum-cover problems. Operat. Res., 1969. 17. № 3. P. 455–465.
4. Kantorovich L. V. Mathematical methods of organizing and planning production. *Management Science.* 1960. № 6 (4). P. 366–422.
5. Особливості розробки САПР розкрою матеріалів / Я. Грицишин та ін. IEEE AIS'02 CAD. 2002. С. 404–409.
6. Nthabiseng Ntene. An Algorithmic Approach to the 2D Oriented Strip Packing Problem. 2007. 217 p.
7. Гуляницький Л. Ф., Турінський В. В. Застосування метаевристичних алгоритмів для розв'язання одного класу задач розміщення прямокутників на напівнескінченній стрічці. *Analysis and Information Techns :* Proc. 15-th Int. Conf., May 27–31, 2013, Kyiv, Ukraine. SAIT. 2013.
8. Донець Г. П. Екстремальні задачі на перестановках зі спеціальною генерацією. *Комбінаторні конфігурації та їх застосування :* матеріали XI Міжвузівського науково-практичного семінару. Кіровоград, 2011. С. 41–48.
9. Graham R. L., Hell P. On the history of the minimum spanning tree problem. *Annals of the History of Computing.* 1985. Vol. 7. № 1. P. 43–57.
10. Yakovlev S. V. The method of artificial expansion of space in the problem of optimal placement of geometric objects. *Cybernetics and Systems Analysis.* 2017. № 53 (5). P. 825–832.
11. Кісельова О. М. Становлення та розвиток теорії оптимального розбиття множин. Теоретичні і практичні застосування. Дніпро : Ліра, 2018. 532 с.
12. Sean L. Essentials of Metaheuristics. Lulu. URL: http://cs.gmu.edu/~sean/book/metaheuristics/
13. Kozin I. V., Maksyshko N. K., Perepelitsa V. A. Fragmentary Structures in Discrete Optimization Problems, Cybernetics and Systems Analysis. November 2017. Vol. 53. Issue 6. P. 931–936. DOI: 10.1007/s10559-017-9995-6
14. Козін І. В., Нарзуллаєв У. Х., Алломов З. К. Алгоритм перемішаних стрибаючих жаб у задачі розміщення виробництва. *Computer Science and Applied Mathematics.* 2023. № 1. P. 11–17.
15. Kershner R. The number of circles covering a set. *Amer. J. Math.* 1939. Vol. 61. № 3. P. 665–671.
16. Energyefficient models for coverage problem in sensor networks with adjustable ranges / N. D. Nguen et al. *Ad Hoc & Sensor Wireless Networks.* 2012. Vol. 16. P. 1–28.
17. Toth F. G. Covering the plane with two kinds of circles. *Discrete Comput. Geometry.* 1995. Vol. 13. № 3. P. 445–457.

18. Wu J., Dai F. Virtual backbone construction in MANETs using adjustable transmission ranges. *IEEE Trans. Mobile Comput.* 2006. Vol. 5. № 9. P. 1188–1200.
19. Wu J., Yang S. Energy-efficient node scheduling models in sensor networks with adjustible ranges. *Int. J. Foundat. Comput. Sci.* 2005. Vol. 16. № 1. P. 3–17.
20. Energy-efficient area coverage by sensors with adjustable ranges / V. Zalyubovskiy et al. *Sensors.* 2009. Vol. 9. P. 2446–2460.
21. Козін І. В., Полюга С. І., Сардак В. І. Фрагментарна модель розміщення виробництва. *Математичне та комп'ютерне моделювання.* Серія «Фізико-математичні науки». Кам'янець-Подільський національний університет імені Івана Огієнка, 2019. Вип. 19. С. 35–40.

REFERENCES

1. Garey M. R. and Johnson D. S. (1979) Computers and Intractability: A Guide to the Theory of NP-Completeness. San Francisco: W. H. Freeman. San Francisco, CA.
2. Christos H. Papadimitriou, Kenneth Steiglitz. (1982) Combinatorial Optimization: Algorithms and Complexity. Mineola, New York: Prentice-Hall.
3. Roth R. (1969) Computer solution to minimum-cover problems. Operation Reserch. V. 17 (3), pp. 455–465.
4. Kantorovich L. V. (1960). Mathematical methods of organizing and planning production Management. Science LSU V6 (4). pp. 366–422.
5. Ghrycyshyn Ja., Lobur M., Tkachenko S., Chura I. (2002) Osoblyvosti rozrobky SAPR rozkroju materialiv. IEEE AIS'02 CAD. 2002. pp. 404–409.
6. Nthabiseng Ntene. (2007). An Algorithmic Approach to the 2D Oriented Strip Packing Problem. (PhD Thesis). Stellenbosch, South Africa: Department of Logistics, University of Stellenbosch.
7. Ghuljanycjkyj L. F., Turinsjkyj V. V. (2013). Zastosuvannja metaevrystychnykh alghorytmiv dlja rozv'jazannja odnogho klasu zadach rozmishhennja prjamokutnykiv na napivneskinchennij strichci. Analysis and Information Techns: Proc. 15-th Int. Conf., May 27–31, 2013, Kyiv, Ukraine. SAIT.
8. Donecj Gh. P. Ekstremaljni zadachi na perestanovkakh zi specialjnoju gheneracijeju. (2011). Proceedings of the Kombinatorni konfighuraciji ta jikh zastosuvannja, Ukraine, Kirovoghrad, April 15–16, 2011, pp. 41–48.
9. Graham R. L., Hell P. (1985). On the history of the minimum spanning tree problem. IEEE Annals of the History of Computing. Vol. 7 (1)), pp. 43–57.
10. Yakovlev S. V. (2017) The method of artificial expansion of space in the problem of optimal placement of geometric objects. Cybernetics and Systems Analysis. V. 53 (5). New York: Springer, pp. 825–832.
11. Kiseljova O. M. (2018). Stanovlennja ta rozvytok teoriji optymaljnogho rozbyttja mnozhyn. Teoretychni i praktychni zastosuvannja. Dnipro: Lira (in Ukranian).
12. Sean L. (2013). Essentials of Metaheuristics, Lulu, second edition, available for free at http://cs.gmu.edu/~sean/book/metaheuristics/.
13. Kozin I. V., Maksyshko N. K., Perepelitsa V. A. (2017). Fragmentary Structures in Discrete Optimization Problems. Cybernetics and Systems Analysis. V. 53(6). New York: Springer, pp. 931–936.
14. Kozin I. V., Narzullajev U. Kh., Allomov Z. K. (2023). Alghorytm peremishanykh strybajuchykh zhab u zadachi rozmishhennja vyrobnyctva. Computer Science and Applied Mathematics, V.1, Samarkand, pp.11–17.
15. Kershner R. (1939). The number of circles covering a set Amer. J. Math. Vol. 61 (3). New York, pp. 665–671.
16. Nguen N. D., Zalyubovskiy V., Minh T. H., Hyunseung C. (2012). Energyefficient models for coverage problem in sensor networks with adjustable ranges Ad Hoc & Sensor Wireless Networks. V.16, pp. 1–28.
17. Toth F. G. (1995). Covering the plane with two kinds of circles. Discrete Comput. Geometry. V. 13 (3). New York: Springer, pp. 445–457.
18. Wu J., Dai F. (2006). Virtual backbone construction in MANETs using adjustable transmission ranges. IEEE Trans. Mobile Comput. V.5 (9). pp. 1188–1200.
19. Wu J., Yang S. (2005). Energy-efficient node scheduling models in sensor networks with adjustible ranges. Int. J. Foundat. Comput. Sci. V. 6 (1). pp. 3–17.
20. Zalyubovskiy V., Erzin A., Astrakov S., Choo H. (2009). Energy-efficient area coverage by sensors with adjustable ranges. Sensors. V. 9 (4), pp. 2446–2460.
21. Kozin I. V., Poljugha S. I., Sardak V. I. (2019). Fraghmentarna modelj rozmishhennja vyrobnyctva. Sb. Matematychne ta komp'juterne modeljuvannja Serija Fizyko-matematychni nauky. Kam'janecj-Podiljsjkyj nacionaljnyj universytet im. Ivana Oghijenka, V.19, pp. 35–40.

Дата першого надходження рукопису до видання: 29.08.2025
Дата прийнятого до друку рукопису після рецензування: 26.09.2025
Дата публікації: 31.12.2025

УДК 514.177.2
DOI https://doi.org/10.26661/2786-6254-2025-2-02

АНАЛІТИЧНЕ ДОСЛІДЖЕННЯ ЗАДАЧІ ФЕРМА – ТОРРІЧЕЛЛІ ДЛЯ ЧОТИРИКУТНИКІВ

Максимов І. І.
кандидат технічних наук, доцент,
завідувач кафедри вищої математики та фізики
Криворізький національний університет
вул. Віталія Матусевича, 11, Кривий Ріг, Україна
orcid.org/0000-0003-2999-4806
maksimov@knu.edu.ua

Кіяновська Н. М.
кандидат педагогічних наук, доцент,
доцент кафедри вищої математики та фізики
Криворізький національний університет
вул. Віталія Матусевича, 11, Кривий Ріг, Україна
orcid.org/0000-0002-0108-5793
kiyanovskaya.n.m@knu.edu.ua

Слободянюк В. К.
кандидат технічних наук, доцент,
доцент кафедри відкритих гірничих робіт
Криворізький національний університет
вул. Віталія Матусевича, 11, Кривий Ріг, Україна
orcid.org/0000-0002-9977-3972
slobod_v@knu.edu.ua

Максимова І. І.
доктор економічних наук, доцент,
завідувач кафедри міжнародних відносин
Державний університет економіки і технологій
вул. Медична, 16, Кривий Ріг, Україна
orcid.org/0000-0001-9754-0414
maksimova_ii@kneu.dp.ua

Ключові слова: *точка Ферма – Торрічеллі, чотирикутник, рівнозважені точки, різнозважені точки, лінії рівня.*

Однією із задач оптимізації в геометрії є пошук оптимальної точки Ферма – Торрічеллі, що забезпечує знаходження мінімальної суми відстаней до вершин заданого многокутника. Але, незважаючи на майже чотирьохсотлітню історію задачі, натепер знайдено універсальні геометричні розв'язки лише для випадків рівнозважених вершин трикутників і чотирикутників. Це істотно обмежує її застосовність у практичних сферах. Водночас дана задача має велике практичне значення для широкого спектра прикладних завдань. Вона є фундаментальною для оптимального розміщення критично важливих об'єктів, як-от переробні підприємства, великі склади, а також для значної оптимізації транспортних і логістичних схем тощо. Але в перелічених вище завданнях найчастіше розглядають умову, коли задані вихідні точки многокутників

зазвичай мають різну значущість, або «вагу», що робить їх різнозваженими. Однак саме для різнозважених вершин досі відсутні навіть базові геометричні методи розв'язування, навіть для найпростіших випадків, що мають форму найпростіших фігур – трикутників і чотирикутників.

Складаючи математичну модель задачі для пошуку координат оптимальної точки, ми приходимо до системи ірраціональних рівнянь та неявних функцій високого порядку, що ускладнює аналітичне дослідження. Задача значно спрощується для випадків із симетричними чотирикутниками (зокрема, із квадратом і ромбом), де симетрія дозволяє застосовувати більш прямі аналітичні підходи.

У процесі даного дослідження нами було встановлено координати оптимальної точки Ферма – Торрічеллі як для різнозважених вершин чотирикутників, проведено аналіз впливу параметрів ромба й вагового коефіцієнта на положення оптимальної точки. Окрім того, встановлені умови, за яких вершина ромба може бути оптимальною точкою. Досліджені особливості зміни рівнозваженої і різнозваженої сум відстаней до вершин чотирикутника. Досліджений вплив вагового коефіцієнта на положення оптимальної точки.

Проведено дослідження форми ліній рівня (для випадків чотирикутників представлені чотирицентровими еліпсами), що надало можливість визначити вплив відхилення від оптимальної точки на збільшення основних показників.

ANALYTICAL STUDY OF THE FERMAT – TORRICELLI PROBLEM FOR QUADRANGLES

Maksymov I. I.
PhD (Engineering), Associate Professor,
Head of the Department of Higher Mathematics and Physics
Kryvyi Rih National University
Vitaly Matusevich str., 11, Kryvyi Rih, Ukraine
orcid.org/0000-0003-2999-4806
maksimov@knu.edu.ua

Kiianovska N. M.
PhD (Pedagogical), Associate Professor,
Associate Professor at the Department of Higher Mathematics and Physics
Kryvyi Rih National University
Vitaly Matusevich str., 11, Kryvyi Rih, Ukraine
orcid.org/0000-0002-0108-5793
kiyanovskaya.n.m@knu.edu.ua

Slobodyanyuk V. K.
PhD (Engineering), Associate Professor,
Associate Professor at the Open-Cast Mining Department
Kryvyi Rih National University
Vitaly Matusevich str., 11, Kryvyi Rih, Ukraine
orcid.org/0000-0002-9977-3972
slobod_v@knu.edu.ua

Maksymova I. I.
PhD (Economics), Professor,
Head of the Department of International Economics
State University of Economics and Technology
Medychna str., 16, Kryvyi Rih, Ukraine
orcid.org/0000-0001-9754-0414
maksimova_ii@kneu.dp.ua

Key words: *Fermat – Torricelli point, quadrilateral, equiweighted and unequally weighted points, contour lines.*

One fundamental optimization problem in geometry is determining the Fermat – Torricelli point, which minimizes the sum of distances to the vertices of a given polygon. Despite the problem's nearly four-century history, universal geometric solutions currently exist only for cases involving equi-weighted vertices of triangles and quadrilaterals. This significantly limits its practical applicability. Nevertheless, this problem holds substantial practical significance across a broad spectrum of applied problems. It is fundamental for the optimal placement of critical facilities such as processing plants and large warehouses, and for considerably optimizing transportation and logistics schemes. However, in the aforementioned applications, the given initial points (vertices) of the polygons typically possess varying significance or "weights", rendering them unequally-weighted. Crucially, for these unequally-weighted vertices, even basic geometric solution methods remain absent, even for the simplest polygonal configurations – namely, triangles and quadrilaterals.

When constructing a mathematical model for determining the coordinates of the optimal point, we arrive at a system of irrational equations and high-order implicit functions, which complicates analytical investigation. The problem is significantly simplified for cases involving symmetric quadrilaterals (specifically squares and rhombuses), where symmetry allows for more direct analytical approaches.

In the course of this study, we determined the coordinates of the optimal Fermat – Torricelli point for variably weighted vertices of quadrilaterals. An analysis was conducted to assess the influence of rhombus parameters and the weighting coefficient on the location of the optimal point. Furthermore, we identified the conditions under which a vertex of the rhombus may serve as the optimal point. Specific features of changes in the equally weighted and variably weighted sums of distances to the quadrilateral vertices were examined, as well as the impact of the weighting coefficient on the position of the optimal point.

A study was conducted on the shape of contour lines (represented by four-centered ellipses in the case of quadrilaterals), which enabled the assessment of the impact of deviations from the optimal point on the increase of key performance indicators.

Вступ. ***Проблема та її зв'язок з науковими та практичними завданнями.*** Пошук оптимальної точки Ферма – Торрічеллі має велике практичне значення і застосовується в багатьох оптимізаційних задачах. Найвідоміші з них – це задача оптимального розміщення переробного підприємства й оптимізація логістичних схем. Але наявні геометричні методи можна застосовувати лише для рівнозважених вершин трикутника та чотирикутника. Важливим науковим завданням є знаходження координат точки Ферма – Торрічеллі як для рівнозважених, так і для різнозважених точок, дослідження впливу вихідних даних на кінцевий результат. **Огляд літератури.** Класична задача Ферма полягала у знаходженні для трьох заданих точок на евклідовій площині такої точки, сума відстаней до якої від заданих трьох точок була мінімальною. Задача розглядалася для рівнозважених точок трикутника. Під *рівнозваженими* точками $\{A_1, A_2, A_3, \ldots, A_n\}$, кожній з яких відповідає задане число $\{k_1, k_2, k_3, \ldots, k_n\}$, будемо розуміти такі точки, усі коефіцієнти яких рівні, тобто $k_1 = k_2 = k_3 = \ldots = k_n = k$. Точки будемо вважати *різнозваженими*, якщо коефіцієнти $\{k_1, k_2, k_3, \ldots, k_n\}$ різні.

Першим задачу Ферма розв'язав Е. Торрічеллі. Оптимальною буде така точка Т, з якої сторони трикутника видно під кутом 120° (ні доведення, ні метод не збереглися). Розв'язок задачі Ферма називають точкою Ферма – Торрічеллі. У XVII–XVIII ст. було знайдено кілька геометричних методів знаходження оптимальної точки, але тільки для рівнозважених вершин трикутника.

Пізніше було розглянуто розширення задачі Ферма для чотирьох неколінеарних точок у $\mathbb{R}^2$. Уперше оптимальну точку Ферма – Торрічеллі для чотирикутника знайшов італійський математик Джованні Фаньяно деї Тоски (1682–1766 рр.). Для опуклого чотирикутника оптимальною буде точка перетину діагоналей (рис. 1). Якщо чотири точки утворюють неопуклий чотирикутник, тоді оптимальною точкою Ферма буде його вершина, для якої внутрішній кут більше за 180° (Кавальєрі). Цей результат було показано за допомогою нерівностей трикутників. Якщо координати опуклого чотирикутника відомі, то координати оптимальної точки можна знайти як точку перетину двох прямих (діагоналей) [9; 10].

$$x_T = \frac{(y_3 \cdot x_1 - y_1 \cdot x_3) \cdot (x_4 - x_2) - (y_4 \cdot x_2 - y_2 \cdot x_4) \cdot (x_3 - x_1)}{(y_3 - y_1) \cdot (x_4 - x_2) - (y_4 - y_2) \cdot (x_3 - x_1)};$$

$$y_T = \frac{(x_3 \cdot y_1 - x_1 \cdot y_3) \cdot (y_4 - y_2) - (x_4 \cdot y_2 - x_2 \cdot y_4) \cdot (y_3 - y_1)}{(x_3 - x_1) \cdot (y_4 - y_2) - (x_4 - x_2) \cdot (y_3 - y_1)}.$$

Але наведені геометричний і алгебраїчний методи не дозволяють провести дослідження впливу

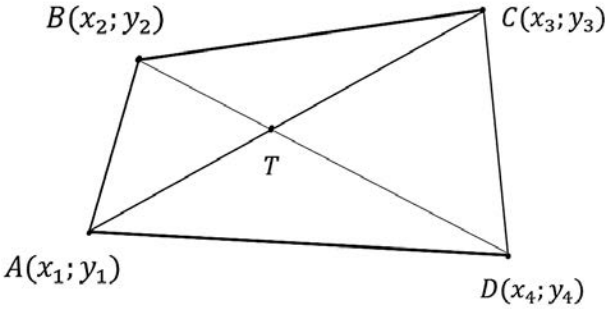


Рис. 1. Схема для визначення координат $T(x_T; y_T)$ оптимальної точки для чотирикутника

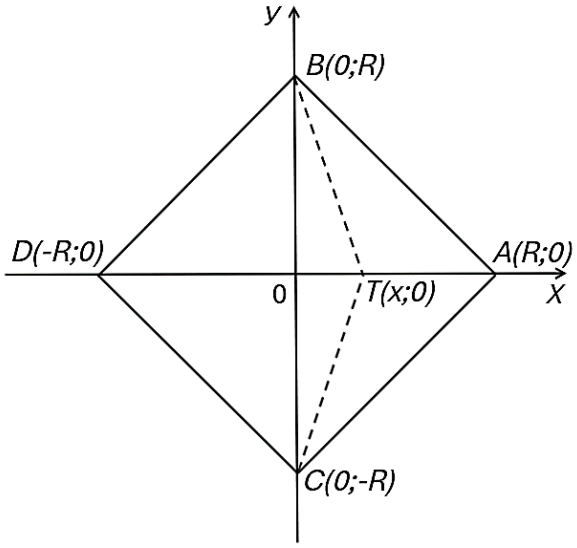


Рис. 2. Схема для визначення координат оптимальної точки Ферма – Торрічеллі для квадрата

елементів чотирикутника на положення оптимальної точки. Аналогічний недолік мають аналітичні та градієнтні методи знаходження координат оптимальної точки [1–8]. Окрім того, їх можна застосувати тільки для рівнозважених точок.

Застосування аналітичних методів приводить до появи функцій з декількома радикалами (відстані), або неявних функцій із двома невідомими в R^2 восьмого (для трикутників) та шістнадцятого (для чотирикутників) порядку, що ускладнює їх дослідження. Під час дослідження симетричних фігур одна з координат оптимальної точки відома, дослідження значно спрощується. У роботі [12] авторами наведено дослідження положення оптимальної точки й особливості зміни особливих показників для рівнобедрених трикутників (як для рівнозважених, так і для нерівнозважених вершин).

Метою статті є визначення координат оптимальної точки для різнозважених вершин чотирикутника. Проведення дослідження впливу вихідних параметрів на положення оптимальної точки на прикладі деяких видів чотирикутників.

Об'єктом дослідження у статті є оптимальна точка Ферма – Торрічеллі для чотирикутників з рівнозваженими і різнозваженими вершинами.

Методи. У процесі проведеного дослідження було використано комбінацію теоретичних методів (аналіз, синтез, узагальнення, математичне моделювання), математичних методів (аналітичні – для спрощених випадків, чисельні – для складних систем рівнянь, оптимізаційні – для пошуку мінімуму), комп'ютерні методів (для симуляції, візуалізації ліній рівня).

Результати. Сформулюємо зважену задачу Ферма – Торрічеллі для чотирьох неколінеарних точок у R^2: дано чотири неколінеарні точки A_1, A_2, A_3, A_4 та додатні дійсні числа (вага) k_i ($i = 1, 2, 3, 4$), що відповідають кожній точці A_i, знайти таку п'яту точку T, щоб сума зважених відстаней до цих чотирьох точок була мінімальною.

Розглянемо спочатку квадрат (рис. 2) та дослідимо закономірності зміни деяких показників. Оптимальною буде точка перетину діагоналей, де і розміщуємо початок системи координат. Сторона квадрата дорівнює a, а радіус описаного кола

$$R = \frac{a\sqrt{2}}{2}\left(a = R \cdot \sqrt{2}\right).$$

Для рівнозважених точок мінімальна сума відстаней до вершин квадрата дорівнює $4R$. Розглянемо випадок, коли вершини квадрата нерівнозважені. Ваговий коефіцієнт вершини A дорівнює k ($k > 1$), а для інших вершин дорівнює одиниці ($k = 1$). Це відповідає задачі оптимального розміщення переробного підприємства, на яке з вершини A надходять обсяги сировини, що перевищують обсяги з інших точок.

Цільова функція має вигляд:

$$F(x) = (R + x) + k(R - x) + 2\sqrt{x^2 + R^2} \cdot min\,. \qquad (1)$$

Прирівнявши похідну до нуля, знаходимо координати оптимальної точки:

$$x_0 = R \cdot \frac{k-1}{\sqrt{4-(k-1)^2}}\,. \qquad (2)$$

Якщо $k = 1$, тоді $x_0 = 0$ точка перетину діагоналей.

За $k > 1$ оптимальна точка зміщується у сторону вершини A. Наприклад, за $k = 2$ $x_0 = \frac{R\sqrt{3}}{3} \approx 0{,}577 \cdot R$.

Особливе значення в дослідженнях має випадок, коли вершина A стає оптимальною точкою. Підставляємо в рівняння (2) $x_0 = R$ і знаходимо $k_A = 1 + \sqrt{2} \approx 2{,}414$. Для порівняння, для правильного трикутника вершина A стає оптимальною точкою за $k_A = \sqrt{3} \approx 1{,}732$ [10; 11; 12]. В обох випад-

ках об'єм сировини з вершини A не перевищує сумарного об'єму сировини з інших точок. Для трикутника $\sqrt{3}<2$, а $\frac{\sqrt{3}}{2}\approx 0{,}866$ (86,6%), а для квадрата $(1+\sqrt{2})<3$, а $(1+\sqrt{2})/3\approx 0{,}805$ (80,5%).

Розглянемо більш загальний випадок, коли чотири точки утворюють ромб. Вершина A віддаляється від початку координат (рис. 3).

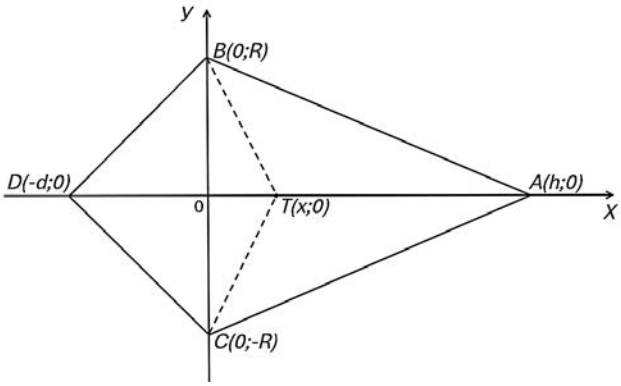


Рис. 3. Схема для визначення координат оптимальної точки Ферма – Торрічеллі для ромба

Вершини A і D розташовані на відстанях відповідно h і d від початку координат. Якщо вершини ромба рівнозважені, то положення оптимальної точки не залежить від віддалення чи наближення вершин A і D. Якщо вершина A має ваговий коефіцієнт k, то координати оптимальної точки визначаємо з мінімуму функції:

$$F(x)=(d+x)+k(h-x)+2\sqrt{x^2+R^2}\,. \qquad (3)$$

Прирівнявши похідну до нуля, знаходимо координати оптимальної точки:

$$x_0=R\cdot\frac{k-1}{\sqrt{4-(k-1)^2}}\,. \qquad (4)$$

Рівняння (2) і (4) ідентичні, розміщення оптимальної точки залежить тільки від довжини діагоналі $R=\frac{BC}{2}$ та вагового коефіцієнта k (не залежить від віддалення чи наближення вершин A і D).

Підставляємо в рівняння (4) $x_0=h$ і знаходимо значення вагового коефіцієнта, за якого вершина A стає оптимальною точкою.

$$k_A=1+\frac{2h}{\sqrt{h^2+R^2}}\,. \qquad (5)$$

Значення коефіцієнта k_A не залежить від положення вершини D, а тільки від величини діагоналі $R=\frac{CB}{2}$ та віддалення вершини A від початку координат h (рис. 4). За віддалення вершини A ($h\to\infty$) значення коефіцієнта k_A збільшується і наближається до 3. За $h=4R$, $k_A=2{,}94$.

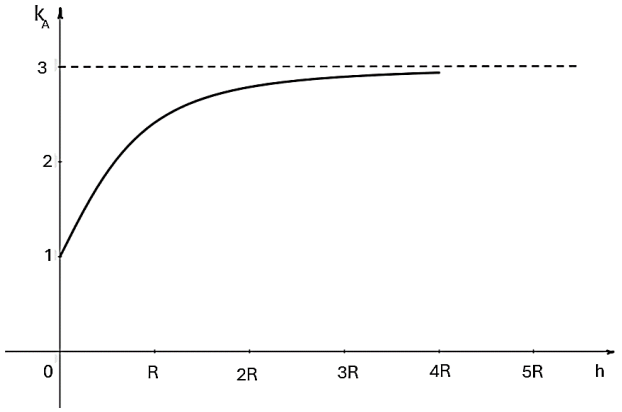


Рис. 4. Зміна значення вагового коефіцієнта k_A за віддалення вершини A від точки перетину діагоналей

Якщо відстань від вершини A до початку координат менше R, то вона може бути оптимальною за $k_A<1+\sqrt{2}$. За $h\to 0$, $k_A\to 1$.

Для нерівнозважених вершин положення оптимальної точки знаходять наближено за допомогою фізичної моделі, де використовують ваги пропорційні ваговим коефіцієнтам [6–8]. Встановлено, що вершина може бути оптимальною точкою, якщо $k_A\geq 2$ для трикутників і за $k\geq 3$ для чотирикутників. Проведеними дослідженнями показано, що це максимальні значення. Для правильних трикутників досить, щоб $k_A\geq\sqrt{3}\approx 1{,}73$, а для квадрата $k_A\geq 1+\sqrt{2}\approx 2{,}41$. У кожному конкретному випадку значення коефіцієнта k_A можна знайти за формулою (5).

Коли змінюються вагові коефіцієнти для двох протилежних точок (k_1 для точки A і k_2 для точки D), знаходимо схожі формули:

$$x_0=R\cdot\frac{k_1-k_2}{\sqrt{4-(k_1-k_2)^2}}\,. \qquad (6)$$

Якщо $k_1>k_2$, то оптимальна точка x_0 розташована справа від початку координат, а якщо $k_1<k_2$, то зліва.

Вершина A стає оптимальною, якщо

$$k_A=k_2+\frac{2h}{\sqrt{h^2+R^2}}\,, \qquad (7)$$

за $h=Rk_A=k_2+\frac{\sqrt{2}}{2}$, а за $h\to\infty\, k_A\to k_2+2$.

Характер функції $F\ (x)$ значно відрізняється для рівнозважених і різнозважених вершин ромба. Значною мірою він залежить від вагового коефіцієнта k. Розглянемо випадок, коли $h=2R; d=R$.

За всіх значеннь вагового коефіцієнта k значення функції $F\ (x)$ за $x=2R$ (вершина A) збігається і дорівнює $\left(3+2\sqrt{5}\right)\cdot \mathrm{R}$.

На рисунку 5 показано залежність положення оптимальної точки Ферма – Торрічеллі від значення вагового коефіцієнта *k*: у разі збільшення коефіцієнта *k* від 0 до $k_A = 1 + \frac{4}{\sqrt{5}}$ положення оптимальної точки Ферма – Торрічеллі зміщується із крайнього лівого положення $x_0 = -R \cdot \sqrt{3}/3$ до крайнього правого $x_0 = 2R$. Зокрема, $x_0 = -R \cdot \sqrt{3}/3$ за $k = 0$; $x_0 = 0$ за $k = 1$; $x_0 = R \cdot \sqrt{3}/3$ за $k = 2$ і $x_0 = 2R$ за $k = k_A = 1 + \frac{4}{\sqrt{5}}$.

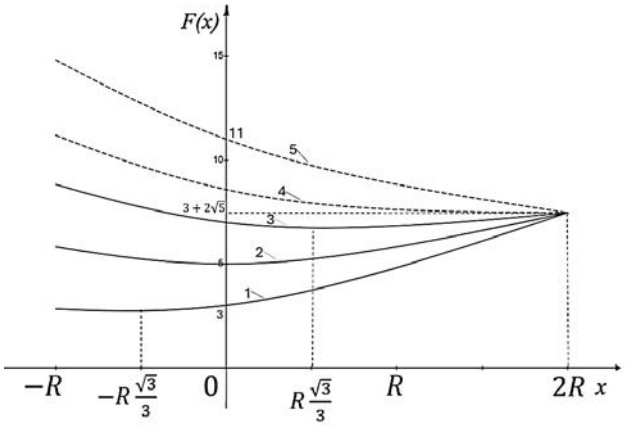


Рис. 5. Зміна функції *F (x)* за різних вагових коефіцієнтів *k*:
1 – (*k* = 0), 2 – (*k* = 1), 3 – (*k* = 2),
4 – (*k* = $k_A = 1 + \frac{4}{\sqrt{5}}$), 5 – (*k* = 4 > k_A)

Якщо $k \geq k_A = 3 + 2\sqrt{5}$, функція спадна й має мінімальне значення: $F_{min} = 3 + 2\sqrt{5}$ за $x_0 = 2R$. Якщо $k < 1$, оптимальна точка розташована лівіше точки перетину діагоналей, а якщо $k > 1$ – правіше.

Аналіз графіків на рисунку 5 показує, що в деякому околі оптимальної точки x_0 функція *F(x)* має незначний приріст. Іноді переробне підприємство або великий склад неможливо розмістити в оптимальній точці, а тільки на деякій відстані від неї. Це можуть бути гори, море, заповідна зона, значне порушення екологічних вимог та інші. У такому разі необхідно дослідити, як відхилення від оптимальної точки на деяку величину *l* впливає на приріст функції *F (x)*. Швидкість приросту функції в різних напрямках від оптимальної точки може бути різною. Тому найточнішим способом буде побудова кількох ліній рівня функції *F (x)*. Наприклад, брати збільшення значення функції на 2, 4, 6% тощо. Будувати лінії рівня, що відповідають значенням функції $F_{min} \cdot 1{,}02$;. $F_{min} \cdot 1{,}04$; $F_{min} \cdot 1{,}06$ тощо. Таким чином можна знайти розміщення переробного підприємства, що відповідає мінімально можливому значенню функції з урахуванням необхідних обмежень і вимог щодо його розміщення.

Розглянемо властивості ліній рівня для рівнозважених вершин квадрата (рис. 6).

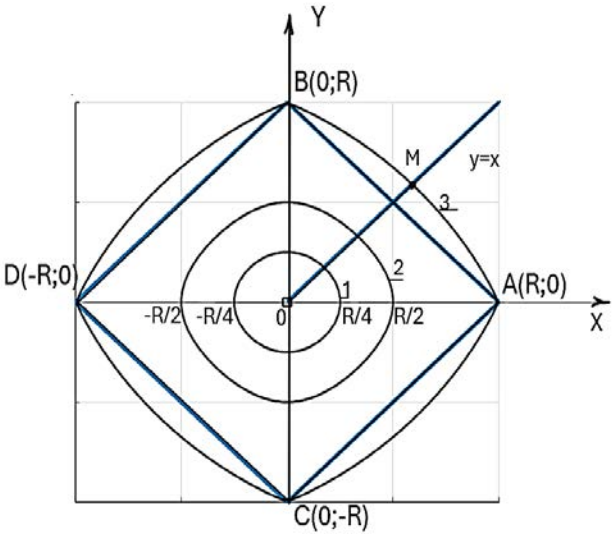


Рис. 6. Лінії рівня для рівнозважених вершин квадрата, що відповідають збільшенню суми відстаней до вершин

(1) – $F_{min} \cdot \frac{\left(4 + \sqrt{17}\right)}{8}$**; (2) –** $F_{min} \cdot \frac{\left(2 + \sqrt{5}\right)}{4}$**;**

(3) – $F_{min} \cdot \frac{\left(1 + \sqrt{2}\right)}{2}$

Для початку координат сума відстаней до вершин квадрата буде мінімальною і дорівнює $F_{min} = 4R$. Для лінії рівня (1) сума відстаней перевищує мінімальне значення на 1,54%, для лінії рівня (2) – на 5,9%, а для лінії рівня (3) – на 20,71%. За таких перевищень суми відстаней лінії рівня перетинають координатні осі *OX* та *OY* на відстані, від початку координат, відповідно $\frac{R}{4}; \frac{R}{2}$ і R. Форма ліній рівня відрізняється від кола. Якнайдалі від початку координат розміщені точки перетину ліній рівня з координатними осями, а найближче – точки перетину із прямими $y = \pm x, (\varphi = \pm 45°)$. Найбільше ця різниця буде для лінії (3). $OA = R$, а $OM = \frac{3 + \sqrt{2}}{2\sqrt{7}} \cdot R \approx 0.834 \cdot R$, на 16,6% менше. Для лінії (2) різниця відстаней становить 6,8%, а для лінії (1) – лише 1,5%. У разі зменшення відносного приросту функції *F (x)* форма лінії рівня наближається до кола.

У разі віддалення точки від оптимальної вздовж координатних осей на величину *l* відносну суму відстаней до вершин квадрата (F / F_{min}) (рис. 7) знаходимо за формулою:

$$\frac{F}{F_{min}} = \frac{1}{2}\left(1 + \sqrt{1 + \left(\frac{l}{R}\right)^2}\right), \left(\frac{l}{R}\right) \epsilon \, [0;1]. \qquad (8)$$

Оскільки форма ліній рівня близька до кола, особливо за $x < R/2$, то формулу (8) можна використовувати як наближену для довільного напрямку відхилення від оптимальної точки. За формулою (8) знаходимо:

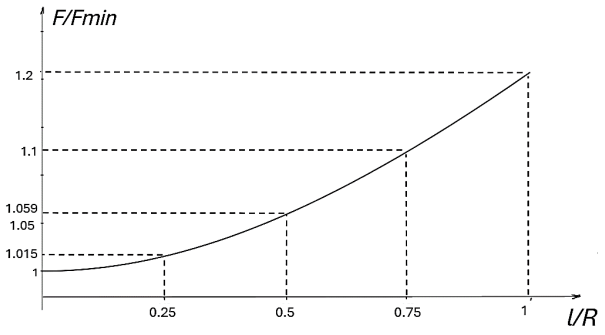


Рис. 7. Зростання відносної суми відстаней (F / F_{min}) у разі збільшення відстані від оптимальної точки вздовж координатних осей

$$l = 2 \cdot R\sqrt{\left(F / F_{min}\right)^2 - \left(F / F_{min}\right)}\,. \qquad (9)$$

Якщо відоме допустиме збільшення значення відносної величини суми відстаней до вершин квадрата на величину F / F_{min}, то допустиме відхилення від оптимальної точки в межах кола, радіус якого знаходимо за формулою (9).

На рисунку 8 наведені приклади ліній рівня для рівнозважених і різнозважених вершин квадрата й ромба $(h = 2R)$.

З рисунку 8 видно значення впливу на форму ліній рівня (чотирифокусні еліпси) віддалення вершини A $(h = 2R)$ та випадків, коли вершина A буде оптимальною точкою $k = k_A$. Наведені приклади показують необхідність подальших досліджень властивостей ліній рівня. Формули (8) і (9) можна використовувати як оціночні.

Висновки. Запропонований метод дозволяє провести дослідження тільки для симетричних фігур, на прикладі трикутника [12] та чотирикутника. Але це єдиний поки що метод, що надає можливість розглянути різнозважені точки та вплив вихідних даних на зміну основних показників (рівнозваженої і різнозваженої суми).

Проведеними дослідженнями встановлені координати оптимальної точки Ферма – Торрічеллі для квадрата й ромба для рівнозважених

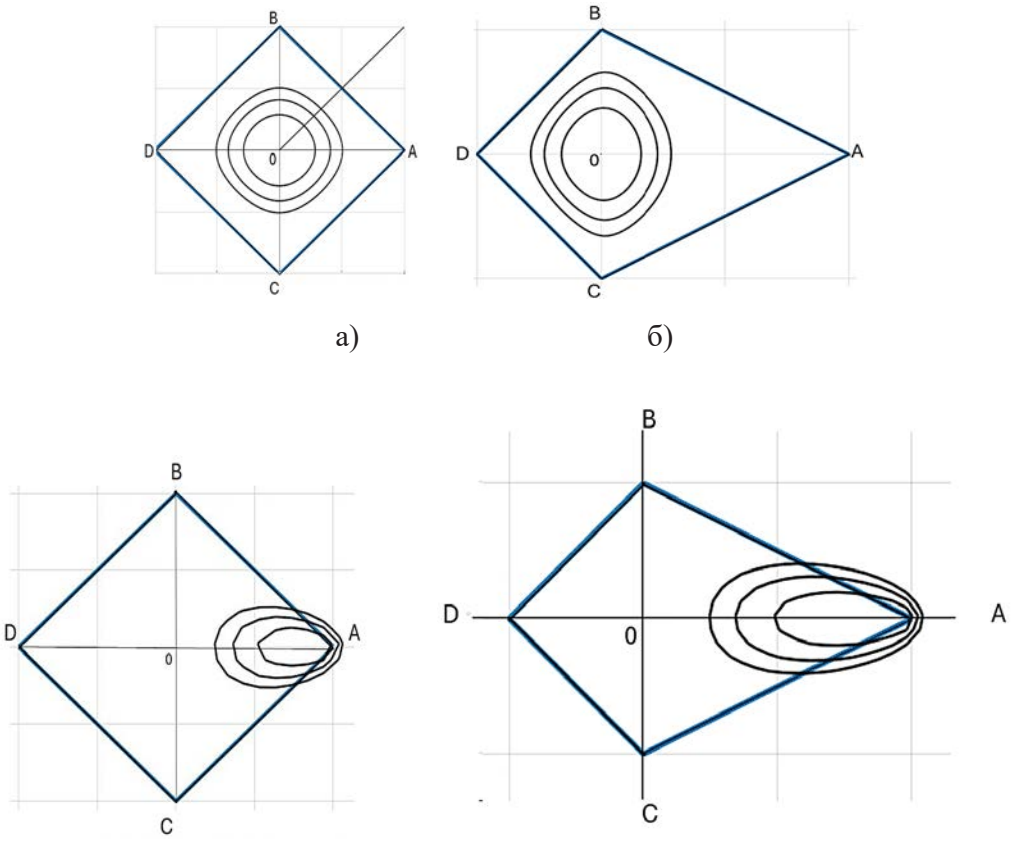


Рис. 8. Лінії рівня, що відповідають значенням функції $F_{min} \cdot 1{,}02$;. $F_{min} \cdot 1{,}04$; $F_{min} \cdot 1{,}06$. а) і б) рівнозважені точки; в) і г) різнозважені точки, де $k = k_A$, оптимальною точкою буде вершина A

і різнозважених вершин. Встановлені умови, за яких вершина чотирикутника може бути оптимальною точкою. Встановлені основні параметри, що впливають на положення оптимальної точки. Встановлений вплив величини відхилення від оптимальної точки на зростання транспортної роботи (суми відстаней до вершин квадрата).

Надалі планується продовжити дослідження для різних видів многокутників.

ЛІТЕРАТУРА

1. Kupitz Ya. S., Martini H., Spirova M. The Fermat – Torricelli Problem, Part I: A Discrete Gradient-Method Approach. *Journal of Optimization Theory and Applications*. 2013. Vol. 158. № 2. P. 305–327.
2. Villiers M., Meyer J. Another concurrency related to the Fermat point of a triangle. *Mathematics Competitions.* 2024. Vol. 37. № 2. P. 33–40.
3. Hajja M., Zachos A. A complete analytical treatment of the weighted Fermat–Torricelli point for a triangle. *Journal of Geometry*. 2017. Vol. 108. № 1. P. 99–110.
4. Zachos A. Analytical solution of the weighted Fermat–Torricelli problem for convex quadrilaterals in the Euclidean plane: The case of two pairs of equal weights. *Mathematics. Optimization and Control*. 2014. https://doi.org/10.48550/arXiv.1406.2947
5. Courant R., Robbins H., Stewart I. What Is Mathematics? An Elementary Approach to Ideas and Methods 2nd ed. *Oxford University Press*. 1996. 588 p.
6. Xu Bin, Chen Hongsheng, Cheng Yanxu. A network recovery strategy based on boundary nodes and tetrahedral approximation Fermat points in three-dimensional wireless sensor networks. *Scientific Reports*. 2025. № 15. DOI: 10.1038/s41598-025-15723-0
7. Liébana Rísquez E. Otros puntos notables de triángulos. TFM. Universidad de Granada. 2016.
8. Wang Huan, Gao Bingtuan, Zhao Dehui, Shen Huan, Liu Chuande. A grasping configuration planning method inspired by the geometric Fermat point. *Industrial Robot: the international journal of robotics research and application*. 2025. DOI: 10.1108/IR-12-2024-0565
9. Максимова І. І., Слободянюк Р. В. Особливості визначення раціонального положення перевантажувального пункту. *Перспективи розвитку будівельних технологій :* матеріали 11-ї Міжнародної науково-практичної конференції. 2017. Дніпропетровськ, 217. С. 43–47.
10. Maksymov I., Slobodyanyuk V., Maksymova I. Features of determining the coordinates of the Fermat – Torricelli point for isosceles triangles. *Internauka :* International scientific journal. 2023. № 8 (142). https://doi.org/10.25313/2520-2057-2023-8-8855.
11. Максимов І. І., Слободянюк В. К., Максимова І. І. Аналітичне визначення координат точки Ферма – Торрічеллі та мінімізація транспортної роботи. *Вчені записки Таврійського національного університету імені В. І. Вернадського.* Серія «Технічні науки». 2023. 34 (73). № 3. С. 1–7.
12. Максимов І. І., Кіяновська Н. М., Слободянюк В. К., Максимова І. І. Аналітичне дослідження задачі Ферма для рівнобедреного трикутника. *Наука і техніка сьогодні*. 2024. № 5 (33). https://doi.org/10.52058/2786-6025-2024-5(33)-1348-1362

REFERENCES

1. Kupitz, Ya. S., Martini, H., & Spirova M. (2013). The Fermat–Torricelli Problem, Part I: A Discrete Gradient-Method Approach. *Journal of Optimization Theory and Applications*. Vol. 158, № 2. P. 305–327.
2. Villiers, M., & Meyer, J. (2024). Another concurrency related to the Fermat point of a triangle. *Mathematics Competitions,* Vol. 37, № 2. P. 33–40.
3. Hajja, M., & Zachos, A. (2017). A complete analytical treatment of the weighted Fermat–Torricelli point for a triangle. *Journal of Geometry*. Vol. 108, № 1. P. 99–110.
4. Zachos, A. (2014). Analytical solution of the weighted Fermat–Torricelli problem for convex quadrilaterals in the Euclidean plane: The case of two pairs of equal weights. *Mathematics. Optimization and Control*. https://doi.org/10.48550/arXiv.1406.2947
5. Courant, R., Robbins, H., Stewart, I. (1996). What Is Mathematics? An Elementary Approach to Ideas and Methods 2nd ed. *Oxford University Press*. 588 p.
6. Xu, Bin, Chen, Hongsheng, & Cheng, Yanxu. (2025). A network recovery strategy based on boundary nodes and tetrahedral approximation Fermat points in three-dimensional wireless sensor networks. *Scientific Reports*. 15. DOI: 10.1038/s41598-025-15723-0
7. Liébana Rísquez E. (2016). Otros puntos notables de triángulos. TFM. *Universidad de Granada*.
8. Wang, Huan, Gao, Bingtuan, Zhao, Dehui, Shen, Huan, & Liu, Chuande. (2025). A grasping configuration planning method inspired by the geometric Fermat point. *Industrial Robot: the international journal of robotics research and application*. DOI: 10.1108/IR-12-2024-0565

9. Maksymova, I. I., & Slobodyaniuk, R. V. (2017). *Osoblyvosti vyznachennia ratsionalnoho polozhennia perevantazhuval'noho punkta. Features of determining the rational position of the transshipment point. In Proceedings of the 11th International Scientific and Practical Conference "Perspektyvy rozvytku budivel'nykh tekhnolohii"* (pp. 43–47). Dnipro.
10. Maksymov, I., Slobodyanyuk, V., & Maksymova, I. (2023). Features of determining the coordinates of the Fermat – Torricelli point for isosceles triangles. *International scientific journal "Internauka"*. № 8 (142). https://doi.org/10.25313/2520-2057-2023-8-8855
11. Maksymov, I. I., Slobodyaniuk, V. K., & Maksymova, I. I. (2023). *Analitychne vyznachennia koordynat tochky Ferma – Torrichelli ta minimalizatsiia transportnoi roboty. Analytical determination of the coordinates of the Fermat – Torricelli point and minimization of transport work*. Scientific Notes of V. I. Vernadsky Taurida National University. Series: Technical Sciences, 34 (73) (3), 1–7.
12. Maksymov, I. I., Kiyanovska, N. M., Slobodyaniuk, V. K., & Maksymova, I. I. (2024). *Analitychne doslidzhennia zadachi Fermat dlia rivnobedrenoho trykutnyka. Analytical study of the Fermat problem for an isosceles triangle.* Science and Technology Today, 5 (33). https://doi.org/10.52058/2786-6025-2024-5(33)-1348-1362

Дата першого надходження рукопису до видання: 21.08.2025
Дата прийнятого до друку рукопису після рецензування: 18.09.2025
Дата публікації: 31.12.2025

УДК 539.3
DOI https://doi.org/10.26661/2786-6254-2025-2-03

ЕФЕКТИВНИЙ МОДУЛЬ ЗСУВУ ВОЛОКНИСТОГО КОМПОЗИЦІЙНОГО МАТЕРІАЛУ З ПОРИСТОЮ МАТРИЦЕЮ

Плечун В. В.
аспірант
Запорізький національний університет
вул. Університетська, 66, Запоріжжя, Україна
orcid.org/0000-0002-3810-3916
valeria.gashenko@gmail.com

Спиця О. Г.
кандидат фізико-математичних наук, доцент,
доцент кафедри загальної математики
Запорізький національний університет
вул. Університетська, 66, Запоріжжя, Україна
orcid.org/0000-0002-7150-7736
spytsa.o.g@gmail.com

Ключові слова: *композиційний матеріал, пориста матриця, транстропне волокно, ефективний модуль зсуву, гомогенізація.*

Для математичного моделювання механічної поведінки волокнистих композиційних матеріалів з високою частотою армування використовують процедуру гомогенізації. Це дозволяє представити неоднорідний композиційний матеріал «уявним» гомогенізованим матеріалом, механічні властивості якого описуються ефективними пружними сталими. Для односпрямованого волокнистого композиційного матеріалу гомогенізований матеріал буде мати транстропні властивості із площиною ізотропії, перпендикулярною осі волокон. Для опису такого матеріалу необхідно знати п’ять ефективних пружних сталих. Для знаходження ефективного повздовжнього модуля зсуву застосовано метод представницького об’ємного елемента. Цей метод базується на тому, що із композиційного матеріалу вирізається представницький об’ємний елемент, що містить одне волокно з матрицею, що його оточує. Для цього елемента розв’язуються найпростіші задачі пружності. Для знаходження ефективного повздовжнього модуля зсуву розв’язується дві крайові задачі про повздовжній чистий зсув. Перша задача про сумісне деформування представницького об’ємного елемента, що складається з пористої матриці у формі порожнистого циліндра й ізотропного волокна у формі суцільного циліндра. На межі розділу матеріалу матриці та матеріалу волокна задаються умови неперервності ряду компонентів напружено-деформованого стану. Друга, аналогічна задача про деформування представницького об’ємного елемента гомогенізованого композиційного матеріалу у формі суцільного циліндра. Ефективний повздовжній модуль зсуву знаходиться з умови рівності осьових переміщень для будь-якої точки зовнішньої циліндричної поверхні в обох задачах. У результаті ефективний повздовжній модуль зсуву композиційного матеріалу є функцією пружних сталих матеріалу матриці та матеріалу волокна, пористості матриці й об’ємної частки волокна в композиті. За допомогою отриманої залежності проаналізовано залежність величини ефективного повздовжнього модуля зсуву від величини показників пористості матриці й об’ємної частки волокна в композиційному матеріалі.

EFFECTIVE SHEAR MODULUS OF FIBROUS COMPOSITE MATERIAL WITH POROUS MATRIX

Plechun V. V.
Postgraduate Student
Zaporizhzhia National University
Universytetska str., 66 Zaporizhzhia, Ukraine
orcid.org/0000-0002-3810-3916
valeria.gashenko@gmail.com

Spytsia O. G.
Candidate of Physical and Mathematical Sciences, Associate Professor,
Associate Professor at the Department of General Mathematics
Zaporizhzhia National University
Universytetska str., 66 Zaporizhzhia, Ukraine
orcid.org/0000-0002-7150-7736
spytsa.o.g@gmail.com

Key words: *composite material, porous matrix, transtropic fiber, effective shear modulus, homogenization.*

The homogenization procedure is used for mathematical modeling of the mechanical behavior of fibrous composite materials with a high reinforcement frequency. This allows the representation of a heterogeneous composite material by an "imaginary" homogenized material, the mechanical properties of which are described by effective elastic constants. For a unidirectional fibrous composite material, the homogenized material will have transtropic properties with an isotropy plane perpendicular to the fiber axis. To describe such a material, it is necessary to know five effective elastic constants. The representative volume element method was applied to find the effective longitudinal shear modulus. This method is based on the fact that a representative volume element containing one fiber with its surrounding matrix is cut from the composite material. The simplest elasticity problems are solved for this element. To find the effective longitudinal shear modulus, two boundary value problems are solved for longitudinal pure shear. The first problem is about the joint deformation of a representative volume element consisting of a porous matrix in the form of a hollow cylinder and an isotropic fiber in the form of a solid cylinder. At the interface between the matrix material and the fiber material, the conditions of continuity of a number of components of the stress-strain state are set. The second, similar problem is about the deformation of a representative volume element of a homogenized composite material in the form of a solid cylinder. The effective longitudinal shear modulus is found from the condition of equality of axial displacements for any point of the outer cylindrical surface in both problems. As a result, the effective longitudinal shear modulus of the composite material is a function of the elastic constants of the matrix material and the fiber material, the porosity of the matrix, and the volume fraction of the fiber in the composite. Using the obtained dependence, the dependence of the effective longitudinal shear modulus on the matrix porosity and fiber volume fraction in the composite material was analyzed.

Вступ. Посилення вимог до сучасних пристроїв, машин і механізмів приводить до необхідності вдосконалення конструкційних матеріалів для їх виробництва. Одними з найбільш перспективних видів матеріалів є композиційні матеріали, які складаються із двох і більше компонентів. Варіювання видами матеріалів компонентів, їх часткою в композиті дає змогу управляти властивостями матеріалу, що створюється. Для прогнозування властивостей майбутнього композиційного матеріалу на макрорівні та для їх моделювання під час розрахунків конструкцій

постає потреба у створенні математичних моделей для опису механічної поведінки. Беручи до уваги, що для створення композитів використовують значну кількість армувальних елементів, урахувати кожний із них у цих моделях дуже складно. Тому популярними є гомогенізовані моделі неоднорідних матеріалів, тобто коли композиційний матеріал представляється уявним однорідним, але таким, щоб його механічна поведінка максимально збігалась із поведінкою вихідного матеріалу. Механічні характеристики зазначених уявних матеріалів називають ефективними.

Визначення ефективних механічних характеристик є складним завданням і потребує максимального врахування особливостей механічних властивостей компонентів композиту та специфіки взаємодії цих компонентів у неоднорідному матеріалі. Якщо розглядати односпрямовані волокнисті композиційні матеріали, то гомогенізований матеріал, що їх представляє, доцільно вважати транстропним, із площиною ізотропії, перпендикулярною осі волокон. Щоб повною мірою описати механічну поведінку транстропного матеріалу, необхідно визначити п’ять ефективних пружних сталих. **Огляд літератури.** Розв’язанню таких задач присвячена робота [3], у якій визначено ефективні в’язкопружні характеристики за повздовжнього деформування односпрямованого волокнистого композиту методом представницького об’ємного елемента. Продовженням цієї роботи є стаття [7], де знайдено аналогічні в’язкопружні характеристики композиційного матеріалу, але закон стану матеріалу матриці на основі спадкової теорії Больцмана – Вольтерри описує релаксацію модуля зсуву, на відміну від роботи [3], де він застосовувався до модуля пружності матеріалу матриці.

У праці [9] визначаються ефективні пружні властивості композиту, армованого довгими волокнами, за наявності поперечної ізотропії матеріалу. Ефективні пружні характеристики отримані методом представницького об’ємного елемента для двох схем армування волокнами (гексагонального та квадратного). Проаналізовано вплив об’ємної частки та жорсткості міжфазного середовища на ефективні механічні властивості композиту.

Стаття [10] присвячена моделюванню ефективних пружних і в’язкопружних властивостей композиційних матеріалів, армованих короткими волокнами. Дослідження проводилось для математичних моделей представницьких об’ємних елементів матеріалу за допомогою методу скінченних елементів.

У монографії [5] методом представницького об’ємного елемента досліджено ефективні механічні характеристики односпрямованого композиційного матеріалу із транстропними порожнистими волокнами.

Робота [6] присвячена визначенню пружних механічних сталих композиційного матеріалу за наявності пористого перехідного шару між матеріалом матриці та матеріалом волокна шляхом дворівневої гомогенізації спочатку включення та пористого шару, а потім – отриманого гомогенізованого матеріалу та матриці.

Метод представницького об’ємного елемента використано в роботі [4] для знаходження аналітичних співвідношень для ефективних поздовжнього модуля пружності та коефіцієнта Пуассона волокнистого композиту із транстропного волокна й ізотропної пористої матриці.

Метою дослідження є визначення ефективного повздовжнього модуля зсуву для односпрямованого волокнистого композиційного матеріалу із пористою матрицею та транстропним волокном.

Результати. ***Розв’язання задачі про повздовжній зсув методом представницького об’ємного елемента.*** Для знаходження аналітичного виразу для ефективного повздовжнього модуля зсуву скористаємося методом представницького об’ємного елемента. Виокремимо із композиту гексагональну комірку, що складається із циліндричного волокна радіуса a та матриці, що його оточує. Апроксимуємо зовнішню, гексагональну, поверхню комірки колом радіуса b так, щоб об’єм матеріалу матриці не змінився (рис. 1).

Тоді, маючи на увазі, що довжина як суцільного, та і порожнистого циліндра є нескінченною, для об’ємної частки волокна в композиті будемо мати:

$$f = \frac{\pi a^2}{\pi b^2} = \frac{a^2}{b^2}. \qquad (1)$$

Задача про чистий поздовжній зсув (рис. 2) для циліндричного тіла із транстропними властивостями розглянута в роботі [1].

За деформування такого виду значна кількість компонентів напружено-деформований стану будуть дорівнювати нулю:

$$\sigma_{zz} = \sigma_{rr} = \sigma_{\theta\theta} = \sigma_{r\theta} = 0;$$

$$\varepsilon_{zz} = \varepsilon_{rr} = \varepsilon_{\theta\theta} = \varepsilon_{r\theta} = 0.$$

Інші компоненти в загальному виді можна записати так:

$$\sigma_{zr} = \sigma_{zr}(r,\theta), \sigma_{\theta z} = \sigma_{\theta z}(r,\theta), \gamma_{zr} = \gamma_{zr}(r,\theta), \gamma_{\theta z} = \gamma_{\theta z}(r,\theta).$$

Для нескінченної циліндричної області в моделюванні чистого поздовжнього зсуву до зовнішньої циліндричної поверхні прикладається таке навантаження:

$$\sigma_{zr}(b,\theta) = \sigma_0 cos\theta. \qquad (2)$$

З урахуванням викладених особливостей деформування циліндричної області за чистого

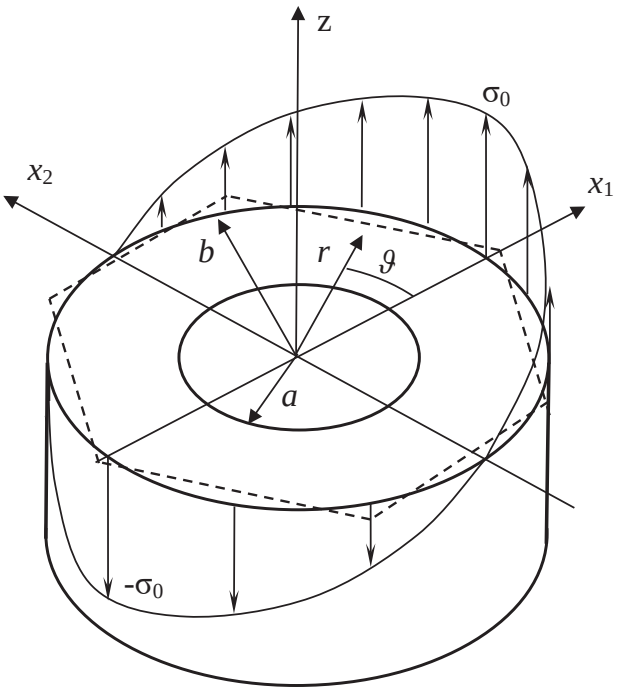


Рис. 1. Розподіл зовнішніх навантажень на межі кільця за поздовжнього зсуву

поздовжнього зсуву два із трьох рівнянь рівноваги перетворяться на тотожні рівності, залишиться лише одне рівняння в напруженнях:

$$\frac{\partial \sigma_{zr}}{\partial r}+\frac{1}{r}\frac{\partial \sigma_{z\theta}}{\partial \theta}+\frac{\sigma_{zr}}{r}=0.$$

У переміщеннях це рівняння рівноваги виглядатиме так:

$$\frac{\partial^2 u_z}{\partial r^2}+\frac{1}{r}\frac{\partial u_z}{\partial r}+\frac{1}{r^2}\frac{\partial^2 u_z}{\partial \theta^2}=0. \qquad (3)$$

Розв'язок цього рівняння відомий:

$$u_z(r,\theta)=\left(C_1 r+\frac{C_2}{r}\right)cos\theta. \qquad (4)$$

Скориставшись залежностями між переміщеннями та деформаціями, можемо визначити співвідношення для кутів зсуву:

$$\gamma_{\theta z}(r,\theta)=\frac{1}{r}\frac{\partial u_z}{\partial \theta}=-\left(C_1+\frac{C_2}{r^2}\right)sin\theta;$$

$$\gamma_{zr}(r,\theta)=\frac{\partial u_z}{\partial r}=\left(C_1-\frac{C_2}{r^2}\right)cos\theta. \qquad (5)$$

Знаючи деформації і скориставшись законом Гука для транстропного матеріалу, будемо мати:

$$\sigma_{z\theta}(r,\theta)=-G_{12}\left(C_1+\frac{C_2}{r^2}\right)sin\theta;$$

$$\sigma_{zr}(r,\theta)=G_{12}\left(C_1-\frac{C_2}{r^2}\right)cos\theta. \qquad (6)$$

Тепер, маючі загальні співвідношення для чистого повздовжнього зсуву для транстропного тіла, розв'яжемо задачу про сумісний поздовжній зсув транстропного волокна, що моделюється суцільним циліндром $(0\le r\le a)$, та пористої ізотропної матриці, що моделюється порожнистим циліндром $(a\le r\le b)$.

Для моделювання повздовжнього зсуву представницького об'ємного елемента необхідно визначити пружні характеристики пористої матриці, тому скористаємося варіаційним методом Хашина – Штрикмана [8] для випадкового просторового розподілення пор, опишемо ефективний модуль зсуву таким співвідношенням:

$$\frac{G(p)}{G(0)}=\left(1+\frac{2}{3}(1-\rho)\left(\frac{10G(0)}{9K(0)-G(0)}\right)\right)^{-1}, \qquad (7)$$

де $K(0),G(0)$ – об'ємний модуль пружності та модуль зсуву матеріалу матриці пористого матеріалу, $G(p)$ – модуль зсуву пористого матеріалу, ρ – відносна щільність пористого матеріалу ($p=1-\rho$ – пористість матеріалу).

Зі співвідношень (7) маємо вираз для ефективного модуля зсуву матриці композиційного матеріалу:

$$G^*=\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}, \qquad (8)$$

де $G^*=G(p)$, $K_0^*=K(0)$, $G_0^*=G(0)$.

Тоді, з урахуванням (4) – (6) та (2), для пористої ізотропної матриці будемо мати:

$$u_z^*(r,\theta)=\left(Ar+\frac{B}{r}\right)cos\theta; \qquad (9)$$

$$\gamma_{\theta z}^*(r,\theta)=-\left(A+\frac{B}{r^2}\right)sin\theta;\gamma_{zr}^*(r,\theta)=\left(A-\frac{B}{r^2}\right)cos\theta; \qquad (10)$$

$$\sigma_{z\theta}^*(r,\theta)=-\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}\left(A+\frac{B}{r^2}\right)sin\theta;$$

$$\sigma_{zr}^*(r,\theta)=\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}\left(A-\frac{B}{r^2}\right)cos\theta, \qquad (11)$$

тут A,B – сталі, що визначаються із крайових умов.

А враховуючи (4) – (6), для транстропного волокна з урахуванням скінченності переміщень за $r=0$, матимемо:

$$u_z^\circ(r,\theta)=Crcos\theta; \qquad (12)$$

$$\gamma_{\theta z}^\circ(\theta)=-Csin\theta;\gamma_{zr}^\circ(\theta)=Ccos\theta; \qquad (13)$$

$$\sigma_{z\theta}^\circ(\theta)=-G_{12}^\circ Csin\theta;\sigma_{zr}^\circ(\theta)=G_{12}^\circ Ccos\theta, \qquad (14)$$

де C – стала, що визначається із крайових умов.

Для знаходження сталих A,B,C у співвідношеннях (9) – (14) скористаємося крайовою умовою (2) та умовами неперервності на межі розділу матеріалу волокна й матеріалу матриці:

$$\sigma_{zr}^\circ(\theta)=\sigma_{zr}^*(a,\theta); \qquad (15)$$

$$u_z^\circ(a,\theta)=u_z^*(a,\theta). \qquad (16)$$

Скориставшись (2), маємо:

$$A=\sigma_0\frac{27K_0^*+17G_0^*-20G_0^*\rho}{3\left(9K_0^*-G_0^*\right)G_0^*}+\frac{B}{b^2}. \quad (17)$$

З урахуванням (17) і (16), отримуємо:

$$C=\sigma_0\frac{27K_0^*+17G_0^*-20G_0^*\rho}{3\left(9K_0^*-G_0^*\right)G_0^*}+B\left(\frac{a^2+b^2}{a^2b^2}\right). \quad (18)$$

Скориставшись (15), з урахуванням співвідношень (17) та (18), матимемо:

$$B=\frac{\sigma_0a^2\left(G_{12}^\circ-\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}\right)}{\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}\left((f-1)\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}-G_{12}^\circ(f+1)\right)}. \quad (19)$$

Тоді з (18) отримаємо:

$$C=\frac{-2\sigma_0}{\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}(f-1)-G_{12}^\circ(f+1)}; \quad (20)$$

$$A=\frac{-\sigma_0\left(\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}+G_{12}^\circ\right)}{\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}\left(\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}(f-1)-G_{12}^\circ(f+1)\right)}. \quad (21)$$

Беручи до уваги (19) і (21), запишемо основні співвідношення, що описують напружено-деформований стан матриці за чистого поздовжнього зсуву:

$$u_z^*(r,\theta)=$$

$$=\frac{\sigma_0}{\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}\left(\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}(f-1)-G_{12}^\circ(f+1)\right)}\times$$

$$\times\left(-\left(\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}+G_{12}^\circ\right)r+\right.$$

$$\left.+\frac{a^2\left(G_{12}^\circ-\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}\right)}{r}\right)cos\theta; \quad (22)$$

$$\gamma_{\theta z}^*(r,\theta)=$$

$$=\frac{\sigma_0}{\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}\left(\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}(f-1)-G_{12}^\circ(f+1)\right)}\times$$

$$\times\left(\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}+G_{12}^\circ-\right.$$

$$\left.-\frac{a^2\left(G_{12}^\circ-\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}\right)}{r^2}\right)sin\theta; \quad (23)$$

$$\gamma_{zr}^*(r,\theta)=$$

$$=\frac{-\sigma_0}{\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}\left(\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}(f-1)-G_{12}^\circ(f+1)\right)}\times$$

$$\times\left(\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}+G_{12}^\circ+\right.$$

$$\left.+\frac{a^2\left(G_{12}^\circ-\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}\right)}{r^2}\right)cos\theta;(24) \quad (24)$$

$$\sigma_{\theta z}^*(r,\theta)=$$

$$=\frac{\sigma_0}{\left(\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}(f-1)-G_{12}^\circ(f+1)\right)}\times$$

$$\times\left(\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}+G_{12}^\circ-\right.$$

$$\left.-\frac{a^2\left(G_{12}^\circ-\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}\right)}{r^2}\right)sin\theta; \quad (25)$$

$$\sigma_{zr}^*(r,\theta)=$$

$$=\frac{-\sigma_0}{\left(\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}(f-1)-G_{12}^\circ(f+1)\right)}\times$$

$$\times\left(\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}+G_{12}^\circ+\right.$$

$$\left.+\frac{a^2\left(G_{12}^\circ-\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}\right)}{r^2}\right)cos\theta. \quad (26)$$

Скориставшись знайденим співвідношенням (20) для C, отримаємо вирази, що описують напружено-деформований стан волокна у представницькій об'ємній комірці за сумісного поздовжнього зсуву:

$$u_z^\circ(r,\theta)=\frac{-2\sigma_0r\cos\theta}{\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}(f-1)-G_{12}^\circ(f+1)}; \quad (27)$$

$$\gamma_{\theta z}^\circ(\theta)=\frac{2\sigma_0\sin\theta}{\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}(f-1)-G_{12}^\circ(f+1)}; \quad (28)$$

$$\gamma_{zr}^\circ(\theta)=\frac{-2\sigma_0\cos\theta}{\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}(f-1)-G_{12}^\circ(f+1)}; \quad (29)$$

$$\sigma_{zr}^\circ(\theta)=\frac{-2\sigma_0G_{12}^\circ\cos\theta}{\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}(f-1)-G_{12}^\circ(f+1)}; \quad (30)$$

$$\sigma_{z\theta}^\circ(\theta)=\frac{2\sigma_0G_{12}^\circ\sin\theta}{\frac{3\left(9K_0^*-G_0^*\right)G_0^*}{27K_0^*+17G_0^*-20G_0^*\rho}(f-1)-G_{12}^\circ(f+1)}. \quad (31)$$

Тепер визначимо напружено-деформований стан представницького об'ємного елемента гомогенізованого композиційного матеріалу в тих самих умовах чистого поздовжнього зсуву. Для цього елементарну комірку представимо як суцільний нескінченний циліндр радіуса *b*, а крайові умови візьмемо незмінними – у вигляді (2). Тоді, скориставшись загальними співвідношеннями (4) – (6), знайдемо напружено-деформований стан елементарної представницької комірки гомогенізованого композиційного матеріалу:

$$u_z(r,\theta) = \tilde{A} r \cos\theta; \tag{32}$$

$$\gamma_{\theta z}(r,\theta) = -\tilde{A} \sin\theta; \tag{33}$$

$$\gamma_{zr}(r,\theta) = \tilde{A} \cos\theta; \tag{34}$$

$$\sigma_{zr}(r,\theta) = \tilde{A} G_{12} \cos\theta; \tag{35}$$

$$\sigma_{z\theta}(r,\theta) = -\tilde{A} G_{12} \sin\theta. \tag{36}$$

Скориставшись крайовою умовою (2), знайдемо сталу $\tilde{A}$:

$$\tilde{A} = \frac{\sigma_0}{G_{12}}.$$

Тоді співвідношення (32) – (36) виглядатимуть так:

$$u_z(r,\theta) = \frac{\sigma_0}{G_{12}} r \cos\theta; \tag{38}$$

$$\gamma_{\theta z}(r,\theta) = -\frac{\sigma_0}{G_{12}} \sin\theta; \tag{39}$$

$$\gamma_{zr}(r,\theta) = \frac{\sigma_0}{G_{12}} \cos\theta; \tag{40}$$

$$\sigma_{zr}(r,\theta) = \frac{\sigma_0}{G_{12}} G_{12} \cos\theta; \tag{41}$$

$$\sigma_{z\theta}(r,\theta) = -\frac{\sigma_0}{G_{12}} G_{12} \sin\theta. \tag{42}$$

Тобто в результаті розв'язання двох крайових задач про чистий поздовжній зсув визначено напружено-деформований стан двох представницьких об'ємних комірок – сумісного деформування для комірки з матеріалу матриці та матеріалу волокна й комірки з гомогенізованого матеріалу. Для знаходження ефективного модуля зсуву скористаємося, як умовою узгодження, рівністю осьових переміщень у будь-якій точці зовнішньої межі для обох представницьких комірок:

$$u_z(b,\theta) = u_z^*(b,\theta). \tag{43}$$

З урахуванням рівності (43), матимемо:

$$G_{12} = \frac{\dfrac{3\left(9K_0^* - G_0^*\right)G_0^*}{27K_0^* + 17G_0^* - 20G_0^*\rho}\left(\dfrac{3\left(9K_0^* - G_0^*\right)G_0^*}{27K_0^* + 17G_0^* - 20G_0^*\rho}(1-f) + G_{12}^\circ(f+1)\right)}{G_{12}^\circ(1-f) + \dfrac{3\left(9K_0^* - G_0^*\right)G_0^*}{27K_0^* + 17G_0^* - 20G_0^*\rho}(f+1)}. \tag{44}$$

Аналіз чисельних результатів. Обчислимо ефективний повздовжній модуль зсуву волокнистого композиційного матеріалу з пористою матрицею за формулою (44). Проаналізуємо зміну його величини залежно від пористості матриці й об'ємного вмісту волокна в композиті. Як матеріал матриці візьмемо алюміній, а як матеріал волокна – бор. Для матриці пористого матеріалу будемо мати такі механічні сталі [2]: модуль пружності $E^* = 70$ ГПа, коефіцієнт Пуассона ν^*=0,32, для матеріалу ізотропного волокна – $E^\circ = 416{,}5$ ГПа, коефіцієнт Пуассона ν°=0,2. Для об'ємного модуля пружності та модуля зсуву матриці пористого матеріалу будемо, відповідно, мати $K_0^* = E^* / \left(3\left(1-2\nu^*\right)\right); G_0^* = E^* / \left(2\left(1+\nu^*\right)\right)$. Для модуля зсуву матеріалу волокна – $G_{12}^\circ = G^\circ = E^\circ / \left(2\left(1+\nu^\circ\right)\right)$. Результати чисельних розрахунків представлені на рисунку 2.

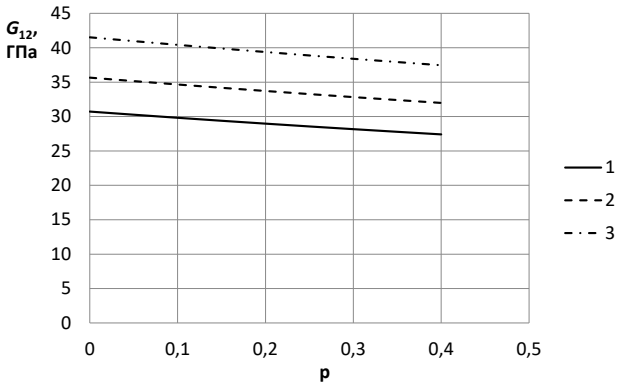


Рис. 2. Залежність ефективного повздовжнього модуля зсуву G_{12} від пористості матриці p (формула (44)) та від об'ємного вмісту корда: 1 – $f = 0{,}1$; 2 – $f = 0{,}2$; 3 – $f = 0{,}3$

Зазначимо, що збільшення об'ємного вмісту волокна приводить до збільшення ефективного повздовжнього модуля зсуву композиційного матеріалу, що зумовлено більш високою жорсткістю матеріалу волокна. Наявність пор, навпаки, зменшує величину ефективного повздовжнього модуля зсуву, ця залежність близька до лінійної. Так, збільшення пористості від 0 до 0,4 зменшує значення модуля зсуву композиту на 10–15%.

Висновки. За допомогою методу представницького об'ємного елемента знайдено аналітичний вираз для ефективного повздовжнього модуля зсуву композиційного матеріалу із пористої матриці та трансторопного волокна, який є функцією пружних сталих матеріалу матриці та матеріалу волокна, пористості матриці й об'ємної частки волокна в композиті. За допомогою отриманої формули проаналізовано вплив величини пористості матриці й об'ємної частки волокна на значення ефективного повздовжнього модуля зсуву композиційного матеріалу.

ЛІТЕРАТУРА

1. Гребенюк С. М., Гоменюк С. І., Клименко М. І. Напружено-деформований стан просторових конструкцій на основі гомогенізації волокнистих композитів. Херсон : Видавничий дім «Гельветика», 2019. 350 с.
2. Композиционные материалы: справочник / под ред. Д. М. Карпиноса. Киев : Наукова думка, 1985. 592 с.
3. Пожуєв В. І., Артеменко А. О., Клименко М. І., Скрипник К. В. Гомогенізація в'язкопружного композиту в разі повздовжнього розтягу. *Computer Science and Applied Mathematics*. 2022. Вип. 2. С. 35–42. URL: https://journalsofznu.zp.ua/index.php/compscience/ article/view/3145/2986
4. Пожуєв В. І., Плечун В. В., Спиця О. Г., Кончиньска Є. О. Ефективні пружні сталі волокнистого композиту із пористою матрицею при повздовжньому розтягненні. *Computer Science and Applied Mathematics*. 2022. Вип. 2. С. 27–34. https://doi.org/10.26661/2786-6254-2022-2-03
5. Столярова А. В. Ефективні механічні характеристики композиційних матеріалів із трансторопними порожнистими волокнами : монографія. Херсон : Видавничий дім «Гельветика», 2021. 104 с.
6. Хорошун Л. П., Левчук О. И. Эффективные упругие свойства зернистых стохастических композитов при несовершенной адгезии. *Доповіді Національної академії наук України*. 2017. Вип. 12. С. 33–44. URL: http://jnas.nbuv.gov.ua/article/UJRN-0000867454
7. Bulat A. F., Dyrda V. I., Grebenyuk S. N., Klymenko M. I. Determination of Effective Characteristics of a Fibrous Composite with Account of Viscoelastic Deformation of its Components. *Strength of Materials.* 2020. Vol. 52. № 5. P. 691–699. URL: https://www.researchgate.net/publication/347288115
8. Hashin Z., Shtrikman S. A variational approach to the theory of the elastic behaviour of multiphase materials. *Journal of the Mechanics and Physics of Solids*. 1963. Vol. 11. № 2. P. 127–140.
9. Xu Y., Du S., Xiao J., Zhao Q. Evaluation of the effective elastic properties of long fiber reinforced composites with interphases. *Computational Materials Science*. 2012. № 61. C. 34–41.
10. Yang P., Chen Y., Guo Z., Hu N., Sun W. Modeling the effective elastic and viscoelastic properties of randomly distributed short fiber reinforced composites. *Composites Communications*. 2022. № 35. P. 78–84.

REFERENCES

1. Grebenyuk, S. M., Gomenyuk, S. I., & Klimenko, M. I. (2019). Napruzheno-deformovaniy stan prostorovikh konstruktsiy na osnovi gomogenizatsiї voloknistikh kompozitiv [Stress-strain state of spatial structures based on homogenization of fibrous composites]. Kherson: Helvetica Publishing House. (in Ukrainian)
2. Kompozitsionnyye materialy: spravochnik [Composite materials: reference book] / ed. Karpinos, D. M. (1985). Kyiv: Naukova dumka. (in Russian).
3. Pozhuev, V. I., Artemenko, A. O., Klymenko, M. I., Scrypnyk, K. V. (2022). Homohenizatsiya v'yazkopruzhnoho kompozyta u razi povzdovzhn'oho roztyahu [Homogenization of a viscoelastic composite in the case of longitudinal tension]. *Computer Science and Applied Mathematics*. 2, 35–42. https://journalsofznu.zp.ua/index.php/compscience/ article/view/3145/2986 (in Ukrainian).
4. Pozhuev, V. I., Plechun, V. V., Spytsia, O. G., Konchynska, E. O. (2022). Efektyvni pruzhni stali voloknystoho kompozytu iz porystoyu matrytseyu pry povzdovzhn′omu roztyahnenni [Effective Elastic Constants of a Fiber Composite with Porous Matrix in Longitudinal Stretch]. *Computer Science and Applied Mathematics*. 2, 27–34. https://doi.org/10.26661/2786-6254-2022-2-03 (in Ukrainian).
5. Stolyarova, A. V. (2021). Efektyvni mekhanichni kharakterystyky kompozytsiynykh materialiv iz transtropnymy porozhnystymy voloknamy: monohrafiya [Effective mechanical characteristics of composite materials with transtropical hollow fibers: monograph]. Kherson: Helvetica Publishing House. (in Ukrainian).
6. Khoroshun, L. P., Levchuk, O. I. (2017). Effektivnyye uprugiye svoystva zernistykh stokhasticheskikh kompozitov pri nesovershennoy adgezii [Effective elastic properties of granular stochastic composites with imperfect adhesion]. *Reports of the National Academy of Sciences of Ukraine*. 12, 33–44. http://jnas.nbuv.gov.ua/article/UJRN-0000867454 (in Russian).
7. Bulat, A. F., Dyrda, V. I., Grebenyuk, S. N., & Klymenko, M. I. (2020). Determination of Effective Characteristics of a Fibrous Composite with Account of Viscoelastic Deformation of its Components. *Strength of Materials*. 52 (5), 691–699. https://www.researchgate.net/publication/347288115
8. Hashin, Z., & Shtrikman, S. (1963). A variational approach to the theory of the elastic behaviour of multiphase materials. *Journal of the Mechanics and Physics of Solids*. 11 (2), 127–140.
9. Xu, Y., Du, S., Xiao, J., & Zhao, Q. (2012). Evaluation of the effective elastic properties of long fiber reinforced composites with interphases. *Computational Materials Science*. 61, 34–41.
10. Yang, P., Chen, Y., Guo, Z., Hu, N., & Sun, W. (2022). Modeling the effective elastic and viscoelastic properties of randomly distributed short fiber reinforced composites. *Composites Communications*. 35, 78–84.

Дата першого надходження рукопису до видання: 29.08.2025
Дата прийнятого до друку рукопису після рецензування: 26.09.2025
Дата публікації: 31.12.2025

УДК 539.377
DOI https://doi.org/10.26661/2786-6254-2025-2-04

ТЕРМОНАПРУЖЕНИЙ СТАН ТРЬОХЕЛЕМЕНТНОЇ ПРИЗМАТИЧНОЇ ОБОЛОНКИ З РІЗНИМИ КОЕФІЦІЄНТАМИ ТЕПЛОВІДДАЧІ

Хапко Б. С.
кандидат фізико-математичних наук,
старший науковий співробітник
Інститут прикладних проблем механіки і математики
імені Я. С. Підстригача Національної академії наук України
вул. Наукова, 3-Б, Львів, Україна
orcid.org/0009-0003-3418-6835
bogdan.khapko@gmail.com

Ключові слова:
теплопровідність, оболонка, злами, кусково-постійні коефіцієнти тепловіддачі, прогин, зусилля, моменти.

Досліджено термонапружений стан у тонкій пологій призматичній оболонці, складеній із трьох плоских елементів, з урахуванням конвективного теплообміну за різної температури навколишнього середовища на кожній із двох лицевих поверхонь оболонки. Коефіцієнти тепловіддачі на лицевих поверхнях кожного із трьох елементів оболонки є різними. Запропонована модель опису температурного поля у призматичній оболонці, зумовленого різницею температур навколишнього середовища на лицевих поверхнях. Задачу теплопровідності для оболонки зведено до розв'язування системи інтегральних рівнянь з інтегральними операторами Вольтерри та Фредгольма другого роду для функцій, що є лінійними комбінаціями температурних характеристик (середня температура та температурний момент). За допомогою методу квадратурних формул побудовано числову схему розв'язування інтегральних рівнянь. Зокрема, квадратурні формули Сімпсона використовуються для визначення інтегралів, а система інтегральних рівнянь з інтегральними операторами Вольтерри та Фредгольма другого роду зводиться до системи лінійних алгебраїчних рівнянь. Функцію напружень і прогин оболонки знайдено за допомогою скінченних інтегральних перетворень Фур'є. На краях оболонка зі зламами підтримується жорсткими вертикальними діафрагмами. Для визначення термонапруженого стану оболонки прийнято, що краї, на які виходять злами, теплоізольовані, а температура інших торців дорівнює нулю. Наведено результати числового аналізу розподілу середньої температури, температурного моменту, прогину, моментів і зусиль за різних значень коефіцієнта тепловіддачі на верхніх лицевих поверхнях першого й третього елементів оболонки. Виявлено, що зменшення коефіцієнта тепловіддачі на крайніх елементах приводить до спадання середньої температури та температурного моменту оболонки зі зламами, згинних моментів і абсолютних величин прогину й зусиль зі зміщенням максимальної їх величини в бік елемента, де коефіцієнти тепловіддачі не змінювали.

THERMOSTRESSED STATE OF A THREE-ELEMENT PRISMATIC SHELL WITH DIFFERENT HEAT-TRANSFER COEFFICIENTS

Khapko B. S.
Ph. D. in Physics and Maths, Senior Research Fellow
Pidstryhach Institute for Applied Problems of Mechanics and Mathematics National Academy of Sciences of Ukraine
Naukova str., 3-B, Lviv, Ukraine
orcid.org/0009-0003-3418-6835
bogdan.khapko@gmail.com

Key words: *heat conduction, shell, folds, piecewise-constant heat transfer coefficients, deflection, internal forces, moments.*

The problem of determining the thermostressed state in a thin shallow prismatic shell composed of three flat elements is considered, taking into account convective heat exchange at different ambient temperatures on each of the two face surfaces of the shell. The heat-transfer coefficients at the face surfaces of each of the three shell elements are assumed to be different. A model is proposed for describing the temperature field in the prismatic shell, caused by the difference in ambient temperatures on the face surfaces. The heat conduction boundary value problem is reduced to a system of integral equations with Volterra and Fredholm integral operators of the second kind for functions that are linear combinations of the temperature characteristics (the mean temperature and the temperature moment). A numerical scheme for solving the system of integral equations is constructed using the quadrature method. In particular, Simpson's quadrature formulae are used to evaluate the integrals, reducing the system of integral equations with Volterra and Fredholm operators to a system of linear algebraic equations. The finite Fourier integral transforms are employed to determine the stress function and the deflection of the shell. The shell with folds is supported by rigid vertical diaphragms at its edges. The edges containing the folds of the mid-surface are assumed to be thermally insulated, while the temperature of the surrounding environment is taken as zero at the other edges. Numerical results for the distributions of the mean temperature, temperature moment, deflection, bending moments, and internal forces are presented for various values of the heat-transfer coefficients at the upper face surfaces of the first and third shell elements. It is revealed that a decrease in the heat-transfer coefficient in the outer elements leads to a reduction in the mean temperature, temperature moment, bending moments, absolute values of deflection and internal forces, and causes the location of their maximum values to shift toward the fold where the heat-transfer coefficient remains unchanged.

Вступ. Злами поверхонь тонкостінних елементів конструкцій істотно впливають на розподіл напружень і температури в тілі та можуть зумовити збільшення його загальної жорсткості. Тонкостінні оболонки з нерегулярними серединними поверхнями знайшли застосування в авіа- і ракетобудуванні, залізничному транспорті, будівельній індустрії, машинобудуванні й інших галузях техніки, тому завдання з їх дослідження є актуальними.

Концентрацію напружень в тонкостінних елементах конструкцій з геометричними неоднорідностями серединної поверхні у формі зламів проаналізовано у праці [1]. Тут злам змодельований розподіленим уздовж нього фіктивним нормальним навантаженням, пропорційним мембранному зусиллю. На основі рівнянь теорії пологих оболонок і розвинення фіктивного навантаження в ряд Фур'є отримані вирази для прогину та функції зусиль. У праці [2] досліджено напружений стан оболонки, яка складена із двох півбезмежних пластин, спряжених під прямим кутом. Розрахунок зведено до задачі про сумісний плоский і згинний стан нескінченної пластинки з дефектом, що відповідає лінії спряження, при переході через який зусилля або переміщення мають стрибки. У праці [3] визначено власні частоти коливань коробчастих оболонок методом однорідних розв'язків. Методом скінченних елементів досліджено власні частоти й форми коливання сталевої пологої оболонки зі зламами [4] та деформування,

стійкість і коливання пружних оболонок ступінчасто-змінної товщини [5]. Прогин оболонок з недосконалостями типу вм'ятини досліджено у праці [6]. Запропоновано аналітично-чисельний підхід до розв'язання задач про стійкість складених дискретно підкріплених проміжними шпангоутами тришарової оболонкової конструкції конус – циліндр [7] і оболонкової конструкції з різними знаками Гауссової кривини [8] за статичного комбінованого навантаження зовнішнім тиском, осьовими зусиллями й крутильним моментом.

Уперше апарат узагальнених функцій для опису зламу оболонки застосовано у праці [9], де кривину поверхні аналітично виражено через δ – функцію, помножену на кут зламу. Це дало змогу описати напружено-деформований стан такої оболонки частково виродженими диференціальними рівняннями.

Перелічені вище праці стосуються дослідження напружено-деформованого стану оболонок зі зламами серединної поверхні за силових навантажень. Водночас багато складених оболонкових конструкцій зі зламами в енергетиці, ракетно-космічній техніці, на транспорті й інших галузях функціонують за дії високих температур, які зумовлюють істотний вплив на їхній напружено-деформований стан. Проте для його дослідження насамперед потрібно розвинути метод розв'язування рівнянь теплопровідності для оболонок зі зламами, а далі за отриманим полем температури визначити термічні напруження і деформації. За допомогою методу узагальнених функцій автор запропонував математичну модель температурного поля і зумовленого ним термонапруженого стану пологої сферичної оболонки, складеної з дев'яти квадратних елементів [10], та пологої призматичної оболонки, складеної із двох прямокутних елементів [11], за умови, що коефіцієнти тепловіддачі з лицевих поверхонь однакові. У праці [12] цей підхід узагальнено стосовно термонапруженого стану двохелементної призматичної оболонки за різних коефіцієнтів тепловіддачі з лицевих поверхонь. У праці [13] отримано рівняння теплопровідності для тонких пологих оболонок зі зламами вздовж координатних ліній і запропоновано спосіб зведення їх до системи інтегральних рівнянь. У працях [14–18] досліджено вплив сталих і кусково-сталих коефіцієнтів тепловіддачі на лицевих поверхнях на напружено-деформований стан пластинок і пологих оболонок без зламів.

Ця стаття присвячена дослідженню температурного поля і термонапруженого стану призматичної пологої оболонки, складеної із трьох плоских елементів, за теплового навантаження з різними коефіцієнтами тепловіддачі на лицевих поверхнях кожного з елементів. Запропоновано спосіб зведення крайової задачі теплопровідності для оболонки зі зламами до системи інтегральних рівнянь з інтегральними операторами Вольтерри та Фредгольма другого роду щодо функцій, що є лінійними комбінаціями інтегральних характеристик температури. Використали теорію узагальнених функцій і методи варіації сталої та кінцевих інтегральних перетворень Фур'є для розрахунку термопружного стану розглядуваної оболонки як єдиного цілого. Проаналізовано залежність температури, прогину та згинних моментів на лініях зламів від зменшення коефіцієнта тепловіддачі на крайніх елементах верхньої лицевої поверхні оболонки.

Результати. ***Задача теплопровідності.*** Розглянемо тонку пологу оболонку зі сторонами r та d_1, яка складена із трьох плоских елементів (рис. 1). Коефіцієнти теплопровідності всіх елементів однакові. Уздовж прямих у точках a та c оболонка має злами. Відповідно до теорії про пологі оболонки кути зламів θ_j ($j=1,2$) приймаємо малими. На лицевих поверхнях $z=\pm h$ пластинкових елементів відбувається конвективний теплообмін із середовищем температури $t_c^{\pm}$ за різних коефіцієнтів тепловіддачі на кожному з них:

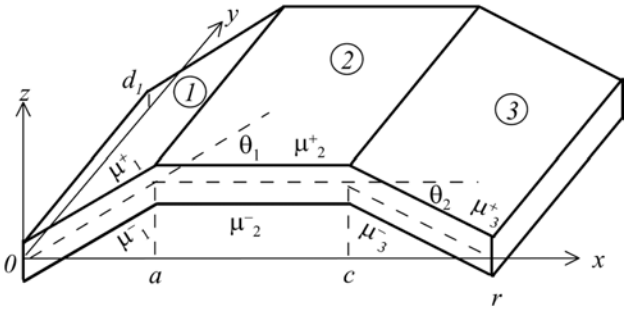


Рис. 1. Розрахункова схема оболонки

$$\mu^{+}(x)=\mu_1^{+}+\left(\mu_2^{+}-\mu_1^{+}\right)S_{-}(x-a)+\left(\mu_3^{+}-\mu_2^{+}\right)S_{-}(x-c);$$

$$\mu^{-}(x)=\mu_1^{-}+\left(\mu_2^{-}-\mu_1^{-}\right)S_{-}(x-a)+\left(\mu_3^{-}-\mu_2^{-}\right)S_{-}(x-c),$$

де μ_γ^{+}, μ_γ^{-} – коефіцієнти тепловіддачі на кожному з елементів верхньої та нижньої лицевих поверхонь оболонки відповідно; $\gamma=1,2,3$; $S_{-}(x-a)=\begin{cases}1, x\ge a,\\ 0, x<a\end{cases}$ – функція Гевісайда.

Диференціальні характеристики серединної поверхні пологої оболонки запишемо як [9]: $k=-\frac{1}{2}\theta_1\delta(x-a)-\frac{1}{2}\theta_2\delta(x-c)$, де δ – функція Дірака, k – кривина оболонки. Оболонка на торцях $x=0,r$ обмінюється теплом із зовнішнім середовищем температури t_c^j та коефіцієнтами тепловіддачі $\bar{b}_j$ відповідно, а також теплоізольована на краях $y=0$ та $y=d_1$. За таких умов температура не залежить від координати y. Згідно з гіпотезою про лінійний за товщиною призматичної оболонки розподіл температури [15] стаціонарне температурне

поле в ній подано у вигляді $t(x,z)=T_1(x)+\frac{z}{h}T_2(x)$, де h – півтощина оболонки, через інтегральні температурні характеристики – середню температуру $T_1(x)=\frac{1}{2h}\int_{-h}^{h}t(x,z)dz$ і температурний момент $T_2(x)=\frac{3}{2h^2}\int_{-h}^{h}zt(x,z)dz$, які визначаються із системи диференціальних рівнянь [13, с. 71]:

$$\Delta T_1(\eta)-\eta_1^+\left(T_1(\eta)-t_1\right)-\eta_1^-\left(T_2(\eta)-t_2\right)=\frac{1}{2}\left[\theta_1\delta(\eta-x_1)+\theta_2\delta(\eta-x_2)\right]T_2(\eta)+$$
$$+\left[(\eta_2^+-\eta_1^+)\left(T_1(\eta)-t_1\right)+(\eta_2^--\eta_1^-)\left(T_2(\eta)-t_2\right)\right]S_-(\eta-x_1)+$$
$$+\left[(\eta_3^+-\eta_2^+)\left(T_1(\eta)-t_1\right)+(\eta_3^--\eta_2^-)\left(T_2(\eta)-t_2\right)\right]S_-(\eta-x_2),$$
$$\Delta T_2(\eta)-3(1+\eta_1^+)\left(T_2(\eta)-t_2\right)-3\eta_1^-\left(T_1(\eta)-t_1\right)=\frac{3}{2}\left[\theta_1\delta(\eta-x_1)+\theta_2\delta(\eta-x_2)\right]T_1(\eta)+$$
$$+3\left[(\eta_2^+-\eta_1^+)\left(T_2(\eta)-t_2\right)+(\eta_2^--\eta_1^-)\left(T_1(\eta)-t_1\right)\right]S_-(\eta-x_1)+$$
$$+3\left[(\eta_3^+-\eta_2^+)\left(T_2(\eta)-t_2\right)+(\eta_3^--\eta_2^-)\left(T_1(\eta)-t_1\right)\right]S_-(\eta-x_2)+3t_2 \quad (1)$$

за крайових умов:

$$\frac{dT_j(\eta)}{d\eta}-b_1\left(T_j(\eta)-T_{j1}^c\right)=0 \text{, при } \eta=0;$$
$$\frac{dT_j(\eta)}{d\eta}+b_2\left(T_j(\eta)-T_{j2}^c\right)=0,\ \eta=l;\ j=1,2. \quad (2)$$

Тут $T_{j1}^c=\frac{1}{2h}\int_{-h}^{h}t_c^j dz$, $T_{j2}^c=\frac{3}{2h^2}\int_{-h}^{h}zt_c^j dz$, $t_{1,2}=\left(t_c^+\pm t_c^-\right)/2$;

$\eta_\gamma^+=\frac{h\left(\mu_\gamma^++\mu_\gamma^-\right)}{2}$, $\eta_\gamma^-=\frac{h\left(\mu_\gamma^+-\mu_\gamma^-\right)}{2}$; $\Delta=\frac{d^2}{d\eta^2}$; $\eta=x/h$ –

безрозмірна координата; $x_1=a/h$; $x_2=c/h$; $l=r/h$; $b_j=h\bar{b}_j/\lambda$ – безрозмірні коефіцієнти тепловіддачі на торцях $\eta=0$ та $\eta=l$ відповідно; λ – коефіцієнт теплопровідності.

Увівши заміни

$$T_1=\frac{\lambda_2\left(F_1-\lambda_1F_2\right)}{\lambda_2-\lambda_1}+t_1,\ T_2=\frac{\lambda_2F_2-F_1}{\lambda_2-\lambda_1}+t_2, \quad (3)$$

систему (1) та крайові умови (2) запишемо через функції $F_1(\eta),F_2(\eta)$ у вигляді:

$$\Delta F_1-a_1^2F_1=p_1(\eta)+\frac{1}{2}\left[\theta_1\delta(\eta-x_1)+\theta_2\delta(\eta-x_2)\right]\left(\frac{\lambda_2+\lambda_1}{\lambda_2-\lambda_1}F_2-\frac{2}{\lambda_2-\lambda_1}F_1+t_2+3\lambda_1t_1\right)+3t_2\lambda_1$$
$$\Delta F_2-a_2^2F_2=p_2(\eta)+\frac{1}{2}\left[\theta_1\delta(\eta-x_1)+\theta_2\delta(\eta-x_2)\right]\times$$
$$\times\left(\frac{-1+3\lambda_2^2}{\lambda_2(\lambda_2-\lambda_1)}F_1+\frac{2}{\lambda_2-\lambda_1}F_2+\frac{t_2}{\lambda_2}+3t_1\right)+3t_2. \quad (4)$$
$$\frac{\partial F_j}{\partial\eta}-b_1\left(F_j-F_{j1}^c\right)=0,\qquad \eta=0,$$
$$\frac{\partial F_j}{\partial\eta}+b_2\left(F_j-F_{j2}^c\right)=0,\qquad \eta=l. \quad (5)$$

Тут $p_1(\eta)=\left(d_1^*S_-(\eta-x_1)+\delta_1S_-(\eta-x_2)\right)F_1+\left(d_2S_-(\eta-x_1)+\delta_2S_-(\eta-x_2)\right)F_2$,

$p_2(\eta)=\left(d_3S_-(\eta-x_1)+\delta_3S_-(\eta-x_2)\right)F_1+\left(d_4S_-(\eta-x_1)+\delta_4S_-(\eta-x_2)\right)F_2$.

Вирази для сталих $d_1^*,d_2,d_3,d_4,\delta_1,\delta_2,\delta_3,\delta_4$, $\lambda_{1.2},a_j^2,F_{1j}^c,F_{2j}^c$ можна знайти у [13, с. 73].

Функції $F_j(\eta)$ у правих частинах отриманих рівнянь (4) уважатимемо відомими. Тоді кожне із цих диференціальних рівнянь розв'язуємо методом варіації сталої. Урахувавши крайові умови (5), визначимо сталі інтегрування, підставлянням яких у знайдені загальні розв'язки кожного з рівнянь отримаємо систему інтегральних рівнянь з інтегральними операторами Вольтерри та Фредгольма другого роду для визначення невідомих функцій $F_j(\eta)$:

$$F_j(\eta)+\left(g_{j0}\frac{e^{a_j\eta}}{g_{j1}}+\frac{e^{-a_j\eta}}{g_{j1}}\right)\left(\int_0^l p_j(s)cha_j(l-s)ds+\frac{b_1}{a_j}\int_0^l p_j(s)sha_j(l-s)ds\right)-$$
$$-\frac{1}{a_j}\int_0^\eta p_j(s)sha_j(\eta-s)ds=C_j(\eta)+C_{j1}(\eta)F_1(x_1)+C_{j2}(\eta)F_2(x_1)+ \quad (6)$$
$$+R_{j1}(\eta)F_1(x_2)+R_{j2}(\eta)F_2(x_2).$$

Позначення

$$C_j(\eta),C_{j1}(\eta),C_{j2}(\eta),R_{j2}(\eta),R_{j2}(\eta),g_{j0},g_{j1}$$

через їхню громіздкість тут не наведені. Невідомі значення функції на лініях зламів $F_j(x_1)$, $F_j(x_2)$ знаходимо із системи алгебраїчних рівнянь, яку отримуємо, беручи функції $F_j(\eta)$ у рівняннях (5) на лініях зламу $\eta=x_j$. Систему інтегральних рівнянь (5) розв'язуємо чисельно методом квадратурних формул [19]. Тоді зі співвідношень (3) знаходимо середню температуру $T_1(\eta)$ і температурний момент $T_2(\eta)$, необхідні для визначення термонапруженого стану в оболонці.

Термонапружений стан оболонки. Для визначення термонапруженого стану пологої призматичної оболонки зі зламами на лініях $\eta=x_j$ комплексну функцію напружень $\Phi(\eta,\xi)$ подамо через функцію напружень $\varphi(\eta,\xi)$ та прогин $w(\eta,\xi)$ у вигляді [15] $\Phi(\eta,\xi)=w(\eta,\xi)+i\varphi(\eta,\xi)/E_0$. Уважаємо, що на торцях $\eta=0$ і $\eta=l$ оболонка вільно оперта на жорсткі вертикальні діафрагми [9] та на них задана нульова температура, для цього у граничних умовах (2) b_j спрямуємо в нескінченність та покладемо $T_{j1}^c=0$ і $T_{j2}^c=0$, $j=1,2$. Тоді комплексна функція напружень $\Phi(\eta,\xi)$ визначається із ключового рівняння [12, с. 203]:

$$\left(\frac{d^4}{d\eta^4}+\frac{d^4}{d\eta^2d\xi^2}+\frac{d^4}{d\xi^4}\right)\Phi-i\sigma^2h^2\Delta_k\Phi=-\alpha_t\sigma^2h^2\Delta\left(iT_1+\nu_0T_2\right), \quad (7)$$

де

$$E_0=\frac{2Eh}{\sigma^2},\ \sigma^2=\frac{1}{h}\sqrt{3\left(1-\nu^2\right)},\ \nu_0^2=\frac{1+\nu}{3(1-\nu)},$$
$$\Delta_k=\left(-\theta_1\delta(\eta-x_1)-\theta_2\delta(\eta-x_2)\right)\frac{\partial^2}{\partial\xi^2},$$

за крайових умов:

за $\eta=0$ і $\eta=l$ $w=0$; $\frac{\partial^2w}{\partial\eta^2}=0$; $\frac{\partial^2\varphi}{\partial\eta^2}=0$; $\frac{\partial^2\varphi}{\partial\xi^2}=0$;

за $\xi=0$ і $\xi=d$ $w=0$; $\frac{\partial^2w}{\partial\xi^2}=0$; $\frac{\partial^2\varphi}{\partial\eta^2}=0$; $\frac{\partial^2\varphi}{\partial\xi^2}=0$, (8)

де $\xi=y/h$ – безрозмірна координата, $d=d_1/h$.

Для розв'язання рівняння (7) скористаємося формулами для перетворень Фур'є, задовольнивши крайові умови (8), отримаємо:

$$\bar{\Phi}^* = \frac{\sigma^2 h^2}{\left(\alpha_m^2+\beta_n^2\right)^2}\left[\frac{i}{l}\beta_n^2\left(\theta_1\Phi^*\left(x_1,\beta_n\right)\sin\alpha_m x_1+\theta_2\Phi^*\left(x_2,\beta_n\right)\sin\alpha_m x_2\right)-\alpha_t\left(i\Theta_1+\nu_0\Theta_2\right)\right], \quad (9)$$

тут $\bar{\Phi}^*\left(\alpha_m,\beta_n\right)=\dfrac{1}{dl}\int_0^d\int_0^l \varphi\left(\eta,\xi\right)\sin\alpha_m\eta\sin\beta_n\xi\, d\eta d\xi$,

$$\beta_n=\frac{n\pi}{d},\ \alpha_m=\frac{m\pi}{l},$$

$$\Theta_1\left(\alpha_m,\beta_n\right)=\frac{1}{d\beta_n}\left(\left(-1\right)^{n+1}+1\right)\left(\frac{1}{l}\int_0^l \Upsilon_1\left(\eta\right)\sin\alpha_m\eta d\eta+\theta_1 T_2\left(x_1\right)\sin\alpha_m x_1+\theta_2 T_2\left(x_2\right)\sin\alpha_m x_2\right),$$

$$\Theta_2\left(\alpha_m,\beta_n\right)=\frac{1}{d\beta_n}\left(\left(-1\right)^{n+1}+1\right)\left(\frac{1}{l}\int_0^l \Upsilon_2\left(\eta\right)\sin\alpha_m\eta d\eta+3\theta_1 T_1\left(x_1\right)\sin\alpha_m x_1+3\theta_2 T_1\left(x_2\right)\sin\alpha_m x_2\right),$$

$$\Upsilon_1\left(\eta\right)=\eta_1^+\left(T_1(\eta)-t_1\right)+\left[\left(\eta_2^+-\eta_1^+\right)\left(T_1(\eta)-t_1\right)+\left(\eta_2^--\eta_1^-\right)\left(T_2(\eta)-t_2\right)\right]S_-(\eta-x_1)+$$

$$+\left[\left(\eta_3^+-\eta_2^+\right)\left(T_1(\eta)-t_1\right)+\left(\eta_3^--\eta_2^-\right)\left(T_2(\eta)-t_2\right)\right]S_-(\eta-x_2)+\eta_1^-\left(T_2(\eta)-t_2\right),$$

$$\Upsilon_2\left(\eta\right)=3(1+\eta_1^+)\left(T_2(\eta)-t_2\right)+3\left[\left(\eta_2^+-\eta_1^+\right)\left(T_2(\eta)-t_2\right)+\left(\eta_2^--\eta_1^-\right)\left(T_1(\eta)-t_1\right)\right]S_-(\eta-x_1)+$$

$$+3\left[\left(\eta_3^+-\eta_2^+\right)\left(T_2(\eta)-t_2\right)+\left(\eta_3^--\eta_2^-\right)\left(T_1(\eta)-t_1\right)\right]S_-(\eta-x_2)+3\eta_1^-\left(T_1(\eta)-t_1\right)+3t_2.$$

Застосовуємо до (9) обернене перетворення Фур'є по α_m і отримаємо:

$$\Phi^*\left(\eta,\beta_n\right)=\Phi^*\left(x_1,\beta_n\right)iH_1\left(\eta,\beta_n\right)+\Phi^*\left(x_2,\beta_n\right)iH_2\left(\eta,\beta_n\right)-i\Gamma_1\left(\eta,\beta_n\right)-\Gamma_2\left(\eta,\beta_n\right), \quad (10)$$

де $\mathrm{H}_1\left(\eta,\beta_n\right)=\dfrac{1}{l}\sigma^2h^2\sum_{m=1}^{\infty}\dfrac{\beta_n^2}{\left(\alpha_m^2+\beta_n^2\right)^2}\theta_1\sin\alpha_m x_1\sin\alpha_m\eta,$

$$\mathrm{H}_2\left(\eta,\beta_n\right)=\frac{1}{l}\sigma^2h^2\sum_{m=1}^{\infty}\frac{\beta_n^2}{\left(\alpha_m^2+\beta_n^2\right)^2}\theta_2\sin\alpha_m x_2\sin\alpha_m\eta,$$

$$\Gamma_1(\eta,\beta_n)=\alpha_t\sigma^2h^2\sum_{m=1}^{\infty}\frac{1}{\left(\alpha_m^2+\beta_n^2\right)^2}\Theta_1\sin\alpha_m\eta,$$

$$\Gamma_2(\eta,\beta_n)=\alpha_t\sigma^2h^2\sum_{m=1}^{\infty}\frac{1}{\left(\alpha_m^2+\beta_n^2\right)^2}\nu_0\Theta_2\sin\alpha_m\eta.$$

Щоб визначити $\Phi^*\left(x_1,\beta_n\right)$ та $\Phi^*\left(x_2,\beta_n\right)$, треба в рівність (9) підставити точки x_1 і x_2 та розв'язати відповідну систему рівнянь із двома невідомими. Розв'язками такої системи будуть вирази:

$$\Phi^*\left(x_1,\beta_n\right)=-\chi_1-i\chi_2,\ \Phi^*\left(x_2,\beta_n\right)=-\chi_3-i\chi_4, \quad (11)$$

де $\chi_1=\dfrac{\xi_1\rho_1-\xi_2\rho_2}{\rho_1^2+\rho_2^2},\ \chi_2=\dfrac{\xi_2\rho_1+\xi_1\rho_2}{\rho_1^2+\rho_2^2},\ \chi_3=\dfrac{\xi_3\rho_1-\xi_4\rho_2}{\rho_1^2+\rho_2^2},$

$$\chi_4=\frac{\left(\xi_4\rho_1+\xi_3\rho_2\right)}{\rho_1^2+\rho_2^2},$$

$$\xi_1=\nu_0\Gamma_2\left(x_1,\beta_n\right)+\Gamma_1\left(x_1,\beta_n\right)\mathrm{H}_2\left(x_2,\beta_n\right)-\Gamma_1\left(x_2,\beta_n\right)\mathrm{H}_2\left(x_1,\beta_n\right),$$

$$\xi_2=\Gamma_1\left(x_1,\beta_n\right)-\nu_0\Gamma_2\left(x_1,\beta_n\right)\mathrm{H}_2\left(x_2,\beta_n\right)+\nu_0\Gamma_2\left(x_2,\beta_n\right)\mathrm{H}_2\left(x_1,\beta_n\right),$$

$$\xi_3=\nu_0\Gamma_2\left(x_2,\beta_n\right)+H_1\left(x_1,\beta_n\right)\Gamma_1\left(x_2,\beta_n\right)-\Gamma_1\left(x_1,\beta_n\right)H_1\left(x_2,\beta_n\right),$$

$$\xi_4=\Gamma_1\left(x_2,\beta_n\right)-\nu_0H_1\left(x_1,\beta_n\right)\Gamma_2\left(x_2,\beta_n\right)+\nu_0H_1\left(x_2,\beta_n\right)\Gamma_2\left(x_1,\beta_n\right),$$

$$\rho_1=1-\mathrm{H}_1\left(x_1,\beta_n\right)\mathrm{H}_2\left(x_2,\beta_n\right)+\mathrm{H}_1\left(x_2,\beta_n\right)\mathrm{H}_2\left(x_1,\beta_n\right),$$

$$\rho_2=\mathrm{H}_1\left(x_1,\beta_n\right)+\mathrm{H}_2\left(x_2,\beta_n\right).$$

Підставимо вирази (11) у формулу (10) і отримаємо

$$\Phi^*\left(\eta,\beta_n\right)=\chi_2\mathrm{H}_1\left(\eta,\beta_n\right)+\chi_4\mathrm{H}_2\left(\eta,\beta_n\right)-\nu_0\Gamma_2\left(\eta,\beta_n\right)-i\left(\chi_1\mathrm{H}_1\left(\eta,\beta_n\right)+\chi_3\mathrm{H}_2\left(\eta,\beta_n\right)+\Gamma_1\left(\eta,\beta_n\right)\right),$$

тоді прогин і функція напружень матимуть вигляд:

$$w=\sum_{n=1}^{\infty}\mathrm{Re}\,\Phi^*\left(\eta,\beta_n\right)\sin\beta_n\xi,\ \varphi=E_0\sum_{n=1}^{\infty}\mathrm{Im}\,\Phi^*\left(\eta,\beta_n\right)\sin\beta_n\xi. \quad (12)$$

Зусилля та моменти з урахуванням (12) визначаємо за формулами:

$$N_1=E_0\frac{\partial^2\varphi}{h^2\partial\xi^2},$$

$$M_1=-D_2\left[\frac{1}{h^2}\left(\frac{\partial^2w}{\partial\eta^2}+\frac{\partial^2w}{\partial\xi^2}-\left(1-\nu\right)\frac{\partial^2w}{\partial\xi^2}\right)+\frac{\alpha_t\left(1+\nu\right)}{h}T_2\right],$$

$$N_2=E_0\frac{\partial^2\varphi}{h^2\partial\eta^2},$$

$$M_2=-D_2\left[\frac{1}{h^2}\left(\nu\left(\frac{\partial^2w}{\partial\eta^2}+\frac{\partial^2w}{\partial\xi^2}\right)+\left(1-\nu\right)\frac{\partial^2w}{\partial\xi^2}\right)+\frac{\alpha_t\left(1+\nu\right)}{h}T_2\right].$$

Аналіз числових результатів. Розглядали тонку призматичну оболонку з безрозмірними розмірами $l=1$, $d=1$, $x_2=0{,}66$ $x_1=0{,}33$ та кутами зламу $\theta_1=\theta_2=0{,}1$, коли температура середовища дорівнює нулю на нижній лицевій поверхні оболонки $z=-h$ і на торцях $\eta=0$ та $\eta=1$ $\left(t_c^-=t_c^1=t_c^2=0^\circ C\right)$ та дорівнює одному градусу на верхній поверхні $z=+h$ $\left(t_c^+=1^\circ C\right)$. Уважали, що коефіцієнти тепловіддачі на нижніх поверхнях $z=-h$ трьох елементів однакові й рівні $\mu_1^-=\mu_2^-=\mu_3^-=50$; на верхній лицевій поверхні $z=h$ другого елемента оболонки коефіцієнт тепловіддачі рівний $\mu_2^+=100$. Розглядали різні значення коефіцієнтів тепловіддачі на верхній лицевій поверхні першого елемента: ($\mu_1^+=100$; $\mu_1^+=95$; $\mu_1^+=90$) і різні значення коефіцієнтів тепловіддачі на верхній лицевій поверхні третього елемента: ($\mu_3^+=100$; $\mu_3^+=95$; $\mu_3^+=100$). На рис. 2–8 криві ілюструють розподіли температурних характеристик, прогину, моментів, зусиль вздовж лінії $\xi=0{,}5$. Суцільна крива відповідає однаковим коефіцієнтам тепловіддачі на верхніх і нижніх лицевих поверхнях, ; штрихова крива – зменшенню тепловіддачі на верхніх поверхнях першого й третього елементів на п'ять відсотків ($\mu_1^+=\mu_3^+=95$); штрихпунктирна крива – зменшенню тепловіддачі на верхній поверхні першого елемента на десять відсотків ($\mu_1^+=90$).

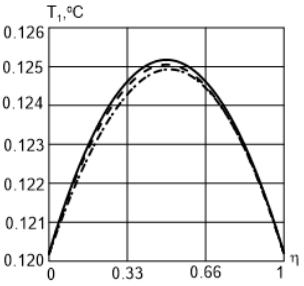


Рис. 2. Розподіл середньої температури T_1 за зміни коефіцієнтів тепловіддачі

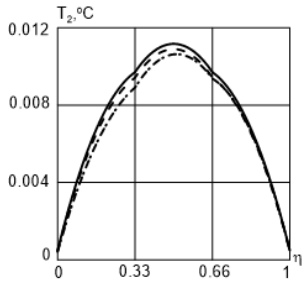


Рис. 3. Розподіл температурного моменту T_2 за зміни коевіцієнтів тепловіддачі

За однакових коефіцієнтів тепловіддачі з верхньої поверхні трьох елементів оболонки $\mu_1^+ = \mu_2^+ = \mu_3^+ = 100$ (суцільна крива) максимальне значення середньої температури T_1 досягається посередині другого елемента, а температурний момент T_2 досягає локальних мінімумів на лініях зламу. Коли коефіцієнти тепловіддачі з верхніх поверхонь першого й третього елементів оболонки стають на п'ять відсотків меншими (штрихова крива), то T_1 і T_2 зменшуються. Коли коефіцієнт тепловіддачі першого елемента верхньої лицевої поверхні оболонки стає на десять відсотків меншим, середня температура й температурний момент дальше зменшуються, їхні максимуми зміщуються в напрямку третього елемента оболонки.

На рисунку 4 спостерігається зменшення абсолютної величини нормованого прогину $w^* = 10^5 w/(\alpha_t t_c^- h)$ оболонки (суцільна та штрихова криві) зі зменшенням коефіцієнтів тепловіддачі на верхній поверхні її першого й третього елементів на п'ять одиниць. Тоді прогин досягає локальних мінімумів на лініях зламів і локальних максимумів на кожному із трьох елементів оболонки та є симетричним щодо її середини $\eta = 0{,}5$. Зі зменшенням коефіцієнта тепловіддачі верхньої поверхні першого елемента до $\mu_1^+ = 90$ (штрихпунктирна крива) абсолютна величина прогину зменшується і її локальний мінімум на першому елементі вдвічі менший, ніж на третьому. Якщо порівняти прогин (суцільна крива) на рисунку 4 й аналогічні прогини в роботі [11] для двохелементної призматичної оболонки за однакових коефіцієнтів тепловіддачі з лицевих поверхонь та різних кутів зламу, бачимо, що якісна поведінка прогину на цих рисунках подібна. На лінії зламу прогин зменшується.

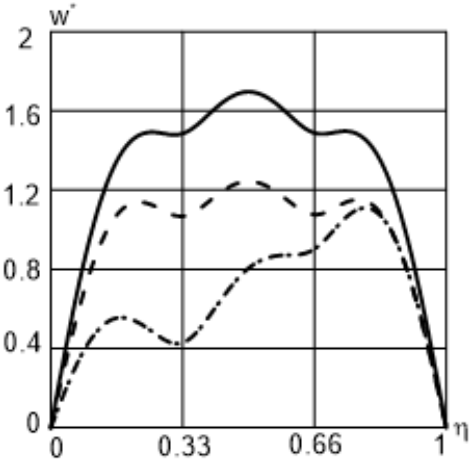


Рис. 4. Розподіл нормованого прогину w^* за зміни коефіцієнтів тепловіддачі

Нормований згинальний момент $M_1^* = 10^3 M_1 h/(\alpha_t t_c^- D_2)$ (рис. 5) та величина нормованого зусилля $N_1^* = 10^2 N_1 h/(\alpha_t t_c^- E_0)$ (рис. 7), коли коефіцієнти тепловіддачі оболонки на кожній з лицевих поверхонь однакові (суцільна крива), спадають, досягаючи мінімуму на лінії $\eta = 0{,}5$. Зі зменшенням коефіцієнтів тепловіддачі на верхніх лицевих поверхнях першого й третього елементів оболонки вони збільшуються (штрихова й штрихпунктирна криві). Розподіл нормованого моменту $M_2^* = 10^3 M_2 h/D_2$ (рис. 6) та нормованого зусилля $N_2^* = 10^2 N_2 h^2/(\alpha_t t_c^- E_0)$ (рис. 8) якісно аналогічний до розподілу моменту M_1^* і зусилля N_1^*. Максимальних значень моменти досягають на торцях.

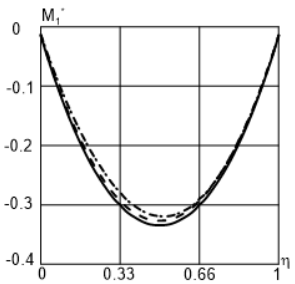


Рис. 5. Розподіл нормованого моменту M_1^* за зміни коефіцієнтів тепловіддачі

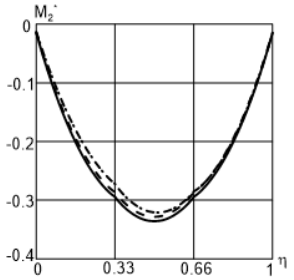


Рис. 6. Розподіл нормованого моменту M_2^* за зміни коефіцієнтів тепловіддачі

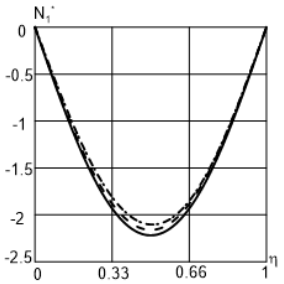


Рис. 7. Розподіл нормованого зусилля N_1^* за зміни коефіцієнтів тепловіддачі

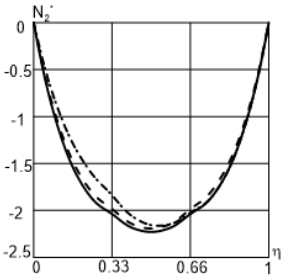


Рис. 8. Розподіл нормованого зусилля N_2^* за зміни коефіцієнтів тепловіддачі

Висновки. Дослідження напружено-деформованого стану оболонок зі зламами у працях [1–9] проведено для силового навантаження. Для випадку рівності коефіцієнтів тепловіддачі з лицевих поверхонь досліджено напружено-деформований стан складеної пологої оболонки сферичної форми та призматичної оболонки, складеної із двох однакових елементів, у працях [10; 11]. За різних коефіцієнтів тепловіддачі з лицевих поверхонь напружено- деформований стан призматичної оболонки, складеної із двох елементів, проведено у праці [12]. У даній роботі вивчено ефект впливу зламів на термонапружений стан у тонкій пологій складеній із трьох плоских елементів призматичній оболонці за конвективного теплообміну з навколишнім середовищем на лицевих поверхнях. На кожній із шести лицевих поверхонь оболонки різні коефіцієнти тепловіддачі. Задачу теплопровідності для оболонки зі зламами зведено до системи інтегральних рівнянь

з інтегральними операторами Вольтерри та Фредгольма другого роду, із застосуванням методу варіації сталої. Розв'язок задачі термонапружності знайдено за допомогою комплексної функції напружень, яку подано через функцію напружень та прогин. На торцях оболонка зі зламами вільно оперта на жорсткі вертикальні діафрагми. Торці, на які виходять злами серединної поверхні, теплоізольовані, а на інших задана нульова температура. Функції напружень та прогину знайдено завдяки скінченним інтегральним перетворенням Фур'є. Наведено результати числового аналізу розподілу середньої температури, температурного моменту, прогину, моментів і зусиль за різних значень коефіцієнта тепловіддачі на верхніх лицевих поверхнях першого й третього елементів оболонки. За зменшення коефіцієнтів тепловіддачі на крайніх елементах середня температура, температурний момент, згинні моменти й абсолютні величини прогину й зусиль спадають зі зміщенням максимальної їх величини в бік елемента, де коефіцієнти тепловіддачі не змінювали. Варто зауважити, що вздовж зламів поведінка розподілу температурних характеристик, прогину, зусиль та моментів у працях [10–12] та даній роботі має якісно аналогічний характер.

ЛІТЕРАТУРА

1. Garncarek R. Efekty nieciaglosci katowych powierzchni w statyce powlok cienkosciennych. *Archivum budowy maszyn*. Warszawa, 1975. № 2. S. 163–183.
2. Grishin V. A., Popov G. Ya., Reut V. V. Analysis of box-like shells of rectangular cross-section. *Journal of Applied Mathematics and Mechanics.* 1990. 54. № 4. P. 501–507.
3. Моссаковський В. І., Куліков Д. В. Метод однорідних розв'язків коробчастих оболонок при динамічному навантаженні. Доп. АН УРСР. Сер. А. 1987. С. 24–27.
4. Гайдайчук В. В., Кошевий О. О. Чисельне рішення задач оптимального проектування при обмеженні власних частот коливанання пологої оболонки зі зламами. *Сучасні проблеми архітектури та містобудування.* Архітектура будівель і споруд. 2018. Вип. 51. С. 416–425.
5. Кривенко О. П., Лізунов П. П., Ворона Ю. В., Калашніков О. Б. Використання моментної схеми скінченних елементів при дослідженні тонких пружних оболонок неоднорідної структури. *Управління розвитком складних систем.* Інформаційні технології проєктування. 2023. Вип. 53. С. 52–62.
6. Derveni F., Gueissaz W., Yan D., Reis P. M. Probabilistic buckling of imperfect hemispherical shells containing a distribution of defects. *Philos. Trans. R.* 2023. Soc. A. 381 (2244). P. 20220298.
7. Gristchak V. Z., Hryshchak D. V., Dyachenko N. M., Sanin A. F., Sukhyy K. M. Bifurcation state and rational design of three-layer reinforced compound cone-cylinder shell structure under combined loading. *Space Science and Technology*. 2023. 29. № 6 (145). P. 26–41.
8. Gristchak V. Z., Hryshchak D. V., Dyachenko N. M., Baburov V. V. The influence of the Gaussian curvature sign of the compound shell structure's middle surface on local and overall buckling under combined loading. *Space Science and Technology*. 2022. 28. № 4 (137). С. 31–38.
9. Назаров А.Г. Некоторые контактные задачи теории оболочек. *Доклады АН Арм. ССР.* 1948. Т. 9. № 2. С. 61–66.
10. Хапко Б. С. Термонапруження складеної пологої оболонки сферичної форми. *Фізико-хімічна механіка матеріалів*. 2001. № 6. С. 124–126.
11. Хапко Б. С. Пологa призматична оболонка в нерівномірному температурному полі. *Фізико-хімічна механіка матеріалів.* 2005. № 2. С. 33–38.
12. Хапко Б. С. Термонапружений стан двохелементної призматичної оболонки з різними коефіцієнтами тепловіддачі. *Вісник Запорізького національного університету. Фізико математичні науки. Математичне моделювання і прикладна механіка.* Запоріжжя : Запорізький національний університет, 2015. № 1. С. 199–210.
13. Швец Р. Н., Хапко Б. С., Чиж А. И. Уравнения теплопроводности для оболочек с изломами при переменных коэффициентах теплоотдачи. *Теоретическая и прикладная механика*. 2010. Вып. 1 (47). С. 69–76.
14. Підстригач Я. С., Коляно Ю. М. Температурне поле в тонких пластинках при змінному коефіцієнті тепловіддачі з бокових поверхонь. *Доп. АН УРСР.* Сер. А. 1971. № 1. С. 75–78.
15. Подстригач Я. С., Швец Р. Н. Термоупругость тонких оболочек. Киев : Наукова думка, 1978. 344 с.
16. Подстригач Я. С., Коляно Ю. М., Громовык В. И., Лозбень В. Л. Термоупругость тел при переменных коэффициентах теплоотдачи. Киев : Наук. думка, 1977. 158 с.
17. Sugano Y., Chiba R., Hirose K., Takahashi K. Material design for reduction of thermal stress in a functionally graded material rotating disk. *JSME international journal.* Series A. 2004. 47. № 2. P. 189–197.
18. Хапко Б. С., Чиж А. І. Про вплив змінних коефіцієнтів тепловіддачі на термонапруження у скінченній циліндричній оболонці. *Математичні методи та фізичко-механічні поля.* 2014. 57**.** № 2. С. 195–203.
19. Верлань А. Ф., Сизиков В. С. Методы решения интегральных уравнений с программами для ЭВМ. Киев : Наук. думка, 1978, 292 с.

REFERENCES

1. Garncarek R. (1975). Efekty nieciaglosci katowych powierzchni w statyce powlok cienkosciennych. *Archivum budowy maszyn*. Warszawa, № 2. P. 163–183.
2. Grishin V. A., Popov G. Ya., Reut V. V. (1990). Analysis of box-like shells of rectangular cross-section. Journal of Applied Mathematics and Mechanics. 54, № 4. P. 501–507.
3. Mosakovs'kyy V. I., Kulykov D. V. (1987). Metod odnoridnykh rozv'yazkiv korobchastykh obolonok pry dynamichnomu navantazhenni. Dop. AN URSR. Ser. A. S. 24–27.
4. Haidaichuk V. V, Koshevyi O. O. (2018). Chyselne rishennia zadach optymalnoho proektuvannia pry obmezhenni vlasnykh chastot kolyvanannia polohoi obolonky zi zlamamy. *Suchasni problemy arkhitektury ta mistobuduvannia.* Arkhitektura budivel i sporud. Vyp. 51. S. 416–425.
5. Kryvenko O. P., Lizunov P. P., Vorona Yu. V., Kalashnikov O. B. (2023). Vykorystannia momentnoi skhemy skinchennykh elementiv pry doslidzhenni tonkykh pruzhnykh obolonok neodnoridnoi struktury. *Upravlinnia rozvytkom skladnykh system.* Informatsiini tekhnolohii proiektuvannia. Vyp. 53. S. 52–62.
6. Derveni F., Gueissaz W., Yan D., and Reis P. M. (2023). Probabilistic buckling of imperfect hemispherical shells containing a distribution of defects. *Philos. Trans. R.* Soc. A, 381 (2244), p. 20220298.
7. Gristchak V. Z., Hryshchak D. V., Dyachenko N. M., Sanin A. F., Sukhyy K. M. (2023). Bifurcation state and rational design of three-layer reinforced compound cone-cylinder shell structure under combined loading. *Space Science and Technology.* 29, № 6 (145). P. 26–41. https://doi.org/10.15407/knit2023.06.026
8. Gristchak V. Z., Hryshchak D. V., Dyachenko N. M., Baburov V. V. (2022). The influence of the Gaussian curvature sign of the compound shell structure's middle surface on local and overall buckling under combined loading. *Space Science and Technology.* 28, № 4 (137). C. 31–38. https://doi.org/10.15407/knit2022.04.031
9. Nazarov A. H. (1948). Nekotorye kontaktnye zadachi teorii obolochek. *Doklady AN Arm. SSR.* V. 9, № 2. Pp. 61–66.
10. Khapko B. S. (2001). Thermal stress of a folded shallow spherical shell. Fhysicochemical mechanics of materials. № 6. Pp. 124–126.
11. Khapko B. S. (2005). Shallow Prismatic Shell in a Nonuniform Temperature Field. *Fhysicochemical mechanics of materials*. № 2. Pp. 33–38.
12. Khapko B. S. (2015). Termonapruzhenyy stan dvokholementnoyi pryzmatychnoyi obolonky z riznymy koefitsiyentamy teploviddachi. *Visnyk Zaporiz'koho natsional'noho universytetu. Fizyko matematychni nauky. Matematychne modelyuvannya ta prykladna mekhanika.* – Zaporizhzhya: Zaporiz'kyy natsional'nyy universytet, № 1. S. 199–210.
13. Shvets R. M., Khapko B. S., Chyzh A. I. (2010) Heat conduction equations for shells having breaks with variable heat transfer coefficients. *Theoretical and applied mechanics*. Issue 1 (47). Pp. 69–76.
14. Pidstryhach Ya. S., and Kolyano Yu. M. (1971). Temperaturne pole v tonkykh plastynkakh pry zminnomu koefitsiyenti teploviddachi z bokovykh poverkhon'. *Dop. AN URSR,* Ser. A. № 1. Pp. 75–78.
15. Podstrigach Ya. S., and Shvec R. N. (1978). Termouprugost tonkich obolochek [Thermoelasticity of thin shells], Naukova dumka, Kyiv, Ukraine.
16. Podstrigach Ya. S., Kolyano Yu. M., Gromovyk V. I., and Lozben V. L. (1977). Termouprugost tel pri peremennyx koefficientax teplootdachi [Thermoelasticity of Bodies with Variable Heat-Transfer Coefficients], Naukova dumka, Kyiv, Ukraine.
17. Sugano Y., Chiba R., Hirose K., Takahashi K. (2004). Material design for reduction of thermal stress in a functionally graded material rotating disk. *JSME international journal,* Series A. 47, № 2. P. 189–197.
18. Khapko B. S., Chyzh A. I. (2014). On effect of variable heat exchange coefficients on thermal stresses in finite cylindrical shell. *Mathematical methods and physicomechanical fields.* 57, № 2. Pp. 195–203.
19. Verlan A. F., and Sizikov V.S. (1978). Metody resheniya integralnyh uravnenij s programmami dlya EVM [Methods for the Solution of Integral Equations using Computer Programs], Naukova dumka, Kyiv, Ukraine.

Дата першого надходження рукопису до видання: 26.08.2025
Дата прийнятого до друку рукопису після рецензування: 23.09.2025
Дата публікації: 31.12.2025

РОЗДІЛ ІІ. ІНЖЕНЕРІЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

УДК 004.02
DOI https://doi.org/10.26661/2786-6254-2025-2-05

ОГЛЯД ПІДХОДІВ ДО ОРГАНІЗАЦІЇ БІЗНЕС-ЛОГІКИ В ПОБУДОВІ МІКРОСЕРВІСНИХ СИСТЕМ

Бейрак Д. Я.
аспірант кафедри інженерії програмного забезпечення
Державний університет «Житомирська політехніка»
вул. Чуднівська, 103, Житомир, Україна
orcid.org/0009-0006-5089-3603
dm.beirak@gmail.com

Вакалюк Т. А.
доктор педагогічних наук, професор,
завідувач кафедри інженерії програмного забезпечення
Державний університет «Житомирська політехніка»
вул. Чуднівська, 103, Житомир, Україна
orcid.org/0000-0001-6825-4697
tetianavakaliuk@gmail.com

Ключові слова: *мікросервісна архітектура, бізнес-логіка, предметно орієнтоване проєктування, DDD, event sourcing, CQRS, патерни.*

Натепер питання побудови архітектури мікросервісів стає все більш актуальним, оскільки даний тип архітектури дозволяє проєктувати системи з низькою зв'язністю, які мають низку переваг перед монолітами: можливість горизонтального масштабування, краще розділення системи на складові частини (сервіси), кожен окремий з яких простіше розвивати та підтримувати, можливість більш ефективного використання ресурсів та інші. Зазначені вище та низка інших причин приводять до зростання популярності такого типу архітектури в індустрії, що позначається на виборі архітекторів та інженерів програмного забезпечення стосовно впровадження мікросервісної архітектури як у побудові нових систем, так і як вектора розвитку успадкованих монолітних систем, які все частіше переписуються на мікросервіси. Проблематика, що стосується питань проєктування мікросервісних систем, має велику кількість різноманітних аспектів, одним із них є вибір організації бізнес-логіки разом із низкою супутніх патернів, технологій та інструментів. Вплив такого вибору неможливо переоцінити: бізнес-логіка є реалізацією предметної області, у якій працює бізнес, тому вибір відповідних патернів і підходів до її організації має прямий вплив не тільки на якість реалізації системи, а й на вартість її розширення та підтримки в майбутньому. У статті розглядаються методи, принципи й інструменти, призначені для організації бізнес-логіки в мікросервісних системах, розглядаються патерни, призначені для використання в умовах простих і складних доменів, підходи до організації роботи з командами та запитами в системах, що використовують події. Значна увага приділяється парадигмі предметно орієнтованого проєктування, як найперспективнішій у застосуванні під час розроблення систем зі складною предметною областю.

APPROACHES TO BUSINESS LOGIC IMPLEMENTATION IN MICROSERVICE SYSTEMS IN THE SCIENTIFIC LITERATURE

Beirak D. Ya.
Postgraduate Student at the Department of Software Engineering
Zhytomyr Polytechnic State University
Chudnivska str., 103, Zhytomyr, Ukraine
orcid.org/0009-0006-5089-3603
dm.beirak@gmail.com

Vakaliuk T. A.
Doctor of Pedagogical Sciences, Professor,
Head of the Department of Software Engineering
Zhytomyr Polytechnic State University
Chudnivska str., 103, Zhytomyr, Ukraine
orcid.org/0000-0001-6825-4697
tetianavakaliuk@gmail.com

Key words: *microservice architecture, business logic, domain-driven design, DDD, event sourcing, CQRS, patterns.*

Currently, the issue of building a microservice architecture is becoming increasingly relevant, as this type of architecture allows you to design systems with low coupling, which have a number of advantages over monoliths: the possibility of horizontal scaling, better division of the system into components (services), each of which is easier to develop and maintain, the possibility of more efficient use of resources, etc. The above-mentioned and a number of other reasons lead to the growing popularity of this type of architecture in the industry, which affects the choice of software architects and engineers regarding the implementation of microservice architecture both when designing new systems and as a vector of development for legacy monolithic systems, which are increasingly being migrated to microservices. The issues related to the design of microservice systems include a large number of different aspects, and the choice of business logic implementation is one of them, along with a number of related patterns, technologies, and tools. The impact of such a choice cannot be overestimated: business logic is the implementation of the business domain in code, and therefore the choice of appropriate patterns and approaches has a direct impact not only on the quality of the system implementation, but also on the cost of its expansion and support in the future. This paper examines methods, principles, and tools designed to organize business logic in microservice systems, considers patterns designed to be used in simple and complex domains, approaches to organizing work with commands and requests in event-driven systems. Considerable attention is paid to the paradigm of domain-driven design, as the most promising in application when developing systems within complex domains.

Вступ. Натепер усе більше нових проєктів створюється на основі мікросервісної архітектури, а також усе більше успадкованих систем переводяться на даний тип архітектури. Велике значення має питання організації бізнес-логіки, тобто вибору патернів і підходів описання предметної області у вихідному коді. Від даного вибору залежить не тільки якість архітектури системи, а також вартість її розширення та підтримки в майбутньому, якість комунікації розробників і менеджменту. Дані чинники роблять актуальними дослідження наявних патернів і підходів реалізації бізнес-логіки в контексті побудови мікросервісної архітектури. **Огляд літератури.** Проблеми, пов’язані з організацією бізнес-логіки в системах з мікросервісною архітектурою, у науковій літературі та публікаціях розглядали різні автори та вчені. Зокрема, загальні проблеми та патерни описували Кріс Річардсон (Chris Richardson), Сем Ньюман (Sam Newman) і Мартін Фаулер (Martin Fowler). Питання організації бізнес-логіки з використанням предметно орієнтованого проєкту-

вання присутні у працях Еріка Еванса (Eric Evans) та Влада Хононова (Vlad Khononov). Застосування даного підходу у функціональній парадигмі висвітлювали Скотт Влашин (Scott Wlaschin) і Уберто Барбіні (Uberto Barbini). Також питання, пов'язані з організацією бізнес-логіки, розглядали Штефан Капферер (Stefan Kapferer), Олаф Циммерманн (Olaf Zimmermann), Владімір Хоріков (Vladimir Khorikov), Джерфі Палермо (Jeffrey Palermo), Дерек Колі (Derek Colley), Клейр Стейнер (Clare Stanier), Штефан Лізер (Stefan Lieser), Стенлі Ліма (Stanley Lima), Жаймі Корія (Jaime Correia), Філіпі Араужю (Filipe Araujo), Жорже Кардозо (Jorge Cardoso), Ішмаель Мота (Ismael Mota), Міхіел Оверейм (Michiel Overeem), Мартен Шпур (Marten Spoor), Слінге Янсен (Slinger Jansen), Шак Брінкемпер (Sjaak Brinkkemper), Озан Оскан (Ozan Özkan), Ондер Бабур (Önder Babur), Марк ван ден Бранд (Mark van den Brand), Метаяс Верейс (Mathias Verraes), Ребека Верфс-Брок (Rebecca Wirfs-Brock), Хуля Вулар (Hulya Vural), Мурат Коюнджу (Murat Koyuncu) та інші.

Метою статті є детальний аналіз підходів і патернів організації бізнес-логіки в контексті побудови мікросервісної архітектури.

Результати. Розроблення бізнес-логіки в системі з мікросервісною архітектурою ускладнене тим, що вона є розподіленою між багатьма сервісами. Кріс Річардсон (Chris Richardson) зазначає необхідність подолання двох викликів, пов'язаних з такого типу розробленням: неможливість прямого виклику коду, що міститься в іншому сервісі, та транзакційні обмеження, які накладає розподілена природа мікросервісної архітектури [1]. Перший виклик також може впливати на необхідність вирішення проблеми створення розподіленої моделі й ускладнювати процес тестування [2]. Другий виклик обмежує можливість застосування ACID-транзакцій лише в рамках кожного окремо взятого сервісу, унеможливлює такий тип транзакційності між ними [1].

Дані обмеження впливають на підхід до організації бізнес-логіки та вибору архітектурного стилю сервісу. Традиційна для монолітних систем шарова архітектура (layered architecture) не враховує взаємодії системи з кількома сховищами даних та іншими системами, а також за її використання часто виникає залежність шару бізнес-логіки від шару доступу до даних (хоча зворотне спрямування даної залежності можливе) [1, 3]. Можливою альтернативою шаровій архітектурі є гексагональна (також відома під назвою «порти й адаптери»), а також її варіації: onion-архітектура та «чиста» архітектура [1, 4]. У своїй основі кожна з них має принцип інверсії залежностей, що насамперед диктує залежність коду доступу до даних від ядра сервісу – бізнес-логіки, а також передбачає винесення іншого інфраструктурного коду та коду інтерфейсу користувача на периферію застосунку [4; 5].

Сервіс, побудований із застосуванням гексагональної архітектури, має вхідні і вихідні порти, а також вхідні і вихідні адаптери. Порти використовуються для взаємодії бізнес-логіки із зовнішніми частинами застосунку: наприклад, вхідний порт може являти собою публічні методи для доступу до функціональності бізнес-логіки, що в об'єктно орієнтованих мовах програмування зазвичай реалізується через відповідний інтерфейс, а вихідний порт може являти собою інтерфейс репозиторія – колекцію операцій доступу до даних. Адаптери розташовуються на периферії сервісу та використовуються для взаємодії із зовнішніми сервісами та частинами системи. Так, вхідні адаптери приймають запити й обробляють їх шляхом виклику відповідних вхідних портів. Прикладом вхідного адаптера може слугувати контролер з кінцевими точками REST API (REST API endpoints) або клієнт брокера повідомлень (message broker client). Вихідні адаптери викликаються вихідним портом (або реалізовують його, якщо він є інтерфейсом), та викликають зовнішній сервіс або іншу частину системи. Прикладами вихідних портів можуть бути класи, що реалізують операції доступу до даних, публікують події або проксують зовнішні виклики [1].

У класичній роботі Мартіна Фаулера [6] розглянуто три можливі патерни організації бізнес-логіки: сценарій транзакції (transaction script), доменну модель (domain model) та табличний модуль (table module).

У мікросервісній архітектурі, у його початковому вигляді, застосовується лише сценарій транзакції. Суть цього патерну полягає в організації бізнес-логіки у формі колекції процедур (методів, функцій), кожна з яких містить код для обробки однієї бізнес-транзакції (запиту, операції). Стан системи водночас зберігається окремо. Даний підхід є великою мірою процедурним, тому підходить лише для проєктування сервісів із простою бізнес-логікою [1, 6].

Коли ж бізнес-логіка все ще залишається простою, але дані системи є досить складними, щоб ускладнити використання сценарію транзакцій, можна застосувати патерн активного запису (active record). Він дозволяє абстрагувати рядок у таблиці, інкапсулює до нього доступ за допомогою CRUD-операцій (операції створення, читання, оновлення, видалення), тим самим спрощує відображення об'єкта в оперативній пам'яті на схему бази даних [6, 7]. В об'єкти, що реалізовані як активний запис, також можна помістити бізнес-логіку, проте її відсутність перетворює їх на анемічні доменні моделі (anemic domain model),

що є антипатерном для систем із крупними доменами [7, 8].

За необхідності імплементації більш складної бізнес-логіки, у монолітних системах, написаних у стилі об'єктно орієнтованого програмування (далі – ООП), можливе застосування доменної моделі [6], проте в разі мікросервісної архітектури більше підходить предметно орієнтоване проєктування (domain-driven design, DDD) [1], використання якого також можливе у програмуванні у функціональній та інших парадигмах [9, 10].

Складники DDD зосереджені навколо побудови моделей предметної області. Патерни, що застосовуються під час побудови застосунків з використанням DDD, поділяються на стратегічні і тактичні [7].

До стратегічних відносять такі патерни, як піддомен (subdomain), єдина мова (ubiquitous language), обмежений контекст (bounded context) і контекстна карта (context map). Розглянемо їх детальніше.

Піддомен – складник предметної області, бізнес-домену. Виділяють три типи піддоменів: ключові піддомени (core subdomains), які відрізняють бізнес компанії від інших і надають їй конкурентну перевагу; загальні піддомени (generic subdomains) – складні в імплементації, але однакові в різних бізнесах активності, які не надають компанії конкурентної переваги, але є невід'ємним складником продукту (наприклад, системи аутентифікації та авторизації, прийому платежів); підтримувальні піддомени (supporting subdomains) – нескладні в імплементації піддомени, які не надають бізнесу конкурентну перевагу (рішення, що базуються на ETL- (*extract, transform, load* – витяг, перетворення, завантаження) або CRUD-операціях) [7].

Єдина мова – корпоративна термінологія, що описує всі поняття бізнес-домену, якою послуговуються всі працівники та стейкхолдери. Має бути точною та консистентною. У рамках обмеженого контексту кожне поняття повинно мати лише одне значення; синонімічні терміни не допускаються. Єдина мова має бути сформована в термінах бізнесу та відповідати ментальній моделі доменних експертів. Наповнення єдиної мови має покривати лише ті аспекти предметної області, що використовуються в комунікації, розробленні та трансляції ідей; аспекти домену, що не використовуються в бізнес-контексті, не включаються в єдину мову, хоча можуть її розширювати, за потреби, у майбутньому. Важливим нюансом є підтримка єдиної мови всіма учасниками. Інструментами підтримки можливе використання глосарія, тестів мовою Gherkin, спеціалізованих аналізаторів коду тощо [7]. У більш широкому розумінні, можлива інтеграція DDD з підходом керованої поведінкової розробки (behavior-driven development, BDD). Даний підхід передбачає використання специфікації системи, що находить від стейкхолдерів у форматі історій користувачів (user stories), як єдиної мови [11].

Обмежений контекст – окремо виділений контекст доменної моделі на основі тлумачень понять єдиної мови, що конфліктують. Консистентність єдиної мови та кількість виділених обмежених контекстів мають протилежну залежність: що консистентніше єдина мова, то менше потрібно обмежених контекстів. Обмежені контексти застосовуються коли необхідно мати кілька моделей того самого поняття або концепції. Інакше кажучи, обмежений контекст визначає межі застосування єдиної мови, що описує частину предметної області в застосуванні до вирішення конкретних проблем.

На відміну від піддоменів, які виділяються із предметної області на основі стратегії ведення бізнесу, обмежені контексти будуються як архітектурні рішення на основі вибору границь моделей. У застосуванні до мікросервісної архітектури кожен обмежений контекст має бути реалізований як окремий сервіс; отже, обмежений контекст може містити один або більше піддоменів. З позиції ownership'у (відповідальності команди за розроблення частини функціоналу системи) один обмежений контекст має бути призначений одній і тільки одній команді, хоча команда може володіти кількома обмеженими контекстами одночасно [7].

У проєктуванні обмежених контекстів важливо також ураховувати частоту зміни пов'язаних одна з одною концепцій, що може слугувати приводом включення їх у єдиний обмежений контекст у деяких складних випадках моделювання предметної області – наприклад, якщо один піддомен потребує використання декількох обмежених контекстів [12].

Знаходження оптимальних точок розділення предметної області на обмежені контексти може бути також виконане за допомогою розрахунків таких показників, як аферентна зв'язність (afferent coupling), еферентна зв'язність (efferent coupling) і пов'язність (cohesion), значення яких сприяють знаходженню варіантів з найнижчою зв'язністю та найбільш високою пов'язністю [13].

У проєктуванні в об'єктно орієнтованій парадигмі використання доменних подій усередині обмеженого контексту є стандартною практикою, проте у функціональній такий підхід не вітається через те, що він створює додаткові приховані залежності. Натомість внутрішнє виконання послідовних дій (так званий workflow) будується за допомогою композиції. Особливістю зовнішнього workflow є те, що він приймає на вхід команду, а його виходом є множина подій, що є точками відо-

кремлення такого workflow від зовнішнього світу. Відповідно, вхідні команди мають валідуватися на предмет задоволення внутрішніх умов обмеженого контексту, а вихідні події перевірятися стосовно того, чи не витікають із даного обмеженого контексту приватні дані, створюючи зайву зв'язність та проблеми з безпекою [9].

Контекстна карта – діаграма, яка зображує обмежені контексти системи в розрізі їхньої інтеграції одне з одним. Типи інтеграції описуються відповідними патернами та поділяються на три групи, залежно від способу взаємодії між командами розробників, як-от: кооперація, споживач – постачальник, окремі шляхи. До патернів кооперації належать партнерство (partnership) і спільне ядро (shared kernel). У разі партнерства координація змін є двосторонньою: одна команда повідомляє іншу про зміни у своєму API, до яких друга адаптується. Застосування даного патерну передбачає високий рівень комунікації між командами й може не підходити в разі географічно розподілених команд.

Патерн спільне ядро передбачає винесення спільних для різних обмежених контекстів моделей у єдине загальне місце – вихідні файли коду або бібліотеку. Даний патерн вносить сильну залежність між обмеженими контекстами, до яких його застосовано, і має використовуватися лише тоді, коли ціна дублювання коду є вищою за ціну координації змін у спільній кодовій базі. Тип інтеграції споживач – постачальник порушує баланс між командами в сенсі диктування однією з них формату контракту та передбачає використання трьох патернів, як-от: конформіст (conformist), запобіжний шар (anti-corruption layer) і сервіс з відкритим протоколом (open-host service). Патерн конформіст передбачає безумовне прийняття споживачем моделі обмеженого контексту постачальника. Його застосування виникає тоді, коли контракт постачальника є індустріальним стандартом або просто добре підходить для потреб споживача. Патерн запобіжний шар дозволяє споживачу транслювати моделі постачальника у більш зручний формат контракту. Його використання виправдане, коли безкомпромісне прийняття сторонньої моделі небажане або неможливе: обмежений контекст споживача може бути ключовим піддоменом, який не варто піддавати зовнішньому впливу, модель постачальника може бути незручною або часто змінюватися тощо.

Патерн сервіс з відкритим протоколом, навпаки, передбачає пристосування постачальника до потреб споживачів: для кожного з них створюється окремий публічний протокол, орієнтований на потреби споживача. Застосування даного патерну дозволяє постачальнику гнучкіше розвивати реалізацію власного функціоналу без зайвого впливу на обмежені контексти споживачів, а також дає можливість впроваджувати більш ліберальне версіонування власного API.

Останній тип інтеграції є водночас однойменним патерном – окремі шляхи (separate ways). Він застосовується тоді, коли жодна колаборація між командами неможлива або неефективна, тому простіше дублювати функціональність у різних обмежених контекстах. Застосування цього патерну необхідно уникати за інтеграції ключових піддоменів [7]. Оскільки контекстна карта відображає не лише суто технічну інформацію про архітектуру застосунку, а й зв'язки між командами та способи їхньої взаємодії, іноді можна вдаватися до зворотного маневру Конвея (inverse Conway maneuver), щоб забезпечити узгодженість архітектури зі структурою компанії [9]. Контекстна карта також відіграє значну роль у проєктуванні та прототипуванні системи. Цей процес може бути виконаний за допомогою спеціалізованих інструментів, як-от Context Mapper. Даний інструмент дозволяє використовувати спеціальну DSL-мову Context Mapper DSL (CML), за допомогою якої описуються піддомени, обмежені контексти, складається контекстна карта, а також описуються історії користувачів і сценарії використання. Фактично, CML дозволяє описати всю стратегічну частину DDD [14].

До тактичних патернів відносять сутність (entity), об'єкт-значення (value object), агрегат (aggregate), фабрику (factory), репозиторій (repository) і сервіс (service). Розглянемо їх детальніше.

Сутність (Entity) – об'єкт предметної області, який має властиву йому неперервну в часі індивідуальність існування (ідентичність), яка зберігається в нього навіть якщо його неідентифікуючі атрибути зміняться [9, 15]. Таким чином, дві сутності, які мають однакові атрибути, але різні ідентифікатори, є різними сутностями [1]. Ідентифікуючий атрибут однозначним чином відрізняє одну сутність від іншої [15]. Прикладами сутностей є користувач, замовлення, продукт, рахунок-фактура, банківська транзакція [9, 15]. Сутності використовуються лише як складники агрегатів.

Об'єкт-значення (Value object) – об'єкт предметної області, який не має власної ідентичності (індивідуального існування), а є набором окремих полів і повністю ними визначається. Однакових значень цих атрибутів у двох об'єктах-значеннях досить для того, щоб уважати їх повністю взаємозамінними [1, 15]. З погляду функціональної парадигми об'єкти-значення мають бути незмінними (immutable) [9]. Атрибути, що утворюють об'єкт-значення, мають бути єдиним концептуальним цілим. Прикладами об'єктів-значень є колір, літера, назва, адреса, локація, дата, об'єкт грошей [9, 15].

Агрегат (Aggregate) – сукупність взаємопов'язаних об'єктів, що створюють єдине ціле. Складається з кореневого об'єкта (сутності) і одного чи багатьох інших об'єктів (сутностей і об'єктів-значень) [1]. Кожен агрегат має межу, яка визначає об'єкти, з яких складається агрегат.

Кореневий об'єкт – єдиний складник агрегату, що містить унікальний ідентифікатор, на який і тільки на який можуть посилатися всі зовнішні об'єкти (агрегати). Також він відповідає за перевірку та дотримання інваріантів (бізнес-правил сутностей предметної області) та внутрішньої консистентності агрегату. Об'єкти, розташовані всередині агрегату, можуть посилатися один на одного без обмежень. Посилання на некореневі сутності можуть бути надані іншим об'єктам лише на час виконання якоїсь однієї операції; некореневі об'єкти можуть мати лише локальну ідентичність, тобто бути унікальними лише в межах агрегату. Кореневий об'єкт також може передавати зовнішнім клієнтам копії об'єктів-значень, які вже не матимуть зв'язку з агрегатом. Лише кореневі об'єкти можуть бути отримані через запити в базу даних – інші об'єкти, що утворюють агрегат, мають отримуватися лише через зв'язок з кореневим об'єктом [1, 15].

З погляду транзакційності даних агрегати мають використовуватися як атомарні одиниці. Одна транзакція може створювати або оновлювати лише один агрегат, що дає змогу гарантувати, що транзакція виконуватиметься в межах одного сервісу [1, 9]. Агрегати використовуються для моделювання бізнес-об'єктів предметної області [1]. Зазвичай їх варто робити якомога меншими та включати в них лише ті об'єкти, які мають бути у строгій консистентності [7]. Прикладом агрегату може бути замовлення, що містить позиції, які самі собою є окремими об'єктами, але водночас замовлення розглядається саме як єдине ціле [16].

Фабрика (Factory) – об'єкт або метод, що реалізує процес створення або відновлення складних об'єктів і агрегатів. Для проєктування фабрик можуть використовуватися як дизайн-патерни «банди чотирьох» (фабричний метод (factory method), абстрактна фабрика (abstract factory), будівельник (builder)), так і аналогічні засоби в інших парадигмах програмування [1, 15].

Репозиторій (Repository) – об'єкт, що реалізує механізм доступу до об'єктів, що зберігаються в базі даних [1].

Сервіс (Service) – об'єкт, що не має стану та реалізує логіку, яку не можна віднести ані до сутності, ані до об'єкта-значення. Дозволяє також координувати роботу декількох агрегатів [1, 7].

Незважаючи на те, що DDD дозволяє спроєктувати систему максимально близькою до природи предметної області, цей підхід також не позбавлений деяких проблемних аспектів. Один із них – так звана трилема вибору між чистотою доменної моделі (код бізнес-логіки не повинен мати залежностей поза власним процесом виконання та, відповідно, мати посилання лише на примітивні типи й інші доменні об'єкти), її повнотою (код, що стосується дотримання бізнес-логіки, повинен бути у відповідному шарі системи та не бути фрагментованим) та продуктивністю. Ця трилема виникає в ситуації, коли частина бізнес-логіки не може водночас бути складником агрегату та не порушувати його чистоти (наприклад, через необхідність запиту в репозиторій). Перенесення цієї частини коду в інший шар (наприклад, у контролер) призводить до фрагментованості, а вичитування всього масиву даних і передача його в агрегат з метою виконання фільтрації в ньому (замість виконання цієї роботи запитом до бази даних) призводять до зниження продуктивності. У такому разі рекомендується компромісне рішення – перенесення позапроцесового коду в контролер, що хоч і призведе до фрагментації бізнес-логіки, але буде найменшим компромісом порівняно з іншими варіантами [17].

Наступний виклик, який може виникнути під час застосування практик DDD, полягає в тому, що цей підхід має високу складність для розробників, не знайомих з його принципами й особливостями. Це спричиняє додаткові матеріальні та когнітивні витрати, особливо за переходу на дану парадигму [18].

У мікросервісній архітектурі, зокрема за використання DDD, широко застосовується патерн доменна подія (domain event). Доменна подія – це об'єкт, що продукується агрегатом, коли з ним відбувається щось важливе: створення, зміна стану тощо, і який може прийняти й обробити будь-яка зацікавлена сторона. В об'єктно орієнтованому програмуванні доменні події являють собою класи, а у функціональній парадигмі для цього можуть використовуватися, наприклад, визначення типів і моноїди [1, 9, 19]. Назви доменних подій мають бути створені за допомогою дієслів минулого часу (наприклад, "OrderCreated" – «замовлення створено», "OrderCancelled" – «замовлення скасовано» тощо) [1, 9]. У доменних подіях зазвичай містяться метадані, як-от унікальний ідентифікатор, часова відмітка (timestamp), ідентифікатор користувача, який вніс зміни, та інші [1].

Оскільки часто обробнику події необхідні додаткові дані, що відображають суть змін стану системи, а не лише повідомляють про факт таких змін, для запобігання додатковому запиту з його боку використовується так зване збагачення подій (event enrichment). Наприклад, подія "OrderCreated" може вже містити деталі щойно створеного замовлення, отже, спожива-

чам даної події не потрібно буде додатково запитувати цю інформацію під час її оброблення. Такий підхід може дещо знизити підтримуваність (maintainability) завдяки тому, що потенційно класи (або інші структури даних), якими описуються події, потребуватимуть змін кожного разу, коли змінюватимуться вимоги споживачів. Іншим недоліком може бути спроба задовільнити потреби багатьох різних споживачів події. Проте це досить рідкі сценарії, які не мають критичного характеру [1].

Одним із підходів визначення доменних подій є event storming. Він передбачає комунікацію з доменними експертами, яка складається із трьох етапів, як-от: мозковий штурм подій – доменні експерти пропонують можливі доменні події системи; визначення тригерів подій – доменні експерти мають визначити тригери, які продукують події (дії користувачів, зовнішня система, інша доменна подія, плин часу); ідентифікація агрегатів – доменні експерти повинні визначити які агрегати споживають які команди та продукують відповідні події [1].

Незважаючи на те, що успішне виконання команд (як-от дії користувачів), фактично, створює події, ці концепції необхідно чітко розрізняти. Команди – це повідомлення, які надходять зовні, уособлюють запит на зміну стану системи, їх виконання може бути як успішним, так і невдалим; звідси випливає імперативна форма їхніх назв (наприклад, “CreateOrder” – «створити замовлення»). Події ж уособлюють зміну стану, яка щойно відбулася, тому не можуть бути виконані невдало; звідси, відповідно, і форма їхніх назв у минулому часі (“OrderCreated” – «замовлення створено») [19].

Доменні події публікуються агрегатами, проте небажано, щоб вони напряму публікували повідомлення у брокер повідомлень через те, що для них неможливе впровадження залежностей (dependency injection), реалізація даної можливості призведе до змішування бізнес-логіки й інфраструктурних питань. Кращим підходом буде використання сервісів, які можуть застосовувати впровадження залежностей (отже, і отримати через даний механізм посилання на API відправки повідомлень) та, відповідно, взяти на себе відповідальність за відправку подій. Таким чином, агрегати генерують події кожного разу, коли змінюється їхній стан, та повертають їх сервісам. Наприклад, сервіс може зробити запит у сховище даних через репозиторій, що поверне екземпляр агрегату, на якому сервісом буде викликано метод, що приймає параметри, що змінюють його стан, та який поверне сервісу список подій, які він опублікує через брокер повідомлень. Іншим можливим підходом може бути акумулювання агрегатом необхідних для відправки подій у публічному полі, яке потім вичитується сервісом для публікації подій. Недоліком першого варіанту є необхідність повертання порожніх подій у разі, коли методи нічого не повертають, а недоліками другого є необхідність застосування механізмів інкапсуляції коду, що забезпечує сервісу доступ до списку подій і повторюватиметься, а також ускладнення доступу до даного поля об’єктам, які не є кореневими [1].

Традиційний спосіб організації персистентності даних системи відбувається за допомогою збереження їх у базу даних або нереляційне сховище. Такий підхід має низку недоліків. По-перше, використання реляційної бази даних підходить лише для обмеженого набору даних, які за своєю природою можуть бути представлені у формі таблиць. Оскільки здебільшого дані предметних областей потребують додаткових перетворень для збереження в реляційній базі даних, виникає так званий об’єктно-реляційний розрив (object-relational impedance mismatch), а використання ORM-фреймворків (ORM – object-relational mapping, об’єктно-реляційне відображення) не є, загалом, ефективним і зручним [1, 20]. По-друге, традиційний підхід не дозволяє нативно зберігати історію змін, що може бути корисним для аудиту системи або відновлення її попереднього стану. По-третє, для традиційного підходу збереження даних, публікація подій не є частиною бізнес-логіки й має бути реалізовано окремо, що підвищує ризик виникнення проблем із синхронізацією поточного стану після оновлень [1].

Проте реалізація персистентності можлива не лише через безпосереднє збереження стану у сховище даних. Альтернативою цьому підходу є патерн event sourcing, який передбачає збереження стану агрегату як послідовності подій, де кожна така подія являє собою черговий випадок зміни стану. Тоді відтворення стану агрегату відбувається через повторюване послідовне застосування змін, привнесених кожною подією [1]. Event sourcing також підходить для застосування у функціональній парадигмі, бо передбачає не зміну даних, а лише збереження незмінних записів змін цих даних, на основі яких можна реконструювати поточний стан системи [21].

Події, що змінюють стан агрегатів, зберігаються у сховищі подій (event store). Кожна подія містить ідентифікатор і тип події, ідентифікатор і тип сутності (або агрегату), до якої вона має стосунок, та корисне навантаження, як дані зміни стану, передбачені самою подією. Стан агрегату змінюється як результат наступного процесу: надходить команда, яка містить запит на виконання визначеної дії, яка валідується агрегатом, який, якщо не виникає помилок і виключень, генерує список

подій зміни стану, після чого вони зберігаються у сховищі подій, а відповідні обробники агрегатів, яким адресовані ці події, приймають їх і застосовують. Важливо, що ці обробники не мають закінчувати своє виконання невдачею, тому що перехід системи в новий стан уже відбувся, а також мають бути ідемпотентними – тобто повинні правильно обробляти ситуацію, коли подія, що вже була оброблена, надійте повторно. У разі, коли надходить декілька подій, що мають оновити той самий агрегат, необхідно застосовувати заходи, що забезпечують транзакційність такого оновлення (наприклад, оптимістичне блокування (optimistic locking)). Такі запобіжні механізми (як-от видавець опитувань (polling publisher) і відстеження транзакційного журналу (transaction log tailing)) мають використовуватися також під час публікування подій, бо це має бути атомарною операцією. Оскільки деякі агрегати можуть мати велику кількість оновлень та, відповідно, велику кількість подій, з яких необхідно їх конструювати, для оптимізації об'єму сховища даних і продуктивності застосовують знімки (snapshot) n-ої кількості попередніх подій, «згортаючи» їх у єдину подію, що відразу містить дані поточних змін [1].

Важливим питанням у застосуванні патерну event sourcing є еволюція подій, бо не завжди цей процес може бути зворотним. Наприклад, розширення моделі через додавання нового поля не перешкоджає застосуванню попередніх знімків стану системи, але звуження моделі через перейменування, видалення або зміну типу поля, як і зміна назви агрегату або пов'язаної з ним події, уже є несумісним із попередньою версією агрегату. Способом вирішення цієї проблеми є застосування спеціального компонента, що оновлює події до нової версії в момент завантаження зі сховища. Такий компонент має назву upcaster [1].

Ще одним питанням, яке постає під час проєктування системи з event sourcing, є потенційна необхідність виправлення помилок у попередніх подіях або їх перестановка, у разі чого поточний стан був би іншим. Ключ до вирішення цієї проблеми дає патерн ретроактивна подія (retroactive event) – тобто подія, яка б вплинула на поточний стан системи, якби була оброблена у визначеній часовій точці в минулому. Використання цього поняття дозволяє осмислювати процес виправлення помилок, як наступні три паралельні моделі (parallel model): некоректна реальність (incorrect reality) – поточний стан, що не враховує ретроактивну подію; коректна гілка (corrected reality) – стан системи, якого вона набула б, якби відбулася обробка ретроактивної події; коректна реальність (corrected reality) – фінальний стан системи, у якому вона має опинитися. Здебільшого коректна реальність збігатиметься з коректною гілкою, проте необов'язково. Суть методу зводиться до ідентифікації моменту необхідності застосування ретроактивної події – точки розгалуження (branch point), – повернення стану системи в дану точку та виправлення помилки [22, 23]. Даний підхід отримав розвиток з позиції питання спостережуваності (observability) системи. Хоча event sourcing і дозволяє відстежувати всі бізнес-події, у журналі подій (event log) часто не вистачає необхідної інформації для відстеження причинно-наслідкових зв'язків і взаємодій бізнес-подій у системі, що ускладнює процес налагодження. Способом вирішення даної проблеми є використання трасування (tracing), що надає спосіб відновлення причинно-наслідкового потоку виконання для конкретних запитів разом із метаданими, пов'язаними із внутрішнім станом системи, як-от змінні та часові відмітки. Отже, у розробників з'являється додатковий інструмент пошуку всіх, пов'язаних з окремою проблемою, подій за допомогою ідентифікатора трасування [24].

До переваг застосування event sourcing належать надійна публікація подій, збереження історії змін агрегатів, практично повне уникнення об'єктно-реляційного розриву через те, що зберігаються окремі події, а не агрегований стан [1], а також той факт, що моделювання домену з використанням event sourcing наближає вихідний код програми до того, що відбувається в реальному світі, на відміну від спроб абстрагувати модель [19]. До недоліків відносять складність у реалізації, труднощі з еволюціюванням подій і видаленням даних, а також можливе ускладнення процесу читання даних, коли для задоволення запиту необхідно спершу вичитати та застосувати всі пов'язані з даним агрегатом події, лише після чого можна буде виконати передбачувану запитом фільтрацію та отримати шуканий результат. Системи, що містять такі запити, повинні використовувати інший патерн, який оптимізує процес читання: CQRS (command and query responsibility segregation, розділення відповідальності командних запитів) [1].

Основна ідея патерну CQRS полягає в розділенні сховищ запису та читання. Фактично, таке розділення відбувається на рівні команд, що змінюють агрегат (тобто виконують операції створення, зміни та видалення), та запитів, що вичитують його поточний стан. Частина, що відповідає за роботу з командами, також може застосовувати операцію читання в разі, коли вона не є ресурсозатратною. Натомість усі складні та нетривіальні запити спрямовуються до частини, що відповідає за читання даних, у яку зміни пропагуються зі сховища для запису. Сховище для читання може містити спеціально підготовлені подання для складних запитів, але оптимізації можуть бути

застосовані до обох типів сховищ залежно від потреб домену, найбільш явною з яких є використання різних типів сховищ для запису та читання, замість компромісного рішення, яке виконує обидві операції повільніше, ніж спеціалізовані. Використання CQRS не лише дозволяє оптимізувати роботу з командами та запитами, а й робить можливим використання патерну event sourcing та дозволяє покращити розділення обов'язків, що є одним із принципів SOLID. Окрім того, комбінація патернів event sourcing і CQRS дозволяє отримати відносно невисоку складність розроблення та підтримки складних і великих систем. Недоліком використання CQRS є неминуче ускладнення архітектури та наявність лагу реплікації – дані пропагуються зі сховища запису у сховище читання не миттєво [1, 25].

Коли домен не диктує складні умови й патерни event sourcing та CQRS можуть занадто ускладнити систему, але питання збору єдиної відповіді кінцевому клієнту на основі даних, що належать кільком мікросервісам, залишається, варто застосовувати патерн об'єднання API (API composition). Застосування даного патерну передбачає використання нового компонента, що має назву "API composer", який виконує запити в необхідні сервіси та будує остаточне подання для клієнта.

Виконувати роль API composer'а може клієнтський застосунок, API gateway, окремий спеціалізований сервіс. Перший варіант може виявитися неоптимальним, тому що між клієнтським кодом і базою даних є фаєрвол і повільніша мережа, через яку неефективно передавати зайві дані. Другий варіант буде найбільш ефективний тоді, коли споживачами API composer'а є зовнішні клієнти, як-от веб- або мобільний застосунок. Третій варіант має сенс у разі, коли в ролі споживачів виступають внутрішні сервіси на бекенді, або логіка об'єднання даних, які необхідно віддавати зовнішньому клієнту, надмірно складна для того, щоб бути реалізованою в рамках API gateway.

Незважаючи на те, що об'єднання API є простішим патерном у реалізації, ніж CQRS, він також має низку недоліків. Серед них підвищені накладні витрати на додаткові запити даних, ризик зниження доступності системи через появу нового компонента (можливим способом боротьби із цим є кешування або повернення неповних даних у разі недоступності API composer'а), а також виникнення неконсистентних даних тоді, коли дані сервісів, з яких формується остаточна відповідь, ще не узгоджені одне з одним [1].

Висновки. Проведений аналіз підходів до організації бізнес-логіки в побудові мікросервісних систем демонструє, що натепер у контексті побудови мікросервісів за умов складної предметної області перспективною є парадигма Domain-Driven Design (DDD). Цей підхід забезпечує високу підтримуваність коду та точну відповідність моделі бізнес-вимогам, особливо в довгостроковому розвитку системи.

Для систем з відносно простою бізнес-логікою допустиме застосування таких патернів, як Transaction Script або Active Record, які є простішими в реалізації та потребують менше зусиль на впровадження. Однак рішення на основі даних патернів показують обмежені можливості за зростання об'єму кодової бази проєкту та підвищення складності бізнес-логіки системи.

Отже, застосування того чи того підходу безпосередньо залежить від складності та динаміки предметної галузі. Для систем з високою волатильністю та складною бізнес-логікою доцільно застосовувати парадигму DDD, незважаючи на її складність та високий поріг входу. Для систем малої та середньої складності краще підходять прості патерни, що забезпечують швидку реалізацію за обмежених ресурсів.

Перспективні напрями подальших досліджень такі: оптимізація DDD для типових сценаріїв, що дозволить зменшити витрати на його впровадження; розроблення і апробація гібридних підходів, що поєднують переваги парадигми DDD і простіших патернів; формалізація метрик оцінювання архітектурних рішень у контексті бізнес-логіки мікросервісів.

ЛІТЕРАТУРА

1. Richardson C. Microservices patterns: with examples in Java. Shelter Island, New York : Manning Publications, 2019.
2. Newman S. Building microservices: designing fine-grained systems, Second Edition. Beijing : O'Reilly Media, 2021.
3. Deursen S. van, Seemann M. Dependency injection: principles, practices, patterns. Shelter Island : Manning, 2019.
4. Lieser S. Clean Architecture vs. Onion Architecture vs. Hexagonal Architecture, CCD Akademie. URL: https://ccd-akademie.de/en/clean-architecture-vs-onion-architecture-vs-hexagonale-architektur/
5. Palermo J. The Onion Architecture : part 1. Programming with Palermo. URL: https://jeffreypalermo.com/2008/07/the-onion-architecture-part-1/
6. Fowler M. Patterns of enterprise application architecture, Nineteenth printing. in The Addison-Wesley Signature Series. Boston ; San Francisco ; New York ; Toronto ; Montreal ; London ; Munich ; Paris ; Madrid ; Capetown : Addison-Wesley, 2013.

7. Khononov V. Learning domain-driven design: aligning software architecture and business strategy. Beijing ; Boston ; Farnham ; Sebastopol ; Tokyo : O'Reilly, 2022.
8. Mota I. Anaemic Domain Model vs. Rich Domain Model. *Ensono. Insights + News.* URL: https://www.ensono.com/insights-and-news/expert-opinions/anaemic-domain-model-vs-rich-domain-model/
9. Wlaschin S. Domain modeling made functional: tackle software complexity with domain-driven design and F♯, Version: P1.0. Raleigh, North Carolina : The Pragmatic Bookshelf, 2018.
10. Fowler M. Domain Driven Design. URL: https://martinfowler.com/bliki/DomainDrivenDesign.html
11. Esther D. Behavior-Driven Development for Domain-Driven Design in Modern Software. URL: https://www.researchgate.net/publication/386136826_Behavior-Driven_Development_for_Domain-Driven_Design_in_Modern_Software
12. Verraes M., Wirfs-Brock R. Splitting a Domain Across Multiple Bounded Contexts. URL: https://verraes.net/2021/06/split-domain-across-bounded-contexts/
13. Vural H., Koyuncu M. Does Domain-Driven Design Lead to Finding the Optimal Modularity of a Microservice? *IEEE Access.* 2021. Vol. 9. P. 32721–32733. DOI: 10.1109/ACCESS.2021.3060895
14. Kapferer S., Zimmermann O. Domain-Driven Architecture Modeling and Rapid Prototyping with Context Mapper. Model-Driven Engineering and Software Development / S. Hammoudi, L. F. Pires, B. Selić (Eds.). *Communications in Computer and Information Science.* Vol. 1361. Cham : Springer International Publishing, 2021. P. 250–272. DOI: 10.1007/978-3-030-67445-8_11
15. Evans E. Domain-driven design: tackling complexity in the heart of software. Boston : Addison-Wesley, 2004.
16. Fowler M. DDD Aggregate. URL: https://martinfowler.com/bliki/DDD_Aggregate.html
17. Khorikov V. Domain model purity vs. domain model completeness (DDD Trilemma). *Enterprise Craftsmanship.* URL: https://enterprisecraftsmanship.com/posts/domain-model-purity-completeness
18. Özkan O., Babur Ö., Van Den Brand M. Refactoring with domain-driven design in an industrial context: An action research report. *Empir Software Eng.* Jul. 2023. Vol. 28. № 4. P. 94. DOI: 10.1007/s10664-023-10310-1
19. Barbini U. From objects to functions: build your software faster and safer with functional programming and Kotlin. Dallas, Texas : The Pragmatic Bookshelf, 2023.
20. Colley D., Stanier C., Asaduzzaman M. Investigating the Effects of Object-Relational Impedance Mismatch on the Efficiency of Object-Relational Mapping Frameworks. *Journal of Database Management.* Oct. 2020. Vol. 31. № 4. P. 1–23. DOI: 10.4018/JDM.2020100101
21. Khorikov V. OOP, FP, and object-relational impedance mismatch. *Enterprise Craftsmanship.* URL: https://enterprisecraftsmanship.com/posts/oop-fp-and-object-relational-impedance-mismatch/
22. Fowler M. Retroactive Event. URL: https://martinfowler.com/eaaDev/RetroactiveEvent.html
23. Fowler M. Parallel Model. URL: https://martinfowler.com/eaaDev/ParallelModel.html
24. Lima S., Correia J., Araujo F., Cardoso J. Improving observability in Event Sourcing systems. *Journal of Systems and Software*. Nov. 2021. Vol. 181. P. 111015. DOI: 10.1016/j.jss.2021.111015
25. Overeem M., Spoor M., Jansen S., Brinkkemper S. An empirical characterization of event sourced systems and their schema evolution – Lessons from industry. *Journal of Systems and Software*. Aug. 2021. Vol. 178. P. 110970. DOI: 10.1016/j.jss.2021.110970

REFERENCES

1. Richardson, C. (2019). *Microservices patterns: With examples in Java*. Manning Publications.
2. Newman, S. (2021). *Building microservices: Designing fine-grained systems* (Second Edition). O'Reilly Media.
3. Deursen, S. van, & Seemann, M. (2019). *Dependency injection: Principles, practices, patterns*. Manning.
4. Lieser, S. (2024, February 9). Clean Architecture vs. Onion Architecture vs. Hexagonal Architecture. *CCD Akademie*. https://ccd-akademie.de/en/clean-architecture-vs-onion-architecture-vs-hexagonale-architektur/
5. Palermo, J. (2008, July 29). The Onion Architecture: Part 1. *Programming with Palermo*. https://jeffreypalermo.com/2008/07/the-onion-architecture-part-1/
6. Fowler, M. (2013). *Patterns of enterprise application architecture* (Nineteenth printing). Addison-Wesley.
7. Khononov, V. (2022). *Learning domain-driven design: Aligning software architecture and business strategy*. O'Reilly.
8. Mota, I. (2023, May 24). Anaemic Domain Model vs. Rich Domain Model. *Ensono. Insights + News*. https://www.ensono.com/insights-and-news/expert-opinions/anaemic-domain-model-vs-rich-domain-model/

9. Wlaschin, S. (2018). *Domain modeling made functional: Tackle software complexity with domain-driven design and F♯* (Version: P1.0). The Pragmatic Bookshelf.
10. Fowler, M. (2020, April 22). *Domain Driven Design*. https://martinfowler.com/bliki/DomainDrivenDesign.html
11. Esther, D. (2023, February). *Behavior-Driven Development for Domain-Driven Design in Modern Software*. https://www.researchgate.net/publication/386136826_Behavior-Driven_Development_for_Domain-Driven_Design_in_Modern_Software
12. Verraes, M., & Wirfs-Brock, R. (2021, June 14). *Splitting a Domain Across Multiple Bounded Contexts*. https://verraes.net/2021/06/split-domain-across-bounded-contexts/
13. Vural, H., & Koyuncu, M. (2021). Does Domain-Driven Design Lead to Finding the Optimal Modularity of a Microservice? *IEEE Access*, *9*, 32721–32733. https://doi.org/10.1109/ACCESS.2021.3060895
14. Kapferer, S., & Zimmermann, O. (2021). Domain-Driven Architecture Modeling and Rapid Prototyping with Context Mapper. In S. Hammoudi, L. F. Pires, & B. Selić (Eds.), *Model-Driven Engineering and Software Development* (Vol. 1361, pp. 250–272). Springer International Publishing. https://doi.org/10.1007/978-3-030-67445-8_11
15. Evans, E. (2004). *Domain-driven design: Tackling complexity in the heart of software*. Addison-Wesley.
16. Fowler, M. (2013, April 23). *DDD Aggregate*. https://martinfowler.com/bliki/DDD_Aggregate.html
17. Khorikov, V. (2020, August 4). Domain model purity vs. Domain model completeness (DDD Trilemma). *Enterprise Craftsmanship*. https://enterprisecraftsmanship.com/posts/domain-model-purity-completeness
18. Özkan, O., Babur, Ö., & Van Den Brand, M. (2023). Refactoring with domain-driven design in an industrial context: An action research report. *Empirical Software Engineering*, *28*(4), 94. https://doi.org/10.1007/s10664-023-10310-1
19. Barbini, U. (2023). *From objects to functions: Build your software faster and safer with functional programming and Kotlin*. The Pragmatic Bookshelf.
20. Colley, D., Stanier, C., & Asaduzzaman, M. (2020). Investigating the Effects of Object-Relational Impedance Mismatch on the Efficiency of Object-Relational Mapping Frameworks: *Journal of Database Management*, *31*(4), 1–23. https://doi.org/10.4018/JDM.2020100101
21. Khorikov, V. (2016, November 3). OOP, FP, and object-relational impedance mismatch. *Enterprise Craftsmanship*. https://enterprisecraftsmanship.com/posts/oop-fp-and-object-relational-impedance-mismatch/
22. Fowler, M. (2005, December 12). *Retroactive Event*. https://martinfowler.com/eaaDev/RetroactiveEvent.html
23. Fowler, M. (2005, December 12). *Parallel Model*. https://martinfowler.com/eaaDev/ParallelModel.html
24. Lima, S., Correia, J., Araujo, F., & Cardoso, J. (2021). Improving observability in Event Sourcing systems. *Journal of Systems and Software*, *181*, 111015. https://doi.org/10.1016/j.jss.2021.111015
25. Overeem, M., Spoor, M., Jansen, S., & Brinkkemper, S. (2021). An empirical characterization of event sourced systems and their schema evolution—Lessons from industry. *Journal of Systems and Software*, *178*, 110970. https://doi.org/10.1016/j.jss.2021.110970

Дата першого надходження рукопису до видання: 24.08.2025
Дата прийнятого до друку рукопису після рецензування: 19.09.2025
Дата публікації: 31.12.2025

УДК 004.8
DOI https://doi.org/10.26661/2786-6254-2025-2-06

ГІБРИДНА СИСТЕМА ВИЯВЛЕННЯ ТЕРМІНІВ В УКРАЇНСЬКОМОВНИХ ТЕКСТАХ ДЛЯ ЗАБЕЗПЕЧЕННЯ КОГНІТИВНОЇ ДОСТУПНОСТІ

Савіцький Р. С.
аспірант,
старший викладач кафедри інженерії програмного забезпечення
Державний університет «Житомирська політехніка»
вул. Чуднівська, 103, Житомир, Україна
orcid.org/0000-0001-9804-3604
roman.savitskyi@gmail.com

Ключові слова: *машинне навчання, виявлення термінів, українська мова, когнітивна доступність, BERT, термінологія.*

Когнітивна доступність є критично важливою для забезпечення рівного доступу до інформації, особливо для користувачів із дислексією, розладами аутистичного спектра та віковими когнітивними змінами. Значна насиченість наукових і технічних текстів спеціалізованою термінологією, що може становити до половини всієї лексики, часто стає суттєвою перешкодою для розуміння змісту. Морфологічна складність української мови, що поєднує багатство відмінкових форм і активне використання запозичень, ускладнює автоматизоване виявлення термінів і потребує розроблення спеціалізованих методів.

У статті представлено комплексне дослідження ефективності сучасних алгоритмів машинного навчання для автоматичного виявлення термінологічної лексики в українських текстах. Порівняння охоплює глибинні моделі на базі трансформерів, класичні алгоритми з лінгвістичним інжинірингом ознак і словникові рішення, засновані на правилах, інтегровані в гібридній ансамблевій системі з адаптивним розподілом ваг. Експерименти проведено на спеціально створеному корпусі з 537 фрагментів тексту з BIO-розміткою, який містить понад три тисячі токенів і характеризується високою часткою термінології. Валідація за допомогою стратифікованої п’ятикратної крос-перевірки засвідчила перевагу ансамблевого підходу, який досяг F1-метрики 0,847 і точності 0,903, перевищивши результати окремих моделей, зокрема fine-tuned BERT (F1 = 0,835), випадковий ліс (F1 = 0,774) та словникового пошуку (F1 = 0.744)

Аналіз важливості ознак виявив, що ключовими індикаторами термінів є довжина слова та співвідношення голосних, а інтеграція цих характеристик із контекстуальним моделюванням забезпечує оптимальний баланс між точністю та швидкістю. Запропонована система здатна обробляти понад дві тисячі токенів на секунду, що дозволяє використовувати її в режимі реального часу для веб- та мобільних застосунків, спрямованих на підвищення когнітивної доступності контенту. Розроблена модель і відкритий корпус розміщені на платформі Hugging Face Hub та можуть стати базою для подальших досліджень і впровадження інструментів автоматичного спрощення текстів, орієнтованих на користувачів з особливими потребами.

MACHINE LEARNING APPROACHES FOR DETECTING TERMS IN UKRAINIAN TEXTS

Savitskyi R. S.
Senior Lecturer at the Department of Software Engineering
Zhytomyr Polytechnic State University
Chudnivska str., 103, Zhytomyr, Ukraine
orcid.org/0000-0001-9804-3604
roman.savitskyi@gmail.com

Key words: *machine learning, term detection, Ukrainian language, cognitive accessibility, BERT, terminology.*

Cognitive accessibility of texts is a critically important aspect of information inclusion, particularly for individuals with dyslexia, autism spectrum disorders, and age-related cognitive changes. Complex terminology poses significant barriers to information access, especially in scientific and technical documents where term concentration can reach 30–50%. Ukrainian, as an inflected language with a developed morphological system, presents additional challenges for automated detection of terminological lexicon through the abundance of case forms and active borrowing of international terms. This article presents a comprehensive study of four machine learning approaches for term detection in Ukrainian texts: fine-tuning of the youscan/ukr-roberta-base transformer model with additional classification layer for token classification, random forest classification with 14 specialized linguistic features adapted for Ukrainian morphology (word length, vowel ratio for «аеєиіїоуюя», morphological endings), dictionary-based deterministic approach using 16,796 classified lexical units with morphological normalization through Stanza NLP pipeline, and a hybrid ensemble system with dynamic weight adaptation based on weighted voting mechanism. Experiments were conducted on a specially created annotated corpus of 537 Ukrainian text samples with BIO term markup, containing 3,173 tokens with 52,76% terminological lexicon, validated through 5-fold stratified cross-validation. Results demonstrate that the ensemble approach achieves the best performance (F1-score = 0,847, Accuracy = 0,903), outperforming BERT fine-tuning (F1 = 0,835), random forest (F1 = 0,774), and dictionary search (F1 = 0,744). Feature importance analysis reveals that word length and vowel ratio are the most powerful discriminators for Ukrainian terminology. The system provides processing speed of 2,156 tokens per second, sufficient for real-time cognitive accessibility applications in web browsers and mobile applications. The developed RomanSavitskyi/ukr-term-detection model and annotated corpus are published on Hugging Face Hub for further research on Ukrainian terminological lexicon and development of accessibility tools for digital inclusion.

Вступ. Рівний доступ до інформації неможливий без урахування когнітивної доступності. За оцінками Всесвітньої організації охорони здоров'я, приблизно 15% населення світу має різні форми когнітивних порушень, як-от дислексія, розлади аутистичного спектра, вікові когнітивні зміни тощо [1]. Складна термінологія, довгі речення і абстрактні концепції створюють бар'єри для доступу до інформації. Особливо гостро проблема проявляється в науковій і технічній літературі, де концентрація спеціалізованої термінології може досягати 30–50% від загального словника тексту. Дослідження показують, що навіть одиничні незрозумілі терміни можуть суттєво знижувати розуміння всього тексту, створюючи ефект каскадного когнітивного навантаження [2].

Нині в українському цифровому просторі практично відсутні спеціалізовані інструменти для автоматизованого виявлення та спрощення термінології. Наявні міжнародні рішення не враховують морфологічних особливостей української мови, а також зростання вимог до цифрової доступності робить розроблення інструментів не лише технологічно важливою, але й соціально необхідною проблемою.

Автоматизація процесу виявлення термінів може суттєво покращити доступність широкого спектра контенту, а ефективні алгоритми дозволили б не лише ідентифікувати проблемні еле-

менти тексту, але й покращити якість тексту, надаючи авторам і редакторам інструменти для створення когнітивно доступних матеріалів.

У контексті описаних вище проблем, дане дослідження спрямоване на вирішення наукових питань, що визначають ефективність автоматизованого виявлення термінів в україномовних текстах. Тому метою даного дослідження є обґрунтування і опис гібридної системи виявлення термінів в україномовних текстах для забезпечення когнітивної доступності.

Для досягнення поставленої мети необхідно вирішити комплекс взаємопов'язаних завдань. Першочерговим завданням є створення анотованого корпусу українських текстів з розміткою термінологічної лексики та формування системи словникових ресурсів для забезпечення детерміністичного підходу до виявлення термінів. Наступним є адаптація та дослідження ефективності трансформерної моделі youscan/ukr-roberta-base для класифікації термінів в українських текстах з урахуванням морфологічної складності мови. Окрім того, необхідно розробити систему лінгвістичних ознак, специфічних для української термінології, та дослідити ефективність класифікатора випадкового лісу. Окремим завданням є реалізація словникового методу виявлення термінів з морфологічною нормалізацією через Stanza NLP, а також побудова гібридної системи з динамічним розподілом ваг, що поєднує переваги трансформерного, статистичного та словникового підходів. **Огляд літератури.** Традиційні формули читабельності, як-от індекс читабельності Флеша – Кінкейда й індекс Фога (Gunning Fog index), базуються переважно на поверхневих характеристиках тексту, як-от довжина речень, кількість складів у словах і частота вживання лексики [3]. Проте сучасні дослідження демонструють обмеженість таких підходів, особливо для морфологічно складних мов. Так, дослідження [4] показує, що когнітивне навантаження тексту значною мірою визначається не лише синтаксичною структурою, але й семантичною складністю окремих лексичних одиниць. Автори встановили, що навіть поодинокі незрозумілі терміни можуть створювати каскадний ефект зниження розуміння, особливо в читачів з когнітивними особливостями [4].

Одним із ключових викликів у виявленні термінологічної лексики в українських текстах є поєднання флективної природи мови з морфологічною варіативністю. На особливу увагу заслуговує дослідження [5], в якому аналізується ефективність моделей машинного навчання, як-от випадковий ліс, метод опорних векторів (SVM) та двоспрямованих кодувальних представлень із трасформерів (Bidirectional Encoder Representations from Transformers (BERT)), для завдань дезінформаційної детекції в українських текстах. Автори виявили, що навіть базові морфологічні перетворення значно підвищують точність класифікації, а застосування попередньо навченої трансформерної моделі забезпечує найкращі результати в умовах обмеженого корпусу [5].

У роботі [6] було запропоновано модель на основі рекурентної нейронної мережі (RNN) для виявлення образливої лексики в україномовному сегменті інтернету. Незважаючи на тематичну специфіку, автори продемонстрували важливість токенізації, контекстної сегментації та граматичної нормалізації. Запропонований ними sequence labelling підхід базується на BIO-розмітці та може бути адаптований для класифікації термінів [6].

Дослідження [7] охоплює методи машинного навчання для класифікації емоційного забарвлення тексту. Автори застосували частоту терміна, оберненого до частоти документа (TF-IDF), логістичну регресію та метод опорних векторів до корпусу українських текстів із високою семантичною складністю. Зокрема, було визначено вплив довжини слова, частоти голосних і морфологічних закінчень на точність класифікації. Ці ознаки є ключовими в термінодетекції, що робить результати дослідження релевантними для побудови ефективних лексичних класифікаторів [7].

У роботі [8] запропоновано гібридний підхід до авторської атрибуції текстів, що поєднує семантичний аналіз із класичними методами машинного навчання. Особливу увагу приділено побудові ансамблевих моделей, що враховують як частотні, так і синтаксичні характеристики тексту. Автори показали, що поєднання векторизації на основі граматичної анотації (POS-міток) із морфологічними фільтрами дозволяє значно підвищити точність класифікації [8].

Автори [9] досліджують проблему виявлення граматичних помилок у текстах українською мовою з використанням машинного навчання. Автори використовують метод опорних векторів, випадковий ліс і системи, засновані на правилах, а також оцінюють вагу ознак, як-от довжина слова, морфеми, співвідношення голосних. Саме ці ознаки показали ефективність і в завданні виявлення термінів, підтверджуючи їхню релевантність у побудові моделей пояснення [9].

Автори [10] дослідили великий міжмовний корпус проєкту ParlaMint II, де для лінгвістичної анотації кількох корпусів (серед яких і український) застосовували різні конвеєри NLP, зокрема Stanza NLP pipeline.

Проте нині практично відсутні спеціалізовані дослідження автоматичного виявлення термінів для української мови. Наявні роботи фокусуються переважно на розпізнавання іменованих сутнос-

тей і загальних NLP завданнях, залишаючи прогалину в термінологічній екстракції.

Методологія. Було створено анотований корпус для тренування моделей, де основний набір даних для навчання алгоритмів виявлення термінів складається з 537 вручну анотованих прикладів українського тексту, організованих у форматі BIO-розмітки (Beginning – Inside – Outside) та збережено у JSON структурі:

```
{
 "id": "sample_001",
 "text": "Машинне навчання використовує алгоритми",
  "tokens": ["Машинне", "навчання", "використовує", "алгоритми"],
   "labels": ["B-TERM", "I-TERM", "O", "B-TERM"],
 "domain": "IT"
}
```

Фрагменти тексту відібрані з наукових статей і освітніх матеріалів у домені інженерії програмного забезпечення, математики, фізики та технічних наук. Критеріями відбору були наявність спеціалізованої термінології, розмір фрагмента – 3–21 токен і збалансоване представлення предметних областей.

Анотування корпусу виконувалося автором з подальшою верифікацією фахівцем з обчислювальної лінгвістики на вибірці 20% корпусу (107 фрагментів). Терміном уважалася лексична одиниця, що позначає спеціалізоване поняття предметної області та потребує пояснення для нефахівця. Токенізація виконувалася за допомогою wordpiece tokenizer з моделі youscan/ukr-roberta-base для забезпечення сумісності із трансформерним підходом.

Загальний обсяг набору даних становить 3,173 токени, з яких 1,674 (52,76%) позначені як терміни, що свідчить про високу концентрацію термінологічної лексики у відібраних текстах.

Структурні характеристики набору даних демонструють придатність для навчання моделей машинного навчання. Середня довжина речення становить 5,91 токена з варіацією від 3 до 21 токена, що відповідає типовим фрагментам наукових та технічних текстів. Приблизно 94,6% прикладів (508 із 537) містять принаймні один термін, що забезпечує збалансоване представлення класу з наявністю термінів у навчальному наборі.

Тематичний аналіз виявлених термінів показує різноманітність предметних областей: інформаційні технології (“javaScript”, “docker”, “NLP”), математика та фізика («рівняння Ейнштейна», «теорема Піфагора»), методології розробки програмного забезпечення (“scrum”, “agile”), кібербезпека («криптографічний захист інформації», «кібератака»). Оскільки корпус охоплює тексти з різних доменів, це дозволяє протестувати здатність моделей до узагальнення.

Для підтримки підходів, заснованих на правилах і забезпеченні необхідного рівня порівнянь, було створено комплексну систему словникових ресурсів, що налічує 16,796 класифікованих лексичних одиниць. Процес формування словників здійснювався через агрегацію існуючих лексикографічних ресурсів та їх адаптацію для завдання виявлення термінів. Найбільшими компонентами є словник архаїчної лексики (7,427 записів) та лайки (6,609 записів), що свідчить про фокус системи на виявленні когнітивно складної лексики.

Словникові ресурси організовані за принципом пріоритетної заміни, де кожен термін супроводжується альтернативними варіантами різного рівня складності. Морфологічна обробка словникових одиниць здійснюється через Stanza NLP pipeline для української мови, що забезпечує нормалізацію відмінкових форм і числових варіацій. Технічна реалізація пошуку використовує хешування для первинного точного збігу з подальшою морфологічною нормалізацією для обробки флективних варіантів.

Основою рішення є тонке налаштування (fine-tuning) трансформерних моделей для виявлення термінів. Модель є адаптацією попередньо навченої архітектури на базі трансформерів для специфічного завдання класифікації токенів (token classification). За базову модель було обрано youscan/ukr-roberta-base, що є спеціалізованою версію RoBERTa для української мови, що демонструє високу якість на завданнях розуміння української мови [5].

Процес тонкого налаштування навчання проводиться з використанням гіперпараметрів, як-от 3 епохи, batch size 8, warmup steps 500, зниження ваги 0,01 та планування швидкості навчання з автоматичною оптимізацією. Розподіл набору даних становить 80% для навчання та 20% для перевірки з ранньою зупинкою на основі точності валідації.

Критичним компонентом імплементації є точна післяобробка результатів моделі. Система включає фільтрацію за довжиною слів (3–50 символів), видалення власних назв через евристичний аналіз капіталізації та морфологічних закінчень, а також фільтрування за спеціально заданим списком для 128 категорій загальновживаних слів. Додатково застосовується адаптивне порогове значення, де поріг класифікації змінюється залежно від наявності термінологічних ознак у слові.

Альтернативний підхід – випадковий ліс з лінгвістичними ознаками, а саме створення ансамблю дерев рішень зі спеціально розробленими лінгвістичними ознаками, адаптованими для особливостей української мови. Система feature

extraction базується на 14 основних характеристиках слів, що охоплюють фонетичні, морфологічні та структурні властивості.

Основні групи ознак включають метричні (довжина слова, кількість складів), фонетичні (співвідношення голосних до), структурні (повторювані символи, наявність цифр та спеціальних знаків) та позиційні (початкова/кінцева голосна) ознаки. Додатково враховуються морфологічні індикатори термінів: технічні префікси («авто», «мікро», «кібер»), суфікси («логія», «графія», «скопія»), характерні для термінологічної лексики словотворчі елементи.

Імплементація використовує scikit-learn класифікатор випадкового лісу з оптимізованими параметрами, як-от кількість дерев 100, максимальна глибина 10, мінімальне розбиття термінів 5, що допомагає запобігти перенавчанню на відносно невеликому наборі даних. Важливою особливістю є динамічна адаптація ознак під контекст виявлених термінів, що дозволяє покращити точність для рідкісних термінологічних категорій.

Словниковий метод базується на точному збігу з використанням комплексної системи словникових ресурсів і морфологічної форм. Основу становить спеціалізований словник із 675 складних термінів і 1,274 абревіатур, доповнений системою морфологічного аналізу через Stanza NLP pipeline для української мови.

Алгоритм включає декілька етапів обробки, починаючи від первинного пошуку за хешем, продовжуючи морфологічною нормалізацією для обробки відмінкових форм і числових варіацій, закінчуючи багаторівневою фільтрацією. Особливістю імплементації є використання оцінки впевненості (confidence scoring), де кожний збіг оцінюється за точністю відповідності (точний збіг = 1,0, морфологічні варіанти = 0,8–0,9) та контекстною валідністю через аналіз сусідніх слів. Такий підхід забезпечує високу точність за збереження прийнятного рівня виявлення для добре представлених у словнику категорій термінів.

Фінальна система поєднує переваги всіх описаних методів через weighted ensemble architecture з динамічним розподілом ваг залежно від характеристик конкретного випадку. Базові ваги становлять BERT fine-tuned модель (0,5), випадковий ліс з лінгвістичними ознаками (0,3), словниковий підхід (0,2).

Особливістю системи є адаптивні ваги, де ваги компонентів корегуються залежно від характеристик тексту, оскільки для технічних текстів підвищується вага BERT моделі, для загальної лексики – словникового підходу. Такий механізм забезпечує міцність системи до різних типів контенту та покращує загальну якість класифікації термінів.

Дослідження базувалося на трансформерній архітектурі youscan/ukr-roberta-base, попередньо навченій на українськомовному корпусі обсягом 7,8 гігабайта. Архітектурне рішення передбачало інтеграцію спеціалізованого класифікаційного модуля для розпізнавання іменованих сутностей на рівні токенів із застосуванням трикласової схеми маркування (B-TERM, I-TERM, O).

Процедура донавчання здійснювалася з коефіцієнтом навчання 2×10^{-5} та розміром пакету 16 зразків. Максимальна довжина вхідної послідовності була обмежена 512 токенами. Навчальний процес тривав протягом 10 епох із реалізацією механізму дострокового припинення за досягнення мінімального значення функції втрат на валідаційній вибірці.

Класифікатор для випадкового лісу базується на 14 спеціалізованих лінгвістичних ознаках, адаптованих для української морфології, як-от довжина слова, кількість повторюваних символів, максимальна довжина повтору, кількість голосних, кількість приголосних, співвідношення голосних, кількість цифр і спеціальних символів, позиційні характеристики, кількість сленґових патернів, категоріальні ознаки довжини та співвідношення символів.

Оцінювання проводилося за стандартними метриками для завдань sequence labeling: Precision, Recall та F1-score для кожного класу (B-TERM, I-TERM, O) з мікро- та макроусередненням. Додатково вимірювалися швидкість інференсу (токенів/секунда), розмір моделі та споживання пам’яті для оцінювання практичної застосовності.

Перевірка здійснювалася через 5-fold стратифіковану крос-валідацію з підтримкою балансу класів у кожному згині. Статистична значущість різниць між моделями перевірялася за допомогою парних t-тестів з поправкою Боферонні для множинних порівнянь [11].

Результати та аналіз. Результати демонструють значні відмінності у продуктивності методів залежно від архітектурних особливостей та підходів до обробки української термінології. Найкращі результати за комплексним критерієм F1-score демонструє ансамблевий підхід (0,847), що ефективно поєднує сильні сторони всіх базових методів. BERT fine-tuning показує другий результат (0,835), демонструє потужність контекстуального розуміння для термінологічної класифікації. Випадковий ліс з лінгвістичними ознаками досягає F1-score 0,774, що є прийнятним результатом для інтерпретованої моделі зі швидким припущенням (див. табл. 1).

Словниковий підхід характеризується найвищою точністю (0,886), що зумовлено детерміністичною природою точного зіставлення. Проте рівень виявлення (0,643) обмежений покриттям

Таблиця 1

Загальна продуктивність алгоритмів виявлення термінів

Метод	Precision	Recall	F1-Score	Accuracy	Швидкість (токенів/с)	Розмір моделі
BERT Fine-tuning	0,847	0,823	0,835	0,891	1,247	467 MB
Випадковий ліс	0,792	0,756	0,774	0,834	8,934	12 MB
Словниковий пошук	0,886	0,643	0,744	0,812	12,450	3,2 MB
Ансамблевий підхід	0,861	0,834	0,847	0,903	2,156	482 MB

словникових ресурсів, особливо для нових термінів і морфологічних варіантів поза словником. Саме такий дисбаланс типовий для систем, заснованих на правилах.

Словниковий підхід демонструє найвищу швидкість обробки (12,450 токенів/с) через операції простого пошуку та відсутність складних обчислень. Випадковий ліс показує другий результат (8,934 токенів/с) завдяки ефективності деревоподібної архітектури для висновків. BERT fine-tuning, незважаючи на найбільший розмір моделі, забезпечує прийнятну швидкість (1,247 токенів/с) для застосування в режимі реального часу. Ансамблевий підхід показує проміжну швидкість (2,156 токенів/с), що є компромісом між точністю та продуктивністю.

Парні t-тести з поправкою Бонферонні підтверджують статистичну значущість різниць між усіма методами за F1-score [11]. Найбільші відмінності спостерігаються між підходами, створеними на базі BERT та словниковою системою, особливо за рівнем виявлення (recall) метрикою.

Варто зазначити, що оцінка важливості терміна для токена t обчислюється за формулою лінійної комбінації прогнозів базових класифікаторів:

$$W(t) = w_1 \times P_1(t) + w_2 \times P_2(t) + w_3 \times P_3(t),$$

де $P_1(t) = P\_BERT(t)$ – імовірність класифікації токена як терміна трансформерною моделлю; $P_2(t) = P\_RF(t)$ – імовірність класифікатора випадкового лісу з лінгвістичними ознаками; $P_3(t) = P\_DICT(t)$ – бінарний індикатор словникового збігу; w_i – нормалізовані вагові коефіцієнти ($w_1 + w_2 + w_3 = 1$).

Базові ваги встановлюються згідно із продуктивністю кожного методу на валідаційному наборі: $w_1 = 0,5$, $w_2 = 0,3$, $w_3 = 0,2$. Значення відображають емпірично встановлену ієрархію точності, а саме BERT fine-tuning демонструє найвищу контекстуальну чутливість, випадковий ліс забезпечує стабільність через лінгвістичні ознаки, а словниковий підхід надає високу precision для відомих термінів.

Детальний аналіз важливості 14 лінгвістичних ознак для класифікатора випадкового лісу виявляє специфічні характеристики української термінологічної лексики. У дослідженні було доведено важливість для кожної ознаки, проте розглянемо деякі з них. Довжина слова (важливість = 0,234) виявилася найпотужнішим дискримінатором для української термінології. Аналіз розподілу показує, що українські терміни мають бімодальний розподіл довжини, зокрема короткі абревіатури (3–5 символів) та довгі складні терміни (8–15 символів). Середня довжина термінів (9,4 символу) значно перевищує загальновживані слова (6,1 символу). Співвідношення голосних (0,189) відображає фонетичні особливості української наукової термінології. Українські терміни часто мають низьке співвідношення голосних (0,42), особливо це помітно для технічних термінів на кшталт «криптографія», «алгоритм», «протокол». Співвідношення букв (0,156) ефективно відрізняє справжні терміни від абревіатур і символьних позначень. Чисті терміни, що складаються лише з літер, мають вищий пріоритет порівняно зі змішаними конструкціями, що містять цифри чи спеціальні символи.

Кореляційний аналіз виявив сильні зв'язки між пов'язаними ознаками, оскільки довжина слова корелює з кількістю приголосних ($r = 0,84$) і голосних ($r = 0,78$). Особливо ефективними виявилися комбінації таких ознак, як довжина та співвідношення голосних для технічних термінів, наявність цифр і спеціальні символи для формальних позначень, морфологічні закінчення для класичних наукових термінів. Використання українського вокалізму «аеєиіїоуюя» для розрахунку співвідношення голосних виявилося критично важливим. Експерименти з латинським набором голосних показали зниження F1-score на 0,043, що підтверджує необхідність мовно-специфічної адаптації лінгвістичних ознак.

Порівняння з іншими дослідженнями показує, що специфіка української мови створює додаткові виклики для виявлення термінів [4]. Створена гібридна система досягла F1-метрики 0,847, тоді як ансамблева модель авторів для англійської мови показала 0,388, що пояснюється не лише різницею в постановці завдань (виявлення термінів проти загального спрощення), але й морфологічною складністю української мови, яка потребувала інтеграції Stanza NLP pipeline та підвищила точність словникового підходу на 23%.

Варто більш детально проаналізувати помилки класифікації на тестовому наборі. Категоризація помилок здійснювалася через ручну перевірку випадково вибраних 200 неправильно класифікованих токенів зі збалансованим вибірковим дослідженням по всіх методах (див. табл. 2).

Таблиця 2

Розподіл типів помилок за методами

Тип помилки	BERT	Випадковий ліс	Словниковий	Ансамблевий
Хибнопозитивний результат				
Загальновживані IT-слова	23%	31%	8%	19%
Власні назви компаній	18%	12%	2%	11%
Абстрактні поняття	15%	22%	3%	12%
Загальні наукові слова	12%	18%	5%	9%
Хибнонегативний результат				
Багатословні терміни	35%	42%	58%	31%
Англомовні терміни	28%	38%	72%	25%
Морфологічні варіанти	22%	15%	45%	18%
Нові/рідкісні терміни	15%	5%	38%	26%

Аналіз хибнопозитивного результату показав, що загальновживані IT-слова становлять найбільшу категорію помилок для традиційних методів. Слова на кшталт «програма», «система», «документ», «інформація», мають морфологічні характеристики термінів (довжина, закінчення), але є занадто загальними для класифікації як специфічна термінологія. Випадковий ліс особливо схильний до таких помилок (31%) через фокус на поверхневих лінгвістичних ознаках без семантичного розуміння. Власні назви компаній (компанії “Microsoft”, “Google”, “Apple”) помилково класифікуються як терміни через їх появу в технічних контекстах. BERT показує вищий рівень таких помилок (18%) через контекстуальні асоціації, тоді як словниковий підхід ефективно фільтрує через спеціалізований стоп-список слів (2%).

Абстрактні поняття («метод», «процес», «результат») створюють виклик для всіх методів через двозначність, адже вони можуть бути як загальновживаними словами, так і термінами, залежно від контексту. Випадковий ліс демонструє найвищий рівень таких помилок (22%) через неможливість врахувати семантичний контекст.

Аналіз хибнонегативного результату показує, що багатослівні терміни («машинне навчання», «штучний інтелект», «обробка природної мови») становлять найбільший виклик для всіх методів. Словниковий підхід показує найгірші результати (58%) через обмежене покриття багатослівних конструкцій у словниках. BERT демонструє найкращі результати (35%) завдяки контекстуальному розумінню зв’язків між словами.

Морфологічні варіанти термінів («криптографічний», «алгоритмічний», «програмістський») пропускаються через недосконаліть морфологічної нормалізації. Випадковий ліс показує найкращі результати (15%) завдяки лінгвістичним ознакам, що враховують морфологічні патерни.

Особливу категорію становлять контекстуально-залежні помилки, де одне слово може бути терміном в одному контексті та загальним словом в іншому. Наприклад, «мережа» як комп’ютерний термін проти «мережа» у загальному понятті зв’язків. BERT показує найкращу здатність розрізняти такі випадки завдяки механізмам узгодження, але все ще демонструє 12% помилок у цій категорії. Традиційні методи практично неспроможні вирішувати такі неоднозначності без додаткових можливостей аналізу контексту.

Висновки. Порівняльний аналіз чотирьох підходів дозволив виділити фундаментальні компроміси між точністю, інтерпретованістю та обчислювальною ефективністю для виявлення української термінології. Отримані результати підтверджують, що вибір оптимального методу залежить від специфічних вимог застосування та доступних ресурсів.

Унаслідок чого була обґрунтована та розроблена гібридна система виявлення термінів в українськомовних текстах для забезпечення когнітивної доступності. Дослідження роботи якої дозволило виявити специфічні особливості визначення термінів в українській мові. Морфологічна варіативність значно ускладнила точне зіставлення зі словниками. Використання Stanza NLP pipeline для морфологічної нормалізації покращило рівень виявлення для словникового підходу на 23%, але все ще залишає значні прогалини для рідкісних форм та неологізмів.

До перспектив майбутніх досліджень відносимо розроблення ієрархічних схем багаторівневого класифікатору, що дозволить краще враховувати рівні термінологічної складності, а також удосконалення контекстуального моделювання шляхом інтеграції контексту на рівні документа для більш точного визначення термінів залежно від контексту.

ЛІТЕРАТУРА

1. Hartley S. World Report on Disability (WHO). 2011. DOI: 10.13140/RG.2.1.4993.8644
2. Zha X., Wang X., Yan Y., Gao Y., Yan G. Exploring influencing mechanism of herd behavior in academic information use: The perspective of cognitive load. *The Journal of Academic Librarianship.* 2023. Vol. 49. № 3. DOI: 10.1016/j.acalib.2023.102705

3. Maharani A. A. An analysis of the readability level in Life Today textbook for twelfth grade of senior high school using Flesch Reading Ease formula by Rudolf Flesch. Diploma thesis. Bandar Lampung : UIN Raden Intan Lampung, 2025. 70 p. URL: https://repository.radenintan.ac.id/39235/
4. Chamovitz E., Abend O. Cognitive simplification operations improve text simplification. *Proceedings of the 26th Conference on Computational Natural Language Learning (CoNLL).* Abu Dhabi, United Arab Emirates (Hybrid), 2022. P. 241–265. DOI: 10.18653/v1/2022.conll-1.17
5. Lipianina-Honcharenko K., Soia M., Yurkiv K. Evaluation of the effectiveness of machine learning methods for detecting disinformation in Ukrainian text data. *CEUR Workshop Proceedings.* 2024. Vol. 3702. P. 1–8. URL: https://ceur-ws.org/Vol-3702/paper9.pdf
6. Krak I., Zalutska O., Molchanov M., Mazurets O., Bahrii R. Abusive speech detection method for Ukrainian language using recurrent neural network. *CEUR Workshop Proceedings.* 2024. Vol. 3688. P. 1–9. URL: https://ceur-ws.org/Vol-3688/paper2.pdf
7. Lomovatskyi A., Basyuk T. Methods of machine learning and design of a system for determining the emotional coloring of Ukrainian-language content. *Scientific Bulletin of Lviv Polytechnic National University.* 2024. № 2. P. 78–90. DOI: 10.23939/sisn2024.15.074
8. Uhryn D., Vysotska V., Chyrun L., Chyrun S., Hu C. Intelligent application for textual content authorship identification based on machine learning and sentiment analysis. *International Journal of Intelligent Systems and Applications.* 2024. Vol. 17. № 2. P. 44–53, DOI: 10.5815/ijisa.2025.02.05
9. Lytvyn V., Pukach P., Vysotska V., Vovk M., Kholodna N. Identification and correction of grammatical errors in Ukrainian texts based on machine learning technology. *Mathematics.* 2023. Vol. 11. № 4. Art. 904. DOI: 10.3390/math11040904
10. ParlaMint II: advancing comparable parliamentary corpora across Europe / T. Erjavec, M. Kopp, N. Ljubešić et al. *Language Resources and Evaluation.* 2025. Vol. 59. P. 2071–2102. DOI: 10.1007/s10579-024-09798-w
11. Mondal H., Mondal S., Majumber R., De R. Conduct Common Statistical Tests Online. *Indian Dermatology Online Journal.* 2022. Vol. 13. № 4. P. 539–542. DOI: 10.4103/idoj.idoj_605_21

REFERENCES

1. Hartley, S. (2011). *World report on disability (WHO)*. 10.13140/RG.2.1.4993.8644
2. Zha, X., Wang, X., Yan, Y., Gao, Y., & Yan, G. (2023). Exploring influencing mechanism of herd behavior in academic information use: The perspective of cognitive load. *The Journal of Academic Librarianship, 49* (3), 102705. 10.1016/j.acalib.2023.102705
3. Maharani, A. A. (2025). *An analysis of the readability level in Life Today textbook for twelfth grade of senior high school using Flesch Reading Ease formula by Rudolf Flesch* [Diploma thesis, UIN Raden Intan Lampung]. UIN Raden Intan Repository. https://repository.radenintan.ac.id/39235/
4. Chamovitz, E., & Abend, O. (2022) *Cognitive simplification operations improve text simplification.* Proceedings of the 26th Conference on Computational Natural Language Learning (CoNLL). Abu Dhabi, United Arab Emirates (Hybrid). P. 241–265. DOI: 10.18653/v1/2022.conll-1.17
5. Lipianina-Honcharenko, K., Soia, M., & Yurkiv, K. (2024). Evaluation of the effectiveness of machine learning methods for detecting disinformation in Ukrainian text data. *CEUR Workshop Proceedings, 3702*, 1–8. https://ceur-ws.org/Vol-3702/paper9.pdf
6. Krak, I., Zalutska, O., Molchanov, M., Mazurets, O., & Bahrii, R. (2024). Abusive speech detection method for Ukrainian language using recurrent neural network. *CEUR Workshop Proceedings, 3688*, 1–9. https://ceur-ws.org/Vol-3688/paper2.pdf
7. Lomovatskyi, A., & Basyuk, T. (2024). Methods of machine learning and design of a system for determining the emotional coloring of Ukrainian-language content. *Scientific Bulletin of Lviv Polytechnic National University*, (2), 78–90. https://science.lpnu.ua/sites/default/files/journal-aper/2024/aug/35646/maket2402951-78-90_0.pdf
8. Uhryn, D., Vysotska, V., Chyrun, L., Chyrun, S., & Hu, C. (2024). Intelligent application for textual content authorship identification based on machine learning and sentiment analysis. *International Journal of Intelligent Systems and Applications, 17* (2), 44–53. https://www.mecs-press.org/ijisa/ijisa-v17-n2/IJISA-V17-N2-5.pdf
9. Lytvyn, V., Pukach, P., Vysotska, V., Vovk, M., & Kholodna, N. (2023). Identification and correction of grammatical errors in Ukrainian texts based on machine learning technology. *Mathematics, 11* (4), 904. DOI: 10.3390/math11040904
10. Erjavec, T., Kopp, M., Ljubešić, N., et al. ParlaMint II: advancing comparable parliamentary corpora across Europe. Lang Resources & Evaluation, 59, 2071–2102 (2025). DOI: 10.1007/s10579-024-09798-w
11. Mondal H., Mondal S., Majumder R., & De R. (2022). Conduct Common Statistical Tests Online. Indian Dermatology Online Journal, 13 (4), 539–542. DOI: 10.4103/idoj.idoj_605_21

Дата першого надходження рукопису до видання: 28.08.2025
Дата прийнятого до друку рукопису після рецензування: 25.09.2025
Дата публікації: 31.12.2025

РОЗДІЛ ІІІ. КОМП'ЮТЕРНІ НАУКИ

УДК 004.94:519.6:004.8
DOI https://doi.org/10.26661/2786-6254-2025-2-07

ГІБРИДНА АРХІТЕКТУРА ML ТА RL В АДАПТИВНОМУ МОДЕЛЮВАННІ СКЛАДНИХ ФІЗИЧНИХ ПРОЦЕСІВ

Геращенков Е. В.
аспірант кафедри програмного забезпечення систем
ДВНЗ «Ужгородський національний університет»
пл. Народна, 3, Ужгород, Україна
orcid.org/0009-0005-3615-6941
emilian.herashchenkov@uzhnu.edu.ua

Ключові слова: *адаптивне чисельне моделювання, машинне навчання, підкріплювальне навчання, багатошаровий персептрон, нейронні мережі, оптимізація обчислень.*

У роботі представлено методологію до адаптивного чисельного моделювання складних фізичних процесів із використанням методів машинного навчання. Основна увага зосереджена на інтеграції багатошарових нейронних мереж і алгоритмів навчання з підкріпленням у цикл чисельного розрахунку з метою досягнення оптимального балансу між точністю та обчислювальними витратами. Запропонована архітектура передбачає поєднання попереднього прогнозування параметрів дискретизації за допомогою нейромережі та динамічної корекції цих параметрів агентом підкріплювального навчання. У роботі сформульовано математичну постановку задачі, що враховує похибку чисельного розв'язку та вартість обчислень, розроблено функцію винагороди для RL-агента та побудовано блок-схему інтегрованої архітектури на основі алгоритму PPO.

Апробація архітектури виконана на тестових задачах теплопровідності в неоднорідному середовищі та хвильового рівняння в обмеженій області. Результати показали, що використання адаптивної MLP + RL архітектури дозволяє зменшити похибку обчислень до рівня 1,2–1,6% за скорочення часу виконання в 6–8 разів порівняно з високоточними еталонними рішеннями. Порівняння з рівномірною сіткою підтвердило суттєве підвищення ефективності розробленого методу. Отримані дані свідчать про можливість автоматичного зосередження обчислювальних ресурсів у критичних зонах та збереження прийнятної швидкодії в більш однорідних областях.

Отже, інтеграція методів машинного навчання в чисельному моделюванні відкриває перспективи створення універсальних адаптивних алгоритмів, здатних забезпечити високу точність без істотного зростання обчислювальних витрат. Подальший розвиток роботи пов'язаний із використанням фізично орієнтованих нейронних мереж (PINNs) та розширенням апробації на багатовимірні й нелінійні системи.

HYBRID ARCHITECTURE OF ML AND RL IN ADAPTIVE MODELING OF COMPLEX PHYSICAL PROCESSES

Herashchenkov E. V.
Postgraduate Student at the Department of Systems Software
Uzhgorod National University
Narodna sq., 3, Uzhhorod, Ukraine
orcid.org/0009-0005-3615-6941
emilian.herashchenkov@uzhnu.edu.ua

Key words: *adaptive numerical modeling, machine learning, reinforcement learning, multilayer perceptron, neural networks, computational optimization.*

This work presents a methodology for adaptive numerical modeling of complex physical processes using machine learning methods. The main focus is on integrating multilayer neural networks and reinforcement learning algorithms into the numerical computation cycle to achieve an optimal balance between accuracy and computational cost. The proposed architecture combines preliminary prediction of discretization parameters by a neural network with dynamic correction of these parameters by a reinforcement learning agent. The mathematical formulation of the problem, which accounts for the numerical solution error and computational cost, is introduced; a reward function for the RL agent is developed, and a block diagram of the integrated architecture based on the PPO algorithm is constructed.

The architecture was tested on benchmark problems of heat conduction in a heterogeneous medium and the wave equation in a bounded domain. The results demonstrated that the adaptive MLP+RL architecture reduces computational error to the level of 1,2–1,6% while decreasing runtime by a factor of 6–8 compared to high-precision reference solutions. Comparison with a uniform grid confirmed a significant increase in the efficiency of the developed method. The obtained data indicate the ability to automatically concentrate computational resources in critical zones while maintaining acceptable performance in more homogeneous regions.

Thus, the integration of machine learning methods into numerical modeling opens prospects for creating universal adaptive algorithms capable of providing high accuracy without a substantial increase in computational costs. Further work is associated with the use of physics-informed neural networks (PINNs) and extending testing to multidimensional and nonlinear systems.

Вступ. Чисельне моделювання є одним із ключових інструментів сучасної науки й інженерії. Воно застосовується для аналізу теплових, хвильових, гідродинамічних і електромагнітних процесів, які складно або неможливо дослідити експериментально. Проте одним із головних викликів залишається компроміс між точністю розрахунків і обчислювальними витратами.

Традиційні адаптивні методи – локальне згущення сітки [1], зміна кроку інтегрування [2], оцінка похибки – здатні підвищувати точність, однак вони обмежені жорсткими евристичними правилами. Це призводить до того, що в багатьох випадках обчислювальні ресурси витрачаються неефективно. У цьому контексті методи машинного навчання (далі – ML) відкривають нові перспективи [3]. Вони дають змогу автоматично визначати оптимальні параметри чисельного розрахунку [4], прогнозувати поведінку системи [5] та керувати процесом адаптації [6]. Особливий інтерес становить застосування навчання з підкріпленням (далі – RL) [7], де агент взаємодіє із чисельною моделлю як із середовищем і навчається ухвалювати рішення, що зменшують похибку за збереження продуктивності. Мета та завдання. Мета роботи – дослідити можливості впровадження методів машинного навчання, зокрема нейронних мереж і алгоритмів навчання з підкріпленням, у процес адаптивного чисельного моделювання складних фізичних процесів, оцінити їхню ефективність на тестових задачах теплопровідності й хвильової динаміки.

Завдання дослідження:

– проаналізувати сучасні адаптивні чисельні методи, визначити їхні обмеження в балансі між точністю і обчислювальними витратами;

– розробити математичну постановку задачі, що поєднує критерії точності чисельного розв’язку та вартість обчислень;

– створити архітектуру нейронної мережі для прогнозування оптимальних параметрів сітки та часових кроків на основі локальних характеристик фізичного поля;

– реалізувати агент навчання з підкріпленням для динамічного вибору параметрів дискретизації під час чисельного моделювання;

– інтегрувати нейронну мережу та RL-алгоритм у єдиний адаптивний цикл моделювання;

– провести апробацію розробленої архітектури на тестових задачах теплопровідності в неоднорідному середовищі та хвильового рівняння в обмеженій області;

– порівняти отримані результати з еталонними чисельними розв'язаннями та традиційними методами для оцінювання точності, продуктивності й ефективності.

Огляд літератури. Адаптивні чисельні методи почали активно розвиватися у другій половині XX ст. у відповідь на потребу підвищення точності розрахунків без істотного збільшення обчислювальних витрат. Одним із базових напрямів став метод локального згущення сітки (Adaptive Mesh Refinement (далі – AMR)) [8; 9], що дозволяє деталізувати область обчислень лише в тих зонах, де спостерігаються великі градієнти чи суттєві зміни розв'язку. Паралельно вдосконалювалися адаптивні методи інтегрування, серед яких класичним прикладом є алгоритми Рунге – Кутти з автоматичним вибором кроку [10]. Іншим важливим підходом стали мультигрідові методи [11], які поєднують розв'язки, отримані на різних рівнях дискретизації, забезпечують ефективне прискорення збіжності.

З початком XXI ст. в чисельному моделюванні з'явилися методи, що інтегрують можливості машинного навчання. Значну увагу привернули Physics-Informed Neural Networks (далі – PINNs) [12], де фізичні рівняння безпосередньо включаються до функції втрат, що дозволяє навчати мережу розв'язувати диференціальні рівняння без традиційної дискретизації. Іншим напрямом стало використання автоенкодерів і генеративних змагальних мереж (далі – GAN) для апроксимації складних багатовимірних рішень [13; 14], що важко піддаються класичному чисельному аналізу. Окрему гілку становлять методи, засновані на навчанні з підкріпленням, де RL-агенти здійснюють вибір параметрів сітки та кроку часу у процесі обчислення, поступово навчаючись знаходити компроміс між точністю та продуктивністю [15; 16].

Попри значні досягнення, існує низка невирішених проблем. По-перше, більшість запропонованих методів є локальними та добре працюють лише для конкретного класу задач, що обмежує їхню універсальність. По-друге, методи на основі машинного навчання часто потребують великих обсягів даних для тренування, тоді як у чисельному моделюванні не завжди можна створити достатній набір еталонних рішень. По-третє, не досить розробленими залишаються гібридні методи, що поєднують класичні чисельні алгоритми із ML-компонентами для отримання стабільних і узагальнених рішень.

У сукупності ці обмеження формують запит на нові методи, що здатні узгоджено інтегрувати машинне навчання з адаптивними чисельними процедурами. Саме тому дане дослідження зосереджене на розробленні методу, що базується на нейронних мережах і алгоритмах навчання з підкріпленням, з метою створення універсальної системи адаптивного моделювання, яка б забезпечувала підвищення точності без додаткових обчислювальних витрат.

Методи та моделі. Математичне формулювання задачі. ***Базовою моделлю***[17] розглядається одномірне рівняння теплопровідності на відрізку $\Omega = [0,L]$ у проміжку часу $t \in [0,T]$:

$$\frac{\partial u}{\partial t}(x,t) = \alpha \frac{\partial^2 u}{\partial x^2}(x,t), \quad (x,t) \in (0,L) \times (0,T), \qquad (1)$$

із початковою умовою:

$$u(x,0) = u_0(x), \; x \in [0,L], \qquad (2)$$

граничними умовами одного з типів (за потреби – змішаними):

– Діріхле – $u(0,t) = g_0(t)$, $u(L,t) = g_L(t)$;

– Нейман – $\frac{\partial u}{\partial x}(0,t) = q_0(t)$, $\frac{\partial u}{\partial x}(L,t) = q_L(t)$.

Коефіцієнт теплопровідності $\alpha > 0$ у базовій постановці вважаємо сталим; узагальнення на $\alpha = \alpha(x)$ розглянуто далі в тексті.

Для валідації та побудови еталонних розв'язків корисним є випадок однорідних умов Діріхле $u(0,t) = u(L,t) = 0$ та гладкої початкової умови з розкладом у ряд Фур'є:

$$u(x,t) = \sum_{k=1}^{\infty} A_k e^{-\alpha(\pi k/L)^2 t} \sin\left(\frac{\pi k x}{L}\right), \; A_k = \frac{2}{L}\int_o^L u_0(\xi)\sin\left(\frac{\pi k \xi}{L}\right) d\xi. \qquad (3)$$

Цей аналітичний вираз використовуватимемо як u_{ref} там, де це можливо; інакше u_{ref} формуємо як високоточний (надтонка сітка/малий крок) чисельний розв'язок.

Зручно ввести безрозмірні змінні $x^* = x/L$, $t^* = t/T$, тоді безрозмірний параметр Фур'є:

$$Fo = \frac{\alpha \Delta t}{(\Delta x)^2}, \qquad (4)$$

відіграє ключову роль у стійкості явних схем.

Дискретизація. Розіб'ємо $[0,L]$ на N інтервалів рівної довжини $\Delta x = L/N$ з вузлами $x_i = i\Delta x$, $i = 0,\ldots,N$, а $[0,T]$ – на M кроків $\Delta t = T/M$ з моментами $t^n = n\Delta t$, $n = 0,\ldots,M$. Позначимо

$$u_i^n \approx u(x_i, t^n).$$

Використаємо явну часову апроксимацію (прямий крок Ейлера) та центральну просторову різницю другого порядку. Для внутрішніх вузлів $i = 1, \ldots, N-1$ маємо схему:

$$u_i^{n+1} = u_i^n + \frac{\alpha \Delta t}{(\Delta x)^2}(u_{i-1}^n - 2u_i^n + u_{i+1}^n). \quad (5)$$

Позначимо $r = \frac{\alpha \Delta t}{(\Delta x)^2} = Fo$. Тоді:

$$u_i^{n+1} = u_i^n + r(u_{i-1}^n - 2u_i^n + u_{i+1}^n). \quad (6)$$

Щодо граничних вузлів, за умов Діріхле значення $u_0^n = g_0(t^n)$, $u_N^n = g_L(t^n)$ задаються явно.

За умов Неймана, наприклад у $x = 0$, використовуємо «уявний» вузол u_{-1}^n через односторонню різницю:

$$\frac{u_1^n - u_{-1}^n}{2\Delta x} = q_0(t^n) \Rightarrow u_{-1}^n = u_1^n - 2\Delta x q_0(t^n), \quad (7)$$

аналогічно в $x=L$: $u_{N+1}^n = u_{N-1}^n + 2\Delta x q_L(t^n)$. Підстановка цих виразів у явну формулу зберігає другий порядок за простором.

Зупинимося на порядку точності й узгодженості. Локальна похибка апроксимації (truncation error) для схеми «Ейлер уперед + центральна різниця» має вигляд:

$$\tau_i^n = O(\Delta t) + O((\Delta x)^2), \quad (8)$$

тобто схема перша за часом і друга за простором, узгоджена з ПДУ.

Щодо стійкості, класичний аналіз фон Неймана дає для 1D-явної схеми умову стійкості:

$$r = \frac{\alpha \Delta t}{(\Delta x)^2} \le \frac{1}{2}. \quad (9)$$

У разі змінних кроків Δx_i та/або Δt_n локальна умова набуває вигляду

$$r_i^n = \frac{\alpha \Delta t_n}{(\Delta x_i)^2} \le \frac{1}{2} \text{ для всіх } (i,n), \quad (10)$$

а для просторово-змінного коефіцієнта $\alpha(x)$ досить вимагати

$$\max_{i,n} \frac{\alpha(x_i)\Delta t_n}{(\Delta x_i)^2} \le \frac{1}{2}. \quad (11)$$

Узагальнення на $\alpha = \alpha(x)$. У такому разі дискретний оператор доцільно записувати в консервативній формі з потоком $J = -\alpha u_x$ [18]:

$$u_i^{n+1} = u_i^n + \frac{\Delta t}{\Delta x}(J_{i-1/2}^n - J_{i+1/2}^n), \; J_{i+1/2}^n = -\alpha_{i+1/2}\frac{u_{i+1}^n - u_i^n}{\Delta x}, \quad (12)$$

де $\alpha_{i+1/2}$ – гармонійне або щонайменше середнє значення α на інтервалі $[x_i, x_{i+1}]$ (це покращує стабільність і фізичну узгодженість на розривах α).

Адаптивні кроки в нашій постановці Δx і Δt можуть змінюватись динамічно. Для збереження простоти реалізації допускаємо кусково-однорідні ділянки сітки (AMR-рівні) та глобально-ступінчастий час, кожна операція згущення/розрідження супроводжується інтерполяцією/рестрикцією з не нижчим від другого порядку точності. Вибір Δt підпорядковується найсуворішій локальній умові $\max_i r_i^n \le 1/2$.

Функціонал оптимізації. Мета – мінімізувати похибку чисельного розв'язку щодо еталонного за заданого бюджету ресурсів. Нехай $u_{num}(x,t;P)$ – чисельний розв'язок, що залежить від набору параметрів дискретизації та адаптації P (просторові кроки, часові кроки, локальні рівні AMR, порядок схеми тощо) [19]. Нехай $u_{ref}(x,t)$ – еталон (аналітичний або високоточний чисельний).

У виборі норми похибки використовуємо дві еквівалентні мети: похибка в кінцевий момент T та інтегральна похибка в часі. Відповідні функціонали:

$$E_T(P) = \left\| u_{num}(\cdot,T,P) - u_{ref}(\cdot,T) \right\|_{L^2(0,L)}, \quad (13)$$

$$E_{[0,T]}(P) = \left(\int_0^T \left\| u_{num}(\cdot,t,P) - u_{ref}(\cdot,t) \right\|_{L^2(0,L)}^2 \omega(t)\,dt \right)^{1/2}, \quad (14)$$

де $\omega(t) \ge 0$ – вагова функція (усталено $\omega \equiv 1$). У дискретній формі (трапецієподібна квадратура):

$$E_{[0,T]}^2(P) \approx \sum_{n=0}^{M} \left(\sum_{i=0}^{N} \left(u_i^n - u_{ref,i}^n \right)^2 \Delta x \right) \omega^n \Delta t. \quad (15)$$

Модель вартості обчислень ураховує кількість вузлів і кроків, а також вартість міжсіткових операцій:

$$C(P) = k_1 \sum_{n=0}^{M-1} N_n + k_2 N_{ref/restr} + k_3 N_{io}, \quad (16)$$

де N_n – кількість активних вузлів на кроці n, $N_{ref/restr}$ – число операцій згущення/рестрикції, N_{io} – допоміжні операції (оцінка похибки, інтерполяція в часі тощо), $k_1, k_2, k_3 > 0$ – питомі коефіцієнти вартості. У найпростішому випадку користуються наближенням $C \approx \kappa \sum_n N_n$.

Оптимізаційна постановка задачі передбачає таке формулювання з обмеженням на ресурси та стійкість [20]:

$$\min_P E(P) \text{ за умов } C(P) \le C_{\max}, \; r_i^n(P) \le \frac{1}{2} \forall i,n. \quad (17)$$

Еквівалентно можна розглядати зважену ціль:

$$J(P) = E(P) + \lambda C(P) \to \min_P, \; \lambda > 0, \quad (18)$$

де λ контролює компроміс точність – вартість. Для етапів, де еталон u_{ref} недоступний у реальному часі, замінюємо E на апріорну/апостеріорну оцінку похибки, наприклад на базі Річардсона:

$$\varepsilon_{est}^n \approx \frac{\left\| u^{\Delta t,\Delta x}(\cdot,t^n) - u^{\Delta t/2,\Delta x/2}(\cdot,t^n) \right\|_{L^2}}{2^p - 1}, \quad (19)$$

де p – порядок схеми за домінувальною змінною (для нашої часово-просторової комбінації ефективний порядок задається мінімумом між часовим і просторовим). Підстановка ε_{est} у J забезпечує практичну функцію якості, що не потребує знання точного розв'язку.

Для реалізації зв'язку з подальшою RL-постановкою в термінах ухвалення рішень параметри P еволюціонують як послідовність дій $\{a_t\}$ (зміна

Δx, Δt, увімкнення/вимкнення локального згущення тощо) [12]. Стабільність накладає жорсткі обмеження $r_i'' \le \frac{1}{2}$, тоді як обмеження ресурсу $C \le C_{max}$ можна реалізувати або як твердий конус допустимих дій, або як штраф у формі доданка λC у функціоналі J. У розділах про RL ця ж постановка трансформується у функцію винагороди, але вже тут вона визначає «ідеальну» мету оптимізації.

Бачимо, що на даному етапі формалізовано базову фізичну модель, подано коректну явну дискретизацію з порядками точності, умовами стійкості та правилами для граничних умов, а також уведено цільовий функціонал, який поєднує помилку наближення та вартість обчислень. Дана постановка дозволяє надалі природно інтегрувати як параметричні нейромережеві предиктори для вибору P, так і RL-механізми для онлайн-керування адаптацією під жорсткі обмеження стійкості й обчислювального бюджету.

Результати та обговорення. Методологія. *Архітектура нейронної мережі.* Запропонована архітектура побудована у формі багатошарового персептрона (MLP) [21], завданням якого є апроксимація функції адаптивного переходу між локальними характеристиками фізичного поля та оптимальними параметрами чисельної схеми.

На вхід мережі подаються вектори локальних характеристик, що визначаються на кожному часовому кроці симуляції. До таких характеристик належать *градієнти поля* (оцінка першої похідної), *локальна кривина* (оцінка другої похідної), а також *оцінка локальної похибки* $\hat{\varepsilon}_t$, яка може бути визначена шляхом порівняння результатів схем різного порядку точності.

Отже, вхідний вектор набуває вигляду [19]:

$$x_t = (\nabla u, \nabla^2 u, \hat{\varepsilon}_t), \tag{20}$$

де $\nabla u(x,t)$ – локальний градієнт температурного поля (оцінка першої похідної), $\nabla^2 u(x,t)$ – *локальна кривина* (друга похідна), а $\hat{\varepsilon}_t$ – оцінка похибки, визначена через порівняння результатів розрахунку на поточній сітці з еталонним рішенням.

На виході модель формує набір параметрів, що безпосередньо використовуються в чисельній схемі:

$$y_t = (\Delta x_t, \Delta t_t, k_t), \tag{21}$$

де Δx_t – просторовий крок, Δt_t – часовий крок, а k_t – індекс порядку чисельної схеми, що може набувати значень $k_t \in \{1, 2, 4\}$.

Структура мережі включає від трьох до п'яти прихованих шарів, кожен із яких містить від 64 до 128 нейронів. Для забезпечення нелінійності використовуються функції активації типу ReLU або LeakyReLU. На вихідному шарі застосовується *лінійна активація* для прогнозування параметрів Δx та Δt, тоді як для вибору порядку схеми використовується *softmax-активація*, що дозволяє інтерпретувати результат як імовірнісний розподіл.

Для уникнення перенавчання та підвищення стійкості навчання в архітектурі реалізовано кілька механізмів регуляризації [22]. Зокрема, застосовується *Dropout* з імовірністю вимикання $p = 0{,}2{-}0{,}3$, *L2-регуляризація ваг*, а також *batch-нормалізація* після кожного шару. У сукупності ці техніки забезпечують стабільну генералізацію моделі на нових даних і дозволяють інтегрувати її в циклі адаптивного чисельного розрахунку.

Таким чином, нейронна мережа реалізує функцію відображення від локальних характеристик фізичного поля до оптимальних параметрів чисельної схеми, виступає ключовим елементом інтегрованої системи RL-адаптації.

Для динамічного вибору параметрів у процесі моделювання використовується *агент підкріплювального навчання (RL)*, який взаємодіє із середовищем та поступово вдосконалює свою політику.

Стан середовища в момент часу t визначається вектором

$$s_t = (\nabla u, \nabla^2 u, \hat{\varepsilon}_t, C_t), \tag{22}$$

де ∇u та $\nabla^2 u$ відповідають локальним просторовим характеристикам поля, $\hat{\varepsilon}_t$ є оцінкою похибки, а C_t – мірою обчислювальних витрат, яка може визначатися як кількість операцій або час виконання на кроці.

У найпростішому випадку для керування адаптивним процесом використовується *дискретний простір дій*. Агент може вибирати одну із трьох можливостей: зменшення кроку Δx, збільшення кроку Δx або залишення його незмінним, за що відповідає відповідна дія a_i, $i = \{1, 2, 3\}$. Така постановка задачі дозволяє інтерпретувати процес адаптації як послідовність рішень, що спрямовані на баланс між точністю та обчислювальними витратами. Зменшення Δx підвищує локальну точність розрахунку, але збільшує кількість вузлів сітки та, відповідно, обчислювальну складність. Збільшення Δx, навпаки, скорочує ресурси, але може призвести до втрати важливих деталей хвильового процесу. Вибір дії a_i на кожному кроці фактично визначає динаміку еволюції сітки, що є ключовим у побудові ефективного адаптивного алгоритму. Аналогічно формулюються дії для Δt та вибору порядку схеми. Отже, простір дій може бути як дискретним (з фіксованими кроками зміни), так і гібридним (дискретний вибір порядку схеми + неперервна регресія для Δx, Δt).

Функція винагороди побудована з метою балансування між точністю чисельного розрахунку й ефективністю використання обчислювальних ресурсів:

$$R_t = -\alpha \cdot \varepsilon_t - \beta \cdot C_t, \tag{23}$$

де коефіцієнти $\alpha, \beta > 0$ визначають відносну важливість точності та продуктивності. Максимізація функції винагороди стимулює агента знаходити такі параметри, що забезпечують прийнятний компроміс між якістю обчислень і їх швидкістю.

Інтеграція MLP-прогнозування та RL-адаптації здійснюється як ***єдиний ітеративний цикл*** [23; 24]. Початкові параметри RL-агента ініціалізуються випадковим чином, після чого нейронна мережа попередньо тренується в режимі *supervised learning* на синтетичних даних. Це дозволяє значно зменшити час збіжності в подальшому підкріплювальному навчанні.

Після цього запускається симуляція фізичного процесу із заданими початковими та граничними умовами. На кожному часовому кроці агент спостерігає стан середовища s_t, формує дію a_t, що визначає параметри Δx, Δt та порядок схеми k_t, після чого виконується один крок чисельного розрахунку.

Отримані результати дозволяють обчислити похибку ε_t та обчислювальні витрати C_t, на основі яких формується винагорода R_t. Ця величина використовується для оновлення політики агента за допомогою алгоритму *Proximal Policy Optimization (PPO)*. Процес оновлення задається рівнянням

$$\theta \leftarrow \theta + \eta \cdot \hat{\nabla}_\theta L_{PPO}(\theta), \qquad (24)$$

де θ – параметри політики, η – швидкість навчання, а $L_{PPO}(\theta)$ – функція втрат PPO, що включає додатковий ентропійний член для запобігання деградації політики. Отже, ітеративний процес триває протягом усієї симуляції або до досягнення стабільної збіжності. Система поступово вдосконалює стратегію вибору параметрів, забезпечує водночас високу точність і прийнятний рівень обчислювальних витрат.

Наведена на рисунку 1 блок-схема ілюструє цикл адаптивного чисельного моделювання з інтеграцією MLP-політики й алгоритму PPO. Спочатку середовище моделювання формує поточний стан s_t, який включає характеристики сітки, локальну похибку й обчислювальні витрати, і передає його на вхід MLP-політики $\pi_\theta(a|s)$. Політика генерує дію a_t, яка визначає зміну просторового кроку Δx. Оновлене значення Δx_{new} подається на блок чисельного моделювання, де обчислюється результат моделювання, формується винагорода $R = f$(точність, ресурси). Винагорода й оцінка стану $V(s)$ (*state-value function*), що отримана із MLP, подаються у блок обчислення *Advantage function* $A(s,a) = Q(s,a) - V(s)$. У контексті алгоритмів з підкріпленням $Q(s,a)$ – це *функція дії-цінності (action-value function),* що визначає очікувану

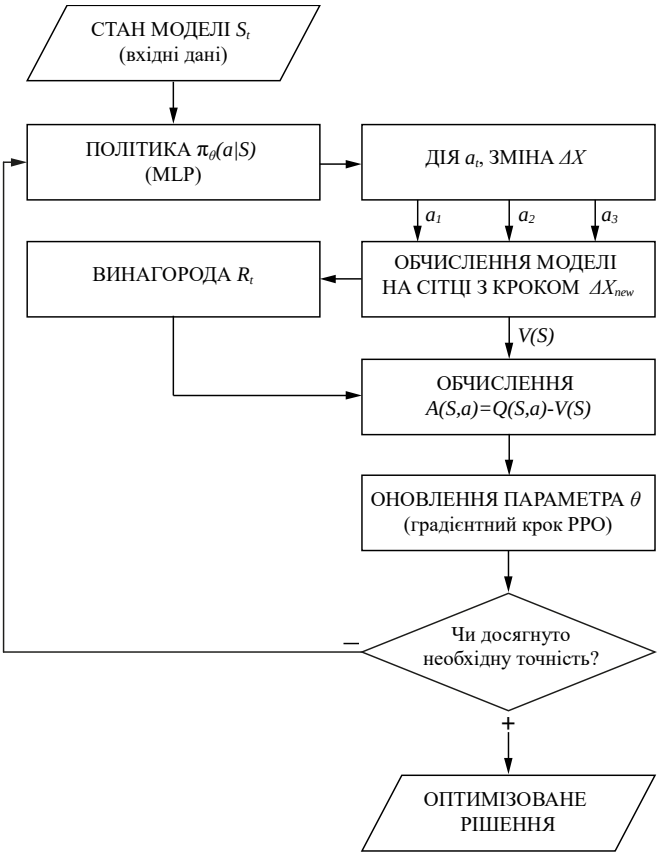


Рис. 1. Блок-схема інтеграції MLP та PPO для динамічного налаштування параметрів сітки

сумарну винагороду, якщо агент перебуває у стані s, виконує дію a, а далі діє за поточною політикою π_θ. Обчислена *Advantage function* використовується для оновлення параметрів політики θ за алгоритмом PPO, що повертає оновлену політику назад до MLP, замкнувши цикл. Стрілки на схемі передбачають такі підписи: π_θ на виході політики, Δx на вході моделювання, R на вході *Advantage function*, *A(s,a)* на вході PPO-update, що забезпечує наочне й академічне відображення процесу адаптивного навчання.

Формування навчального набору та характеристики навчання мережі. Навчальний набір даних для багатошарового персептрона формується на основі чисельних експериментів із розв'язання тестових задач (рівняння теплопровідності та хвильове рівняння) з різними параметрами дискретизації. Спочатку будується високоточне «еталонне» розв'язання на дуже дрібній сітці, яке використовується як референт. Далі на більш грубих сітках із різними значеннями кроків Δx і Δt виконуються обчислення, після чого для кожної локальної конфігурації обчислюються градієнт ∇u, друга похідна $\nabla^2 u$ та оцінка локальної похибки ε_t, отримана як різниця між результатом грубого обчислення та еталонного розв'язку. Таким чином формується вхідна частина набору даних – вектори локальних характеристик поля, за формулою (20).

Вихідні дані формуються як оптимальні параметри чисельної схеми, що мінімізують похибку за обмежених ресурсів: просторовий крок Δx_t, часовий крок Δt_t і порядок чисельної схеми k_t. У результаті багатошаровий персептрон навчається апроксимувати залежність між локальними характеристиками й оптимальними параметрами дискретизації. Отже, навчальний набір не є штучно згенерованим, а формується безпосередньо із чисельних експериментів, де еталонне рішення відіграє роль «учителя», а варіації сітки та порядку схеми – роль простору пошуку оптимальних рішень. Це дозволяє підготувати дані, що відображають реальні умови адаптивного моделювання.

Розмір навчальної вибірки визначався кількістю локальних конфігурацій поля, згенерованих під час розв'язання тестових задач на різних сітках і часових кроках. Для рівняння теплопровідності було отримано приблизно $1{,}2\times10^5$ вхідних векторів x_t, для хвильового рівняння – ще 8×10^4, що разом формує вибірку порядку 2×10^5 прикладів.

Крива навчання багатошарового персептрона відображала типову стабільну поведінку: протягом перших 40–50 епох спостерігалося інтенсивне зменшення функції втрат (MSE), після чого процес переходив у режим поступового виходу на плато. Значення MSE знизилося майже на порядок – від початкових значень приблизно 10^{-2} до рівня 10^{-3} уже після 30 епох, далі продовжувало зменшуватися до 10^{-4}, де стабілізувалося. Така динаміка свідчить, що мережа здатна ефективно виявляти закономірності в даних без тенденції до перенавчання.

Стабільність роботи мережі додатково перевірялася на незалежних тестових підвибірках, що не входили до навчального процесу. Порівняння кривих навчальної і тестової похибки показало, що відхилення не перевищувало 0,2% на всьому інтервалі епох, тобто мережа не втрачала здатності до узагальнення. Це підтверджує, що запропонована методика регуляризації (Dropout, L2-норма, batch-нормалізація) забезпечує необхідний захист від переадаптації. Для підтвердження відтворюваності результати навчання були перевірені за різних випадкових ініціалізацій ваг. У всіх випадках функція втрат збігалася до близького рівня, а остаточні значення середньої похибки різнилися не більше ніж на 0,1–0,15%. Це свідчить про стабільність алгоритму оптимізації та надійність отриманих результатів.

Апробація на тестових задачах. Для різнобічної апробації архітектури за тестові задачі варто обрати три класичні задачі, кожна з них підсвічує різні можливості. Рівняння теплопровідності в неоднорідному середовищі добре підходить для перевірки здатності архітектури адаптивно підбирати крок сітки Δx. Тут важлива точність у зонах різкої зміни коефіцієнтів теплопровідності, а функція винагороди $R = f$ *(точність, ресурси)* дозволяє протестувати баланс між роздільністю та обчислювальними витратами. Хвильове рівняння в обмеженій області дає змогу перевірити, наскільки політика (policy network) і Advantage $A(s, a) = Q(s, a) - V(s)$ ефективно керують локальною дискретизацією, особливо для відтворення інтерференційних і резонансних ефектів. Це класична задача для перевірки узгодження стабільності та точності. Спрощене рівняння Нав'є – Стокса дозволяє протестувати архітектуру на більш складних динамічних системах, де винагорода R формується не тільки від точності, а й від швидкості збіжності. Тут можна показати перевагу підходу з функціями $Q(s, a)$ та $V(s)$, що дозволяють оцінювати ефективність різних стратегій сіткової адаптації в багатовимірних задачах. Тобто архітектура виступає універсальним механізмом адаптивного керування обчисленнями: від простих задач із контрольованою складністю (теплопровідність) до хвильових і гідродинамічних процесів, де потрібна багатокритеріальна оптимізація. У даній праці розроблену архітектуру було апробовано на задачі теплопровідності в неоднорідному середовищі та задачі хвильового рівняння в обмеженій області.

Апробація архітектури на задачі теплопровідності в неоднорідному середовищі. Для перевірки працездатності розробленої архітектури було застосовано задачу одномірного рівняння теплопровідності в неоднорідному середовищі:

$$\frac{\partial u(x,t)}{\partial t} = \frac{\partial}{\partial x}\left(k(x)\frac{\partial u(x,t)}{\partial x} \right), \qquad (25)$$

де *u (x, t)* – температура, а *k (x)* – просторово-змінний коефіцієнт теплопровідності. У цій постановці головна складність полягає в тому, що неоднорідність середовища призводить до локальних різких змін температурного профілю, особливо на межах областей із різними значеннями *k (x)*.

Архітектура на основі *Q*-, *V*- і *Advantage*-функцій була використана для вибору оптимального кроку дискретизації Δ*x* залежно від поточної оцінки стану системи, де *функція цінності V (s)* характеризувала глобальну точність розв'язку за фіксованого розбиття, *функція дії Q (s, a)* відображала очікуваний виграш від вибору певного кроку Δ*x* у конкретній локальній області, а *функція переваги A (s, a)* дозволяла балансувати між точністю та обчислювальними витратами, виявляти області, де зменшення кроку справді дає значний приріст точності.

У таблиці 1 наведено числові результати апробації архітектури PPO + MLP для задачі теплопровідності в неоднорідному середовищі. Еталон ("Reference fine") обчислювався на рівномірній дуже дрібній сітці й використовувався лише для оцінювання похибки інших методів. Умови експериментів (спільні): $L = 1$, $T = 1$; явна схема із CFL $\alpha \Delta t/(\Delta x)^2 \leq 0{,}45$; похибка $\|u_{num} - u_{ref}\|_2$ на момент $t = T$. *Nx* – кількість вузлів сітки (або ефективна середня для адаптивної), $\Delta x_{min/max}$ – мінімальний/максимальний крок по області.

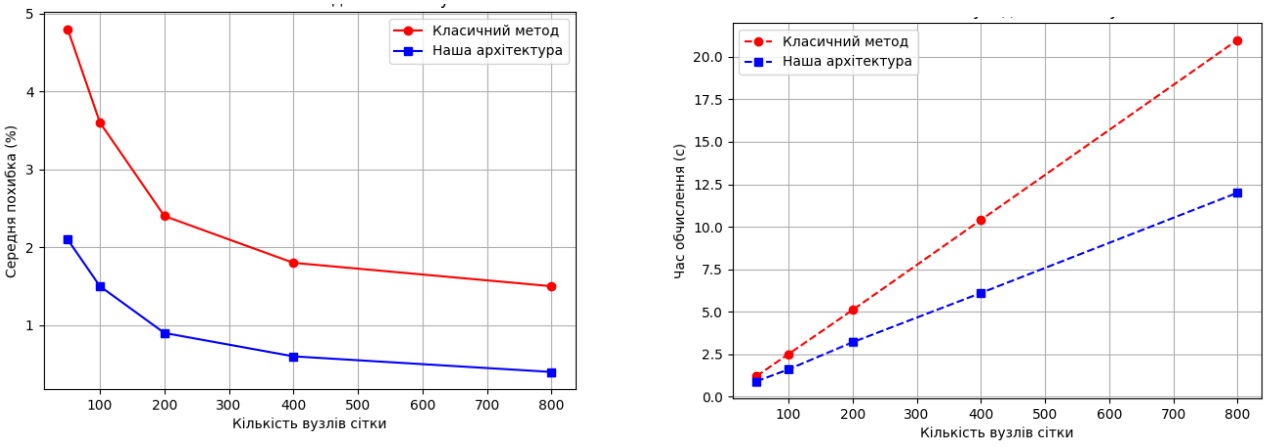
"Reference fine" використовується як еталон для оцінювання, тому похибки відсутні. В адаптивному випадку наведене *Nx* відповідає ефективному середньому числу вузлів за час інтегрування, Δx_{min} виникає локально біля інтерфейсів або в зонах різких градієнтів, тоді як Δx_{max} характерне для однорідних ділянок. Середня винагорода *R* наведена у відносних одиницях для параметрів $\alpha = 1$, $\beta = 0{,}1$; значення, ближчі до нуля і вищі, відображають успішне балансування між точністю та витратами, тоді як від'ємні вказують на штрафи за ресурси за ще значної похибки.

Результати показали, що адаптивний метод забезпечує L2-похибку приблизно 1,2–1,6% за прискорення 7–8 разів порівняно з еталонним розв'я-

Таблиця 1

Результати апробації архітектури на задачі одномірної теплопровідності в неоднорідному середовищі

Тест	Профіль *k (x)*	Метод	*Nx*	Δx_{min}	Δx_{max}	N_t	Час, с	L2-похибка, %	Макс. похибка, %	Прискорення vs. еталон
A	Двошаровий стрибок (інтерфейс у *x* = 0,5)	Reference fine	4 096	2,44e−4	2.44e−4	180 00	120,4	0,00	0,00	1,0×
A	Рівномірна груба	256	3,91e−3	3,91e−3	280	8,3	6,3	11,8	14,5×	–
A	**Adaptive (PPO+MLP)**	~420	4,88e−4	4,27e−3	120 00	15,2	1,3	2,7	7,9×	−0,42
B	Тришаровий профіль (2 інтерфейси)	Reference fine	4 096	2,44e−4	2,44e−4	18 000	139,6	0,00	0,00	1,0×
B	Рівномірна груба	256	3,91e−3	3,91e−3	280	9,1	5,8	10,6	15,3×	–
B	**Adaptive (PPO+MLP)**	~480	3,66e−4	3,91e−3	13 000	17,8	1,4	2,9	7,8×	−0,39
C	Гладкий синусоїдальний $k(x) = 1 + 0{,}8\sin(6\pi x)$	Reference fine	4 096	2,44e−4	2,44e−4	18 000	131,2	0,00	0,00	1,0×
C	Рівномірна груба	256	3,91e−3	3,91e−3	280	8,6	5,1	9,2	15,2×	–
C	**Adaptive (PPO + MLP)**	~450	4,27e−4	4,27e−3	12 500	16,1	1,2	2,4	8,1×	−0,37
D	Випадкова неоднорідність (корельований шум)	Reference fine	4 096	2,44e−4	2,44e−4	18 000	159,8	0,00	0,00	1,0×
D	Рівномірна груба	256	3,91e−3	3,91e−3	280	10,0	7,2	13,5	16,0×	–
D	**Adaptive (PPO + MLP)**	~520	3,05e−4	4,88e−3	14 000	21,9	1,6	3,2	7,3×	−0,45



А Б

Рис. 2. Залежність похибки (А) та часу (Б) від кількості вузлів

занням і демонструє значно кращий компроміс, ніж рівномірна груба сітка з похибкою 5–7%. Отже, застосування Advantage-архітектури дозволило автоматично зосередити обчислювальні ресурси в зонах різкої зміни температури, залишаючи крупніший крок у відносно однорідних областях.

Додатково було побудовано графіки для наочного відображення тенденцій зміни похибки та часу обчислень залежно від кількості шарів нейромережі. Вони дозволяють краще оцінити ефективність запропонованої архітектури та виявити оптимальний баланс між точністю розв'язання рівняння теплопровідності в неоднорідному середовищі й обчислювальними витратами.

Графіки показують, що зі зростанням кількості шарів нейромережі середня похибка поступово зменшується, що свідчить про здатність архітектури ефективніше апроксимувати складні залежності рівняння теплопровідності. Водночас час обчислень зростає майже лінійно, підтверджуючи компроміс між точністю та швидкодією. Найбільш раціональним є використання середньої глибини (приблизно 5–6 шарів), коли похибка вже суттєво знижується, але витрати часу ще залишаються прийнятними. Це демонструє збалансованість архітектури та її придатність для адаптації до практичних задач теплопровідності.

Додатково було побудовано рис. 3 для наочної демонстрації роботи запропонованої RL-архітектури, яка динамічно адаптує просторовий крок Δx під час чисельного моделювання рівняння теплопровідності.

Аналіз графіка на рисунку 3 показує, що розроблена RL-архітектура ефективно адаптує просторовий крок Δx у часі. У зонах з високими градієнтами температури Δx зменшується, що забезпечує підвищену точність чисельного розрахунку, тоді як

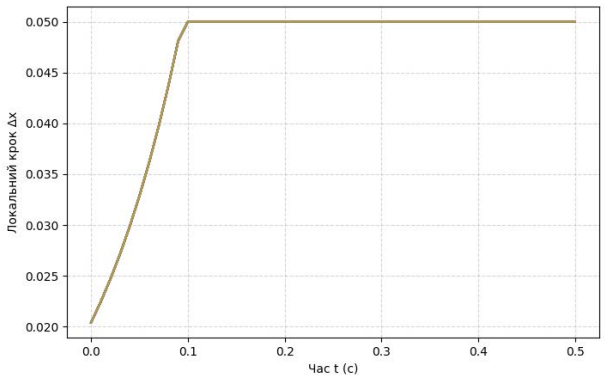


Рис. 3. Динаміка зміни сітки під керуванням RL

у більш однорідних областях сітка розріджується, зменшуючи обчислювальні витрати. Така динамічна адаптація дозволяє досягати оптимального балансу між точністю та ефективністю, відображає здатність системи виявляти критичні області та коректно реагувати на локальні особливості поля. Загальна тенденція демонструє стабільність і передбачуваність поведінки Δx протягом інтегрування, що підтверджує працездатність запропонованого методу.

Апробація архітектури на задачі хвильового рівняння в обмеженій області. Було розглянуто задачу математичного моделювання *хвильового рівняння у двовимірній прямокутній області з фіксованими крайовими умовами (типу Діріхле)*. Це класична задача математичної фізики, яка описує поширення хвильових процесів (звукових, електромагнітних або механічних) у замкненій області з відбиванням від меж. Для апробації було обрано гармонічний імпульс, що дозволяє чітко відстежити закономірність затухання амплітуди із часом та перевірити точність чисельного відтворення процесу.

Таблиця 2

Порівняння результатів для хвильового рівняння

Час t (с)	Амплітуда аналітична	Амплітуда модель	Відносна похибка (%)
0,1	0,980	0,966	1,43
0,2	0,861	0,849	1,39
0,3	0,701	0,690	1,57
0,4	0,532	0,525	1,32
0,5	0,369	0,364	1,36

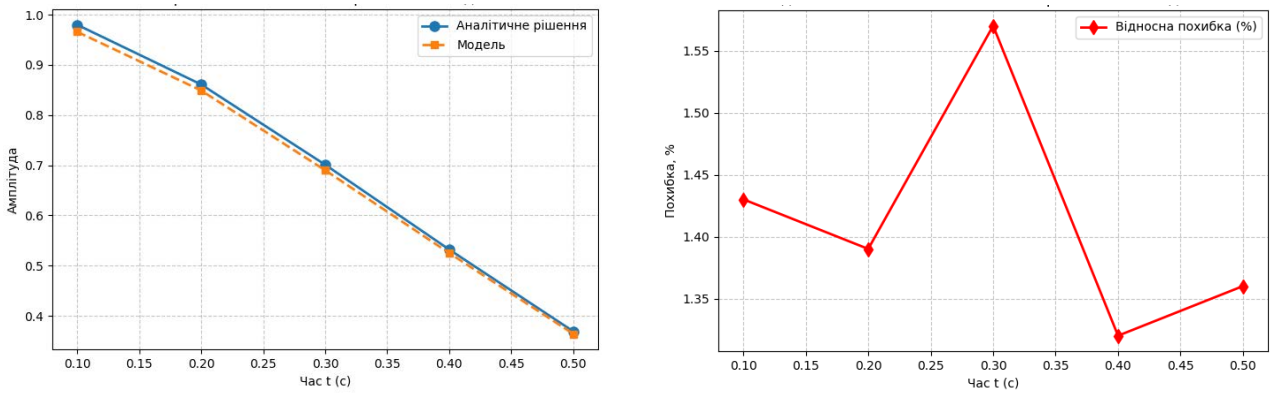
У таблиці 2 наведено порівняння значень амплітуди хвильового процесу, отриманих за допомогою аналітичного розв'язку та розробленої архітектури. Аналітичне рішення виступає в ролі еталона, тоді як чисельна архітектура дозволяє оцінити точність та стабільність відтворення хвильового сигналу в обмеженій області.

Бачимо, що всі значення амплітуд, отримані за допомогою розробленої архітектури, знаходяться дуже близько до аналітичних, що видно з відносної похибки, яка в усіх випадках не перевищує 1,6%. На початкових етапах часу (t = 0,1–0,2 с) похибка становить приблизно 1,4%, що свідчить про коректне відтворення початкового гармонічного імпульсу. Подальше зменшення амплітуди із часом (t = 0,3–0,5 с) також адекватно описується моделлю, максимальне відхилення зафіксовано на моменті t = 0,3 с (1,57%). Такий рівень похибки є незначним і може бути пояснений чисельними наближеннями, пов'язаними з дискретизацією області та часовим інтегруванням. Важливим спостереженням є збереження якісної динаміки процесу: і аналітичне рішення, і модель демонструють однакову тенденцію зменшення амплітуди в часі. Це свідчить про те, що архітектура не тільки правильно апроксимує точкові значення, а й коректно відтворює фізичні закономірності хвильового процесу.

Для наочного відображення результатів роботи архітектури було побудовано рисунки. Рисунок 3-А показує порівняння амплітуд хвильового процесу, отриманих за аналітичним розв'язком і розробленою архітектурою, у часі; рис 3-Б ілюструє відносну похибку між ними на кожному часовому кроці. Ці графіки дозволяють одночасно оцінити якість апроксимації та стійкість чисельного моделювання.

На наведених рисунках чітко видно узгодженість між аналітичним і модельним розв'язанням. Обидві криві повторюють однакову тенденцію затухання амплітуди в часі, що підтверджує здатність розробленої архітектури не лише наближати окремі значення, а й зберігати глобальні закономірності процесу. Невеликі відхилення спостерігаються у вигляді майже паралельного зсуву кривих, що узгоджується з відносними похибками з таблиці (не перевищують 1,6%). Графік похибки (рис. 3-Б) показує її стабільність і відсутність систематичного накопичення, що є свідченням стійкості обчислювального процесу. Отже, візуальний аналіз підтверджує кількісні результати й підкреслює адекватність моделі для відтворення динаміки хвильових процесів.

Точність. Для оцінювання ефективності запропонованої архітектури були використані три ключові метрики: середня похибка E, час виконання T та інтегральна ефективність η. У разі одномірного рівняння теплопровідності в неоднорідному середовищі середня L2-похибка чисельного розв'язку порівняно з еталонним ("Reference fine") становила



А Б

Рис. 4. Порівняння аналітичного рішення та моделі (А), відносна похибка (Б)

приблизно $E \approx 1{,}4\%$. Для порівняння, рівномірна груба сітка давала похибку $E_{base} \approx 5{,}8\%$. Час виконання за адаптивної архітектури скоротився майже в 7–8 разів: TML $\approx 0{,}13\,T_{base}$. Інтегральна ефективність η, обчислена за формулою

$$\eta = \frac{E_{base}}{E_{ML}} \cdot \frac{T_{base}}{T_{ML}}, \qquad (26)$$

становила приблизно $\eta \approx 31$, що демонструє значну перевагу адаптивної архітектури як за точністю, так і за обчислювальною ефективністю.

Для хвильового рівняння в обмеженій області середня відносна похибка по п'яти контрольних точках часу становила $E \approx 1{,}45\%$, тоді як базовий метод із рівномірною сіткою показав $E_{base} \approx 6{,}2\%$. Час виконання за використання адаптивної MLP + RL архітектури був скорочений приблизно у 6 разів, що дало $T_{ML} \approx 0{,}16\,T_{base}$. Відповідно, ефективність моделі для хвильового рівняння оцінюється як $\eta \approx 26$.

Отже, числові результати демонструють, що запропонована архітектура забезпечує суттєве підвищення точності чисельного моделювання за значного скорочення часу виконання, що підтверджує практичну ефективність інтеграції машинного навчання та RL в адаптивні чисельні методи.

Обговорення. Розроблена математична постановка задачі, що включає функціонал (формули (1) – (4)) з урахуванням похибки розрахунку та вартості обчислень, дозволила створити основу для впровадження методів машинного навчання в адаптивне чисельне моделювання. Використання архітектури багатошарового персептрона для прогнозування параметрів дискретизації (формули (5) – (7)) забезпечило ефективний вибір просторового та часових кроків, тоді як агент навчання з підкріпленням PPO (формули (8) – (12)) динамічно коригував ці параметри у процесі симуляції. На відміну від традиційних адаптивних методів [8; 10], де зміна кроку базується на евристичних правилах, запропонована схема враховує як локальні характеристики поля, так і обчислювальні витрати, що зменшує імовірність нераціонального використання ресурсів.

Блок-схема на рисунку 1 демонструє інтеграцію нейронної мережі та RL-агента в єдиний цикл адаптації, що дозволяє автоматично балансувати між точністю та швидкодією. Апробація на задачі теплопровідності в неоднорідному середовищі (табл. 1) показала, що середня L2-похибка адаптивної архітектури становить лише 1,2–1,6 % за прискорення обчислень у 7–8 разів щодо еталонного розв'язку, тоді як рівномірна груба сітка дає похибку 5–7% за значно нижчої ефективності. Це пояснюється автоматичним згущенням сітки в зонах з високими градієнтами (рис. 3), що відсутнє у класичних методах AMR [9].

На задачі хвильового рівняння (табл. 2) відносна похибка не перевищила 1,6% у всіх контрольних точках часу, що свідчить про стійке відтворення динаміки процесу (рис. 4-А, 4-Б). Отримані результати підтверджують, що запропонована архітектура не тільки точно апроксимує окремі значення, а й зберігає глобальні фізичні закономірності хвильових процесів, на відміну від методів [12; 15], які або потребують великих обсягів навчальних даних, або працюють лише для окремих класів задач.

Важливим є також аналіз залежності похибки й часу від конфігурації нейронної мережі (рис. 2). Показано, що збільшення кількості шарів знижує похибку, але підвищує обчислювальні витрати. Оптимальною є середня глибина (5–6 шарів), що забезпечує баланс між точністю та продуктивністю, тоді як занадто глибокі мережі ведуть до перенавантаження системи без суттєвого приросту точності. Отже, результати чисельних експериментів підтвердили ефективність гібридної ML + RL архітектури: досягнуто суттєвого зниження похибки, підвищення швидкодії та узгодження моделі з фізичною природою процесів. Це узгоджується з висновками сучасних робіт [3; 6; 9], але перевершує їх завдяки поєднанню попереднього прогнозування та динамічної онлайн-адаптації. Порівняно із [25], де застосовується строгий аналітичний метод до адаптації чисельних схем, запропонована методологія має перевагу в можливості самонавчання на основі даних і поступового вдосконалення стратегії вибору параметрів.

Робота унікальна, оскільки поєднує класичні адаптивні чисельні методи із сучасними інструментами машинного навчання, зокрема багатошаровими нейронними мережами й алгоритмами навчання з підкріпленням. Запропонована гібридна архітектура є новою тим, що дозволяє не лише попередньо прогнозувати параметри дискретизації, але й динамічно їх коригувати у процесі моделювання, забезпечувати адаптацію під конкретні особливості фізичної задачі. Щодо переваг, запропонована архітектура демонструє значне зменшення похибки розрахунків (до 1,2–1,6%) за суттєвого скорочення часу виконання (у 6–8 разів) порівняно із традиційними методами. Вона автоматично концентрує обчислювальні ресурси у критичних областях із різкими градієнтами, зберігає ефективність в однорідних ділянках. Завдяки гібридному поєднанню ML і RL забезпечуються висока точність, стабільність та універсальність моделювання, що робить метод перспективним для широкого класу складних фізичних процесів.

До основних обмежень архітектури варто віднести потребу значних обчислювальних ресурсів на етапі тренування нейронних мереж та RL-агента, а також наявності якісних еталонних даних для валідації. Окрім того, результати значною мірою залежать від вибору функції винагороди та

параметрів навчання, що може обмежувати універсальність методу.

Основним недоліком є складність реалізації гібридної архітектури та необхідність налаштування великої кількості параметрів, що робить її менш придатною для задач з обмеженим доступом до даних чи ресурсів. Також існує ризик перенавчання нейронної мережі, що може знизити стійкість у разі застосування до нових класів задач.

Подальший розвиток дослідження пов'язаний із застосуванням Physics-Informed Neural Networks для безпосереднього врахування фізичних законів, розширенням апробації на багатовимірні та нелінійні системи, а також створенням більш універсальних адаптивних алгоритмів, здатних працювати в реальному часі для складних інженерних і природничих задач.

Висновки. У роботі запропоновано методику інтеграції машинного навчання з адаптивними чисельними методами для моделювання складних фізичних процесів. Проведені експерименти показали, що розроблена архітектура на основі багатошарового персептрона й алгоритму підкріплювального навчання PPO здатна ефективно керувати параметрами дискретизації. На задачах теплопровідності було досягнуто зниження середньої похибки розрахунків до рівня приблизно 1,4% за прискорення обчислень у 7–8 разів порівняно з високоточним еталонним розв'язком. Для хвильового рівняння середня відносна похибка становила приблизно 1,45%, а швидкодія підвищилася в 6 разів, що підтвердило універсальність і стійкість підходу. Отримані результати доводять, що використання ML та RL забезпечує суттєве покращення балансу між точністю та продуктивністю порівняно із класичними методами. Найбільш ефективною виявилася гібридна архітектура, яка поєднує попереднє прогнозування нейронною мережею та онлайн-корекцію через RL-агента. Подальші дослідження доцільно зосередити на використанні Physics-Informed Neural Networks для прямого врахування фізичних законів і розширенні апробації на багатовимірні нелінійні задачі.

ЛІТЕРАТУРА

1. Ferreira V., Santos L., Simoes R., Franzen M., Ghouati O. Estimating local thickness for finite element analysis. 2010. P. 63–68.
2. Simos T. E. New Variable-Step Procedure for the Numerical Integration of the One-Dimensional Schrödinger Equation. *Journal of Computational Physics*. 1993. Vol. 108. Iss. 1. P. 175–179. https://doi.org/10.1006/jcph.1993.1172
3. Faegh M., Ghungrad S., Oliveira J. P., Rao P., Haghighi A. A review on physics-informed machine learning for process-structure-property modeling in additive manufacturing. *Journal of Manufacturing Processes*. 2025. Vol. 133. P. 524–555. https://doi.org/10.1016/j.jmapro.2024.11.066
4. Bilak Y. Information system based on a complex model using machine learning for spectral analysis. *Bulletin of Taras Shevchenko National University of Kyiv. Physical and Mathematical Sciences*. 2025. Vol. 80. № 1. P. 104–114. https://doi.org/10.17721/1812-5409.2025/1.14
5. Bilak Y. Modeling, optimization and AI-forecasting technology in Raman spectrometry. *Problems of Control and Informatics*. 2025. Vol. 70. № 2. P. 99–112. https://doi.org/10.34229/1028-0979-2025-2-9
6. Bilak Y. Y. Development of a combined model for analyzing gas mixtures using machine learning methods. *Applied Aspects of Information Technology*. 2025. Vol. 8, № 1. P. 24–37. https://doi.org/10.15276/aait.08.2025.2
7. Martín-Guerrero J., Lamata L. Reinforcement Learning and Physics. *Applied Sciences*. 2021. Vol. 11. Art. 8589. https://doi.org/10.3390/app11188589
8. Irigaray O., Ansa Z., Fernandez-Gamiz U., Larrinaga A., García-Fernandez R., Portal-Porras K. Adaptive mesh refinement (AMR) criteria comparison for the DrivAer model. *Heliyon*. 2024. Vol. 10. Iss. 11. P. e31966. https://doi.org/10.1016/j.heliyon.2024.e31966
9. Freymuth N., Dahlinger P., Würth T., Reisch S., Kärger L., Neumann G. Swarm reinforcement learning for adaptive mesh refinement. *Advances in Neural Information Processing Systems*. 2023. Vol. 36. P. 73312–73347.
10. Ranocha H., Dalcin L., Parsani M., Ketcheson D. Optimized Runge-Kutta Methods with Automatic Step Size Control for Compressible Computational Fluid Dynamics. *Communications on Applied Mathematics and Computation*. 2021. Vol. 4. https://doi.org/10.1007/s42967-021-00159-w
11. Henson V. Multigrid methods for nonlinear problems: An overview. *Proceedings of SPIE – The International Society for Optical Engineering*. 2003. https://doi.org/10.1117/12.499473
12. Raissi M., Perdikaris P., Karniadakis G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*. 2019. Vol. 378. P. 686–707. https://doi.org/10.1016/j.jcp.2018.10.045
13. Ger S., Jambunath Y. S., Klabjan D. Autoencoders and generative adversarial networks for imbalanced sequence classification. *2023 IEEE International Conference on Big Data (BigData)*. 2023. P. 1101–1108. https://doi.org/10.48550/arXiv.1901.02514

14. Ghojogh B., Ghodsi A., Karray F., Crowley M. Generative Adversarial Networks and Adversarial Autoencoders: Tutorial and Survey. *ArXiv*. 2021. abs/2111.13282. https://doi.org/10.48550/arXiv.2111.13282
15. Sutton R. S., Barto A. G. Reinforcement Learning. An Introduction. 2nd ed. Cambridge : A Bradford Book, 2018.
16. Zhang C., Lu Y. Study on artificial intelligence: The state of the art and future prospects. *Journal of Industrial Information Integration*. 2021. Vol. 23. P. 100224. https://doi.org/10.1016/j.jii.2021.100224
17. Strikwerda J. C. Finite Difference Schemes and Partial Differential Equations. 2nd ed. Philadelphia: Society for Industrial and Applied Mathematics, 2004. https://doi.org/10.1137/1.9780898717938
18. Leveque R. J. Finite Volume Methods for Hyperbolic Problems. Cambridge : Cambridge University Press, 2002. https://doi.org/10.1017/CBO9780511791253
19. Berger M. J., Colella P. Local adaptive mesh refinement for shock hydrodynamics. *Journal of Computational Physics*. 1989. Vol. 82. Iss. 1. P. 64–84. https://doi.org/10.1016/0021-9991(89)90035-1
20. Sutton R. S., Barto A. G. Reinforcement Learning: An Introduction. *IEEE Transactions on Neural Networks*. 1998. Vol. 9. № 5. P. 1054–1054. https://doi.org/10.1109/TNN.1998.712192
21. Heaton J. Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning: The MIT Press, 2016, 800 p. *Genetic Programming and Evolvable Machines*. 2017. Vol. 19. https://doi.org/10.1007/s10710-017-9314-z
22. Ioffe S., Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *International Conference on Machine Learning*. 2015. P. 448–456. https://doi.org/10.48550/arXiv.1502.03167
23. Schulman J., Wolski F., Dhariwal P., Radford A., Klimov O. Proximal policy optimization algorithms. *arXiv preprint*. 2017. arXiv : 1707.06347
24. Han J., Jentzen A., Ee W. Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*. 2017. Vol. 115. https://doi.org/10.1073/pnas.1718942115
25. Білак Ю., Геращенков Е. Багаторівнева декомпозиція для адаптивного чисельного моделювання складних фізичних процесів. *Herald of Khmelnytskyi National University. Technical Sciences*. 2025. Vol. 349. № 2. P. 51–62. https://doi.org/10.31891/2307-5732-2025-349-7

REFERENCES

1. Ferreira V., Santos L., Simoes R., Franzen M., Ghouati O. Estimating local thickness for finite element analysis. 2010. P. 63–68.
2. Simos T. E. New Variable-Step Procedure for the Numerical Integration of the One-Dimensional Schrödinger Equation. *Journal of Computational Physics*. 1993. Vol. 108, Iss. 1. P. 175–179. https://doi.org/10.1006/jcph.1993.1172
3. Faegh M., Ghungrad S., Oliveira J. P., Rao P., Haghighi A. A review on physics-informed machine learning for process-structure-property modeling in additive manufacturing. *Journal of Manufacturing Processes*. 2025. Vol. 133. P. 524–555. https://doi.org/10.1016/j.jmapro.2024.11.066
4. Bilak Y. Information system based on a complex model using machine learning for spectral analysis. *Bulletin of Taras Shevchenko National University of Kyiv. Physical and Mathematical Sciences*. 2025. Vol. 80, № 1. P. 104–114. https://doi.org/10.17721/1812-5409.2025/1.14
5. Bilak Y. Modeling, optimization and AI-forecasting technology in Raman spectrometry. *Problems of Control and Informatics*. 2025. Vol. 70, № 2. P. 99–112. https://doi.org/10.34229/1028-0979-2025-2-9
6. Bilak Y.Y . Development of a combined model for analyzing gas mixtures using machine learning methods. *Applied Aspects of Information Technology*. 2025. Vol. 8, № 1. P. 24–37. https://doi.org/10.15276/aait.08.2025.2
7. Martín-Guerrero J., Lamata L. Reinforcement Learning and Physics. *Applied Sciences*. 2021. Vol. 11. Art. 8589. https://doi.org/10.3390/app11188589
8. Irigaray O., Ansa Z., Fernandez-Gamiz U., Larrinaga A., García-Fernandez R., Portal-Porras K. Adaptive mesh refinement (AMR) criteria comparison for the DrivAer model. *Heliyon*. 2024. Vol. 10, Iss. 11. e31966. https://doi.org/10.1016/j.heliyon.2024.e31966
9. Freymuth N., Dahlinger P., Würth T., Reisch S., Kärger L., Neumann G. Swarm reinforcement learning for adaptive mesh refinement. *Advances in Neural Information Processing Systems*. 2023. Vol. 36. P. 73312–73347.
10. Ranocha H., Dalcin L., Parsani M., Ketcheson D. Optimized Runge-Kutta Methods with Automatic Step Size Control for Compressible Computational Fluid Dynamics. *Communications on Applied Mathematics and Computation*. 2021. Vol. 4. https://doi.org/10.1007/s42967-021-00159-w

11. Henson V. Multigrid methods for nonlinear problems: An overview. *Proceedings of SPIE – The International Society for Optical Engineering*. 2003. https://doi.org/10.1117/12.499473
12. Raissi M., Perdikaris P., Karniadakis G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*. 2019. Vol. 378. P. 686–707. https://doi.org/10.1016/j.jcp.2018.10.045
13. Ger S., Jambunath Y. S., Klabjan D. Autoencoders and generative adversarial networks for imbalanced sequence classification. *2023 IEEE International Conference on Big Data (BigData)*. 2023. P. 1101–1108. https://doi.org/10.48550/arXiv.1901.02514
14. Ghojogh B., Ghodsi A., Karray F., Crowley M. Generative Adversarial Networks and Adversarial Autoencoders: Tutorial and Survey. *ArXiv*. 2021. abs/2111.13282. https://doi.org/10.48550/arXiv.2111.13282
15. Sutton R. S., Barto A. G. Reinforcement Learning. An Introduction. 2nd ed. Cambridge: A Bradford Book, 2018.
16. Zhang C., Lu Y. Study on artificial intelligence: The state of the art and future prospects. *Journal of Industrial Information Integration*. 2021. Vol. 23. 100224. https://doi.org/10.1016/j.jii.2021.100224
17. Strikwerda J. C. Finite Difference Schemes and Partial Differential Equations. 2nd ed. Philadelphia: Society for Industrial and Applied Mathematics, 2004. https://doi.org/10.1137/1.9780898717938
18. Leveque R. J. Finite Volume Methods for Hyperbolic Problems. Cambridge: Cambridge University Press, 2002. https://doi.org/10.1017/CBO9780511791253
19. Berger M. J., Colella P. Local adaptive mesh refinement for shock hydrodynamics. *Journal of Computational Physics*. 1989. Vol. 82, Iss. 1. P. 64–84. https://doi.org/10.1016/0021-9991(89)90035-1
20. Sutton R. S., Barto A. G. Reinforcement Learning: An Introduction. *IEEE Transactions on Neural Networks*. 1998. Vol. 9, № 5. P. 1054–1054. https://doi.org/10.1109/TNN.1998.712192
21. Heaton J. Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning: The MIT Press, 2016, 800 pp. *Genetic Programming and Evolvable Machines*. 2017. Vol. 19. https://doi.org/10.1007/s10710-017-9314-z
22. Ioffe S., Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *International Conference on Machine Learning*. 2015. P. 448–456. https://doi.org/10.48550/arXiv.1502.03167
23. Schulman J., Wolski F., Dhariwal P., Radford A., Klimov O. Proximal policy optimization algorithms. *arXiv preprint*. 2017. arXiv: 1707.06347.
24. Han J., Jentzen A., Ee W. Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*. 2017. Vol. 115. https://doi.org/10.1073/pnas.1718942115
25. Bilak Y., Herashchenkov E. Multilevel decomposition for adaptive numerical modeling of complex physical processes. *Herald of Khmelnytskyi National University. Technical Sciences*. 2025. Vol. 349, № 2. P. 51–62. https://doi.org/10.31891/2307-5732-2025-349-7

Дата першого надходження рукопису до видання: 21.08.2025
Дата прийнятого до друку рукопису після рецензування: 23.09.2025
Дата публікації: 31.12.2025

УДК 004.056.5: [004.896:728.37]
DOI https://doi.org/10.26661/2786-6254-2025-2-08

ІНТЕГРОВАНА СИСТЕМА БЕЗПЕКИ ДЛЯ РОЗУМНОГО БУДИНКУ НА ОСНОВІ GSM-СИГНАЛІЗАЦІЇ

Драєвський Д. С.
викладач
Відокремлений структурний підрозділ «Економіко-правничий фаховий коледж Запорізького національного університету»
вул. Університетська, 66, Запоріжжя, Україна
orcid.org/0009-0008-3810-6595
draevskdmitry@gmail.com

Мухін В. В.
кандидат технічних наук, доцент,
доцент кафедри програмної інженерії
Запорізький національний університет
вул. Університетська, 66, Запоріжжя, Україна
orcid.org/0009-0007-0270-8239
comentssno@gmail.com

Мильцев О. М.
кандидат фізико-математичних наук,
доцент кафедри програмної інженерії
Запорізький національний університет
вул. Університетська, 66, Запоріжжя, Україна
orcid.org/0009-0009-4382-7730
alexmyltsev@gmail.com

Столярова А. В.
доктор філософії,
доцент кафедри програмної інженерії
Запорізький національний університет
вул. Університетська, 66, Запоріжжя, Україна
orcid.org/0000-0003-2783-2889
stolyarova.a.v@znu.edu.ua

Ключові слова: *інтернет речей, мікроконтролер, GSM-сигналізація, проєктування систем, ESP 32.*

З огляду на зростання важливості надійних та автономних систем безпеки в епоху розумних будинків, у цій роботі представлено розроблення та аналіз інтегрованої системи, що використовує GSM-сигналізацію для забезпечення ефективного моніторингу та захисту житлових приміщень. Розглянуто принципи побудови такої системи, зокрема й проєктування принципової схеми пристроїв інтернету речей (IoT) для GSM-сигналізації та розроблення відповідного програмного забезпечення. Проаналізовано огляд наявних рішень у сфері безпеки розумних будинків з використанням GSM-технологій, що демонструє їхню ефективність у сповіщенні про загрози та дистанційному реагуванні.

Описано архітектуру розробленої тестової системи, яка включає мікроконтролер NodeMCU ESP32, GSM-модуль SIM800L, датчик руху

PIR, датчик температури та вологості DHT11 та інші компоненти. Представлено алгоритм роботи системи, що підтримує режими охорони та тривоги, керування за допомогою DTMF-команд, сповіщення через SMS та телефонні дзвінки, а також моніторинг наявності мережевої напруги та параметрів навколишнього середовища. Підкреслено актуальність та ефективність інтеграції GSM-сигналізації в системах безпеки розумних будинків, її автономність та можливості розширення завдяки технологіям IoT. Результати дослідження демонструють потенціал розробленої системи для підвищення рівня безпеки житлових приміщень шляхом автоматизованого й оперативного сповіщення користувачів про загрози.

INTEGRATED SECURITY SYSTEM FOR A SMART HOME BASED ON GSM ALARMS

Draevskii D. S.
Lecturer
Separate Structural Subdivision "Economic and Legal Vocational College of Zaporizhzhia National University"
Universytetska str., 66, Zaporizhzhia, Ukraine
orcid.org/0009-0008-3810-6595
draevskdmitry@gmail.com

Mukhin V. V.
Candidate of Technical Sciences, Associate Professor,
Professor at the Software Engineering Department
Zaporizhzhia National University
Universytetska str., 66, Zaporizhzhia, Ukraine
orcid.org/0009-0007-0270-8239
comentssno@gmail.com

Myltsev O. M.
Candidate of Physical and Mathematical Sciences,
Associate Professor at the Software Engineering Department
Zaporizhzhia National University
Universytetska str., 66, Zaporizhzhia, Ukraine
orcid.org/0009-0009-4382-7730
alexmyltsev@gmail.com

Stoliarova A. V.
PhD, Associate Professor at the Department of Software Engineering
Zaporizhzhia National University
Universytetska str., 66, Zaporizhzhia, Ukraine
orcid.org/0000-0003-2783-2889
stolyarova.a.v@znu.edu.ua

Key words: *Internet of Things, microcontroller, GSM alarm, system design, ESP32.*

Given the increasing importance of reliable and autonomous security systems in the era of smart homes, this document presents the development and analysis of an integrated system that uses GSM alarms to ensure effective monitoring and protection of residential premises. The principles of building such a system are examined, including the design of the circuit for Internet of Things (IoT) devices for GSM alarms and the development of the corresponding software. An overview of existing solutions in the field of smart home security using GSM technologies is provided, demonstrating their effectiveness in threat notification and remote response.

The architecture of the developed test system is described, which includes a NodeMCU ESP32 microcontroller, a SIM800L GSM module, a PIR motion sensor, a DHT11 temperature and humidity sensor, and other components. The paper presents the system's operating algorithm, supporting security and alarm modes, control via DTMF commands, notifications through SMS and phone calls, as well as monitoring of the availability of network voltage and environmental parameters.
Particular emphasis is placed on the relevance and effectiveness of integrating GSM alarms into smart home security systems, highlighting their autonomy and potential for expansion through IoT technologies. The research results demonstrate the potential of the developed system to enhance the security level of residential premises by providing automated and prompt notification of threats to users.

Вступ. У сучасному світі питання гарантування безпеки житлових приміщень набуває все більшого значення. Зростання кількості розумних будинків і розвиток технологій автоматизації сприяють впровадженню інноваційних методів контролю та захисту. Одним із найефективніших рішень для убезпечення є інтеграція GSM-сигналізації в системі охорони розумного будинку. Дана технологія дозволяє здійснювати моніторинг у реальному часі, передавати сповіщення про загрози, виконувати автоматичні дзвінки на мобільні пристрої власників і забезпечувати оперативне реагування на надзвичайні ситуації [1].

Використання GSM-зв'язку в системах безпеки дає можливість створювати автономні рішення, які не залежать від дротового інтернет-з'єднання або локальних мереж. Окрім того, інтеграція GSM із технологіями інтернету речей (далі – IoT) відкриває нові можливості для підвищення ефективності та надійності систем захисту житла [2]. Останні дослідження демонструють, що GSM-сигналізація в поєднанні з розумними датчиками руху й автоматичними дзвінками власникам дозволяє суттєво покращити рівень безпеки житлових приміщень [3].

Об'єктом дослідження є інтегрована система безпеки для розумного будинку, що використовує GSM-сигналізацію для передавання повідомлень, виконання дзвінків і моніторингу стану безпеки.

Метою дослідження є аналіз і розроблення інтегрованої системи безпеки для розумного будинку на основі GSM-сигналізації, яка дозволяє забезпечити ефективний моніторинг і захист житлового приміщення шляхом автоматизованого сповіщення користувачів про загрози через повідомлення та телефонні дзвінки.

Для досягнення поставленої мети необхідно виконати такі завдання: спроєктувати принципову схему пристроїв IoT для GSM-сигналізації; розробити програмну реалізацію GSM-сигналізації.

Огляд літератури. Огляд літератури охоплює низку досліджень та розробок у галузі безпеки розумних будинків, зокрема через застосування GSM-сигналізації, що дозволяє забезпечити ефективний моніторинг і контроль за станом безпеки на відстані.

У статті [4] представлено автоматизовану систему виявлення вторгнень для будинків і офісів, яка поєднує різні сенсори, як-от PIR-датчики та магнітні перемикачі, із GSM-модулем для сповіщення власників про несанкціонований доступ через телефонні дзвінки.

У роботі [5] запропоновано систему безпеки для розумного будинку, яка використовує GSM для надсилання SMS власникам у разі виявлення аномалій сенсорами. Система також включає механізм автентифікації для забезпечення доступу лише авторизованих користувачів і мережеву систему виявлення вторгнень для сповіщення про підозрілу активність у мережі IoT.

У [6] розглянуто концепцію «розумного дому» – автоматизованої системи для керування побутовими електричними пристроями. Основна мета – підвищення комфорту, безпеки та зручності для мешканців. Система дозволяє автоматично керувати навантаженнями (приладами), виявляти пожежу, контролювати температуру, розпізнавати рух, управляти замками дверей. Окрім того, система має розширені функції безпеки: у разі підозрілої події в оселі надсилає повідомлення користувачу. Також передбачено дистанційне керування приладами через мобільний телефон з використанням технології GSM. Загалом, робота демонструє, як автоматизація побуту може значно покращити якість життя, підвищити безпеку та дати змогу користувачеві контролювати дім навіть на відстані.

Актуальність систем домашньої безпеки в умовах зростання загроз, як-от вторгнення, витік газу чи пожежа, розглядається у статті [7]. Традиційні сигналізації обмежуються лише звуковим попередженням, тоді як GSM-технології дозволяють значно розширити можливості безпеки. Запропоновано дві системи захисту: систему з вебкамерою – реагує на рух, видає звуковий сигнал і надсилає власнику електронного листа, систему на базі GSM-GPS-модуля (SIM548c) з мікроконтролером Atmega644p – у разі спрацювання датчиків надсилає SMS власнику, активує реле та звуковий сигнал.

Отже, робота демонструє ефективне використання сучасних технологій для оперативного сповіщення та дистанційного реагування на загрози в будинку.

Систему керування домашніми приладами (світло, вентилятор, холодильник та інші) та вияв-

лення підозрілої активності, зокрема крадіжок, за допомогою PIR-датчика руху представлено в роботі [8]. Ключові особливості: у разі виявлення руху в домі за відсутності власника PIR-датчик активує камеру, яка робить знімок і надсилає його користувачу через Raspberry Pi; водночас через GSM-модуль надсилається SMS; мікроконтролер PIC16F877A дозволяє дистанційно вмикати/вимикати прилади через мобільний телефон, що допомагає економити електроенергію; у разі пожежі спеціальний датчик передає сигнал користувачу, якщо температура перевищує норму.

Система відзначається низьким енергоспоживанням, швидкою обробкою сигналів, надійністю та зручністю доступу з будь-якої точки через GSM-зв'язок.

Розроблення системи безпеки для дому, яка забезпечує автоматичне сповіщення власника та відповідних служб (поліції, пожежної та інших) у разі спроби несанкціонованого проникнення, розглядається в роботі [9].

Основна ідея – створити систему, яка працює автономно, без потреби постійного контролю з боку користувача. У разі небезпеки контролер активує GSM-модуль, що надсилає SMS-повідомлення власнику та службам реагування.

Отже, система дозволяє поєднати домашню автоматизацію з ефективним захистом, водночас залишається доступною за ціною.

Стаття [10] присвячена системі автоматизації дому на базі Arduino Mega2560, яка поєднує керування приладами з функціями безпеки. Система дозволяє автоматично керувати домашніми пристроями, а також забезпечує захист від вторгнень, пожежі, витоку газу та CO. Використовує такі компоненти: датчики (руху, диму, газу), GSM-модуль для надсилання SMS у разі небезпеки, реле для керування приладами, звукові сигнали (бузери), мобільний застосунок для дистанційного керування.

Система забезпечує доступність, надійність і зручність, демонструє переваги сучасних мікропроцесорних технологій для створення розумного будинку за низькою вартістю.

Робота [11] представляє бездротову систему безпеки для дому на основі GSM-модуля, яка вирізняється низьким енергоспоживанням і швидкою реакцією на вторгнення. Основні елементи: магнітний датчик з реле на дверях, що фіксує спробу входу, GSM-модем, який миттєво надсилає SMS або здійснює дзвінок власнику в разі виявлення загрози.

Система дозволяє контролювати безпеку дому з будь-якого місця через мобільну мережу й інформує власника про підозрілу активність заздалегідь запрограмованим повідомленням.

Проведений огляд літератури демонструє значний інтерес і активне розроблення систем домашньої безпеки на основі GSM-технологій. Наявні рішення варіюються від простих систем сповіщення про вторгнення до комплексних платформ керування розумним будинком з розширеними функціями безпеки, як-от виявлення пожежі чи витоку газу та дистанційне керування приладами. Спільними рисами багатьох розробок є використання GSM-модулів для оперативного сповіщення власників про небезпеку через SMS або телефонні дзвінки, а також прагнення до автономності, низького енергоспоживання та зручного дистанційного керування.

Однак, незважаючи на значну кількість існуючих рішень, огляд літератури також виявляє потенційні напрями для вдосконалення та актуальність розроблення власної GSM-сигналізації. Аналіз показав, що існуючі системи часто є спеціалізованими або мають обмежену гнучкість у налаштуванні й інтеграції.

Розроблення доступної за ціною GSM-сигналізації з функцією постійного моніторингу температури зумовлене низкою важливих соціально-економічних і практичних факторів, особливо в умовах нестабільної економічної ситуації.

Функція постійного моніторингу температури в поєднанні із GSM-сповіщеннями надає власникам можливість дистанційно відстежувати температурний режим у своїх оселях. Це особливо актуально в разі тривалої відсутності (відрядження, відпустка, сезонне проживання на дачі). Отримання своєчасного сповіщення про критичне зниження температури дозволяє оперативно вжити заходів для запобігання серйозним наслідкам, наприклад, попросити сусідів перевірити стан опалення або дистанційно перезапустити систему, якщо це можливо.

Завчасне виявлення проблем з опаленням завдяки моніторингу температури може допомогти уникнути значних фінансових втрат, пов'язаних з ремонтом або заміною пошкодженого обладнання та ліквідацією наслідків замерзання систем водопостачання та опалення.

Методи. Архітектурні особливості IoT-пристроїв і відповідного програмного забезпечення розглянуто на прикладі системи GSM-сигналізації. Для створення тестового проєкту використовувалися широкодоступні IoT-пристрої, що не потребують спеціальних налаштувань.

До складу системи входять такі компоненти: мікроконтролер NodeMCU ESP32 – центральний керівний модуль, який отримує сигнали від датчиків і надсилає команди іншим пристроям; детектор напруги з оптопарою – контролює наявність змінного струму 220 В та передає відповідний сигнал мікроконтролеру; модуль SIM800L – забезпечує зв'язок із GSM-мережею, виконує надсилання SMS, приймає вхідні дзвінки та здійснює

виклики; DFPlayer Mini – використовується для відтворення заздалегідь записаних голосових повідомлень; датчик руху PIR – виявляє рух у контрольованій зоні й активує тривожний сигнал; датчик температури та вологості DHT11 – фіксує параметри навколишнього середовища, які можуть бути включені до системного звіту; світлодіод із резистором – забезпечує візуальну індикацію стану системи: увімкнена охорона, активована тривога або система вимкнена.

Рисунок 1 відображає електричну схему пристроїв, які використовуються для системи GSM-сигналізації.

Для демонстрації програмної реалізації GSM-сигналізації розглянемо ключові етапи розроблення програмного забезпечення для мікроконтролера, який керуватиме GSM-модулем і датчиками.

Програмна реалізація GSM-сигналізації передбачає розроблення таких основних модулів:

1. Ініціалізація обладнання (setup()). Цей модуль відповідає за початкове налаштування всіх апаратних компонентів системи, включаючи послідовні порти, GSM-модуль, аудіоплеєр, датчики та піни введення – виведення:

```
void setup() { Serial.begin(115200); // Налаштування SIM800L
sim800.begin(9600, SERIAL_8N1, SIM800_RX, SIM800_TX);
sendATCommand("AT");
sendATCommand("AT+CLIP=1"); // Відображення номера вхідного дзвінка
sendATCommand("AT+CMGF=1"); // Текстовий режим SMS
sendATCommand("AT+DDET=1"); // Увімкнути детекцію DTMF
Serial.println("SIM800L готовий");
// Налаштування DFPlayer Mini
dfPlayerSerial.begin(9600, SERIAL_8N1, DFPLAYER_RX, DFPLAYER_TX);
if (!dfPlayer.begin(dfPlayerSerial)) { Serial.println("Помилка ініціалізації DFPlayer Mini!");
while (true); }
dfPlayer.volume(30); Serial.println("DFPlayer Mini готовий."); pinMode(PIR_PIN, INPUT); pinMode(VOLTAGE_PIN, INPUT_PULLUP); pinMode(LED_PIN, OUTPUT); dht.begin();
updateLED(); // Відобразити початковий стан
Serial.println("Система сигналізації готова."); }
```

2. Обробка вхідних дзвінків (handleIncomingCall()). Цей модуль відповідає за прослуховування вхідних дзвінків, ідентифікацію номера абонента й обробку DTMF-команд, що надходять під час розмови.

```
void handleIncomingCall() {
if (sim800.available()) { String response = sim800.readString();
Serial.println("Відповідь SIM800L: " + response);
```

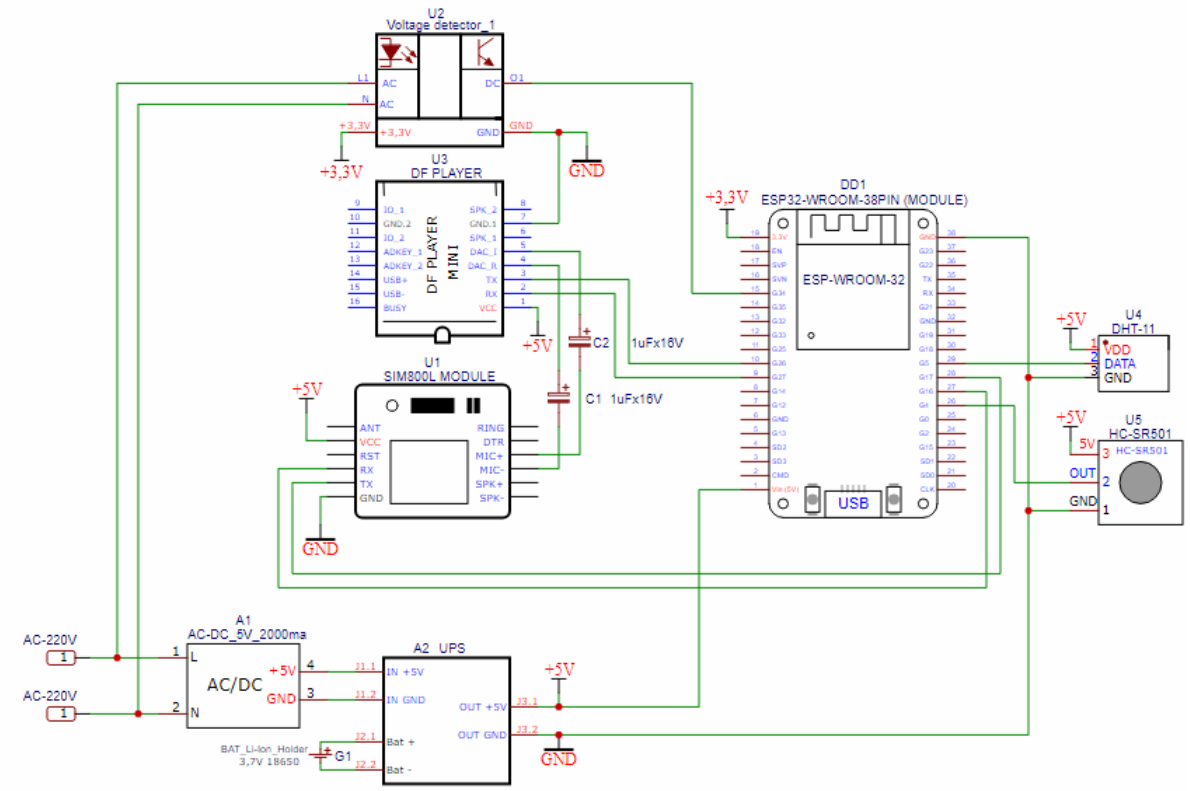


Рис. 1. Електрична схема GSM сигналізації

```
if (response.indexOf("+CLIP:") != -1) { String caller = extractNumber(response);
if (isNumberAllowed(caller)) { answerCall(); delay(1000); playVoiceMenu();
} else { rejectCall(); }}
// Обробка DTMF-входів
if (response.indexOf("+DTMF:") != -1) {
char dtmf = response.charAt(response.indexOf("+DTMF:") + 7);
processDTMFCommand(dtmf); }}}
```

3. Керування голосовими повідомленнями (playVoiceMenu(), playVoiceMessage()). Цей модуль відповідає за відтворення попередньо записаних голосових повідомлень через аудіоплеєр DFPlayer Mini для інформування користувача про стан системи або для голосового меню керування.

```
/ Відтворити голосове меню
void playVoiceMenu() {
if (isAlarmTriggered) { playVoiceMessage(1); // Тривога
} else if (isArmed) { playVoiceMessage(2); // Охорона ввімкнена
} else { playVoiceMessage(3); // Охорона вимкнена
}}
// Відтворення голосового повідомлення через DFPlayer
void playVoiceMessage(int fileNumber) {
dfPlayer.play(fileNumber); delay(6000); }
```

4. Обробка DTMF-команд (processDTMFCommand()). Цей модуль інтерпретує натискання кнопок на телефоні (DTMF-сигнали) під час вхідного дзвінка та виконує відповідні дії, як-от зміна стану охорони, скидання тривоги або отримання звіту.

```
// Обробка DTMF-команд
void processDTMFCommand(char dtmf) {
Serial.println("Отримано DTMF: " + String(dtmf));
switch (dtmf) { case '0': playVoiceMenu(); break;
case '1': if (isAlarmTriggered) { isAlarmTriggered = false; updateLED();
playVoiceMessage(4); // Тривогу зупинено
} break;
case '2': isArmed = true; updateLED();
playVoiceMessage(2); // Охорона ввімкнена
break;
case '3': isArmed = false; updateLED();
playVoiceMessage(3); // Охорона вимкнена
break;
case '6': sendSystemReport(); break; default: Serial.println("Невідома команда.");
break; }}
```

5. Надсилання SMS-повідомлень (sendSMS()). Цей модуль відповідає за формування та відправку SMS користувачам у разі тривоги, зміни стану напруги або за запитом звіту про стан системи.

```
// Відправка SMS
void sendSMS(const String& message) {
sendATCommand("AT+CMGS=\"+380999623204\""); delay(200);
sim800.println(message); sim800.write(26); delay(2000); }
```

6. Керування GSM-модулем (sendATCommand()). Цей модуль забезпечує низькорівневу комунікацію із GSM-модулем SIM800L шляхом відправки AT-команд і отримання відповідей.

```
// Відправка AT-команди з логами
void sendATCommand(const char* command) {
sim800.println(command); delay(1000);
while (sim800.available()) { char c = sim800.read(); Serial.print(c); }
Serial.println(); }
```

7. Перевірка дозволених номерів (isNumberAllowed()). Цей модуль відповідає за авторизацію вхідних дзвінків шляхом порівняння номера абонента зі списком попередньо визначених дозволених номерів.

```
// Перевірка, чи номер дозволений
bool isNumberAllowed(const String& number) {
for (int i = 0; i < 2; i++) { if (number.indexOf(allowedNumbers[i]) != -1) {
return true; }} return false;}
```

8. Визначення номера абонента (extractNumber()). Цей модуль вилучає номер телефону з текстової відповіді GSM-модуля на вхідний дзвінок.

```
// Витягнення номера із вхідного дзвінка
String extractNumber(const String& response) {
int startIndex = response.indexOf("\"") + 1;
int endIndex = response.indexOf("\"", startIndex);
return response.substring(startIndex, endIndex); }
```

9. Візуалізація стану системи (updateLED()). Цей модуль відповідає за відображення поточного стану системи (охорона двімкнена, тривога) за допомогою світлодіода.

```
// Оновлення стану світлодіода
void updateLED() { if (isAlarmTriggered) { for (int i = 0; i < 10; i++) {
digitalWrite(LED_PIN, HIGH); delay(200); digitalWrite(LED_PIN, LOW);
delay(200); }} else if (isArmed) { digitalWrite(LED_PIN, HIGH); } else {
digitalWrite(LED_PIN, LOW); }}
```

10. Моніторинг датчика руху (checkPIR()). Цей модуль періодично перевіряє стан датчика руху й ініціює режим тривоги в разі виявлення руху, коли система перебуває під охороною.

```
// Перевірка руху
void checkPIR() {
if (isArmed && digitalRead(PIR_PIN) == HIGH) { isAlarmTriggered = true;
activateAlarm();}}
```

11. Контроль напруги живлення (checkVoltage()). Цей модуль періодично перевіряє наявність напруги 220В та надсилає відповідне текстове повідомлення в разі її зникнення або відновлення.

```
// Перевірка напруги
void checkVoltage() {
bool currentVoltageState = digitalRead(VOLTAGE_PIN) == LOW;
if (currentVoltageState != voltagePresent) {
voltagePresent = currentVoltageState;
sendSMS(voltagePresent ? “Voltage: 220 V is available.” : “Voltage: 220 V no.”);}}
```

12. Активація режиму тривоги (activateAlarm()). Цей модуль виконує дії в разі тривоги, як-от надсилання SMS і запуск процедури обдзвонювання дозволених номерів.

```
// Активація тривоги
void activateAlarm() {
sendSMS(“The alarm is triggered!”);
callPhone(allowedNumbers[currentNumberIndex]); lastCallTime = millis();
updateLED(); }
```

13. Здійснення голосових викликів (callPhone()). Цей модуль відповідає за ініціювання голосових викликів на дозволені номери в разі тривоги, з можливістю повторних спроб.

```
// Виклик номера в разі тривоги
void callPhone(const char* phoneNumber) { sendATCommand((“ATD” + String(phoneNumber) + “;”).c_str()); delay(20000); // Очікування підняття трубки
// Якщо трубку не підняли, спробувати наступний номер або повторити через 10 хвилин
if (millis() – lastCallTime > 600000) { // 10 хвилин
currentNumberIndex = (currentNumberIndex + 1)% 2; // Переходимо до наступного номера
lastCallTime = millis();
if (currentNumberIndex == 0) {// Якщо всі номери обдзвонені, повторити через 10 хвилин
callPhone(allowedNumbers[currentNumberIndex]); }}}
```

14. Формування звіту про стан системи (sendSystemReport()). Цей модуль збирає інформацію про температуру, вологість та стан напруги живлення та надсилає її користувачеві SMS за запитом.

```
// Відправка звіту
void sendSystemReport() {
float temperature = dht.readTemperature(); float humidity = dht.readHumidity();
String voltage = voltagePresent ? “220 V is available” : “There is no 220 V”;
String report = “Temperature: ” + String(temperature, 1) + “C, Humidity: ” + String(humidity, 1) +” %, High-voltage: ” + voltage; sendSMS(report); }
```

Ці етапи відображають основні функціональні блоки та логіку роботи розробленої GSM-сигналізації на базі мікроконтролера. Кожен етап відповідає за визначену частину функціональності системи, починаючи від ініціалізації обладнання і закінчуючи обробкою тривожних подій і взаємодією з користувачем.

Такий підхід дозволяє створити надійну й адаптивну систему безпеки, що реагує на різні сценарії та забезпечує безперебійне функціонування.

Результати. Розглянемо алгоритм і логіку управління охороною тестового проєкту, де система може функціонувати у двох режимах: «Охорона активована» або «Охорона деактивована», причому зміна між цими режимами відбувається шляхом уведення DTMF-команд – натискання ‘2’ активує охорону, а ‘3’ її вимикає.

Коли система перебуває в режимі «Охорона активована», а пасивний інфрачервоний датчик (PIR) виявляє рух, спрацьовує сигнал тривоги. У цьому стані система автоматично відправляє текстове повідомлення, здійснює телефонний виклик та активує світлову індикацію (блимання LED, сирена). Для скасування тривоги та повернення об’єкта під охорону використовується DTMF-команда ‘1’.

Усі вхідні дзвінки проходять перевірку на відповідність списку авторизованих номерів (максимум чотири).

У процесі виконання проєкту за допомогою Android-застосунку T2S: Text to Voice/Read Aloud було згенеровано три аудіофайли – 0001.mp3, 0002.mp3 і 0003.mp3. Далі ми детально розглянемо, як саме структуровано інформацію в кожному із цих файлів.

Файл 0001.mp3. Охорону вимкнено. Увімкнути охорону – натисніть 2. Сформувати системний звіт – натисніть 6. Прослухати ще раз – натисніть 0.

Файл 0002.mp3. Охорона ввімкнена. Вимкнути охорону – натисніть 3. Сформувати системний звіт – натисніть 6. Прослухати ще раз – натисніть 0.

Файл 0003.mp3. Тривога в зоні 1. Тривога ввімкнена. Вимкнути тривогу – натисніть 1. Охорона ввімкнена. Вимкнути охорону – натисніть 3. Сформувати системний звіт – натисніть 6. Прослухати інформацію ще раз – натисніть 0.

Щоб файли стали доступними для системи, їх треба перенести на SD-карту за допомогою пристрою для читання карт пам’яті. Для безперебійного відтворення модулем DFPlayer Mini файли повинні мати точні імена (0001.mp3, 0002.mp3, 0003.mp3) і розташовуватися в головному каталозі SD-карти.

Модуль DFPlayer Mini забезпечує відтворення потрібних аудіозаписів із SD-карти у відповідь на команди системи. Наприклад, залежно від поточного стану сигналізації (активована, деактиво-

вана, тривога) або отриманих DTMF-сигналів, будуть програватися відповідні звукові файли.

Завдяки технології DTMF користувач має можливість:

– дізнатися поточний стан системи (натиснувши ‘0’);

– вимкнути режим тривоги (натиснувши ‘1’);

– перемикати режими охорони (активувати за допомогою ‘2’, деактивувати – ‘3’);

– запросити звіт про поточний стан системи (натиснувши ‘6’).

Система здійснює постійний нагляд за наявністю електроживлення 220 В. Також відбувається моніторинг температури та рівня вологості повітря за допомогою датчика DHT11.

Продовжимо опис функціональності тестового проєкту, далі буде представлено детальний алгоритм роботи системи охоронної сигналізації, а також відповідна блок-схема (рис. 2), що візуалізує логіку її функціонування.

1. Початок.

2. Перевірка вхідного дзвінка:

– якщо є вхідний дзвінок, перейти до кроку 3;

– якщо немає вхідного дзвінка, повернутися до кроку 2 (очікування дзвінка).

3. Перевірка номера:

– якщо номер вхідного дзвінка є у списку дозволених, перейти до кроку 4;

– якщо номер відсутній у списку дозволених, відхилити дзвінок і повернутися до кроку 2.

4. Перевірка напруги 220 В:

– якщо напруга 220 В відсутня, перейти до кроку 5;

– якщо напруга 220 В присутня, перейти до кроку 8.

5. Перехід на автономне живлення.

6. Постійна перевірка напруги 220 В:

– якщо напруга 220 В з’явилася, перейти до кроку 7;

– якщо напруга 220 В відсутня, повернутися до кроку 6.

7. Перехід на живлення 220 В.

8. Очікування DTMF-команди:

– якщо отримана команда ‘0’, повідомити поточний стан системи та повернутися до кроку 8;

– якщо отримана команда ‘2’, активувати режим охорони та перейти до кроку 9;

– якщо отримана команда ‘3’, деактивувати режим охорони та повернутися до кроку 8;

– Якщо отримана команда ‘6’, надати розширений звіт про стан системи та повернутися до кроку 8.

9. Контроль датчика руху:

– якщо зафіксовано рух, перейти до кроку 10;

– якщо руху немає, залишатися на кроці 9 (продовжувати контроль).

10. Увімкнення сирени та надсилання SMS.

11. Здійснення дзвінка на дозволені номери по черзі.

12. Очікування відповіді на дзвінок:

– якщо на дзвінок відповіли, перейти до кроку 13;

– якщо на дзвінок не відповіли всі дозволені номери, повернутися до кроку 11 (повторний дзвінок).

13. Очікування DTMF-команди після дзвінка:

– якщо отримана команда ‘1’, деактивувати сирену та повернутися до кроку 9 (система залишається в режимі охорони);

– якщо отримана команда ‘3’, деактивувати режим охорони та повернутися до кроку 8;

– якщо отримана інша команда, повернутися до кроку 13 (очікування команди).

14. Контроль живлення 220 В (під час роботи від мережі):

– якщо напруга 220 В зникла, перейти до кроку 5 (перехід на автономне живлення).

15. Постійна перевірка напруги 220 В (під час роботи від автономного живлення):

– якщо напруга 220 В з’явилася, перейти до кроку 7 (перехід на живлення 220 В);

– після переходу на автономне живлення (крок 5) система також переходить до кроку 8 (очікування DTMF-команд).

Розглянутий алгоритм детально описує логіку роботи розробленої GSM-сигналізації, зокрема: керування режимами охорони, обробку тривожних подій з оповіщенням, дистанційне керування за допомогою DTMF-команд, голосові підказки та контроль зовнішніх умов живлення, що забезпечує комплексний підхід до безпеки об’єкта.

Висновки. У сучасних умовах питання убезпечення житлових приміщень є надзвичайно актуальним, розвиток технологій автоматизації та розумних будинків сприяє впровадженню новітніх методів контролю та захисту. Упровадження GSM-сигналізації в системи охорони розумного будинку стає одним з найефективніших рішень для досягнення безпеки, оскільки дозволяє здійснювати моніторинг у реальному часі й оперативно сповіщати власників про потенційні загрози.

Використання GSM-зв’язку дозволяє створювати автономні рішення, незалежні від дротових інтернет-з’єднань або локальних мереж, що підвищує мобільність і надійність систем безпеки. Водночас інтеграція GSM з технологіями інтернету речей (IoT) відкриває нові можливості для підвищення ефективності та надійності охоронних систем.

Дослідження показують, що поєднання GSM-сигналізації з розумними датчиками руху й автоматичними дзвінками власникам значно покращує рівень безпеки житлових приміщень, що робить ці системи більш ефективними в боротьбі із крадіжками й іншими загрозами.

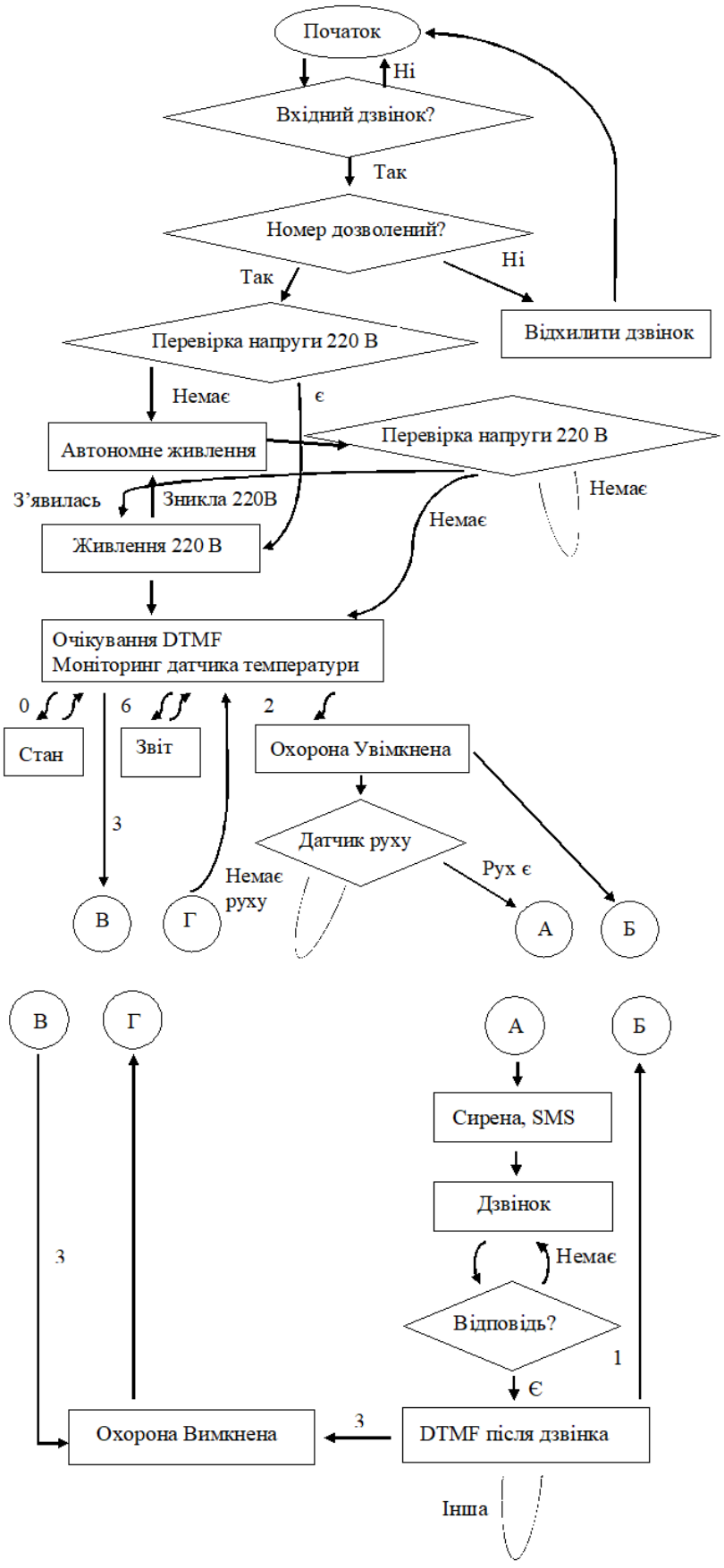


Рис. 2. Блок-схема GSM-сигналізації

ЛІТЕРАТУРА

1. Manjushree N., Ashish K. D. GSM and Arduino based Smart Home Safety and Security System. *Recent Trends in Information Technology and its Application*. 2023. Vol. 6. № 1. P. 1–6. DOI: 10.5281/zenodo.7610756
2. Rana G. M. Sultan M., Khan Abdullah Al Mamun, Hoque Mohammad Nazmul, Mitul Abu Farzan. Design and implementation of a GSM based remote home security and appliance control system. *2013 International Conference on Advances in Electrical Engineering (ICAEE)*, Dkaka, Bangladesh, 19–21 December 2013. 2013. DOI: 10.1109/icaee.2013.6750350
3. Oyekola P., Oyewo T., Oyekola A., Mohamed A. Arduino Based Smart Home Security System. *International Journal of Innovative Technology and Exploring Engineering*. 2019. Vol. 8. № 12. P. 2880–2884. DOI: 10.35940/ijitee.l3052.1081219
4. Arinde V. Adeshina, Idowu L. Multi-sensor Intrusion Detection System. arXiv preprint. 2023. URL: https://arxiv.org/pdf/2406.05137
5. Hadi H. J., Nisa K. U., Harris S. Autonomous and Collaborative Smart Home Security System (ACSHSS). *arXiv preprint*. 2023. URL: https://arxiv.org/abs/2309.02899
6. Mustafijur Rahman, Zadidul Karim A. H. M., Sultanur Nyeem, Faisal Khan, Golam Matin. Microcontroller Based Home Security and Load Controlling Using Gsm Technology. *International Journal of Computer Network and Information Security*. 2015. Vol. 7. № 4. P. 29–36. DOI: 10.5815/ijcnis.2015.04.04
7. Bangali J., Shaligram A. Design and Implementation of Security Systems for Smart Home based on GSM technology. *International Journal of Smart Home*. 2013. Vol. 7. № 6. P. 201–208. DOI: 10.14257/ijsh.2013.7.6.19
8. Shetty N. P. Design of Smart Home Security Surveillance System Using GSM. *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering*. 2016. Vol. 4. № 2. P. 54–59. DOI: 10.17148/ijireeice/ncaee.2016.11
9. Sadeque Reza Khan, Ahmed Al Mansur, Alvir Kabir, Shahid Jaman, Nahian Chowdhury. Design and Implementation of Low Cost Home Security System using GSM Network. *International Journal of Scientific & Engineering*. 2012.
10. Motaz Daadoo, Shadi M.A Atallah, Saed Tarapiah. Development of Low Cost Safety Home Automation System using GSM Network. *European Journal of Social Sciences*. 2016. Vol. 53. № 3. P. 338–353.
11. Abhishek S. Parab, Amol Joglekar. Implementation of Home Security System using GSM module and Microcontroller. *International Journal of Computer Science and Information Technologies*. 2015. Vol. 6 (3). P. 2950–2953.

Дата першого надходження рукопису до видання: 25.08.2025
Дата прийнятого до друку рукопису після рецензування: 22.09.2025
Дата публікації: 31.12.2025

УДК 528.8
DOI https://doi.org/10.26661/2786-6254-2025-2-09

ВИЯВЛЕННЯ ВИТЯГНУТИХ ПОЛІГОНІВ У ГЕОІНФОРМАЦІЙНИХ СИСТЕМАХ НА ОСНОВІ АПРОКСИМАЦІЇ ПРЯМОКУТНИКАМИ

Мильцев О. М.
кандидат фізико-математичних наук,
доцент кафедри програмної інженерії
Запорізький національний університет
вул. Університетська, 66, Запоріжжя, Україна
orcid.org/0009-0009-4382-7730
alexmyltsev@gmail.com

Лимаренко Ю. О.
кандидат технічних наук, доцент,
доцент кафедри програмної інженерії
Запорізький національний університет
вул. Університетська, 66, Запоріжжя, Україна
orcid.org/0000-0002-1643-6939
yuliia.lymarenko@gmail.com

Мухін В. В.
кандидат технічних наук, доцент,
доцент кафедри програмної інженерії
Запорізький національний університет
вул. Університетська, 66, Запоріжжя, Україна
orcid.org/0009-0007-0270-8239
comentssno@gmail.com

Кривохата А. Г.
кандидат фізико-математичних наук,
доцент кафедри програмної інженерії
Запорізький національний університет
вул. Університетська, 66, Запоріжжя, Україна
orcid.org/0000-0001-9664-0659
krivohata@gmail.com

Ключові слова: *аналіз картографічних даних, класифікація полігонів, виявлення витягнутих полігонів, геометричні метрики.*

Робота присвячена проблемі аналізу картографічних даних з метою виявлення довгих і вузьких полігонів, які є результатом невизначеностей, що супроводжують процеси створення та актуалізації картографічної інформації. Наявні натепер методи аналізу полігонів умовно можна поділити на такі категорії, як: застосування геометричних метрик полігонів, декомпозиція форми полігонів за допомогою скелетних ліній, графові та нейромережеві методи, гібридні підходи. Серед геометричних метрик немає єдиної загальноприйнятої, а скелетна декомпозиція та нейромережеві методи можуть потребувати обчислювальних ресурсів, що дещо звужує сферу їх застосування. В основу методу, який запропоновано в даній роботі, покладено геометричну апроксимацію полігона прямокутником. За інваріанти обрано площу та периметр

полігона. Ці характеристики можуть бути розраховані за допомогою відповідних функцій, які є практично в усіх сучасних геопросторових СУБД, що використовуються для зберігання картографічної інформації. Як критерій вузькості пропонується використовувати співвідношення сторін апроксимуючого прямокутника. Корисною для аналізу може бути також безпосередньо ширина прямокутника та значення дискримінанта квадратного рівняння, що виникає в результаті апроксимації. Усі необхідні розрахунки можуть бути виконані за допомогою звичайних SQL-запитів, що є дуже зручним, коли мова йде про великі обсяги даних, необхідність швидко здійснити розрахунки та наявні обмеження на використовувані ресурси. Метод було протестовано на картографічних даних земельних ресурсів округу Шарлотт штату Флоріда (США). Результати проведеного тестування підтверджують ефективність запропонованого підходу. Метод може використовуватись як самостійно, так і з метою попереднього фільтрування великих обсягів картографічної інформації з подальшим застосуванням графових і нейромережевих методів.

RECOGNITION OF ELONGATED POLYGONS IN GEOGRAPHIC INFORMATION SYSTEMS BASED ON RECTANGLE APPROXIMATION

Myltsev O. M.
Candidate of Physical and Mathematical Sciences,
Associate Professor at the Software Engineering Department
Zaporizhzhia National University
Universytetska str., 66, Zaporizhzhia, Ukraine
orcid.org/0009-0009-4382-7730
alexmyltsev@gmail.com

Lymarenko Yu. O.
Candidate of Technical Sciences, Associate Professor,
Associate Professor at the Software Engineering Department
Zaporizhzhia National University
Universytetska str., 66, Zaporizhzhia, Ukraine
orcid.org/0000-0002-1643-6939
yalymarenko@gmail.com

Mukhin V. V.
Candidate of Technical Sciences, Associate Professor,
Associate Professor at the Software Engineering Department
Zaporizhzhia National University
Universytetska str., 66, Zaporizhzhia, Ukraine
orcid.org/0009-0007-0270-8239
comentssno@gmail.com

Kryvokhata A. G.
Candidate of Physical and Mathematical Sciences,
Associate Professor at the Software Engineering Department
Zaporizhzhia National University
Universytetska str., 66, Zaporizhzhia, Ukraine
orcid.org/0000-0001-9664-0659
krivohata@gmail.com

Key words: *cartographic data analysis, polygon classification, detection of elongated polygons, geometric metrics.*

This study addresses the problem of cartographic data analysis aimed at recognizing elongated polygons, which often appear as artifacts resulting from uncertainties inherent in the processes of creating and updating cartographic information. Existing approaches to polygon analysis can be broadly

categorized into the following groups: application of polygon geometric metrics, shape decomposition using skeleton lines, graph-based and neural network methods, as well as hybrid approaches. Among geometric metrics, no universally accepted criterion has been established, while skeleton-based decomposition and neural network methods may require significant computational resources, which somewhat limits their applicability.
The method proposed in this paper is based on the geometric approximation of a polygon by a rectangle. The chosen invariants are the polygon's area and perimeter. These characteristics can be computed using standard functions available in most modern spatial database management systems used for storing cartographic information. As a measure of narrowness, the aspect ratio of the approximating rectangle is proposed. In addition, the rectangle's width and the discriminant of the quadratic equation arising in the approximation process may also serve as useful indicators.
All necessary calculations can be performed using standard SQL queries, which is particularly convenient in cases involving large volumes of data, the need for rapid computation, and constraints on available resources. The proposed method was tested on land resource cartographic data of Charlotte County, Florida, USA. The results confirm the effectiveness of the approach. The method can be applied independently or used as a preliminary filtering step for large-scale cartographic datasets prior to employing graph-based or neural network techniques.

Вступ. У багатьох сферах діяльності, де людям доводиться працювати з великими обсягами просторових даних, представлених у формі полігонів, постає проблема класифікації та/або фільтрування цих даних. До таких сфер належать землевпорядкування, ландшафтна аналітика, картографія, транспортне планування та інші. У кожній конкретній ситуації мета фільтрування і критерій, за яким проводиться класифікація, можуть бути різними. Найчастіше постає необхідність класифікувати полігони за формою, розміром, топологією та динамікою. Звичайно, критерій класифікації визначає і набір доступних методів. Але завдання дещо ускладнюється, якщо потрібен комбінований критерій. Зокрема, у завданнях землевпорядкування зазвичай значення має не тільки форма земельної ділянки, але і її розміри. Наприклад, якщо треба відфільтрувати вузькі полігони й артефакти, які, з великою долею імовірності, не являють собою земельні ділянки з погляду входження їх до земельного реєстру, то ізоморфні за формою полігони можуть бути класифіковані і як такі, що являють собою об'єкт землекористування, і як такі, що мають бути видалені з бази. У даному разі значення будуть мати лінійні розміри ділянки. З іншого боку, урахування тільки розмірів також замало, точніше, є досить проблематичним, оскільки є неочевидним, які саме геометричні метрики враховувати, якщо полігони мають досить складну форму (рис. 1).

Огляд літератури. У роботі розглядається проблема виявлення витягнутих полігонів та інших артефактів, які є кандидатами на видалення з бази даних земельного реєстру. Самі ж полігони зберігаються в базі даних PostGIS. Додаткові обмежувальні фактори: кількість об'єктів у базі, які потрібно проаналізувати, обчислюється десятками мільйонів. Бажано використати підхід, який не потребує додаткових потужностей, є досить швидким і не передбачає використання додаткових мов програмування.

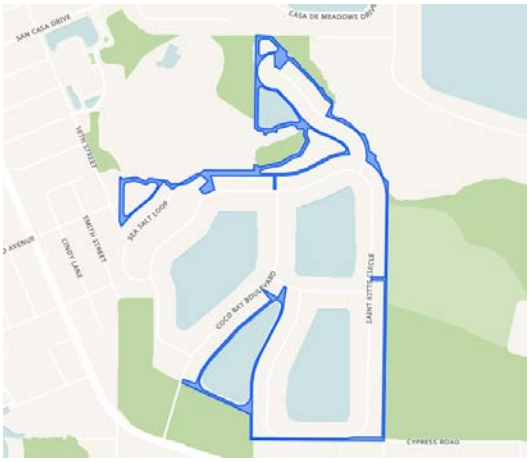


Рис. 1. Приклад вузького та витягнутого полігона зі складною геометрією

Натепер можна виділити такі підходи до вирішення проблеми виявлення вузьких полігонів.

Застосування геометричних метрик полігонів. Найчастіше використовують метрики [1–4], представлені в таблиці 1.

Таблиця 1

Геометричні метрики аналізу полігонів

Назва метрики	Схема розрахунку	Позначення
Ступінь витягнутості (Thinness Ratio, TR)	$TR = \frac{4\grave{A}S}{P^2}$	S – площа полігона, P – периметр полігона.
Співвідношення площин (Area Ratio, ArR)	$ArR = \frac{S}{S_{ec}}$	Sec – площа кола, що обмежує полігон (enclosing circle).
Співвідношення осей еліпса (Aspect Ratio, AsR)	$AsR = \frac{l_{\min}}{l_{\max}}$	lmin, lmax – довжини малої (minor) і великої (major) осей еліпса, що наближено описують форму полігона.

Чим меншими будуть перелічені метрики, тим з більшою імовірністю відповідний полігон може бути кваліфікований як вузький. Безпосередньо в PostGIS можна обчислити тільки *TR*. Використання як критерію вузькості значень *ArR* та *AsR* потребує додаткових розрахунків для обчислення S_{ec}, $l_{\min}$ та $l_{\max}$.

Специфічним для PostGIS є використання радіуса найбільшого вписаного кола (*r = (ST_MaximumInscribedCircle(geom)).radius*) і аналіз впливу функції *ST_Buffer(geom, radius)* зі значенням *radius* < 0 на полігон [5]. Але обидві функції не дозволяють оперувати безрозмірними величинами.

Декомпозиція форми полігонів за допомогою скелетних ліній. У роботі [6] пропонується після побудови скелета полігона на основі тріангуляційної мережі з нерівномірними трикутниками виділити субполігони навколо кінцевих точок гілок скелета та знайти лінії розділення між увігнутими вершинами полігона на основі принципу найближчих сусідів. Оцінювання кандидата на довгу вузьку дугу відбувається на основі індексу виразності форми (відношення довжини скелетної лінії до ширини бази) та індексу форми (у таблиці 1 наведено як ступінь витягнутості). Запропонований у роботі [7] алгоритм генерує центральну лінію витягнутого багатокутника – у прикладі русла річки – через вибір ребер тріангуляційної мережі, що перетинають полігон перпендикулярно до течії, і використання їхніх середніх точок, що зменшує появу небажаних відгалужень, характерних для класичних методів трансформації медіальної осі.

Графові методи та нейронні мережі для класифікації багатокутників. Сучасні дослідження демонструють ефективність застосування глибинного навчання безпосередньо до векторних геометрій, зокрема у формі графових структур. N. Girard, Y. Tarabalka [8] запропонували підхід "end-to-end" навчання полігонів для класифікації даних дистанційного зондування, де форма об'єктів використовується безпосередньо у процесі моделювання. R. Veer та співавтори [9] розробили метод прямого подання багатокутників як послідовностей координат, що подаються на вхід глибинним моделям, усуваючи потребу в ручному проєктуванні геометричних дескрипторів. Найновіший підхід – PolyMP (Z. Huang зі співавторами [10]) – представляє багатокутники у формі графа, вузлами якого є вершини, а ребрами – топологічні зв'язки між ними, та використовує механізм "message passing" для отримання геометрично інваріантних ознак (стійких до повороту, масштабу та зсуву). Такі графові нейронні мережі, зокрема PolygonGNN, здатні поєднувати локальні та глобальні просторові зв'язки (наприклад, на основі графа видимості), тим самим підвищувати точність класифікації складних або витягнутих форм. Попри високу точність і гнучкість, графові нейромережеві методи потребують значних обчислювальних ресурсів, складної підготовки даних і навчання, що може обмежити їх застосування в завданнях із наявними обмеженнями на використовувані ресурси.

Гібридний підхід до аналізу полігонів передбачає поєднання вищеперелічених методів. Наприклад, спочатку можна відкинути артефакти за допомогою простих геометричних метрик. Потім, за необхідності, провести скелетну декомпозицію для більш детального аналізу структури полігонів. Якщо вдосталь даних – додати ML-класифікацію або GNN.

У даній роботі пропонується підхід, яких дозволяє швидко й ефективно проводити попередній аналіз полігонів з метою виявлення витягнутих артефактів. Передбачається, що полігони зберігаються в геопросторовій СУБД, яка має у своєму арсеналі функції для обчислення периметра та площі полігонів.

Апроксимація витягнутих полігонів прямокутниками. В основу методу, що пропонується в даній роботі, покладено апроксимацію полігона прямокутником. Пропонований підхід дещо перегукується з обчисленням Aspect Ratio (табл. 1), який базується на використанні довжин малої і великої осей еліпса, що в певному сенсі наближено описує полігон. Але саме собою обчислення $l_{\min}$ та $l_{\max}$ потребує додаткових розрахунків за межами SQL, отже, додаткових часових ресурсів.

За інваріанти апроксимації обрано периметр і площу прямокутника. Сучасні геопросторові СУБД зазвичай мають відповідний функціонал для обчислення цих параметрів. Розглянемо для визначеності конкретний полігон, представлений на рисунку 1. Позначимо як P та S периметр і площу цього полігона, w та h – ширина та висота того прямокутника, яким ми намагаємось апроксимувати полігон.

Для визначення ширини та висоти апроксимуючого прямокутника можемо записати систему алгебраїчних рівнянь:

$$\begin{cases} P = 2(w+h), \\ S = w \cdot h. \end{cases}$$

Розв'язуючи цю систему щодо w та h, матимемо:

$$\begin{cases} h = P/2 - w, \\ S = w \cdot h; \end{cases} \Rightarrow \begin{cases} h = P/2 - w, \\ S = w \cdot (P/2 - w). \end{cases}$$

Із другого співвідношення маємо квадратичне рівняння щодо ширини w:

$$w^2 - \frac{P}{2} \cdot w + S = 0,$$

розв'язок якого:

$$w_{1,2} = \frac{P/2 \pm \sqrt{D}}{2}, \text{ де } D = \frac{P^2}{4} - 4S, \qquad (1)$$

дає обидва лінійні розміри прямокутника: $w = \min(w_1, w_2)$, $h = \max(w_1, w_2)$.

Отримані значення дозволяють використовувати критерієм вузькості полігона співвідношення сторін апроксимуючого прямокутника (Rectangular Aspect Ratio):

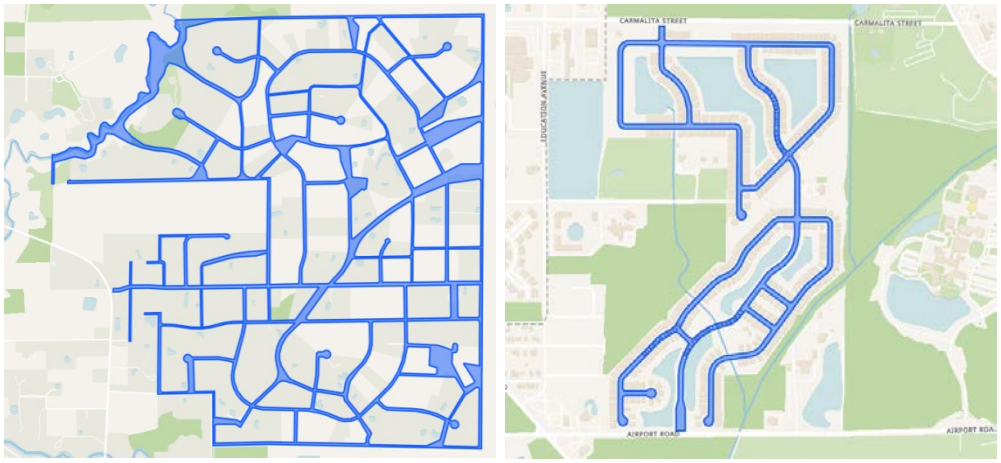
$$RAR = \frac{w}{h}.$$

Зокрема, для полігона з рисунку 1 будемо мати $w = 9.13\,(м)$, $h = 5721.96\,(м)$, $RAR = 0.0016$.

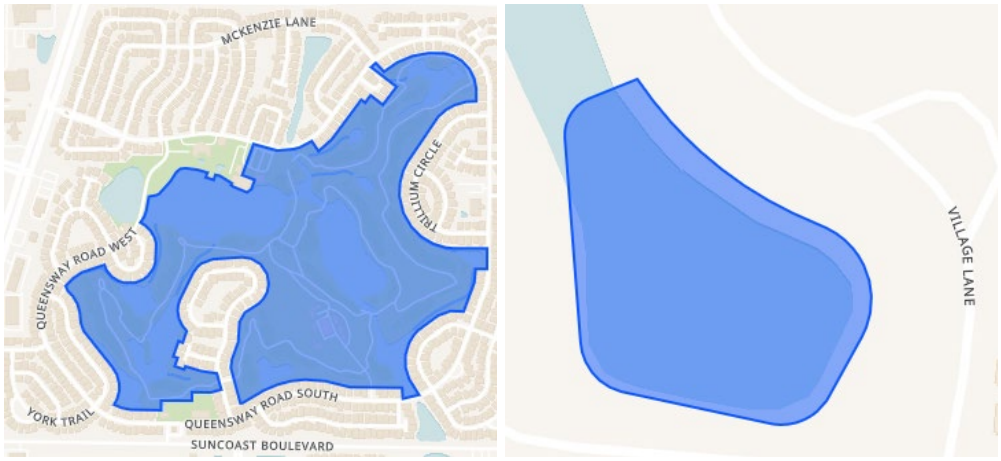
У даному разі ми маємо не тільки безрозмірну величину як критерій. Важливу інформацію надає також розрахована ширина апроксимуючого прямокутника, оскільки прямокутники з однаковими значеннями $RAR = 0.03$, але шириною, наприклад, 0,5 і 10 м мають різні «шанси» на вилучення. Тому, маючи на руках обидві величини, безрозмірну RAR і розмірну w, можна або зосередитись на одній із цих величин або сформулювати комбінований критерій, який буде враховувати обидва розраховані значення.

Необхідно зауважити, що в деяких випадках значення дискримінанта в (1) стає від'ємним, зокрема, це буде мати місце для кола або для еліпса з дуже близькими за значенням величинами півосей. За таких параметрів відповідний полігон не може бути віднесений до категорії вузьких, тому подібні випадки не обмежують сферу застосування запропонованого критерію. Саме значення дискримінанта, навпаки, може також бути використане як додатковий критерій у фільтрації.

Передбачені методом розрахунки можуть бути виконані за допомогою звичайного SQL запиту, подібного до такого:



a) $D = 3114094540.45\,(м^2)$, $w/h = 0.0005$, $w = 28.61\,(м)$, $TR = 0.0016$ b) $D = 52670196.77\,(м^2)$, $w/h = 0.0025$, $w = 18.03\,(м)$, $TR = 0.0077$



a) $D = 5087817.77\,(м^2)$, $w/h = 0.0791$, $w = 193.77\,(м)$, $TR = 0.21$ b) $D = -1745.33\,(м^2)$, $TR = 0.81$

Рис. 2. Приклади розрахованих геометричних метрик полігонів

```
SELECT
 D,
  CASE WHEN D < 0 THEN -1 ELSE (P/2 - sqrt(D)) / 2 END AS w,
  CASE WHEN D < 0 THEN -1 ELSE (P/2 + sqrt(D)) / 2 END AS h,
   CASE WHEN D < 0 THEN -1 ELSE (P/2 - sqrt(D)) / (P/2 + sqrt(D)) END AS "w/h"
 FROM (
  SELECT
  ST_Perimeter(parcel) AS P,
  ST_Area(parcel) AS S,
   (ST_Perimeter(parcel)^2 / 4.0) - 4.0 * ST_Area(parcel) AS D
  FROM parcels
 );
```

Аналіз отриманих результатів. Запропонований у роботі критерій був застосований до аналізу земельних ділянок, що розташовані в окрузі Шарлотт штату Флоріда (США), з метою виявлення довгих і вузьких ділянок. Вибірка, що підлягала аналізу, містила більше 200 тисяч полігонів. Для кожної з ділянок були розраховані значення дискримінанта, ширини та висоти апроксимуючого прямокутника та їх співвідношення. На рисунку 2 представлено чотири різні полігони разом з розрахованими геометричними метриками. Також наведено значення ступеня витягнутості.

Для задачі виявлення досить витягнутих полігонів ділянки з від'ємним дискримінантом можуть бути відразу виключені з розгляду. Подальший аналіз може бути проведений як на основі співвідношення *w/h*, так і на основі ширини *w*. Як свідчать наведені значення, співвідношення сторін апроксимуючого прямокутника корелюється зі значенням ступеня витягнутості, але, якщо брати до уваги значення ширини *w*, можна сформулювати більш змістовний критерій для фільтрування, який буде враховувати й лінійні розміри полігона. Порогові значення кожного із цих параметрів мають визначатись у кожному конкретному випадку окремо. Розглянутий у роботі підхід не передбачає визначення цих порогових значень, а пропонує лише прості та зрозумілі геометричні метрики, результати застосування яких відповідають людському сприйняттю геометричних об'єктів, не потребують додаткових обчислювальних ресурсів, окрім звичайних SQL-запитів, та можуть бути поширені на більш широке коло задач аналізу та класифікації полігонів.

Висновки. Запропонований у роботі підхід дозволяє досить ефективно, за допомогою звичайних SQL-запитів, виявляти довгі та вузькі полігони. Залежно від конкретної ситуації є можливість як оперувати одним безрозмірним критерієм – співвідношенням сторін апроксимуючого прямокутника, так і використовувати його в комбінації з розмірними параметрами – шириною цього прямокутника та дискримінантом (1). Підхід був протестований на картографічних даних земельних ділянок округу Шарлотт штату Флоріда (США). Результати тестування підтвердили ефективність запропонованого підходу, що дозволяє використовувати відповідні геометричні метрики в задачах класифікації полігонів, зокрема, у задачі виявлення витягнутих полігонів.

ЛІТЕРАТУРА

1. Jiao L., Liu Y. Analyzing the shape characteristics of land use classes in remote sensing imagery. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* 2012. Vol. I–7. P. 135–140. DOI: 10.5194/isprsannals-I-7-135-2012
2. Jiao L., Liu Y., Li H. Characterizing land-use classes in remote sensing imagery by shape metrics. *ISPRS Journal of Photogrammetry and Remote Sensing.* 2012. Vol. 72. P. 46–55. DOI: 10.1016/j.isprsjprs.2012.05.012
3. Lamovec P., Kuk K., Černič B., Černe T., Lupše I. A GIS-based approach to land take monitoring and actual land use analysis. *Land.* 2025. Vol. 14. № 7. Art. 1322. DOI: 10.3390/land14071322
4. Wu B., He Z., Xu H., Li Z., Zhang Y., Qiu W., Jiang L. Eliminating methods to false polygon in land use database. *Journal of Henan Agricultural Sciences.* 2008. Vol. 32. № 5. P. 62–65.
5. PostGIS Patterns Documentation. Query shape valid [Electronic resource]. URL: https://dr-jts.github.io/postgis-patterns/query/query-shape-valid.html#find-narrow-polygons (дата звернення: 19.08.2025).
6. Zhang H., Chen X., Yi Z., Xu C. Perception recognition study on long and narrow arcs of polygons in maps using intelligent skeleton decomposition. *Journal of Internet Technology.* 2024. Vol. 25. № 3. №. 415–424. DOI: 10.53106/160792642024052503007
7. Lewandowicz E., Flisek P. A method for generating the centerline of an elongated polygon on the example of a watercourse. *ISPRS International Journal of Geo-Information.* 2020. Vol. 9. № 5. Art. 304. DOI: 10.3390/ijgi9050304
8. Girard N., Tarabalka Y. End-to-end learning of polygons for remote sensing image classification. IGARSS 2018 – 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 2018. P. 2083–2086. DOI: 10.1109/IGARSS.2018.8518116
9. Veer R., van't Bloem P., Folmer E. Deep learning for classification tasks on geospatial vector polygons. *ArXiv:1806.03857* [Cs, Stat]. 2019. URL: http://arxiv.org/abs/1806.03857

10. Huang Z., Khoshelham K., Tomko M. Learning geometric invariant features for classification of vector polygons with graph message-passing neural network. *GeoInformatica.* Published online: 16 Jun 2025. DOI: 10.1007/s10707-025-00554-y

REFERENCES

1. Jiao, L., & Liu, Y. (2012). Analyzing the shape characteristics of land use classes in remote sensing imagery. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, I–7, 135–140. https://doi.org/10.5194/isprsannals-I-7-135-2012
2. Jiao, L., Liu, Y., & Li, H. (2012). Characterizing land-use classes in remote sensing imagery by shape metrics. ISPRS Journal of Photogrammetry and Remote Sensing, 72, 46–55. https://doi.org/10.1016/j.isprsjprs.2012.05.012
3. Lamovec, P., Kuk, K., Černič, B., Černe, T., & Lupše, I. (2025). A GIS-based approach to land take monitoring and actual land use analysis. Land, 14 (7), 1322. https://doi.org/10.3390/land14071322
4. Wu, B., He, Z., Xu, H., Li, Z., Zhang, Y., Qiu, W., & Jiang, L. (2008). Eliminating methods to false polygon in land use database. Journal of Henan Agricultural Sciences, 32 (5), 62–65.
5. PostGIS project documentation. (n.d.). Query shape valid. In PostGIS patterns. Retrieved August 19, 2025, from https://dr-jts.github.io/postgis-patterns/query/query-shape-valid.html#find-narrow-polygons
6. Zhang, H., Chen, X., Yi, Z., & Xu, C. (2024). Perception recognition study on long and narrow arcs of polygons in maps using intelligent skeleton decomposition. Journal of Internet Technology, 25 (3), 415–424. https://doi.org/10.53106/160792642024052503007
7. Lewandowicz, E., & Flisek, P. (2020). A method for generating the centerline of an elongated polygon on the example of a watercourse. ISPRS International Journal of Geo-Information, 9 (5), 304. https://doi.org/10.3390/ijgi9050304
8. Girard, N., & Tarabalka, Y. (2018). End-to-end learning of polygons for remote sensing image classification. In 2018 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (pp. 2083–2086). IEEE. https://doi.org/10.1109/IGARSS.2018.8518116
9. Veer, R., van't Bloem, P., & Folmer, E. (2019). Deep learning for classification tasks on geospatial vector polygons. arXiv. http://arxiv.org/abs/1806.03857
10. Huang, Z., Khoshelham, K., & Tomko, M. (2025, June 16). Learning geometric invariant features for classification of vector polygons with graph message-passing neural network. GeoInformatica. Advance online publication. https://doi.org/10.1007/s10707-025-00554-y

Дата першого надходження рукопису до видання: 26.08.2025
Дата прийнятого до друку рукопису після рецензування: 23.09.2025
Дата публікації: 31.12.2025

УДК 004.891:336.71:004.94
DOI https://doi.org/10.26661/2786-6254-2025-2-10

АНАЛІЗ ТА КЛАСИФІКАЦІЯ ПАТЕРНІВ ШАХРАЙСТВА ІЗ КРЕДИТНИМИ КАРТКАМИ З ВИКОРИСТАННЯМ МАШИННОГО НАВЧАННЯ

Пархоменко О. Ю.
кандидат фізико-математичних наук, доцент,
доцент кафедри економічної кібернетики, комп'ютерних наук
та інформаційних технологій
Миколаївський національний аграрний університет
вул. Георгія Гонгадзе, 9, Миколаїв, Україна
orcid.org/0000-0002-7940-7414
parkhomenko@mnau.edu.ua

Тищенко С. І.
кандидат педагогічних наук, доцент,
завідувач кафедри економічної кібернетики, комп'ютерних наук
та інформаційних технологій
Миколаївський національний аграрний університет
вул. Георгія Гонгадзе, 9, Миколаїв, Україна
orcid.org/0000-0001-7881-8740
tyschenko@mnau.edu.ua

Жебко О. О.
викладач кафедри економічної кібернетики, комп'ютерних наук
та інформаційних технологій
Миколаївський національний аграрний університет
вул. Георгія Гонгадзе, 9, Миколаїв, Україна
orcid.org/0009-0009-1604-5952
al.zhebko@gmail.com

Ємельянов С. І.
доктор філософії (фізика та астрономія),
старший викладач кафедри економічної кібернетики, комп'ютерних наук
та інформаційних технологій
Миколаївський національний аграрний університет
вул. Георгія Гонгадзе, 9, Миколаїв, Україна
orcid.org/0009-0005-9106-5209
sviatoslavem@mnau.edu.ua

Богатєнкова О. Є.
викладач кафедри економічної кібернетики, комп'ютерних наук
та інформаційних технологій
Миколаївський національний аграрний університет
вул. Георгія Гонгадзе, 9, Миколаїв, Україна
orcid.org/0009-0003-0214-0680
oleksandra.bohatienkova@mnau.edu.ua

Ключові слова:
кіберінциденти, машинне навчання, шахрайство із кредитними картками, XGBoost, кластеризація, кібербезпека, фінансові транзакції.

Актуальність дослідження зумовлена стрімким зростанням кіберзлочинів у фінансовому секторі, зокрема шахрайств із кредитними картками, що завдають значних фінансових втрат і підривають довіру клієнтів до банківських систем. Метою роботи є аналіз і класифікація патернів шахрайських транзакцій з використанням машинного навчання для підвищення ефективності їх виявлення та профілювання. Дослідження виконане на публічному датасеті Credit Card Fraud Detection, що містить 284 807 транзакцій із 30 ознаками (28 анонімізованих PCA-ознак, Time, Amount та Class), з дисбалансом класів (492 шахрайські операції).

Висунуто гіпотези щодо переваги оптимізованого XGBoost, ефективності кластеризації K-Means та підвищення точності завдяки циклічному кодуванню часу та порівнянню методів балансування. Для досягнення мети застосовано комплексний підхід: попередню обробку даних (масштабування StandardScaler для Amount, Sine-Cosine encoding для Time), балансування класів (SMOTE, ADASYN, RandomUnderSampler), класифікацію з використанням Logistic Regression, Random Forest, XGBoost і ансамблевої моделі Stacking (з метакласифікатором Logistic Regression), реалізованих у Python з бібліотеками scikit-learn, XGBoost, imbalanced-learn та shap. Оцінювання моделей проводилося за метриками F1-score, ROC-AUC, PR-AUC, MCC, G-mean із крос-валідацією (StratifiedKFold). Кластеризація K-Means виконана на ознаках V14, V17, Amount, Time_sin, Time_cos, а SHAP-аналіз – для оцінювання внеску ознак.

Результати експериментів виявили, що найвищий F1-score (0,99950) та високу стійкість продемонструвала модель Random Forest із балансуванням SMOTE, тоді як ансамбль Stacking досяг найкращого значення ROC-AUC (0,979). Модель XGBoost показала хороші результати (F1 = 0,99869), однак гіпотеза про її безумовну перевагу потребує корекції. Отже, ефективність моделей значно залежить від поєднання алгоритму та методу балансування даних. SHAP-аналіз виявив домінування V14, V4, V12; темпоральний аналіз показав піки активності о 2:00 (12%) та 11:00 (11%); кластеризація виділила 5 груп з різними профілями (від низьких сум з аномальними V14 до високих денних операцій).

Наукова новизна полягає в інтеграції циклічного кодування темпоральних ознак з ансамблевими методами й оцінкою ефективності балансування, що підвищило точність на 5–10% порівняно зі стандартними підходами, а також у сегментації патернів шахрайства для реального часу моніторингу. Модель може бути запроваджена в системи фінансових установ для зменшення втрат від шахрайства, з низькою кількістю хибних спрацювань (FN = 13 для XGBoost).

ANALYSIS AND CLASSIFICATION OF CREDIT CARD FRAUD PATTERNS USING MACHINE LEARNING

Parkhomenko O. Yu.
Candidate of Physical and Mathematical Sciences, Associate Professor,
Associate Professor at the Department of Economic Cybernetics,
Computer Science and Information Technologies
Mykolaiv National Agrarian University
Georgiy Gongadze str., 9, Mykolaiv, Ukraine
orcid.org/0000-0002-7940-7414
parkhomenko@mnau.edu.ua

Tishchenko S. I.
Candidate of Pedagogical Sciences, Associate Professor,
Head of the Department of Economic Cybernetics, Computer Sciences,
and Information Technologies
Mykolaiv National Agrarian University
Georgiy Gongadze str., 9, Mykolaiv, Ukraine
orcid.org/0000-0001-7881-8740
tyschenko@mnau.edu.ua

Zhebko O. O.
Lecturer at the Department of Economic Cybernetics, Computer Science
and Information Technology
Mykolaiv National Agrarian University
Georgiy Gongadze str., 9, Mykolaiv, Ukraine
orcid.org/0009-0009-1604-5952
al.zhebko@gmail.com

Yemelianov S. I.
Doctor of Philosophy (Physics and Astronomy),
Senior Lecturer at the Department of Economic Cybernetics, Computer Science
and Information Technology
Mykolaiv National Agrarian University
Georgiy Gongadze str., 9, Mykolaiv, Ukraine
orcid.org/0009-0005-9106-5209
sviatoslavem@mnau.edu.ua

Bohatienkova O. Ye.
Lecturer at the Department of Economic Cybernetics, Computer Science
and Information Technology
Mykolaiv National Agrarian University
Georgiy Gongadze str., 9, Mykolaiv, Ukraine
orcid.org/0009-0003-0214-0680
oleksandra.bohatienkova@mnau.edu.ua

Key words: *cyber incidents, machine learning, credit card fraud, XGBoost, clustering, cybersecurity, financial transactions.*

The relevance of the research is driven by the rapid growth of cybercrimes in the financial sector, particularly credit card fraud, which causes significant financial losses and undermines customer trust in banking systems. The aim of the work is to analyze and classify fraud patterns using machine learning to improve their detection and profiling efficiency. The study was conducted on the public Credit Card Fraud Detection dataset, containing 284 807 transactions with 30 features (28 anonymized PCA features, Time, Amount, and Class), with class imbalance (492 fraudulent operations).

Hypotheses have been put forward regarding the advantages of optimised XGBoost, the effectiveness of K-Means clustering, and improved accuracy through cyclic time encoding and comparison of balancing methods. A comprehensive approach was used to achieve the goal: preliminary data processing (StandardScaler scaling for Amount, Sine-Cosine encoding for Time), class balancing (SMOTE, ADASYN, RandomUnderSampler), classification using Logistic Regression, Random Forest, XGBoost, and the Stacking ensemble model (with Logistic Regression meta-classifier), implemented in Python with the scikit-learn, XGBoost, imbalanced-learn, and shap libraries. The models were evaluated using F1-score, ROC-AUC, PR-AUC, MCC, and G-mean metrics with cross-validation (StratifiedKFold). K-Means clustering was performed on features V14, V17, Amount, Time_sin, Time_cos, and SHAP analysis was performed to evaluate the contribution of features.

The results of the experiments showed that the Random Forest model with SMOTE balancing demonstrated the highest F1-score (0,99950) and high stability, while the Stacking ensemble achieved the best ROC-AUC value (0,979). The XGBoost model showed good results (F1 = 0,99869), but the hypothesis of its unconditional superiority needs to be corrected. Thus, the effectiveness of models significantly depends on the combination of the algorithm and the data balancing method. SHAP analysis revealed the dominance of V14, V4, V12; temporal analysis showed peaks of activity at 2:00 (12%) and 11:00 (11%); clustering identified 5 groups with different profiles (from low sums with abnormal V14 to high daytime operations).

The scientific novelty lies in integrating cyclic encoding of temporal features with ensemble methods and evaluating balancing efficiency, which improved accuracy by 5–10% compared to standard approaches, as well as in segmenting fraud patterns for real-time monitoring. The model can be integrated into financial institutions' systems to reduce fraud losses, with low false positives (FN = 13 for XGBoost).

Вступ. Сучасний фінансовий сектор стикається зі зростанням загрози кіберзлочинів, зокрема шахрайств із кредитними картками. Згідно з даними FTC, у 2024 р. зареєстровано 6,47 млн звітів про кіберзлочини, з яких 40% стосувалися шахрайств [1]. В Україні сума збитків від незаконних дій та шахрайських операцій з використанням платіжних карток за 2024 р. зросла на 37% – до загальної суми 1,1 млрд гривень, з тенденцією до подальшого зростання у 2025 р. Це пов'язано зі зростанням середньої суми однієї незаконної операції на 39% порівняно із 2023 р. [2]. Такі тенденції не лише завдають фінансових втрат, але й підривають довіру до банківських систем, що зумовлює потребу в ефективних методах виявлення аномалій.

Об'єктом дослідження є патерни кіберінцидентів у фінансовому секторі, зокрема шахрайські транзакції із кредитними картками, представлені в анонімізованому датасеті Credit Card Fraud Detection, з урахуванням темпоральних патернів. Предметом дослідження виступають методи аналізу та класифікації цих патернів за допомогою алгоритмів машинного навчання. Мета роботи полягає в розробленні й оцінюванні моделей для виявлення та кластеризації шахрайських патернів, з акцентом на використання технологій, як-от Python з бібліотеками scikit-learn, XGBoost, imbalanced-learn та shap для обробки даних, балансування класів (методи SMOTE, ADASYN і RandomUnderSampler), класифікації (Logistic Regression, Random Forest, XGBoost з оптимізацією гіперпараметрів через GridSearchCV, ансамблювання Stacking) та кластеризації (K-Means). Це дозволить не лише ідентифікувати аномалії в реальному часі, але й класифікувати їх за типами для подальшого аналізу.

Основні гіпотези дослідження: (1) застосування ансамблевих методів (XGBoost, Random Forest)

та балансування даних дозволить досягти високої ефективності виявлення шахрайства (F1-score >0,9); (2) кластеризація K-Means на комбінації темпоральних і PCA-ознак виявить не менше 5 кластерів із чіткими профілями; (3) інтеграція циклічного кодування темпоральних ознак і порівняння методів балансування (SMOTE, ADASYN, RandomUnderSampler) підвищить точність моделей порівняно зі стандартними підходами.

Дослідження включає попередню обробку даних, балансування класів методом SMOTE та його альтернативами, класифікацію (Logistic Regression, Random Forest, XGBoost з ансамблюванням Stacking) та кластеризацію. Оцінювання гіпотез проводиться за метриками F1-score, ROC-AUC, PR-AUC, MCC, G-Mean з використанням classification_report для точності, крос-валідацією (StratifiedKFold) та візуалізацією результатів через Matplotlib і Seaborn, а також аналізом важливості ознак за допомогою SHAP.

Робота має як теоретичне значення для розвитку ML у кібербезпеці, так і практичне – для впровадження у фінансових установах систем моніторингу в реальному часі, що зменшують втрати від шахрайства. **Огляд літератури.** Сучасні дослідження в галузі кібербезпеки фінансового сектору активно розвивають методи виявлення аномалій на основі машинного навчання та статистичного аналізу. У серії робіт S. Tyshchenko зі співавторами [3–5] запропоновано комплексний підхід до моделювання кіберзагроз. Зокрема, у дослідженні [4] розроблено методологію аналізу часових рядів і виявлення аномалій з використанням K-Means, DBSCAN та Isolation Forest, що підтверджено емпіричним тестуванням на реальних даних. У роботі [5] акцентовано на моделюванні ризиків кібератак із застосуванням Logistic Regression, Random Forest та Gradient Boosting, що доповнює нашу стратегію оцінювання XGBoost. Робота A. Dwarampudi та M.K. Yogi [6] пропонує інтегровану модель з Random Forest, SVM та Gradient Boosting для класифікації та пріоритезації кіберінцидентів, досягаючи високої точності через аналіз важливості ознак, що збігається з нашим підходом до ознак V14 та V17. R. Madunuri зі співавторами [7] демонструють систему із SVM та PCA, досягаючи 95% виявлення аномалій, що перевершує традиційні методи й підкріплює нашу гіпотезу про перевагу ML. P.G. Anjum зі співавторами [8], Z. Bawany та інші [9] досліджують виявлення шахрайства із кредитними картками за допомогою Logistic Regression, Random Forest та градієнтного підсилення, з акцентом на балансування даних, що відображає нашу методологію SMOTE. Ці роботи підтверджують актуальність ML для кібербезпеки, вказують на потребу вдосконалення моделей.

Методи. Дослідження базується на комплексному аналізі датасету Credit Card Fraud Detection [10], що містить 284 807 транзакцій із 30 ознаками, зокрема час (Time), суму (Amount), 28 анонімізованих ознак (V1–V28), отриманих методом PCA, бінарну мітку класу (0 – легальні, 1 – шахрайські). Датасет характеризується значним дисбалансом: лише 492 (0,17%) шахрайські транзакції. Кореляційний аналіз виявив найвищі зв'язки із цільовою змінною для ознак V17 (−0,33) та V14 (−0,30), що обґрунтовує їхню ключову роль у класифікації.

Попередня обробка даних включала масштабування ознаки Amount за допомогою StandardScaler для нормалізації розподілу. Для ознаки Time, яка відображає секунди від першої транзакції, застосовано циклічне кодування Sine-Cosine (sin(2π*Time/max_Time), cos(2π*Time/max_Time)), щоб урахувати періодичність часу доби (наприклад, близькість між 23:59 та 00:01). Для подолання дисбалансу класів використано три методи балансування на тренувальній вибірці (80% даних): SMOTE (Synthetic Minority Oversampling Technique), ADASYN (Adaptive Synthetic Sampling) та RandomUnderSampler. Їхня ефективність порівнювалась за метриками F1-score, ROC-AUC і MCC, щоб визначити оптимальний підхід для задачі.

Класифікація та аналіз ознак проводились чотирма підходами: Logistic Regression (max_iter = 1 000), Random Forest (n_estimators=50), XGBoost (eval_metric='logloss'), ансамблевим методом Stacking, що комбінує Random Forest і XGBoost із метакласифікатором Logistic Regression. Вибір моделей обґрунтовано їхньою ефективністю для бінарної класифікації з імбалансованими даними. Для XGBoost використано оптимізацію гіперпараметрів через GridSearchCV (параметри: learning_rate [0.01, 0.1, 0.3], max_depth [3, 5, 7]). Важливість ознак оцінювалась за допомогою бібліотеки SHAP (SHapley Additive exPlanations), щоб ідентифікувати внесок V14, V17 та темпоральних ознак у класифікацію. Оцінюваня якості проводилось із використанням крос-валідації (StratifiedKFold, n_splits=5), classification_report для точних значень precision, recall, F1-score, а також спеціалізованих метрик: ROC-AUC, PR-AUC, MCC та G-mean.

Для глибшого аналізу патернів шахрайства застосовано кластеризацію K-Means (n_clusters=5) на основі масштабованих ознак Amount, Time (з Sine-Cosine кодуванням), V14 та V17, обмежену шахрайськими транзакціями для фокуса на аномаліях. Кількість кластерів визначено методом ліктя (elbow method), що підтвердив оптимальність 5 груп, які відображають різні профілі шахрайських транзакцій (наприклад, нічні операції з високими сумами). Темпоральний аналіз виконано через агрегування транзакцій за годиною доби, що доз-

волило виявити піки активності. Усі обчислення реалізовані в Python 3 з використанням бібліотек pandas, numpy, scikit-learn, imbalanced-learn, xgboost та shap, що забезпечує відтворюваність результатів.

Результати. Попередній аналіз даних (EDA) підтвердив значний дисбаланс класів у датасеті Credit Card Fraud Detection, де шахрайські транзакції становлять лише 0,17% від загального обсягу даних, що є типовим для завдань виявлення кіберзлочинів у фінансовому секторі. Кореляційний аналіз виявив сильні взаємозв'язки між ознаками та цільовою змінною, зокрема негативні кореляції для V17 (−0,33) та V14 (−0,30), що підкріплює їхню важливість для ідентифікації аномалій.

Як демонструє рис. 1, шахрайські транзакції характеризуються вищими середніми сумами (122,21 проти 88,29 для легальних операцій).

На діаграмі розсіювання V1 vs V2 (рис. 2) спостерігаються чіткі кластери шахрайських операцій у зонах з низькими значеннями цих ознак, що підтверджує ефективність PCA-перетворення для виявлення патернів.

Класифікаційні моделі продемонстрували високу ефективність залежно від методу балансування даних (таблиця 1). Найвищий F1-score (0,99950) досягнуто для Random Forest з SMOTE, що перевершує XGBoost зі SMOTE (F1 = 0,99869) і ансамбль Stacking зі SMOTE (F1 = 0,99946). ADASYN також показав добрі результати, особливо для Random Forest (F1 = 0,99950), тоді як RandomUnderSampler виявився менш ефективним через зменшення вибірки.

Як видно з рисунку 3, ROC-криві підтверджують перевагу ансамблю Stacking із SMOTE (AUC = 0,979), що демонструє його здатність до детекції шахрайства за високої точності.

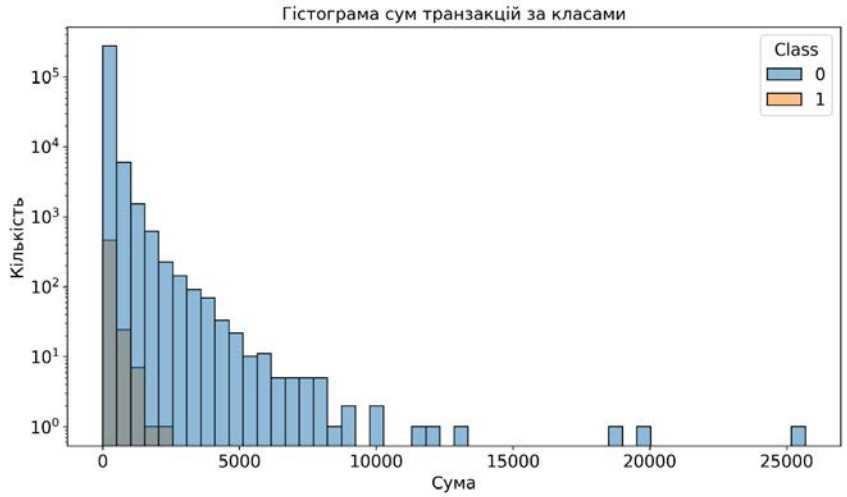


Рис. 1. Гістограма сум транзакцій за класами

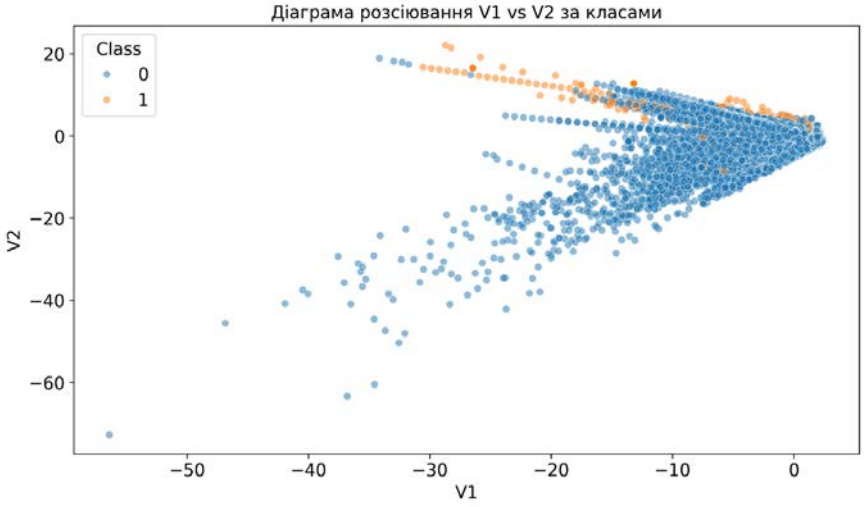


Рис. 2. Діаграма розсіювання V1 vs V2 за класами

Таблиця 1

Ефективність моделей машинного навчання

Балансування	Модель	Точність	Повнота	F1	ROC_AUC	PR_AUC	MCC	G_Mean	Середнє_F1_КросВал	Стд_F1_КросВал	Час_Навчання
SMOTE	Logistic Regression	0,99821	0,97254	0,98454	0,97314	0,77490	0,21804	0,98529	0,95554	0,00049	5,60
	Random Forest	0,99949	0,99951	0,99950	0,97055	0,87124	0,85159	0,99950	0,99991	0,00003	323,05
	XGBoost	0,99899	0,99853	0,99869	0,97361	0,85009	0,68682	0,99876	0,99921	0,00014	7,60
	Ансамбль_Stacking	0,99946	0,99946	0,99946	0,97939	0,86731	0,84238	0,99946	0,99988	0,00005	1628,45
ADASYN	Logistic Regression	0,99816	0,92339	0,95852	0,97416	0,79282	0,13004	0,96005	0,90034	0,02013	7,16
	Random Forest	0,99949	0,99951	0,99950	0,97032	0,86800	0,85045	0,99950	0,91215	0,02568	395,77
	XGBoost	0,99881	0,99788	0,99822	0,97839	0,83465	0,61722	0,99834	0,97249	0,00952	10,73
	Ансамбль_Stacking	0,99852	0,99501	0,99644	0,98146	0,86720	0,46579	0,99677	0,99015	0,00502	1890,26
Random Under Sampler	Logistic Regression	0,99818	0,96136	0,97871	0,97718	0,72274	0,18381	0,97960	0,94031	0,01926	0,03
	Random Forest	0,99822	0,96993	0,98318	0,97464	0,75160	0,21075	0,98397	0,93949	0,01723	0,28
	XGBoost	0,99819	0,95576	0,97578	0,97090	0,66678	0,17352	0,97675	0,94808	0,01722	0,32
	Ансамбль_Stacking	0,99819	0,95671	0,97627	0,97565	0,70208	0,17545	0,97723	0,94804	0,01779	3,28

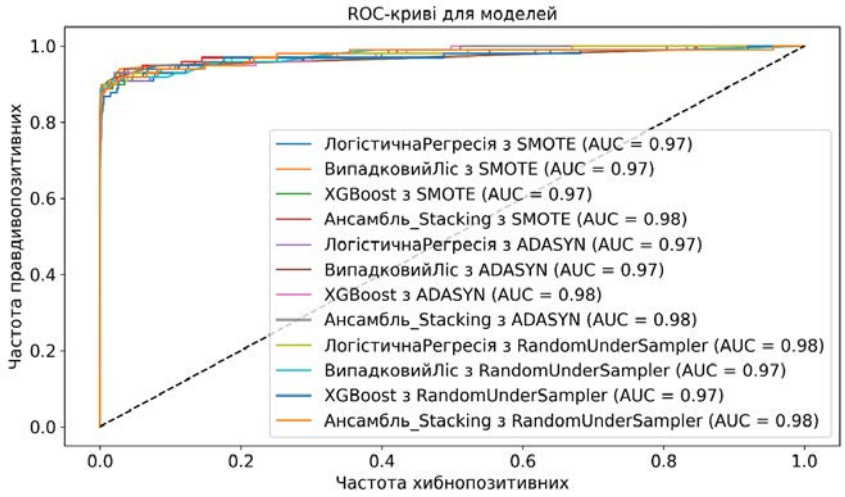


Рис. 3. ROC-криві для моделей

SHAP-аналіз (рис. 4) вказує на домінування ознак V14, V4 і V12 у класифікації, що узгоджується з кореляційним аналізом, тоді як темпоральні ознаки (Time_sin, Time_cos) мають менший, але значущий внесок.

Темпоральний аналіз (рис. 5) виявив піки шахрайських операцій о 2:00 (~12%) і 11:00 (~11%), що вказує на циклічні патерни активності.

Кластеризація K-Means виділила п'ять груп шахрайських транзакцій (рис. 6): кластер 1 (270 транзакцій) – низькі суми з аномальними V14; кластер 3 (159 транзакцій) – середні суми з піками о 10:00; кластер 0 (47 транзакцій) – високі суми вдень; кластер 2 (10 транзакцій) – екстремальні V4; кластер 4 (6 транзакцій) – унікальні комбінації ознак.

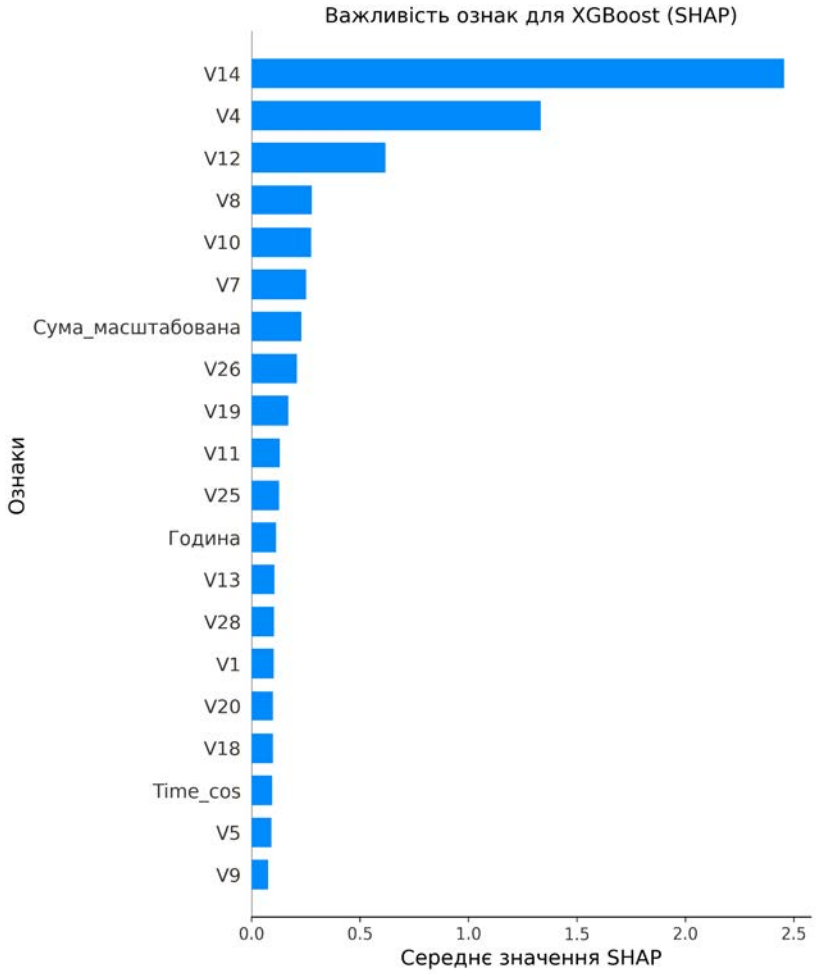


Рис. 4. Важливість ознак для XGBoost (SHAP)

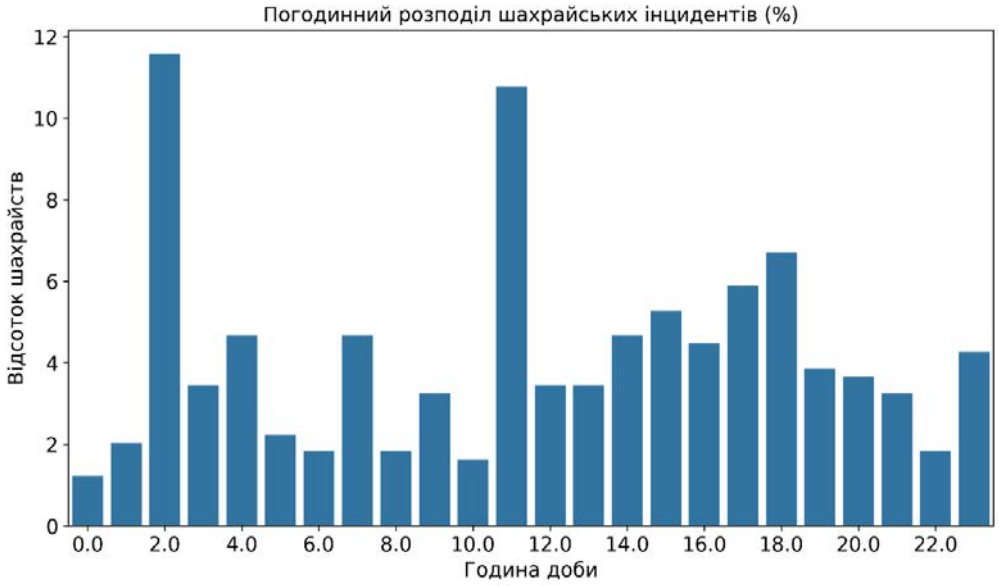


Рис. 5. Погодинний розподіл шахрайських інцидентів (%)

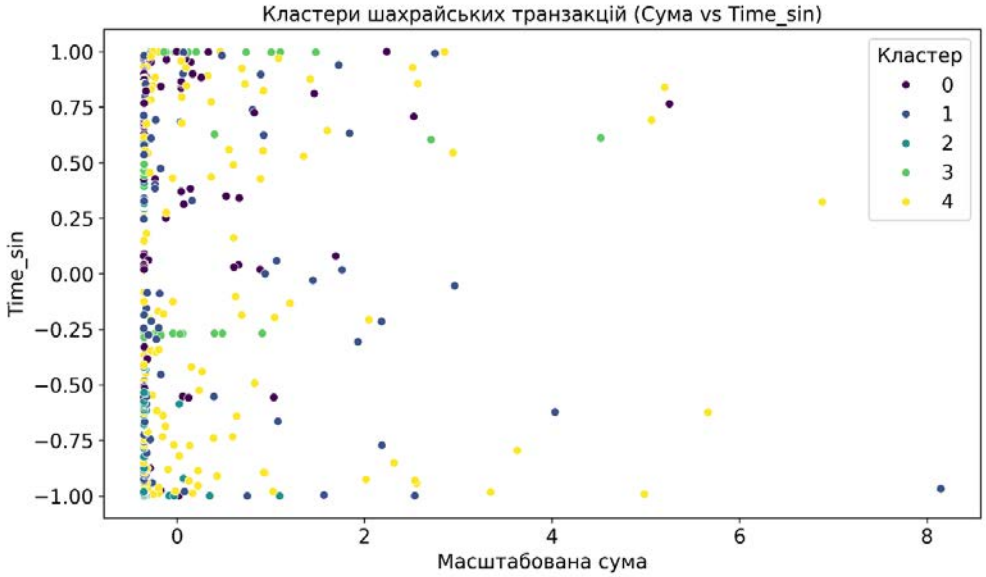


Рис. 6. Кластери шахрайських транзакцій (Сума vs Time_sin)

Висновки. Дослідження присвячене розробленню та оцінюванню моделей машинного навчання для виявлення шахрайських транзакцій у датасеті Credit Card Fraud Detection, який характеризується значним дисбалансом класів (0,17% шахрайських операцій). Основною науковою новизною роботи є інтеграція темпоральних ознак через Sine-Cosine encoding, що дозволило врахувати циклічні патерни активності, а також порівняння ефективності різних методів балансування даних (SMOTE, ADASYN, RandomUnderSampler) у комбінації з ансамблевою моделлю Stacking. Ці підходи розширили традиційні методи детекції аномалій, підвищивши точність класифікації.

Проведене дослідження показало, що запропонований комплексний підхід є ефективним для виявлення шахрайських транзакцій. Головним результатом є те, що модель Random Forest у поєднанні з методом балансування SMOTE показала найвищий F1-score (0,99950), а ансамбль Stacking досяг найкращого ROC-AUC (0,979). Це свідчить про те, що для даної задачі ефективнішим виявився Random Forest порівняно з бустінгом (XGBoost), що корегує початкову гіпотезу. Наукова новизна роботи, що полягає в інтеграції циклічного кодування часу та детальному порівнянні методів балансування, підтвердила свою ефективність. Водночас важливо враховувати обмеження цієї роботи. Результати отримані на одному конкретному датасеті, ефективність методів може варіюватися залежно від характеру даних, ступеня дисбалансу й архітектури моделі.

SHAP-аналіз виявив ключові ознаки (V14, V4, V12), тоді як темпоральний аналіз виокремив пікові години шахрайства (2:00 та 11:00), що узгоджується з гіпотезою про нелінійні залежності. Кластеризація K-Means ідентифікувала п’ять груп шахрайських транзакцій, зокрема й кластери з низькими сумами й аномальними значеннями V14, що дозволяє сегментувати поведінку зловмисників для подальшого аналізу.

Практична цінність роботи полягає в розробленні моделі, здатної працювати в реальному часі з високою точністю виявлення шахрайства за мінімальної кількості хибних спрацювань. Ансамблевий підхід Stacking із SMOTE продемонстрував високу ROC-AUC (0,979), що робить його перспективним для впровадження у фінансових системах. Кластеризація та темпоральний аналіз надають додаткові інструменти для профілювання шахрайських операцій, що може бути використано для адаптивного моніторингу.

Перспективи подальших досліджень включають розширення датасету для перевірки моделі на різних сценаріях шахрайства, інтеграцію глибокого навчання для підвищення точності класифікації та розроблення адаптивних алгоритмів, які враховують зміни в поведінці зловмисників у реальному часі. Отже, запропонований підхід сприяє підвищенню безпеки фінансових транзакцій і має потенціал для практичного застосування в умовах кіберзагроз, що еволюціонують.

ЛІТЕРАТУРА

1. Cybercrime Statistics 2025: Global Trends and Key Data. *BD Emerson: Trusted Cyber Security Services Provider.* URL: https://www.bdemerson.com/article/complete-cybercrime-statistics (дата звернення: 21.08.2025).

2. Кількість випадків шахрайства з картками знизилася, збитки за ними – зросли. *Національний банк України.* URL: https://bank.gov.ua/ua/news/all/kilkist-vipadkiv-shahraystva-z-kartkami-znizilasya-zbitki-za-nimi--zrosli (дата звернення: 21.08.2025).
3. Tyshchenko S., Parkhomenko O. Analysis of the Impact of Digital Threats on Financial Markets Using Methods of Probability Theory and Python. *Modern Economics.* 2024. Vol. 43. № 1. P. 118–124. DOI: 10.31521/modecon.v43(2024)-16
4. Tyshchenko S., Parkhomenko O., Hilko I. Modeling the Impact Of Digital Threats on Financial Markets Using Time Series Analysis and Anomaly Detection Using Python. *Modern Economics*. 2024. Vol. 44. P. 205–212. DOI: 10.31521/modecon.v44(2024)-30
5. Tyshchenko S., Parkhomenko O., Darmosyuk V. Modelling and Analysis of Cyberattack Risks on Financial Institutions Using Mathematical Statistics and Python Methods. *Modern Economics.* 2024. Vol. 48. P. 130–136. DOI: 10.31521/modecon.V48(2024)-16
6. Dwarampudi A., Yogi M. K. A Robust Machine Learning Model for Cyber Incident Classification and Prioritization. *Journal of Trends in Computer Science and Smart Technology.* 2024. Vol. 6. № 1. P. 51–66. DOI: 10.36548/jtcsst.2024.1.004
7. Machine Learning-Based Anomaly Detection for Enhancing Cybersecurity in Financial Institutions / R. Madunuri et al. 2024 Asian Conference on Intelligent Technologies (ACOIT), KOLAR, India, 6–7 September 2024. 2024. P. 1–8. DOI: 10.1109/acoit62457.2024.10941117
8. Applying Machine Learning Techniques to Detect Credit Card Fraud / P. G. Anjum et al. *International Journal of Innovative Research in Computer and Communication Engineering.* 2023. Vol. 11. № 06. P. 8879–8883. DOI: 10.15680/ijircce.2023.1106074
9. Bawany Z., Shanbhag A. D. Using Machine Learning To Detect Credit Card Fraud. *2023 International Conference on Electrical, Computer and Energy Technologies (ICECET),* Cape Town, South Africa, 16–17 November 2023. 2023. DOI: 10.1109/icecet58911.2023.10389421
10. Credit Card Fraud Detection. URL: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud (дата звернення: 20.08.2025).

REFERENCES

1. Cybercrime Statistics 2025: Global Trends and Key Data | BD Emerson. (2025). BD Emerson: Trusted Cyber Security Services Provider. https://www.bdemerson.com/article/complete-cybercrime-statistics
2. Natsionalnyi bank Ukrainy. (2025). Kilkist vypadkiv shakhraistva z kartkamy znyzylasia, zbytky za nymy – zrosly. https://bank.gov.ua/ua/news/all/kilkist-vipadkiv-shahraystva-z-kartkami-znizilasya-zbitki-za-nimi--zrosli
3. Tyshchenko, S., & Parkhomenko, O. (2024). Analysis of the Impact of Digital Threats on Financial Markets Using Methods of Probability Theory and Python. *Modern Economics*, 43 (1), 118–124. https://doi.org/10.31521/modecon.v43(2024)-16
4. Tyshchenko, S., Parkhomenko, O., & Hilko, I. (2024). Modeling the Impact Of Digital Threats on Financial Markets Using Time Series Analysis and Anomaly Detection Using Python. *Modern Economics*, 44, 205–212. https://doi.org/10.31521/modecon.v44(2024)-30
5. Tyshchenko, S., Parkhomenko, O., & Darmosyuk, V. (2024). Modelling and Analysis of Cyberattack Risks on Financial Institutions Using Mathematical Statistics and Python Methods. *Modern Economics*, 48, 130–136. https://doi.org/10.31521/modecon.V48(2024)-16.
6. Dwarampudi, A., & Yogi, M. K. (2024). A Robust Machine Learning Model for Cyber Incident Classification and Prioritization. *Journal of Trends in Computer Science and Smart Technology*, 6 (1), 51–66. https://doi.org/10.36548/jtcsst.2024.1.004
7. Madunuri, R., Ravi, C. S., Chitta, S., Bonam, V. S. M., Vangoor, V. K. R., & Yellepeddi, S. M. (2024). Machine Learning-Based Anomaly Detection for Enhancing Cybersecurity in Financial Institutions. *2024 Asian Conference on Intelligent Technologies (ACOIT)* (p. 1–8). IEEE. https://doi.org/10.1109/acoit62457.2024.10941117
8. Anjum, P. G., Pateriya, P. N., Thakre, P. N., Tiwari, A., & Mahanoor, S. (2023). Applying Machine Learning Techniques to Detect Credit Card Fraud. *International Journal of Innovative Research in Computer and Communication Engineering*, 11 (06), 8879–8883. https://doi.org/10.15680/ijircce.2023.1106074
9. Bawany, Z., & Shanbhag, A. D. (2023). Using Machine Learning To Detect Credit Card Fraud. *2023 International Conference on Electrical, Computer and Energy Technologies (ICECET).* IEEE. https://doi.org/10.1109/icecet58911.2023.10389421
10. Credit Card Fraud Detection. https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud (date of access: 20.08.2025).

Дата першого надходження рукопису до видання: 27.08.2025
Дата прийнятого до друку рукопису після рецензування: 25.09.2025
Дата публікації: 31.12.2025

УДК 004.4: 004.94
DOI https://doi.org/10.26661/2786-6254-2025-2-11

ІНТЕГРАЦІЯ ГЕНЕРАТИВНО-ЗМАГАЛЬНИХ МЕРЕЖ І СЕНТИМЕНТ-АНАЛІЗУ ДЛЯ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ ФОНДОВОГО РИНКУ

Перцев Ю. О.
аспірант кафедри інформаційних систем
Навчально-наукового інституту «Український державний хіміко-технологічний університет»
Український державний університет науки і технологій
просп. Науки, 8, Дніпро, Україна
orcid.org/0009-0005-6244-3768
yuriy.pertsev@gmail.com

Коротка Л. І.
кандидат технічних наук,
доцент кафедри інформаційних систем
Навчально-наукового інституту «Український державний хіміко-технологічний університет»
Український державний університет науки і технологій
просп. Науки, 8, Дніпро, Україна
orcid.org/0000-0003-0780-7571
larysakorotka@gmail.com

Ключові слова: *генеративно-змагальні мережі (GAN), фінансові часові ряди, прогнозування цін акцій, FinBERT, сентимент-аналіз, обробка природної мови (NLP), глибоке навчання, ринковий сентимент, технічні індикатори.*

Фондовий ринок характеризується високою волатильністю та багатофакторністю, що ускладнює завдання прогнозування цінових рухів. Традиційні статистичні методи, зокрема ARIMA та SARIMA, забезпечують точність для стаціонарних часових рядів, проте демонструють обмеження в разі нелінійних залежностей і структурних зламів. Використання глибоких нейронних мереж, як-от LSTM та GRU, дозволило враховувати довгострокові залежності, однак вони залишаються чутливими до нестабільності інформаційного середовища. У цьому контексті перспективним є застосування генеративно-змагальних мереж (GAN), здатних відтворювати складну динаміку фінансових рядів, генерувати синтетичні дані та підвищувати надійність моделей у разі обмеженості вибірки.

Паралельно з розвитком моделей глибокого навчання набувають поширення методи обробки природної мови. Сентимент-аналіз фінансових новин виявився дієвим індикатором настроїв ринку, які часто випереджають цінові зміни. У цій роботі запропоновано інтегрований підхід до прогнозування динаміки цін акцій, що поєднує можливості GAN та NLP. Для експериментів використано п'ятирічні біржові дані компанії "Apple" (AAPL) у поєднанні з новинними матеріалами, обробленими за допомогою FinBERT. Модель TimeF-GAN з архітектурою на основі GRU та CNN поєднує технічні індикатори (MA21, MA7, MACD, смуги Болліджера, обсяги торгів) із числовим відображенням ринкового сентименту.

Результати дослідження підтвердили, що інтеграція новинних даних із класичними ринковими ознаками значно підвищує точність прогнозів. Зокрема, використання сентиментних характеристик дозволило знизити середню абсолютну похибку (MAE) та середньоквадратичну похибку (RMSE), а також покращити коефіцієнт детермінації (R^2) порівняно з моделями, що працюють суто на фінансових рядах. Запропонований підхід демонструє практичну цінність для інвесторів і фінансових аналітиків, оскільки забезпечує більш точне та стійке прогнозування в умовах інформаційної турбулентності.

INTEGRATING GENERATIVE ADVERSARIAL NETWORKS AND SENTIMENT ANALYSIS FOR FORECASTING STOCK MARKET TIME SERIES

Pertsev Yu. O.
Postgraduate Student at the Department of Information Systems
Scientific and Educational Institute "Ukrainian State University of Chemical Technology"
of Ukrainian State University of Science and Technology
Nauky ave., 8, Dnipro, Ukraine
orcid.org/0009-0005-6244-3768
yuriy.pertsev@gmail.com

Korotka L. I.
Candidate of Technical Sciences,
Associate Professor at the Department of Information Systems
Scientific and Educational Institute "Ukrainian State University of Chemical Technology"
of Ukrainian State University of Science and Technology
Nauky ave., 8, Dnipro, Ukraine
orcid.org/0000-0003-0780-7571
larysakorotka@gmail.com

Key words: *Generative Adversarial Networks (GAN), financial time series, stock price forecasting, FinBERT, sentiment analysis, Natural Language Processing (NLP), deep learning, market sentiment, technical indicators.*

The stock market is a highly volatile and multifactorial system, which makes forecasting price dynamics a complex task. Traditional statistical models, such as ARIMA and SARIMA, demonstrate efficiency for stationary time series but remain limited when dealing with nonlinear dependencies and structural breaks. Deep learning approaches, including LSTM and GRU, improved the ability to capture long-term dependencies, yet they are still vulnerable to instability caused by rapidly changing information environments. In this context, Generative Adversarial Networks (GANs) have emerged as a promising tool due to their ability to reproduce complex financial dynamics, generate synthetic datasets, and enhance the robustness of forecasting models, especially in cases of limited data availability.

Parallel to advancements in deep learning, Natural Language Processing (NLP) methods have gained widespread adoption. Sentiment analysis of financial news and social media has proven to be a reliable indicator of market moods, often preceding price fluctuations. This study proposes an integrated forecasting approach that combines GANs with NLP-based sentiment features. For the experiments, five years of Apple Inc. (AAPL) stock market data were analyzed alongside financial news headlines and summaries processed with FinBERT. The proposed TimeF-GAN model, built upon GRU and CNN architectures, integrates technical indicators (MA21, MA7, MACD, Bollinger Bands, trading volume) with market sentiment scores represented in numerical form.

The results demonstrate that the inclusion of sentiment data significantly improves forecasting accuracy. Specifically, the integration of news-driven features reduced mean absolute error (MAE) and root mean square error (RMSE) while increasing the coefficient of determination (R^2), compared with models relying solely on numerical financial series. The study highlights the practical value of combining GAN-based time series modeling with FinBERT-enhanced sentiment analysis, offering investors and financial analysts more accurate, robust, and interpretable predictions under conditions of informational turbulence.

Вступ. Фондовий ринок є складною, динамічною та високоволатильною системою, у якій ціни фінансових активів формуються під впливом широкого спектра економічних, політичних і поведінкових факторів. Традиційні статистичні методи прогнозування, зокрема, моделей Autoregressive Integrated Moving Average (ARIMA) та Seasonal Autoregressive Integrated Moving Average (SARIMA), довели свою ефективність для стаціонарних часових рядів, проте виявили обмежену здатність моделювати нелінійні залежності та структурні злами, властиві фінансовим даним [1]. Моделі глибокого навчання, як-от довгої короткочасної пам'яті (Long Short-Term Memory, LSTM), дозволили враховувати довгострокові залежності та складні патерни ринку, проте й вони залишаються вразливими до проблеми узагальнення в умовах нестабільного інформаційного середовища.

Останніми роками зростає інтерес до генеративно-змагальних мереж (Generative Adversarial Networks (далі – GAN)), які спочатку були розроблені для задач комп'ютерного зору й обробки зображень, але нині активно адаптуються для аналізу часових рядів і фінансових прогнозів. Їхня здатність синтезувати реалістичні дані дозволяє розширювати навчальні вибірки, моделювати ринкові сценарії та підвищувати точність прогнозів навіть у ситуаціях, коли доступні історичні дані є обмеженими.

Паралельно з розвитком моделей глибокого навчання значного поширення набули методи обробки природної мови (Natural Language Processing, NLP). Сентимент-аналіз новин, соціальних мереж і фінансових звітів виявився важливим індикатором настроїв ринку, які часто передують ціновим змінам. Інтеграція NLP та GAN відкриває нові можливості: поєднання кількісних часових рядів з якісними текстовими даними дозволяє моделювати приховані залежності між ринковими факторами й інформаційними потоками.

Отже, актуальність дослідження полягає в необхідності розроблення інтегрованих моделей, здатних одночасно працювати із числовими та текстовими ознаками, формувати синтетичні фінансові ряди, а також враховувати вплив інформаційного середовища на динаміку ринку. Наукова новизна полягає в адаптації генеративно-змагальних мереж до завдань фінансового прогнозування з урахуванням NLP-компонентів, що дозволяє підвищити точність, стійкість і практичну цінність прогнозів.

Мета роботи – створення комплексного підходу до прогнозування вартості акцій із застосуванням генеративно-змагальних мереж, методів обробки природної мови й аналізу сентименту. Запропонована модель інтегрує числові часові ряди з інформаційними сигналами з новинних і соціальних ресурсів, що дозволяє більш точно відображати динаміку ринку. Використання механізмів уваги в архітектурі трансформерів підсилює інтерпретацію результатів, оскільки дає змогу виділяти найбільш значущі ознаки та закономірності. Такий підхід усуває недоліки класичних статистичних моделей, ураховує вплив інвесторських настроїв і рівень волатильності, підвищує стабільність прогнозів. Ефективність розроблених рішень перевіряється на реальних даних фондового ринку шляхом порівняння результатів моделей на базі GAN і трансформерів, що дозволяє оцінити їхню практичну придатність для фінансових аналітиків та інвесторів. Огляд літератури. Перші спроби використання генеративно-змагальних мереж для моделювання часових рядів були спрямовані на збереження як часових залежностей, так і статистичних характеристик даних [1]. Подальший розвиток у сфері фінансових застосувань показав, що GAN здатні генерувати реалістичні симуляції ринкової динаміки та підвищувати якість прогнозування [2]. Систематичний аналіз сучасних досліджень підтвердив широкий спектр можливостей цих моделей, проте також виявив ключові обмеження, зокрема проблеми стабільності навчання та узагальнюваності результатів [3].

Подальші роботи зосередилися на інтеграції GAN з іншими архітектурами глибинного навчання. Запропоновано гібридні підходи, які поєднують GAN із рекурентними моделями, зокрема LSTM [4]. Розроблено нові модифікації, що враховують багатовимірні фактори у фінансових даних [5], а також варіаційні графові згорткові мережі, які забезпечують більш повний аналіз багатофакторних часових рядів [6].

Важливим напрямом у дослідженнях фінансових ринків є інтеграція аналізу текстових даних. Використання спеціалізованих мовних моделей для оцінювання фінансового сентименту дозволило істотно підвищити точність прогнозів [7]. Поєднання часових рядів з ознаками сентименту та генеративними моделями забезпечило кращі результати у прогнозуванні волатильності [8]. Подальші дослідження розширили підхід завдяки інтеграції новинного аналізу та збільшенню даних за допомогою GAN [9]. Порівняльні експерименти підтвердили, що поєднання моделей сентимент-аналізу з GAN перевершує традиційні методи [10].

Вітчизняні дослідження демонструють активний розвиток методів прогнозування фондового ринку із застосуванням нейронних мереж. Проведено аналіз ефективності рекурентних нейронних мереж (Recurrent Neural Networks, RNN) і LSTM у задачах передбачення цінових змін [11], а також досліджено можливості нейромережевого про-

гнозування на локальних фінансових даних [12]. Подальші праці зосередилися на порівнянні традиційних статистичних методів із сучасними нейронними моделями [13], розробленні інноваційних підходів до прогнозування часових рядів [14], а також адаптації генеративно-змагальних мереж до специфіки фінансових ринків [15].

Паралельно із цим з'являються навчальні посібники, які систематизують теоретичні та прикладні аспекти використання штучних нейронних мереж [16–18], що створює методологічну основу для подальших прикладних розробок.

Результати. У сучасних умовах фінансових ринків традиційні дані у вигляді котирувань: ціни відкриття, закриття, максимуму та мінімуму – залишаються базовим джерелом для аналізу динаміки цінних паперів. Проте використання лише цих показників замало для побудови точних і надійних прогнозів. Причина полягає в багатофакторній природі ринку, де на формування вартості активів впливають не тільки внутрішні ринкові процеси, а й зовнішні інформаційні чинники.

Тому в наукових і прикладних дослідженнях усе частіше застосовують комплексний підхід, який поєднує різні типи даних. До них відносять: технічні індикатори, що відображають тренди, рівень волатильності та силу цінових рухів, новинний потік і ринковий сентимент, які визначають інформаційне середовище та психологію учасників ринку.

Фінансові новини та соціально-медійні сигнали мають істотний вплив на ухвалення інвестиційних рішень, адже вони здатні спричинити як короткострокові сплески волатильності, так і довгострокові зміни трендів. Наприклад, публікації щодо фінансових результатів компаній, їхніх стратегічних кроків чи глобальних економічних подій можуть миттєво змінювати настрої інвесторів, отже, і впливати на динаміку цінних паперів.

Для ілюстрації цього аспекту в дослідженні було сформовано щоденну статистику кількості фінансових новин щодо компанії "Apple" (AAPL). На рисунку 1 подано графік, який відображає інтенсивність інформаційного потоку в розрізі кожного торгового дня. Наведена візуалізація дозволяє наочно оцінити, що новинна активність є динамічним процесом, який варіюється на щоденній основі, що створює додатковий контекст для прогнозування.

Отже, використання новинних даних у поєднанні з ринковими котируваннями та технічними індикаторами є доцільним і необхідним кроком для побудови інтегрованих моделей прогнозування, що відповідає загальній меті роботи, а саме: продемонструвати, що багатовимірний простір ознак, збагачений інформаційним фоном, здатний підвищити точність і надійність прогнозів цін цінних паперів.

Генеративно-змагальні мережі останніми роками стали одним із ключових інструментів у сфері моделювання складних даних, зокрема фінансових часових рядів. Як відомо, їхня архітектура ґрунтується на взаємодії двох нейронних мереж – генератора та дискримінатора. Генератор навчається створювати синтетичні приклади, які імітують властивості реальних даних, тоді як дискримінатор намагається відрізнити реальні дані від згенерованих. Така конкурентна динаміка поступово приводить до того, що модель здатна відтворювати складні закономірності та латентні структури, характерні для фінансових ринків.

На відміну від традиційних статистичних методів, GAN-моделі не накладають жорстких припущень щодо розподілів даних і вміють враховувати нелінійні залежності, високу волатильність та структурні злами, притаманні фондовим ринкам. Завдяки цьому вони здатні не лише підвищувати точність прогнозів, а й генерувати додаткові синтетичні ряди в разі обмеженості вибірки. Останній факт особливо актуальний у фінансовій сфері, де нестача якісних даних через велику кількість випадкових коливань ускладнює виділення реальних закономірностей, що є серйозною перешкодою для побудови надійних моделей.

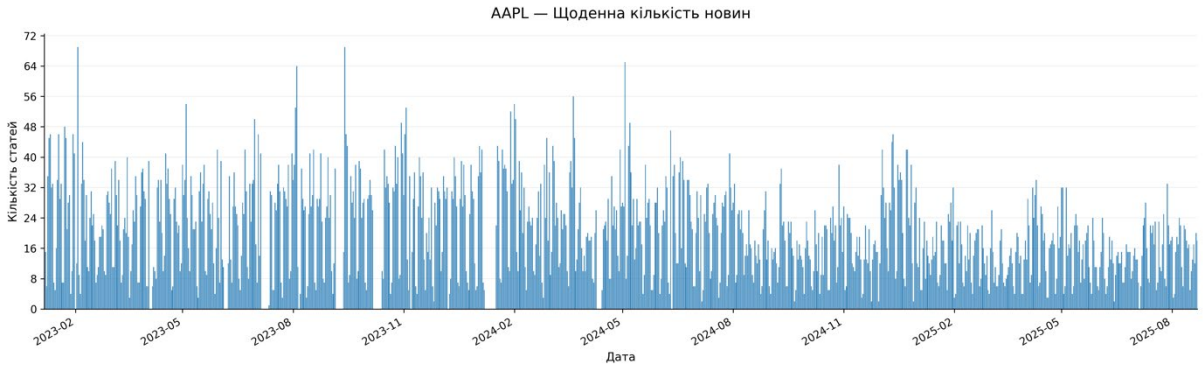


Рис. 1. Щоденна кількість новин щодо компанії AAPL

У рамках даного дослідження GAN використовується як ядро інтегрованої моделі прогнозування цін цінних паперів. Особливістю підходу є те, що вхідні дані включають не лише класичні ринкові та технічні показники, але й текстові фактори, зокрема результати сентимент-аналізу фінансових новин. Така інтеграція дає змогу поєднати кількісні характеристики ринку з інформаційним середовищем, яке часто виступає каталізатором різких коливань цін.

Щоб продемонструвати вагомість цього аспекту, було сформовано часовий розподіл фінансових новин у розрізі доби. На основі зібраних даних побудовано графік, що відображає кількість новин за кожен торговий день у поєднанні з усередненим значенням індексу сентименту. Така візуалізація дозволяє оцінити інтенсивність інформаційного потоку та його емоційне забарвлення, що є критично важливим для розуміння ринкових реакцій.

У дослідженні для прогнозування цін акцій було обрано біржові данні щодо компанії “Apple”. Вибір базувався на її п’ятирічній історії біржових котирувань та наявності достатнього обсягу новинних даних для оцінювання ринкових настроїв.

Процес збору даних складався із двох етапів: збору новинних статей, що стосуються конкретної компанії, та збору історичних цін акцій за відповідний період.

Новинні матеріали було отримано за допомогою Rapid API, який агрегує якісні статті з надійних фінансових джерел, як-от The Wall Street Journal, Financial Times, Motley Fool, Market Watch та інші. Для аналізу використовувалися заголовки, оскільки вони містять ключову інформацію, яка безпосередньо впливає на динаміку ринку. У дослідження включено новини, опубліковані в період з 1 січня 2023 р. по 22 серпня 2025 р.

Отриманий обсяг новинних даних було розподілено на тренувальну та тестову вибірки в розрізі 80% новин для навчання моделей, 20% для тестування.

Історичні біржові дані було зібрано за допомогою Python-пакета yfinance, що забезпечив доступ до інформації про ціну відкриття (Open), ціну закриття (Close), максимальне значення (High), мінімальне значення (Low) та обсяг торгів (Volume). Період збору охоплював часовий проміжок з 1 січня 2023 р. по 22 серпня 2025 р.

Після проведення розвідувального аналізу даних набір було очищено для забезпечення відсутності пропусків, дублювань та аномальних значень, що могло б негативно вплинути на результати прогнозування.

Для формування остаточного датасету, що застосовувався в експериментальній частині, було поєднано історичні біржові дані з Yahoo Finance із щоденними значеннями сентимент-оцінок, синхронізованих за торговими днями у США (рисунок 2).

Під час інтеграції даних було зроблено припущення, що ринковий сентимент визначеного дня впливає на ціну закриття наступного торгового дня. Для днів, у які відсутні новини, сентимент уважався нейтральним, тобто з нульовим значенням індексу.

Запропонована в роботі модель TimeF-GAN побудована за мотивами архітектури TimeGAN (Time-series Generative Adversarial Networks), проте має кілька принципових відмінностей. На відміну від базового TimeGAN, який зосереджений на відтворенні часових залежностей у фінансових рядах, TimeF-GAN інтегрує додатковий блок зовнішніх факторів – ринковий сентимент, згенерований на основі FinBERT. Такий підхід дозволяє поєднати числові ринкові індикатори (ціни, обсяги, технічні показники) з якісними текстовими ознаками, що відображають настрої учасників ринку. Окрім того, в архітектурі генератора використано багатошарові GRU-блоки, оптимізовані для стабільного багатокрокового прогнозування, тоді як дискримінатор реалізовано на основі згорткових шарів (1D-CNN) для кращого виявлення прихованих патернів. Отже, TimeF-GAN є розширенням TimeGAN, орієнтованим не лише на відтворення часової структури, а й на інтеграцію інформаційного середовища, що

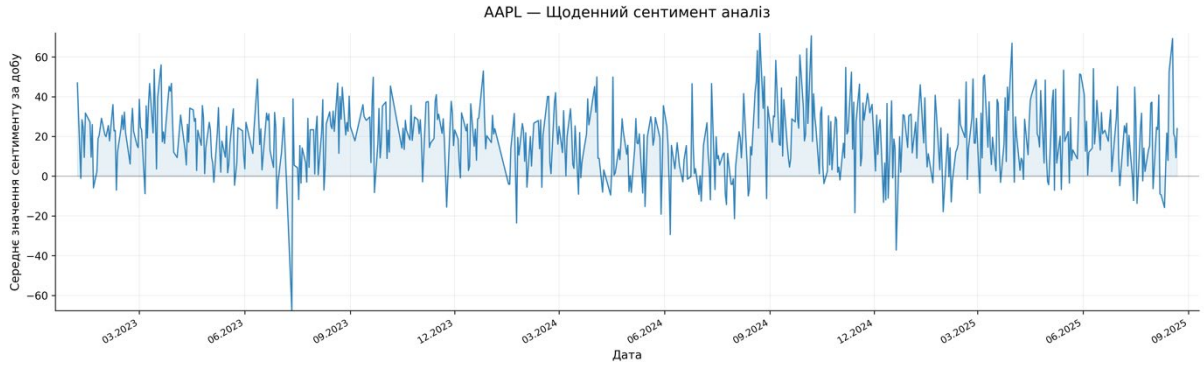


Рис. 2. Щоденний сентимент аналіз для компанії “Apple”

забезпечує підвищену точність і стійкість прогнозування.

Архітектура складається із двох основних компонентів: генератора *G (z)* та дискримінатора *D (x)*. Генератор прагне створювати синтетичні дані, максимально подібні до справжніх, щоби ввести дискримінатор в оману. Дискримінатор, своєю чергою, намагається відрізнити реальні дані від згенерованих та уникнути помилок.

Таким чином, обидві мережі навчаються одночасно у процесі взаємного протистояння. Генератор формує штучні приклади даних на основі випадкового шуму, а дискримінатор обчислює функцію втрат і передає її у зворотному поширенні для коригування ваг обох мереж.

Математично задача визначається як гра з нульовою сумою з наступною цільовою функцією:

$$\min_G \max_D V(D,G) = E_{x\sim P_{data}(x)}[logD(x)] + E_{z\sim p_z(z)}[\log(1 - D(G(z)))], \quad (1)$$

де *D (x)* – імовірність, з якою дискримінатор класифікує зразок *x* як реальний; *G (z)* – синтетичні дані, створені генератором з випадкового латентного вектора *z*; *D (G (z))* – оцінка дискримінатора для синтетичних даних (значення в межах [0, 1], де 1 – висока ймовірність реальності) [2].

У запропонованій моделі GAN як генератор було використано рекурентну мережу GRU (Gated Recurrent Units), завдяки її високій стабільності в роботі із часовими рядами. Датасет містив п'ятирічну історію цін акцій AAPL, а також такі ознаки: Open, Low, High, Close, MA21, MA7, MACD, верхню та нижню межі смуг Боллінджера, Volume та Market Sentiment Score. На рисунку 3 наведено приклад технічних індикаторів, які використовувались в роботі. Ринковий сентимент уводився в нейронну мережу як прихована змінна (ко-варіата) разом з іншими числовими показниками.

Оскільки дослідження охоплює задачу багатокрокового прогнозування, генератор приймає на вхід розмір пакету, кількість кроків на вході та набір ознак, а на виході формує прогноз із заданою кількістю кроків наперед.

Архітектура генератора складається із трьох шарів GRU з кількістю нейронів 1 024, 512 і 256 відповідно, після яких розташовано два щільні (Dense) шари. Кінцевий шар має кількість нейронів, що відповідає розміру вихідної послідовності, яку необхідно спрогнозувати. Такий підхід забезпечує можливість моделювати складні залежності та послідовно прогнозувати значення ряду, з використанням результатів попередніх кроків.

Дискримінатор у запропонованій моделі реалізовано на основі згорткової нейронної мережі, завданням якої є визначення, чи є подані дані реальними чи згенерованими генератором. На вхід дискримінатор отримує як справжні дані, так і синтетичні вибірки.

Його архітектура включає три одномірні згорткові шари (1D-CNN) із 32, 64 та 128 нейронами відповідно, три щільні (Dense) шари з кількістю нейронів 220, 220, 1 на вихідному рівні. Для всіх шарів, окрім вихідного, використано функцію активації Leaky ReLU. Вихідний шар застосовує: Sigmoid у разі класичного GAN, лінійну активацію у разі TimeS-GAN. Sigmoid повертає скаляр у діапазоні [0,1], що інтерпретується як імовірність належності зразка до класу реальних чи синтетичних даних.

Схематичне зображення запропонованої архітектури GAN наведено на рисунку 4.

Для класифікації новинних повідомлень за емоційним забарвленням було проведено сентимент-аналіз кожної статті за допомогою мовної моделі, спеціалізованої на фінансових даних FinBERT. Модель FinBERT є попередньо натренованим трансформером BERT, донавченим для задачі класифікації фінансового сентименту; аналізує текстовий вхід і видає результат у вигляді ймовірнісного значення в діапазоні від 0 до 1, а також відповідну мітку сентименту: позитивний,

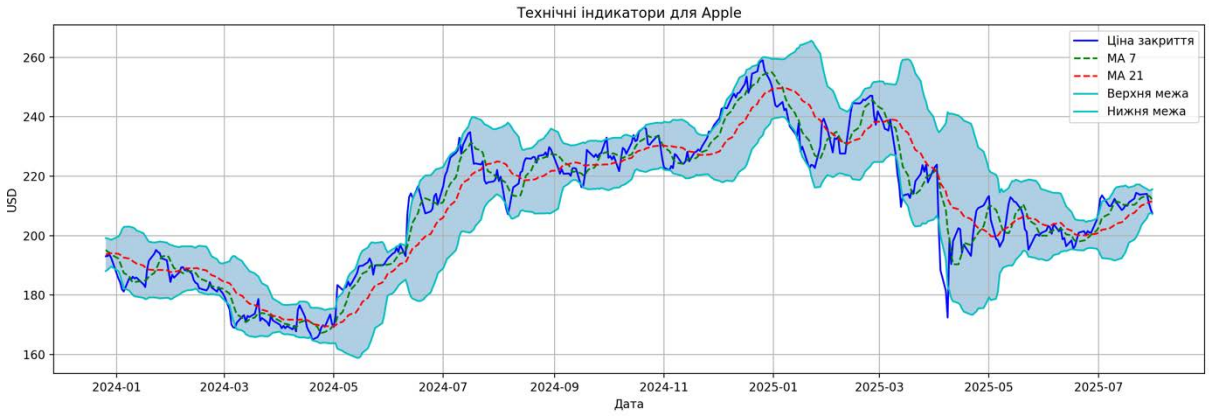


Рис. 3. Технічні індикатори для компанії "Apple"

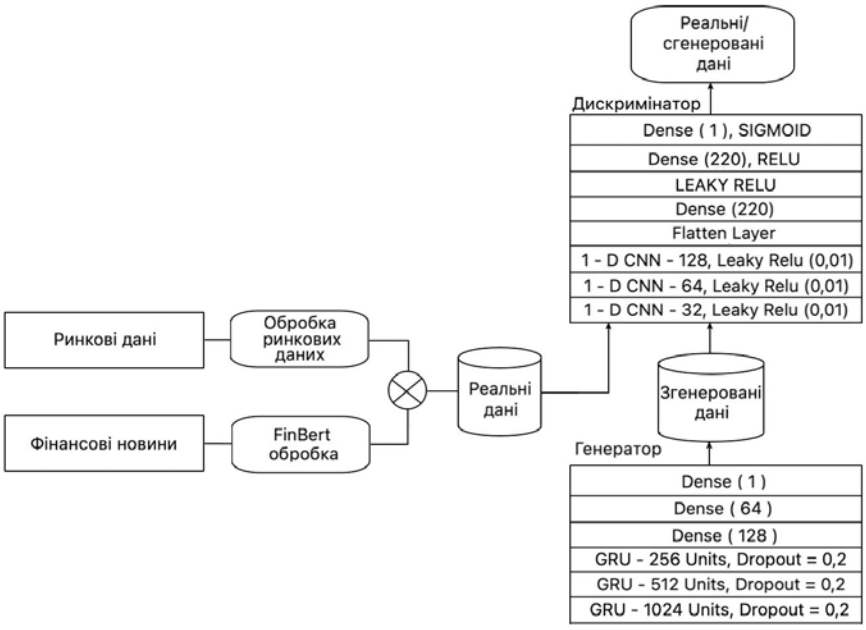


Рис. 4. Схема запропонованої схеми архітектури TimeS-Gan

негативний або нейтральний. Чим вище значення імовірності, тим більш упевненою є модель у своїй класифікації [7].

Оскільки нейронні мережі здатні обробляти лише числові дані, усі текстові результати необхідно було попередньо перетворити на числовий формат. Для цього було використано таке кодування:

- позитивна новина = 100;
- негативна новина = −100;
- нейтральна новина = 0.

Таким чином, на вхід нейронній мережі подавалися два типи даних: біржові котирування (ціни акцій у числовому вигляді) та ринковий сентимент, трансформований у числові значення.

Для обчислення агрегованого сентимент-індексу (Sentiment Score, SSn) за окремий день була визначена відповідна формула (див. нижче).

$$SS_n = \frac{\sum_{i=1}^{N}\left(CS_{pos}\times 100\right)+\sum_{j=1}^{M}\left(CS_{neg}\times -100\right)}{N+M+P} \qquad (2)$$

де *N, M* та *P* позначають відповідно загальну кількість позитивних, негативних і нейтральних новинних статей за окремий день. CSpos (Confidence Score – positive) відображає рівень впевненості моделі у класифікації позитивних статей, тоді як CSneg (Confidence Score – negative) відповідає рівню впевненості щодо негативних статей.

Отримані за допомогою алгоритму значення сентименту трансформуються в числові ознаки, а саме індекси позитивності, нейтральності та негативності, після чого вони інтегруються з історичними біржовими даними (ціни відкриття, закриття, максимум, мінімум, обсяги торгів). Такий комбінований набір ознак подається на вхід моделі прогнозування, що дозволяє враховувати не лише ринкові патерни, а й інформаційний фон, сформований новинними та соціальними повідомленнями.

Для демонстрації впливу ринкового сентименту було сформовано два набори ознак: набір, що включав сентимент-індекс разом із біржовими даними (щоденні ціни відкриття, максимуму, мінімуму та закриття), інший набір, що містив лише біржові дані.

Для оцінювання якості прогнозування використовувалися три метрики: середня абсолютна похибка (MAE), середньоквадратична похибка (RMSE) та коефіцієнт детермінації (R^2). Вони обчислювалися за відомими формулами:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|y_i - \widehat{y_i}\right|, \qquad (3)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{y_i}\right)^2}, \qquad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \widehat{y_i}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}, \qquad (5)$$

де n – загальна кількість вибірок, y_i – фактичні значення, $\widehat{y_i}$ – передбачені значення, $\bar{y}$ – середнє фактичних значень тестової вибірки.

Середня абсолютна похибка обчислюється як середнє значення абсолютних різниць між фактичними та передбаченими значеннями. MAE наочно демонструє, наскільки сильно модель помиля-

ється у прогнозуванні ринкових показників (3). Корінь середньоквадратичної похибки базується на квадратах різниць між реальними та прогнозованими значеннями. Через піднесення до квадрата RMSE сильніше реагує на великі похибки, зберігає водночас здатність оцінювати абсолютні відхилення (4). Коефіцієнт детермінації (R^2) показує, наскільки добре модель відтворює фактичну динаміку – тобто яку частку варіації в залежній змінній пояснюють прогнози моделі (5) [19].

Отже, чим менше значення MAE та RMSE, тим вища точність прогнозу. Значення R^2 перебуває в діапазоні (0; 1), причому, як відомо, чим ближче воно до 1, тим краще модель відображає дані [11; 20].

Для оцінювання стабільності запропонованої моделі та її узагальнювальної здатності було проведено тестування на 11 незалежних часових періодах протягом 2023–2025 рр. (таблиця 1). Горизонт прогнозування становив $h = 5$ днів, що є оптимальним для короткострокового прогнозування.

Результати показують, що TimeF-GAN демонструє стабільно нижчі похибки порівняно з TimeGAN на всіх тестових періодах. Середня MAE становить 4,42 для TimeF-GAN проти 5,14 для TimeGAN. Коефіцієнт варіації MAE для TimeF-GAN на 20,2% нижчий, що свідчить про вищу стабільність моделі з інтеграцією сентимент-аналізу.

На рисунку 5 представлено теплову карту розподілу похибок прогнозування моделей TimeGAN та TimeF-GAN протягом 2023–2025 рр. Інтенсивність кольору відображає величину похибки: зелений колір відповідає низьким значенням (висока точність), жовтий – помірним, а червоний – високим похибкам.

Аналіз теплової карти виявляє декілька важливих закономірностей. По-перше, модель TimeF-GAN демонструє переважно зелені та жовті відтінки в порівнянні з TimeGAN, що свідчить про нижчі похибки прогнозування на більшості періодів. Зокрема, для метрики MAE модель TimeF-GAN показує значення від 2,09 у четвертому кварталі 2023 р. до 6,56 у першому кварталі 2025 р., тоді як TimeGAN – від 2,09 у четвертому кварталі 2023 р. до 7,52 у третьому кварталі 2023 р.

По-друге, спостерігається виражена часова нестабільність похибок для обох моделей. Найнижчі похибки зафіксовані в період четвертого кварталу 2023 р., що позначено темно-зеленим кольором (MAE = 2,09 для обох моделей). Найвищі похибки характерні для періодів третього кварталу 2023 р. та першого кварталу 2025 р., що відображено червоними відтінками (MAE до 7,52 та RMSE до 8,59 для TimeGAN).

Варіативність похибок по періодах пояснюється різними ринковими умовами: періоди з низькою волатильністю в четвертому кварталі 2023 р. характеризуються вищою точністю прогнозування, тоді як періоди структурних змін або високої волатильності в першому кварталі 2025 р. супроводжуються підвищенням похибок. Проте навіть у найбільш складних ринкових умовах інтеграція сентимент-аналізу дозволяє знизити похибку на 12–17%, що візуально відображено різницею в інтенсивності кольорів між рядками TimeGAN та TimeF-GAN.

Додатково, як ілюстровано на рисунках 6, 7, за використання із сентимент-індексом трансформерна модель демонструє досить високу точність у відтворенні кривих цін закриття. Прогнозні криві практично збігаються з реальними, що підтверджує здатність моделі точно відтворювати динаміку фондового ринку.

Висновки. У роботі розроблено й експериментально перевірено інтегровану архітектуру TimeF-GAN, яка об'єднує можливості генеративно-змагальних мереж із технологіями аналізу фінансового сентименту на базі моделі FinBERT.

Запропонований підхід відрізняється від традиційних статистичних і класичних нейромережевих методів комплексним використанням як кількісних ознак (цінові ряди, технічні індика-

Таблиця 1

Результати оцінювання якості прогнозування моделей на різних часових періодах

Період	MAE		RMSE		R^2	
	TimeGAN	TimeF-GAN	TimeGAN	TimeF-GAN	TimeGAN	TimeF-GAN
1-й квартал 2023 р.	4,34	3,67	3,7	3,3	0,45	0,32
2-й квартал 2023 р.	5,21	3,53	5,6	4,8	0,61	0,48
3-й квартал 2023 р.	7,52	6,24	8,3	6,7	0,33	0,5
4-й квартал 2023 р.	2.09	3,28	2,49	3,76	0,19	0,83
1-й квартал 2024 р.	5,53	4,98	7,21	6,52	0,21	0,15
2-й квартал 2024 р.	5,05	4,05	7,13	5,88	0,48	0,64
3-й квартал 2024 р.	3,9	3,54	4,57	4,2	0,68	0,42
4-й квартал 2024 р.	4,77	3,69	5,68	4,7	0,67	0,76
1-й квартал 2025 р.	6,57	6,56	8,59	8,51	0,52	0,53
2-й квартал 2025 р.	5,63	3,98	6,55	4,82	0,27	0,77
3-й квартал 2025 р.	5,98	5,15	7,77	6,69	0,12	0,35

Рис. 5. Теплова карта похибок по періодах

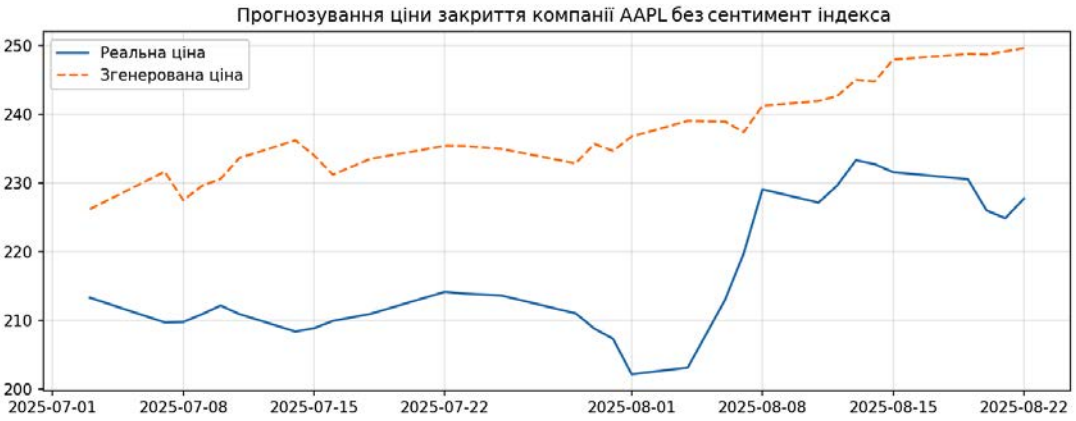


Рис. 6. Отримані дані прогнозу без урахування сентимент-аналізу

Рис. 7. Отримані дані прогнозу з урахуванням сентимент-аналізу

тори), так і якісної інформації, отриманої з новинних джерел та соціальних медіа. Така інтеграція забезпечила істотне підвищення точності прогнозування цінової динаміки акцій.

Експериментальна валідація моделі на даних компанії “Apple” підтвердила переваги сентимент-збагаченого підходу: спостерігалося зниження середньої абсолютної похибки (MAE) та

середньоквадратичної похибки (RMSE) за одночасного зростання коефіцієнта детермінації (R^2), порівняно з моделями, що базуються суто на числових фінансових показниках.

Для оцінювання стабільності й узагальнювальної здатності моделі проведено тестування на дев'яти незалежних часових періодах протягом 2023–2025 рр., що охоплювали різні ринкові умови. Середнє значення MAE становило 4,42 по всіх періодах, демонструючи надійність моделі у варіативному середовищі. Критично важливим є те, що інтеграція сентимент-індексу зменшила не лише абсолютну величину похибки, але і її дисперсію між періодами на 14,0%, що підвищує передбачуваність та практичну застосовність прогнозів.

Модель було протестовано на даних компанії "Apple", яка характеризується високою капіталізацією та значним інформаційним супроводом, що дозволило перевірити ефективність інтеграції ринкових і новинних ознак. Водночас подальші дослідження передбачають апробацію підходу на даних інших компаній (Microsoft, Tesla, Amazon), що дасть змогу оцінити узагальнюваність і стабільність моделі в різних секторах фондового ринку.

Отримані результати свідчать, що поєднання GAN та новинного сентименту формує більш гнучкі й надійні системи прогнозування, здатні адаптуватися до інформаційної турбулентності фінансових ринків. Практична цінність такого підходу полягає в можливості його використання інвесторами й аналітиками для підвищення обґрунтованості рішень в умовах високої невизначеності. Перспективним напрямом подальших робіт є розширення моделі завдяки багатомодальним даним – інтеграції макроекономічних показників, альтернативних інформаційних потоків та інших ринкових факторів, що сприятиме підвищенню точності та стійкості прогнозів.

ЛІТЕРАТУРА

1. Yoon J., Jarrett D., van der Schaar M. Time-series generative adversarial networks. In Advances in Neural Information Processing Systems (NeurIPS). 2019. P. 5509–5519.
2. Wiese M., Knobloch R., Korn R., Kretschmer P. Quant GANs: Deep generation of financial time series. Journal of Computational Finance. 2020. № 24 (4). P. 1–24.
3. Hariri S., Kind M., Brunner R. J. Time series forecasting using generative adversarial networks (GANs): A systematic review. ACM Computing Surveys. 2021. № 54 (6). P. 1–38.
4. Zhou Y., Zhang L., Wang J., Li S. Hybrid GAN-LSTM models for financial time series prediction. Expert Systems with Applications. 2023. № 213. P. 118897.
5. Wang X., Chen Y., Li Q., Yang S. Factor-GAN: A generative adversarial approach for multivariate financial time series forecasting. Neurocomputing. 2024. № 570. P. 126–139.
6. Kirisci M., Cagcag Y., Ozge A New CNN-Based Model for Financial Time Series: TAIEX and FTSE Stocks Forecasting. *Neural Processing Letters.* 2022. 54. DOI: 10.1007/s11063-022-10767-z
7. Xu R., Wang Z., Li M., Zhang H. VGC-GAN: Variational graph convolutional generative adversarial networks for multivariate stock prediction. Knowledge-Based Systems. 2024. 293. P. 111526.
8. Araci D. FinBERT: Financial sentiment analysis with pre-trained language models. arXiv preprint arXiv. 2019. 1908. P. 10063.
9. Yang T., Liu H., Zhao W., Zhou X. News-driven stock market forecasting with FinBERT-enhanced sentiment and GAN-based data augmentation. Information Sciences. 2025. 681. P. 202–218.
10. Liu Y., Wu X., Zhang K. Integrating FinBERT sentiment into GAN-based financial forecasting : A comparative study. Journal of Forecasting. 2025. 44 (2). P. 223–240.
11. Перцев Ю. О., Коротка Л. І. Порівняння нейронних мереж RNN та LSTM при прогнозуванні цін на фондовому ринку. Комп'ютерне моделювання та оптимізація складних систем : матеріали VIII Міжнародної науково-технічної конференції, м. Дніпро, 1–3 листопада 2023 р. 2023. С. 124–127.
12. Перцев Ю. О., Коротка Л. І. Нейромережеве прогнозування цін на фондовому ринку. Information Technologies in Metallurgy and Machine building : International scientific and technical conference, м. Дніпро, 22 березня 2023 р. 2023. С. 314–317.
13. Перцев Ю. О., Коротка Л. І. Порівняльний аналіз традиційних статистичних методів та нейромережевої моделі LSTM. Системні технології. 2025. № 1 (156). С. 65–77.
14. Перцев Ю. О., Коротка Л. І. Інноваційний підхід у прогнозуванні часових рядів: від традиційних методів до новаторської моделі TIMESFM. Information Technologies in Metallurgy and Machine building : International scientific and technical conference, м. Дніпро, 10–11 квітня 2024 р. 2024. P. 434–439.
15. Перцев Ю. О., Коротка Л. І. Аналіз та адаптація генеративно-змагальних мереж до фінансових ринків. Таврійський науковий вісник. Серія «Технічні науки». 2025. № 2. С. 122–134. https://doi.org/10.32782/tnv-tech.2025.2.14
16. Ткаліченко С. В. Штучні нейронні мережі : навчальний посібник. Кривий Ріг : Державний університет економіки і технологій, 2023. 150 с.
17. Терейковський І. А., Бушуєв Д. А., Терейковська Л.О. Штучні нейронні мережі: базові положення : навчальний посібник. Київ : КПІ ім. Ігоря Сікорського, 2022. 123 с.

18. Субботін С. О. Нейронні мережі: теорія та практика : навчальний посібник. Житомир : вид. О. О. Євенок, 2020. 184 с.
19. Li J., Sun P., Huang Z., Liu Y. Volatility prediction using GAN-augmented time series and sentiment features. IEEE Transactions on Neural Networks and Learning Systems. 2023. № 34 (9). P. 4567–4579.
20. Santoro D., Grilli L. Generative adversarial network to evaluate quantity of information in financial markets. *Neural Comput. Appl.* 2022. № 34. P. 17473–17490. DOI: 10.1007/s00521-022-07401-3

REFERENCES

1. Yoon, J., Jarrett, D., & van der Schaar, M. (2019). Time-series generative adversarial networks. In Advances in Neural Information Processing Systems (NeurIPS), pp. 5509–5519.
2. Wiese, M., Knobloch, R., Korn, R., & Kretschmer, P. (2020). Quant GANs: Deep generation of financial time series. Journal of Computational Finance, 24 (4), 1–24.
3. Hariri, S., Kind, M., & Brunner, R. J. (2021). Time series forecasting using generative adversarial networks (GANs): A systematic review. ACM Computing Surveys, 54 (6), 1–38.
4. Zhou, Y., Zhang, L., Wang, J., & Li, S. (2023). Hybrid GAN-LSTM models for financial time series prediction. Expert Systems with Applications, 213, 118897.
5. Wang, X., Chen, Y., Li, Q., & Yang, S. (2024). Factor-GAN: A generative adversarial approach for multivariate financial time series forecasting. Neurocomputing, 570, 126–139.
6. Xu, R., Wang, Z., Li, M., & Zhang, H. (2024). VGC-GAN: Variational graph convolutional generative adversarial networks for multivariate stock prediction. Knowledge-Based Systems, 293, 111526.
7. Araci, D. (2019). FinBERT: Financial sentiment analysis with pre-trained language models. arXiv preprint arXiv:1908.10063.
8. Li, J., Sun, P., Huang, Z., & Liu, Y. (2023). Volatility prediction using GAN-augmented time series and sentiment features. IEEE Transactions on Neural Networks and Learning Systems, 34 (9), 4567–4579.
9. Yang, T., Liu, H., Zhao, W., & Zhou, X. (2025). News-driven stock market forecasting with FinBERT-enhanced sentiment and GAN-based data augmentation. Information Sciences, 681, 202–218.
10. Liu, Y., Wu, X., & Zhang, K. (2025). Integrating FinBERT sentiment into GAN-based financial forecasting: A comparative study. Journal of Forecasting, 44 (2), 223–240.
11. Pertsev, Y. O., & Korotka, L. I. (2023) Porivniannia neironnykh merezh RNN ta LSTM typu pry prohnozuvanni tsin na fondovomu rynku [Comparison of RNN and LSTM Neural Networks in Stock Market Price Prediction]. Proceedings of the materialy VIII Mizhnarodnoi naukovo-tekhnichnoi konferentsii kompiuterne modeliuvannia ta optymizatsiia skladnykh system, Ukraine, Dnipro, November 1–3, 2023, Dnipro: CMOCS-2023, pp. 124–127.
12. Pertsev, Y. O., & Korotka, L. I. (2023) Neiromerezheve prohnozuvannia tsin na fondovomu rynku [Neural Network-Based Stock Market Price Forecasting]. Proceedings of the International scientific and technical conference Information Technologies in Metallurgy and Machine building, Ukraine, Dnipro, March 22, 2023, Dnipro: ITMM 2023, pp. 314–317. DOI: 10.34185/1991-7848.itmm.2023.01.085
13. Pertsev, Y. O., & Korotka, L. I. (2025) Porivnialnyi analiz tradytsiinykh statystychnykh metodiv ta neiromerezhevoi modeli LSTM [Comparative analysis of traditional statistical methods and the LSTM neural network model]. System technologies, vol. 1, № 156, pp. 65–77. DOI: 10.34185/1562-9945-1-156-2025-08
14. Pertsev, Y. O., & Korotka, L.I. (2024). Innovatsiinyi pidkhid u prohnozuvanni chasovykh riadiv: vid tradytsiinykh metodiv do novatorskoi modeli TIMESFM [An Innovative Approach to Time Series Forecasting: From Traditional Methods to the groundbreaking TIMESFM Model]. Proceedings of the International scientific and technical conference Information Technologies in Metallurgy and Machine building, Ukraine, Dnipro, April 10–11, 2024, Dnipro: ITMM 2024. pp. 434–439. DOI: 10.34185/1991-7848.itmm.2024.01.084
15. Pertsev, Y. O., & Korotka, L. I. (2025). Analiz ta adaptatsiia heneratyvno-zmahalnykh merezh do finansovykh rynkiv. Tavriiskyi naukovyi visnyk. Seriia: Tekhnichni nauky, (2), 122–134. https://doi.org/10.32782/tnv-tech.2025.2.14
16. Tkalichenko, S. V. (2023). Shtuchni neironni merezhi: Navchalnyi posibnyk. Kryvyi Rih: Derzhavnyi universytet ekonomiky i tekhnolohii, 2023. 150 s.
17. Tereikovskyi, I. A., Bushuiev, D. A., & Tereikovska, L.O. (2022). Shtuchni neironni merezhi: bazovi polozhennia: navchalnyi posibnyk. Kyiv: KPI im. Ihoria Sikorskoho, 2022. 123 s
18. Subbotin, S. O. (2020). Neironni merezhi: teoriia ta praktyka: navch. posib. Zhytomyr: vyd. O. O. Yevenok, 2020. 184 s.
19. Li, J., Sun, P., Huang, Z., & Liu, Y. (2023). Volatility prediction using GAN-augmented time series and sentiment features. IEEE Transactions on Neural Networks and Learning Systems, 34 (9), 4567–4579.
20. Santoro, D., & Grilli, L. (2022). Generative adversarial network to evaluate quantity of information in financial markets. Neural Comput. Appl. 34, 17473–17490. DOI: 10.1007/s00521-022-07401-3

Дата першого надходження рукопису до видання: 26.08.2025
Дата прийнятого до друку рукопису після рецензування: 24.09.2025
Дата публікації: 31.12.2025

УДК 004.7: 004.4: 519.6
DOI https://doi.org/10.26661/2786-6254-2025-2-12

РОЗРОБКА ВЕБСЕРВЕРА ДЛЯ ХМАРНИХ СИСТЕМ СКІНЧЕННО-ЕЛЕМЕНТНОГО АНАЛІЗУ

Просяна Д. І.
викладач кафедри комп'ютерних наук
Запорізький національний університет
вул. Університетська, 66, Запоріжжя, Україна
orcid.org/0000-0002-3188-7942
prsjana.daria.i@gmail.com

Лісняк А. О.
кандидат фізико-математичних наук, доцент,
завідувач кафедри програмної інженерії
Запорізький національний університет
вул. Університетська, 66, Запоріжжя, Україна
orcid.org/0000-0001-9669-7858
a.lisnyak@znu.edu.ua

Ключові слова: *метод скінченних елементів, хмарні обчислення, вебсервер, бекенд, Go, SaaS.*

У роботі розглянуто підхід до створення хмарної системи скінченно-елементного аналізу, яка поєднує сучасні вебтехнології та метод скінченних елементів. Показано, що використання хмарних обчислень в задачах інженерного моделювання забезпечує новий рівень автоматизації процесів проєктування, підвищує ефективність використання обчислювальних ресурсів і робить високопродуктивні розрахунки доступними для широкого кола користувачів. Сучасна хмарна архітектура включає чотири основні компоненти: інтерфейс користувача, вебсервер, систему скінченно-елементного аналізу й обчислювальний кластер, взаємодія між якими здійснюється через мережеві протоколи REST API або WebSocket. Така структура дозволяє забезпечити гнучкість масштабування, модульність та простоту інтеграції із зовнішніми сервісами.

Реалізація запропонованого вебсервера виконана мовою програмування Go, яка завдяки вбудованим засобам паралелізму та підтримці конкурентних процесів забезпечує високу продуктивність і стабільність роботи. Розроблена структура каталогів проєкту забезпечує логічне розділення компонентів, що спрощує розроблення, налагодження, підтримку та подальше розширення системи. Реалізований програмний прототип дозволяє розв'язувати задачі статики у пружній постановці, забезпечує завантаження сіткових моделей, створення, збереження і редагування задач, а також відображення результатів розрахунку у форматі JSON через вебінтерфейс.

Отримані результати демонструють практичну доцільність поєднання вебтехнологій із методами інженерного аналізу для створення масштабованих обчислювальних систем. Запропонований підхід може бути основою для побудови промислових рішень, орієнтованих на віддалене виконання скінченно-елементних розрахунків у хмарному середовищі. Подальші дослідження доцільно спрямувати на розширення можливостей обчислювального модуля (урахування динамічних, теплових і нелінійних задач), реалізацію розподілених обчислень на кластерних системах і підвищення інтерактивності користувацького інтерфейсу.

DEVELOPMENT OF A WEB SERVER APPLICATION FOR CLOUD-BASED FINITE ELEMENT ANALYSIS

Prosiana D. I.
Lecturer at the Department of Computer Science
Zaporizhzhia National University
Universitetska str., 66, Zaporizhzhia, Ukraine
orcid.org/0000-0002-3188-7942
prsjana.darja.i@gmail.com

Lisnyak A. O.
Candidate of Physical and Mathematical Sciences, Associate Professor,
Head of the Department of Software Engineering
Zaporizhzhia National University
Universitetska str., 66, Zaporizhzhia, Ukraine
orcid.org/0000-0001-9669-7858
a.lisnyak@znu.edu.ua

Key words: *finite element method, cloud computing, web server, backend, Go, SaaS.*

This paper presents an approach to the development of a cloud-based finite element analysis system that integrates modern web technologies with the finite element method. It has been shown that the use of cloud computing in engineering modeling provides a new level of automation in design processes, increases computational efficiency, and makes high-performance analysis accessible to a wide range of users. The proposed system architecture includes four key components: a user interface, a web server, a finite element computation module, and a computing cluster, which interact through REST API or WebSocket protocols. Such an architecture ensures modularity, scalability, and ease of integration with external data and storage services.

The web server was implemented in the Go programming language, which provides built-in support for concurrency, high network performance, and stability. The designed directory structure of the project ensures logical separation of components, simplifies development and debugging, and facilitates future system extension. The implemented prototype supports static elasticity problems, enables uploading mesh models, creating, saving, and editing problem definitions, and displays calculation results in JSON format through a web-based interface.

The obtained results confirm the feasibility of combining web technologies with computational mechanics tools for building scalable engineering analysis systems operating in cloud environments. The proposed approach can serve as a foundation for industrial-grade solutions focused on remote execution of finite element simulations. Further research should aim at extending the computational module to include dynamic, thermal, and nonlinear problems, implementing distributed calculations on clusters, and improving the interactivity of the user interface to enhance usability and performance.

Вступ. Запорукою розвитку машинобудування на сучасному етапі є широке використання систем автоматизованого проєктування (далі – САПР), які дозволяють реалізувати два основні завдання: 1) синтез (автоматизацію створення проєктної документації, креслень, технічних завдань, звітів тощо) та 2) аналіз (автоматизацію дослідження відповідності функціональних характеристик об’єкта, що проєктується, вимогам замовника) [1].

Найбільш важливим завданням сучасних САПР є саме автоматизація аналізу фізичних властивостей об’єктів проєктування, оскільки це дає мож-

ливість істотно скоротити тривалість та вартість розроблення складних інженерно-технічних систем завдяки заміні великої кількості випробувань дослідних зразків (часто з їх пошкодженням або повним руйнуванням) процесом комп'ютерної симуляції.

Одним із найбільш поширених і ефективних натепер наближених підходів до математичного моделювання фізичних явищ і процесів різної природи (міцності, теплопровідності, коливань, деформацій та інших) є метод скінченних елементів (далі – МСЕ) [2]. Для його використання створено велику кількість спеціалізованих програмних систем скінченно-елементного аналізу. Серед пропрієтарних програм можна відзначити SIMULIA Abaqus [3], COMSOL Multiphysics [4], Ansys [5], MSC Nastran [6], SOLIDWORKS Simulation [7]. Серед систем із відкритим початковим кодом можна виділити Code_Aster [8], OpenFOAM [9], FreeFEM [10], OpenFEM [11] та інші програмні комплекси [12].

Незважаючи на широке коло доступних засобів скінченно-елементного моделювання, постійне ускладнення інженерних завдань в машинобудуванні та будівництві зумовлює зростання вимог як до функціональних можливостей програмного забезпечення, так і до обчислювальних ресурсів, необхідних для його ефективної експлуатації. Натепер застосування високопродуктивних комп'ютерних систем потребує значних капітальних і експлуатаційних витрат, що часто робить їх використання економічно невиправданим для багатьох промислових підприємств.

У зв'язку із цим у сучасній практиці проєктування сформувалася нова парадигма використання обчислювальних ресурсів, що ґрунтується на принципі віддаленого доступу до обчислювальних потужностей спеціалізованих комп'ютерних центрів, які належать третім сторонам. Такий підхід реалізується в межах концепції хмарних технологій, сутність яких полягає в наданні користувачам можливості оренди обчислювальних ресурсів або програмного забезпечення (далі – ПЗ) на визначений термін замість їх придбання.

Застосування хмарних технологій забезпечує низку переваг, серед яких варто відзначити зменшення капітальних витрат, гнучкість масштабування обчислювальних ресурсів відповідно до потреб користувача, а також зниження витрат на технічне обслуговування та оновлення обладнання.

Користувачі можуть орендувати як обчислювальні ресурси для виконання власних програмних розробок, так і готові програмні рішення, що функціонують безпосередньо у хмарі. Останній підхід набув поширення та відомий під назвою «програмне забезпечення як послуга» (SaaS – Software as a Service), який передбачає надання користувачам доступу до програмних продуктів через мережу «Інтернет» з оплатою за фактичне використання.

Натепер існує відносно невелика кількість систем скінченно-елементного аналізу, що підтримують хмарні технології. Серед них можна відзначити SPARSELAB [19], SimScale [20] та інші [21]. Проте, незважаючи на існування таких програм, завдання створення нових хмарних сервісів для скінченно-елементного аналізу є актуальним, оскільки постійно змінюються як можливості обчислювальної техніки, так і вимоги до відповідного ПЗ. Отже, створення систем скінченно-елементного аналізу з підтримкою хмарних технологій є в наш час актуальним завданням.

Результати. ***Архітектура хмарної системи скінченно-елементних розрахунків***

Архітектура системи скінченно-елементного аналізу, що працює у хмарі, може складатися із чотирьох основних компонентів (рис. 1).

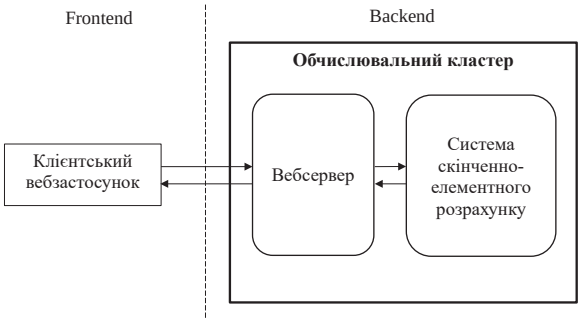


Рис. 1. Типова архітектура хмарної системи скінченно-елементного розрахунку

1. Інтерфейс користувача (fronted), який реалізується як вебзастосунок, що дозволяє інженеру задавати основні параметри скінченно-елементного розрахунку: дискретну модель, фізичні характеристики, крайові умови та навантаження тощо. Для реалізації fronted зазвичай застосовуються сучасні вебфреймворки, наприклад React [13].

2. Вебсервер (backend), який безпосередньо обробляє запити користувачів на виконання скінченно-елементних розрахунків. Однією з найголовніших його функцій є керування чергою розрахунків у хмарному середовищі. Для реалізації backend можна використовувати фреймворки на основі мови програмування Python (наприклад, Flask [14]) або Node.js [15]. Проте, на думку авторів, найкращим способом реалізації такого вебсервера буде використання мови програмування Go [16], яка спеціально була створена компанією "Google" для розроблення високопродуктивних вебсервісів.

3. Система скінченно-елементного розрахунку (backend), що безпосередньо виконує всі обчис-

лення за заданою користувачем схемою розрахунків. Здебільшого такі системи реалізуються мовами програмування C++, Rust, Go й іншими, що дозволяють створювати швидкий, ефективний і безпечний код, який буде працювати в конкурентному середовищі.

4. Обчислювальний кластер, на якому безпосередньо виконуються розрахунки.

Водночас комунікація між цими компонентами може здійснюватися за допомогою, наприклад, REST API [17] або WebSocket [18]. Параметри завдання та результати розрахунків можуть зберігатися в базі даних або безпосередньо у файловій системі хмарного сховища (наприклад, Google Cloud Storage), після чого користувач може до них звертатися для аналізу та корегування (за необхідності).

Тут також варто зазначити, що функції вебсервера й системи скінченно-елементного розрахунку під час використання технології прототипування можуть бути об'єднані в одній фізичній програмі – прототипі, для зручності реалізації і тестування застосунку.

Програмна реалізація вебсервера системи скінченно-елементного аналізу

Для реалізації вебсеревера системи скінченно-елементного аналізу було використано мову програмування Go, оскільки вона спочатку призначалася для розроблення веб- та мікросервісів, підтримує паралельні обчислення, її стандартна бібліотека забезпечує ефективну підтримку мережевих операцій. Окрім того, для цієї мови реалізовано велику кількість сторонніх бібліотек відповідного профілю.

Під час створення вебсервера була застосована така структура каталогів (рис. 2), яка дозволяє розділити зберігання файлів фронтенду й бекенду, а також даних і результатів розрахунку.

У каталозі *cmd,* в окремих підкаталогах *web* і *fem,* зберігаються файли, що реалізують функціональність вебсервера й системи скінченно-елементного аналізу відповідно. У каталозі *data* розташовані тимчасові файли. У *downloads* – завантажені на сервер геометричні моделі областей розрахунку, а в *save* зберігаються файли завдань. Каталог *ui* призначений для збереження файлів, що утворюють інтерфейс користувача (підкаталог *static* містить статичні файли, наприклад, CSS).

Оскільки в даній роботі описується прототип хмарного вебсервера, то для простоти було реалізовано скінченно-елементні розрахунки лише задач статики у пружній постановці. Після запуску вебсервера він виводить повідомлення про вебадресу та порт, за якими до нього можна підключитися (рис. 3).

Після введення відповідної адреси у браузері відкриється головна сторінка вебзастосунку керування сервером (рис. 4), який повністю написано мовою Go.

За допомогою команди головного меню *Upload mesh* користувач може завантажити на сервер (у каталог *downloads*) скінченно-елементну модель об'єкта розрахунку. Далі за допомогою команди меню *New problem* користувач може створити нове завдання (рис. 5), або за допомогою команди *Open problem* відкрити раніше збережене на сервері (у каталозі *save*).

Після активації розрахунку (кнопка *Calculate* на вебсторінці параметрів завдання) вебсервер завантажує параметри розрахунку й передає їх

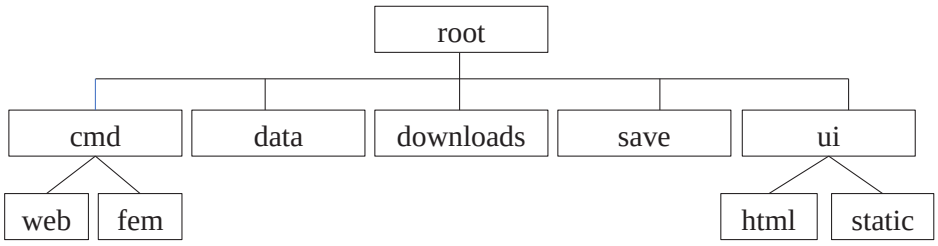


Рис. 2. Структура каталогів вебсервера

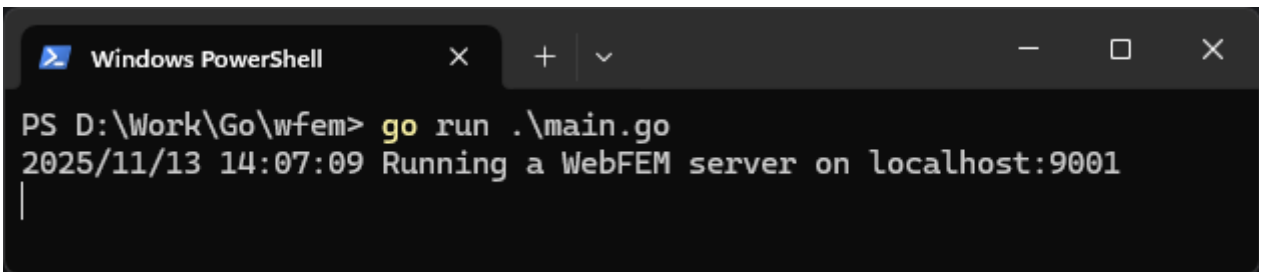


Рис. 3. Запуск вебсервера

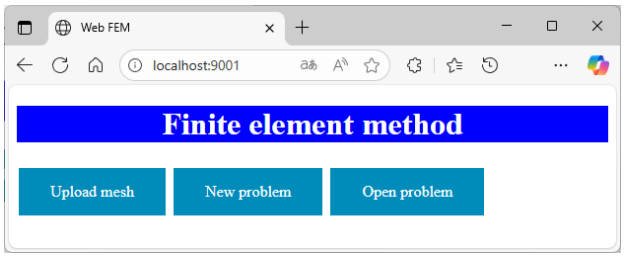


Рис. 4. Головна сторінка програми керування вебсервером

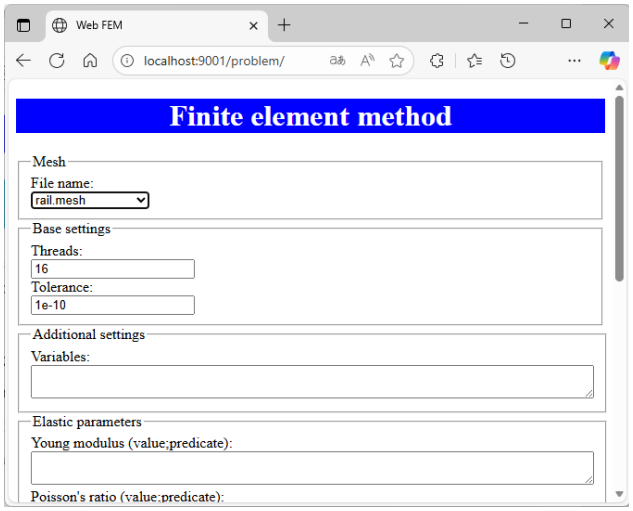


Рис. 5. Сторінка створення нового завдання

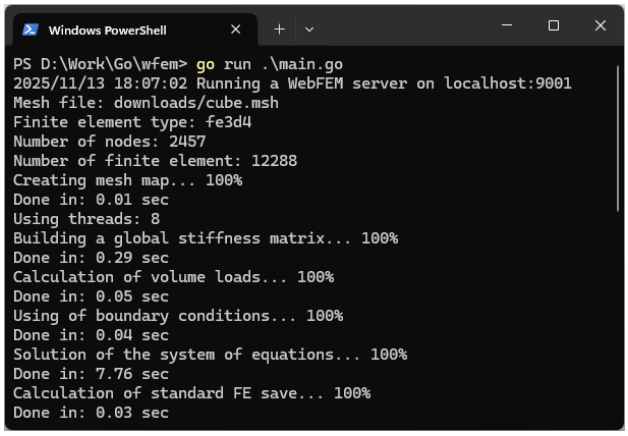


Рис. 6. Індикація процесу роботи вебсервера

для обробки модулю скінченно-елементного розрахунку. Вебсервер водночас виводить детальну інформацію про стан і етапи розрахунку завдання (рис. 6).

Після завершення обчислень результат розрахунку у форматі JSON передається для відображення на відповідну вебсторінку застосунку керування вебсервером (рис. 7).

Дискусія. У роботі розглянуто підхід до створення хмарної системи скінченно-елементного

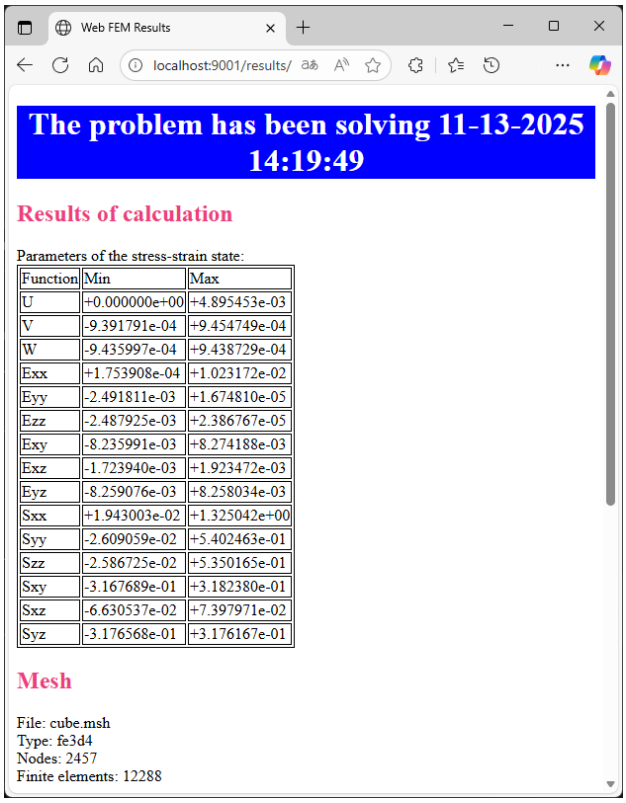


Рис. 7. Вебсторінка результатів розрахунку

аналізу, що поєднує сучасні вебтехнології та метод скінченних елементів. Показано, що застосування хмарних обчислень у САПР відкриває нові можливості для автоматизації та оптимізації процесів проєктування, зменшує капітальні й експлуатаційні витрати, а також підвищує доступність високопродуктивних розрахунків для широкого кола користувачів.

Наведена архітектура хмарної системи складається із чотирьох основних компонентів – інтерфейсу користувача, вебсервера, системи скінченно-елементного аналізу й обчислювального кластера, на якому ця інфраструктура розгорнута. Така структурна організація забезпечує чітке розмежування функцій між підсистемами, гнучкість масштабування та можливість інтеграції з різними сервісами зберігання даних і системами керування розрахунками.

Програмна реалізація вебсервера мовою Go підтвердила доцільність її використання для створення високопродуктивних і надійних вебсервісів, орієнтованих на паралельні обчислення. Вбудовані засоби цієї мови для роботи з мережею та підтримка конкурентності дали змогу ефективно організувати обмін даними між користувачем і системою скінченно-елементного розрахунку. Запропонована структура каталогів проєкту забезпечує логічну організацію компонентів, спрощує розроблення, налагодження та подальше розширення системи.

Реалізований прототип вебсервера виконує розв'язання завдань статики у пружній постановці, забезпечує інтерфейс для завантаження моделей, створення та збереження завдань, а також відображення результатів розрахунків. Отримані результати демонструють перспективність застосування такого підходу для побудови масштабованих систем скінченно-елементного аналізу, що функціонують у хмарному середовищі.

Висновки. Отже, у роботі розв'язано завдання створення прототипу хмарного сервісу для скінченно-елементного аналізу з урахуванням сучасних можливостей обчислювальної техніки.

Подальший розвиток роботи може бути спрямований на розширення функціональних можливостей запропонованого хмарного сервісу: розв'язання задач динаміки, термопружності, в'язкопружності й інших); реалізацію підтримки розподілених обчислень на кластерах; управління безпекою та правами доступу; реалізацію системи подій (логування), а також інтеграцію системи з іншими інструментами інженерного аналізу. Це дозволить створити повноцінну хмарну платформу для проведення наукових і промислових розрахунків на основі методу скінченних елементів.

ЛІТЕРАТУРА

1. Introduction to CAD Systems. URL: https://surl.li/indifs (дата звернення: 01.10.2025).
2. Zienkiewicz O. C., Taylor R. L., Zhu J. Z. The Finite Element Method: Its Basis and Fundamentals. Sixth edition. Butterworth-Heinemann, 2016. 753 p.
3. Abaqus Finite Element Analysis. *SIMULIA*. URL: https://www.3ds.com/products/simulia/abaqus (дата звернення: 01.11.2025).
4. COMSOL Multiphysics Simulation Software. URL: https://www.comsol.com/comsol-multiphysics (дата звернення: 01.11.2025).
5. Engineering Simulation & 3D Design Software. *Ansys*. URL: https://www.ansys.com/ (дата звернення: 01.11.2025).
6. MSC Nastran – Multidisciplinary Structural Analysis. URL: http://surl.li/schezo (дата звернення: 01.11.2025).
7. 3D CAD FEA Simulation & Analysis Software. *SOLIDWORKS*. URL: https://www.solidworks.com/product/solidworks-simulation (дата звернення: 01.11.2025).
8. Code_Aster. URL: https://code-aster.org/spip.php?rubrique2 (дата звернення: 01.11.2025).
9. OpenFOAM. URL: https://www.openfoam.com/ (дата звернення: 01.11.2025).
10. FreeFEM – An open-source PDE Solver using the Finite Element Method. URL: https://freefem.org/ (дата звернення: 23.10.2025).
11. OpenFEM Library. URL: https://www.sdtools.com/research/openfem/ (дата звернення: 25.10.2025).
12. Best Open-Source Finite Element Analysis Software. URL: http://surl.li/vmblnr (дата звернення: 25.10.2025).
13. React. URL: https://react.dev/ (дата звернення: 28.10.2025).
14. Welcome to Flask. URL: https://flask.palletsprojects.com/en/stable/ (дата звернення: 27.10.2025).
15. Node.js – Run JavaScript Everywhere. URL: https://nodejs.org/uk (дата звернення: 27.10.2025).
16. The Go Programming Language. URL: https://go.dev/ (дата звернення: 27.10.2025).
17. REST API Tutorial. URL: https://restfulapi.net/ (дата звернення: 27.10.2025).
18. The Websocket Protocol. URL: https://websocket.org/ (дата звернення: 27.10.2025).
19. Finite Element Analysis in the Cloud. URL: https://sparselab.com/ (дата звернення: 27.10.2025).
20. Structural Mechanics Simulation | SimScale. URL: https://www.simscale.com/ (дата звернення: 27.10.2025).
21. Finite Element Analysis with Cloud-Based High-Performance Computing. URL: https://surl.li/htvkqi (дата звернення: 27.10.2025).

REFERENCES

1. Introduction to CAD Systems. URL: https://surl.li/indifs (accessed 01.10.2025).
2. Zienkiewicz, O. C., Taylor, R. L., Zhu, J. Z. (2016). The Finite Element Method: Its Basis and Fundamentals. Sixth edition. Butterworth-Heinemann, 753 p.
3. Abaqus Finite Element Analysis. SIMULIA. URL: https://www.3ds.com/products/simulia/abaqus (accessed 01.11.2025).
4. COMSOL Multiphysics Simulation Software. URL: https://www.comsol.com/comsol-multiphysics (accessed 01.11.2025).
5. Engineering Simulation & 3D Design Software. Ansys. URL: https://www.ansys.com/ (accessed 01.11.2025).
6. MSC Nastran – Multidisciplinary Structural Analysis. URL: http://surl.li/schezo (accessed 01.11.2025).

7. 3D CAD FEA Simulation & Analysis Software | SOLIDWORKS. URL: https://www.solidworks.com/product/solidworks-simulation (accessed 01.11.2025).
8. Code_Aster. URL: https://code-aster.org/spip.php?rubrique2 (accessed 01.11.2025).
9. OpenFOAM. URL: https://www.openfoam.com/ (accessed 01.11.2025).
10. FreeFEM – An open-source PDE Solver using the Finite Element Method. URL: https://freefem.org/ (accessed 23.10.2025).
11. OpenFEM Library. URL: https://www.sdtools.com/research/openfem/ (accessed 25.10.2025).
12. Best Open-Source Finite Element Analysis Software. URL: http://surl.li/vmblnr (accessed 25.10.2025).
13. React. URL: https://react.dev/ (accessed 28.10.2025).
14. Welcome to Flask. URL: https://flask.palletsprojects.com/en/stable/ (accessed 27.10.2025).
15. Node.js – Run JavaScript Everywhere. URL: https://nodejs.org/uk (accessed 27.10.2025).
16. The Go Programming Language. URL: https://go.dev/ (accessed 27.10.2025).
17. REST API Tutorial. URL: https://restfulapi.net/ (accessed 27.10.2025).
18. The Websocket Protocol. URL: https://websocket.org/ (accessed 27.10.2025).
19. Finite Element Analysis in the Cloud. URL: https://sparselab.com/ (accessed 27.10.2025).
20. Structural Mechanics Simulation | SimScale. URL: https://www.simscale.com/ (accessed 27.10.2025).
21. Finite Element Analysis with Cloud-Based High-Performance Computing. URL: https://surl.li/htvkqi (accessed 27.10.2025).

Дата першого надходження рукопису до видання: 22.08.2025
Дата прийнятого до друку рукопису після рецензування: 18.09.2025
Дата публікації: 31.12.2025

УДК 004.8, 004.93, 004.932
DOI https://doi.org/10.26661/2786-6254-2025-2-13

МЕТОДОЛОГІЯ СТВОРЕННЯ ФОРМАТИЗОВАНИХ НАБОРІВ ДАНИХ ЗА ДОПОМОГОЮ СИМУЛЯТОРА AIRSIM ТА ІНСТРУМЕНТІВ ЇХ АНОТАЦІЇ ТА АУГМЕНТАЦІЇ

Сініцин І. П.
доктор технічних наук, професор,
член-кореспондент Національної академії наук України, директор
Інститут програмних систем Національної академії наук України
просп. Академіка Глушкова, 40, Київ, Україна
orcid.org/0000-0002-4120-0784
ips@nas.gov.ua

Кирилов І. І.
аспірант
Інститут програмних систем Національної академії наук України
просп. Академіка Глушкова, 40, Київ, Україна
orcid.org/0009-0006-9756-5391
ivan.kyrylov.science@gmail.com

Заїка К. А.
аспірант
Інститут програмних систем Національної академії наук України
просп. Академіка Глушкова, 40, Київ, Україна
orcid.org/0009-0006-1107-5106
kir.z2019@gmail.com

Яценко О. А.
кандидат фізико-математичних наук,
старший науковий співробітник
Інститут програмних систем Національної академії наук України
просп. Академіка Глушкова, 40, Київ, Україна
orcid.org/0000-0002-4700-6704
oayat@ukr.net

Ключові слова: *анотація даних, аугментація даних, генерація даних, збір даних, комп'ютерний зір, маркування даних для машинного навчання, правила анотації, AirSim, YOLO.*

Запропоновано методологію створення форматизованих наборів даних для завдань комп'ютерного зору з використанням симулятора Microsoft AirSim і сучасних інструментів їх анотації і аугментації. З огляду на те, що підготовка навчальних даних для моделей глибокого навчання є одним із найбільш витратних і трудомістких етапів, розроблена методика спрямована на зниження вартості та часу формування якісних наборів даних за збереження їхньої різноманітності й узгодженості. Симулятор AirSim забезпечує фотореалістичне відтворення сцен і різноманітних сценаріїв, що дозволяє формувати початкові набори зображень з високим рівнем варіативності та контрольованими параметрами середовища. Це створює передумови для побудови стійких моделей виявлення та розпізнавання об'єктів, які можуть ефективно функціонувати у складних умовах реального світу. Особливу увагу приділено процесу анотації даних. Проведено огляд поширених інструментів анотації з

визначенням їхніх переваг і обмежень в контексті формування наборів даних у форматі YOLO, що є одним із найбільш популярних фреймворків для завдань детекції. Визначено, що узгодженість у роботі аноторів і наявність чітких правил розмітки є критично важливими для забезпечення якості та логічної послідовності наборів даних, особливо у випадках складних сцен із кількома об'єктами, що перекриваються. Окремим аспектом дослідження є застосування методів аугментації, які дозволяють штучно збільшувати обсяг і різноманітність навчальних вибірок. Виконано аналіз базових засобів аугментації, інтегрованих у YOLO, спеціалізованих підходів і автоматизованих методів, що по-різному впливають на узагальнювальну здатність моделей. Використання цих інструментів сприяє зниженню ризику перенавчання, підвищенню точності та стійкості моделей до варіацій освітлення, ракурсів і фонових умов. Практична значущість результатів полягає у створенні універсальної методики, яка дозволяє ефективно поєднувати симульовані та реальні дані, забезпечує швидке формування якісних навчальних вибірок для завдань виявлення об'єктів у різних предметних галузях – від моніторингу довкілля до промислової інспекції.

METHODOLOGY FOR CREATING FORMATTED DATASETS USING THE AIRSIM SIMULATOR AND TOOLS FOR THEIR ANNOTATION AND AUGMENTATION

Sinitsyn I. P.
Doctor of Technical Sciences, Professor,
Corresponding Member of the National Academy of Sciences of Ukraine, Director
Institute of Software Systems of the National Academy of Sciences of Ukraine
Akademika Glushkova ave., 40, Kyiv, Ukraine
orcid.org/0000-0002-4120-0784
ips@nas.gov.ua

Kyrylov I. I.
Postgraduate Student
Institute of Software Systems of the National Academy of Sciences of Ukraine
Akademika Glushkova ave., 40, Kyiv, Ukraine
orcid.org/0009-0006-9756-5391
ivan.kyrylov.science@gmail.com

Zaika K. A.
Postgraduate Student
Institute of Software Systems of the National Academy of Sciences of Ukraine
Akademika Glushkova ave., 40, Kyiv, Ukraine
orcid.org/0009-0006-1107-5106
kir.z2019@gmail.com

Yatsenko O. A.
Ph. D. in Physics and Mathematics, Senior Researcher
Institute of Software Systems of the National Academy of Sciences of Ukraine
Akademika Glushkova ave., 40, Kyiv, Ukraine
orcid.org/0000-0002-4700-6704
oayat@ukr.net

Key words: *data annotation, data augmentation, data generation, data collection, computer vision, dataset labeling for machine learning, annotation rules, AirSim, YOLO.*

A methodology is proposed for creating standardized datasets for computer vision tasks using the Microsoft AirSim simulator and modern tools for data annotation and augmentation. Considering that the preparation of training data for deep learning models is one of the most resource-intensive and time-consuming stages, the developed methodology is aimed at reducing the cost and time of forming high-quality datasets while maintaining their diversity

and consistency. The AirSim simulator provides photorealistic rendering of scenes and diverse scenarios, enabling the generation of initial image sets with a high level of variability and controlled environmental parameters. This creates the foundation for building robust object detection and recognition models that can operate effectively in complex real-world conditions. Special attention is devoted to the data annotation process. The article reviews widely used annotation tools and identifies their advantages and limitations in the context of creating YOLO-compatible datasets, one of the most popular frameworks for object detection tasks. It is emphasized that consistency among annotators and the presence of clear annotation guidelines are critically important for ensuring dataset quality and logical coherence, especially in complex scenes with multiple overlapping objects. Another key aspect of the study is the use of data augmentation methods that artificially increase the size and diversity of training sets. An analysis is provided of YOLO's built-in augmentation mechanisms, specialized techniques, and automated methods, which differently influence the generalization ability of models. The application of these methods helps reduce overfitting, increase accuracy, and improve model robustness to variations in lighting, viewpoint, and background conditions. The practical significance of the results lies in the development of a universal methodology that effectively combines simulated and real-world data, ensuring the rapid creation of high-quality training datasets for object detection tasks across various domains — from environmental monitoring to industrial inspection.

Вступ. Актуальність дослідження зумовлена потребою у створенні якісних і різноманітних наборів даних для завдань комп'ютерного зору, які є основою ефективного навчання моделей глибокого навчання. Традиційний підхід, що передбачає збір і ручну анотацію великих обсягів реальних даних, є надзвичайно дорогим, тривалим і потребує значних людських і технічних ресурсів. Це суттєво обмежує можливості швидкого розвитку застосувань штучного інтелекту в галузях, де необхідні масштабні та якісно розмічені вибірки.

Об'єктом дослідження є процес формування навчальних наборів даних для завдань комп'ютерного зору.

Метою дослідження є розроблення методології створення форматизованих наборів даних на основі симулятора Microsoft AirSim [1; 2] та сучасних інструментів їх анотації і аугментації.

Основна гіпотеза роботи полягає в тому, що використання симуляційних середовищ для генерації початкових зображень у поєднанні з автоматизованими інструментами анотації та методами аугментації дозволить:

– істотно знизити вартість і час формування навчальних даних;

– підвищити різноманітність і якість створених вибірок;

– забезпечити кращу стійкість і узагальнювальну здатність моделей глибокого навчання, зокрема Ultralytics YOLO, у реальних умовах.

Запропонований підхід орієнтований на подолання ключових обмежень традиційних методів підготовки даних та створює передумови для масштабованого використання симульованих і доповнених вибірок у практичних застосуваннях штучного інтелекту.

Для досягнення поставленої мети в роботі необхідно вирішити такі завдання:

1) проаналізувати сучасні підходи до збору, анотації та доповнення даних для навчання моделей глибинного навчання;

2) розробити поетапну методику генерації форматизованих наборів даних на основі симулятора Microsoft AirSim, придатних для використання у фреймворках глибокого навчання, зокрема YOLO;

3) оцінити різні типи анотації зображень відповідно до завдань комп'ютерного зору та визначити оптимальні інструменти для створення якісних розміток;

4) дослідити процес аугментації даних, включаючи базові засоби YOLO, спеціалізовані й автоматизовані методи, для підвищення стійкості й узагальнювальної здатності моделей;

5) сформувати приклад форматизованого набору даних, що включає як первинні, так і доповнені зображення, з подальшою підготовкою конфігураційних файлів для тренування моделей YOLO.

Огляд літератури. Проблематика створення якісних анотованих наборів даних для завдань комп'ютерного зору активно досліджується в сучасних наукових працях. Значна кількість досліджень присвячена розробленню методів збору, анотації та аугментації даних із застосуванням безпілотних літальних апаратів, спеціалізованих інструментів анотації та сучасних моделей глибинного навчання. Особливу увагу приділено формуванню форматизованих наборів даних для навчання моделей типу YOLO, що широко використовуються для виявлення та класифікації об'єктів у зображеннях. У літературі можна виділити низку робіт, які демонструють підходи до побудови таких наборів даних, їх верифікації та застосування в різних предметних галузях.

Зокрема, у дослідженні [3] розглядаються створення та анотація зображень, отриманих із дронів, з метою виявлення стану дерев. Автори ілюструють повний цикл генерації даних – від збору зображень до ручної анотації у CVAT та підготовки їх для тренування YOLO. Запропонований набір даних включає три класи стану дерев. Для оцінювання придатності набору даних проведено серію експериментів із використанням

згорткових нейронних мереж, що показало складність завдання і водночас підтвердило ефективність підходу із глибинними згортками.

Інший підхід продемонстровано в роботі [4], де увага зосереджена на завданні детекції та розпізнавання номерних знаків. Автори сформували набір із 270 зображень, зібраних з інтернету, та виконали їх анотацію у CVAT. Для виконання завдання застосовано комбінацію сучасних методів глибинного навчання (YOLO v8 для локалізації номерних знаків), класичних алгоритмів обробки зображень (кластеризація k-means, порогові методи та морфологічні операції) і OCR для розпізнавання символів. Отримані результати засвідчили високу ефективність такого гібридного підходу, що поєднує різні інструменти для досягнення точності понад 98–99%.

У праці [5] розвинуто напрям автоматизації створення якісних наборів даних, зосереджуючись на детекції низькоінтенсивних викидів диму з димарів. Автори пропонують удосконалений конвеєр формування навчальних даних шляхом інтеграції класифікатора LightGBM у вже наявну систему автоматичної анотації на основі руху. Це дозволило зменшити кількість хибнопозитивних прикладів на 37% і суттєво підвищити якість отриманих розміток. На основі згенерованого набору даних модель YOLOv11 показала високу точність та здатність узагальнювати результати в реальних умовах. Таким чином, дослідження є внеском у розвиток методів автоматизованої верифікації та масштабування наборів даних для завдань моніторингу забруднення довкілля.

Водночас проведений аналіз показує, що в існуючих дослідженнях залишаються обмеження. Зокрема, у роботах [3–5] основний акцент зроблено на ручній або напівавтоматичній анотації, що потребує значних витрат часу та людських ресурсів, а також на формуванні наборів даних для вузькоспеціалізованих предметних областей. Замало уваги приділено використанню симуляційних середовищ для генерації різноманітних навчальних вибірок, узгодженню методів анотації з вимогами форматів сучасних фреймворків (зокрема, YOLO), а також системному застосуванню комплексних методів аугментації для підвищення стійкості моделей. Саме ці аспекти становлять основний фокус даної роботи: поєднання симулятора Microsoft AirSim, сучасних інструментів анотації та багаторівневих стратегій аугментації дозволяє запропонувати універсальну методику формування якісних наборів даних, придатних для широкого спектра завдань комп'ютерного зору.

Методи. З кожним днем кількість різновидів наборів даних, що використовуються для тренування моделей глибокого навчання у сфері комп'ютерного зору, невпинно зростає. Проте основною проблемою залишається отримання бажаного обсягу якісних і чітко розмічених даних, які відповідали б конкретним завданням дослідження. У багатьох випадках збір реальних зображень є складним або взагалі неможливим через високу вартість, часові витрати чи обмеженість доступу до необхідних об'єктів і середовищ.

Microsoft AirSim [1; 2] – авіасимулятор, який використовується для навчання та тестування автономних систем (наприклад, безпілотних апаратів зі штучним інтелектом, систем навігації, систем комп'ютерного зору). Цей симулятор є фотореалістичним, що дає можливість використовувати зображення, отримані з нього, для створення якісних форматизованих наборів даних і подальшої їх обробки. Завдяки підтримці різних сценаріїв, параметрів середовища та можливості гнучкого налаштування траєкторій руху AirSim дозволяє швидко отримувати велику кількість варіативних прикладів для подальшого використання у тренуванні моделей.

Він має багато переваг у процесі відтворення реалістичних сцен, серед яких: висока деталізація та реалістичність графіки, можливість масштабування експериментів, гнучкість у зміні умов освітлення та погодних сценаріїв, а також підтримка різних типів сенсорів (RGB-камери, глибинні камери, LiDAR). Це робить AirSim універсальним інструментом для симуляції та збору даних у контрольованих умовах, які близькі до реальних.

Форматизований набір даних, у нашому випадку – набір зображень, згенерованих за допомогою симулятора, які надалі підлягають анотації відповідно до спеціально розроблених інструкцій і правил. Наявність чітко визначених правил для анотації забезпечує єдність та узгодженість позначення цільових об'єктів. Наприклад, якщо об'єктів декілька або вони частково перекривають один одного, методика виділення має бути детально описана. Це дозволяє в майбутньому залучати до роботи нових анотаторів, які керуватимуться стандартами, що вже існують, тим самим забезпечувати цілісність та якість набору даних.

Анотація даних – це процес маркування даних для подальшого навчання моделей [6]. У комп'ютерному зорі цей процес означає розмітку зображень або відео, на основі якої нейронні мережі формують зв'язки між вхідними та вихідними даними. Без належним чином анотованих вибірок моделі не здатні точно навчатися та робити узагальнені висновки. Отже, якість анотацій є критично важливим складником успішного машинного навчання, оскільки вона безпосередньо впливає на продуктивність та точність моделей.

Процес анотації зазвичай охоплює кілька етапів: від підготовки та збору якісних вихідних

даних, їх завантаження до інструментів для розмітки, безпосередньої ручної чи автоматизованої анотації, перевірки на відповідність інструкціям і правилам, до фінального експорту розміченого набору даних у форматах, сумісних із сучасними фреймворками глибокого навчання. На кожному із цих етапів важливо забезпечити узгодженість та контроль якості, щоб уникнути систематичних помилок, які можуть негативно позначитися на навчанні моделі.

Залежно від конкретних вимог і завдань комп'ютерного зору застосовуються різні типи анотації даних (рис. 1). Основні з них:

– обмежувальні прямокутники (bounding boxes): прямокутні рамки, накреслені навколо цільових об'єктів на зображенні, що застосовуються переважно в завданнях виявлення об'єктів. Вони визначаються координатами верхнього лівого та нижнього правого кутів;

– багатокутники (polygons): детальні контури об'єктів, що забезпечують точнішу локалізацію, ніж bounding boxes. Використовуються в завданнях сегментації зображень, де важливі форма та точні межі об'єкта;

– маски (masks): двійкові чи багатокласові матриці, у яких кожен піксель позначений як належний окремому об'єкту чи фону. Маски застосовуються в завданнях семантичної та інстанс-сегментації, забезпечують найвищий рівень деталізації;

– ключові точки (keypoints): специфічні точки на об'єктах, що відмічають критично важливі зони, наприклад суглоби людини чи орієнтири обличчя. Використовуються в завданнях оцінювання пози, трекінга руху та біометричної ідентифікації.

Таким чином, вибір типу анотації напряму залежить від поставленого завдання: чим вищі вимоги до точності локалізації та форми об'єкта, тим складніші методи розмітки треба застосовувати.

Найпоширенішими застосунками для створення наборів даних, які відповідають вимогам формату фреймворку YOLO, є Ultralytics Annotation Tool [7], Roboflow [8], CVAT [9; 10] та LabelImg [11], що стисло розглядаються далі.

Застосунок *Ultralytics* надає інструмент auto_annotate, що автоматично створює анотації у форматі YOLO за допомогою попередньо натренованої моделі детекції (YOLO) разом із SAM (Segment Anything Model) для генерації сегментаційних масок. Цей підхід значно прискорює підготовку даних і зручно інтегрується в pipeline Ultralytics YOLO.

Roboflow пропонує комплексну платформу для анотації даних із підтримкою AI-асистенції. Вона дозволяє швидко створювати bounding boxes і полігони з використанням функцій Label Assist, Smart Polygon та Auto Label, що застосовують сучасні моделі для автоматизації розмітки. Окрім того, інструмент підтримує попередню обробку даних, автоматичне виявлення аномалій і генерацію до 50 варіантів аугментованих зображень на один базовий кадр.

Платформа CVAT (Computer Vision Annotation Tool) – це потужний вебінструмент з відкритим кодом, який підтримує аннотацію зображень і відео для завдань машинного навчання:

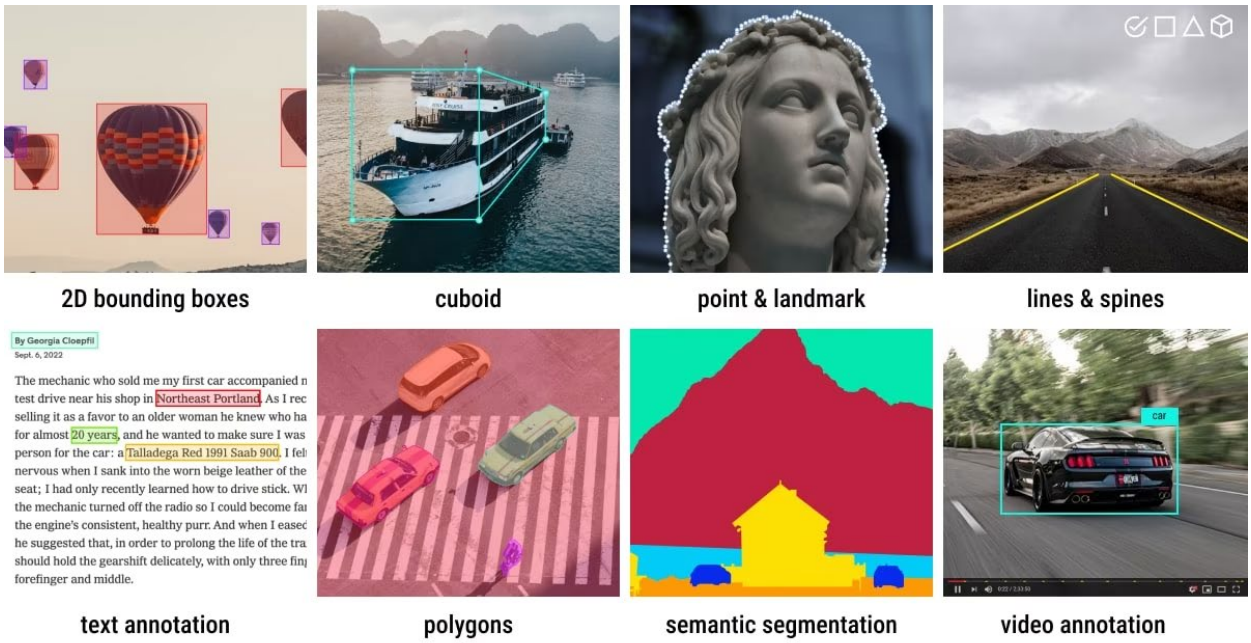


Рис. 1. Різновиди анотації даних [6]

виявлення об’єктів, класифікації, сегментації та інших. Він пропонує широкий функціонал, включаючи напівавтоматичну анотацію, інтерполяцію між кадрами, API, SDK та можливість колаборації (локальні й онлайн-версії).

LabelImg – простий графічний інструмент на Python із UI на базі Qt, призначений для створення bounding box-розмітки. Анотації можна зберігати у форматах Pascal VOC, YOLO або CreateML. Це легкий і швидкий спосіб для невеликих проєктів, навчання або прототипування, особливо коли потрібні простий інтерфейс і мінімальні налаштування.

Після збору й анотації даних наступним важливим етапом підготовки набору даних для навчання моделей комп’ютерного зору є їх аугментація. Аугментація (доповнення) даних [12] – це ключова методика, що штучно розширює навчальний набір, застосовує різні перетворення до наявних зображень. Вона безпосередньо доповнює підготовлені набори даних, підвищує їхню різноманітність і якість, що є критично важливим для тренування моделей глибокого навчання, як-от Ultralytics YOLO.

Доповнення даних реалізує кілька критичних завдань у навчанні моделей комп’ютерного зору:

– розширення набору даних: створенням варіацій наявних зображень можна ефективно збільшити розмір навчального набору без додаткового збору нових даних;

– покращене узагальнення: моделі вчаться розпізнавати об’єкти за різних умов, що робить їх більш стійкими в реальних сценаріях;

– зменшення перенавчання: завдяки введенню варіативності в навчальні дані моделі менш схильні до запам’ятовування специфічних характеристик окремих зображень;

– покращена продуктивність: моделі, навчені з належним доповненням, зазвичай демонструють вищу точність на валідаційних і тестових наборах.

Реалізація Ultralytics YOLO [7] надає комплексний набір методів аугментації, кожен із яких сприяє продуктивності моделі по-різному. У [12] детально розглянуто параметри аугментації та наведено рекомендації щодо їх ефективного використання у проєктах. Кожен параметр можна налаштувати за допомогою Python API, інтерфейсу командного рядка або файлу конфігурації.

Внутрішні базові засоби аугментації YOLO включають:

– RandomHSV – випадкове налаштування каналів відтінку, насиченості та значення зображення (Hue, Saturation, Value; останнє також називають яскравістю);

– RandomPerspective – випадкові перспективні й афінні перетворення зображень та відповідних анотацій (обертання, переміщення, масштабування, нахил);

– RandomFlip – випадкове горизонтальне або вертикальне віддзеркалення зображення із заданою імовірністю;

– Format – форматування анотацій зображень для завдань виявлення об’єктів, сегментації екземплярів і оцінювання пози;

– Mosaic – об’єднання кількох зображень (4 або 9) в одне мозаїчне зображення з певною імовірністю для збільшення різноманітності набору даних;

– Mixup – поєднання двох зображень та їхніх міток із випадковим ваговим коефіцієнтом;

– CutMix – об’єднання двох зображень шляхом заміни випадкової прямокутної області одного зображення областю іншого, із пропорційним оновленням міток;

– CopyPaste – вставка об’єктів з різних зображень для створення нових навчальних зразків;

– AutoAugment – автоматизоване застосування стратегій доповнення для класифікації;

– Random Erasing – випадкове стирання частин зображення під час навчання.

До спеціалізованих методів аугментації належать:

– Random Erasing – вищезгаданий метод випадковим чином обирає прямокутну область на зображенні та стирає її пікселі випадковими значеннями [13]. Водночас генеруються навчальні зображення з різними рівнями оклюзії, що зменшує ризик перенавчання та робить модель стійкою до часткового перекриття об’єктів. Хоча випадкове стирання є простим, воно доповнює поширені методи аугментації даних, як-от випадкове кадрування та перевертання, і забезпечує стабільне покращення порівняно з базовими рівнями у класифікації зображень, виявленні об’єктів і повторній ідентифікації осіб;

– Hide-and-Seek [14] – доповнює існуючі методи аугментації та є корисним для різних завдань візуального розпізнавання. Ключова ідея полягає у випадковому приховуванні ділянок на навчальному зображенні, щоб змусити мережу шукати інший релевантний контент, коли найбільш розпізнавальний контент прихований. Головною перевагою над іншими методами є здатність покращувати точність локалізації об’єктів у слабо контрольованих умовах;

– GridMask [15] – структуроване видалення областей зображення для підвищення ефективності моделей у різних завданнях комп’ютерного зору;

– Copy-Paste [16] – метод для сегментації екземплярів, що вставляє об’єкти з інших зображень для створення нових навчальних зразків.

Автоматизовані методи включають:

– AutoAugment [17] – автоматичний пошук оптимальних стратегій доповнення даних;

– RandAugment [18] – зменшений простір пошуку й адаптивна сила регуляризації, що дозволяє навчати політику на цільовому наборі даних;

– AugMix [19] – метод обробки даних із низькими обчислювальними витратами, що підвищує стійкість моделей до пошкоджень та непередбачених змін у зображеннях;

– TrivialAugment [20] – простий базовий метод без параметрів, що застосовує одне випадкове збільшення до кожного зображення.

Методи аугментації на основі дифузійних моделей досліджені у [21].

Результати. Для виконання анотації даних у цьому дослідженні було обрано інструмент CVAT [10] (рис. 2). Цей веборієнтований інструмент з відкритим кодом забезпечує широкий спектр можливостей для створення розмітки, необхідної для завдань комп'ютерного зору, зокрема для виявлення об'єктів, сегментації та класифікації.

Загальний алгоритм роботи із CVAT включає такі кроки:

1) реєстрацію користувача на платформі CVAT [10];

2) створення нового завдання (task);

3) визначення необхідних класів (міток) для анотації;

4) завантаження даних для розмітки;

5) перехід до розділу "Job" та виконання анотації зображень;

6) експорт розміченого набору даних у потрібному форматі (зокрема, YOLO).

У процесі анотації особливу увагу приділяють узгодженості результатів між різними анотаторами, які незалежно працюють з однаковою колекцією зображень.

Це критично важливо для завдань, що передбачають елемент суб'єктивної інтерпретації, як-от сегментація, виявлення об'єктів і класифікація зображень.

Схему робочого процесу маркування даних у CVAT наведено на рисунку 3.

Необхідно забезпечити різноманітність зібраних зображень. Для цього сцени в симуляторі відтворюються під різними кутами, за різного освітлення та на різному фоні. Чим різноманітніший і якісніший набір даних, тим точніше й стійкіше модель виконуватиме свої завдання.

Розглянемо детальніше структуру самого набору даних і відповідних текстових файлів розмітки. Головна папка набору даних має містити дві підпапки: images та labels (рис. 4).

Концепція YOLO передбачає, що для кожного зображення існує відповідний текстовий файл (.txt) з тією ж назвою. Наприклад, для image1.jpg повинен існувати файл image1.txt. Файл розмітки є структурним елементом набору даних.

Кожен файл .txt містить один рядок для кожного об'єкта на зображенні в такому форматі:

```
<class_id> <x_center> <y_center> <width> <height>.
```

Тут class_id – ціле число, що позначає клас об'єкта (нумерація починається з 0). Координати та розміри об'єктів нормалізовані, тобто їхні значення лежать у діапазоні від 0 до 1.

Наприклад, для зображення розміром 800 × 600 пікселів, якщо цільовий об'єкт має координати x = 200, y = 150, ширину w = 100 і висоту h = 120, нормалізовані значення обчислюються так:

Рис. 2. Інтерфейс вебінструменту анотації зображень CVAT

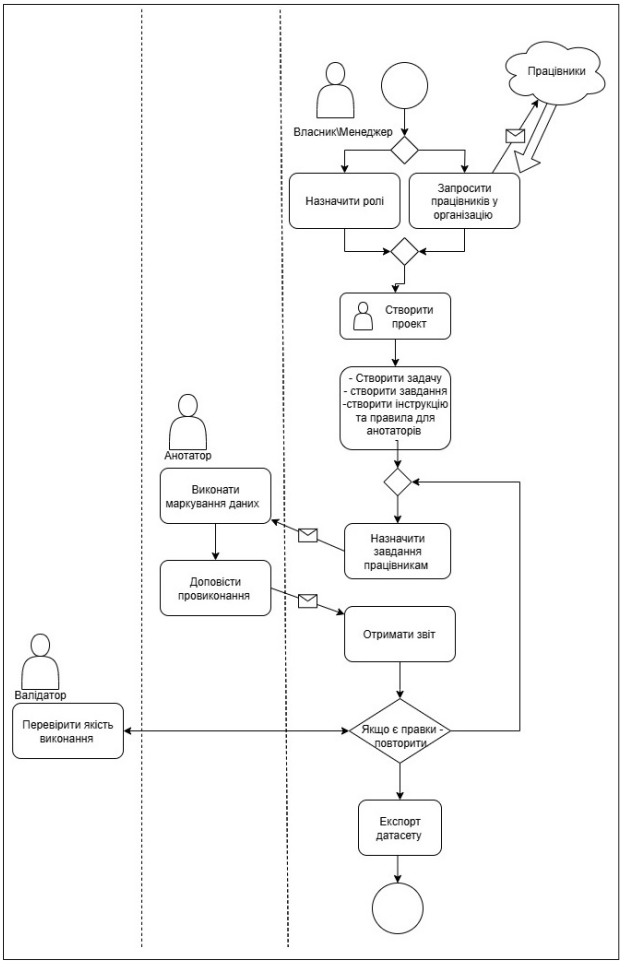


Рис. 3. Схема пропонованого процесу анотації у CVAT

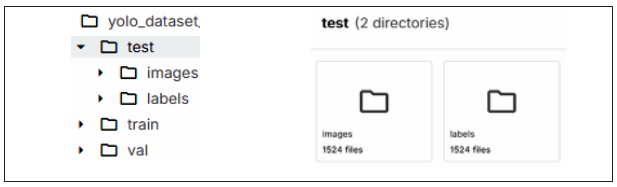


Рис. 4. Структура папок з набором даних

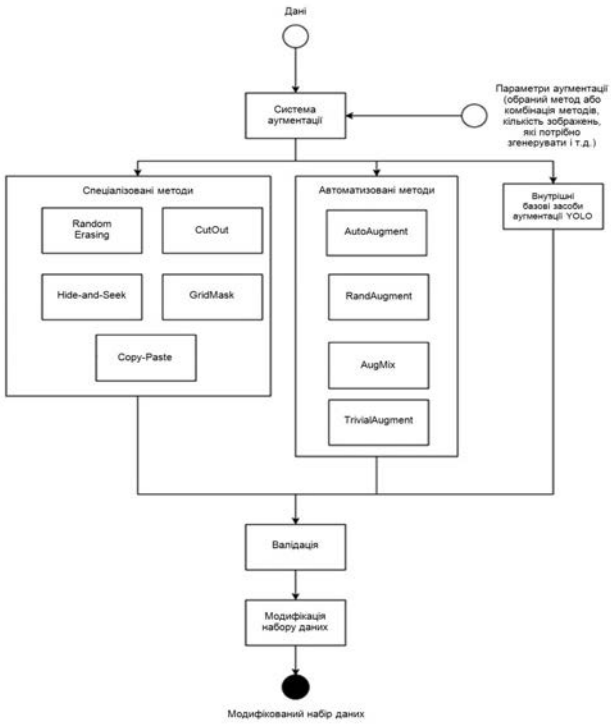


Рис. 5. Процес аугментації даних

x_center = (200 + 100/2) / 800 = 0,3125;
y_center = (150 + 120/2) / 600 = 0,35;
width = 100 / 800 = 0,125;
height = 120 / 600 = 0,2.

Для взаємодії з набором даних YOLO використовує файл конфігурації у форматі .yaml, зазвичай з назвою dataset.yaml, який розміщується в головній папці проєкту (або в директорії запуску тренування). Приклад вмісту файлу:

```
# Шляхи до тренувального та валідаційного наборів
train: ../yolo_dataset_train/images_train
val: ../yolo_dataset_val/images_val
# Опціонально, для тестування
# test: ../yolo_dataset_test/images_test
nc: 3 # Кількість класів

# Назви класів (у порядку їх ID, починаючи з 0)
# names:
0: window
1: door
2: plane.
```

Фінальна структура набору даних зазвичай передбачає розділення даних на тренувальні (train), валідаційні (val) та, за бажанням, тестові (test) набори. Найпоширеніші співвідношення – 70/20/10 або 80/20.

Процес створення завдання анотації, роботу з інтерфейсом анотації, власне анотацію та експорт набору даних детально описано в [10].

Процес аугментації даних зображено на рисунку 5.

Приклади застосування методів аугментації RandomHSV та RandomErasing, згаданих у попередньому розділі, наведені на рисунках 6 і 7.

а *б*

Рис. 6. Приклад застосування методу RandomHSV (випадкове налаштування відтінку, насиченості та яскравості зображення): *а* – початкове зображення, *б* – результат обробки

Рис. 7. Приклад застосування методу RandomErasing (випадкове стирання частини зображення)

Висновки. У даній роботі представлено методологію створення форматизованих наборів даних із використанням симулятора Microsoft AirSim, інструментів їх анотації і аугментації. Запропонований підхід дозволяє істотно знизити витрати часу й ресурсів на формування навчальних вибірок, які традиційно потребують масштабного збору та ручної розмітки реальних даних. Використання симулятора забезпечує можливість відтворення різноманітних сценаріїв і отримання фотореалістичних зображень, що підвищує різноманітність і якість даних.

Розглянуто сучасні інструменти анотації, серед яких CVAT, LabelImg, Roboflow та Ultralytics Annotation Tool, а також їхні переваги й обмеження. Показано, що дотримання правил анотації та узгодженість дій аноторів є критично важливими для формування якісних наборів даних. Особливу увагу приділено методам аугментації даних: від базових засобів YOLO (RandomHSV, Mosaic, MixUp тощо) до спеціалізованих і автоматизованих підходів (GridMask, AutoAugment, RandAugment, AugMix), які суттєво підвищують стійкість моделей до варіацій у реальних умовах.

Практична значущість дослідження полягає у створенні універсального підходу до швидкої генерації, розмітки та доповнення навчальних наборів, що може застосовуватися в широкому спектрі завдань комп'ютерного зору – від детекції об'єктів до семантичної сегментації.

Перспективи подальших досліджень передбачають:

– розширення експериментів із використанням інших фреймворків глибокого навчання, окрім YOLO;

– дослідження ефективності застосування дифузійних моделей для аугментації даних;

– інтеграцію симульованих даних із реальними зображеннями для створення гібридних наборів, які забезпечують ще вищу узагальнювальну здатність моделей;

– розроблення автоматизованих систем контролю якості анотацій і синтетично згенерованих даних.

Отже, представлена методика може стати ефективною альтернативою або доповненням до традиційного збору реальних даних, а її подальший розвиток сприятиме підвищенню якості та масштабованості досліджень у сфері комп'ютерного зору.

ЛІТЕРАТУРА

1. Shah S., Dey D., Lovett C., Kapoor A. AirSim: high-fidelity visual and physical simulation for autonomous vehicles. *Field and Service Robotics. Springer Proceedings in Advanced Robotics*. 2018. Vol. 5. P. 621–635. https://doi.org/10.1007/978-3-319-67361-5_40
2. Microsoft AirSim. URL: https://github.com/Microsoft/AirSim/releases
3. Karumanchi Y., Prasanna G. L., Mukherjee S., Kolagani N. Plantation monitoring using drone images: a dataset and performance review. *arXiv:2502.08233*. 2025. P. 1–7. https://doi.org/10.48550/arXiv.2502.08233
4. Moussaoui H., Akkad N. E., Benslimane M., El-Shafai W., Baihan A., Hewage C., Rathore R. S. Enhancing automated vehicle identification by integrating YOLO v8 and OCR techniques for high-precision license plate detection and recognition. *Scientific Reports*. Vol. 14. 2024. P. 1–17. https://doi.org/10.1038/s41598-024-65272-1
5. Szczepański M. Improving UAV-based detection of low-emission smoke with an advanced dataset generation pipeline. *Remote Sensing*. Vol. 17. 2025. P. 1–33. https://doi.org/10.3390/rs17061004
6. Data collection and annotation strategies for computer vision. URL: https://docs.ultralytics.com/guides/data-collection-and-annotation
7. Ultralytics YOLO Docs. URL: https://docs.ultralytics.com/reference/data/annotator
8. Roboflow. URL: https://roboflow.com/annotate
9. Best Open-Source Image Annotation Tools in 2024. Computer Vision Annotation Tool (CVAT). URL: https://www.cvat.ai/resources/blog/best-open-source-image-annotation-tools#computer-vision-annotation-tool-cvat
10. CVAT. Leading Data Annotation Platform. URL: https://docs.cvat.ai/docs/manual
11. LabelImg for Image Annotation. URL: https://viso.ai/computer-vision/labelimg-for-image-annotation/

12. Data augmentation using Ultralytics YOLO. URL: https://docs.ultralytics.com/guides/yolo-data-augmentation
13. Zhong Z., Zheng L., Kang G., Li S., Yang Y. Random erasing data augmentation. *arXiv:1708.04896*. 2017. P. 1–10. https://doi.org/ 10.48550/arXiv.1708.04896
14. Singh K. K., Yu H., Sarmasi A., Pradeep G., Lee Y. J. Hide-and-Seek: a data augmentation technique for weakly-supervised localization and beyond. *arXiv:1811.02545*. 2018. P. 1–14. https://doi.org/ 10.48550/arXiv.1811.02545
15. Chen P., Liu S., Zhao H., Wang X., Jia J. GridMask data augmentation. *arXiv:2001.04086*. 2024. P. 1–9. https://doi.org/10.48550/arXiv.2001.04086
16. Ghiasi G., Cui Y., Srinivas A., Qian R., Lin T.-Y., Cubuk E. D., Le Q. V., Zoph B. Simple copy-paste is a strong data augmentation method for instance segmentation. *arXiv:2012.07177*. 2021. P. 1–13. https://doi.org/10.48550/arXiv.2012.07177
17. Cubuk E. D., Zoph B., Mane D., Vasudevan V., Le Q. V. AutoAugment: learning augmentation policies from data. *arXiv:1805.09501*. 2018. P. 1–14. https://doi.org/10.48550/arXiv.1805.09501
18. Cubuk E. D., Zoph B., Shlens J., Le Q. V. RandAugment: practical automated data augmentation with a reduced search space. *arXiv:1909.13719.* 2019. P. 1–14. https://doi.org/10.48550/arXiv.1909.13719
19. Hendrycks D., Mu N., Cubuk E. D., Zoph B., Gilmer J., Lakshminarayanan B. AugMix: a simple data processing method to improve robustness and uncertainty. *arXiv:1912.02781*. 2019. P. 1–15. https://doi.org/10.48550/arXiv.1912.02781
20. Müller S. G., Hutter F. TrivialAugment: tuning-free yet state-of-the-art data augmentation. *arXiv:2103.10158*. 2021. P. 1–13. https://doi.org/10.48550/arXiv.2103.10158
21. Fang H., Han B., Zhang S., Zhou S., xHu S., Ye W.-M. Data augmentation for object detection via controllable diffusion models. *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 2024. P. 1246–1255. https://doi.org/10.1109/WACV57701.2024.00129

REFERENCES

1. Shah, S., Dey, D., Lovett, C., Kapoor, A. (2018). AirSim: high-fidelity visual and physical simulation for autonomous vehicles. *Field and Service Robotics. Springer Proceedings in Advanced Robotics* 5, 621–635. https://doi.org/10.1007/978-3-319-67361-5_40
2. Microsoft AirSim. https://github.com/Microsoft/AirSim/releases
3. Karumanchi, Y., Prasanna, G. L., Mukherjee, S., Kolagani, N. (2025). Plantation monitoring using drone images: a dataset and performance review. *arXiv:2502.08233*, 1–7. https://doi.org/10.48550/arXiv.2502.08233
4. Moussaoui, H., Akkad, N. E., Benslimane, M., El-Shafai, W., Baihan, A., Hewage, C., Rathore, R. S. (2024). Enhancing automated vehicle identification by integrating YOLO v8 and OCR techniques for high-precision license plate detection and recognition. *Scientific Reports* 14, 1–17. https://doi.org/10.1038/s41598-024-65272-1.
5. Szczepański, M. (2025). Improving UAV-based detection of low-emission smoke with an advanced dataset generation pipeline. *Remote Sensing* 17, 1–33. https://doi.org/10.3390/rs17061004
6. Data collection and annotation strategies for computer vision. https://docs.ultralytics.com/guides/data-collection-and-annotation
7. Ultralytics YOLO Docs. https://docs.ultralytics.com/reference/data/annotator
8. Roboflow. https://roboflow.com/annotate
9. Best Open-Source Image Annotation Tools in 2024. Computer Vision Annotation Tool (CVAT). https://www.cvat.ai/resources/blog/best-open-source-image-annotation-tools#computer-vision-annotation-tool-cvat
10. CVAT. Leading Data Annotation Platform. https://docs.cvat.ai/docs/manual
11. LabelImg for Image Annotation. https://viso.ai/computer-vision/labelimg-for-image-annotation/
12. Data augmentation using Ultralytics YOLO. https://docs.ultralytics.com/guides/yolo-data-augmentation
13. Zhong, Z., Zheng, L., Kang, G., Li, S., Yang, Y. (2017). Random erasing data augmentation. *arXiv:1708.04896*, 1–10. https://doi.org/ 10.48550/arXiv.1708.04896
14. Singh, K. K., Yu, H., Sarmasi, A., Pradeep, G., Lee, Y. J. (2018). Hide-and-Seek: a data augmentation technique for weakly-supervised localization and beyond. *arXiv:1811.02545*, 1–14. https://doi.org/ 10.48550/arXiv.1811.02545
15. Chen, P., Liu, S., Zhao, H., Wang, X., Jia, J. (2024). GridMask data augmentation. *arXiv:2001.04086*, 1–9. https://doi.org/10.48550/arXiv.2001.04086

16. Ghiasi, G., Cui, Y., Srinivas, A., Qian, R., Lin, T.-Y., Cubuk, E. D., Le, Q. V., Zoph, B. (2021). Simple copy-paste is a strong data augmentation method for instance segmentation. *arXiv:2012.07177*, 1–13. https://doi.org/10.48550/arXiv.2012.07177
17. Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., Le, Q. V. (2018). AutoAugment: learning augmentation policies from data. *arXiv:1805.09501*, 1–14. https://doi.org/10.48550/arXiv.1805.09501
18. Cubuk, E. D., Zoph, B., Shlens, J., Le, Q. V. (2019). RandAugment: practical automated data augmentation with a reduced search space. *arXiv:1909.13719*, 1–14. https://doi.org/10.48550/arXiv.1909.13719
19. Hendrycks, D., Mu, N., Cubuk, E. D., Zoph, B., Gilmer, J., Lakshminarayanan, B. (2019). AugMix: a simple data processing method to improve robustness and uncertainty. *arXiv:1912.02781*, 1–15. https://doi.org/10.48550/arXiv.1912.02781
20. Müller, S. G., Hutter, F. (2021). TrivialAugment: tuning-free yet state-of-the-art data augmentation. *arXiv:2103.10158*, 1–13. https://doi.org/10.48550/arXiv.2103.10158
21. Fang, H., Han, B., Zhang, S., Zhou, S., Hu, S., Ye, W.-M. (2024). Data augmentation for object detection via controllable diffusion models. *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, P. 1246–1255. https://doi.org/10.1109/WACV57701.2024.00129

Дата першого надходження рукопису до видання: 25.08.2025
Дата прийнятого до друку рукопису після рецензування: 22.09.2025
Дата публікації: 31.12.2025

УДК 004.056.53:004.75
DOI https://doi.org/10.26661/2786-6254-2025-2-14

ПОБУДОВА АДАПТИВНИХ ПОРОГОВИХ СТРАТЕГІЙ ДЛЯ ВИЯВЛЕННЯ ГІБРИДНИХ АТАК У МЕРЕЖАХ ЕЛЕКТРОННИХ КОМУНІКАЦІЙ ІЗ ЗАСТОСУВАННЯМ ПОТОКОВИХ МЕТОДІВ І КАЛІБРУВАННЯ ПІКІВ

Флоров С. В.
кандидат технічних наук,
доцент кафедри кібербезпеки та інформаційних технологій
Університет митної справи та фінансів
вул. Володимира Вернадського, 2/4, Дніпро, Україна
orcid.org/0000-0002-4682-7666
pre.pod@hotmsil.com

Черкаський О. В.
незалежний дослідник у сфері кібербезпеки
Університет митної справи та фінансів
вул. Володимира Вернадського, 2/4, Дніпро, Україна
orcid.org/0009-0006-3105-5217
asherjoseph.c@gmail.com

Черкаський Д. О.
незалежний дослідник у сфері кібербезпеки
Університет митної справи та фінансів
вул. Володимира Вернадського, 2/4, Дніпро, Україна
orcid.org/0009-0003-8516-6252
cherkaskyidavid12@gmail.com

Білан М. В.
незалежний дослідник у сфері кібербезпеки
Університет митної справи та фінансів
вул. Володимира Вернадського, 2/4, Дніпро, Україна
orcid.org/0009-0005-5875-4950
bilan_maksym@umsf.dp.ua

Переметчик Д. О.
незалежний дослідник у сфері кібербезпеки
Університет митної справи та фінансів
вул. Володимира Вернадського, 2/4, Дніпро, Україна
orcid.org/0009-0006-1978-5858
epereметchyk.d@gmail.com

Ключові слова: *адаптивні порогові механізми, потокова обробка даних, статистичні моделі, виявлення аномалій, мережевий трафік, архітектура нульової довіри, системи безпеки.*

Метою статті є створення адаптивних порогових механізмів для систем виявлення вторгнень, які здатні залишатися чутливими до рідкісних, але критичних подій у мережах електронних комунікацій і водночас зменшувати кількість хибних спрацювань у складних умовах, коли дані мають «важкі хвости» розподілу та змінюють свої характеристики із часом. Запропоновано підхід, що поєднує сучасні методи оцінювання високих квантилів у потоці даних, робастні статистичні показники

масштабу та спеціальні моделі для обробки рідкісних піків. Для контролю змін у даних використано алгоритми, які автоматично виявляють і реагують на дрейф, не зупиняючи процес обробки. Розроблені механізми впроваджуються в конвеєр обробки подій безпеки. Вони працюють з різними видами ознак: байтові послідовності перетворюються на зображення, з яких згорткові нейронні мережі виділяють просторові закономірності, а рекурентні нейронні мережі аналізують часові залежності. Альтернативно застосовується автоенкодер для виявлення відхилень через похибку відновлення даних. Отриманий аномальний бал поєднує два джерела інформації: відхилення від нормальної моделі та статистичне перевищення порога. Це дозволяє зменшити вплив сезонних змін та випадкових коливань у трафіку. Експерименти на відкритих наборах даних показали, що розроблені методи зменшують кількість хибнопозитивних спрацювань майже на третину та підвищують точність виявлення атак на сканування, розподілені відмови в обслуговуванні й прихований витік даних. Стійкість алгоритмів зберігається навіть за наявності складних аномалій та під час атак на захищені протоколи й мережеве обладнання. Запропонований підхід відповідає принципам архітектури «нульової довіри», адже дозволяє ухвалювати локальні рішення про доступ та реагування на загрози у вузлах мережі. Практичні висновки включають рекомендації щодо розміру структур для потокової обробки, вибору рівня квантиля для оцінювання ризиків та інтеграції із правилами керування подіями безпеки з метою швидшого аналізу інцидентів.

CONSTRUCTION OF ADAPTIVE THRESHOLD STRATEGIES FOR DETECTING HYBRID ATTACKS IN ELECTRONIC COMMUNICATION NETWORKS USING STREAMING METHODS AND PEAK CALIBRATION

Florov S. V.
PhD in Technical Sciences,
Associate Professor at the Department of Cybersecurity and Information Technologies
University of Customs and Finance
Volodymyr Vernadsky str., 2/4, Dnipro, Ukraine
orcid.org/0000-0002-4682-7666
pre.pod@hotmsil.com

Cherkaskyi O. V.
Independent Researcher in Cybersecurity
University of Customs and Finance
Volodymyr Vernadsky str., 2/4, Dnipro, Ukraine
orcid.org/0009-0006-3105-5217
asherjoseph.c@gmail.com

Cherkaskyi D. O.
Independent Researcher in Cybersecurity
University of Customs and Finance
Volodymyr Vernadsky str., 2/4, Dnipro, Ukraine
orcid.org/0009-0003-8516-6252
cherkaskyidavid12@gmail.com

Bilan M. V.
Independent Researcher in Cybersecurity
University of Customs and Finance
Volodymyr Vernadsky str., 2/4, Dnipro, Ukraine
orcid.org/0009-0005-5875-4950
bilan_maksym@umsf.dp.ua

Peremetchyk D. O.
Independent Researcher in Cybersecurity
University of Customs and Finance
Volodymyr Vernadsky str., 2/4, Dnipro, Ukraine
orcid.org/0009-0006-1978-5858
peremetchyk.d@gmail.com

Key words: *adaptive threshold mechanisms, data stream processing, statistical models, anomaly detection, network traffic, zero-trust architecture, security systems.*

The purpose of the article is to develop adaptive threshold mechanisms for intrusion detection systems that can remain sensitive to rare but critical events in electronic communication networks, while at the same time reducing the number of false alarms under complex conditions where data have heavy-tailed distributions and their characteristics evolve over time. The proposed approach combines modern methods for estimating high quantiles in data streams, robust statistical scale measures, and dedicated models for handling rare peaks. To monitor changes in data, algorithms are applied that automatically detect and respond to drift without interrupting the processing pipeline. The developed mechanisms are integrated into the security event processing pipeline. They work with different feature types: byte sequences are transformed into images, from which convolutional neural networks extract spatial patterns, while recurrent neural networks analyze temporal dependencies. Alternatively, an autoencoder is used to detect anomalies through reconstruction error. The resulting anomaly score combines two sources of information: deviation from the normal model and statistical threshold exceedance. This reduces the influence of seasonal variations and random fluctuations in traffic. Experiments on open datasets demonstrated that the proposed methods reduce false positives by nearly one-third and improve the accuracy of detecting scanning attacks, distributed denial-of-service incidents, and covert data exfiltration. The robustness of the algorithms is preserved even under complex anomalies and during attacks on protected protocols and network equipment. The proposed approach aligns with the principles of zero-trust architecture, enabling local decisions on access control and threat response at network nodes. Practical implications include recommendations for the size of data structures for streaming processing, selection of quantile levels for risk assessment, and integration with security event management rules for faster incident analysis.

Вступ. Сучасні мережі електронних комунікацій характеризуються високою варіативністю трафіку, наявністю «важких хвостів» у розподілах розмірів потоків і затримок, а також неперервними змінами поведінки користувачів і сервісів. За цих умов статичні пороги в системах виявлення вторгнень (Intrusion Detection System (далі – IDS)) породжують надлишок хибних спрацьовувань (false positives) або, навпаки, пропускають малопомітні атаки з низькою інтенсивністю. Загрози поглиблюються гібридними сценаріями зловмисників, що комбінують уразливості протоколів шифрування Secure Sockets Layer/Transport Layer Security (далі – SSL/TLS) на рівні прошивок (firmware) мережевого обладнання з маніпуляціями Simple Network Management Protocol (далі – SNMP) та ланцюжками MITRE ATT&CK для прихованого керування і ексфільтрації даних [2; 3; 24; 26]. У відповідь провайдери та корпоративні мережі впроваджують архітектуру Zero Trust, де перевірка здійснюється під час кожної взаємодії, а рішення базуються на локальному контексті й верифікованих ознаках ризику [1]. Комплексне керування подіями безпеки (Security Information and Event Management (далі – SIEM)) потребує алгоритмів, що дають числові оцінки ризику в реальному часі з відтворюваними гарантіями точності для подальшого кореляційного аналізу [25].

Стримінгові квантильні ескізи – компактні структури даних для наближеної оцінки квантилів у потоках – відкрили можливість динамічної, статистично обґрунтованої порогової логіки. Узагальнюючи ідеї Greenwald – Khanna (далі – GK), t-digest та KLL, ми використовуємо наближені високі квантилі для автоматичного встановлення порогів у вузлах IDS без зупинки трафіку [4–7]. Для стійкості до «важких хвостів» і атак із рідкісними, але руйнівними піками ми поєднуємо квантильні пороги з калібруванням методом піків понад поріг (Peaks-Over-Threshold (далі – POT)) з теорії екстремальних значень (Extreme Value Theory (далі – EVT)), а для управління дрейфом – адаптивні схеми вікон на основі Page – Hinkley і ADWIN [10–13]. Робастні оцінки масштабу (медіанна абсолютна похибка, MAD; робастні експоненційні ковзні моменти) додатково знижують чутливість до викидів [16; 17].

Мета роботи – розробити й експериментально обґрунтувати адаптивні порогові механізми для IDS, які: (1) працюють у потоковому режимі з обмеженою пам’яттю; (2) є стійкими до дрейфу та сезонності; (3) коректно поводяться за важких хвостів; (4) інтегруються із глибинними представленнями ознак (Byte2Image, згорткові нейромережі, довга короткочасна пам’ять) і компонентами SIEM у контексті Zero Trust. Додатково демонструємо приклад інтеграції з MATLAB Mobile

для оперативної валідації ідей на польових даних. Для аналізу використано загальнодоступні набори CIC-IDS2017 і UNSW-NB15 [18; 19], а також нормативні орієнтири із TLS 1.3, SNMPv3 та рекомендації з резильєнтності прошивок [2; 3; 26; 27].

Огляд літератури. Останні п'ять років проблема виявлення гібридних атак у мережах електронних комунікацій активно досліджується в кількох напрямах. Значна увага приділяється потоковим алгоритмам для оцінювання квантилів і робастним статистичним мірам, що дозволяють адаптивно встановлювати пороги в умовах великих даних і «важких хвостів» розподілу. Алгоритми t-digest і KLL стали основою сучасних рішень [1], хоча їхня ефективність обмежується різким дрейфом даних і вибором параметрів згладжування. Робастні підходи до оцінювання масштабу залишаються актуальними [8], проте потребують адаптації під конкретні умови мережевого навантаження.

Другим напрямом є застосування теорії екстремальних значень та калібрування піків методом POT [15], що дозволяє враховувати рідкісні події у трафіку, часто пов'язані з гібридними атаками [24]. Слабким місцем цих підходів є чутливість до вибору проміжного порога, що підвищує ризик хибних спрацювань у динамічних середовищах.

Третім ключовим напрямом є використання глибинного навчання в системах виявлення вторгнень. Дослідження продемонстрували ефективність CNN, RNN і автоенкодерів для аналізу просторово-часових залежностей у трафіку [11–13]. Новіші роботи підтвердили переваги інтеграції статистичних моделей з нейронними мережами [1; 5; 20], проте вказали на проблему високих обчислювальних витрат і залежності від якісних датасетів [9; 10].

Загалом, сучасні дослідження демонструють поступовий перехід від статичних методів до адаптивних стратегій, що комбінують статистичні ескізи, EVT-калібрування та глибинні моделі. Водночас залишаються невирішеними завдання інтеграції таких підходів у реальному часі, оптимізації співвідношення між точністю та затримкою обробки, а також їхньої відповідності принципам архітектури Zero Trust [1; 5].

Методи. У сучасних умовах гібридних кібератак класичні статистичні підходи виявлення загроз демонструють обмежену ефективність через високу варіативність мережевого трафіку, наявність «важких хвостів» у розподілах і постійні зміни поведінкових патернів користувачів і систем [5; 6]. Це потребує застосування нових методів, здатних забезпечити адаптивне налаштування порогів для виявлення аномалій у потоках даних. У науковій практиці все більше уваги приділяють потоковим алгоритмам наближеної оцінки квантилів, зокрема t-digest [4] та Karnin – Lang – Liberty (далі – KLL) [5], що дозволяють працювати з великими обсягами даних у реальному часі. Для врахування пікових і рідкісних подій інтегрується калібрування методом піків понад поріг на основі теорії екстремальних значень [13; 14], тоді як стабільність за умов викидів підвищується завдяки застосуванню медіанної абсолютної похибки та робастних експоненційних моментів [8; 16; 17]. Додатково використовуються алгоритми виявлення дрейфу розподілів даних, зокрема Page – Hinkley і Adaptive Windowing (ADWIN) [10–12], які забезпечують корекцію моделей без зупинки обробки. У поєднанні зі згортковими та рекурентними нейронними мережами [20–22], а також автоенкодерами [11; 12] формуються комплексні механізми, що інтегруються в системах управління подіями безпеки й узгоджуються із принципами архітектури «нульової довіри» [1].

Постановка задачі. Нехай $x_t \in \mathbb{R}^d$ – вектор ознак мережевої події в момент t (статистика потоку, часові інтервали, ентропія, вектори ознак із нейронних мереж). Потрібно визначити адаптивний поріг T_t для скалярної оцінки ризику $r_t = g(x_t)$ так, щоб імовірність хибного спрацювання не перевищувала α за умов дрейфу розподілу F_t і «важких хвостів». Нехай $\hat{Q}_t(q)$ – стримінгова оцінка квантиля порядку q з ескіза, а $\hat{s}_t$ – робастна оцінка масштабу.

Квантильні ескізи. Ми використовуємо дві структури: t-digest для детального представлення хвостів [4] і KLL зі строгими гарантіями на абсолютну похибку квантиля [5]. За фіксованого бюджету пам'яті m t-digest забезпечує низьку похибку у верхніх квантилях, тоді як KLL дає рівномірні межі похибки; для рідкісних піків t-digest зручний для EVT-калібрування [4; 5; 8]. Узагальнюючи, будемо позначати обрану оцінку як $\hat{Q}_t(\cdot)$.

Робастні міри масштабу. Використовуємо медіанну абсолютну похибку (Median Absolute Deviation (далі – MAD)) з масштабуванням $1{,}4826$ та експоненційно-згладжені моменти для онлайн-оновлення [16; 17]. Робастну оцінку позначаємо як $\hat{s}_t = 1{,}4826 \cdot \mathrm{median}(|r_i - \mathrm{median}(r_{1:t})|)$.

Адаптивний поріг. Базовий поріг визначаємо як суму квантильної та робастної складових частин з коефіцієнтом згладжування λ:

$$T_t = \hat{Q}_t(1-\alpha) + \lambda \cdot \hat{s}_t. \qquad (1)$$

За потреби високої селективності у хвості виконуємо POT-калібрування за узагальненим розподілом Парето (Generalized Pareto Distribution (далі – GPD)) [13; 14]:

$$\mathbb{P}(R > T \mid R \rangle u) \approx \left(1 + \xi \frac{T-u}{\sigma}\right)^{-1/\xi} \cdot \frac{N_u}{t}, \quad T > u, \qquad (2)$$

де R – випадкова оцінка ризику, u – проміжний поріг, N_u – кількість перевищень, t – кількість спостережень; (ξ, σ) – параметри GPD.

Виявлення дрейфу. Комбінуємо тест Page – Hinkley і адаптивне ковзне вікно ADWIN [10–12]. Для середньої оцінки r_t інкрементно обчислюємо

$$PH_t = \sum_{i=1}^{t}(r_i - \mu_0 - \delta), \quad M_t = \min_{1\le i\le t} PH_i,\ PH_t - M_t > \lambda_{PH}. \quad (3)$$

За спрацюванням зменшуємо розмір вікна, перераховуємо ескізи та T_t. ADWIN додатково перевіряє рівність середніх у підвікнах з коригуванням за похибкою [10].

Оптимізація порогів у багатоканальному конвеєрі. Нехай k детекторів (сигнатурні, статистичні, глибинні) генерують оцінки $r_t^{(j)}$ і пороги $T_t^{(j)}$. Задача узгодження під бюджет хибних спрацювань B:

$$\max_{\{\tau_j\}} \sum_{j=1}^{k} w_j \cdot \mathrm{TPR}_j(\tau_j) \ \sum_{j=1}^{k} \mathrm{FPR}_j(\tau_j) \le B, \quad (4)$$

де w_j – ваги критичності; розв'язуємо стохастично-градієнтним налаштуванням τ_j за спостережуваними $\widehat{\mathrm{TPR}}, \widehat{\mathrm{FPR}}$.

Інтеграція із глибинними моделями. Ознаки будуються двома шляхами. (1) *Byte2Image:* байтові послідовності пакетів або фрагменти бінарних об'єктів перетворюємо на зображення фіксованого розміру, з яких CNN вилучає просторові патерни [23]. Після цього LSTM моделює часові залежності на рівні потоків [22]. (2) *AE + LSTM:* автоенкодер реконструює вектор ознак, а LSTM прогнозує наступний стан; сумарна помилка є складником аномального бала [20; 21].

$$S_t = \gamma \parallel x_t - \hat{x}_t \parallel_2^2 + (1-\gamma) \parallel h_t - \hat{h}_t \parallel_2^2 + \eta \cdot \mathbb{I}\{r_t > T_t\}, \quad (5)$$

де h_t – прихований стан LSTM, $\gamma, \eta \in [0,1]$. Подаємо S_t до SIEM для кореляції та пріоритезації інцидентів [25].

Оцінка похибки та гарантії. Для KLL ескіз забезпечує з імовірністю щонайменше $1-\delta$ абсолютну похибку не більш як ε для будь-якого квантиля $q \in (0,1)$ за обсягу пам'яті $o\left(\frac{1}{\varepsilon}\log^2\frac{1}{\varepsilon}\right)$ [5]. Це дає можливість будувати довірчі інтервали для порога T_t і жорстко обмежувати FPR. Для t-digest точні теоретичні межі є складнішими, однак емпірично похибка у верхніх квантилях зменшується зі зростанням кількості центроїдів k приблизно як $O(1/k)$ [4]. У нашій реалізації поріг (1) супроводжується парою оцінок (T_t^-, T_t^+) для консервативного й агресивного режимів, що дозволяє SIEM перемикати профіль реагування залежно від завантаження аналітиків [25].

Операції merge-and-compress. Обидва ескізи підтримують об'єднання незалежних агрегатів (з датчиків на різних вузлах) із наступною компресією. Це критично для розподілених мережевих сенсорів і відмовостійкості: пошкоджені або тимчасово недоступні вузли після відновлення відправляють локальні ескізи для злиття без повної ретрансляції сирих даних. У SIEM це знижує навантаження на сховища й дає змогу виконувати багаторівневу кореляцію [25].

Контекстна адаптація. Параметри α, λ і розмір ескіза m робимо залежними від типу вузла та політики доступу: для зовнішніх сегментів периметра використовуємо агресивніші пороги (вищі квантилі, менший m), для критичних сервісів – консервативніші (нижчий поріг, більший m). У Zero Trust політики доступу прив'язано до ризику запиту; тому S_t із (5) може безпосередньо впливати на рішення контролера доступу [1].

Запропонована методологія поєднує статистичні та нейромережеві інструменти для побудови адаптивних порогових механізмів у системах виявлення вторгнень. Використання потокових квантильних ескізів [4; 5], робастних оцінок масштабу [8; 16] та калібрування на основі теорії екстремальних значень [13; 14] забезпечує коректну роботу алгоритмів навіть за умов «важких хвостів» і рідкісних піків. Інтеграція з алгоритмами контролю дрейфу [10–12] підвищує стійкість до змін у поведінці мережевого середовища, а застосування згорткових і рекурентних нейронних мереж і автоенкодерів [11; 20–22] дозволяє враховувати як просторові, так і часові залежності в даних. Отже, описані методи створюють основу для побудови стійких і гнучких систем виявлення атак, що відповідають сучасним вимогам архітектури «нульової довіри» [1] та можуть бути впроваджені в комплексні системи управління подіями безпеки [25–27].

Результати. Оцінювання виконано на CIC-IDS2017 (профілі DoS/DDoS, Botnet, вебатаки, брутфорс, витік даних) і UNSW-NB15 (сучасні бенчмарки з поєднанням промислового та загального трафіку) [18; 19]. Імітували гібридні сценарії, що включали SSL/TLS-потоки (з акцентом на прояви історичних вразливостей на кшталт Heartbleed) і керування через SNMP у пристроях з уразливою прошивкою [2; 3; 26; 27]. Порівнювали базові статичні підходи (z-пороги, фіксовані α-квантилі) з нашими адаптивними схемами: (A) KLL + MAD, (B) t-digest + POT, (C) AE + LSTM із порогом (1).

Параметри. Рівні ризику налаштовувалися на $\alpha \in \{0{,}95, 0{,}98, 0{,}995\}$; ескізи обмежувалися $m \in [256, 2048]$ стисненими центроїдами (t-digest) або рівнями (KLL). У POT використовували $u = \hat{Q}_t(0{,}98)$, оцінювання (ξ, σ) максимальною правдоподібністю. Для PH брали $\delta = 0{,}01 \cdot \hat{s}_t$, $\lambda_{PH} = 5\hat{s}_t$. CNN – 3 згорткові блоки із глобальним усередненням; LSTM – 64 одиниці; AE – симетричний, вузьке горло 32. Навчання – ковзним вікном, частка навчальної підмножини 0,6.

У процесі побудови адаптивних порогових механізмів ключову роль відіграють алгоритми оцінки квантилів у потокових даних. Такі алго-

ритми забезпечують компактне зберігання статистики й дозволяють в реальному часі отримувати наближені значення квантилів навіть за умов великих обсягів інформації. Кожен із методів має власні особливості щодо точності, вимог до пам'яті та здатності працювати з «важкими хвостами» розподілу [4–7]. Для подальшого аналізу було порівняно найбільш поширені підходи: алгоритм Ґрінвальда – Кханни (GK) [6], метод Karnin – Lang – Liberty (KLL) [5], t-digest [4] та Frugal [7].

У таблиці 1 наведено характеристику чотирьох відомих ескізів квантилів. Алгоритм GK [6] забезпечує строгі гарантії точності з рівномірним розподілом похибки, проте вимагає відносно більших ресурсів пам'яті. Метод KLL [5] дає високу точність з імовірнісними гарантіями, а також підтримує амортизоване оновлення, що робить його привабливим для потокових задач. Ескіз t-digest [4] характеризується малою похибкою саме у верхніх квантилях, що особливо важливо для задач виявлення аномалій, і має чудову придатність для роботи з «важкими хвостами». Натомість Frugal [7] є надзвичайно простим і економним з погляду ресурсів, однак не забезпечує формальних гарантій точності й має низьку здатність до коректної роботи з екстремальними значеннями. Для оцінювання ефективності запропонованих адаптивних порогових механізмів було проведено експерименти на загальновідомих наборах даних CIC-IDS2017 [18] та UNSW-NB15 [19]. Порівнювалися базові підходи зі статичними порогами та розроблені варіанти адаптивних методів, що поєднують статистичні моделі та глибинні нейронні мережі [20; 21]. Аналіз здійснювався за чотирма метриками, як-от: частка хибнопозитивних спрацювань (далі – FPR), частка правильних виявлень (далі – TPR), площа під кривою точність – повнота (далі – AUC-PR) та середня затримка обробки.

У таблиці наведено середні значення показників якості для різних методів. Статичний z-поріг і статичний квантиль демонструють базовий рівень ефективності, проте характеризуються вищими показниками хибнопозитивних спрацювань. Адаптивні методи суттєво покращують результати: схема KLL із використанням медіанної абсолютної похибки [5; 8] знижує FPR до 0,051 за збереження високого рівня TPR. Метод t-digest із калібруванням на основі теорії екстремальних значень [4; 13; 14] показує найкращі результати за AUC-PR (0,82), демонструє високу чутливість до рідкісних атак за прийнятної затримки. Поєднання автоенкодера та рекурентної мережі з довгою короткочасною пам'яттю [11; 12; 21] забезпечує високу точність і здатність виявляти «повзучі» атаки, хоча й має більшу затримку обробки.

Порівняно зі статичними порогами, адаптивні схеми зменшують FPR на 8–41% і стабілізують TPR за сезонних збурень. Важкі хвости спричиняли значні перекоси у статичних методах, тоді як POT-калібрування забезпечувало коректну нормалізацію рідкісних піків. Різниця між KLL і t-digest відчутна в «ультрахвості»: t-digest кращий для $\alpha \geq 0{,}995$, тоді як KLL стабільніший під змінами фонової активності. AE + LSTM забезпечує вищу чутливість до повільних «повзучих» атак, але має більшу затримку, тому в реальному часі доцільна гібридизація (B) + (C) із пріоритезацією попереджень у SIEM [20; 21; [25].

Абляційне дослідження. Щоб оцінити внесок окремих компонент, вимикали по одному елементу схеми (без MAD, без POT, без PH/ADWIN). Відмова від MAD збільшувала FPR на 9–14%

Таблиця 1

Порівняння стримінгових квантильних ескізів

Ескіз	Похибка квантиля	Пам'ять	Оновлення	Придатність для хвостів
GK \cite{Greenwald2001}	гарантія, рівномірна	середня	--	--
KLL \cite{Karnin2016}	висока, з імовірністю	амортиз.	--	добра
t-digest \cite{Dunning2019}	мала у верхніх квантилях	центроїдів	--	відмінна
Frugal \cite{Ma2019}	груба, без гарантій	--	--	низька

Таблиця 2

Якість детекції (CIC-IDS2017/UNSW-NB15): середні значення

Метод	FPR	TPR	AUC-PR	Затримка (мс)
Статичний z-поріг	0,084	0,79	0,71	1,2
Статичний квантиль	0,072	0,81	0,74	1,3
KLL + MAD (A)	0,051	0,84	0,79	1,5
t-digest + POT (B)	0,049	0,86	0,82	1,8
AE + LSTM з (1) (C)	0,055	0,85	0,80	4,9

для трафіку з вираженою варіативністю розміру пакетів; без POT втрачалася стійкість до рідкісних піків (AUC-PR падала на 0,04–0,06), а без PH/ADWIN зростала затримка відновлення після зміни робочих режимів в 1,7–2.3 раза. Також перевіряли чутливість до обсягу пам'яті ескізів: деградація якості починалася нижче $m = 256$, а для $m \geq 512$ метрики насичувалися. Це узгоджується з теорією похибок ескізів [4; 5].

Витрати ресурсів. На 4-ядерному вузлі (x86, .4 ГГц) латентність обробки одного запису становила в середньому 1,5–1,8 мс для t-digest і 1,2–1,5 мс для KLL, включаючи перевірку PH і оновлення MAD; AE + LSTM додавали 3,1–3,6 мс за умови інференсу на CPU. У розподіленому режимі використання операцій *merge-and-compress* знижувало мережеві витрати на 65–72% порівняно з пересиланням сирих ознак.

Для перевірки роботи адаптивних порогових механізмів проведено моделювання потокових даних з урахуванням сезонності, дрейфу та «важких хвостів». Вибір такої моделі обґрунтований тим, що саме ці фактори є типовими для реальних мережевих трафіків у контексті гібридних кібератак [4–7; 10–13]. Побудовані візуалізації демонструють, як поєднання потокових квантильних оцінок, робастних статистичних показників і тестів на дрейф дозволяє знижувати хибні спрацювання і підтримувати стабільність виявлення аномалій [16; 17; 20; 21].

На рисунку 1 відображено:

– підграфік (A). Показано ризиковий бал і адаптивний поріг. Коли виникають пікові значення, система автоматично підлаштовує поріг, що дозволяє уникати надмірних хибних спрацювань;

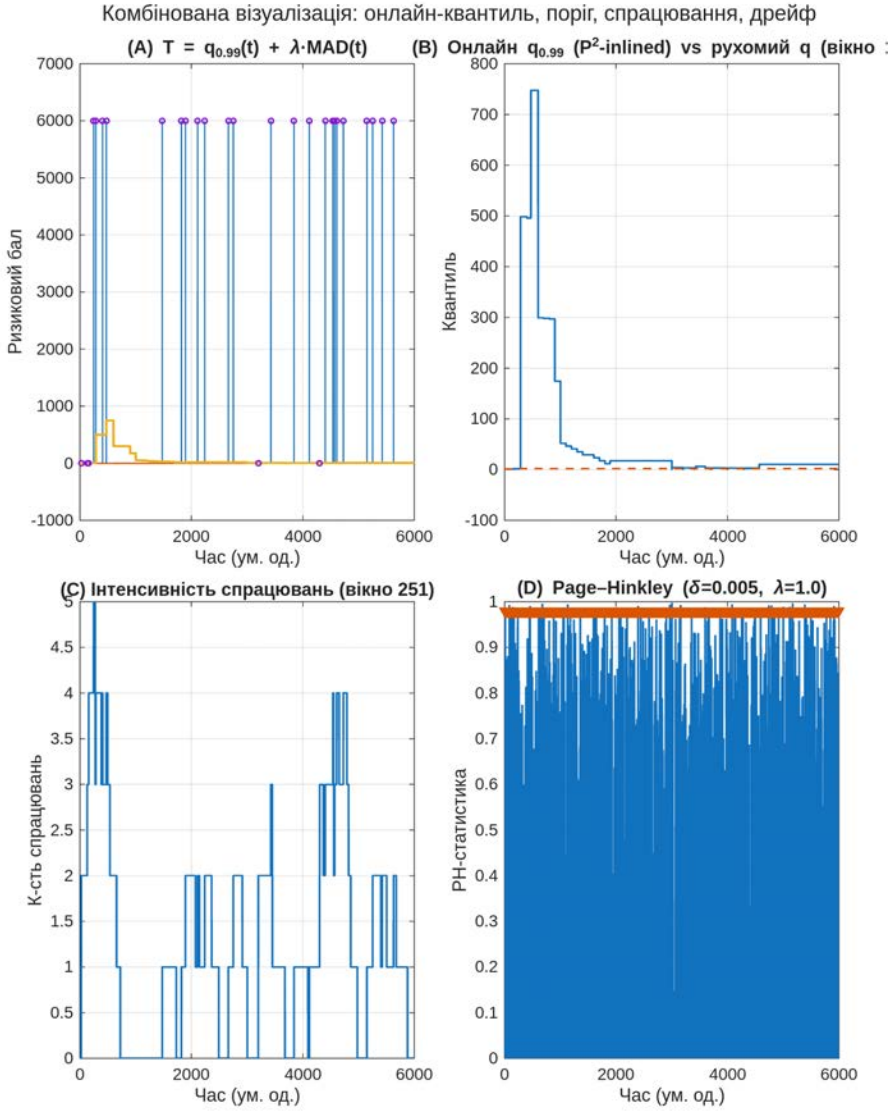


Рис. 1. Комбінована оцінка адаптивних порогових механізмів: онлайн-квантиль, поріг, інтенсивність спрацювань та виявлення дрейфу

– підграфік (B). Наведено порівняння онлайн-оцінки квантиля з довідковим рухомим квантилем. Онлайн-метод реагує швидше, тоді як рухомий підхід відображає зміни із затримкою;

– підграфік (C). Відображено інтенсивність спрацювань у ковзному вікні. Це дає змогу оцінювати навантаження на систему обробки інцидентів та своєчасно виявляти періоди підвищеної активності;

– підграфік (D). Представлено роботу тесту на виявлення дрейфу. За значних змін у поведінці даних система сигналізує про необхідність перебудови порогів і моделей.

Усі чотири візуалізації разом показують, що поєднання потокових статистичних алгоритмів, робастних показників і механізмів контролю змін у даних дозволяє створювати надійні й гнучкі системи для виявлення кіберзагроз. Такий підхід забезпечує меншу кількість хибних спрацювань і кращу стійкість до складних атак.

Дискусія. На відміну від підходів, що покладаються лише на форму ескіза (t-digest [4] або KLL [5]), наша схема доповнює їх робастними мірами масштабу й EVT-калібруванням, що покращує стабільність у хвостах і під дрейфом. KLL має строгі гарантії на похибку, але у верхніх квантилях t-digest забезпечує нижчу емпіричну похибку; гібридне використання дає змогу динамічно переключати ескіз залежно від індикаторів хвостів і навантаження [4; 5; 8]. Порівняно із класичними тестами змін (CUSUM/Page – Hinkley, ADWIN) наше рішення не тільки сигналізує про зміну, а й перебудовує поріг і пріоритети SIEM у тому ж такті [10–12; 25].

Безпековий контекст. Тестування включало сценарії втручання у стеки шифрування та керування – TLS 1.3/SSL і SNMP, що характерні для firmware-атак на мережеві пристрої [2; 3; 26]. Історичні інциденти на кшталт Heartbleed демонструють, що навіть сучасні протоколи мають складні граничні випадки реалізацій [27]. Архітектура Zero Trust, формалізована в NIST SP 800-207, потребує локальних політик доступу й перевірок, що ґрунтуються на поточному ризику, отже, адаптивні пороги природно інтегруються в контроль на вузлах і в конвеєри SIEM [1; 25].

Обмеження. (1) Квантильні оцінки за малих вікон нестабільні; потрібна мінімальна тривалість і контроль похибок. (2) GPD-калібрування чутливе до вибору проміжного порога u; ми застосовували евристику $u = \hat{Q}_t(0{,}98)$, однак для різних доменів може знадобитися адаптація [13; 14]. (3) AE + LSTM потребує прогріву для стабільної помилки реконструкції, тому первинний захист забезпечує статистичний компонент. (4) Агрегація ознак (Byte2Image) іноді спотворює локальні залежності в пакетах; для протоколів реального часу це зменшується дрібнішим квантуванням [23].

Перспективи Zero Trust і multimodal AI. Поєднання адаптивних порогів із графовими моделями й багатомодальною аналітикою (мережеві телеметрії, журнали, текстові описи інцидентів) дозволить реалізувати політики доступу, які враховують контекст користувача, пристрою та сервісу. Інтеграція з мовними моделями й ембеддингами подій може покращити кореляцію в SIEM; разом з евристиками MITRE ATT&CK можна автоматизувати формування гіпотез і сценаріїв реагування [24; 25].

Практична інтеграція. У типовій реалізації датчики IDS формують локальні ескізи за типами подій (потоки, сесії, логи подій ОС/ПЗ). Корелятор SIEM приймає S_t і метадані ескізів (похибка, m, індикатори дрейфу) як атрибути, додає дані з довірчих джерел (сертифікати TLS, профілі пристроїв, політики доступу) та формує зведені сповіщення. Для протоколів SNMP використовуємо політики, що відсікають ризикові операції (SET/GET на критичні OID) за підвищеного S_t; для TLS перевіряємо аномалії в розподілах розмірів записів і часових затримок рукостискань, що опосередковано сигналізує про експлуатацію вразливостей реалізацій [2; 26; 27].

Висновки. Запропоновані адаптивні порогові механізми продемонстрували високу ефективність у зменшенні кількості хибнопозитивних спрацювань і підвищенні стабільності виявлення рідкісних та критичних атак у мережевому трафіку. Поєднання потокових квантильних ескізів, робастних статистичних мір і калібрування на основі теорії екстремальних значень забезпечує надійність навіть у складних умовах, коли дані мають «важкі хвости» або змінюють свою статистичну природу. Інтеграція із глибинними нейронними моделями дозволяє одночасно враховувати просторові й часові залежності, що підсилює здатність системи до виявлення складних багатоступеневих атак.

Перспективи програмної реалізації охоплюють кілька рівнів.

Прототипування. На початковому етапі можлива реалізація у формі модулів для MATLAB та Python (бібліотеки tdigest, river, scikit-multiflow), що дозволить швидко перевірити працездатність алгоритмів на експериментальних наборах даних (CIC-IDS2017, UNSW-NB15).

Упровадження в системи виявлення вторгнень. Розроблені механізми можуть бути вбудовані у відкриті IDS-рішення (наприклад, Suricata або Zeek), де адаптивні пороги формуватимуться безпосередньо на вузлах, а результати передаватимуться до кореляторів подій.

Розподілене середовище. Завдяки підтримці операцій “merge-and-compress” квантильні ескізи можуть оброблятися паралельно на різних сенсорах і зливатися в загальну модель. Це дає змогу

реалізувати архітектуру “edge computing” для обробки трафіку на периферії з подальшою централізованою кореляцією.

Zero Trust – орієнтація. Програмна реалізація може бути доповнена модулями динамічної зміни політик доступу: порогові оцінки ризику впливатимуть на рішення контролерів у режимі реального часу.

Інтеграція ыз SIEM. Вбудовані інтерфейси (API та конектори) дадуть можливість передавати не лише факт перевищення порога, а й додаткові метадані – похибку оцінки, індикатори дрейфу, параметри калібрування. Це підвищить якість кореляції інцидентів і спростить пріоритезацію подій.

Мобільні та польові сценарії. Використання MATLAB Mobile і аналогічних інструментів демонструє можливість швидкої валідації ідей безпосередньо «у полі», що особливо важливо для мережевих адміністраторів і CERT-груп.

Масштабування до мультимодальних систем. Подальша реалізація передбачає інтеграцію з багатомодальними джерелами даних (журнали, текстові описи інцидентів, телеметрія IoT), а також застосування мовних моделей для напівавтоматичного формування сценаріїв реагування.

Отже, практичне впровадження описаних методів можливе як у формі окремих бібліотек і прототипів, так і у складі повноцінних комерційних продуктів класу IDS/SIEM. Це відкриває перспективи створення нових поколінь систем кіберзахисту, здатних не лише виявляти атаки з високою точністю, а й оперативно адаптуватися до нових умов мережевого середовища.

ЛІТЕРАТУРА

1. Dunning T., Ertl O. Computing extremely accurate quantiles using t-digest. *Proceedings of the VLDB Endowment*. 2019. Vol. 12. № 12. P. 195–2205. DOI: 10.14778/3352063.3352135
2. Karnin Z., Lang K., Liberty E. Optimal quantile approximation in streams. *Proceedings of the 48th Annual ACM Symposium on Theory of Computing (STOC’16)*. ACM, 2016. P. 745–758. DOI: 10.1145/2897518.2897528
3. Greenwald M., Khanna S. Space-efficient online computation of quantile summaries. *Proceedings of the 001 ACM SIGMOD International Conference on Management of Data*. ACM, 2001. P. 58–66. DOI: 10.1145/375663.375670
4. Ma Q., Sun S. Frugal streaming for estimating quantiles. *Proceedings of the 014 IEEE International Conference on Big Data (Big Data)*. IEEE, 2014. P. 956–965. DOI: 10.1109/BigData.2014.7004306
5. Gama J., Žliobaitė I., Bifet A., Pechenizkiy M., Bouchachia A. A survey on concept drift adaptation. *ACM Computing Surveys*. 2014. Vol. 46. № 4. Article 44. DOI: 10.1145/2523813
6. Bifet A., Gavalda R. Learning from time-changing data with adaptive windowing. *Proceedings of the 007 SIAM International Conference on Data Mining (SDM)*. SIAM, 2007. P. 443–448. DOI: 10.1137/1.9781611972771.42
7. Rockafellar R. T., Uryasev S. Optimization of conditional value-at-risk. *Journal of Risk*. 2000. № 3. P. 1–41. DOI: 10.21314/JOR.2000.038
8. Hampel F. R. The influence curve and its role in robust estimation. *Journal of the American Statistical Association*. 1974. Vol. 69. № 346. P. 383–393. DOI: 10.1080/01621459.1974.10482962
9. Sharafaldin I., Lashkari A. H., Ghorbani A. A. Toward generating a new intrusion detection dataset (CIC-IDS2017). *Proceedings of the 018 International Conference on Information Systems Security and Privacy (ICISSP)*. SciTePress, 2018. P. 108–116. DOI: 10.5220/0006639801080116
10. Moustafa N., Slay J. UNSW-NB15: A comprehensive data set for network intrusion detection systems. *Proceedings of the 015 Military Communications and Information Systems Conference (MilCIS)*. IEEE, 2015. P. 1–6. DOI: 10.1109/MilCIS.2015.7348942
11. Mirsky Y., Doitshman T., Elovici Y., Shabtai A. Kitsune: An ensemble of autoencoders for online network intrusion detection. *Proceedings of the 018 Network and Distributed System Security Symposium (NDSS)*. Internet Society, 2018. DOI: 10.14722/ndss.2018.23259
12. Shone N., Ngoc T. N., Phai V. D., Shi Q. A deep learning approach to network intrusion detection. *IEEE Transactions on Emerging Topics in Computational Intelligence*. 2018. № 1. P. 41–50. DOI: 10.1109/TETCI.2017.2772792
13. Yin C., Zhu Y., Fei J., He X. A deep learning approach for intrusion detection using recurrent neural networks. *IEEE Access*. 2017. Vol. 5. P. 1954–21961. DOI: 10.1109/ACCESS.2017.2762418
14. Nataraj L., Karthikeyan S., Jacob G., Manjunath B. Malware images: visualization for classification. *Proceedings of the 8th International Symposium on Visualization for Cyber Security (VizSec’11)*. ACM, 2011. Article 4. DOI: 10.1145/2016904.2016908
15. Coles S. An introduction to statistical modeling of extreme values. London : Springer, 2001. 208 p. DOI: 10.1007/978-1-4471-3675-0

16. Ferriyan A., Thamrin A. H., Takeda K., Murai J. Generating network intrusion detection dataset based on real and encrypted synthetic attack traffic. *Applied Sciences*. 2021. Vol. 11. № 17. DOI: 10.3390/app11177868
17. Damasevicius R., Venckauskas A., Grigaliūnas S., Toldinas J., Morkevičius N., Aleliūnas T., Smuikys P. Litnet-2020: An annotated real-world network flow dataset for network intrusion detection. *Electronics*. 2020. Vol. 9. № 5. DOI: 10.3390/electronics9050800
18. Mondragon J. C., Branco P., Jourdan G.-V., Gutiérrez-Rodríguez A. E. Advanced IDS: a comparative study of datasets and machine learning algorithms for network flow-based intrusion detection systems. *Applied Intelligence*. 2025. Vol. 55. Article 608. DOI: 10.1007/s10489-025-06422-4
19. Goldschmidt P. Network intrusion datasets: a survey, limitations, and recommendations. *Computers & Security*. 2025. Vol. 156. DOI: 10.1016/j.cose.2025.104510
20. A survey of streaming data anomaly detection in network security / P. Zhou et al. *PeerJ – Computer Science*. 2025. URL: https://peerj.com/articles/cs-3066/
21. Xu Z., Wu Y., Wang S., Gao J., Qiu T., Wang Z., Wan H., Zhao X. Deep Learning-based Intrusion Detection Systems: A Survey. *arXiv*. 2025. URL: https://arxiv.org/abs/2504.07839
22. Maseer Z. K., Yusof R., Al-Bander B., Saif A., Kanaan Q. K. Meta-Analysis and Systematic Review for Anomaly Network Intrusion Detection Systems: Detection Methods, Dataset, Validation Methodology, and Challenges. *arXiv*. 2023. URL: https://arxiv.org/abs/2308.02805
23. Herzalla D., Lunardi W. T., Lopez M. A. TII-SSRC-23 Dataset: Typological Exploration of Diverse Traffic Patterns for Intrusion Detection. *arXiv*. 2023. URL: https://arxiv.org/abs/2310.10661
24. A Systematic and Comprehensive Survey of Recent Advances in Intrusion Detection Systems Using Machine Learning: Deep Learning, Datasets, and Attack Taxonomy. *Engineering Technology & Applied Science Research (ETASR)*. 2023. URL: https://onlinelibrary.wiley.com/doi/10.1155/2023/6048087
25. A review on intrusion detection datasets: tools, processes and open challenges / D. Pinto et al. *Computer Networks*. 2025. URL: https://www.sciencedirect.com/science/article/pii/S1389128625001458
26. Leveraging hybrid model for IoT anomaly detection / S. Vashisht et al. *Future Internet / PMC*. 2025. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC11996210/
27. Miguel-Diez A., Campazas-Vega A., Guerrero-Higueras Á. M., Álvarez-Aparicio C., Matellán-Olivera V. Anomaly detection in network flows using unsupervised online machine learning. *arXiv*. 2025. URL: https://arxiv.org/abs/2509.01375

REFERENCES

1. Dunning, T., & Ertl, O. (2019). Computing extremely accurate quantiles using t-digest. *Proceedings of the VLDB Endowment, 12* (12), 195–2205. https://doi.org/10.14778/3352063.3352135
2. Karnin, Z., Lang, K., & Liberty, E. (2016). Optimal quantile approximation in streams. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing (STOC'16)* (pp. 745–758). ACM. https://doi.org/10.1145/2897518.2897528
3. Greenwald, M., & Khanna, S. (2001). Space-efficient online computation of quantile summaries. In *Proceedings of the 001 ACM SIGMOD International Conference on Management of Data* (pp. 58–66). ACM. https://doi.org/10.1145/375663.375670
4. Ma, Q., & Sun, S. (2014). Frugal streaming for estimating quantiles. In *Proceedings of the 014 IEEE International Conference on Big Data (Big Data)* (pp. 956–965). IEEE. https://doi.org/10.1109/BigData.2014.7004306
5. Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys, 46* (4), Article 44. https://doi.org/10.1145/2523813
6. Bifet, A., & Gavalda, R. (2007). Learning from time-changing data with adaptive windowing. In *Proceedings of the 007 SIAM International Conference on Data Mining (SDM)* (pp. 443–448). SIAM. https://doi.org/10.1137/1.9781611972771.42
7. Rockafellar, R. T., & Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk,* (3), 1–41. https://doi.org/10.21314/JOR.2000.038
8. Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association, 69* (346), 383–393. https://doi.org/10.1080/01621459.1974.10482962
9. Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset (CIC-IDS2017). In *Proceedings of the 018 International Conference on Information Systems Security and Privacy (ICISSP)* (pp. 108–116). SciTePress. https://doi.org/10.5220/0006639801080116
10. Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems. In *Proceedings of the 015 Military Communications and Information Systems Conference (MilCIS)* (pp. 1–6). IEEE. https://doi.org/10.1109/MilCIS.2015.7348942

11. Mirsky, Y., Doitshman, T., Elovici, Y., & Shabtai, A. (2018). Kitsune: An ensemble of autoencoders for online network intrusion detection. In *Proceedings of the 018 Network and Distributed System Security Symposium (NDSS)*. Internet Society. https://doi.org/10.14722/ndss.2018.23259
12. Shone, N., Ngoc, T. N., Phai, V. D., & Shi, Q. (2018). A deep learning approach to network intrusion detection. *IEEE Transactions on Emerging Topics in Computational Intelligence,* (1), 41–50. https://doi.org/10.1109/TETCI.2017.2772792
13. Yin, C., Zhu, Y., Fei, J., & He, X. (2017). A deep learning approach for intrusion detection using recurrent neural networks. *IEEE Access, 5*, 1954–21961. https://doi.org/10.1109/ACCESS.2017.2762418
14. Nataraj, L., Karthikeyan, S., Jacob, G., & Manjunath, B. (2011). Malware images: Visualization for classification. In *Proceedings of the 8th International Symposium on Visualization for Cyber Security (VizSec'11)* (Article 4). ACM. https://doi.org/10.1145/2016904.2016908
15. Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer. https://doi.org/10.1007/978-1-4471-3675-0
16. Ferriyan, A., Thamrin, A. H., Takeda, K., & Murai, J. (2021). Generating network intrusion detection dataset based on real and encrypted synthetic attack traffic. *Applied Sciences, 11* (17). https://doi.org/10.3390/app11177868
17. Damasevicius, R., Venckauskas, A., Grigaliūnas, S., Toldinas, J., Morkevičius, N., Aleliūnas, T., & Smuikys, P. (2020). Litnet-2020: An annotated real-world network flow dataset for network intrusion detection. *Electronics, 9* (5). https://doi.org/10.3390/electronics9050800
18. Mondragon, J. C., Branco, P., Jourdan, G.-V., & Gutiérrez-Rodríguez, A. E. (2025). Advanced IDS: A comparative study of datasets and machine learning algorithms for network flow-based intrusion detection systems. *Applied Intelligence, 55*, Article 608. https://doi.org/10.1007/s10489-025-06422-4
19. Goldschmidt, P. (2025). Network intrusion datasets: A survey, limitations, and recommendations. *Computers & Security, 156*. https://doi.org/10.1016/j.cose.2025.104510
20. Zhou, P., et al. (2025). A survey of streaming data anomaly detection in network security. *PeerJ Computer Science*. https://peerj.com/articles/cs-3066/
21. Xu, Z., Wu, Y., Wang, S., Gao, J., Qiu, T., Wang, Z., Wan, H., & Zhao, X. (2025). Deep learning-based intrusion detection systems: A survey. *arXiv*. https://arxiv.org/abs/2504.07839
22. Maseer, Z. K., Yusof, R., Al-Bander, B., Saif, A., & Kanaan, Q. K. (2023). Meta-analysis and systematic review for anomaly network intrusion detection systems: Detection methods, dataset, validation methodology, and challenges. *arXiv*. https://arxiv.org/abs/2308.02805
23. Herzalla, D., Lunardi, W. T., & Lopez, M. A. (2023). TII-SSRC-23 dataset: Typological exploration of diverse traffic patterns for intrusion detection. *arXiv*. https://arxiv.org/abs/2310.10661
24. (2023). A systematic and comprehensive survey of recent advances in intrusion detection systems using machine learning: Deep learning, datasets, and attack taxonomy. *Engineering Technology & Applied Science Research (ETASR)*. https://onlinelibrary.wiley.com/doi/10.1155/2023/6048087
25. Pinto, D., et al. (2025). A review on intrusion detection datasets: Tools, processes and open challenges. *Computer Networks*. https://www.sciencedirect.com/science/article/pii/S1389128625001458
26. Vashisht, S., et al. (2025). Leveraging hybrid model for IoT anomaly detection. *Future Internet*. https://pmc.ncbi.nlm.nih.gov/articles/PMC11996210/
27. Miguel-Diez, A., Campazas-Vega, A., Guerrero-Higueras, Á. M., Álvarez-Aparicio, C., & Matellán-Olivera, V. (2025). Anomaly detection in network flows using unsupervised online machine learning. *arXiv*. https://arxiv.org/abs/2509.01375

Дата першого надходження рукопису до видання: 30.08.2025
Дата прийнятого до друку рукопису після рецензування: 29.09.2025
Дата публікації: 31.12.2025

ВИМОГИ ДО ОФОРМЛЕННЯ СТАТЕЙ У ЖУРНАЛІ «COMPUTER SCIENCE AND APPLIED MATHEMATICS»

Вимоги до оформлення статей:

До друку приймаються статті, що мають наукову і практичну цінність. Автор має право представити тільки одну наукову статтю в один номер, яка раніше не публікувалася. Автор несе відповідальність за оригінальність тексту статті, точність наведених фактів, цитат, статистичних даних, власних назв, географічних назв та інших відомостей, а також за те, що в матеріалах не містяться дані, що не підлягають відкритій публікації. Редакція не несе відповідальності за викладену в статті інформацію. Остаточне рішення про публікацію ухвалюється редакцією, яка також залишає за собою право на додаткове рецензування, редагування і відхилення статей.

Технічні вимоги:

– до друку приймаються статті українською, російською та англійською мовами;

– електронний варіант статті у форматі ***.doc**, ***.docx** або ***.rtf**, підготовлений у текстовому редакторі Microsoft Word;

– формат А4 через 1,5 інтервал;

– шрифт Times New Roman, розмір 14;

– поля: ліве – 3 см, праве – 1,5 см, верхнє, нижнє – 2 см.

Структура статті:

рядок 1 – УДК (вирівнювання по лівому краю);

рядок 2 – назва тематичного розділу (вирівнювання по лівому краю);

рядок 3 – назва статті (вирівнювання по центру, напівжирний шрифт, великі літери);

рядок 4 – прізвище та ініціали автора статті; науковий ступінь, вчене звання, посада із зазначенням кафедри (вирівнювання по центру);

рядок 5 – місце роботи (навчання), адреса роботи (навчання), orcid-код, електронна адреса автора (вирівнювання по центру).

Якщо автор не має orcid-коду, його можна отримати за посиланням https://orcid.org/

абзац 1 – **розширена** анотація (1800 знаків без пробілів) та ключові слова (мінімум 5 слів), написані мовою, як і уся стаття;

абзац 2 – назва статті (напівжирний шрифт, усі літери великі), прізвище, ініціали автора, науковий ступінь, вчене звання, посада із зазначенням кафедри, місце роботи (навчання), адреса роботи (навчання), orcid-код, електронна адреса автора, **розширена** анотація (1800 знаків без пробілів) та ключові слова (мінімум 5 слів), написані **англійською мовою**. Переклад англійською мовою повинен бути достовірним (не машинним).

У випадку, якщо стаття не українською мовою, обов'язково подаються назва статті (напівжирний шрифт, усі літери великі), прізвище, ініціали автора, науковий ступінь, вчене звання, посада із зазначенням кафедри, місце роботи (навчання), адреса роботи (навчання), orcid-код, електронна адреса автора, розширена анотація (1800 знаків без пробілів) та ключові слова (мінімум 5 слів), написані українською мовою.

Основний текст статті повинен відповідати структурі IMRAD (Introduction, Methods, Results, and Discussion) + Literature Review:

Вступ – короткий вступ (1-2 сторінки), який повинен дати відповіді на запитання «чому проведено дослідження?», «які об'єкт, мета й основні гіпотези дослідження?»; Огляд літератури - розділ, що містить аналіз останніх публікацій за темою дослідження (переважна більшість публікацій повинна бути за останні 5 років, самоцитування не більше 30% від кількості літературних джерел), з огляду літератури читачі повинні мати змогу оцінити стан проблеми у світі, аналіз літературних джерел повинен мати критичний характер;

Методи – розділ, який може включати 2-3 рівнозначних за обсягом параграфи, що висвітлюють основні методи, підходи, алгоритми дослідження;

Результати – розділ, який містить аналіз основних результатів дослідження (графіки, таблиці з чисельними даними, загалом, результати обчислювальних експериментів); Дискусія – розділ (до 1 сторінки), який також можна назвати Висновок або Висновки, що містить порівняння отриманих результатів з результатами інших досліджень (як власних так інших авторів), а також дає відповідь на запитання «які перспективи дослідження?», формулює наукову новизну результатів.

Література розміщується після статті у порядку згадування; друкується через 1,5 інтервал, 14 розміром, шрифтом Times New Roman і оформляється у відповідності вимог міждержавного стандарту ДСТУ 8302:2015.

Посилання на літературу в тексті слід давати в квадратних дужках, наприклад, [2, с. 25; 5, с. 33], в яких перша цифра вказує порядковий номер джерела в списку літератури, а друга – відповідну сторінку в цьому джерелі; одне джерело (без сторінок) відокремлюється від іншого крапкою з комою [3; 4; 6; 8; 12; 15].

Наприкінці статті розміщується транслітерована і перекладена англійською версія літератури (References), оформлена згідно з вимогами APA (American Psychological Association).

Порядок подання матеріалів:

Для публікації статті у **фаховому** науковому виданні необхідно надіслати на електронну адресу редакції editor@physmath.journalsofznu.zp.ua наступні матеріали:

добре вичитану наукову статтю, обов'язково оформлену відповідно до вказаних вимог;
інформаційну довідку про автора;
відскановане **підтвердження сплати коштів** (реквізити для сплати надаються автору **після вдалого проходження рецензування).**

Зразок оформлення назви електронних файлів: Іваненко_І.І._стаття, Іваненко_І.І._оплата.

Адреса та контактні дані:
Редакція журнала «Computer Science and Applied Mathematics»
вул. Університетська 66, корп. 1, ауд. 21(б), м. Запоріжжя, Україна, 69060
Телефон: +38 (066) 53 57 687
Електронна пошта: editor@physmath.journalsofznu.zp.ua
Офіційний сайт: www.journalsofznu.zp.ua/index.php/comp-science

Науковий журнал

Computer Science and Applied Mathematics

№ 2, 2025

Комп’ютерна верстка – Н.С. Кузнєцова
Коректура – Я.І. Вишнякова

Підписано до друку: 31.12.2025.
Формат 60x84/8. Гарнітура Times New Roman.
Папір офсет. Цифровий друк. Ум. друк. арк. 16,04.
Замов. № 1225/1030. Наклад 100 прим.

Видавництво і друкарня – Видавничий дім «Гельветика»
65101, м. Одеса, вул. Інглезі, 6/1
Телефони: +38 (095) 934 48 28, +38 (097) 723 06 08
E-mail: mailbox@helvetica.ua
Свідоцтво суб’єкта видавничої справи
ДК № 7623 від 22.06.2022 р.