Генеративні алгоритми розширюють можливості автономних агентів, дозволяючи їм самостійно створювати нові дії, адаптуватися до невизначених середовищ і підвищувати власну ефективність. Подальші дослідження мають бути спрямовані на поєднання генеративного навчання з когнітивними моделями для створення агентів, здатних до логічного мислення та пояснюваного ухвалення рішень.

Annotation. The article considers algorithmic structures and data processing methods used in developing intelligent agents. It reviews the evolution from hierarchical systems to generative models that implement adaptive learning and prediction. A simplified example of a generative agent based on reinforcement algorithms is presented, showing how agents can autonomously create new strategies. The importance of hybrid architectures combining analytical and generative methods is emphasized.

Keywords: intelligent agent, generative algorithm, reinforcement learning, optimization, artificial intelligence.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Russell S., Norvig P. Artificial Intelligence: A Modern Approach. Pearson, 2021. 1166 p. URL: http://lib.ysu.am/disciplines_bk/efdd4d1d4c2087fe1cbe03d9ced67f34.pdf

2. Sutton R. S., Barto A. G. Reinforcement Learning: An Introduction. *Stanford University.* Cambridge, London: MIT Press, 2014. 352 p. URL: https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf

3. Goodfellow I., Bengio Y., Courville A. Deep Learning. MIT Press, 2016.

4. Schmidhuber J. Generative AI and Autonomous Systems. Neural Networks, 2024.

5. Бойко С. І., Корольов Д. В. Генеративні методи в інтелектуальних системах. Київ: КПІ ім. І. Сікорського, 2023.

6. Nathalia N., Alencar P., Cowan D. Self-Adaptive Large Language Model (LLM)-Based Multiagent Systems. *Cornell University.* 2023. DOI: 10.48550/arXiv.2307.06187.

7. Thuy Ngoc Nguyen, Duy Nhat Phan, Cleotilde Gonzalez. Learning in Cooperative Multiagent Systems Using Cognitive and Machine Models. *Cornell University.* 2023. DOI: 10.48550/arXiv.2308.09219.

8. Helie S., Sun R. Cognitive Architectures and Agents. Part of the book series: Springer Handbooks. *Springer Nature Link.* P. 683–696. URL: https://link.springer.com/chapter/10.1007/978-3-662-43505-2_36

9. Casals A., Brandão A. A. F. HECATE: An ECS-based Framework for Teaching and Developing Multi-Agent Systems. *Cornell University.* 2025. DOI: 10.48550/arXiv.2509.06431.

УДК 004.056.55:004.032.26(477)

АНАЛІЗ ЗАСТОСУВАННЯ НЕЙРОННИХ МЕРЕЖ ДЛЯ ЗАПОБІГАННЯ ТА ВИЯВЛЕННЯ РІЗНИХ ТИПІВ КІБЕРАТАК НА МЕРЕЖЕВІ РЕСУРСИ В ПЕРІОД ВІЙСЬКОВОГО СТАНУ

А. І. Юкальчук, Л. В. Загоруйко

Анотація. У роботі досліджено застосування та порівняння ефективності різних архітектур штучних нейронних мереж (MLP, CNN, SOM, Hopfield та кілька гібридних конфігурацій, зокрема CNN+MLP і Transformer+ Hopfield) для задачі виявлення мережевих вторгнень на базі стандартного датасету KDD'99. Усі моделі реалізовано мовою *Python.*

Ключові слова: нейронна мережа, виявлення вторгнень, DOS, R2L, U2R.

Вступ. У сучасних умовах швидкого зростання кіберзагроз та все більшої залежності критичних об'єктів від інформаційних мереж постає задача створення надійних і адаптивних систем виявлення вторгнень. У роботі розглядається застосування штучних нейронних мереж для автоматичного розпізнавання аномалій у мережевому трафіку, зокрема для виявлення складних і рідкісних сценаріїв атак, які становлять найбільшу загрозу для безпеки інформаційних систем.

Метою роботи було порівняти ці архітектури в контексті задачі виявлення вторгнень на основі загальнодоступного набору даних KDD'99, встановити переваги та обмеження кожного підходу та сформулювати практичні рекомендації щодо їх застосування в середовищах із підвищеними вимогами до безпеки. Для оцінки моделей використовувалися стандартні метричні підходи, що дозволяють порівнювати їх здатність розрізняти різні типи мережевих подій.

Сучасні дослідження у сфері виявлення вторгнень. У літературі останніх років існує значна кількість робіт, пов'язаних із виявленням вторгнень на основі глибокого навчання. Сучасні дослідження у сфері кібербезпеки зосереджуються на інтеграції штучного інтелекту та методів машинного навчання для покращення здатності виявляти та протидіяти кіберзагрозам. Хоча ці дослідження досягли значних результатів, більшість із них все ще обмежуються конкретними тестовими наборами даних, зокрема NSL-KDD, UNSW-NB15 та CIC-IoT2023. Однак нові дослідження дедалі частіше використовують події безпеки та журнали, пов'язані з реальними випадками, що робить їх результати більш практично застосовними [1–3].

Saini та співавтори провели всебічний огляд дуальності адверсаріального навчання в мережевому виявленні вторгнень, досліджуючи атаки та контрзаходи. Їх дослідження показало, що графові нейронні мережі (GNN) демонструють особливу перспективність для виявлення мережевих вторгнень, завдяки їх здатності обробляти складні залежності та взаємозв'язки в мережевих даних. Автори підкреслили важливість розвитку стійких механізмів захисту для протидії потенційним порушенням мережевої безпеки та приватності, викликаним адверсаріальними атаками [3].

Zhong та співавтори провели всебічне дослідження застосування графових нейронних мереж для систем виявлення вторгнень. Їх систематична таксономія класифікує існуючі та майбутні дослідження методів виявлення вторгнень на основі GNN, підкреслюючи переваги цих архітектур у захопленні складних взаємозв'язків у структурованих графом даних [4].

Chhetri та Namin дослідили застосування трансформерних моделей для прогнозування наслідків кібератак. Їх робота демонструє ефективність архітектури BERT та ієрархічних мереж уваги (HAN) для аналізу текстових описів кібератак та прогнозування їх потенційного впливу, що є важливим кроком у напрямі автоматизації оцінки загроз [5].

Sharma запропонував реалізацію моделі, яка використовує аналіз мережевого трафіку (NTA) та дані лічильників продуктивності апаратного забезпечення (HPC) для точного позначення zero-day атак. Дослідження спрямоване на створення моделі на основі автокодера, яка об'єднує апаратні та мережеві характеристики для ефективної класифікації, використовуючи стандартні набори даних кібербезпеки CICIDS2017, NSL-KDD та D.A.V.I.D.E HPC [6].

Rasikha та співавтори розробили ансамблевий метод глибокого навчання для виявлення кібератак і підкреслили, що методи глибокого навчання є надзвичайно бажаними для виявлення кібератак, оскільки вони не потребують інженерії ознак або припущень про розподіл даних. Їх підхід продемонстрував високу ефективність у виявленні різноманітних типів атак в IoT-мережах [7].

Змодельовані типи нейронних мереж. У процесі дослідження для моделювання та подальшого аналізу ефективності у виявленні кібератак було відібрано кілька різних підходів до побудови нейронних мереж. Серед них розглядалися класичні багатошарові персептрони, що традиційно застосовуються для задач класифікації; згорткові нейронні мережі, які добре працюють з послідовними та структурованими даними; самоорганізовані карти Кохонена, що належать до несупервізійних методів та дозволяють виконувати кластеризацію і візуалізацію складних даних; асоціативна мережа Хопфілда, яка імітує механізми пам'яті та здатна відновлювати зразки за частковими вхідними даними; а також гібридні архітектури, що поєднують властивості кількох підходів із метою досягнення більш збалансованої продуктивності у виявленні як масових, так і рідкісних типів атак. Такий вибір моделей дозволив комплексно оцінити можливості різних парадигм нейронних мереж у контексті задач кіберзахисту.

Класичний багатошаровий персептрон (MLP) використано як базову модель для порівняння. MLP – це послідовність щільних (fully connected) шарів із нелінійними активаціями, що навчаються відображати вхідні вектори ознак у простір, де класи відокремлюються лінійно. Під час навчання застосовувався стохастичний градієнтний спуск із крос-ентропійною функцією втрат для багатокласової класифікації; стандартні методи регуляризації (dropout, L2) та нормалізації допомагали запобігти перенавчанню. MLP працює добре, коли зв'язки між ознаками можна захопити через комбінації лінійних перетворень і простих нелінійностей; проте у випадку сильно локалізованих або структурованих шаблонів у даних (наприклад, ча-

сові або корельовані блоки ознак) він може вимагати великої кількості параметрів і довшого тренування.

Згорткові нейронні мережі (CNN), хоча з самого початку і розроблені для роботи із зображеннями, виявилися корисними для виявлення локальних закономірностей у векторі ознак мережевого трафіку за умови правильного відтворення структури вхідного простору. Замість повного з'єднання, CNN застосовують локальні фільтри, що шукають повторювані підструктури та взаємозв'язки між сусідніми ознаками; це дозволяє ефективніше вчитися на великому числі ознак і знижує кількість параметрів. У контексті IDS це дає перевагу в розпізнаванні характерних шаблонів пакетних послідовностей або груп ознак, що супроводжують певні типи атак. Для тренування в CNN використовувалися операції пулингу та нормалізації, а також адаптація навчальної швидкості та батч-нормалізація для стабільності навчання.

Самоорганізуючі карти (SOM) застосовували як інструмент ненадзорного виявлення аномалій і візуалізації простору ознак. SOM виставляє багатовимірні вектори на двовимірну карту, зберігаючи топологічні відношення між прикладами; це добре підходить для попереднього виявлення кластерів типового трафіку й виявлення відхилень, які можуть відповідати аномаліям. SOM не є класифікатором, але може бути у ролі попереднього фільтра або джерела додаткових ознак (наприклад, відстань до найближчого нейрона карти) для наступних стадій класифікації. У поєднанні з наглядовими моделями SOM підвищує інтерпретованість та чутливість до нових, раніше не бачених типів вторгнень.

Мережі Хопфілда розглядались у ролі асоціативної пам'яті для збереження прототипів нормального й атакуючого трафіку. Така мережа здатна до автодоведення та відновлення раніше засвоєних шаблонів за частково пошкодженого ввідного вектора, тому її використовували для перевірки відповідності нового зразка наборам «нормальних» або «атакувальних» станів. Практично Hopfield-підхід корисний для виявлення очевидних відповідностей або для побудови механізмів «порогу схожості», але має обмежену здатність масштабуватись до високорозмірних, сильно варіативних наборів ознак без додаткової обробки та квантизації.

Окрему групу становлять гібридні архітектури, що поєднують сильні сторони різних підходів. Модель CNN+MLP поєднує згорткові шари, які витягують локальні шаблони, із щільними шарами, що реалізують остаточне нелінійне відображення в класи. Така комбінація дозволяє структурувати навчання: перші шари виділяють інформативні ознаки, а фінальні шари виконують складну дискурсію класів. Інший варіант – поєднання трансформерних блоків з асоціативними мережами (Transformer+Hopfield), де механізм уваги трансформера захоплює глобальні залежності між ознаками, а асоціативний компонент забезпечує швидку перевірку прототипів і підвищену чутливість до рідкісних, але критичних патернів. Гібриди виявилися корисними, адже дозволили одночасно зберегти чутливість до дрібних відмінностей та узагальнювальну здатність моделі.

Із практичної точки зору важливими компонентами кожної моделі були методи боротьби з дисбалансом класів та стійкості моделі до шуму. Для цього використовувалися техніки перезапису ваг класів у функції втрат, застосування oversampling / undersampling до тренувальної множини та підхід focal loss у разі потреби підвищити вагу рідкісних прикладів. Регуляризація, рання зупинка та крос-валідація залишались стандартними механізмами контролю якості моделі. Розроблена система передбачає багаторівневу конвеєрну архітектуру: попередній фільтр (SOM, прості правила), основний класифікатор (MLP / CNN / гібрид) і модуль асоціативного контролю (Hopfield або порогова логіка) для остаточного ухвалення рішення.

Узагальнено, моделювання показало, що різні архітектури відповідають різним ролям у системі виявлення вторгнень: простіші щільні мережі придатні для базової класифікації, згорткові – для виявлення локальних патернів у багатовимірних ознаках, самоорганізуючі карти – для виявлення аномалій та візуалізації, а асоціативні мережі – для прототипного контролю і підвищення надійності спрацьовувань. Гібридні підходи дозволяють поєднати ці властивості й досягти більш гнучкої системи, яку можна налаштовувати під конкретні пріоритети захисту. Далі в статті розглядаються методологічні деталі реалізації і критерії порівняння моделей, що дозволяє обґрунтувати вибір архітектури під конкретні експлуатаційні умови.

Результати дослідження.

1. Self-Organizing Map (SOM) мережі

SOM показала дуже сильну здатність відокремлювати масивні класи: «dos» і «normal» мають винятково високі recalls (99.76 % і 99.50 % відповідно). Також SOM добре виділяє «probe» (88.26 %). Проте для рідкісних і тонких класів результат слабкий: «r2l» – 57.64 %, «u2r» – 0.00 %. Це типовий профіль: достатнє групування великих кластерів і помітно відмінні візуалізаційні кордони (U-matrix), але слабка чутливість до дуже рідкісних шаблонів, якщо вони не представлені достатньо в топології карти або після прив'язки BMU→лейбл.

2. Гібридні (CNN + MLP) мережі

Ця модель демонструє найкращі загальні результати в наборі: усі класи мають дуже високі recalls – всі близькі до 99 % і більше (u2r = 100.00 %, r2l = 99.56 %, probe ≈ 99.88 %, dos ≈ 99.34 %, normal ≈ 99.22 %), overall accuracy 99.32 %. Це вказує на те, що архітектура ефективно вловлює як масивні, так і рідкісні патерни. Такий профіль свідчить про дуже хорошу збалансованість навчання і високу дискримінаційну здатність мережі.

3. MLP (багатошаровий персептрон)

MLP дає високу загальну точність (99.20 %) і дуже добрий результат по «dos» і «normal» (≈99+ %), але абсолютно провалює класи r2l та u2r (обидва 0.00 %) і має лише 95.08 % для probe. Це означає, що модель навчилась «переводити» майже всі приклади в кілька домінуючих класів (особливо dos / normal), і втрачає рідкісні класи – ймовірно через дисбаланс і недостатньо складну архітектуру / функцію втрат для підсилення рідкісних класів.

4. Гібридні (Hopfield + Transformer) мережі

Гібридна модель дає змішаний результат: достатні показники для атакних класів (dos 98.66 %, r2l 99.47 %, u2r 98.08 %, probe 99.71 %), але значно гірший результат по «normal» – 88.16 %, через що overall падає до 96.60 %. Тобто модель добре розрізняє атаки, але часто зараховує нормальний трафік до атак (зниження recall (normal)). Це може вказувати на те, що комбінований вплив прототипної (біполярної) складової і трансформерного енкодера зміщує межу ухвалення рішень у бік більш «агресивного» виявлення атак, підвищивши recall атак, але за рахунок великого числа помилкових спрацьовувань на нормальні приклади.

5. Hopfield (мережа Хопфілда)

Hopfield-підхід дає середні / нижчі результати: загальна точність ~72.79 %, по атаках recall в межах 63–92 % (найкраще u2r ≈ 92.5 %, найгірше – dos/probe ≈ 64–72 %). Архітектура добре зберігає деякі кореляційні шаблони (через зовнішні добутки), але в цілому не дозволяє досягти конкурентних показників у задачі мультикласової класифікації великого, дисбалансного реального датасету: висока вартість обчислень для кожного зразка за умови одночасного відносно низького узагальнення.

6. CNN (згорткова мережа)

Чиста CNN дуже добре розпізнає «normal» (99.90 %) і «dos» (99.29 %) та має гарні результати для probe (93.86 %), але майже не виявляє r2l і u2r (r2l ≈ 0.98 %, u2r = 0.00 %). Отже, як і MLP, CNN фокусується на найбільш «помітних» шаблонах і слабко розпізнає дуже рідкісні атаки без додаткових механізмів балансування або архітектурних підсилень.

За поданими результатами, найзбалансованішою та найефективнішою моделлю є гібрид CNN+MLP, яка демонструє майже стовідсоткові показники у всіх п'яти класах (≈99 %+), тобто найкраще поєднує виявлення масових і рідкісних атак. Модель Transformer+Hopfield також дуже сильна в розпізнаванні атак (високі recalls для dos, r2l, u2r, probe), проте за рахунок цього суттєво страждає виявлення нормального трафіку (normal ≈88 %), що робить її «агресивним» детектором атак із підвищеною кількістю хибних тривог. Чисті архітектури (MLP і CNN) демонструють добрі результати для домінуючих класів (dos, normal), але виявилися неробочими для рідкісних атак (r2l, u2r близько 0 %), тобто недостатньо чутливі до небагаточисельних шаблонів.

SOM показала відмінне групування для dos і normal, а також пристойний probe, але слабко пропрацьовує дуже рідкісні u2r і частково r2l. Hopfield-підхід дав помірні результати в цілому (середній рівень точності), проте не конкурує з найкращими сучасними архітектурами

за загальною ефективністю. У підсумку, для універсального застосування найкращий вибір – гібрид CNN+MLP; якщо ж пріоритет – «не пропустити» атак будь-якою ціною (навіть ціною фп), варто розглянути Transformer+Hopfield.

На рис. 1 представлений графік результатів дослідження у вигляді матриці інтенсивностей, згенерований за отриманими даними найсучаснішим інструментом штучного інтелекту Gemini, який був наданий українським студентам абсолютно безкоштовно компанією Google.

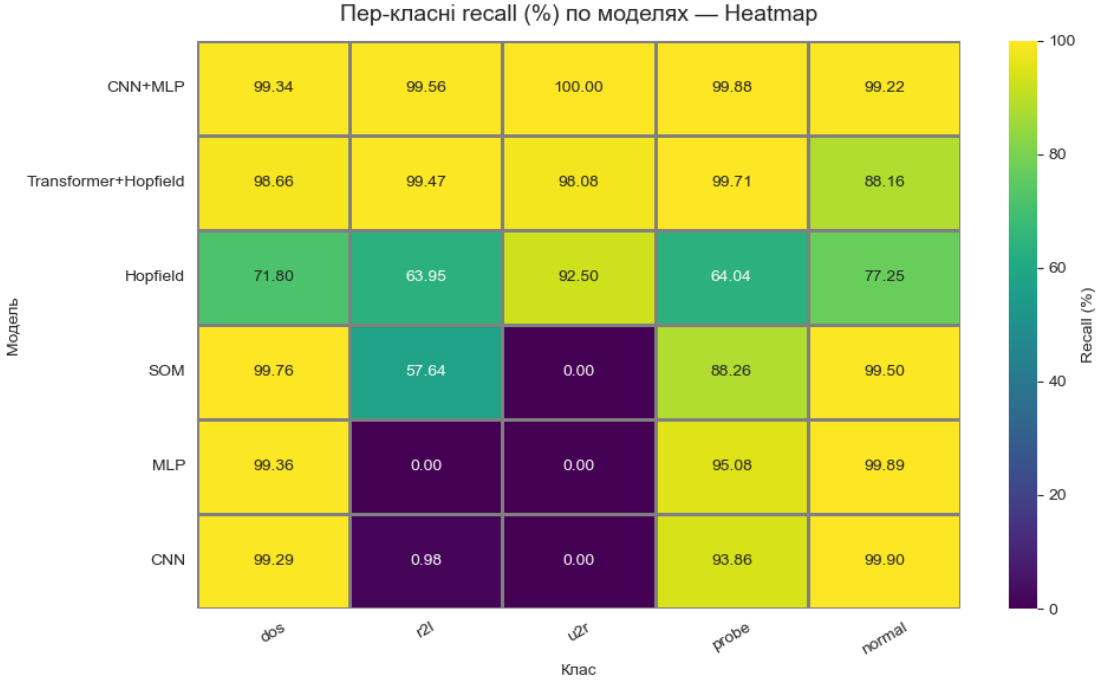

Рис. 1. Графік результатів дослідження у вигляді матриці інтенсивностей

Abstract. The paper investigates the application of various artificial neural network architectures (MLP, CNN, SOM, Hopfield and several hybrid configurations, in particular CNN+MLP and Transformer+Hopfield) to the problem of network intrusion detection based on the standard KDD'99 dataset. All models are implemented in *Python*.

Keywords: neural network, intrusion detection, KDD'99, CNN+MLP, Hopfield, SOM, R2L, U2R.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Malik M. Advancing Network Security Through Artificial Intelligence and Machine Learning: A Comprehensive Survey and Future Directions. *International Journal of Network Security Advances.* 2025. URL: https://surl.li/fscapi

2. Bensaoud A., Jugal Kalita. Optimized detection of cyber-attacks on IoT networks via hybrid deep learning models. *Ad Hoc Networks.* 2025. Vol. 170. DOI: 10.1016/j.adhoc.2025.103770.

3. Shalini S., Chennamaneni A., Sawyerr B. A Review of the Duality of Adversarial Learning in Network Intrusion: Attacks and Countermeasures. *Arhiv.* 2023. URL: https://arxiv.org/html/2412.13880v1

4. Zhong Meihui, Mingwei Lin, Chao Zgang. A survey on graph neural networks for intrusion detection systems: Methods, trends and challenges. *Computers and Security.* 2024. Vol. 141. DOI: 10.1016/j.cose.2024.103821.

5. Chhetri Bipin, Akbar Siami Namin. The Application of Transformer-Based Models for Predicting Consequences of Cyber Attacks. Proceedings of the IEEE International Conference on Computers, Software and Applications (COMPSAC). Toronto, Canada, 2025. 21 p. DOI: 10.48550/arXiv.2508.13030.

6. Neural Network Based Zero-Day-Attack Detection: A Machine Learning Approach to Enhancing Cybersecurity / Abhijeet Sharma, Akash Nigam, Jatin Chhaudhary, Sonal Shukla. *Proceedings of the International Conference on Innovative Computing and Communication (ICICC 2024).* URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4848658

7. Rasikha V., Marikkannu P. An ensemble deep learning-based cyber-attack detection system using optimization strategy. Knowledge-Based Systems. 2024. Vol. 301. DOI: 10.1016/j.knosys.2024.112211.