УДК 004.732.056 **doi: 10.26906/SUNZ.2026.1.113**

В. В. Челак, О. А. Горносталь

Національний технічний університет “Харківський політехнічний інститут”, Харків, Україна

УДОСКОНАЛЕНИЙ МЕТОД ПОБУДОВИ ДЕРЕВ З БАГАТОВИМІРНИМИ ВУЗЛАМИ РІШЕНЬ ДЛЯ ІДЕНТИФІКАЦІЇ СТАНУ КОМП’ЮТЕРНИХ СИСТЕМ

Анотація. Об’єктом дослідження є процес ідентифікації стану комп’ютерних систем на основі аналізу багатовимірних даних їх функціонування. **Предметом дослідження** є методи побудови дерев рішень із багатовимірними вузлами та алгоритми оптимізації їх структури для задач класифікації станів комп’ютерних систем. **Метою дослідження** є розробка та дослідження удосконаленого методу побудови дерев з багатовимірними вузлами рішень, який забезпечує підвищення точності та узагальнюючої здатності моделей ідентифікації стану комп’ютерних систем у складних і невизначених умовах. **Методи, що використовуються**: методи машинного навчання, деревні моделі класифікації, процедури нелінійного розділення простору, ітераційна оптимізація параметрів на основі звуження області пошуку, підходи до оцінювання якості класифікації. **Отримані результати**: розроблено процедуру автоматизованого вибору типу гіперфігури у вузлі дерева, що дозволяє адаптувати форму розділення до локальної структури даних. Показано, що використання узагальнених поверхонь зменшує необхідну глибину дерева та скорочує кількість послідовних поділів. В експериментальних дослідженнях встановлено підвищення показників Accuracy, Precision, Recall, F1-score та MCC у порівнянні з класичними деревами рішень і моделями з фіксованою геометрією вузлів. **Висновки**: запропонований метод забезпечує більш точне та стабільне моделювання меж між станами комп’ютерних систем, зберігаючи інтерпретованість прийнятих рішень і прийнятний рівень обчислювальних витрат. Отримані результати підтверджують доцільність використання адаптивних багатовимірних вузлів у системах моніторингу та інформаційної безпеки.

Ключові слова: ідентифікація стану, комп’ютерні системи, дерево рішень, багатовимірний вузол, нелінійне розділення, вирішальне правило, машинне навчання.

Вступ

Сучасні комп’ютерні системи характеризуються високим рівнем складності, гетерогенністю компонентів та динамічністю процесів, що в них відбуваються. Вони інтегрують апаратні, програмні та мережеві підсистеми, які взаємодіють між собою в режимі реального часу та функціонують в умовах постійних змін навантаження, конфігурації й зовнішніх впливів. За таких умов зростає ймовірність виникнення нештатних режимів роботи, прихованих відмов, деградації продуктивності або аномальної поведінки, зумовленої як внутрішніми збоями, так і цілеспрямованими атаками. Тому задача своєчасної та достовірної ідентифікації стану комп’ютерних систем є однією з ключових у сфері інформаційної безпеки, системного моніторингу та управління надійністю.

Традиційні підходи до оцінювання стану комп’ютерних систем, що ґрунтуються на жорстко заданих правилах, порогових значеннях або класичних статистичних моделях, мають обмежену ефективність у складних і слабко формалізованих умовах. Такі методи, як правило, потребують точного визначення меж між нормальними та аномальними станами, що не завжди можливо через стохастичний характер системних процесів, наявність шумів у вимірюваних даних та неповноту інформації. Крім того, із зростанням кількості контрольованих параметрів суттєво ускладнюється побудова універсальних правил, здатних адекватно описувати поведінку системи в усіх режимах її функціонування.

Огляд пов'язаних наукових публікацій. У роботі [1] запропоновано підхід до виявлення кіберзагроз, що поєднує попередню фільтрацію подій на основі правил та двоетапну модель машинного навчання. На першому етапі здійснюється відбір подій із використанням інформації про відомі вразливості та класифікацію атак, що дозволяє зменшити обсяг даних для подальшого аналізу. Другий етап передбачає застосування моделей машинного навчання для ідентифікації аномальної поведінки. Особливістю підходу є врахування додаткового контексту подій, зокрема джерела, середовища виконання та мережевих характеристик, що підвищує точність атрибуції загроз. Запропоноване рішення забезпечує зменшення обчислювальних витрат і покращення показників точності класифікації. Водночас використані методи класифікації базуються на класичних механізмах поділу простору ознак, що обмежує їх ефективність у випадках складних нелінійних залежностей та перекривання станів комп’ютерних систем.

У роботах [2, 3] розглянуто підходи до автоматизованої ідентифікації кіберзагроз на основі аналізу відкритих інформаційних джерел та потокових даних у режимі реального часу. Автори [2] запропонували комплексну архітектуру, що поєднує методи машинного та глибинного навчання з тематичним аналізом текстових даних для виявлення відомих і нових типів загроз, зокрема атак нульового дня. У роботі [3] основну увагу приділено напівкерованому навчанню в умовах обмеженої кількості розмічених даних, де поєднання генеративних моделей і контрастного навчання дозволяє підвищити якість класифікації загроз. Незважаючи на високу ефективність запропонованих підходів у задачах аналізу текстових потоків та інтелектуальної обробки відкритих даних, вони орієнтовані переважно на семантичну інтерпретацію інформаційних повідомлень і не враховують особливості багатовимірних телеметричних даних комп’ютерних систем, що обмежує їх застосування для безпосередньої ідентифікації станів систем на основі показників їх функціонування.

У роботах [4-6] досліджуються сучасні підходи до виявлення та ідентифікації кіберзагроз із використанням методів штучного інтелекту в різних класах складних кіберфізичних і мережевих систем. У дослідженні [4] запропоновано структурований підхід до проєктування систем виявлення вторгнень для SCADA-середовищ, що включає попередню обробку даних, балансування вибірок, автоматизований відбір ознак та порівняльний аналіз моделей машинного навчання, де найвищі результати продемонстрували ансамблеві методи. Робота [5] має оглядовий характер і систематизує напрями застосування штучного інтелекту в кібербезпеці, підкреслюючи зростаючу роль алгоритмів аналізу аномалій та інтелектуальної обробки багатовимірних даних для підвищення адаптивності систем захисту. У [6] розглянуто автоматизований фреймворк ідентифікації загроз у мережах взаємодії транспортних засобів і дорожньої інфраструктури, де глибинні моделі використовуються для виявлення прихованих шаблонів і класифікації типів атак. Водночас спільною рисою наведених підходів є орієнтація на складні нейромережеві або ансамблеві моделі з обмеженою інтерпретованістю, що ускладнює їх використання в задачах пояснюваної ідентифікації станів комп'ютерних систем та обґрунтовує доцільність подальших досліджень у напрямі розвитку інтерпретованих деревних моделей з багатовимірними вузлами рішень.

У роботах [7–10] розглянуто методи виявлення аномалій і кіберзагроз у складних розподілених середовищах із використанням машинного навчання та аналізу великих обсягів даних. Дослідження [7] зосереджене на задачі виявлення внутрішніх загроз у незбалансованих вибірках, де застосування щільнісних методів аналізу дозволяє ефективно ідентифікувати рідкісні аномальні стани. У роботі [8] запропоновано інтелектуальну архітектуру захисту середовищ Інтернету речей, що поєднує класифікаційні моделі машинного навчання з обчисленнями на периферії мережі для підвищення точності та оперативності виявлення атак. У [9] розглянуто превентивний підхід до виявлення аномалій і управління ризиками в бездротових комунікаційних системах, де акцент зроблено на аналізі часових і поведінкових характеристик у реальному часі. Робота [10] присвячена автоматизованій ідентифікації нових кіберзагроз на основі аналізу текстових потоків із відкритих джерел та їх подальшій класифікації за цілями і рівнем ризику. Водночас наведені підходи орієнтовані або на окремі аспекти аномальної поведінки, або на специфічні класи даних, що ускладнює їх безпосереднє використання для комплексної ідентифікації станів комп'ютерних систем у багатовимірному просторі параметрів і підкреслює актуальність розвитку інтерпретованих моделей, здатних враховувати взаємозв'язки між множиною ознак.

У роботах [11–14] методи машинного навчання застосовуються для ідентифікації станів і аномальних режимів у різнорідних інформаційних та кіберфізичних системах, що характеризуються багатовимірністю та високою динамікою даних. Дослідження [11] присвячене задачі виявлення та розпізнавання безпілотних апаратів на основі аналізу акустичних сигналів, де класифікація здійснюється з урахуванням складного впливу фонових факторів, що підкреслює важливість моделювання взаємозалежних ознак. У роботі [12] систематизовано підходи до забезпечення конфіденційності в інтерфейсах «мозок–комп'ютер», акцентуючи увагу на ризиках, пов'язаних з обробкою чутливих багатовимірних сигналів, та необхідності побудови адаптивних і контрольованих моделей аналізу. У [13] і [14] розглянуто методи виявлення атак підвищення привілеїв і кіберзагроз у хмарних та IoT-середовищах із використанням ансамблевих і оптимізаційних алгоритмів, орієнтованих на досягнення високої точності класифікації. Водночас зазначені підходи ґрунтуються переважно на складних комбінованих моделях, що ускладнює інтерпретацію результатів і не завжди дозволяє явно описати логіку прийняття рішень у багатовимірному просторі ознак, що є критично важливим для задач пояснюваної ідентифікації станів комп'ютерних систем.

У попередніх дослідженнях авторів було закладено теоретичні та методологічні основи застосування дерев рішень із багатовимірними вузлами для задач ідентифікації стану комп'ютерних систем. У роботі [15] запропоновано підхід до формування багатовимірних областей прийняття рішень, що дозволив перейти від ортогональних розділень простору ознак до більш гнучкого опису меж між класами. Подальший розвиток цієї ідеї отримано в дослідженнях [16], де для налаштування структури дерева та параметрів вузлів застосовано метаевристичну оптимізацію, що сприяло підвищенню узагальнюючої здатності моделей. У роботах [17] увагу зосереджено на проблемі дисбалансу навчальних даних під час класифікації атак, що дозволило адаптувати деревні моделі до умов нерівномірного представлення станів та зменшити кількість помилок для міноритарних класів.

Важливим напрямом розвитку стало вдосконалення процедур оцінювання якості класифікації та інтеграція ансамблевих механізмів прийняття рішень. У роботі [18] запропоновано розширений апарат аналізу ефективності моделей на основі багатовимірного подання матриці помилок і ROC-орієнтованих характеристик, що забезпечило більш глибоке розуміння поведінки класифікаторів у задачах із високою ціною помилки. Подальше поєднання дерев з багатовимірними вузлами та нечіткої логіки реалізовано в дослідженні [19], де стекінгова організація ансамблю дозволила підвищити стабільність і точність ідентифікації станів у складних умовах невизначеності. Отримані результати сформували підґрунтя для подальших досліджень, спрямованих на удосконалення методів побудови самих багатовимірних вузлів та підвищення адаптивності деревних структур до внутрішньої геометрії даних.

З огляду на це актуальною науковою задачею є подальший розвиток методів побудови дерев з багатовимірними вузлами рішень шляхом удосконалення алгоритмів формування таких вузлів, підвищення їх адаптивності до структури даних та зменшення чутливості до шумових впливів. Особливого значення

набуває створення підходів, здатних ефективно працювати з великими масивами телеметричних і системних даних, що описують функціонування комп'ютерних систем у реальних умовах експлуатації.

Постановка проблеми. У цій роботі пропонується удосконалений метод побудови дерев з багатовимірними вузлами рішень, орієнтований на підвищення точності ідентифікації стану комп'ютерних систем. Запропонований підхід спрямований на більш адекватне моделювання складних областей розділення у багатовимірному просторі ознак та врахування внутрішньої структури даних під час формування вузлів дерева. Удосконалення методу полягає у зміні процедури пошуку параметрів багатовимірних вузлів, що дозволяє зменшити класифікаційну помилку та підвищити стійкість моделі до варіацій вхідних даних.

Об'єктом дослідження є процес ідентифікації стану комп'ютерних систем на основі аналізу багатовимірних даних їх функціонування.

Предметом дослідження є методи побудови дерев рішень із багатовимірними вузлами та алгоритми оптимізації їх структури для задач класифікації станів комп'ютерних систем.

Метою роботи є розробка та дослідження удосконаленого методу побудови дерев з багатовимірними вузлами рішень, який забезпечує підвищення точності та узагальнюючої здатності моделей ідентифікації стану комп'ютерних систем у складних і невизначених умовах.

Досягнення поставленої мети передбачає вирішення таких основних завдань:

1. Аналіз обмежень існуючих підходів до побудови дерев з багатовимірними вузлами;
2. Розробку удосконаленого алгоритму формування вузлів рішень з урахуванням структури даних;
3. Експериментальну оцінку ефективності запропонованого методу на реальних наборах даних, що описують нормальні та аномальні стани комп'ютерних систем;
4. Порівняння отриманих результатів з класичними деревними методами та раніше запропонованими підходами.

Реалізація цих завдань дозволить обґрунтувати доцільність застосування удосконаленого методу та визначити перспективи його подальшого розвитку і практичного використання.

Удосконалений метод побудови дерев з багатовимірними вузлами рішень

Як модифікацію раніше запропонованого підходу [15] пропонується розширити спосіб формування багатовимірних вузлів шляхом використання різних типів гіперфігур, не обмежуючись лише гіперсферичними областями. Такий підхід забезпечує більш гнучке моделювання меж між класами у просторі ознак та дозволяє точніше відображати реальну структуру розподілу даних.

Застосування узагальнених геометричних форм потенційно дає змогу зменшити кількість вузлів і глибину дерева, оскільки складні області можуть бути описані меншою кількістю розділень. Це, у свою чергу, сприяє підвищенню узагальнюючої здатності моделі та покращує інтерпретованість прийнятих рішень за рахунок формування компактнішої структури класифікатора.

Водночас підвищення гнучкості опису супроводжується зростанням обчислювальної складності етапу навчання. Пошук оптимального вигляду гіперфігури та визначення її параметрів стають значно складнішими порівняно з випадком гіперсфери, для якої необхідно оцінити лише координати центру та радіус. Для складніших моделей меж розділення, зокрема поліноміальних або квадратичних поверхонь, кількість параметрів суттєво зростає, що призводить до ускладнення процедури оптимізації та збільшення часу навчання.

Таким чином, запропоноване удосконалення передбачає компроміс між точністю опису багатовимірного простору ознак і витратами на побудову моделі, що потребує розробки ефективних алгоритмів пошуку параметрів та стратегій обмеження складності гіперфігур.

В роботі використовуються такі рівняння:

1. Гіперплощина (1):

$$P(X,1) \equiv \left(\sum_{i=1}^{N} a_i x_i + d \le 0 \right). \quad (1)$$

2. Гіперсфера (2):

$$P(X,2) \equiv \left(\sum_{i=1}^{N} (x_i - c_i) \le r^2 \right). \quad (2)$$

3. Гіпереліпсоїд (3):

$$P(X,3) \equiv \left(\sum_{i=1}^{N} \frac{(x_i - c_i)^2}{a_i^2} \le 1 \right). \quad (3)$$

4. Параболоїд (4):

$$P(X,4) \equiv \left(x_N - \sum_{i=1}^{N-1} a_i {x_i}^2 - d \le 0 \right). \quad (4)$$

5. Поліноміальна крива (5):

$$P(X,5) \equiv \left(\sum_{i=1}^{N} \sum_{j=1}^{M} a_{ji} {x_i}^j + d \le 0 \right). \quad (5)$$

де: X – об'єкт що описується сукупністю ознак x_i в гіперпросторі; k – *індекс рівняння з множини {1,2,3,4,5};* $P(X, k)$ – предикат; a_i, d, r, c_i – параметри рівнянь, що зберігаються незалежно один від одного в системі; N – кількість ознак; M – ступінь поліному.

Процедура формування багатовимірного вузла передбачає автоматизований вибір типу розділяючої гіперфігури, яка найкраще відображає локальну структуру даних та забезпечує мінімальне значення функції помилки класифікації $E(A)$.

$$E(A) = \sum_{i=1}^{N} [y_i \ne t_i]' ; \quad (6)$$

$$[y_i \ne t_i]' = \begin{cases} 1.5\,, & y_i \ne t_i \cap t_i = 1; \\ 1\,, & y_i \ne t_i \cap t_i = 0\,; \\ 0\,, & y_i = t_i. \end{cases} \quad (7)$$

Крок 1. Формування підвибірки вузла. Із навчальної множини виділяється підмножина зразків TS_v, що потрапили до поточного вузла дерева відповідно до рішень попередніх розділень.

Крок 2. Ініціалізація набору моделей-кандидатів. Задається множина можливих типів рівнянь $\Theta = \{HPlane,\ HSph, HEli, HPara, HPoly\}$.

Крок 3. Вибір розмірності на простору. Правила вибору базуються на використання на основі методу кореляційних плеяд – обчислюється повна матриця кореляційних коефіцієнтів, будується граф, в якому ребра відображають в якості вагів абсолютні значення кореляційних коефіцієнтів та проводиться фільтрація розроблених методів. Після чого визначається потенційні багатовимірні ознаки.

Крок 4. Оцінювання параметрів кожного типу. Для кожної фігури $\theta_k \in \Theta$ виконується процедура підбору параметрів (A, C, r, d) у поточному вузлі.

Безпосередній перебір параметрів у багатовимірному просторі є обчислювально складною задачею, оскільки кількість можливих комбінацій експоненційно зростає зі збільшенням числа ознак та порядку рівняння. Тому в запропонованому методі використовується ітераційна процедура спрямованого звуження області пошуку, що базується на принципі дихотомії.

Крок 4.1. Ініціалізація інтервалів. Для кожного параметра задається початковий діпазон допустимих значень

$$p_j \in [L_j, U_j]. \tag{8}$$

Межі визначаються на основі статистичних характеристик підвибірки (мінімальних, максимальних значень та дисперсії).

Крок 4.2. Обчислення центральної точки. На поточній ітерації для кожного параметра визначається середина інтервалу:

$$p_j^{mid} = \frac{L_j + U_j}{2}. \tag{9}$$

На основі отриманого набору параметрів формується кандидатна гіперфігура та виконується розділення зразків.

Крок 4.3. Оцінювання якості. Для сформованого розділення розраховується значення функції помилки *E(A)*

Крок 4.4. Звуження області пошуку. Для кожного параметру визначається напрям покращення. Якщо зміщення параметру в одну з половин інтервалу призводить до зменшення помилки, обирається відповідна частина інтервалу. Таким чином, після кожної ітерації ширина діапазону пошуку зменшується у два рази:

$$[L_j, U_j] \to \frac{1}{2}(U_j - L_j). \tag{10}$$

Крок 4.5. Перевірка критерію зупинки. Ітерації продовжуються до виконання однієї з умов:

1. Інтервал став меншим за задану точність ε.
2. Зміна функції помилки стала незначною або не змінюється декілька ітерацій поспіль;
3. Досягнуто максимальну кількість ітерацій.

У результаті отримується набір параметрів, що забезпечує локально оптимальне розділення у поточному вузлу.

Крок 5. Вибір найкращого типу гіперфігури. Після завершення процедури оптимізації для кожної фігури $\theta_k \in \Theta$ отримується набір параметрів, що забезпечує мінімальне значення функції помилки у межах розглянутих інтервалів пошуку. На цьому етапі необхідно визначити модель, яка забезпечує найкращу якість розділення з урахуванням не лише точності, але й складності опису, яка призводить до зменшення швидкодії результуючої моделі.

Крок 5.1. Обчислення узагальненого критерію якості. Для кожної фігури формується інтегральна оцінка

$$Q_k = E_k + \lambda \cdot C_k\,, \tag{11}$$

де E_k – значення помилки класифікації для k-ї фігури; C_k – показник складності моделі; λ – коефіцієнт регуляції.

Крок 5.2. Оцінка складності. Складність може визначатись як кількість параметрів у рівнянні, порядок поліному (при k=5) та необхідна кількість операцій для перевірки предикату. Це дозволяє віддавати перевагу простішим моделям у випадках, коли вони демонструють близьку якість розподілення між собою.

Крок 5.3. Вибір оптимальної моделі. Тип гіперфигури визначається за правилом:

$$\theta^* = arg\min_k Q_k. \tag{12}$$

Крок 5.4. Перевірка доцільності поділу. Якщо отримане покращення відносно батьківського вузла є незначним або кількість зразків у підмножинах стає меншою за встановлений поріг, або досягнуто розподіл з нульовим значенням помилки, подальше розгалуження припиняється, а такий вузол оголошується листом.

Крок 6. Формування дочірніх вузлів та рекурсивне продовження побудови дерева. Після визначення оптимального типу гіперфігури θ^* та відповідного набору параметрів виконується безпосереднє розділення підвибірки поточного вузла на нові підмножини відповідно до значення предикату *P(X,k)*.

Крок 6.1. Розподіл зразків. Кожен об'єкт X_i із підмножини TS_v перевіряється на виконання умови належності до області, визначеної обраною гіперфігурою. У результаті формуються дві підмножини: ліва гілка для зразків з істинним значенням предикату та права гілка для яких умова не виконується.

Крок 6.2. Проведення аналізу отриманих підмножин. Обчислюється кількість елементів для кожної гілки, однорідність класів та можливість подальшого зменшення помилки.

Крок 6.3. Перевірка критеріїв зупинки. Рекурсивне розгалуження припиняється, якщо виконується хоча б одна з умов:

1. Досягнуто максимально допустиму глибину дерева;
2. Кількість зразків у вузлах менша за порогове значення;
3. Зменшення функції помилки є незначним.
4. Усі об'єкти належать одному класу.

У цьому випадку вузол оголошується листом, а його клас визначається за правилом більшості.

Крок 6.4. Рекурсивний виклик. Якщо умови зупинки не виконуються, для кожної підмножини процеду-

ра вибору типу гіперфігури та її параметрів повторюється, починаючи з Кроку 2. Виклики додаються в чергу з пріоритетом на обсяг помилки – чим більше значення помилки тим пріоритетніше вузол дерева.

Аналіз обчислювальної складності запропонованого методу та очікуваний вплив на показники якості класифікації

Запропонований метод побудови дерев із багатовимірними вузлами рішень передбачає виконання кількох вкладених процедур, основними з яких є вибір типу гіперфігури, оптимізація її параметрів та рекурсивне формування структури дерева. Обчислювальна складність алгоритму визначається кількістю об'єктів навчальної вибірки, розмірністю простору ознак, числом кандидатних моделей і необхідною точністю підбору параметрів.

Нехай N – кількість зразків у поточному вузлі, M – кількість ознак, K – число можливих типів гіперфігур, P_k – кількість параметрів для моделі типу k.

Складність оцінювання предиката. Перевірка належності одного зразка до області визначеної гіперфігурою, потребує $O(M)$ або $O(P_k)$ операцій залежно від виду рівняння. Відповідно оцінювання розділення для всіх об'єктів вузла має складність $O(N \cdot Pk)$.

Складність підбору параметрів. У процедурі оптимізації використовується ітераційне звуження інтервалів пошуку. Якщо для кожного параметра потрібно досягти точності ε, а початкова ширина інтервалу становить D, кількість ітерацій для одного параметра дорівнюватиме:

$$I = \log_2(D/\varepsilon)\ . \qquad (13)$$

Тоді повна складність пошуку параметрів однієї гіперфігури може бути оцінена як:

$$O(I \cdot N \cdot P_k). \qquad (14)$$

Складність вибору типу фігури. Оскільки процедура виконується для кожного з K кандидатів, сумарні витрати становитимуть:

$$O(K \cdot I \cdot N \cdot P_k). \qquad (15)$$

Рекурсивна складність побудови дерева. Глибина дерева у середньому є логарифмічною відносно кількості зразків. Тому повну складність навчання можна наближено оцінити як:

$$O(K \cdot I \cdot P_k \cdot N \cdot \log N). \qquad (16)$$

Отримані співвідношення свідчать, що основні витрати пов'язані з оптимізацією параметрів багатовимірних вузлів. Водночас використання процедури дихотомічного звуження інтервалів забезпечує логарифмічну залежність від точності пошуку, що суттєво зменшує обчислювальні витрати порівняно з повним перебором. Крім того, зменшення кількості вузлів і глибини дерева завдяки більш точному моделюванню меж між класами частково компенсує зростання складності окремого розділення. Таким чином, запропонований підхід забезпечує прийнятний баланс між точністю класифікації та ресурсами, необхідними для навчання моделі.

В табл. 1 наведено порівняльний аналіз обчислювальної складності методів побудови дерев рішень. Запропоноване удосконалення процедури формування багатовимірних вузлів дозволяє сформулювати низку теоретичних положень щодо очікуваної поведінки моделі під час навчання та узагальнення.

Таблиця 1 – **Порівняльний аналіз з класичними методами та першою версією**

Характеристика	Класичні методи побудови дерев рішень (одновимірні вузли)	Метод побудови дерев з багатовимірними вузлами рішень	Модифікований метод побудови дерев з багатовимірними вузлами рішень
Тип розділення простору ознак	Порогове значення по одній ознаці	Гіперсферами або пороговим значенням по одній ознаці	Адаптивний вибір гіперфігури (площина, сфера, поліном тощо)
Кількість параметрів у вузлі	1 поріг	Центр та радіус (≈ M)	Залежить від типу фігури, може перевищувати M
Пошук параметрів	Лінійний або бінарний по відсортованих значеннях	Ітераційна оптимізація	Ітераційна оптимізація з вибором моделі
Вартість оцінювання одного кандидата	$O(N)$	$O(N \cdot M)$	$O(N \cdot P_k)$
Кількість кандидатів у вузлі	≈ M	1 тип фігури	K типів фігур
Складність оптимізації	Низька	Середня	Вища, але контрольована регуляризацією
Типова глибина дерева	Відносно велика	Менша	Найменша за рахунок гнучких меж
Ризик перенавчання	Високий при великих глибинах	Помірний	Контрольований через штраф складності
Інтерпретованість	Висока	Висока	Зберігається, але залежить від типу поверхні
Орієнтовна складність навчання	$O(M \cdot N \cdot \log N)$	$O(I \cdot M \cdot N \cdot \log N)$	$O(K \cdot I \cdot P_k \cdot N \cdot \log N)$
Компенсація зростання складності	Відсутня	Менша кількість вузлів	Менша кількість вузлів та краща якість розділення

Розширення класу можливих розділяючих поверхонь підвищує ймовірність побудови більш точного локального поділу простору ознак у кожному вузлі. Унаслідок цього для досягнення заданого рівня по-

милки може знадобитися менша кількість рекурсивних розгалужень у порівнянні з одновимірними розділеннями.

Використання нелінійних гіперфігур забезпечує кращу відповідність геометрії розділення реальному розподілу даних. Це особливо важливо у випадках наявності корельованих ознак або складних форм кластерів, де ортогональні поділи призводять до фрагментації простору.

Хоча кількість параметрів у вузлі збільшується, зменшення числа самих вузлів і глибини дерева частково або повністю компенсує додаткові витрати. Таким чином, загальна складність моделі може зростати повільніше, ніж складність окремого розділення.

Більш адекватне моделювання локальної структури даних сприяє зменшенню потреби у глибоких ієрархіях та великій кількості послідовних розщеплень. Це знижує ризик накопичення локальних помилок і позитивно впливає на стабільність моделі на тестових вибірках.

Незважаючи на ускладнення математичного опису вузлів, рішення все ще можуть бути представлені у вигляді явних аналітичних умов належності до області. Це дозволяє зберігати властивості інтерпретації результату, притаманні деревним моделям.

Очікуваний вплив удосконаленого методу на показники якості класифікації. Запропоноване розширення класу розділяючих поверхонь у багатовимірних вузлах рішень створює передумови для покращення основних характеристик ефективності класифікації. Теоретичні міркування, наведені у попередньому підрозділі, дозволяють сформувати низку очікувань щодо поведінки моделі під час експериментальної перевірки.

Першим критерієм розглянемо класичну точність Accuracy. Більш гнучке представлення меж між класами дозволяє зменшити кількість помилково класифікованих зразків у прикордонних областях. За рахунок цього очікується зростання загальної частки правильних передбачень, особливо у випадках складних або перекривних розподілів. Наступним виступають показники повноти та точності позитивних класів (Recall, Precision). Можливість точніше локалізувати області, характерні для аномальних станів, повинна сприяти зменшенню як хибно-негативних, так і хибно-позитивних рішень. Це має позитивно відобразитися на балансі між повнотою виявлення загроз та кількістю помилкових спрацьовувань. Також слід розглянути інтегральні метрики (F_1-score, MCC). Оскільки зазначені показники враховують співвідношення різних типів помилок, їх покращення очікується як наслідок більш узгодженого поділу простору ознак. Особливо це важливо для задач із дисбалансом класів, де локальні неточності можуть суттєво впливати на підсумковий результат.

Крім підвищення оцінок якості слід також розглянути очікування щодо стабільності моделі. Зменшення необхідної глибини дерева та скорочення кількості послідовних розділень мають сприяти зниженню варіативності результатів при зміні навчальних підвибірок. Це дозволяє очікувати кращої відтворюваності та більш передбачуваної поведінки моделі на нових даних. Загалом, очікується, що модифікований метод забезпечить покращення як локальних, так і глобальних характеристик класифікації, зберігаючи інтерпретованість і керованість моделі. Остаточна перевірка сформульованих припущень здійснюється у межах експериментальних досліджень, результати яких наведено в наступних розділах.

Експериментальні дослідження

Експериментальна перевірка запропонованого удосконаленого методу побудови дерев із багатовимірними вузлами рішень спрямована на оцінювання його здатності підвищувати точність ідентифікації станів комп'ютерних систем у порівнянні з класичними та раніше розробленими підходами. Основною метою експериментів є підтвердження того, що розширення класу розділяючих поверхонь дозволяє більш адекватно описувати структуру даних без критичного збільшення обчислювальних витрат.

Для забезпечення коректності порівняння було використано набір даних, сформований у межах попередніх досліджень [15–19] і сумісний із роботами, присвяченими застосуванню дерев із багатовимірними вузлами. Вибір саме цієї вибірки дозволяє безпосередньо оцінити внесок запропонованого удосконалення, виключаючи вплив сторонніх факторів.

Дані відображають функціонування комп'ютерних систем у нормальних та аномальних режимах і включають показники завантаження процесора, використання оперативної пам'яті, активності дискових підсистем, мережевих характеристик та поведінкових параметрів процесів. Такий набір ознак забезпечує багатовимірне представлення стану системи та створює умови для формування складних меж між класами. Початковий масив даних було розділено на начальну та тестову частини. Навчальна вибірка використовується для побудови моделей і оптимізації параметрів вузлів, тоді як тестова – для оцінювання узагальнюючої здатності. Розподіл здійснювався з урахуванням збереження пропорцій між нормальними та аномальними станами. Для визначення ефективності модифікованого підходу проводилось порівняння з такими алгоритмами:

1. Класичне дерево рішень з одновимірними пороговими вузлами (DT);
2. Дерево з багатовимірними вузлами на основі гіперсфер (MDT);
3. Удосконалений метод з адаптивним вибором типу гіперфігури (AMDT, який пропонується в статті).

Таке порівняння дозволить оцінити як абсолютний виграш у якості, а також приріст відносно попередніх етапів розвитку методу.

Для аналізу результатів використовувались показники, що є стандартними для задач класифікації: Accuracy, Precision, Recall, F_1-Score та коефіцієнт кореляції Метьюза (MCC).

Використання набору взаємодоповнюючих метрик дозволяє уникнути викривленої інтерпретації результатів та забезпечить комплексну оцінку поведінки моделі. Оптимізація параметрів гіперфігур у вузлах здійснювалася за процедурою ітераційного звуження області пошуку. Для забезпечення справе-

дливості порівняння всі методи реалізовано в однаковому програмному середовищі та запускались на однакових апаратних ресурсах.

Результати методів з основними метриками якості продемонстровано у табл. 2 та табл. 3.

Таблиця 2 – **Порівняння якості ідентифікації стану КС**

Метод ідентифікації стану КС	Acc	F_1-Score	MCC
DT	0,679	0,647	0,364
MDT	0,909	0,916	0,831
AMDT	0,979	0,980	0,959

Таблиця 3 – **Порівняння точності та повноти**

Метод ідентифікації стану КС	Precision	Recall
DT	0,718	0,589
MDT	0,847	0,998
AMDT	0,963	0,997

Починаючи з узагальнених показників видно стабільне покращення результатів при переході від одновимірних до багатовимірних та далі до адаптивних вузлів. Підвищення повноти свідчить про зменшення пропущених аномалій, тоді як зростання точності вказує на скорочення кількості хибних спрацьовувань. Отримані результати підтверджують, що адаптивний вибір форми розділяючої поверхні дозволяє ефективніше описувати складні області у просторі ознак, ніж використання фіксованої геометрії. Візуалізації отриманих метрик зазначено на рис. 1 для F_1-Score, рис. 2 для MCC а також рис. 3-4 для частки хибних спрацьовувань та пропущених аномалій.

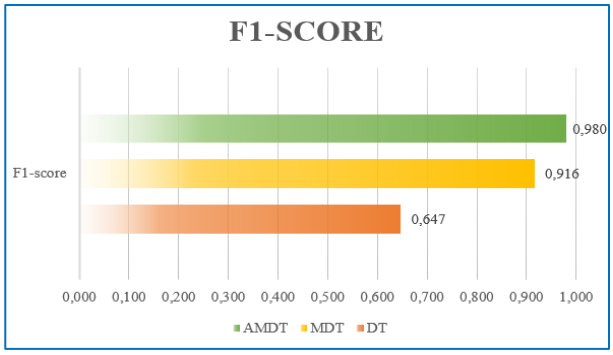

Рис. 1. Порівняння за метрикою F_1-Score

Отримані дані демонструють, що використання складніших гіперфігур дозволяє точніше описувати прикордонні області між класами. У класичних деревах для досягнення подібної точності необхідно значно збільшувати глибину, що призводить до накопичення помилок у нижніх ярусах. Базовий варіант із гіперсферичними вузлами вже забезпечує суттєве покращення, однак обмеження форми не дозволяє повністю адаптуватися до реальної геометрії даних.

Удосконалений метод демонструє найкращі результати завдяки можливості локально підбирати тип поверхні, що мінімізує як помилки класифікації, так і необхідну кількість розгалужень.

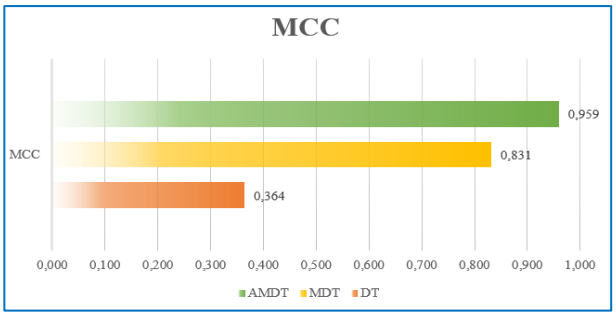

Рис. 2. Порівняння за метрикою кореляційного коефіцієнта Метьюза

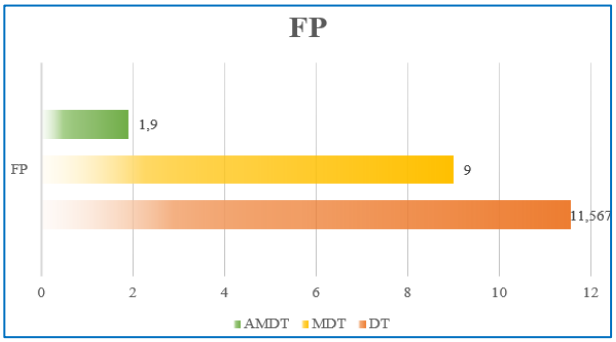

Рис. 3. Порівняння за часткою хибних спрацьовувань або помилок першого роду (у відсотках)

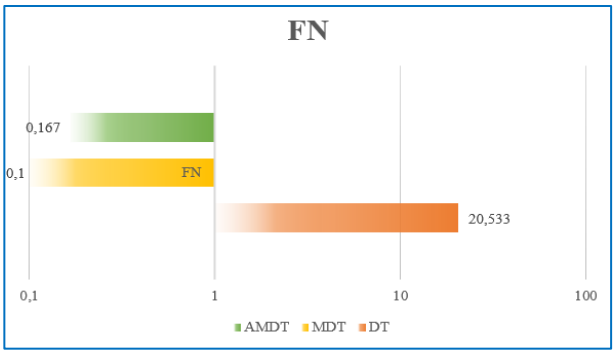

Рис. 4. Порівняння за часткою пропусків загроз або помилок другого роду (у відсотках, логарифмічна шкала)

Оцінка результатів та формування рекомендацій для подальших досліджень

Проведені експериментальні дослідження підтвердили доцільність розширення класу розділяючих поверхонь у багатовимірних вузлах рішень. У порівнянні з класичними підходами та попередньою реалізацією дерев із фіксованою геометрією вузлів запропонований метод продемонстрував стабільне покращення інтегральних показників якості класифікації, зокрема F_1-Score та коефіцієнта кореляції Метьюза.

Найбільший виграш спостерігається у прикордонних областях між класами, де використання гнучкіших поверхонь дозволило зменшити як кількість пропущених аномалій, так і частоту хибних спрацьовувань. Це свідчить про здатність методу точніше враховувати внутрішню геометрію простору ознак і формувати більш узгоджені правила прийняття рішень. Важливим результатом є також зменшення

варіативності моделей при повторних запусках навчання. Скорочення глибини дерев та кількості послідовних поділів дозволило знизити накопичення локальних похибок, що позитивно вплинуло на відтворюваність результатів і стабільність роботи алгоритму на нових даних.

Разом із тим встановлено, що підвищення точності супроводжується зростанням витрат часу на оптимізацію параметрів вузлів. Проте за рахунок використання процедур спрямованого звуження області пошуку та зменшення кількості необхідних розгалужень загальні витрати залишаються прийнятними для практичного застосування в системах моніторингу стану комп'ютерних систем.

Отримані результати дозволяють рекомендувати запропонований метод для використання у задачах, де важливими є інтерпретованість моделей, висока точність виявлення аномалій та стійкість до зміни розподілу даних.

Подальший розвиток підходу може бути спрямований на автоматизацію вибору набору кандидатних гіперфігур, розробку більш ефективних методів оптимізації параметрів у високорозмірних просторах, а також інтеграцію механізмів адаптивного оновлення структури дерева під час надходження нових даних. Перспективним є також дослідження можливостей комбінування запропонованого методу з ансамблевими та гібридними архітектурами, що можуть додатково підвищити стійкість до шумів і невизначеності.

Висновки

У роботі запропоновано удосконалений метод побудови дерев рішень із багатовимірними вузлами для задачі ідентифікації стану комп'ютерних систем. Основною особливістю підходу є розширення класу розділяючих поверхонь та впровадження процедури адаптивного вибору типу гіперфігури у кожному вузлі на основі мінімізації функції помилки з урахуванням складності моделі.

Розроблено алгоритм ітераційного підбору параметрів, що базується на спрямованому звуженні області пошуку та забезпечує контрольоване зростання обчислювальних витрат.

Запропонована стратегія дозволяє ефективно працювати у багатовимірному просторі ознак без необхідності повного перебору можливих конфігурацій.

Теоретичний аналіз показав, що використання гнучкіших геометричних моделей дозволяє зменшити необхідну глибину дерева, покращити апроксимацію меж між класами та знизити ризик накопичення локальних помилок.

При цьому зберігається інтерпретованість рішень, що є важливою вимогою для систем моніторингу та інформаційної безпеки.

Результати експериментальних досліджень підтвердили сформульовані очікування. Удосконалений метод продемонстрував підвищення показників Accuracy (0.97), Precision (0.96), Recall (0.99), F1-score (0.98) та MCC (0.96) у порівнянні з класичними деревами та базовою реалізацією багатовимірних вузлів.

Отже, запропонований підхід є перспективним для практичного використання у задачах автоматизованої ідентифікації станів комп'ютерних систем, особливо в умовах складної структури даних, наявності шумів та перекривання класів.

Конфлікт інтересів

Автори декларують, що не мають конфлікту інтересів стосовно даного дослідження, в тому числі фінансового, особистісного характеру, авторства чи іншого характеру, що міг би вплинути на дослідження та його результати, представлені в даній статті.

Використання засобів штучного інтелекту

Автори підтверджують, що не використовували технології штучного інтелекту при створенні представленої роботи.

Список літератури

1. K. Thirasak, T. Chuaphanngam, D. Chainarong and S. Fugkeaw, "TF2ML: Threat Filtering With Two-Stage Machine Learning for Efficient Provenance-Aware Threat Detection and Response," in *IEEE Open Journal of the Computer Society*, vol. 6, pp. 1751-1762, 2025, doi: https://doi.org/10.1109/OJCS.2025.3618157
2. A. T. Haile, S. L. Abebe and H. M. Melaku, "Real-Time Automated Cyber Threat Classification and Emerging Threat Detection Framework," in *IEEE Open Journal of the Computer Society*, vol. 6, pp. 921-930, 2025, doi: https://doi.org/10.1109/OJCS.2025.3580235
3. O. Cherqi, Y. Moukafih, M. Ghogho and H. Benbrahim, "Enhancing Cyber Threat Identification in Open-Source Intelligence Feeds Through an Improved Semi-Supervised Generative Adversarial Learning Approach With Contrastive Learning," in *IEEE Access*, vol. 11, pp. 84440-84452, 2023, doi: https://doi.org/10.1109/ACCESS.2023.3299604
4. M. Zaman, D. Upadhyay and C. -H. Lung, "Validation of a Machine Learning-Based IDS Design Framework Using ORNL Datasets for Power System With SCADA," in *IEEE Access*, vol. 11, pp. 118414-118426, 2023, doi: https://doi.org/10.1109/ACCESS.2023.3326751
5. A. Al Siam, M. Alazab, A. Awajan and N. Faruqui, "A Comprehensive Review of AI's Current Impact and Future Prospects in Cybersecurity," in *IEEE Access*, vol. 13, pp. 14029-14050, 2025, doi: https://doi.org/10.1109/ACCESS.2025.3528114
6. P. Kumar, R. Kumar, A. Jolfaei and N. Mohammad, "An Automated Threat Intelligence Framework for Vehicle–Road Cooperation Systems," in *IEEE Internet of Things Journal*, vol. 11, no. 22, pp. 35964-35974, 15 Nov.15, 2024, doi: https://doi.org/10.1109/JIOT.2024.3397652.
7. T. A. Al-Shehari *et al*., "Enhancing Insider Threat Detection in Imbalanced Cybersecurity Settings Using the Density-Based Local Outlier Factor Algorithm," in *IEEE Access*, vol. 12, pp. 34820-34834, 2024, doi: 10.1109/ACCESS.2024.3373694
8. A. M. Almasabi, M. Khemakhem, F. E. Eassa, A. Ahmed Abi Sen, A. B. Alkhodre and A. Harbaoui, "A Smart Framework to Detect Threats and Protect Data of IoT Based on Machine Learning," in *IEEE Access*, vol. 12, pp. 176833-176844, 2024, doi: https://doi.org/10.1109/ACCESS.2024.3498603

9. A. Algarni, T. Acarer and Z. Ahmad, "An Edge Computing-Based Preventive Framework With Machine Learning- Integration for Anomaly Detection and Risk Management in Maritime Wireless Communications," in *IEEE Access*, vol. 12, pp. 53646-53663, 2024, doi: https://doi.org/10.1109/ACCESS.2024.3387529
10. R. Marinho and R. Holanda, "Automated Emerging Cyber Threat Identification and Profiling Based on Natural Language Processing," in *IEEE Access*, vol. 11, pp. 58915-58936, 2023, doi: https://doi.org/10.1109/ACCESS.2023.3260020
11. S. Lee, A. Abdulboriy Abdulkhay Ugli and J. S. Shin, "More Realistic Audio-Based Drone Detection and Identification Approaches With Machine Learning," in *IEEE Access*, vol. 13, pp. 170328-170350, 2025, doi: https://doi.org/10.1109/ACCESS.2025.3613683
12. K. Xia *et al*., "Privacy-Preserving Brain–Computer Interfaces: A Systematic Review," in *IEEE Transactions on Computational Social Systems*, vol. 10, no. 5, pp. 2312-2324, Oct. 2023, doi: https://doi.org/10.1109/TCSS.2022.3184818
13. M. Mehmood, R. Amin, M. M. A. Muslam, J. Xie and H. Aldabbas, "Privilege Escalation Attack Detection and Mitigation in Cloud Using Machine Learning," in *IEEE Access*, vol. 11, pp. 46561-46576, 2023, doi: https://doi.org/10.1109/ACCESS.2023.3273895
14. R. Allafi and I. R. Alzahrani, "Enhancing Cybersecurity in the Internet of Things Environment Using Artificial Orca Algorithm and Ensemble Learning Model," in *IEEE Access*, vol. 12, pp. 63282-63291, 2024, doi: https://doi.org/10.1109/ACCESS.2024.3390093
15. S. Y. Gavrylenko, V. V. Chelak, S. G. Semenov Development of Method for Identification the Computer System State based on the Decision Tree with Multi-Dimensional Nodes / Radio Electronics, Computer Science, Control (RECSC), Zaporizhzhia, No. 2 (2022), P. 113-122, doi: https://doi.org/10.15588/1607-3274-2022-2-11
16. Chelak V., Hornostal O., Chelak Y., Gavrylenko S. "Decision Tree Construction Method using Cuckoo Search for Computer System State Identification", 2025 IEEE 5th KhPI Week on Advanced Technology (KhPIWeek), Kharkiv, Ukraine, 2025 doi: https://doi.org/10.1109/KhPIWeek61436.2025.11288613
17. О. Горносталь, В. Челак. Класифікація мережевих атак методами машинного навчання в умовах дисбалансу тренувальних даних / Системи управління, навігації та зв'язку, Полтава, 2025, Том 3 (81), С. 64-71, doi: https://doi.org/10.26906/SUNZ.2025.3.064
18. V. Chelak, O. Hornostal, Ye. Chelak, S. Gavrylenko. Advanced Methods for Classification Quality Assessment Leveraging ROC Analysis and Multidimensional Confusion Matrix. Advanced Information Systems, 2025, Vol 9(1), pp. 24–34, doi: https://doi.org/10.20998/2522-9052.2025.1.03
19. В. Челак, О. Горносталь, Нечіткий ансамбль дерев рішень для ідентифікації стану комп'ютерних систем/ Системи управління, навігації та зв'язку, Полтава, 2025, Том 4 (82), С. 144-150, doi: https://doi.org/10.26906/SUNZ.2025.4.144

Received (Надійшла) 03.01.2026
Accepted for publication (Прийнята до друку) 11.02.2026
Publication date (Дата публікації) 27.02.2026

ВІДОМОСТІ ПРО АВТОРІВ / ABOUT THE AUTHORS

Челак Віктор Володимирович – PhD, доцент кафедри "Комп'ютерна інженерія та програмування", Національний технічний університет "Харківський політехнічний інститут", Харків, Україна;
Viktor Chelak – PhD, Associate Professor of Department of "Computer Engineering and Programming", National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine;
e-mail: victor.chelak@gmail.com; ORCID ID: https://orcid.org/0000-0001-8810-3394;
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=57216331944&origin=resultslist.

Горносталь Олексій Андрійович – PhD, асистент кафедри "Комп'ютерна інженерія та програмування", Національний технічний університет "Харківський політехнічний інститут", Харків, Україна;
Oleksii Hornostal – PhD, Assistant Professor of Department of "Computer Engineering and Programming", National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine;
e-mail: gornostalaa@gmail.com; ORCID ID: https://orcid.org/0000-0001-5820-9999;
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=57189040595.

Improved method for constructing trees with multidimensional decision nodes for computer system state identification

Viktor Chelak, Oleksii Hornostal

Abstract. The object of the research is the process of computer system state identification based on the analysis of multidimensional operational data. **The subject of the research** is methods for constructing decision trees with multidimensional decision nodes and algorithms for optimizing their structure in computer system state classification tasks. **The goal of the research** is to develop and investigate an improved method for constructing decision trees with multidimensional decision nodes that provides higher accuracy and better generalization ability of computer system state identification models under complex and uncertain conditions. **Methods**: machine learning techniques, tree-based classification models, nonlinear space partitioning procedures, iterative parameter optimization based on search interval reduction, and classification quality assessment approaches. **Results**: a procedure for automated selection of the hyperfigure type within a tree node has been developed, enabling the adaptation of the decision boundary to the local structure of data. It has been shown that the use of generalized surfaces reduces the required tree depth and decreases the number of consecutive splits. Experimental studies demonstrate improvements in Accuracy, Precision, Recall, F1-score, and MCC compared with classical decision trees and models with fixed node geometry. **Conclusions**: the proposed method provides more accurate and stable modeling of boundaries between computer system states while preserving interpretability and maintaining acceptable computational costs. The obtained results confirm the feasibility of applying adaptive multidimensional nodes in monitoring and information security systems.

Keywords: state identification, computer systems, decision tree, multidimensional node, nonlinear partitioning, decisive rule, machine learning.