



DOI 10.28925/2663-4023.2026.33.1158

УДК 004.056:[004.8:004.4]

Шляхова Анастасія Станіславівна

студент

Інститут спеціального зв'язку та захисту інформації

Національного технічного університету України

“Київський політехнічний інститут імені Ігоря Сікорського”, Київ, Україна

ORCID: 0009-0009-5926-2105

slahovaanastasia33@gmail.com

Шевчук Ольга Сергіївна

викладач

Інститут спеціального зв'язку та захисту інформації

Національного технічного університету України

“Київський політехнічний інститут імені Ігоря Сікорського”, Київ, Україна

ORCID: 0000-0002-2866-439X

olia13511@gmail.com

Онiщенко Володимир Олександрович

викладач

Інститут спеціального зв'язку та захисту інформації

Національного технічного університету України

“Київський політехнічний інститут імені Ігоря Сікорського”, Київ, Україна

ORCID: 0009-0000-1355-9178

v.o.onishchenko@ukr.net

ВИЯВЛЕННЯ КІБЕРАТАК У МЕРЕЖЕВОМУ ТРАФІКУ НА ОСНОВІ АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ

Анотація. У статті розглянуто роль та значення методів машинного навчання (ML) як методологічного фундаменту сучасного штучного інтелекту. Обґрунтовано, що здатність інформаційних систем до самонавчання та адаптації в динамічних середовищах є ключовим фактором їх ефективності. Проаналізовано стрімке зростання попиту на ML-технології в усіх сферах людської діяльності, що неминуче призводить до акумуляції та обробки колосальних обсягів чутливої інформації. Концентрація таких даних створює нові вектори загроз, оскільки стає пріоритетною ціллю для кіберзловмисників. Особливу увагу приділено впровадженню ML-алгоритмів у сучасні екосистеми безпеки. Розглянуто досвід використання провідних індустріальних рішень, що замінюють традиційні сигнатурні підходи інтелектуальним аналізом. Детально описано механізми реалізації мережових атак, спрямованих на підміну початкових даних та маніпуляцію процесом навчання. Результати проведених експериментальних досліджень підтверджують, що використання неактуальних або скомпрометованих моделей у системах кіберзахисту створює ілюзію безпеки, залишаючи критичну інфраструктуру вразливою до цілеспрямованих атак. Стаття пропонує концептуальний погляд на необхідність створення захищених протоколів навчання для забезпечення стійкості інтелектуальних систем.

Ключові слова: кібербезпека; машинне навчання; системи виявлення вторгнень; виявлення аномалій; виявлення новизни; моніторинг у реальному часі.

ВСТУП

Сучасне інформаційне суспільство характеризується стрімкою цифровізацією всіх сфер діяльності, що супроводжується активним впровадженням інформаційно-комунікаційних технологій. Підвищення рівня цифрової взаємодії та довіри в соціальному середовищі водночас зумовлює зростання ролі людського фактору, який розглядається як одна з найбільш критичних вразливостей у ландшафті кіберзагроз [1]. Унаслідок цього людський фактор дедалі частіше стає об'єктом цілеспрямованих кібератак, що охоплюють усі типи інфокомунікаційних мереж. У свою чергу Україна разом зі США, Південною Кореєю та Китаєм входить до четвірки країн, які зазнають найбільшої кількості кібератак [2].



За даними IT Ukraine Association, у 2022 році на Україну було здійснено 2194 кібератаки, у 2023-2544, а у 2024 році їх кількість зросла до 4315 [2]. З початку 2025 року CERT-UA фіксує в середньому близько 15 кіберінцидентів на добу та здійснює моніторинг понад 150 кластерів кіберзагроз (UAC) [3].

Постановка проблеми. Одним із найбільш поширених та впливових напрямів сучасного технологічного розвитку став штучний інтелект, який за відносно короткий період привернув значну увагу та здобув високий рівень довіри з боку користувачів. Водночас поява й активне впровадження нових технологій у кіберпросторі супроводжуються формуванням принципово нових загроз. Завдяки широкій доступності інструментів штучного інтелекту вони активно використовуються не лише легітимними користувачами, але й кіберзловмисниками. Зокрема, спостерігається поява нових рівнів фішингових атак, реалізованих шляхом імітації людського голосу різними методами, застосування інтелектуальних чат-ботів, автоматизованого генерування шкідливого програмного коду, у тому числі варіантів шкідливих PowerShell-скриптів, а також створення вірусів зі змінною структурою, адаптованою до середовища конкретної жертви (DeepLocker, IBM Research, 2018-2019). Окрему загрозу становлять автономні системи проникнення, засновані на методах навчання з підкріпленням (RL-based Autonomous Cyber Agents, DARPA, 2020-2023), маскування шкідливих команд під звичайну адміністративну активність (SunBurst / SolarWinds, 2020 – частково AI-driven), а також генерування мультимедійного та текстового контенту з використанням засобів штучного інтелекту.

Використання штучного інтелекту в зазначених кібератаках та методах їх реалізації на сучасному етапі правового регулювання тісно пов'язане з питаннями захисту персональних даних і правового режиму об'єктів інтелектуальної власності. Хоча відповідні відносини регулюються окремими нормативно-правовими законами [4, 5], чинні правові норми часто не встигають за темпами технологічного розвитку, зокрема у сфері штучного інтелекту та великих даних, що призводить до виникнення правових прогалів і зумовлює потребу подальшого законодавчого вдосконалення. Особливої актуальності ця проблема набуває з огляду на ризики неконтрольованого доступу до великих масивів даних про майновий стан і родинні зв'язки громадян, що може створювати загрози національній безпеці та використовуватися іноземними спеціальними службами з недоброчесними цілями.

Методологічною основою штучного інтелекту є методи машинного навчання, які забезпечують здатність інформаційних систем до самонавчання та адаптації до динамічних умов функціонування середовища. Попит на застосування таких методів охоплює практично всі сфери людської діяльності, що зумовлює обробку значних обсягів чутливої інформації, яка, у свою чергу, становить підвищений інтерес для кіберзловмисників. Алгоритми машинного навчання (Machine Learning, ML) використовуються компаніях для побудови моделей на основі даних з метою автоматизованого та незалежного прийняття рішень. У сучасних системах кіберзахисту методи ML дедалі частіше замінюють або доповнюють традиційні підходи до аналізу безпеки. Зокрема, рутинний моніторинг мережевого трафіку на наявність аномалій реалізується за допомогою інструментів типу Cisco Secure Network Analytics, аналіз журналів подій – із використанням Splunk with ML Toolkit, виявлення шкідливого програмного забезпечення – засобами Microsoft Defender for Endpoint, класифікація спаму та фішингових повідомлень – у межах Google Workspace Security, а пріоритизація інцидентів – у платформах класу Cortex XDR (Palo Alto Networks), а також у низці інших IDS/IPS-рішень.

Водночас загрози для таких систем можуть виникати вже на етапах навчання або тестування моделей машинного навчання. Реалізація різних типів мережевих атак, зокрема підміна початкових даних або маніпуляція навчальними наборами даних (datasets), здатна призвести до обману моделі, її перенавчання, некоректної класифікації подій та хибної оцінки показників точності. Як наслідок, відбувається використання неактуальних або ненадійних моделей машинного навчання в системах кіберзахисту, що підтверджено результатами експериментальних досліджень.

Аналіз останніх досліджень і публікацій. Експериментальне наукове дослідження Reshamlal Pradhan (2022) опубліковане в журналі Neuroquantology, показало застосування алгоритмів Decision Tree для побудови системи виявлення вторгнень (ML IDS) [6]. У роботі розглянуто використання алгоритмів Decision Tree на основі датасету CICIDS2017, який є одним із найбільш репрезентативних наборів мережевого трафіку та містить як нормальну активність, так і різні типи атак, зокрема DDoS, Botnet та Web Attack. Зазначений датасет широко використовується для оцінювання ефективності методів машинного навчання в задачах IDS. Для навчання та тестування моделей автор застосовував програмний інструмент Weka, використовуючи різні реалізації дерев рішень (J48, REP Tree, Random Tree). Результати дослідження засвідчили здатність моделей Decision Tree до швидкої сегментації мережевого трафіку та чіткого розмежування між легітимною активністю і характерними патернами атак (DDoS, Botnet, Web Attack) завдяки ієрархічній структурі прийняття рішень. Під час тестування на CICIDS2017 моделі, зокрема J48 та Random Tree, продемонстрували високу адаптивність до складної структури даних і стабільно низький рівень хибнопозитивних спрацювань. Загальна точність



класифікації атак DDoS, Botnet та Web Attack становила близько 98–99 %, що підтверджується високими значеннями Precision (точність), Recall (повнота/чутливість) та F1-score (баланс між точністю і повнотою). Отримані результати свідчать про високу ефективність методів Decision Tree у задачах IDS при роботі з реалістичними мережевими даними.

Водночас, попри високі показники точності, системи машинного навчання залишаються вразливими до цілеспрямованих атак на етапах навчання та експлуатації. Так, у дослідженні Alshahrani E., Alghazzawi D., Alotaibi R. та Rabie O. (2022) було проаналізовано вплив GAN-базованих атак типу evasion та poisoning на системи IDS, побудовані з використанням алгоритмів Decision Tree (DT) та Logistic Regression (LR) [7].

Evasion-атаки реалізуються на фазі експлуатації моделі шляхом модифікації шкідливого трафіку з метою його маскування під легітимну активність, тоді як poisoning-атаки здійснюються на фазі навчання через навмисну підміну або «отруєння» навчальних даних. Результати експерименту показали різну чутливість моделей до таких впливів: алгоритм Decision Tree зазнав більшого зниження точності під час обробки evasion-трафіку, тоді як Logistic Regression продемонструвала значніше падіння точності при poisoning-атаках. Зокрема, evasion-атаки знижували точність DT та LR до приблизно 94 % і 96 % відповідно, тоді як poisoning-атаки призводили до зниження точності LR до близько 95 %, у той час як DT зберігав показник на рівні близько 97 % [7]. Обидва алгоритми відносяться до типу навчання з учителем та потребують числових і категоріальних типів даних. Хоча рівень точності понад 90 % традиційно вважається високим, у контексті систем виявлення вторгнень навіть похибка на рівні 3-4 % може становити критичну вразливість. У цьому аспекті результати дослідження Прадхана (98-99 %), суттєво зменшують так звану «сліпу зону» IDS та знижують ризик успішного проникнення зловмисників через невиявлені вектори атак.

Додатково, у дослідженні 2020 року показано, що змагальні атаки у моделях «сірого» та «чорного ящика» здатні обходити різні ML-моделі в IDS із імовірністю понад 97 % (evasion rate) [8]. Такий рівень обходу суттєво підриває довіру до систем машинного навчання та створює загрози витоку інформації, компрометації даних і порушення цілісності інформаційних систем.

Сукупність наведених досліджень підтверджує актуальність удосконалення процесів побудови, тестування та експлуатації систем кіберзахисту на основі машинного навчання. Аналітична оцінка стійкості конкретних моделей, а також постійний моніторинг нових вразливостей залишаються ключовими науковими та практичними завданнями у сфері кібербезпеки. Водночас алгоритми ML продовжують активно застосовуватися як зловмисниками для реалізації атак, так і фахівцями з безпеки – для їх виявлення та протидії. Точність та надійність моделей машинного навчання безпосередньо залежать від якості й актуальності даних, рівня міждисциплінарної взаємодії між інженерами-програмістами та фахівцями з аналізу даних, а також від своєчасного налаштування й супроводу моделей. У цьому контексті машинне навчання має розглядатися не як ізольований технічний інструмент, а як складова стратегічної архітектури кіберзахисту, що вимагає трансформації корпоративної культури та методологій безпеки.

У комплексному дослідженні, опублікованому у 2025 році, наголошується на критичній потребі впровадження інтелектуальних систем оцінювання якості даних, оскільки низька якість даних безпосередньо обмежує продуктивність і надійність ML-систем [9].

Фундаментом для побудови таких систем є технології машинного навчання, які традиційно поділяються на чотири категорії: навчання з учителем (supervised), без учителя (unsupervised), напівкероване навчання (semi-supervised) та навчання з підкріпленням (reinforcement learning) [10]. Вибір конкретної категорії залежить від структури вхідних даних (зокрема, наявності маркування), а також від функціональних завдань, які має виконувати система захисту.

Метою статті є аналіз підходів до розв'язання актуальних проблем у сфері кібербезпеки, а також оцінка ефективності та стійкості моделей машинного навчання до сучасних кібератак. У роботі запропоновано орієнтовну модель побудови системи виявлення кібератак на основі ML та сформульовано вимоги до сучасних IDS з урахуванням актуальних реалій кіберпростору.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Інтеграція модулів машинного навчання в рішення класу IDS/IPS та SIEM розпочалася у період 2010-2015 років і на сьогодні досягла стадії технологічної зрілості. Найбільш поширеними в таких системах є методи навчання з учителем та без учителя, що зумовлено специфікою задач моніторингу та аналізу мережевого трафіку.

Навчання з учителем ґрунтується на використанні заздалегідь визначених шаблонів, які виконують роль «вчителя», відповідно до яких модель зіставляє ознаки трафіку для виявлення



активності, нехарактерної для нормальної поведінки мережі. Реалізація такого підходу потребує маркованих даних, підготовка яких є трудомістким і ресурсоємним процесом, що обмежує його застосування в умовах необхідності швидкого реагування на нові атаки. Попри це, алгоритми навчання з учителем широко використовуються в системах моніторингу мережевого трафіку, зокрема в рішеннях класу Next-Generation Firewalls (NGFW) [11], де застосовуються методи Random Forest або Gradient Boosting (XGBoost) [12]. У системах виявлення фішингу, таких як Email Security Gateways, поширеним є використання сучасних алгоритмів на основі трансформерних моделей (наприклад, BERT або RoBERTa) у поєднанні з Logistic Regression, що забезпечує ефективну класифікацію спаму та фішингових повідомлень.

Метод навчання без учителя застосовується в системах, орієнтованих на виявлення аномалій без використання заздалегідь визначених шаблонів. У цьому випадку модель самостійно формує уявлення про нормальну поведінку системи та виявляє відхилення за заданими параметрами. Такий підхід є особливо ефективним для моніторингу та виявлення нових або раніше невідомих атак, а також характеризується меншою обчислювальною складністю, оскільки не потребує маркованих даних. Прикладами реалізації є системи поведінкового аналізу користувачів і сутностей (UBA/UEBA), а також механізми виявлення атак нульового дня (Zero-day). У платформах класу SIEM, NDR та Flow Analysis цей підхід зазвичай реалізується з використанням алгоритмів Isolation Forest або K-Means, які дозволяють ефективно виявляти аномалії в мережевому трафіку.

Напівкероване навчання використовується як компромісний підхід у випадках, коли повна розмітка даних є надто дорогою або складною, проте доступна обмежена кількість маркованих прикладів. Такий метод дозволяє поєднувати марковані та немарковані дані в процесі навчання, зменшуючи потребу в ручному маркуванні та скорочуючи час підготовки моделей [13]. Напівкеровані підходи застосовуються, зокрема, в системах Sandboxing та Malware Analysis, де використовуються методи згорткових нейронних мереж (CNN) або глибоких нейронних мереж (DNN) [12, 10].

Навчання з підкріпленням застосовується в динамічних середовищах, де відсутні заздалегідь відомі правильні відповіді, а модель навчається шляхом взаємодії з середовищем і отримання винагород за бажані дії. До таких систем належать SOAR (Security Orchestration, Automation and Response), автономні SOC та Honeypots, які реалізуються з використанням алгоритмів Deep Q-Learning (DQN) або Proximal Policy Optimization (PPO) [12, 14].

У таблиці 1 наведено спрощену класифікацію типів машинного навчання залежно від характеру доступних даних та способу отримання зворотного зв'язку табл. 1.

Таблиця 1

Класифікація методів машинного навчання за способом формування даних

Тип навчання	Спосіб формування навчальної вибірки
Навчання з вчителем (Supervised)	Розмічені дані (X, Y)
Навчання без учителя (Unsupervised)	Немає міток
Напівкероване навчання (Semi-supervised)	Частково є мітки
Навчання з підкріпленням (Reinforcement)	Навчання через винагороду

Загалом машинне навчання суттєво трансформує підходи до виконання завдань у різних галузях, дозволяючи системам навчатися на основі даних і вдосконалюватися з часом без явного програмування, що забезпечує ефективну роботу таких рішень, як віртуальні асистенти, системи виявлення шахрайства та інші інтелектуальні сервіси [9]. У системах виявлення кібератак алгоритми машинного навчання не завжди виступають базовим компонентом, проте є важливим доповненням для підвищення рівня безпеки інформаційних систем.

Результати наукових досліджень свідчать, що для досягнення високого рівня захисту мережевих ресурсів доцільно застосовувати методи, засновані на виявленні аномалій [15]. Це пояснюється обмеженою ефективністю традиційних сигнатурних підходів, які здатні протидіяти лише відомим загрозам і не забезпечують своєчасного реагування на нові або модифіковані атаки. Натомість методи виявлення аномалій дозволяють ідентифікувати відхилення від нормальної поведінки мережі та забезпечують більш адаптивний і проактивний рівень захисту.

Порівняльні дослідження методів виявлення кібератак із використанням машинного навчання показали, що підходи без учителя є більш гнучкими та швидшими за методи навчання з учителем. Зокрема, модель на основі логістичної регресії продемонструвала вищу точність класифікації, проте потребувала значно більшого часу на попередню обробку даних, включно з обробкою пропущених значень, кодуванням категоріальних ознак методом One-Hot Encoding та стандартизацією даних за



допомогою StandardScaler. Така підготовка є необхідною для масштабування даних і безпосередньо впливає на якість навчання та точність оцінки моделі.

Натомість модель на основі алгоритму Isolation Forest показала вищу швидкість під час обробки мережевого трафіку. У кожному вузлі ізоляційного дерева випадковим чином обирається одна ознака та значення для розбиття даних, що зменшує кореляцію між деревами, підвищує узагальнюючу здатність моделі та прискорює процес навчання. За своєю логікою цей підхід є спорідненим із Random Forest, проте характеризується простішою реалізацією.

Алгоритм Isolation Forest краще пристосований до потокової обробки даних, оскільки дозволяє виявляти аномалії без попереднього маркування та може бути адаптований до інкрементального або часткового оновлення моделі, що є критично важливим для аналізу мережевого трафіку в режимі реального часу. Водночас логістична регресія, як метод навчання з учителем, потребує маркованих даних і для роботи в потоковому режимі вимагає повного перенавчання або використання спеціалізованих онлайн-алгоритмів оптимізації, що ускладнює її практичне застосування. Крім того, моделі з учителем можуть бути менш ефективними у виявленні рідкісних атак і загроз нульового дня.

Таким чином, вибір конкретного підходу машинного навчання має визначитися характером даних і вимогами до системи захисту. Для статичних або структурованих даних доцільним є застосування методів навчання з учителем, наприклад у виробничих середовищах для автоматизації повторюваних процесів, тоді як для динамічних мережевих середовищ більш ефективними є методи без учителя, орієнтовані на аналіз аномалій у реальному часі.

У подальшому дослідженні Fuhrman S., Gungor O. та Rosing T. запропоновано нетрадиційний підхід до обробки мережевих даних у задачах виявлення вторгнень. Автори розділили вхідні дані на три компоненти: нормальні дані (10 %), навчальний досвід та тестовий досвід, при цьому навчальний і тестовий досвід додатково поділялися на m окремих підбірок. Запропонована архітектура орієнтована на аналіз потокового мережевого трафіку та складається з безперервного екстрактора ознак (Continuous Feature Extractor, CFE) і детектора новизни, побудованого на основі методу головних компонент (PCA).

CFE формує динамічні представлення вхідних даних, які надалі використовуються модулем виявлення новизни для розрізнення нормального та атакуючого трафіку. Завдяки безперервному навчанню CFE система здатна поступово формувати нові представлення ознак, що дозволяє ідентифікувати раніше невідомі атаки ще до їх масової появи у потоці даних. PCA-реконструкція використовується для обчислення аномального бали на основі ознак, сформованих CFE, причому PCA навчається виключно на нормальних даних, що забезпечує ефективне виявлення невідомих та модифікованих атак. Кінцева класифікація здійснюється шляхом порівняння аномального бали з пороговим значенням τ : у разі його перевищення трафік класифікується як атакуючий. Запропонована система не потребує використання розмічених даних [16].

Такий підхід є перспективним напрямом у протидії складним кіберзагрозам, зокрема APT-атакам і модифікованим DoS/DDoS-атакам. Водночас налаштування гіперпараметрів подібних систем ускладнюється під час тестування на більшості стандартних наборів даних, які переважно орієнтовані на статичні моделі. Це обмежує можливість отримання об'єктивної оцінки ефективності та ускладнює коректне порівняння з альтернативними рішеннями.

Експериментальні результати, отримані на реалістичних наборах даних, засвідчили, що система CND IDS забезпечує суттєве покращення якості виявлення та продуктивності: F-міра зросла у 6,1 рази, а пропускна здатність – у 6,5 рази порівняно з сучасними методами безперервного навчання (SOTA) [15]. Це підтверджує перспективність підходу в умовах динамічного ландшафту кіберзагроз, попри складність налаштування моделей.

Загалом потокове навчання в машинному навчанні є ключовим елементом побудови сучасних IDS і означає здатність алгоритмів обробляти та аналізувати дані, що надходять у безперервному режимі реального часу, без необхідності зберігання всього датасету та повного перенавчання моделі. Такий підхід забезпечує інкрементальне навчання, за якого модель донавчається на нових даних, зберігаючи раніше набуті знання, що дозволяє оперативно ідентифікувати інциденти та своєчасно реагувати на них. Важливою перевагою є мінімальні витрати пам'яті, оскільки зберігаються лише поточні дані та параметри моделі.

Разом із тим потокове навчання супроводжується ризиком так званого «катастрофічного забування», коли модель втрачає раніше засвоєні знання у випадку суттєвої зміни розподілу нових даних. Це явище є характерним для низки алгоритмів навчання без учителя, зокрема Isolation Forest. Подолання цієї проблеми можливе шляхом застосування стратегій безперервного навчання, що обґрунтовує доцільність використання методів виявлення новизни в системах IDS [16]. Саме такий підхід уперше був системно описаний у роботі «CND IDS: Continual Novelty Detection for Intrusion Detection Systems» авторами Fuhrman S., Gungor O. та Rosing T [16].



Різні типи потокового навчання узагальнено в таблиці 2. Однією з ключових проблем у роботі з поточними даними є зміна розподілу в часі (concept drift), яка суттєво впливає на ефективність моделей. Особливо чутливими до цього явища є нейронні мережі, адаптація яких до динамічних змін часто потребує повторного перенавчання. У зв'язку з цим у системах виявлення кібератак перевага надається алгоритмам, спеціально розробленим для поточних даних, зокрема Very Fast Decision Tree (VFDT), також відомому як Hoeffding Tree.

Таблиця 2

Класифікація методів навчання в умовах динамічного надходження даних

Типи потокового навчання	Принцип роботи
Online learning	Оновлення після кожного прикладу
Incremental learning	Оновлення малими порціями
Streaming learning	Обробка нескінченного потоку
Continual learning	Навчання без катастрофічного забування

Алгоритм VFDT ґрунтується на використанні нерівності Гефдінга, що дозволяє приймати рішення щодо розгалуження вузлів дерева на основі обмеженої кількості спостережень без необхідності повного аналізу всього набору даних. Такий підхід забезпечує інкрементальне навчання та адаптацію моделі в режимі реального часу, що є критично важливим для аналізу мережевого трафіку.

Зокрема, у дослідженні 2025 року показано, що класифікатор Hoeffding VFDT ефективно застосовується в IDS-сценаріях для поточних даних, забезпечуючи швидке оновлення моделі та адаптацію до нових шаблонів атак [15]. У задачах кібербезпеки перевага часто надається деревоподібним алгоритмам, які поєднують інтерпретованість, високу швидкодію та здатність працювати з нетиповими патернами. Хоча більшість таких методів належать до навчання з учителем, вони мають модифікації для напівкерованого та безкерованого навчання.

Попри актуальність методів навчання з учителем, вони дедалі меншою мірою відповідають викликам сучасного кіберпростору. Автоматична адаптація моделей і виявлення аномалій за відсутності повної розмітки даних є ключовими чинниками ефективної протидії атакам нульового дня. У зв'язку з цим напівкеровані та безкеровані підходи набувають дедалі більшої популярності в сучасних системах виявлення кібератак.

Потоковість у сучасному світі, що характеризується великими обсягами сирих даних, відіграє ключову роль, у зв'язку з чим необхідність формування гнучких архітектур перебуває на етапі інтенсивного розвитку та активного впровадження. Ґрунтуючись на результатах аналізу, Rios та інших [17], представили систему CND IDS, що базується на принципах безперервного виявлення детектора новизни. У межах запропонованої архітектури було інтегровано процеси динамічного оновлення ознак і виявлення аномалій, що дозволило суттєво підвищити показники точності (accuracy) та здатність системи адаптуватися до нових загроз у режимі реального часу. Зазначений підхід виявився ефективним для розв'язання поставленої задачі.

Не менш важливим аспектом є процес відбору найбільш релевантних змінних для побудови та навчання моделей машинного навчання. Метою застосування аналітичних методів є зменшення розмірності простору ознак та ідентифікація ключових інформативних параметрів з урахуванням специфіки різних алгоритмів. Це забезпечує істотне прискорення обробки та навчання моделей (у два рази і більше), підвищення точності, зростання продуктивності та зменшення ресурсоемності за рахунок усунення надмірності даних, що є критично важливим для систем, орієнтованих на забезпечення інформаційної безпеки. Одним із перших відомих алгоритмів відбору ознак є Relief, запропонований Kenji Kira та Larry Rendell у 1992 році як метод оцінювання важливості ознак у задачах класифікації [18]. Подальший розвиток систем виявлення кібератак зумовив потребу у впровадженні підходів Incremental Feature Encoding та інших online / continual learning методів, у межах яких система не обмежується одноразовим відбором ознак, а адаптує їх у режимі реального часу відповідно до змін у потоці даних. Саме перехід від статичних до інкрементальних і динамічних методів аналізу є предметом сучасних наукових досліджень і логічно продовжує еволюцію мережевих систем виявлення вторгнень – Network Intrusion Detection Systems (NIDS). Концептуально NIDS були вперше запропоновані Дороті Денінг (Dorothy E. Denning) у 1987 році як підхід до виявлення аномальної та зловмисної активності в комп'ютерних мережах, а згодом практично реалізовані в роботах і проєктах Snort (Martin Roesch, 1998), Bro/Zeek (Vern Paxson, 1999) та Suricata, які заклали основу для розвитку адаптивних і машинно-орієнтованих NIDS у середовищах реального часу [18].

У сучасному дослідженні «A Generalized and Real-Time Network Intrusion Detection System Through Incremental Feature Encoding and Similarity Embedding Learning» (Sensors, 2025) [18] автори



запропонували новий підхід до NIDS, що ґрунтується на двох ключових ідеях, які суттєво відрізняються від традиційних методів. Перша полягає у використанні механізму інкрементального кодування ознак, здатного фіксувати зміни в мережевому трафіку в режимі реального часу. Друга – у застосуванні стратегії навчання на основі семантичного вбудовування подібності для формування компактного та дискримінативного простору представлення, що покращує виявлення як відомих, так і нових типів атак [18].

На початковому етапі автори розглядали NIDS на основі сесій, де ключовою ознакою був міжпакетний час (Inter-Arrival Time, IAT) у потоках даних між пристроєм і користувачем. Однак такий підхід мав низку недоліків. Зокрема, протокол TCP у разі повторної передачі пакетів збільшував тривалість сесії, що призводило до штучного зростання значень IAT. Як наслідок, значення цього показника доводилося встановлювати на завищеному рівні, що негативно впливало на якість моделі.

У подальшому було обрано підхід NIDS на основі аналізу окремих пакетів, у межах якого ознаки вилучалися з кожного пакета незалежно. Проте такий метод призводив до зниження точності та зростання рівня хибних тривог (False Alarms). З урахуванням цього було зроблено спробу врахування часових залежностей між пакетами в межах сесії, однак такий підхід виявився неефективним для обробки як коротких, так і довгих сеансів.

Для розв'язання зазначеної проблеми система GR-IDS пропонує модуль інкрементального кодування ознак, який автоматично та поступово витягує інформативні характеристики з послідовності пакетів поточного сеансу [18].

Проблема маркованих даних залишається актуальною для багатьох фахівців з кібербезпеки. Окрім труднощів, пов'язаних із формуванням репрезентативних наборів даних, особлива увага приділяється вибору коректних алгоритмічних рішень, що відповідають реальній природі кіберзагроз. Низка сучасних досліджень продемонструвала перспективні результати порівняння різних алгоритмів і методів машинного навчання для систем виявлення кібератак. У межах цієї роботи розглядаються два дослідження, у яких порівнюються статичний та динамічний експерименти з метою оцінювання різниці в точності, стабільності та адаптивності до нових типів кібератак.

У середовищі фахівців з кібербезпеки триває активна наукова дискусія щодо ефективності традиційних сигнатурних методів виявлення вторгнень, оскільки більшість таких систем здатні розпізнавати лише відомі шаблони атак і не забезпечують належного рівня виявлення нових або модифікованих кіберзагроз. Для оцінювання продуктивності та адаптивності алгоритмів машинного навчання було проведено статичне дослідження на готовому наборі мережевих даних [19]. Як наведено в Таблиці 3, найвищу точність (ассурагу) продемонстрував алгоритм XGBoost – 98 %, що свідчить про його ефективність у класифікації відомих і складних типів атак. Алгоритм GAN-based IDS показав дещо нижчу точність – близько 96 %, водночас характеризується високим рівнем адаптивності до потенційно нових та еволюційних загроз.

Інші моделі продемонстрували нижчі результати: ANN – 91 %, SVM – 88 %, Decision Tree – 85 %, тоді як Random Forest показав найнижчу точність – 78 %, що свідчить про обмежену здатність цієї моделі до розпізнавання нових сценаріїв атак. Окрім ассурагу, у Таблиці 3 наведено показник recall, який відображає здатність алгоритмів коректно виявляти всі реальні атаки. Так, XGBoost і GAN-based IDS мають recall на рівні приблизно 98 % та 96 % відповідно, тоді як SVM, Decision Tree та Random Forest демонструють нижчі значення (~76-86 %). Отримані результати підтверджують, що адаптивні алгоритми з високими показниками ассурагу та recall є більш придатними для побудови сучасних NIDS, здатних ефективно реагувати на нові та складні кіберзагрози.

Таблиця 3

Результати статичного дослідження алгоритмів машинного навчання

Алгоритм моделі	Метрика ассурагу (%)	Метрика recall (%)
XGBoost	98 %	~ 98 %
GAN based IDS	96 %	~ 96 %
ANN	91 %	~90 %
SVM	88 %	~ 86 %
Decision Tree	85 %	~ 82 %
Random Forest	78 %	~ 76 %

Поштовхом до реалізації першого дослідження стало прагнення здійснити систематичне порівняння ефективності різних алгоритмів машинного навчання для виявлення кібератак у мережевому трафіку. Основними завданнями роботи були оцінювання показників точності та повноти виявлення, аналіз здатності алгоритмів обробляти дані в режимі реального часу, а також визначення моделей, що



забезпечують оптимальний баланс між продуктивністю та обчислювальними витратами. Автори наголошують на доцільності подальших досліджень у напрямі гібридних і ансамблевих моделей, застосування підходів навчання з урахуванням дій противника (adversarial learning), а також розвитку механізмів реального часу в системах IDS з метою поєднання високої точності та швидкості виявлення сучасних кіберзагроз.

У межах динамічного дослідження [20], орієнтованого на аналіз потокових мережевих даних, було продемонстровано здатність системи зберігати попередні знання, що характеризується показником Backward Transfer (BWT) на рівні приблизно 20,76 %. Експерименти проводилися на значних обсягах мережевого трафіку та даних із сенсорів IoT-пристроїв у режимі реального часу, за умов динамічної зміни характеристик потоків. Водночас дослідження виявило притаманну таким системам проблему високої ймовірності забування раніше засвоєних знань. Незважаючи на це, запропонований підхід дозволив ефективно усунути обмеження, пов'язані з виявленням нових типів кібератак.

За показником адаптивності Forward Transfer (FWT) система безперервного навчання досягла значення близько 70,05 %, що свідчить не лише про високий рівень адаптації до нових даних, а й про суттєвий прогрес у розв'язанні ключових проблем систем виявлення кібератак. Загальна оцінка ефективності показала, що система забезпечила точність класифікації на рівні 70,6 %, демонструючи збалансовану здатність до виявлення аномалій у динамічному потоці даних. Для порівняння, традиційні статичні системи в аналогічних умовах, як правило, демонструють знижену ефективність. Узагальнені результати застосованих методів наведено в Таблиці 4.

Таблиця 4

Результати досліджень статичного дослідження методів машинного навчання

Метод	Оцінка (%)	Коментар
Average LL PR AUC	~70.6 %	Середня здатність моделі правильно класифікувати як нормальні, так і атакуючі потоки протягом навчання (по 6 наборах даних).
Backward Transfer (BWT)	~20.76 %	Позитивне значення показника свідчить про те, що модель зберегла здатність коректно виявляти раніше відомі типи атак і, крім того, продемонструвала підвищення ефективності їх розпізнавання після донавчання на нових даних.
Forward Transfer (FWT)	~70.05 %	Високий показник FWT демонструє, що модель ефективно використовує знання, набуті на попередніх даних, для швидшого та точнішого виявлення нових атак, забезпечуючи адаптивність системи до динамічного потоку мережевих даних

Дослідження CITADEL [20] зосереджене на розв'язанні ключових проблем обробки потокових даних у режимі реального часу та формуванні адаптивної моделі виявлення нових кібератак і аномалій без втрати раніше набутих знань. Запропонований підхід продемонстрував високу точність, однак залишається відкритим питання оптимального балансу між збереженням старих знань і засвоєнням нових, зокрема в умовах дисбалансу класів.

Порівняння результатів статичного дослідження Sharma, Kumar (2025) [19] та динамічного підходу CITADEL (Li et al., 2025) [20] є ускладненим через використання різних парадигм і метрик оцінювання (табл. 5). Загалом перше дослідження підтвердило актуальність традиційних методів машинного навчання та їхню високу ефективність у контрольованих умовах. Натомість друге дослідження відкрило нові перспективи в побудові адаптивних IDS, здатних працювати зі складним і мінливим потоком мережевого трафіку в реальному часі.

Таблиця 5

Підсумкове порівняння статичного та динамічного дослідження

	Статичне дослідження	Коментар
Назва дослідження	Comparative Analysis of Machine Learning Models for Intrusion Detection Systems	CITADEL: Continual Anomaly Detection for Enhanced Learning in IoT Intrusion Detection
Дата дослідження	05 грудня 2025	26 серпня 2025
Алгоритми/Методи дослідження	(XGBoost, GAN-based IDS, ANN, SVM, Decision Tree, Random Forest	Self-supervised CL + Memory-aware Autoencoder + Novelty detection
Розбиття даних	Train/Test	-



Продовження таблиці 5

Метрика оцінювання	Accuracy до 98 %	~+72.9 % improvement vs VLAD
Тип машинного навчання	Supervised ML	Continual learning
Особливості	Шум, дисбаланс класів	Проблема з балансом старого та нового навчання
Еволюція загроз	Залишилася викликом	Подолала
Ресурсозатратність	Деякі алгоритми потребують	Потребує

Статичний підхід, заснований на керованому машинному навчанні (XGBoost, GAN-based IDS, ANN, SVM, Decision Tree, Random Forest), продемонстрував високу точність виявлення відомих атак – до 98 % за умови чітко сформованих навчальних і тестових вибірок. Його перевагами є відносна простота реалізації, інтерпретованість результатів і можливість інтеграції в наявні IDS-платформи. Водночас такі моделі залишаються чутливими до шуму та дисбалансу класів і мають обмежену здатність до адаптації в умовах еволюції загроз.

Натомість динамічний підхід CITADEL, побудований на концепціях continual learning, self-supervised learning та novelty detection, продемонстрував суттєво вищу адаптивність до змін мережевої поведінки й появи нових типів атак без повного перенавчання моделі. Використання memory-aware autoencoder дозволило значною мірою зменшити ефект катастрофічного забування та досягти приросту ефективності приблизно на 72,9 % порівняно з базовими методами. Разом із тим така система є більш ресурсомісткою та складною в реалізації й потребує ретельного контролю компромісу між збереженням попередніх знань і навчанням на нових даних. З практичної точки зору результати обох досліджень свідчать, що статичні ML-IDS доцільно застосовувати в корпоративних і датацентрових середовищах зі стабільним профілем трафіку.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Таким чином, аналіз двох підходів підтверджує загальну тенденцію переходу сучасних IDS від статичних моделей до адаптивних безперервно навчених систем, а також обґрунтовує доцільність використання гібридних архітектур, що поєднують високу точність керованого навчання з гнучкістю continual learning.

Для систем виявлення кібератак у динамічних середовищах доцільно переглянути традиційний підхід до поділу даних. Класичне співвідношення тренувальної та тестової вибірок (80/20) є ефективним для офлайн-оцінювання, однак обмежено придатним для аналізу мережевого трафіку в реальному часі. Рекомендовано використовувати трикомпонентний поділ даних (нормальні, навчальні та тестові), що забезпечує гнучкіше навчання та коректнішу оцінку ефективності IDS.

Запровадження концепції continual learning дозволяє реалізувати безперервну адаптацію до змін мережевого середовища та появи нових атак, підтримуючи накопичення знань і зменшуючи вплив катастрофічного забування. Такі системи здатні працювати з постійним потоком даних, потребують мінімального обсягу пам'яті для збереження попередньої інформації та перевершують традиційні підходи, які не підтримують інкрементальне навчання.

Удосконалення методів кореляційного аналізу та інкрементального кодування ознак сприяло зменшенню обсягу вхідних даних і підвищенню точності класифікації. Виділення релевантних ознак забезпечує компактне й дискримінативне представлення трафіку, що є ключовим чинником ефективної роботи IDS у реальному часі. Водночас застосування методів машинного навчання має відповідати вимогам інформаційної безпеки. Статичні системи залишаються ефективними для аналізу логів, історичних інцидентів, аудиту та тестування моделей. У таких умовах рекомендовано використання XGBoost, GAN-based IDS та ANN, які забезпечують високу точність і детальний аналіз помилок, хоча й поступаються динамічним підходам у здатності реагувати на нові загрози.

Сучасна модель системи виявлення кібератак базується на інтеграції алгоритмів машинного навчання, які забезпечують адаптивну та ефективну класифікацію мережевого трафіку. Вона передбачає збір та попередню обробку даних із мережевого середовища з одночасним виділенням релевантних ознак, що мають найбільший вплив на ідентифікацію атак. Для підвищення точності та узагальнюваності моделі застосовується поєднання статичних методів керованого навчання для відомих загроз та безперервного навчання (continual learning) для адаптації до нових або еволюціонованих атак у реальному часі. Інкрементальне кодування ознак та методи кореляції дозволяють компактно та дискримінативно представляти дані, зменшуючи обсяг оброблюваної інформації та прискорюючи роботу



системи. Така інтегрована структура забезпечує безперервне оновлення знань, підтримку стабільної ефективності при зміні мережевого середовища та здатність своєчасно виявляти як відомі, так і нові кіберзагрози, створюючи основу для надійної та самокерованої системи виявлення вторгнень.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. ДНУ «Інститут інформації, безпеки і права НАПрН України», & Національна бібліотека України імені В. І. Вернадського. (2024). *Кібербезпека в інформаційному суспільстві: Інформаційно-аналітичний дайджест*, (5), 29.
2. BBC News Україна. (2025, January 24). *Ресстри відновили: Які наслідки кібератаки для України*. <https://www.bbc.com/ukrainian/articles/c5ye75y8415o>
3. Державна служба спеціального зв'язку та захисту інформації України. (2025). *Огляд кіберзагроз та стратегій захисту в 2025 році: досвід CERT-UA*. <https://cip.gov.ua/ua/faqs/cyber-threat-overview-and-defense-strategies-in-2025-cert-ua-s-experience>
4. Закон України «Про захист персональних даних» № 2297-VI. (2010, June 1; ред. 2025). <https://zakon.rada.gov.ua/laws/card/2297-17/ed20250101>
5. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 (Artificial Intelligence Act). (2024). *Official Journal of the European Union*. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
6. Pradhan, R. (2022). Decision tree based classifications on CICIDS 2017 dataset for the identification of DDoS, botnet, and web attack. *NeuroQuantology*, 20(12).
7. Alshahrani, E., Alghazzawi, D., Alotaibi, R., & Rabie, O. (2022). Adversarial attacks against supervised machine learning-based network intrusion detection systems. *Peer-to-Peer Networking and Applications*. <https://doi.org/10.1007/s12083-024-01859-9>
8. Han, D., Wang, Z., Zhong, Y., Chen, W., Yang, J., Lu, S., Shi, X., & Yin, X. (2020). Evaluating and improving adversarial robustness of machine learning-based network intrusion detectors [Preprint]. arXiv. <https://arxiv.org/abs/2005.07519>
9. Omdena. (2025, July 30). *Top machine learning issues for businesses in 2025*. <https://www.omdena.com/blog/machine-learning-issues-businesses-2025>
10. Denovo. (2025). *Що make machine learning?* <https://denovo.ua/resources/what-is-machine-learning>
11. Palo Alto Networks. (2023). *Machine learning in the next-generation firewall* (White paper).
12. Fuhrman, S., Gungor, O., & Rosing, T. (2025). CND IDS: Continual novelty detection for intrusion detection systems [Preprint]. arXiv. <https://arxiv.org/abs/2502.14094>
13. Li, E., Gungor, O., Shang, Z., & Rosing, T. (2025). CITADEL: Continual anomaly detection for enhanced learning in IoT intrusion detection [Preprint]. arXiv. <https://arxiv.org/abs/2508.19450>
14. Domingos, P., & Hulten, G. (2000). Mining high-speed data streams. In *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '00)* (pp. 71-80). ACM. <https://doi.org/10.1145/347090.347107>
15. Rios, A., Ahuja, N., Ndiour, I., Genc, U., Itti, L., & Tickoo, O. (2022). incDFM: Incremental deep feature modeling for continual novelty detection. In *European Conference on Computer Vision* (pp. 588-604). Springer.
16. *A generalized and real-time network intrusion detection system through incremental feature encoding and similarity embedding learning*. (2025). *Sensors*, 25(16), Article 4961. <https://doi.org/10.3390/s25164961>
17. Sharma, V., & Kumar, M. (2025). Comparative analysis of machine learning models for intrusion detection systems. *Panamerican Mathematical Journal*, 35(3s), 273-285. <https://doi.org/10.52783/pmj.v35.i3s.3891>

**Anastasiia Shlyakhova**

Student

Institute of Special Communications and Information

Protection of Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine

ORCID: 0009-0009-5926-2105

*slahovaanastasia33@gmail.com***Olha Shevchuk**

Lecturer

Institute of Special Communications and Information

Protection of Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine

ORCID: 0000-0002-2866-439X

*olia13511@gmail.com***Volodymyr Onishchenko**

Lecturer

Institute of Special Communications and Information

Protection of Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine

ORCID: 0009-0000-1355-9178

*v.o.onishchenko@ukr.net***DETECTION OF CYBERATTACKS IN NETWORK TRAFFIC BASED ON MACHINE LEARNING ALGORITHMS**

Abstract. The article examines the role and significance of machine learning (ML) methods as the methodological foundation of modern artificial intelligence. It is substantiated that the ability of information systems to self-learn and adapt in dynamic environments is a key factor in their effectiveness. The rapid growth in demand for ML technologies across all spheres of human activity is analyzed, which inevitably leads to the accumulation and processing of vast volumes of sensitive information. The concentration of such data creates new threat vectors, as it becomes a priority target for cyber adversaries. Special attention is paid to the implementation of ML algorithms in modern security ecosystems. The experience of using leading industrial solutions that replace traditional signature-based approaches with intelligent analysis is reviewed. The mechanisms for implementing network attacks aimed at poisoning initial data and manipulating the training process are described in detail. The results of the conducted experimental studies confirm that the use of irrelevant or compromised models in cybersecurity systems creates an illusion of security, leaving critical infrastructure vulnerable to targeted attacks. The article offers a conceptual outlook on the necessity of developing secure training protocols to ensure the resilience of intelligent systems.

Keywords: cybersecurity; machine learning; intrusion detection systems; anomaly detection; novelty detection; real-time monitoring.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Instytut informatsii, bezpeky i prava NAPrN Ukrainy, & Natsionalna biblioteka Ukrainy imeni V. I. Vernadskoho. (2024). *Kiberbezpeka v informatsiinomu suspilstvi: Informatsiino-analitychnyi daidzhest* (No. 5, p. 29). (in Ukrainian)
2. BBC News Україна. (2025, January 24). Реєстри відновили: Які наслідки кібератаки для України. <https://www.bbc.com/ukrainian/articles/c5ye75y8415o>
3. Derzhavna sluzhba spetsialnoho zviazku ta zakhystu informatsii Ukrainy. (2025). *Ohliad kiberzahroz ta stratehii zakhystu v 2025 rotsi: dosvid CERT-UA*. <https://cip.gov.ua/ua/faqs/cyber-threat-overview-and-defense-strategies-in-2025-cert-ua-s-experience>
4. Zakon Ukrainy “Pro zakhyst personalnykh danykh” No. 2297-VI. (2010, June 1; rev. 2025). <https://zakon.rada.gov.ua/laws/card/2297-17/ed20250101>
5. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 (Artificial Intelligence Act). (2024). *Official Journal of the European Union*. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>



6. Pradhan, R. (2022). Decision tree based classifications on CICIDS 2017 dataset for the identification of DDoS, botnet, and web attack. *NeuroQuantology*, 20(12).
7. Alshahrani, E., Alghazzawi, D., Alotaibi, R., & Rabie, O. (2022). Adversarial attacks against supervised machine learning-based network intrusion detection systems. *Peer-to-Peer Networking and Applications*. <https://doi.org/10.1007/s12083-024-01859-9>
8. Han, D., Wang, Z., Zhong, Y., Chen, W., Yang, J., Lu, S., Shi, X., & Yin, X. (2020). Evaluating and improving adversarial robustness of machine learning-based network intrusion detectors. *arXiv*. <https://arxiv.org/abs/2005.07519>
9. Omdena. (2025, July 30). Top machine learning issues for businesses in 2025. <https://www.omdena.com/blog/machine-learning-issues-businesses-2025>
10. Denovo. (2025). Що таке machine learning? <https://denovo.ua/resources/what-is-machine-learning>
11. Palo Alto Networks. (2023). *Machine learning in the next-generation firewall* (White paper).
12. Fuhrman, S., Gungor, O., & Rosing, T. (2025). CND IDS: Continual novelty detection for intrusion detection systems. *arXiv*. <https://arxiv.org/abs/2502.14094>
13. Li, E., Gungor, O., Shang, Z., & Rosing, T. (2025). CITADEL: Continual anomaly detection for enhanced learning in IoT intrusion detection. *arXiv*. <https://arxiv.org/abs/2508.19450>
14. Domingos, P., & Hulten, G. (2000). Mining high-speed data streams. In *Proceedings of the sixth ACM SIGKDD international conference on knowledge discovery and data mining (KDD '00)* (pp. 71–80). ACM. <https://doi.org/10.1145/347090.347107>
15. Rios, A., Ahuja, N., Ndiour, I., Genc, U., Itti, L., & Tickoo, O. (2022). incDFM: Incremental deep feature modeling for continual novelty detection. In *European conference on computer vision* (pp. 588–604). Springer.
16. A generalized and real-time network intrusion detection system through incremental feature encoding and similarity embedding learning. (2025). *Sensors*, 25(16), Article 4961. <https://doi.org/10.3390/s25164961>
17. Sharma, V., & Kumar, M. (2025). Comparative analysis of machine learning models for intrusion detection systems. *Panamerican Mathematical Journal*, 35(3s), 273–285. <https://doi.org/10.52783/pmj.v35.i3s.3891>

Отримано редакцією журналу / Received: 10.02.26

Прорецензовано / Revised: 25.02.26

Схвалено до друку / Accepted: 25.06.26



This work is licensed under Creative Commons Attribution-noncommercial-sharealike 4.0 International License.