DOI 10.34185/1562-9945-2-163-2026-09
УДК 004.056.55

Б.В. Семенюк, Б.О. Савенко

МЕТОД ПІДВИЩЕННЯ ТОЧНОСТІ СИСТЕМИ ВИЯВЛЕННЯ АТАК У ЗОНАХ НЕВИЗНАЧЕНОСТІ НА ОСНОВІ АНАЛІЗУ ПОМИЛОК ТА СЕЛЕКТИВНОГО ПЕРЕВИЗНАЧЕННЯ РІШЕНЬ

Анотація. У статті запропоновано метод підвищення точності системи виявлення атак (IDS) у задачі бінарної класифікації мережевого трафіку за умов дисбалансу класів і наявності прикордонних прогнозів. Підхід реалізує двоступеневу схему прийняття рішень, що поєднує адаптацію навчальних даних на основі аналізу помилок і селективне перевизначення рішень у попередньо визначеній зоні невизначеності. На першому етапі базова згорткова нейронна мережа (CNN) навчається на спектрограмному поданні мережевих з'єднань, сформованому за допомогою Short-Time Fourier Transform (STFT). Після навчання виконується аналіз помилок класифікації на навчальній і валідаційній підвибірках на рівні підкласів атак. Підкласи, що формують найбільшу кількість хибних рішень, використовуються для формування розширеної навчальної вибірки шляхом цільового синтетичного доповнення (ErrorBoost). Далі на оновленому наборі даних з нуля навчається допоміжна модель IDS з ідентичною архітектурою.

Для обробки прикордонних прогнозів вводиться зона невизначеності як інтервал апостеріорних імовірностей, близьких до порогового значення прийняття рішення. Для зразків, імовірність яких потрапляє до цього інтервалу, рішення базової моделі вибірково перевіряється допоміжною моделлю. Остаточна класифікація визначається з урахуванням порогів упевненості, встановлених виключно на валідаційній підвибірці без використання тестових даних під час налаштування параметрів. Такий механізм забезпечує кероване регулювання компромісу між хибнопозитивними та хибнонегативними рішеннями, що є критичним для практичного застосування IDS.

Експериментальне оцінювання виконано на наборі даних NSL-KDD із дотриманням фіксованого протоколу розбиття на навчальну, валідаційну та тестову підвибірки. Використання лише ErrorBoost без селективного перевизначення не забезпечило стабільного покращення як самостійне рішення. Натомість комбінований підхід, що поєднує аналіз помилок та вибірковий перегляд рішень у зоні невизначеності, продемонстрував покращення показників. У кращому експериментальному запуску значення accuracy на тестовій підвибірці досягло 0,8522 за збереження збалансованих значень precision і recall для класу атак. Отримані результати підтверджують, що врахування фактичних помилок моделі та селективне перевизначення невизначених прогнозів дозволяє підвищити ефективність IDS без ускладнення архітектури та порушення коректності експериментальної процедури.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

Ключові слова: система виявлення атак, бінарна класифікація, дисбаланс класів, аналіз помилок, зона невизначеності, селективне перевизначення рішень, NSL-KDD, STFT, згорткова нейронна мережа.

Постановка проблеми. Системи виявлення атак у комп'ютерних мережах є одним із ключових компонентів забезпечення інформаційної безпеки, оскільки дозволяють автоматизовано ідентифікувати несанкціоновану активність та аномальну поведінку мережевого трафіку. У більшості прикладних сценаріїв intrusion detection розглядається як задача бінарної класифікації, у межах якої необхідно віднести кожен зразок до нормального або атакуючого класу. Особливістю такої задачі є висока вартість хибнонегативних рішень, коли атака помилково класифікується як нормальна активність.

Дисбаланс призводить до деградації якості виявлення атак навіть за використання складних моделей, зокрема глибоких нейронних мереж [1, 6]. Систематичні огляди також підкреслюють, що метрика загальної точності не відображає реальну здатність IDS виявляти рідкісні та складні атаки і суттєво залежить від протоколу оцінювання та структури набору даних [12, 18].

Поєднання синтетичного балансування з CNN може покращувати показники для міноритарних класів [17]. Однак у типових схемах балансування акцент робиться на вирівнюванні часток класів, тоді як «важкі» для моделі підкласи атак і прикордонні приклади залишаються недостатньо опрацьованими. Саме тому актуальними є адаптивні механізми навчання, що спираються на фактичні помилки моделі та коректний протокол експериментів [8, 10].

Окремою проблемою є наявність зон невизначеності прогнозу, у яких імовірність належності зразка до класу атаки наближається до порогового значення. У таких випадках моделі можуть демонструвати нестабільність, що призводить до зростання кількості хибнопозитивних або хибнонегативних рішень. Показано, що зменшення хибнонегативної складової можливе за рахунок додаткових механізмів обробки складних прикладів, однак у багатьох підходах не вводиться явне виділення зони невизначеності та вибірковий перегляд рішень [13].

Таким чином, актуальною є задача розроблення методу підвищення точності IDS, який поєднує адаптивне навчання на основі аналізу помилок моделі та механізм селективного перевизначення рішень у прикордонних (невизначених) випадках. Такий підхід має забезпечити зменшення кількості хибних рішень без використання тестових даних на етапах налаштування та без суттєвого ускладнення архітектури.

Аналіз останніх досліджень і публікацій. У роботі [1] досліджено виявлення та пом'якшення DDoS-атак у програмно-конфігурованих мережах на основі адаптивної ентропійної метрики, яка відстежує зміни статистичних характеристик потоків. Обмеженням такого підходу є орієнтація на конкретний клас атак і залежність від якості оцінювання ентропійних показників у сценаріях зі змішаним трафіком, що ускладнює перенесення на загальну IDS-задачу з різнорідними типами атак.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

У дослідженні [2] розглянуто застосування глибокого навчання для побудови IDS в IoT-середовищі з акцентом на автоматизоване виділення ознак. Водночас не запропоновано механізмів, які явно враховують дисбаланс підкласів атак, а також не проаналізовано поведінку моделі у випадках прогнозів із близькими до порогових значеннями ймовірності.

У роботі [3] досліджено багатошарові нейромережеві архітектури для підвищення якості IDS в IoT, однак підхід потребує ретельного налаштування та не містить явної стратегії, спрямованої на підкласи атак, що формують найбільшу кількість помилок.

Окремі прикладні роботи підкреслюють, що неоднорідність трафіку та відмінності сценаріїв експлуатації ускладнюють побудову універсальних детекторів без хибних спрацьовувань [6]. Систематичний огляд [7] наголошує на проблемах відтворюваності результатів і браку аналізу помилок на рівні підкласів атак. У порівняльному дослідженні [8] показано чутливість метрик до вибору датасету та протоколу розбиття, що робить принциповим суворе розмежування навчальної, валідаційної й тестової підвибірок у роботах з IDS.

Узагальнювальні роботи акцентують, що якість IDS визначається не лише архітектурою, а й передобробкою даних та узгодженим протоколом оцінювання [10]. Окремо підкреслюється вплив дисбалансу й асиметрії помилок (насамперед хибнонегативних), а також нестача алгоритмічних стратегій, які одночасно використовують інформацію про помилки моделі для адресного переформування навчальної вибірки та вводять явні критерії виділення зони невизначеності [12, 18].

Для синтетичного доповнення даних, окрім SMOTE, застосовують генеративні моделі (GAN) для формування зразків рідкісних атак або доменно-специфічних сценаріїв [11, 15]. Водночас якість синтетичних прикладів істотно залежить від стабільності навчання генератора та може супроводжуватися артефактами, а також не завжди визначено, для яких саме підкласів атак доповнення є доцільним.

У підсумку, сучасні підходи охоплюють глибоке навчання, синтетичне балансування та генерацію прикладів, а також окремі прийоми зниження хибнонегативної складової [13, 14]. Разом із тим у розглянутій літературі бракує методів, які водночас (1) адресно «підсилюють» саме ті підкласи атак, де базова модель помиляється найчастіше, і (2) переглядають рішення лише в прикордонній зоні прогнозу, не залучаючи тестові дані під час налаштування. Саме цю прогалину заповнює запропонований у роботі підхід.

Метою цієї статті є розроблення методу підвищення точності системи виявлення атак у задачі бінарної класифікації мережевого трафіку на основі поєднання аналізу помилок базової моделі та механізму селективного перевизначення рішень у зоні невизначеності прогнозу. Метод спрямований на зменшення кількості хибних рішень шляхом цільового формування навчальної вибірки для підкласів атак, на яких модель демонструє найбільшу частоту помилок, а також на контроль прийняття рішень у прикордонних випадках без використання тестових даних на етапах налаштування.

Викладення основного матеріалу. У даному дослідженні задача виявлення атак формулюється як бінарна класифікація мережевого трафіку, у межах якої кожен зразок

x_iнеобхідно віднести до одного з двох класів $y_i \in \{0,1\}$, де значення 0 відповідає нормальному трафіку, а значення 1 – атакуючій активності. Сукупність навчальних даних подається у вигляді множини

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}, \quad (1)$$

де N— кількість зразків. Вхідні дані представлені табличними ознаками набору NSL-KDD, що характеризують статистичні та протокольні властивості мережевих з'єднань.

Для забезпечення узгодженості масштабу всі числові ознаки підлягають нормалізації, яка виконується відповідно до виразу

$$x_i^{(norm)} = 2 \cdot \frac{x_i - \min(x)}{\max(x) - \min(x)} - 1, \quad (2)$$

що дозволяє привести значення ознак до симетричного інтервалу та зменшити вплив різних масштабів на подальші перетворення. Нормовані ознаки використовуються для формування синтетичного сигналу, параметри якого визначаються значеннями вхідних характеристик.

З метою підвищення виразності ознак застосовано перетворення мережевого трафіку в частотно-часове представлення. Для цього використовується Short-Time Fourier Transform (STFT), який для дискретного сигналу $s(t)$визначається відповідно до виразу

$$\mathrm{STFT}(t, f) = \sum_{\tau=-\infty}^{\infty} s(\tau)\, w(\tau - t)\; e^{-j2\pi f \tau}, \quad (3)$$

де $w(\cdot)$є віконною функцією. Отримане спектральне подання дозволяє сформувати двовимірну спектрограму, що відображає локальні частотні зміни сигналу в часі.

Сформовані спектрограми використовуються як вхідні дані для Convolutional Neural Network (CNN), параметри якої оптимізуються шляхом мінімізації функції втрат для бінарної класифікації. Ймовірність належності зразка до класу атаки визначається вихідним шаром мережі відповідно до сигмоїдної функції

$$p_i = \sigma(z_i) = \frac{1}{1 + e^{-z_i}}, \quad (4)$$

де z_iє агрегованим виходом нейронної мережі для зразка x_i.

Навчання базової моделі здійснюється шляхом мінімізації бінарної крос-ентропійної функції втрат

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \ln(p_i) + (1 - y_i)\ln(1 - p_i)], \quad (5)$$

що забезпечує узгоджене оновлення параметрів мережі для коректного розділення нормального та атакуючого трафіку. Після завершення навчання формується прогноз базо-

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

вої моделі для навчальної та валідаційної вибірок, який використовується для аналізу помилок класифікації.

Аналіз помилок базової моделі полягає у визначенні підкласів атак, для яких кількість хибних рішень є максимальною. Для цього вводиться індикатор помилки

$$e_i = \begin{cases} 1, & y_i \neq \hat{y}_i, \\ 0, & y_i = \hat{y}_i, \end{cases} \quad (6)$$

де $\hat{y}_i$ є прогнозованим класом. Агрегація значень e_i для кожного підкласу атак дозволяє сформувати множину проблемних типів, що роблять найбільший внесок у загальну похибку системи.

На основі отриманої статистики формується розширена навчальна вибірка шляхом додаткового синтетичного збільшення представлення проблемних підкласів. Для кожного такого підкласу визначається цільова кількість зразків

$$N_k^* = \min(\alpha \cdot N_k, N_{\max}), \quad (7)$$

де N_k є початковою кількістю зразків відповідного підкласу, а $N_{\max}$ обмежує максимальний розмір класу у вибірці. Отриманий підхід, надалі позначений як ErrorBoost, дозволяє адаптувати структуру навчальних даних до реальної поведінки моделі.

На розширеній вибірці з нуля навчається додаткова модель IDS з ідентичною архітектурою, яка використовується для аналізу складних випадків класифікації. Для формалізації таких випадків вводиться зона невизначеності прогнозу, що визначається інтервалом

$$\tau = \{ p_i \mid \tau_L \leq p_i \leq \tau_U \}, \quad (8)$$

де τ_L та τ_U є нижньою та верхньою межами зони невизначеності, визначеними на валідаційній вибірці.

Для зразків, що потрапляють до зони невизначеності, застосовується механізм селективного перевизначення рішень. Остаточне рішення формується шляхом порівняння прогнозів базової та додаткової моделей і визначається відповідно до правила

$$\hat{y}_i^{(final)} = \begin{cases} 1, & p_i^{(boost)} \geq \theta_U, \\ 0, & p_i^{(boost)} \leq \theta_L, \\ \hat{y}_i^{(base)}, & \text{в інших випадках}, \end{cases} \quad (9)$$

де θ_L та θ_U є порогами впевненості допоміжної моделі. Такий підхід дозволяє зменшити кількість хибних рішень у прикордонних випадках без неконтрольованого зростання хибнопозитивних спрацьовувань.

Запропонований метод інтегрує аналіз помилок, цільове синтетичне збільшення навчальних даних і двоступеневе прийняття рішень у єдиний конвеєр побудови системи виявлення атак. Усі параметри методу визначаються виключно на навчальній та валідаційній вибірках, що забезпечує коректність оцінювання та можливість практичного застосування підходу.

Результати експериментів. Експериментальні дослідження запропонованого методу виконано з використанням набору даних NSL-KDD, який було розділено на навчальну, валідаційну та тестову підвибірки відповідно до фіксованого протоколу. Тестова вибірка не використовувалася на жодному з етапів налаштування параметрів методу, зокрема під час формування ErrorBoost-вибірки та визначення порогів селективного перевизначення рішень. Оцінювання якості класифікації здійснювалося за стандартними метриками precision, recall, f1-score та accuracy.

Результати застосування моделі, навченої на ErrorBoost-вибірці без використання механізму селективного перевизначення рішень, показали, що така модель не забезпечує стабільного покращення загальної точності порівняно з базовою. Для цього варіанту значення accuracy на тестовій вибірці становило 0,8210, при цьому спостерігалося зростання кількості хибнонегативних рішень для класу атак, що підтверджується значенням recall, рівним 0,7886. Отриманий результат свідчить про те, що додаткове синтетичне збільшення навчальної вибірки для проблемних підкласів атак не може розглядатися як самодостатня заміна базової моделі та потребує інтеграції з механізмами контролю прийняття рішень.

Для оцінювання впливу селективного перевизначення рішень було введено зону невизначеності прогнозу з межами $\tau_L = 0{,}25$ та $\tau_U = 0{,}75$. На валідаційній вибірці обсяг зони невизначеності становив 944 зразки, для яких було виконано автоматичний підбір порогів упевненості допоміжної моделі. За отриманих значень порогів частка перевизначених рішень у межах зони невизначеності склала 90,25 %. При цьому кількість хибнонегативних рішень зменшилася з 127 до 36, тоді як кількість хибнопозитивних зросла з 174 до 226, що призвело до зменшення сумарної кількості помилок у зоні невизначеності з 301 до 262.

Аналогічний аналіз було виконано на тестовій вибірці, де зона невизначеності охоплювала 3122 зразки. Частка перевизначених рішень у цій зоні становила 89,08 %. У межах зони невизначеності спостерігалося зменшення кількості хибнонегативних рішень з 1038 до 949 за одночасного зростання кількості хибнопозитивних з 385 до 596. Така динаміка підтверджує наявність керованого компромісу між різними типами помилок, який формується внаслідок застосування селективного перевизначення рішень і залежить від обраних порогів упевненості.

Загальна оцінка якості класифікації для комбінованого підходу, що поєднує базову модель та допоміжну модель, застосовану в зоні невизначеності, показала значення accuracy 0,8117 на тестовій вибірці. Для класу нормального трафіку значення recall становило 0,9155, тоді як для класу атак — 0,7683, що відображає перерозподіл помилок у напрямі зменшення частки пропущених атак у прикордонних випадках порівняно з базовим рішенням.

У кращому експериментальному запуску, отриманому за оптимального поєднання параметрів ErrorBoost та порогів селективного перевизначення рішень, загальна точність класифікації на тестовій вибірці досягла 0,8522. Для цього варіанту значення precision та recall для класу атак становили 0,8907 та 0,8438 відповідно, а середнє зна-

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

чення f1-score склало 0,8504. Отриманий результат демонструє підвищення показника accuracy приблизно на 1 відсотковий пункт порівняно з базовим підходом, що підтверджує доцільність поєднання аналізу помилок та селективного перевизначення рішень у зонах невизначеності прогнозу.

Таблиця 1

Порівняння результатів для різних варіантів

Варіант	Accuracy	Precision (Attack)	Recall (Attack)	F1-score (Attack)
Базова модель	0,8117	—	—	—
ErrorBoost без перевизначення	0,8210	—	0,7886	—
Комбінований (кращий запуск)	0,8522	0,8907	0,8438	0,8504

Джерело: сформовано автором

Висновки. У роботі запропоновано метод підвищення точності системи виявлення атак у задачі бінарної класифікації мережевого трафіку, що ґрунтується на поєднанні аналізу помилок базової моделі та механізму селективного перевизначення рішень у зонах невизначеності прогнозу. Метод орієнтований на врахування фактичної поведінки моделі під час навчання та прийняття рішень і не потребує використання тестових даних на етапах налаштування.

Показано, що застосування цільового синтетичного збільшення навчальної вибірки для підкласів атак, на яких базова модель демонструє найбільшу кількість помилок, дозволяє сформувати допоміжну модель, чутливішу до складних і рідкісних типів атак. Водночас експериментально встановлено, що така модель не забезпечує стабільного покращення загальної точності при використанні як самостійне рішення, що обґрунтовує доцільність її застосування виключно у складі комбінованого підходу.

Запропонований механізм селективного перевизначення рішень у зоні невизначеності дозволяє керовано впливати на співвідношення хибнопозитивних і хибнонегативних рішень. Результати експериментів на наборі даних NSL-KDD показали, що за рахунок вибіркового залучення допоміжної моделі у прикордонних випадках можливо зменшити кількість пропущених атак без неконтрольованого зростання загальної кількості помилок. У кращому експериментальному запуску досягнуто підвищення показника accuracy з 0,84 до 0,85 на тестовій вибірці.

Отримані результати підтверджують ефективність запропонованого підходу для задач виявлення атак у комп’ютерних мережах з дисбалансом класів і наявністю зон невизначеності прогнозу. Подальші дослідження можуть бути спрямовані на аналіз чутливості методу до вибору меж зони невизначеності, адаптацію підходу до багатокласових сценаріїв та перевірку його узагальнювальної здатності на інших наборах даних мережевого трафіку.

ЛІТЕРАТУРА

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

2. Joseph, J. E., Aleke, N. T., & Onyeanisi, O. P. (2025). Deep learning based intrusion detection system for network security in IoT system. *International Journal of Education, Management, and Technology*, 3(1), 119–138. doi: 10.58578/ijemt.v3i1.4539.

3. Farooq, M., & Ahmad, F. (2024). Improved intrusion detection in IoT using multi-layered neural architectures. *International Journal of Computing, 23*(2), 268–273 doi:10.47839/ijc.23.2.3546.

4. Kashtalian, A., Sergii, L., Sachenko, A., Savenko, B., Savenko, O. & Nicheporuk, A. (2025). Evaluation criteria of centralization options in the architecture of multicomputer systems with traps and baits. *Radioelectronic and Computer Systems, 2025*(1), 264–297. doi:10.32620/reks.2025.1.18.

5. Denysiuk, D., Savenko, O., Lysenko, S., Savenko, B., & Kashtalian, A. (2023). Method for detecting steganographic changes in images using machine learning. In Proceedings of the 13th International Conference on Dependable Systems, Services and Technologies (DESSERT) (pp. 1–6). Athens: IEEE. doi: 10.1109/DESSERT61349.2023.10416453.

6. Alladi, T., Chamola, V., Sikdar, B., & Choo, K.-K. R. (2020). Consumer IoT: Security Vulnerability Case Studies and Solutions. IEEE Consumer Electronics Magazine, 9(2), 17–25. doi:10.1109/MCE.2019.2953740

7. Hussain, A., Sharif, H., Rehman, F., Kirn, H., Sadiq, A., & Khan, M. S. (2023). A Systematic Review of Intrusion Detection Systems in Internet of Things Using ML and DL. 2023 4th International Conference on Computing, Mathematics and Engineering Technologies (iCoMET). doi:10.1109/iCoMET57998.2023.10099142

8. Kilincer, I. F., Ertam, F., & Sengur, A. (2021). Machine learning methods for cyber security intrusion detection: Datasets and comparative study. Computer Networks, 188, article 107840. doi:10.1016/j.comnet.2021.107840

9. Li, G., & Jung, J. J. (2023). Deep learning for anomaly detection in multivariate time series: Approaches, applications, and challenges. Information Fusion, 91, 93–102. doi:10.1016/j.inffus.2022.10.008

10. Sheikh, M. S., & Peng, Y. (2022). Procedures, Criteria, and Machine Learning Techniques for Network Traffic Classification: A Survey. IEEE Access, 10, 64806–64829. doi:10.1109/access.2022.3181135

11. Mari, A.-G., Zinca, D., & Dobrota, V. (2023). Development of a Machine-Learning Intrusion Detection System and Testing of Its Performance Using a Generative Adversarial Network. Sensors, 23(3), 1315. doi:10.3390/s23031315

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

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

13. Mijalkovic, J., & Spognardi, A. (2022). Reducing the False Negative Rate in Deep Learning Based Network Intrusion Detection Systems. Algorithms, 15(8), 258. doi:10.3390/a15080258
14. Shahriar, M. H., Haque, N. I., Rahman, M. A., & Alonso, M. (2020). G-IDS: Generative Adversarial Networks Assisted Intrusion Detection System. 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC). doi:10.1109/compsac48688.2020.0-218
15. Cai, Z., Du, H., Wang, H., Zhang, J., Si, Y., & Li, P. (2023). One-Dimensional Convolutional Wasserstein Generative Adversarial Network Based Intrusion Detection Method for Industrial Control Systems. Electronics, 12(22), 4653. doi:10.3390/electronics12224653
16. Baich, M., & Sael, N. (2025). Enhancing Machine Learning Model Prediction with Feature Selection for Botnet Intrusion Detection. Engineering Proceedings, 112(1), 55. doi:10.3390/engproc2025112055
17. Hassannataj Joloudari, J., Marefat, A., Nematollahi, M. A., Oyelere, S. S., & Hussain, S. (2023). Effective Class-Imbalance Learning Based on SMOTE and Convolutional Neural Networks. Applied Sciences, 13(6), 4006. doi:10.3390/app13064006
18. Yang, Z., Liu, X., Li, T., Wu, D., Wang, J., Zhao, Y., & Han, H. (2022). A systematic literature review of methods and datasets for anomaly-based network intrusion detection. Computers & Security, 116, 102675. doi:10.1016/j.cose.2022.102675
19. Maniriho, P., Niyigaba, E., Bizimana, Z., Twiringiyimana, V., Mahoro, L. J., & Ahmad, T. (2020). Anomaly-based Intrusion Detection Approach for IoT Networks Using Machine Learning. 2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM). doi:10.1109/CENIM51130.2020.9297958
20. Sheibani, M., Konur, S., Awan, I., & Qureshi, A. (2024). A Multi-Layered Defence Strategy against DDoS Attacks in SDN/NFV-Based 5G Mobile Networks. Electronics, 13(8), 1515. doi:10.3390/electronics13081515
21. Sathaporn, P., Krungseanmuang, W., Chaowalittawin, V., Benjangkaprasert, C., & Purahong, B. (2025). DDoS Detection Using a Hybrid CNN–RNN Model Enhanced with Multi-Head Attention for Cloud Infrastructure. Applied Sciences, 15(21), 11567. doi:10.3390/app152111567

REFERENCES

1. Dalou', J., Al-Duwairi, B., & Al-Jarrah, M. (2020). Adaptive entropy-based detection and mitigation of DDoS attacks in SDN networks. *International Journal of Computing*, 19(3), 399–410. doi: 10.47839/ijc.19.3.1889.
2. Joseph, J. E., Aleke, N. T., & Onyeanisi, O. P. (2025). Deep learning based intrusion detection system for network security in IoT system. *International Journal of Education, Management, and Technology*, 3(1), 119–138. doi: 10.58578/ijemt.v3i1.4539.
3. Farooq, M., & Ahmad, F. (2024). Improved intrusion detection in IoT using multi-layered neural architectures. *International Journal of Computing, 23*(2), 268–273. doi:10.47839/ijc.23.2.3546.

4. Kashtalian, A., Sergii, L., Sachenko, A., Savenko, B., Savenko, O. & Nicheporuk, A. (2025). Evaluation criteria of centralization options in the architecture of multicomputer systems with traps and baits. *Radioelectronic and Computer Systems, 2025*(1), 264–297. doi:10.32620/reks.2025.1.18.
5. Denysiuk, D., Savenko, O., Lysenko, S., Savenko, B., & Kashtalian, A. (2023). Method for detecting steganographic changes in images using machine learning. In Proceedings of the 13th International Conference on Dependable Systems, Services and Technologies (DESSERT) (pp. 1–6). Athens: IEEE. doi: 10.1109/DESSERT61349.2023.10416453.
6. Alladi, T., Chamola, V., Sikdar, B., & Choo, K.-K. R. (2020). Consumer IoT: Security Vulnerability Case Studies and Solutions. IEEE Consumer Electronics Magazine, 9(2), 17–25. doi:10.1109/MCE.2019.2953740
7. Hussain, A., Sharif, H., Rehman, F., Kirn, H., Sadiq, A., & Khan, M. S. (2023). A Systematic Review of Intrusion Detection Systems in Internet of Things Using ML and DL. 2023 4th International Conference on Computing, Mathematics and Engineering Technologies (iCoMET). doi:10.1109/iCoMET57998.2023.10099142
8. Kilincer, I. F., Ertam, F., & Sengur, A. (2021). Machine learning methods for cyber security intrusion detection: Datasets and comparative study. Computer Networks, 188, article 107840. doi:10.1016/j.comnet.2021.107840
9. Li, G., & Jung, J. J. (2023). Deep learning for anomaly detection in multivariate time series: Approaches, applications, and challenges. Information Fusion, 91, 93–102. doi:10.1016/j.inffus.2022.10.008
10. Sheikh, M. S., & Peng, Y. (2022). Procedures, Criteria, and Machine Learning Techniques for Network Traffic Classification: A Survey. IEEE Access, 10, 64806–64829. doi:10.1109/access.2022.3181135
11. Mari, A.-G., Zinca, D., & Dobrota, V. (2023). Development of a Machine-Learning Intrusion Detection System and Testing of Its Performance Using a Generative Adversarial Network. Sensors, 23(3), 1315. doi:10.3390/s23031315
12. Ahmad, Z., Khan, A. S., Shiang, C. W., Abdullah, J., & Ahmad, F. (2021). Network intrusion detection system: A systematic study of machine learning and deep learning approaches. Emerging Telecommunications Technologies, 32(1), e4150. doi:10.1002/ett.4150
13. Mijalkovic, J., & Spognardi, A. (2022). Reducing the False Negative Rate in Deep Learning Based Network Intrusion Detection Systems. Algorithms, 15(8), 258. doi:10.3390/a15080258
14. Shahriar, M. H., Haque, N. I., Rahman, M. A., & Alonso, M. (2020). G-IDS: Generative Adversarial Networks Assisted Intrusion Detection System. 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC). doi:10.1109/compsac48688.2020.0-218
15. Cai, Z., Du, H., Wang, H., Zhang, J., Si, Y., & Li, P. (2023). One-Dimensional Convolutional Wasserstein Generative Adversarial Network Based Intrusion Detection Method for Industrial Control Systems. Electronics, 12(22), 4653. doi:10.3390/electronics12224653

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

16. Baich, M., & Sael, N. (2025). Enhancing Machine Learning Model Prediction with Feature Selection for Botnet Intrusion Detection. Engineering Proceedings, 112(1), 55. doi:10.3390/engproc2025112055
17. Hassannataj Joloudari, J., Marefat, A., Nematollahi, M. A., Oyelere, S. S., & Hussain, S. (2023). Effective Class-Imbalance Learning Based on SMOTE and Convolutional Neural Networks. Applied Sciences, 13(6), 4006. doi:10.3390/app13064006
18. Yang, Z., Liu, X., Li, T., Wu, D., Wang, J., Zhao, Y., & Han, H. (2022). A systematic literature review of methods and datasets for anomaly-based network intrusion detection. Computers & Security, 116, 102675. doi:10.1016/j.cose.2022.102675
19. Maniriho, P., Niyigaba, E., Bizimana, Z., Twiringiyimana, V., Mahoro, L. J., & Ahmad, T. (2020). Anomaly-based Intrusion Detection Approach for IoT Networks Using Machine Learning. 2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM). doi:10.1109/CENIM51130.2020.9297958
20. Sheibani, M., Konur, S., Awan, I., & Qureshi, A. (2024). A Multi-Layered Defence Strategy against DDoS Attacks in SDN/NFV-Based 5G Mobile Networks. Electronics, 13(8), 1515. doi:10.3390/electronics13081515
21. Sathaporn, P., Krungseanmuang, W., Chaowalittawin, V., Benjangkaprasert, C., & Purahong, B. (2025). DDoS Detection Using a Hybrid CNN–RNN Model Enhanced with Multi-Head Attention for Cloud Infrastructure. Applied Sciences, 15(21), 11567. doi:10.3390/app152111567

Received 10.03.2026
Accepted 13.03.2026
Published 31.03.2026

A method for improving the accuracy of an intrusion detection system in uncertainty zones based on error analysis and selective decision revision

The paper proposes a method for improving the accuracy of an intrusion detection system (IDS) in the task of binary classification of network traffic under class imbalance and the presence of borderline predictions. The approach is based on a two-stage decision scheme that combines error-driven data adaptation and selective decision revision within a predefined uncertainty zone. At the first stage, a baseline convolutional neural network (CNN) model is trained using spectrogram-based representations of network connections obtained via Short-Time Fourier Transform (STFT). After training, classification errors on the training and validation subsets are analyzed at the level of attack subclasses. Subclasses that contribute the largest number of false decisions are identified and used to form an extended training set through targeted synthetic oversampling (ErrorBoost). A secondary IDS model with the same architecture is then trained from scratch on the extended dataset.

To handle borderline predictions, an uncertainty zone is introduced as an interval of posterior probabilities close to the decision threshold. For samples whose predicted probability falls within this interval, the decision of the baseline model is selectively re-evaluated using the auxiliary model. Final classification is determined according to confidence thresholds defined exclusively on the validation subset, without using test data during parameter tuning. Such a mechanism enables controlled adjustment of the trade-off between false positive and false negative rates, which is critical in practical IDS deployment.

Experimental evaluation was conducted on the NSL-KDD dataset using a fixed protocol with separate training, validation, and test subsets. The ErrorBoost strategy alone did not provide stable improvements when applied as an independent solution. However, the combined approach integrating error-based oversampling and selective decision revision achieved improved performance. In the best experimental run, the overall accuracy reached 0.8522 on the test subset, while maintaining balanced precision and recall for the attack class. The results confirm that incorporating model-specific error analysis and selective re-evaluation of uncertain predictions can enhance IDS performance without increasing architectural complexity or violating experimental validity.

Keywords: *intrusion detection system, binary classification, class imbalance, error analysis, uncertainty zone, selective decision revision, NSL-KDD, STFT, convolutional neural network.*

Семенюк Богдан Васильович - аспірант, спеціальність 122 Комп'ютерні науки, Хмельницький національний університет.
ORCID: https://orcid.org/0009-0001-8831-8835
Савенко Богдан Олегович - старший викладач, доктор філософії з комп'ютерної інженерії, Хмельницький національний університет.
ORCID: https://orcid.org/0000-0001-5647-9979

Semenyuk Bohdan Vasylovych -- Postgraduate Student, Specialty 122 Computer Science, Khmelnytskyi National University.
ORCID: https://orcid.org/0009-0001-8831-8835
Savenko Bohdan Olehovych - Senior Lecturer, PhD in Computer Engineering, Khmelnytskyi National University.
ORCID: https://orcid.org/0000-0001-5647-9979

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)