ХМЕЛЬЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МІНІСТЕРСТВА ОСВІТИ І НАУКИ УКРАЇНИ

Кваліфікаційна наукова праця
на правах рукопису

СЕМЕНЮК БОГДАН ВАСИЛЬОВИЧ

УДК 004.3:004.75:004.49

ДИСЕРТАЦІЯ

МОДЕЛІ ТА МЕТОДИ ВИЯВЛЕННЯ КОМП'ЮТЕРНИХ АТАК В КОРПОРАТИВНИХ МЕРЕЖАХ НА ОСНОВІ ЧАСТОТНО-ЧАСОВОГО ПРЕДСТАВЛЕННЯ МЕРЕЖНОГО ТРАФІКУ

122 Комп'ютерні науки
(шифр і назва спеціальності)
12 Інформаційні технології
(галузь знань)

Подається на здобуття наукового ступеня доктора філософії
Дисертація містить результати власних досліджень. Подані до захисту наукові положення є власним напрацюванням. Всі використані ідеї, наукові результати, цитати супроводжуються належними посиланнями на їх авторів та джерела опублікування. Всі частини тексту дисертації, під час написання яких використовувались технології штучного інтелекту, перевірені та відредаговані особисто.

_____ Б. В. Семенюк
(підпис)

Наукові керівники:
Корецька Людмила Олександрівна, кандидат технічних наук, доцент
Савенко Богдан Олегович, доктор філософії

Хмельницький – 2026

АНОТАЦІЯ

Семенюк Б. В. Моделі та методи виявлення комп'ютерних атак в корпоративних мережах на основі частотно-часового представлення мережного трафіку. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії з галузі знань 12 Інформаційні технології за спеціальністю 122 Комп'ютерні науки. – Хмельницький національний університет, Хмельницький, 2026.

Розв'язання науково-прикладної задачі підвищення точності та збалансованості виявлення комп'ютерних атак у корпоративних мережах на основі частотно-часового представлення мережного трафіку, адаптивного балансування навчальної вибірки та селективного перевизначення рішень у зоні невизначеності прогнозу є актуальним напрямом досліджень у галузі інформаційної безпеки. Актуальність задачі зумовлена зростанням інтенсивності мережного обміну, ускладненням структури атак, появою нових і комбінованих сценаріїв вторгнень, а також суттєвим дисбалансом між нормальним трафіком і рідкісними типами атак, що знижує ефективність традиційних систем виявлення вторгнень (СВВ).

У дисертації здійснено аналіз сучасних систем і методів виявлення вторгнень, способів подання мережних даних для задач інтелектуального аналізу та підходів до застосування машинного і глибинного навчання в СВВ. Обґрунтовано доцільність використання соніфікації мережного трафіку як засобу перетворення табличного опису мережних з'єднань у частотно-часове подання, придатне для подальшого спектрального аналізу та автоматизованого розпізнавання атак за допомогою двовимірних згорткових нейронних мереж. Розроблено модель процесу виявлення комп'ютерних атак, метод частотно-часового перетворення мережного трафіку, метод сигнатурозбережного адаптивного балансування навчальної вибірки, метод підвищення точності виявлення атак у зоні невизначеності прогнозу, а також виконано постановку та проведення експериментальних досліджень.

Об'єкт дослідження – процес виявлення комп'ютерних атак у корпоративних мережах.

Предмет дослідження – моделі та методи виявлення комп'ютерних атак у корпоративних мережах на основі частотно-часового представлення мережного трафіку, соніфікації, адаптивного балансування навчальної вибірки та аналізу помилок класифікації.

Метою дисертаційного дослідження є підвищення точності та збалансованості виявлення комп'ютерних атак у корпоративних мережах шляхом розроблення моделей і методів частотно-часового представлення мережного трафіку, формування інформативних спектральних ознак, адаптивного балансування навчальної вибірки та селективного перевизначення рішень у зоні невизначеності прогнозу.

Наукова новизна отриманих результатів полягає в наступному:

1) вперше розроблено модель процесу виявлення комп'ютерних атак на основі соніфікації мережного трафіку, особливістю якої є інтеграція звукового перетворення багатовимірного простору ознак, спектрального аналізу, адаптивного балансування навчальної вибірки та каскадного механізму прийняття рішень у межах узагальненої моделі процесу виявлення атак, що дає змогу узгодити етапи навчання і функціонування моделі та підвищити збалансованість класифікації в умовах дисбалансу класів;

2) розроблено новий метод частотно-часового перетворення мережного трафіку, який, на відміну від відомих векторних підходів до представлення ознак, передбачає інтерпретацію багатовимірного простору параметрів з'єднання як параметрів сигналу з лінійним частотно-амплітудним відображенням і подальшим спектральним аналізом, що дає змогу формувати структуроване 2D-подання трафіку та підвищити інформативність вхідних даних для задачі класифікації атак;

3) розроблено новий метод сигнатурозбережного адаптивного балансування навчальної вибірки, який, на відміну від відомих процедур надсемплінгу та підвибірки, характеризується збереженням структурних характеристик міноритарного класу та врахуванням помилок базової моделі під час формування

розширеної навчальної множини, що дає змогу зменшити вплив дисбалансу класів на функцію ризику, підвищити повноту виявлення атак і збалансованість класифікації;

4) розроблено новий метод підвищення точності виявлення атак у зоні невизначеності прогнозу, який, на відміну від відомих порогових схем прийняття рішень, передбачає формалізацію τ-інтервалу невизначеності та селективне перевизначення рішень на основі допоміжної моделі, навченої на помилкових прикладах базового класифікатора, що дає змогу зменшити кількість хибнонегативних рішень без порушення принципу незалежності тестової вибірки.

Практичне значення отриманих результатів. Розроблено модель процесу виявлення комп'ютерних атак у корпоративних мережах та методи частотно-часового перетворення мережного трафіку, сигнатурозбережного адаптивного балансування навчальної вибірки і селективного підвищення точності в зоні невизначеності прогнозу, які дають змогу формувати інформативне спектральне подання мережних з'єднань, застосовувати двовимірні згорткові нейронні мережі для автоматизованого виявлення атак та зменшувати негативний вплив дисбалансу класів на якість класифікації. Особливістю розробленого підходу є поєднання в єдиному соніфікаційному конвеєрі етапів перетворення мережного трафіку у хвильову форму, побудови спектрограм, навчання базового класифікатора, адаптивного балансування навчальної вибірки та селективного перевизначення рішень у τ-зоні невизначеності, що забезпечує можливість використання запропонованого рішення під час створення і модернізації систем моніторингу безпеки корпоративних мереж.

У результаті проведених експериментальних досліджень підтверджено працездатність соніфікаційного підходу як альтернативного каналу подання мережних з'єднань для задачі виявлення вторгнень, а також ефективність каскадної схеми, що поєднує балансування SPAS, помилково-орієнтоване підсилення проблемних підкласів атак, визначення τ-зони та селективне перевизначення рішень. Застосування запропонованого підходу забезпечило відтворюване

покращення якості розпізнавання, а значення F1-міри на тестовій множині зросло приблизно на 1 відсоток порівняно з базовою соніфікованою моделлю.

Теоретичні та практичні результати дослідження впроваджені в ТОВ «ІТТ» (м. Хмельницький), ПП «Авіві» (м. Хмельницький), а також в освітньому процесі Хмельницького національного університету при викладанні дисциплін на кафедрі комп'ютерної інженерії та інформаційних систем для спеціальності F7 Комп'ютерна інженерія, зокрема в курсах «Безпека та захист комп'ютерних систем», «Моделювання та методи оптимізації в наукових та експериментальних дослідженнях», «Методології забезпечення якості, надійності, гарантоздатності та безпеки комп'ютерних систем та мереж».

У вступі представлено обґрунтування актуальності науково-прикладної задачі підвищення точності та збалансованості виявлення комп'ютерних атак у корпоративних мережах в умовах зростання інтенсивності мережного обміну, ускладнення структури атак, появи нових комбінованих сценаріїв вторгнень та наявності суттєвого дисбалансу між нормальним трафіком і рідкісними типами атак. Визначено зв'язок тематики дисертаційного дослідження з актуальними напрямами розвитку систем виявлення вторгнень, методів машинного та глибинного навчання, а також засобів інтелектуального аналізу мережного трафіку. Сформульовано мету, об'єкт, предмет, завдання та методи дослідження, наведено відомості про наукову новизну, практичне значення отриманих результатів, їх апробацію, публікації та впровадження. Також у вступі обґрунтовано доцільність використання частотно-часового представлення мережного трафіку як перспективного напряму підвищення інформативності вхідних даних для інтелектуального виявлення атак.

У першому розділі здійснено аналіз предметної області дослідження, сучасних систем і методів виявлення вторгнень, а також підходів до класифікації мережного трафіку в задачах інформаційної безпеки. Розглянуто особливості сигнатурних, статистичних, поведінкових та інтелектуальних СВВ, проаналізовано переваги і недоліки використання традиційного векторного подання мережних ознак, а також досліджено можливості застосування машинного і глибинного

навчання для розпізнавання комп’ютерних атак. Окрему увагу приділено проблемам дисбалансу класів, недостатньої відокремлюваності мережних подій у просторі ознак, ризику завищення показників точності та зниженню якості виявлення рідкісних підкласів атак. За результатами проведеного аналізу сформульовано постановку задачі дослідження та визначено основні напрями її розв’язання.

У другому розділі представлено модель процесу виявлення комп’ютерних атак на основі частотно-часового представлення мережного трафіку, що поєднує етапи попередньої обробки даних, кодування мережних ознак у хвильову форму, побудови спектральних образів і класифікації атак за допомогою двовимірних згорткових нейронних мереж. Розроблено метод соніфікації мережного трафіку, який дає змогу перетворювати табличний опис мережних з’єднань у сигнал, придатний для подальшого спектрального аналізу. Подано математичну модель класифікації, що формалізує процес прийняття рішення щодо належності мережного з’єднання до нормального або атакувального класу. Наведено опис основних етапів частотно-часового перетворення, побудови спектрограм та формування ознак для подальшого навчання моделі.

У третьому розділі представлено метод сигнатурозбережного адаптивного балансування навчальної вибірки, орієнтований на зменшення негативного впливу дисбалансу класів і покращення представлення проблемних підкласів атак у навчальних даних. Розроблено підхід, який поєднує збереження структурних властивостей міноритарного класу з урахуванням помилок базової моделі, що дає змогу формувати більш інформативну та збалансовану навчальну множину. Також представлено метод підвищення точності виявлення атак у зоні невизначеності прогнозу, який базується на аналізі помилок базового класифікатора, виділенні τ-зони невизначеності та селективному перевизначенні рішень за допомогою допоміжної моделі. Запропонований підхід орієнтовано на зменшення кількості хибнонегативних рішень і підвищення якості прийняття рішень у складних випадках прогнозування.

У четвертому розділі наведено організацію експериментального дослідження, описано використаний набір даних, етапи формування навчальної, валідаційної та тестової вибірок, а також особливості програмної реалізації запропонованих моделей і методів. Подано результати експериментального оцінювання базової соніфікованої моделі, методу сигнатурозбережного адаптивного балансування та методу підвищення точності в зоні невизначеності прогнозу. Виконано порівняння отриманих результатів з базовими підходами, наведено оцінювання точності, повноти, F1-міри та інших показників ефективності. На основі проведених експериментів підтверджено доцільність використання частотно-часового представлення мережного трафіку, а також ефективність запропонованих методів для покращення якості виявлення комп'ютерних атак.

У висновках наведено основні наукові та практичні результати дисертаційного дослідження, сформульовано підсумкові положення щодо досягнення поставленої мети та розв'язання науково-прикладної задачі підвищення точності й збалансованості виявлення комп'ютерних атак у корпоративних мережах.

У додатках подано публікації, що відображають наукові результати дисертації, акти впровадження, фрагменти програмної реалізації, таблиці з результатами експериментів та інші допоміжні матеріали, необхідні для повнішого висвітлення змісту і результатів проведеного дослідження.

Ключові слова: системи виявлення вторгнень, комп'ютерні атаки, корпоративні мережі, мережний трафік, соніфікація, частотно-часове представлення, спектрограма, двовимірної ЗНМ, дисбаланс класів, адаптивне балансування, зона невизначеності прогнозу.

ANNOTATION

Semeniuk B. V. Models and Methods for Detecting Computer Attacks in Corporate Networks Based on the Time-Frequency Representation of Network Traffic. – Qualifying scientific work as a manuscript.

Dissertation for the degree of Doctor of Philosophy in the field of knowledge 12 Information Technologies, specialty 122 Computer Science. – Khmelnytskyi National University, Khmelnytskyi, 2026.

Solving the scientific and applied problem of improving the accuracy and balance of computer attack detection in corporate networks based on the time-frequency representation of network traffic, adaptive balancing of the training dataset, and selective reclassification of decisions in the prediction uncertainty zone is a relevant area of research in the field of information security. The relevance of this problem is determined by the growing intensity of network traffic exchange, the increasing complexity of attack structures, the emergence of new and combined intrusion scenarios, as well as a significant imbalance between normal traffic and rare attack types, which reduces the effectiveness of traditional intrusion detection systems.

The dissertation analyzes modern intrusion detection systems and methods, approaches to representing network data for intelligent analysis tasks, and the use of machine learning and deep learning in intrusion detection systems. The feasibility of using network traffic sonification as a means of transforming the tabular description of network connections into a time-frequency representation suitable for further spectral analysis and automated attack recognition using two-dimensional convolutional neural networks is substantiated. A model of the computer attack detection process, a method for time-frequency transformation of network traffic, a method for signature-preserving adaptive balancing of the training dataset, and a method for improving attack detection accuracy in the prediction uncertainty zone have been developed. Experimental studies were also designed and conducted.

The object of research is the process of detecting computer attacks in corporate networks.

The subject of research is models and methods for detecting computer attacks in corporate networks based on the time-frequency representation of network traffic, sonification, adaptive balancing of the training dataset, and classification error analysis.

The purpose of the dissertation research is to improve the accuracy and balance of computer attack detection in corporate networks by developing models and methods for the time-frequency representation of network traffic, forming informative spectral features, adaptively balancing the training dataset, and selectively redefining decisions in the prediction uncertainty zone.

The scientific novelty of the obtained results is as follows:

1) a new method for the time-frequency transformation of network traffic has been developed which, unlike known vector-based approaches to feature representation, provides for the interpretation of the multidimensional space of connection parameters as signal parameters with linear frequency-amplitude mapping and subsequent spectral analysis, which made it possible to form a structured 2D representation of traffic and increase the informativeness of input data for attack classification tasks;

2) a new method for signature-preserving adaptive balancing of the training dataset has been developed which, unlike known oversampling and undersampling procedures, is characterized by preserving the structural characteristics of the minority class and taking into account the errors of the base model when forming an expanded training set, which made it possible to reduce the impact of class imbalance on the risk function, increase attack detection recall, and improve classification balance;

3) a new method for improving attack detection accuracy in the prediction uncertainty zone has been developed which, unlike known threshold-based decision-making schemes, provides for the formalization of the τ-uncertainty interval and selective reclassification of decisions based on an auxiliary model trained on the erroneous examples of the base classifier, which made it possible to reduce the number of false-negative decisions without violating the principle of independence of the test dataset;

4) for the first time, a model of the computer attack detection process based on network traffic sonification has been proposed which, unlike known IDS models, integrates the audio transformation of a multidimensional feature space, spectral analysis,

adaptive balancing of the training dataset, and a cascading decision-making mechanism within a generalized attack detection process model, which made it possible to coordinate the training and operation stages of the model and improve the balance of classification under class imbalance conditions.

Practical significance of the obtained results. A model of the computer attack detection process in corporate networks and methods for time-frequency transformation of network traffic, signature-preserving adaptive balancing of the training dataset, and selective accuracy improvement in the uncertainty zone have been developed. These make it possible to form an informative spectral representation of network connections, apply two-dimensional convolutional neural networks for automated attack detection, and reduce the negative impact of class imbalance on classification quality. A distinctive feature of the developed approach is the integration, within a single sonification pipeline, of the stages of network traffic transformation into waveform signals, spectrogram construction, base classifier training, adaptive balancing of the training dataset, and selective reclassification of decisions in the τ-uncertainty zone, which ensures the possibility of using the proposed solution in the creation and modernization of corporate network security monitoring systems.

As a result of the experimental studies, the operability of the sonification approach as an alternative channel for presenting network connections for the IDS task was confirmed, as well as the effectiveness of the cascade scheme combining SPAS-balancing, error-oriented amplification of problematic attack subclasses (ErrorBoost), definition of the τ-zone and selective redefinition of solutions. The application of the proposed approach ensured reproducibility of the recognition quality, and the value of F1-score on the test set increased by about 1 percent compared to the base sonified model.

The theoretical and practical results of the research are implemented in LLC "ITT" (Khmelnytskyi), PE "Avivi" (Khmelnytskyi), as well as in the educational process of the Khmelnytskyi National University when teaching disciplines at the Department of Computer Engineering and Information Systems for the specialty F7 Computer Engineering, in particular in the courses "Security and protection of computer systems", "Modeling and optimization methods in scientific and experimental research",

"Methodologies for ensuring the quality, reliability, warranty and security of computer systems and networks".

The introduction substantiates the relevance of the scientific and applied problem of improving the accuracy and balance of computer attack detection in corporate networks under conditions of increasing network traffic intensity, increasingly complex attack structures, the emergence of new combined intrusion scenarios, and a significant imbalance between normal traffic and rare attack types. The relationship between the dissertation topic and current trends in the development of intrusion detection systems, machine learning and deep learning methods, as well as intelligent network traffic analysis tools is identified. The purpose, object, subject, objectives, and research methods are formulated, and information on the scientific novelty, practical significance of the obtained results, their approbation, publications, and implementation is presented. The introduction also substantiates the feasibility of using the time-frequency representation of network traffic as a promising direction for increasing the informativeness of input data for intelligent attack detection.

The first chapter analyzes the subject area of the research, modern intrusion detection systems and methods, as well as approaches to network traffic classification in information security tasks. The features of signature-based, statistical, behavioral, and intelligent IDS are considered, the advantages and disadvantages of traditional vector representation of network features are analyzed, and the possibilities of using machine learning and deep learning for recognizing computer attacks are examined. Particular attention is paid to the problems of class imbalance, insufficient separability of network events in the feature space, the risk of overstated accuracy indicators, and reduced quality of detecting rare attack subclasses. Based on the results of the analysis, the research problem is formulated and the main directions for its solution are determined.

The second chapter presents a model of the computer attack detection process based on the time-frequency representation of network traffic, which combines the stages of data preprocessing, encoding network features into waveform signals, constructing spectral images, and classifying attacks using two-dimensional convolutional neural networks. A method for network traffic sonification is developed, making it possible to

transform the tabular description of network connections into a signal suitable for further spectral analysis. A mathematical classification model is presented which formalizes the process of deciding whether a network connection belongs to the normal or attack class. The main stages of time-frequency transformation, spectrogram construction, and feature formation for subsequent model training are described.

The third chapter presents a method for signature-preserving adaptive balancing of the training dataset aimed at reducing the negative impact of class imbalance and improving the representation of problematic attack subclasses in the training data. An approach is developed that combines preservation of the structural properties of the minority class with consideration of the errors of the base model, which makes it possible to form a more informative and balanced training set. The chapter also presents a method for improving attack detection accuracy in the prediction uncertainty zone, based on the analysis of errors of the base classifier, identification of the τ-uncertainty zone, and selective reclassification of decisions using an auxiliary model. The proposed approach is focused on reducing the number of false-negative decisions and improving decision quality in complex prediction cases.

The fourth chapter presents the organization of the experimental study, describes the dataset used, the stages of forming the training, validation, and test sets, as well as the features of software implementation of the proposed models and methods. The results of the experimental evaluation of the baseline sonified model, the signature-preserving adaptive balancing method, and the method for improving accuracy in the prediction uncertainty zone are presented. A comparison of the obtained results with baseline approaches is performed, and the evaluation of accuracy, recall, F1-score, and other performance indicators is provided. Based on the conducted experiments, the feasibility of using the time-frequency representation of network traffic, as well as the effectiveness of the proposed methods for improving the quality of computer attack detection, is confirmed.

The conclusions present the main scientific and practical results of the dissertation research and formulate the final provisions regarding the achievement of the stated goal

and the solution of the scientific and applied problem of improving the accuracy and balance of computer attack detection in corporate networks.

The appendices contain publications reflecting the scientific results of the dissertation, implementation certificates, fragments of software implementation, tables with experimental results, and other supplementary materials necessary for a more complete presentation of the content and results of the conducted research.

Keywords: intrusion detection systems, computer attacks, corporate networks, network traffic, sonification, time-frequency representation, spectrogram, 2D-CNN, class imbalance, adaptive balancing, prediction uncertainty zone.

СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Наукові праці, в яких опубліковані основні наукові результати дисертації

1. Семенюк Б., Каштальян А. Виявлення комп’ютерних атак із використанням 2D-CNN та соніфікації мережного трафіку. *Information Technology: Computer Science, Software Engineering and Cyber Security*. 2025. С. 193–201. DOI: https://doi.org/10.32782/IT/2025-1-26

2. Семенюк Б. В., Савенко Б. О. Метод виявлення комп’ютерних атак на основі адаптивності та балансування навчальної вибірки. *Вісник Черкаського державного технологічного університету*. 2026. № 1. С. 34-43. DOI: https://doi.org/10.62660/bcstu/1.2026.3

3. Семенюк Б. В., Савенко Б. О. Метод підвищення точності системи виявлення атак у зонах невизначеності на основі аналізу помилок та селективного перевизначення рішень. *Системні технології*. 2026. № 2. С. 99-110. DOI: https://doi.org/10.34185/1562-9945-2-163-2026-09

4. Семенюк Б., Корецька Л. Особливості проєктування та дослідження системи виявлення вторгнень на основі соніфікації мережевого трафіку. *Measuring and Computing Devices in Technological Processes*. 2026. № 1. С. 246–254. DOI: https://doi.org/10.31891/2219-9365-2026-85-31

5. Семенюк Б., Стецюк Ю. Дослідження впливу синтетичного балансування навчальної вибірки на точність виявлення комп'ютерних атак. *Measuring and Computing Devices in Technological Processes*. 2026. № 2. С. 410–417. DOI: https://doi.org/10.31891/2219-9365-2026-86-48

Праці, які засвідчують апробацію матеріалів дисертації

6. Semeniuk B., Kashtalian A., Martiniuk D., Drozd A., Abdel-Badeeh M. Salem. Detection of computer attacks based on sonification of network traffic. *Intelitsis'25: The 6th International Workshop on Intelligent Information Technologies & Systems of Information Security*, April 04, 2025, Khmelnytskyi, Ukraine. Pp. 252-263. URL: https://ceur-ws.org/Vol-3963/paper21.pdf *(Scopus)*

Публікації, які додатково відображають наукові результати дисертації

7. Свідоцтво про реєстрацію авторського права на твір (програму) № 145847 Україна. Комп'ютерна програма «Sonified IDS» / Семенюк Б.В., Нічепорук А.О., Савенко О.С. Дата реєстрації 24.04.2026. URL: https://sis.nipo.gov.ua/uk/search/detail/1925939/

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ

КА – комп’ютерна атака

КМ – корпоративна мережа

СВВ – система виявлення вторгнень

СВКА – система виявлення комп’ютерних атак

МН – машинне навчання

ГН – глибинне навчання

НМ – нейронна мережа

СВВ – система виявлення вторгнень

ШІ – штучний інтелект

ІР – Інтернет речей

ЗНМ – згорткова нейронна мережа

ІКМ – імпульсно-кодова модуляція

КПФ – короткочасне перетворення Фур’є

AI – Artificial Intelligence

AUC – Area Under the Curve

CNN – Convolutional Neural Network

DL – Deep Learning

FFT – Fast Fourier Transform

FNR – False Negative Rate

FPR – False Positive Rate

IDS – Intrusion Detection System

IoT – Internet of Things

ML – Machine Learning

NIDS – Network Intrusion Detection System

NSL-KDD – Network Security Laboratory Knowledge Discovery in Databases

PCM – Pulse Code Modulation

ROC – Receiver Operating Characteristic

SMOTE – Synthetic Minority Over-sampling Technique

SMOTENC – Synthetic Minority Over-sampling Technique for Nominal and Continuous Features

STFT – Short-Time Fourier Transform

TPR – True Positive Rate

ВСТУП

Обґрунтування вибору теми дослідження. Сучасний розвиток корпоративних інформаційних систем супроводжується стрімким зростанням обсягів мережного обміну, ускладненням архітектури цифрової інфраструктури, широким використанням хмарних сервісів, віртуалізованих середовищ, віддаленого доступу та розподілених прикладних платформ. Зазначені процеси істотно розширюють функціональні можливості корпоративних мереж (КМ), але одночасно створюють додаткові передумови для реалізації комп'ютерних атак (КА), прихованого несанкціонованого доступу, розгортання багатокрокових зловмисних сценаріїв і порушення безперервності функціонування інформаційних систем.

У сучасному цифровому середовищі проблема виявлення КА набуває особливої актуальності, оскільки своєчасне виявлення зловмисної активності є однією з базових умов забезпечення кібербезпеки корпоративної інфраструктури. В умовах безперервного мережного обміну, великої кількості сесій, високої швидкості передавання даних і різнорідності джерел трафіку зловмисні дії можуть маскуватися під нормальну мережну активність, що ускладнює їх ідентифікацію традиційними засобами захисту. Це зумовлює необхідність створення інтелектуальних методів автоматизованого аналізу мережного трафіку, здатних виявляти не лише явні, а й приховані або слабко виражені ознаки КА.

Задача підвищення ефективності виявлення КА відповідає сучасним потребам розвитку інформаційної безпеки державних, виробничих і корпоративних інформаційних систем, а також загальній тенденції до впровадження інтелектуальних інформаційних технологій у системи моніторингу та захисту мережної інфраструктури. Надійне виявлення КА сприяє зменшенню ризику компрометації інформаційних ресурсів, запобіганню порушенням безперервності бізнес-процесів, зниженню ймовірності витоку конфіденційних даних і підвищенню загального рівня довіри до цифрових сервісів. У цьому контексті особливої актуальності набувають підходи, орієнтовані на

автоматизоване виявлення вторгнень засобами штучного інтелекту, машинного та глибинного навчання.

Сьогодні ефективними засобами автоматизованого виявлення аномальної та зловмисної мережної активності є системи виявлення вторгнень, що використовують статистичний аналіз, методи розпізнавання образів, машинне навчання та нейромережеву класифікацію. Проте, незважаючи на активний розвиток СВВ, універсального підходу, який би забезпечував високу точність виявлення КА за умов складної структури мережного трафіку, дисбалансу класів, слабкої відокремлюваності зловмисних і нормальних зразків та невизначеності прогнозів у прикордонних випадках класифікації, наразі не існує. Особливої складності набуває виявлення рідкісних або структурно неоднорідних підкласів атак, які в реальних даних представлені недостатньо та часто втрачаються на фоні домінантного нормального трафіку.

Виявлення КА належить до задач інтелектуального аналізу даних, які активно досліджуються та представлені у вітчизняній та зарубіжній науковій літературі у сферах комп'ютерних наук та кібербезпеки. Розвитком методів виявлення zero-day та аномальних атак, застосуванням машинного і глибокого навчання в системах виявлення вторгнень, аналізом мережного трафіку, дослідженням проблем дисбалансу класів, генеративного розширення вибірки та підвищення узагальнювальної здатності моделей займаються Mbona I. [1], Aldweesh A. [2], Asharf J. [4], Moustafa N. [4], Abdelmoumin G. [10], Whitaker J. [10], Rawat D.B. [10], Ahmad Z. [20], Kocher G. [24], Elmasry W. [52], Tang C. [54], Hwang R.-H. [102], Jamoos M. [107], Mari A.-G. [114]. Питанням класифікації мережного трафіку, побудови інформативних вхідних подань і аналізу часових закономірностей мережної активності займаються Sheikh M.S. [7], Abbasi M. [88], Shahraki A. [88], Hardegen C. [93], Malik A. [87].

Значну увагу сигнатурним і аномальним СВВ, методам статистичного аналізу мережних з'єднань, застосуванню класичних класифікаторів, ансамблевих моделей, згорткових і рекурентних нейронних мереж, а також процедурам балансування навчальних вибірок приділено в роботах Abdelmoumin G. [10], Whitaker J. [10], Rawat D.B. [10], Rahman A. [10], Ahmad Z. [20], Kocher G. [24],

Jamoos M. [107]. Однак більшість відомих підходів орієнтована на пряме векторне подання мережних ознак, у межах якого міжознакові взаємозв'язки та приховані структурні закономірності часто втрачаються або виявляються недостатньо виразно для ефективного використання засобів глибинного навчання.

Аналіз сучасних методів і засобів виявлення комп'ютерних атак показав їх значний потенціал для розв'язання задач кіберзахисту та можливість використання в різних контекстах корпоративної мережної інфраструктури. Разом із тим ці підходи не завжди забезпечують достатньо інформативне представлення мережного трафіку для моделей глибинного навчання, не повною мірою враховують структурні особливості міноритарних класів атак і, як правило, не містять спеціалізованих механізмів корекції рішень у зоні невизначеності прогнозу. Використання лише векторного опису мережних з'єднань обмежує можливості виявлення прихованих просторово-частотних закономірностей, а класичні процедури надсемплінгу не завжди зберігають характерні риси рідкісних атак, що негативно впливає на якість класифікації.

Особливої уваги потребує проблема дисбалансу класів у наборах даних для СВВ. У більшості практичних випадків обсяг нормального трафіку суттєво перевищує обсяг атакуючого, а окремі типи або підкласи атак представлені незначною кількістю прикладів. Це призводить до зміщення процесу навчання в бік домінантних класів, формування завищених оцінок загальної точності та погіршення здатності моделі виявляти рідкісні, але критично важливі атаки. Крім того, навіть за наявності достатньо потужного базового класифікатора зберігається група прикладів, для яких вихідні оцінки моделі перебувають у зоні невизначеності, що створює передумови для хибнонегативних рішень і потребує розроблення спеціальних каскадних механізмів перевизначення прогнозу.

Перспективним напрямом розв'язання зазначених проблем є використання частотно-часового представлення мережного трафіку, сформованого на основі його соніфікації. Такий підхід забезпечує перехід від традиційного табличного опису мережних ознак до структурованого сигналоподібного подання, придатного для подальшого спектрального аналізу. Перетворення багатовимірного простору параметрів мережного з'єднання у хвильовий сигнал із наступним формуванням

двовимірного подання створює передумови для застосування моделей аналізу зображень, зокрема двовимірних згорткових нейронних мереж, до задачі виявлення атак. У такому разі класифікація базується не лише на окремих числових ознаках, а й на виявленні просторово-частотних структур, що підвищує інформативність вхідних даних.

Разом із тим використання частотно-часового представлення саме по собі не усуває проблему недостатньої представленості міноритарних класів та не гарантує коректної обробки складних або сумнівних випадків класифікації. Тому підвищення точності СВВ потребує поєднання кількох взаємопов'язаних компонентів: інформативного перетворення мережного трафіку; спеціалізованого адаптивного балансування навчальної вибірки, орієнтованого на збереження сигнатурних характеристик рідкісних атак; механізму аналізу помилок базової моделі та селективного перевизначення рішень у τ-зоні невизначеності прогнозу.

Таким чином, наразі існує суперечність між необхідністю високоточного виявлення КА у корпоративних мережах (КМ) та недосконалістю існуючих методів і засобів, які не забезпечують достатньо інформативного представлення мережного трафіку, не враховують повною мірою структурні особливості рідкісних КА і не містять цілісного механізму корекції рішень у складних випадках класифікації. Це особливо важливо для систем, що функціонують в умовах значного дисбалансу класів і високої ціни помилки пропуску атаки. Відтак розроблення моделей і методів виявлення комп'ютерних атак у корпоративних мережах на основі частотно-часового представлення мережного трафіку, які забезпечують інформативне подання даних, адаптивне формування навчальної множини та підвищення точності прийняття рішень у зоні невизначеності, є актуальною науково-прикладною задачею.

Зазначена науково-прикладна задача відповідає предметній області стандарту вищої освіти України зі спеціальності 122 Комп'ютерні науки для третього освітньо-наукового рівня, зокрема такому теоретичному змісту предметної області, як сучасні моделі, методи, …, процеси та способи отримання, представлення, обробки, аналізу, передачі, … даних в інформаційних системах.

Зв’язок роботи з науковими програмами, планами, темами. Зв’язок роботи з науковими програмами, планами, темами. Дисертаційне дослідження виконувалось у рамках науково-дослідної тематики Хмельницького національного університету: держбюджетної науково-дослідної теми «Система забезпечення стійкості до витоку конфіденційної інформації в корпоративних мережах в умовах впливів комп’ютерних атак» (номер держреєстрації 0126U002082), в якій автор дисертації був виконавцем.

Мета і задачі дослідження.

Об’єкт дослідження – процес виявлення комп’ютерних атак у корпоративних мережах.

Предмет дослідження – моделі та методи виявлення комп’ютерних атак у корпоративних мережах на основі частотно-часового представлення мережного трафіку, соніфікації, адаптивного балансування навчальної вибірки та аналізу помилок класифікації.

Метою дисертаційного дослідження є підвищення точності та збалансованості виявлення комп’ютерних атак у корпоративних мережах за рахунок розроблення моделей і методів частотно-часового представлення мережного трафіку, формування інформативного двовимірного подання даних, сигнатурозбережного адаптивного балансування навчальної вибірки та селективного перевизначення рішень у зоні невизначеності прогнозу.

Для досягнення поставленої мети необхідно вирішити такі задачі:

1. Провести аналіз методів, засобів і технологій виявлення комп’ютерних атак у корпоративних мережах, а також дослідити обмеження існуючих підходів до подання мережного трафіку та класифікації атак в умовах дисбалансу класів.

2. Розробити модель процесу виявлення комп’ютерних атак, яка поєднує етапи попередньої обробки мережного трафіку, його соніфікації, спектрального аналізу, базової класифікації, адаптивного балансування та каскадного прийняття рішень.

3. Розробити метод частотно-часового перетворення мережного трафіку, придатний для побудови структурованого двовимірного подання даних і

застосування нейромережевих моделей аналізу зображень до задачі бінарної класифікації атак.

4. Розробити метод сигнатурозбережного адаптивного балансування навчальної вибірки, який забезпечує зменшення впливу дисбалансу класів і підвищення репрезентативності міноритарних підкласів атак.

5. Розробити метод підвищення точності виявлення атак у зоні невизначеності прогнозу на основі аналізу помилок базового класифікатора та селективного перевизначення рішень за допомогою допоміжної моделі.

6. Розробити програмно-алгоритмічне забезпечення для валідації запропонованих моделей і методів та провести експериментальні дослідження їх ефективності.

Методи дослідження. Для розв’язання поставлених задач використовуються методи математичної статистики і теорії ймовірностей, методи машинного та глибинного навчання, методи аналізу мережного трафіку, методи цифрової обробки сигналів, методи спектрального та частотно-часового аналізу, методи формування та балансування навчальних вибірок, а також методи емпіричного дослідження і порівняльного аналізу результатів експериментів.

Наукова новизна отриманих результатів полягає в наступному:

1) вперше розроблено модель процесу виявлення комп’ютерних атак на основі соніфікації мережного трафіку, особливістю якої є інтеграція звукового перетворення багатовимірного простору ознак, спектрального аналізу, адаптивного балансування навчальної вибірки та каскадного механізму прийняття рішень у межах узагальненої моделі процесу виявлення атак, що дає змогу узгодити етапи навчання і функціонування моделі та підвищити збалансованість класифікації в умовах дисбалансу класів;

2) розроблено новий метод частотно-часового перетворення мережного трафіку, який, на відміну від відомих векторних підходів до представлення ознак, передбачає інтерпретацію багатовимірного простору параметрів з’єднання як параметрів сигналу з лінійним частотно-амплітудним відображенням і подальшим спектральним аналізом, що дає змогу формувати структуроване 2D-подання трафіку та підвищити інформативність вхідних даних для задачі класифікації атак;

3) розроблено новий метод сигнатурозбережного адаптивного балансування навчальної вибірки, який, на відміну від відомих процедур надсемплінгу та підвибірки, характеризується збереженням структурних характеристик міноритарного класу та врахуванням помилок базової моделі під час формування розширеної навчальної множини, що дає змогу зменшити вплив дисбалансу класів на функцію ризику, підвищити повноту виявлення атак і збалансованість класифікації;

4) розроблено новий метод підвищення точності виявлення атак у зоні невизначеності прогнозу, який, на відміну від відомих порогових схем прийняття рішень, передбачає формалізацію τ-інтервалу невизначеності та селективне перевизначення рішень на основі допоміжної моделі, навченої на помилкових прикладах базового класифікатора, що дає змогу зменшити кількість хибнонегативних рішень без порушення принципу незалежності тестової вибірки.

Практичне значення отриманих результатів. Розроблено модель процесу виявлення комп'ютерних атак у корпоративних мережах та методи частотно-часового перетворення мережного трафіку, сигнатурозбережного адаптивного балансування навчальної вибірки і селективного підвищення точності в зоні невизначеності прогнозу, які дають змогу формувати інформативне спектральне подання мережних з'єднань, застосовувати двовимірні згорткові нейронні мережі для автоматизованого виявлення атак та зменшувати негативний вплив дисбалансу класів на якість класифікації. Особливістю розробленого підходу є поєднання в єдиному соніфікаційному конвеєрі етапів перетворення мережного трафіку у хвильову форму, побудови спектрограм, навчання базового класифікатора, адаптивного балансування навчальної вибірки та селективного перевизначення рішень у τ-зоні невизначеності, що забезпечує можливість використання запропонованого рішення під час створення і модернізації систем моніторингу безпеки корпоративних мереж.

У результаті проведених експериментальних досліджень підтверджено працездатність соніфікаційного підходу як альтернативного каналу подання мережних з'єднань для задачі виявлення вторгнень, а також ефективність каскадної схеми, що поєднує балансування SPAS, помилково-орієнтоване підсилення

проблемних підкласів атак, визначення τ-зони та селективне перевизначення рішень. Застосування запропонованого підходу забезпечило відтворюване покращення якості розпізнавання, а значення F1-міри на тестовій множині зросло приблизно на 1 відсоток порівняно з базовою соніфікованою моделлю.

Теоретичні та практичні результати дослідження впроваджені в ТОВ «ІТТ» (акт від 30.06.2026, м. Хмельницький), ПП «Авіві» (довідка від 29.06.2026, м. Хмельницький), а також в освітньому процесі Хмельницького національного університету (акт від 26.06.2026) при викладанні дисциплін на кафедрі комп'ютерної інженерії та інформаційних систем для здобувачів спеціальності F7 Комп'ютерна інженерія, зокрема в курсах «Безпека та захист комп'ютерних систем», «Моделювання та методи оптимізації в наукових та експериментальних дослідженнях», «Методології забезпечення якості, надійності, гарантоздатності та безпеки комп'ютерних систем та мереж».

Особистий внесок здобувача. У працях, опублікованих у співавторстві, автору належать такі наукові результати: [127, 132] – розроблено модель та метод частотно-часового перетворення мережного трафіку на основі соніфікації та формування двовимірного спектрального подання для виявлення КА із застосуванням двовимірної ЗНМ; [128] – розроблено метод сигнатурозбережного адаптивного балансування навчальної вибірки; [129] – розроблено метод підвищення точності виявлення КА у зоні невизначеності прогнозу; [130] – розроблено модель процесу виявлення КА на основі частотно-часового представлення мережного трафіку; [131] – досліджено вплив синтетичного балансування навчальної вибірки на точність виявлення КА; [133] – розроблено архітектуру програмного забезпечення та програмний код.

У роботах, опублікованих у співавторстві, співавторам належать такі результати: [127, 132] – розроблення методу виявлення КА, обґрунтування доцільності соніфікації мережного трафіку, уточнення постановки експериментального дослідження, апробація результатів (Каштальян А., Мартинюк Д., Дрозд А., Abdel-Badeeh M. Salem); [128, 129] – формалізація адаптивного балансування та каскадної схеми селективного уточнення рішень (Савенко Б.); [130] –визначення особливостей проєктування системи виявлення

вторгнень і методики її дослідження (Корецька Л.); [131] – організація експериментального дослідження щодо впливу синтетичного балансування на точність виявлення КА (Стецюк Ю.); [133] – розроблено вимоги до програмного забезпечення та здійснено його верифікацію (Савенко О.), розроблено алгоритми (Нічепорук А.).

Апробація результатів дисертації. Апробацію основних положень, ідей, висновків дисертаційної роботи проведено на науковому семінарі кафедри комп'ютерної інженерії та інформаційних систем у Хмельницькому національному університеті. Наукові результати роботи доповідалися на таких конференціях: 6th International Workshop on Intelligent Information Technologies & Systems of Information Security (IntelITSIS), Khmelnytskyi, Ukraine, April 04, 2025; 15th International Conference on Dependable Systems, Services and Technologies (DeSSerT-2025; https://www.dessert-conf.org/dessert-2025/), Athens, Greece, December 19-21, 2025; 16th International Conference on Advanced Computer Information Technologies (ACIT-2026; https://www.acit.tech/), Zlin, Czech Republic, September 2–4, 2026.

Публікації. За результатами проведених досліджень основні наукові результати опубліковано у 5 наукових статтях [127–131]. Апробацію результатів дисертації засвідчено публікацією однієї праці у матеріалах міжнародної конференції [132], проіндексованої в наукометричній базі Scopus, та доповідями (наявні відомості в програмах конференцій) на двох міжнародних конференціях. Отримано одне свідоцтво про реєстрацію авторського права на твір (комп'ютерна програма) [133].

Структура та обсяг дисертації. Дисертаційна робота складається з анотації, змісту, переліку умовних скорочень, вступу, чотирьох розділів, висновків, списку використаних джерел та п'яти додатків. Повний обсяг дисертації становить 222 сторінки, з них анотація – 11 с., зміст – 2 с., перелік скорочень – 2 с., основний текст – 150 с., список використаних джерел із 133 найменувань – 20 с., додатки – 66 с. Дисертація містить 41 рисунок і 20 таблиць.

РОЗДІЛ 1
АНАЛІЗ СУЧАСНИХ СИСТЕМ ТА МЕТОДІВ ВИЯВЛЕННЯ ВТОРГНЕНЬ ТА ЗАСТОСУВАННЯ ЗАСОБІВ ШТУЧНОГО ІНТЕЛЕКТУ

1.1 Актуальність задачі виявлення комп’ютерних атак у корпоративних мережах

Сучасний розвиток інформаційно-комунікаційних технологій супроводжується ускладненням архітектури КМ, зростанням інтенсивності обміну даними та розширенням взаємодії між внутрішніми й зовнішніми цифровими сервісами. До складу таких середовищ входять хмарні платформи, мобільні клієнти, розподілені прикладні сервіси, віртуалізовані ресурси, елементи Інтернету речей та інші компоненти, що формують багаторівневу динамічну інфраструктуру. За цих умов збільшується кількість потенційних векторів атаки, а задача своєчасного виявлення порушень безпеки залишається однією з центральних у кібербезпеці [3, 4].

Особливої актуальності ця задача набуває через те, що традиційні засоби захисту, які ґрунтуються переважно на периметровій фільтрації, сегментації, контролі доступу та сигнатурних правилах, не забезпечують повного виявлення складних і нових загроз. У реальних корпоративних мережах дедалі більшого значення набувають модифіковані атаки, багатокрокові сценарії компрометації, прихована зловмисна активність у межах уже встановлених з’єднань, а також zero-day загрози, для яких відсутні готові сигнатурні описи [1, 2]. За таких умов виникає потреба не лише у блокуванні відомих шаблонів атак, а й у виявленні аномальних станів мережного середовища на основі аналізу його поведінки.

У системі кіберзахисту КМ таку функцію виконують системи виявлення вторгнень. СВВ доцільно розглядати як спеціалізований аналітичний компонент, призначений для безперервного спостереження за мережним або вузловим середовищем, аналізу подій безпеки та виявлення ознак потенційно небезпечної активності [124, 125]. На відміну від засобів, що реалізують переважно жорстке

обмеження доступу за наперед визначеними правилами, СВВ виконує функцію раннього розпізнавання інцидентів, забезпечує формування інформаційної основи для їх інтерпретації та підтримує подальше реагування. Саме тому СВВ посідає важливе місце між механізмами запобігання та механізмами реагування в загальному контурі кіберзахисту.

Функціонально СВВ здійснює збір і попередню обробку телеметрії, виділення інформативних характеристик, аналіз мережного трафіку або системних подій, формування рішення про наявність або відсутність загрози та передавання результатів до суміжних систем контролю і реагування. У сучасних корпоративних інфраструктурах ця роль розширюється за рахунок інтеграції з платформами централізованого моніторингу, потокової аналітики та сервісами обробки великих масивів даних [18]. Подібний підхід реалізується не лише у класичних мережних середовищах, а й у спеціалізованих кіберфізичних та промислових системах, де своєчасне виявлення вторгнень є важливою умовою безпечного функціонування технологічних процесів [106]. У програмно-керованих та високодинамічних мережах СВВ також може виступати джерелом даних для адаптивного керування безпекою, переналаштування політик доступу та ізоляції сегментів мережі [117].

Практична значущість СВВ у корпоративному середовищі визначається і характером наслідків помилок виявлення. Пропуск КА, тобто помилка типу хибнонегативна, може спричинити подальше поширення загрози мережею, компрометацію сервісів, порушення конфіденційності даних або доступності критичних бізнес-процесів. Водночас надмірна кількість хибнопозитивних спрацьовувань знижує практичну цінність системи, перевантажує операторів безпеки та ускладнює виділення дійсно критичних подій. Тому підвищення точності СВВ повинно розглядатися не ізольовано, а разом зі зменшенням помилково позитивних і помилково негативних, що є однією з ключових умов ефективного використання систем виявлення вторгнень [20, 21].

Однією з базових ознак класифікації СВВ є принцип формування рішення щодо належності спостережуваної активності до нормального або небезпечного стану. Найпоширенішими є сигнатурний, аномальний і гібридний підходи.

Сигнатурний підхід ґрунтується на порівнянні поточної активності з наперед визначеними правилами та шаблонами відомих атак. Його перевагою є висока точність щодо добре формалізованих загроз, однак основним недоліком залишається обмежена здатність до виявлення нових і модифікованих атак [55]. Аномальний підхід, навпаки, базується на моделюванні нормального стану системи та виявленні відхилень від нього, що робить його перспективним для розпізнавання загроз нульового дня і складних поведінкових відхилень [44, 45]. Разом із тим його ефективність істотно залежить від якості побудови моделі норми та здатності системи відокремлювати реальні аномалії від нестандартної, але допустимої активності. Гібридний підхід поєднує елементи сигнатурного й аномального аналізу або інтегрує декілька різнорідних моделей у межах єдиного механізму прийняття рішення, що дозволяє частково компенсувати недоліки окремих класів систем [12, 13]. У сучасних дослідженнях межа між аномальними та гібридними СВВ поступово стирається, оскільки дедалі частіше використовуються ансамблеві, генеративні та багатокомпонентні моделі, результативність яких визначається не лише алгоритмом, а й способом подання вхідних даних [76, 77, 102].

Важливим чинником ефективності СВВ є також рівень представлення даних, на основі яких формується рішення. У задачах виявлення атак найчастіше використовуються пакетний, потоковий, сесійний і журнальний рівні представлення [7]. Пакетний рівень забезпечує найвищу деталізацію і є інформативним для виявлення атак, що проявляються на рівні окремих повідомлень, однак супроводжується значним обсягом даних і високими вимогами до обробки [83, 84]. Потоковий рівень спирається на агреговані характеристики мережної взаємодії, краще масштабується і широко застосовується як компроміс між інформативністю та практичною придатністю [82], хоча частина локальних особливостей при цьому втрачається [86]. Сесійне представлення дозволяє враховувати часову логіку розвитку взаємодії, а журнальне - аналізувати внутрішні події на рівні систем і сервісів, для чого дедалі частіше використовуються моделі аналізу послідовностей і контекстних залежностей [30]. Отже, вибір рівня

представлення даних визначається типом атак, доступністю телеметрії, архітектурою моделі та вимогами до швидкодії й масштабованості СВВ [88].

Поширення методів машинного та глибокого навчання суттєво розширило можливості сучасних СВВ. Такі підходи дозволяють автоматизовано виявляти складні залежності в мережних даних, враховувати багатовимірні взаємозв'язки між параметрами трафіку та підвищувати повноту виявлення аномальної активності. У сучасних оглядах підкреслюється, що інтеграція інтелектуальних методів стала одним із провідних напрямів розвитку СВВ [10]. Однак саме використання машинного або глибокого навчання не усуває всіх наявних обмежень. Результативність таких систем істотно залежить від якості та структури навчальних даних, способу представлення мережного трафіку, наявності дисбалансу класів, коректності розділення навчальних і тестових вибірок та здатності моделі до узагальнення.

Особливо важливою є залежність результатів СВВ від контрольного набору даних. Доступні набори даних відрізняються за складом ознак, співвідношенням класів, способом формування навчальної та тестової частин і ступенем відповідності сучасним умовам функціонування мереж. Тому одна й та сама модель може демонструвати різну ефективність на різних наборах даних, що надає питанням коректності експериментального оцінювання самостійного методологічного значення [56]. Не менш суттєвою є проблема дисбалансу класів: у реальному мережному середовищі нормальна активність кількісно домінує над аномальною, тоді як окремі небезпечні типи атак представлені незначною кількістю прикладів. За таких умов модель може формально демонструвати високі інтегральні показники, але виявляти недостатню чутливість до рідкісних і складних загроз [24].

Окремої уваги потребує і спосіб подання мережного трафіку. Більшість відомих СВВ працює з табличними ознаками або агрегованими характеристиками потоку, що є зручним для формалізації та побудови класифікаторів. Разом із тим така форма не завжди повно відображає часову структуру процесу, внутрішні взаємозв'язки між параметрами та складні шаблони зміни мережної активності. У

сучасних дослідженнях підкреслюється, що ефективність виявлення визначається не лише вибором моделі, а й тим, у якій формі мережні дані подаються на її вхід [100, 101]. Це зумовлює актуальність пошуку альтернативних способів подання трафіку, здатних краще відображати структурні, часові та частотні особливості аномальної поведінки.

Таким чином, проведений аналіз показує, що СВВ є однією з базових підсистем сучасного кіберзахисту КМ, а їх ефективність визначається не лише принципом формування рішення, а й характером вхідних даних, способом їх подання, якістю навчальної вибірки та здатністю моделі працювати в умовах мінливого мережного середовища. Це обґрунтовує доцільність подальшого розгляду інтелектуальних методів виявлення КА як напряму розвитку СВВ, орієнтованого на підвищення точності, здатності та спроможності виявляти як відомі, так і нові типи загроз.

1.2 Аналіз інтелектуальних методів виявлення комп’ютерних атак

Розвиток систем виявлення вторгнень упродовж останніх років значною мірою пов’язаний із переходом від жорстко заданих сигнатурних правил до інтелектуальних методів аналізу мережної активності. Якщо у класичних системах рішення про наявність загрози формувалося переважно на основі наперед визначених шаблонів, то методи машинного навчання дали змогу будувати правило класифікації безпосередньо з даних, виявляючи статистичні залежності між ознаками мережного трафіку та станами безпеки [19]. Саме цей перехід означав зміну самої логіки побудови СВВ: центр ваги змістився від експертного ручного формування правил до аналізу емпіричних закономірностей, наявних у мережних даних. У результаті система виявлення вторгнень почала розглядатися не лише як механізм перевірки відповідності відомим сигнатурам, а як адаптивна аналітична підсистема, здатна навчатися на прикладах нормальної та аномальної поведінки.

Найбільш поширеними в задачі СВВ стали дерева рішень, випадкові ліси, метод опорних векторів, k-найближчих сусідів, баєсівські класифікатори й

ансамблеві схеми, що працюють переважно з табличним поданням мережних ознак [5, 6]. Їх практична цінність пояснюється помірною обчислювальною складністю, відносною інтерпретованістю та придатністю до роботи в умовах обмежених ресурсів, а також можливістю підвищення якості виявлення за рахунок оптимізації ознакового простору й налаштування параметрів моделі [25]. Для багатьох практичних сценаріїв саме такі властивості є принципово важливими, оскільки СВВ повинна не лише забезпечувати прийнятну точність, а й функціонувати з допустимими витратами часу та ресурсів. Класичні моделі машинного навчання виявилися придатними для тих випадків, коли мережний трафік уже перетворено на структурований вектор ознак і коли основна задача полягає у відокремленні добре виражених статистичних шаблонів нормальної й аномальної поведінки. У багатьох дослідженнях саме належна підготовка вхідних даних, а не радикальне ускладнення алгоритму, забезпечувала відчутний приріст якості класифікації. Це стосується і процедур нормалізації, і кодування категоріальних змінних, і відбору найбільш інформативних характеристик, без яких навіть достатньо якісний класифікатор може втрачати здатність до коректного розмежування між класами.

Для еталонних наборів на кшталт NSL-KDD особливо важливими виявилися процедури відбору релевантних характеристик і зменшення розмірності, оскільки саме вони дають змогу зменшити вплив шумових та дубльованих ознак на підсумкове рішення [126]. Така обставина підтверджує, що в задачі СВВ якість класифікації визначається не лише вибором алгоритму, а й тим, наскільки змістовно сформовано сам простір вхідних даних. Якщо ознаки дублюють одна одну, містять випадкові коливання або лише опосередковано пов’язані з атакувальною активністю, то модель починає спиратися на випадкові статистичні залежності. Навпаки, компактне й інформативне подання підвищує шанси на формування більш узагальнюваних правил класифікації. Саме тому в рамках класичного машинного навчання важливе місце посіли підходи, у яких відбір ознак, редукція розмірності, кластеризація та оптимізація гіперпараметрів розглядаються як складові єдиного аналітичного конвеєра.

Разом із тим навіть за високих інтегральних показників точності класичні моделі машинного навчання нерідко виявляють обмежену чутливість до складних і рідкісних підкласів КА, а їх результативність істотно залежить від способу попередньої обробки, поділу вибірки та налаштування гіперпараметрів [91], [95]. Саме ця залежність від сформованого вектора ознак і стала однією з причин переходу до глибших моделей аналізу, здатних автоматично будувати внутрішнє представлення мережної поведінки [99]. Іншими словами, класичне машинне навчання показало, що аналітичне виявлення атак на основі даних є принципово можливим, але водночас продемонструвало межі підходів, які працюють лише з наперед заданими ознаками. Якщо кроки загрози проявляються у складних нелінійних взаємозалежностях, часових конфігураціях або багаторівневих контекстах, то модель, яка аналізує лише статичний вектор характеристик, виявляється недостатньо чутливою.

Подальший етап розвитку СВВ пов’язаний із широким упровадженням методів глибокого навчання, орієнтованих на роботу зі складними нелінійними залежностями в мережних даних [22, 23]. На відміну від класичних алгоритмів, глибокі моделі дозволяють формувати внутрішнє багаторівневе представлення даних, у межах якого прості первинні ознаки перетворюються на складніші й більш інформативні абстракції. У контексті аналізу трафіку найбільшого поширення набули згорткові, рекурентні, автоенкодерні та комбіновані архітектури, які дозволяють працювати відповідно з двовимірними структурами, часовими послідовностями, стисненими латентними поданнями та багаторівневими схемами виявлення [57]. Саме поява таких підходів змінила і уявлення про те, яким може бути вхідне подання для СВВ. Якщо раніше мережний трафік переважно розглядався як набір агрегованих параметрів з’єднання чи потоку, то тепер він дедалі частіше подається як часовий ряд, матриця, спектральна карта або інша структура, що краще узгоджується з архітектурою моделі.

Перевага глибокого навчання (DL) полягає в суттєвому зменшенні залежності від ручного конструювання ознак і в можливості автоматизовано виділяти приховані просторові, часові та контекстні закономірності мережної

активності [52]. Це особливо важливо для задач, у яких атака проявляється не окремим статистичним відхиленням, а ритмікою трафіку, послідовністю подій або складною комбінацією параметрів [54]. Для таких сценаріїв простого табличного опису часто недостатньо, тоді як глибокі моделі здатні виявляти локальні шаблони, повторювані структури та довгі залежності в даних. Наприклад, згорткові мережі є природними для аналізу двовимірних подань трафіку, де інформативне значення мають локальні конфігурації ознак. Рекурентні архітектури краще пристосовані до послідовнісних сценаріїв, коли важливим є порядок подій, тривалість залежностей та зміна режиму в часі. Автоенкодери ж дозволяють не лише стискати ознаковий простір, а й виявляти відхилення на основі помилки реконструкції, що робить їх придатними для аномалійного аналізу за неповної інформації про всі типи атак.

Водночас ефективність глибоких СВВ істотно зростає лише тоді, коли архітектура узгоджується з процедурами балансування вибірки, організацією функції втрат і характером вхідного представлення [27, 28]. Іншими словами, складність моделі сама по собі не гарантує високої якості. Якщо дані подано у формі, що не відображає змістовних закономірностей процесу, або якщо процедура навчання не враховує сильний перекіс між класами, то навіть архітектурно потужна мережа може демонструвати локально високі, але результати, залежні від локальних властивостей вибірки. Для аналізу багатовимірних часових рядів принципового значення набуває здатність враховувати довгі залежності, зміну режимів функціонування та взаємозв’язки між каналами спостереження, що пояснює поширення довга короткочасна пам'ять (LSTM), замкнута рекурентна одиниця (GRU) та споріднених архітектур у задачі виявлення аномального трафіку [29, 109]. Такі моделі роблять акцент не на статичному зрізі ознак, а на процесуальній природі мережної активності, що є особливо важливим для виявлення поступово сформованих аномалій або атак із вираженою часовою структурою.

Окремий напрям становлять підходи на основі перенесення навчання, які дають змогу зменшити вимоги до обсягу розмічених даних і прискорити адаптацію моделі до нового домену [58, 59]. Для СВВ це має безпосереднє значення, оскільки

в багатьох практичних середовищах кількість високоякісно розмічених прикладів атак є обмеженою, а повторне навчання глибокої моделі з нуля виявляється дорогим або малоефективним. Використання вже сформованих внутрішніх представлень створює можливість скорочувати витрати на адаптацію системи до нових сценаріїв, нових типів трафіку або спеціалізованих мережних середовищ. Проте підвищення складності архітектури не усуває автоматично проблем узагальнення, робастності та обчислювальної вартості, а тому питання узагальнення моделей глибокого навчання у нових умовах функціонування залишається відкритим [63, 64, 40]. Це означає, що навіть на етапі глибокого навчання проблема правильної побудови вхідного подання і коректного експерименту не зникає, а навпаки стає ще більш значущою.

Окремий напрям підвищення якості СВВ сформували генеративні та комбіновані підходи, у яких результат досягається не лише вибором класифікатора, а поєднанням декількох процедур – балансування, генерації прикладів, ансамблювання та багатоступеневої обробки вхідних даних. Поширення генеративно-змагальних мереж відкрило можливість синтезувати додаткові приклади для рідкісних класів атак і частково компенсувати перекіс вибірки [8, 9]. У роботах, присвячених застосуванню генеративно-змагальний мереж (GAN) у задачах виявлення КА, показано, що правильно організований генеративний механізм здатний покращувати не лише загальну точність, а й повноту виявлення слабко представлених підкласів атак [107, 114]. Теоретична привабливість такого підходу полягає в тому, що модель не просто повторює наявний дисбаланс, а цілеспрямовано посилює представленість тих класів, які інакше губляться на фоні домінуючих шаблонів мережної поведінки.

Разом із тим генеративні моделі не можна розглядати як універсальне вирішення проблеми, оскільки їх ефективність залежить від якості вихідних даних, стабільності навчання та того, наскільки синтетичні зразки зберігають суттєві властивості реального трафіку [11, 14]. Якщо генерація формує лише зовнішньо подібні, але змістовно спрощені приклади, модель може покращувати результати на контрольному наборі, однак не набувати реальної чутливості до складних типів

атак. Додатково виникає питання інтерпретованості приросту точності: стає неочевидним, чи покращення зумовлено реально кращим виявленням, чи лише адаптацією до штучно спрощеного ознакового простору. Паралельно розвиваються комбіновані та ансамблеві схеми, де приріст точності досягається завдяки узгодженому використанню гібридного семплінгу, кластеризації, ієрархічної обробки або кількох класифікаторів [74, 75]. Такі моделі є особливо перспективними для складного трафіку, коли одна архітектура не забезпечує повного охоплення всіх інформативних залежностей [78]. Подальшим розвитком цієї логіки стали розподілені й федеративні схеми СВВ, у яких об'єднуються локальні моделі або локальні оцінки без повної централізації даних, що має практичне значення для неоднорідних корпоративних інфраструктур [96]. У цьому разі аналітичний результат формується не лише на основі локальної вибірки, а й з урахуванням множини джерел даних, що частково підвищує здатність системи працювати за доменних варіацій.

Таким чином, інтелектуальні методи суттєво розширили можливості СВВ порівняно з традиційними сигнатурними системами, бо дали змогу працювати не лише з відомими шаблонами КА, а й зі складними закономірностями мережної поведінки [33]. Їх важливою перевагою є здатність до адаптивного виділення ознак, гнучкість щодо різних типів вхідних даних і потенціал використання в розподілених мережних середовищах [39]. Разом із тим залишаються низка принципових обмежень. По-перше, результативність моделі істотно залежить від способу подання мережних даних, оскільки навіть складна архітектура не гарантує високої якості, якщо вхідне представлення не відображає суттєвих зв'язків між подіями, потоками або вузлами [41]. По-друге, дисбаланс класів зберігає системний характер і не зникає автоматично після переходу до глибших або генеративних моделей [104]. Крім того, некоректне чи надмірно агресивне балансування може саме по собі викликати перекручення оцінки ефективності [69]. По-третє, багато моделей демонструють високу точність у межах конкретного контрольного набору, але втрачають якість при перенесенні на інші домени, що ставить проблему здатності до узагальнення на один рівень важливості з проблемою точності [66, 67].

По-четверте, для глибинного навчання СВВ істотною є вразливість до навмисно модифікованих прикладів, які обходять межу виявлення [113]. Нарешті, практичне впровадження високоточних моделей обмежується вимогами до швидкодії, масштабованості та сумісності з реальною інфраструктурою, особливо в промислових і критичних середовищах [115, 105, 116].

Оскільки ефективність інтелектуальних СВВ безпосередньо залежить від властивостей вибірки, окремого значення набуває аналіз контрольних наборів даних. У задачі виявлення атак набір даних визначає не лише технічну основу експерименту, а й межі інтерпретації результатів: структуру ознак, тип подання трафіку, набір класів КА, рівень дисбалансу та характер розділення на навчальну й тестову частини. Іншими словами, набір даних виступає не нейтральним носієм прикладів, а експериментальною моделлю мережного середовища, від властивостей якої залежить, які висновки взагалі можна зробити щодо придатності СВВ. У сучасній практиці поряд із класичними еталонними наборами активно використовуються спеціалізовані набори даних для інтернету речей (Internet of Things), транспортних, промислових і кіберфізичних середовищ, що краще відображають особливості конкретного домену, але водночас ускладнюють пряме порівняння моделей [61, 70, 71]. Для системи управління інцидентами та суміжних середовищ особливо важливим виявляється не лише факт наявності аномалії, а й структура зібраних даних, яка визначає здатність моделі виявляти атаки змішаної кіберфізичної природи [110, 111].

Іншим критично важливим параметром є характер розподілу класів: багато відкритих СВВ-наборів даних характеризуються істотним перекосом між нормальними зразками та окремими типами атак, що спотворює значення інтегральних метрик і стимулює побудову більш збалансованих версій наборів [68]. Крім того, відрізняється і репрезентативність загроз: одні набори охоплюють базові категорії атак, інші орієнтовані на вузькі домени, наприклад IP або ботнет-активність, що підвищує їх прикладну цінність, але обмежує перенесення результатів на інші типи мереж [73]. Отже, універсального набору даних, однаково придатного для всіх постановок задачі СВВ, фактично не існує. Саме тому вибір

набору даних повинен визначатися не тільки його популярністю чи обсягом, а й відповідністю меті, типу моделі, способу подання трафіку та вимогам до відтворюваності результатів. Це особливо важливо, бо неправильно обрана експериментальна база може підмінити справжні закономірності доменно-специфічними особливостями конкретного набору.

Не менш важливим є питання коректного експериментального оцінювання моделей СВВ. Достовірність висновків визначається не лише вибором архітектури, а й тим, наскільки правильно організовано весь конвеєр дослідження, тобто від формування навчальної і тестової вибірок до вибору метрик і правил попередньої обробки. Однією з найбільш поширених проблем є витік інформації між навчальною та тестовою частинами вибірки, який виникає внаслідок некоректного семплінгу, нормалізації, балансування або відбору ознак. За таких умов модель фактично адаптується і до статистичних властивостей тестових даних, що призводить до завищення підсумкових метрик [72]. Другою принциповою проблемою є некоректне трактування загальна точність як головного критерію ефективності, особливо в умовах дисбалансованих даних. Якщо домінуючий клас становить основну частину вибірки, модель може демонструвати переконливу загальну точність навіть тоді, коли вона погано виявляє рідкісні, але критично важливі типи атак. Отже, оцінювання СВВ має залежати від сукупності показників.

Наступною проблемою є відсутність єдиного стандарту порівняння моделей між різними предметними середовищами, оскільки результати, отримані на наборі даних загального призначення, не можна механічно переносити на диспетчерський контроль та збір даних (SCADA), IP чи інші спеціалізовані мережі [103, 108]. Четвертий ризик пов’язаний із порівнянням моделей, що працюють у різних умовах попередньої обробки, коли вклад семплінгу, відбору ознак або ієрархічної організації даних змішується з вкладом самого класифікатора [85]. Додатково на інтерпретацію результатів впливає доменна специфіка сучасних 5G, програмно-конфігурована мережа (SDN) та інших мережних середовищ, у яких точність залежить не лише від моделі, а й від конфігурації інфраструктури та характеру трафіку [118, 119, 120, 121]. Також, без повної фіксації параметрів вибірки, способу

кодування ознак, порядку балансування та добору гіперпараметрів експеримент втрачає відтворюваність, що є критичним недоліком для порівняння СВВ-підходів [60].

Вибір експериментальної бази повинен відповідати не лише вимогам актуальності загроз, а й вимогам відтворюваності, структурованості та коректного поділу даних. Саме з цієї точки зору використання NSL-KDD є доцільним. Попри історичний характер, цей набір забезпечує фіксований набір ознак мережного з'єднання, придатний як для бінарної, так і для багатокласової постановки задачі, а також дозволяє зберігати відокремлення навчальної та тестової частин без додаткового випадкового змішування. Для дослідження, у якому ключовим є порівняння способів подання трафіку та аналіз їх впливу на якість виявлення атак, така формалізованість має принципове значення. На відміну від вузькодоменних наборів даних, NSL-KDD не прив'язує експеримент до специфіки окремого предметного середовища, що полегшує інтерпретацію результатів як загальнішого дослідження підходу до представлення і аналізу мережного трафіку. Водночас обмеження цього набору, тобто застарілість частини сценаріїв і наявність дисбалансу класів, повинні враховуватися під час інтерпретації результатів, а сам набір доцільно розглядати не як повне відображення сучасного трафіку, а як контрольну експериментальну основу для порівняння підходів [54, 99].

Проведений аналіз інтелектуальних методів та наборів даних показує, що результативність СВКА визначається не лише вибором класифікатора, а й формою вхідного подання мережного трафіку. Саме спосіб репрезентації задає межі подальшого аналізу, оскільки від нього залежить, які залежності можуть бути виявлені, тобто лише статичні табличні співвідношення, чи також часова динаміка, ритміка та локальні структурні зміни. У суміжних галузях аналізу аномалій показано, що якість виявлення істотно залежить від того, наскільки вдало побудоване представлення даних, особливо якщо процес має багатовимірну і виражено часову природу [35, 36]. Подібний висновок простежується і в задачах моніторингу технічних та транспортних процесів, де ефективність моделі визначається не лише алгоритмом, а й формою сигналу, яка повинна зберігати

змістовну динаміку спостережуваного явища [42, 43]. Для мережного трафіку це має особливе значення, оскільки він є процесом, що розгортається у часі, характеризується зміною інтенсивності, повторюваністю подій і структурними переходами [93, 94].

Одним із перспективних напрямів альтернативного подання мережних даних є соніфікація, тобто перетворення числових, часових або багатовимірних характеристик процесу у звукову форму. Соніфікація є не простим озвучуванням даних, а способом побудови подання, у якому параметри джерела відображаються через характеристики звуку, тобто висоту тону, амплітуду, ритм, тембр або тривалість. Ефективність такого подання визначається збереженням інформативності, часовою узгодженістю та здатністю робити помітними зміни режиму функціонування [37]. У методологічному сенсі це означає, що соніфікація повинна розглядатися не як ілюстративний або допоміжний засіб, а як формальний механізм трансформації мережного трафіку в інший інформаційний простір, придатний для подальшого аналізу. За логікою побудови можна виділити параметричний і подієвий підходи. Перший забезпечує безперервне відображення змін стану системи в акустичних характеристиках, другий - звукове маркування окремих значущих подій або переходів. Для складних динамічних середовищ найбільш доцільним є їх поєднання, оскільки воно дозволяє одночасно відображати загальний перебіг процесу і локалізувати критичні відхилення [79, 92].

Доцільність використання соніфікації в задачі СВКА зумовлюється насамперед динамічною природою самого мережного трафіку. Аномальна активність часто проявляється не у значенні окремої ознаки, а в ритміці, повторюваності, нерівномірності потоку або зміні співвідношення між його характеристиками. Для таких випадків форма подання стає критично важливою. У дослідженнях аномалійного аналізу мережного та IP-трафіку показано, що здатність моделі виділяти приховані закономірності безпосередньо залежить від організації вхідного представлення [46]. Додатковою передумовою є обмеженість суто візуального підходу до інтерпретації складних потокових даних: зі зростанням кількості каналів спостереження й швидкості зміни станів графічні панелі

втрачають частину своєї чутливості до короткочасних або малоконтрастних відхилень [15]. У такій ситуації альтернативний канал представлення може мати не допоміжне, а аналітичне значення. Разом із розвитком глибоких моделей, орієнтованих на обробку сигналів і двовимірних подань, це створює можливість розглядати соніфікацію не як допоміжний засіб оповіщення, а як формальний етап побудови альтернативного вхідного представлення для подальшого аналізу [87, 98]. Особливо це узгоджується з атаками, для яких самостійне діагностичне значення мають часові конфігурації та зміна режимів навантаження, зокрема при DDoS і координованих сценаріях мережної активності [26, 122].

Водночас привабливість соніфікації полягає ще й у тому, що вона відкриває шлях до подальшого спектрального аналізу. Якщо мережний трафік перетворено на звуковий сигнал, то він може бути досліджений як часовий ряд, спектр або двовимірне часово-частотне зображення. Для моделей типу згорткова нейронна мережа (CNN) саме такі структури часто є більш природними, ніж традиційний вектор ознак. У цьому разі виграш може бути пов’язаний не лише зі зміною форми подання, а й із переходом до іншого типу аналітичного простору, де приховані закономірності аномальної поведінки стають більш виразними. Саме така логіка особливо важлива для дисертаційного дослідження, орієнтованого на використання соніфікації не для людського прослуховування як такого, а для подальшого інтелектуального аналізу мережного трафіку.

Разом із тим використання соніфікації в СВВ не можна вважати методологічно повністю усталеним. Основними обмеженнями цього напряму є відсутність уніфікованого принципу відображення мережних характеристик у звуковий простір, ризик втрати частини інформації під час перетворення, складність забезпечення надійної розрізнюваності між нормальним трафіком і різними типами атак, а також відсутність загальноприйнятих еталонних процедур оцінювання саме для соніфікованих моделей. Додатковими проблемами залишаються залежність результату від типу мережного середовища, неоднозначне співвідношення між соніфікацією як засобом людського сприйняття і як формальним поданням для машинного аналізу, а також те, що саме перетворення

трафіку у звук не усуває автоматично класичні проблеми СВВ, які включають дисбаланс класів, ризик перенавчання та залежність від якості вибірки. Отже, соніфікацію доцільно розглядати не як готове вирішення проблеми, а як перспективний спосіб формування вхідних даних, ефективність якого потребує окремого формального обґрунтування та експериментальної перевірки.

Необхідно також враховувати, що в задачі виявлення вторгнень будь-яке альтернативне представлення даних має оцінюватися не само по собі, а порівняно з традиційними формами організації трафіку. Інакше кажучи, цінність соніфікації визначається не фактом новизни, а здатністю забезпечити такий приріст інформативності, який виправдовує ускладнення аналітичного конвеєра. Для цього необхідно зберігати методологічну чистоту експерименту, контролювати поділ на навчальну й тестову вибірки та уникати змішування внеску самого представлення і внеску класифікатора. Саме тому вибір контрольного набору даних, придатного для відтворюваного порівняльного дослідження, є принципово важливим.

Таким чином, проведений аналіз показує, що сучасні інтелектуальні підходи суттєво розширили можливості СВВ, однак їх результативність залишається тісно пов’язаною з характеристиками набору даних, коректністю експериментальної постановки та способом подання мережного трафіку. У цьому контексті вибір NSL-KDD як контрольної експериментальної бази та розгляд соніфікації як альтернативного представлення мережних даних є методологічно обґрунтованими. Водночас наявні підходи не усувають низки принципових проблем, пов’язаних із залежністю результатів від структури вхідних даних, дисбалансом класів, обмеженою узагальнювальною здатністю моделей і ризиком завищених оцінок якості.

1.3 Основні проблеми сучасних підходів до виявлення комп’ютерних атак

Проведений аналіз сучасних систем виявлення вторгнень, інтелектуальних методів аналізу мережного трафіку, контрольних наборів даних та альтернативних способів подання інформації показав, що попри суттєвий розвиток СВВ задача

підвищення точності виявлення КА у КМ залишається невирішеною в повному обсязі. У наявних підходах простежуються суттєві методологічні та прикладні обмеження, які знижують здатність системи надійно розпізнавати аномальну активність, особливо в умовах мінливості трафіку, нерівномірного представлення класів атак і необхідності коректного експериментального оцінювання.

Першою з таких проблем є залежність результатів виявлення КА від способу представлення вхідних даних. У значній частині досліджень основну увагу приділено вибору класифікатора, оптимізації його параметрів або ускладненню архітектури, тоді як форма подання мережного трафіку часто розглядається як другорядний технічний етап. Проте для СВВ саме представлення даних визначає, які закономірності мережної поведінки стають доступними для моделі, а які залишаються прихованими. Найпоширенішим у практиці залишається табличне або векторне подання трафіку, коли кожне з'єднання чи потік описується фіксованим набором числових і категоріальних ознак. Такий підхід є зручним для формалізації, добре узгоджується з класичними алгоритмами машинного навчання й полегшує відтворюване експериментальне оцінювання. Саме на цьому принципі побудовано значну частину автоенкодерних і класифікаційних схем для СВВ [53], а також підходів до інтелектуальної класифікації трафіку в програмно-керованих мережах, де результат істотно залежить від вибору вхідних параметрів та способу їх нормалізації [87]. Разом із тим статичний вектор ознак не завжди дозволяє повноцінно відобразити внутрішню часову організацію процесу, ритміку подій, зміну режимів функціонування або локальні аномальні шаблони. Якщо КА проявляється не абсолютним значенням окремої характеристики, а послідовністю дій, структурою обміну чи повторюваністю нетипових комбінацій параметрів, то плоский ознаковий опис виявляється недостатнім для її надійного розпізнавання. У такому випадку проблема полягає не лише в потужності класифікатора, а в самому характері вхідного подання, яке не містить повного опису структури, що підлягає виявленню.

Зазначене пояснює інтерес до альтернативних форм репрезентації мережних даних, зокрема до послідовнісних, багатоканальних, сигналоподібних та доменно-

специфічних представлень. Для середовищ Інтернету речей, 5G та інших розподілених мереж питання форми вхідних даних набуває ще більшої ваги, оскільки нормальна й аномальна активність відрізняються не лише значеннями ознак, а й структурою взаємодії, динамікою звернень і контекстом обміну. У дослідженнях intrusion detection для 5G Інтернету речей підкреслюється, що сама постановка задачі передбачає адаптацію моделі до особливостей середовища, а отже й до відповідного способу організації вхідної інформації [62]. Це означає, що одна й та сама модель може демонструвати різну результативність не тільки через зміну домену, а й через зміну форми подання трафіку. Для СВВ така залежність має системний характер: різні представлення зберігають різні аспекти мережної поведінки, а відтак або відкривають, або приховують для моделі ті закономірності, на яких ґрунтується розпізнавання атак. Саме тому підвищення якості СВВ не може зводитися лише до пошуку більш потужного алгоритму, бо воно потребує обґрунтування такого способу репрезентації мережного трафіку, який зберігає часову, структурну та контекстну природу аномальної поведінки.

Другою фундаментальною проблемою є дисбаланс класів, що істотно впливає на якість виявлення рідкісних типів атак. У більшості реальних і відкритих наборів даних нормальний трафік або окремі поширені класи загроз кількісно переважають над малочисельними підкласами аномалій. Для СВВ це має не лише статистичний, а й безпосередній прикладний зміст, оскільки саме рідкісні класи часто відповідають найбільш небезпечним сценаріям вторгнення, виявлення яких і становить основну практичну цінність системи. За наявності такого перекосу більшість алгоритмів навчання, орієнтованих на мінімізацію середньої помилки на всій вибірці, природно зміщуються в бік домінуючих станів. У результаті формуються нечіткі межі відокремлення для рідкісних атак, зростає кількість хибнонегативних рішень і погіршується чутливість до тих класів, які представлені найменшою кількістю прикладів [48]. Особливо гостро ця проблема проявляється в багатокласовій постановці, коли малочисельні типи атак не лише рідко зустрічаються, а й частково перекриваються з нормальним трафіком або з іншими підкласами аномалій. За таких умов дисбаланс класів поєднується з низькою

міжкласовою відокремлюваністю, і навіть складні глибокі моделі виявляються недостатньо чутливими без спеціалізованих механізмів балансування або модифікації процедури навчання [49].

Проблема дисбалансу додатково ускладнюється тим, що її вплив не обмежується лише етапом побудови класифікатора. Вона безпосередньо спотворює інтерпретацію результатів, оскільки висока загальна точність може досягатися за рахунок правильного прогнозування домінуючих класів, тоді як якість розпізнавання рідкісних і критичних загроз залишається низькою. У такій ситуації інтегральна оцінка ефективності не відображає реального рівня захищеності системи. Для СВВ це є принципово важливим, бо практична цінність системи визначається не середнім успіхом на типових прикладах, а здатністю не пропускати саме ті аномалії, які є найнебезпечнішими. Тому дисбаланс класів слід враховувати не лише під час навчання моделі, а й під час вибору метрик, побудови процедури порівняння результатів та організації вибірки. Сучасні підходи до компенсації дисбалансу, тобто повторне формування вибірки, синтетичне розширення вибірки, модифікація функції втрат, ансамблеві схеми, безумовно покращують ситуацію, але не усувають проблему автоматично. Якщо балансування виконується без урахування внутрішньої структури даних, можливе формування штучних зразків, які погано відповідають реальним мережним шаблонам. У такому випадку модель демонструє покращення на контрольному наборі, але не набуває достатньої здатності виявляти рідкісні атаки в нових умовах. Дослідження, присвячені побудові збалансованих наборів даних для виявлення вторгнень, підтверджують, що якість компенсації дисбалансу визначається не стільки кількістю доданих прикладів, скільки збереженням змістовної структури класу [80, 81]. Це має особливе значення, оскільки проблема дисбалансу тісно пов’язана зі способом представлення трафіку: якщо рідкісний клас на рівні вхідних даних подано недостатньо виразно, то навіть формальне балансування не гарантує помітного покращення його виявлення [47].

Третьою ключовою проблемою є обмежена узагальнювальна здатність моделей та ризик перенавчання. Для СВВ це означає, що система може

демонструвати високі результати в межах конкретного контрольного набору даних, але втрачати ефективність після перенесення в інше мережне середовище, за зміни характеру трафіку, структури ознак або співвідношення між класами. Така ситуація є особливо небезпечною, оскільки формально успішна модель створює враження надійності, проте фактично виявляється налаштованою на статистичні особливості конкретної вибірки, а не на фундаментальні ознаки поведінки під час КА. Причина цього полягає в навчанні на обмежених або доменно-залежних даних, тобто якщо набір прикладів не охоплює достатнього різноманіття нормальних режимів і аномальних сценаріїв, то модель починає відтворювати локальні закономірності навчальної вибірки, які не мають узагальнювального значення поза її межами. Для СВВ це проявляється як різке зниження якості після переходу від контрольованого еталонного оцінювання-середовища до реального трафіку, де варіативність мережної поведінки значно вища, а межі між нормою та аномалією менш очевидні.

Ризик перенавчання особливо зростає при використанні глибоких архітектур, здатних формувати складні внутрішні залежності. За недостатнього різноманіття даних така модель починає підлаштовуватися до випадкових або другорядних особливостей навчальної вибірки. Це може проявлятися в надмірній чутливості до окремих ознак, чутливості до зміни розподілу даних і різкому погіршенні результатів на зовнішніх наборах. Подібні труднощі є характерними не лише для СВВ, а й для задач аномального аналізу в інших динамічних середовищах, де модель повинна розрізняти нормальні варіації процесу та справжні відхилення, а отже проблема узагальнення набуває центрального методологічного значення [32]. Для систем виявлення атак ця проблема тісно пов’язана зі способом представлення трафіку. Якщо модель навчається на ознаковому просторі, який відображає лише часткові або поверхневі характеристики мережної взаємодії, то вона неминуче орієнтується на випадкові шаблони. Навпаки, чим ближче представлення даних до внутрішньої структури процесу, тим більша ймовірність, що модель сформує змістовніші правила розмежування між нормальним і аномальним режимом. Отже, ризик перенавчання посилюється не лише складністю архітектури, а й невідповідністю між формою подання трафіку та реальною природою

закономірностей, які необхідно виявити. Одним із напрямів послаблення цієї проблеми є використання підходів, орієнтованих на формування більш узагальнених внутрішніх представлень, зокрема перенесення навчання та повторне використання узагальнених репрезентацій [65]. Однак і ці підходи не усувають повністю проблему, якщо не вирішено питання інформативного вхідного подання та коректної методології оцінювання.

Четвертою проблемою є завищення показників точності в дослідженнях СВВ. У наукових роботах часто повідомляється про дуже високі значення загальної точності, що на перший погляд створює враження майже повного розв’язання задачі класифікації мережних атак. Проте для СВВ високе значення загальної точності ще не означає, що модель набула достатньої здатності розпізнавати атаки в реальному середовищі. Однією з причин завищених оцінок є надмірно локальний характер експерименту. Якщо модель перевіряється в межах вузького домену, на обмеженому класі загроз або в середовищі зі стабільною структурою трафіку, то висока точність може відображати не загальну ефективність СВВ, а лише добру адаптацію до конкретної постановки [34]. Подібна ситуація виникає і в роботах, орієнтованих на виявлення атак у вузько визначених мережних архітектурах, де сам контекст дослідження суттєво спрощує задачу розмежування між нормальним і аномальним трафіком [122]. Іншою причиною є невідповідність між реальною складністю задачі та способом її експериментальної формалізації. Якщо багатокласову проблему фактично зведено до грубої бінарної класифікації або з дослідження виключено найбільш складні та малочисельні типи атак, то підсумкова точність закономірно зростає. Проте це не означає, що модель стала краще розпізнавати аномалії як такі; лише спрощується сама постановка задачі.

Суттєвий внесок у завищення показника загальної точності робить і некоректна організація навчальної та тестової частин вибірки. Якщо під час нормалізації, балансування, масштабування або відбору ознак використовується інформація з усієї вибірки без чіткого розмежування між навчальною і тестовою частинами, модель непрямо одержує доступ до характеристик тестових даних ще до фінальної перевірки. У такому випадку підсумковий результат втрачає

властивість незалежної оцінки й починає відображати ефект методологічного витоку інформації. Додатковим чинником завищення показників виступає і дисбаланс класів. Він полягає в тому, що за домінування одного стану модель може досягати високої загальної точності навіть за умови слабкого виявлення рідкісних, але критичних типів атак. У такій ситуації загальна точність перестає бути надійним критерієм якості СВВ, оскільки маскує систематичні помилки щодо малочисельних класів. Нарешті, висока точність може бути наслідком адаптації моделі до специфічних статистичних закономірностей конкретного набору даних. Якщо контрольний набір має обмежений спектр варіацій трафіку або стабільні штучні шаблони, модель починає використовувати саме їх як основу для рішення, а не фундаментальні ознаки атакувальної поведінки. Саме тому високе значення загальної точності не може розглядатися як самодостатній доказ ефективності СВВ. Для отримання достовірних висновків необхідно оцінювати модель у межах методологічно коректної постановки, аналізувати результати для окремих класів загроз і розглядати точність лише як одну з часткових характеристик системи.

Узагальнення наведених проблем дає змогу сформулювати протиріччя. Воно полягає в тому, що сучасні підходи до побудови СВВ здебільшого орієнтовані або на вдосконалення архітектури класифікатора або на локальні процедури балансування й оптимізації, але недостатньо враховують взаємозв’язок між способом подання мережного трафіку, чутливістю до рідкісних класів атак, узагальнювальною здатністю моделі та коректністю експериментального оцінювання. Інакше кажучи, наявні рішення часто покращують окрему складову задачі, але не формують цілісного підходу, у якому представлення даних, процедура навчання та перевірки, робота з дисбалансом і вибір моделі були б узгодженими елементами єдиної системи підвищення точності СВВ. Особливо виразно це проявляється тоді, коли модель працює з традиційним табличним описом трафіку, що не повною мірою відображає часову й структурну природу аномальної поведінки. У такій ситуації навіть складні алгоритми виявляються обмеженими інформаційним потенціалом самого вхідного подання.

Для подолання протиріччя перспективний підхід до підвищення точності СВВ повинен відповідати кільком узгодженим вимогам. По-перше, він має ґрунтуватися на більш інформативному способі представлення мережного трафіку, який зберігав би не лише статичні значення окремих ознак, а й часову динаміку, ритміку, структурні співвідношення між параметрами та локальні шаблони зміни стану мережі. По-друге, такий підхід повинен бути менш чутливим до дисбалансу класів і забезпечувати підвищення чутливості саме до рідкісних і складних підкласів загроз, а не лише покращення загальної точності за рахунок домінуючих станів. По-третє, він має спиратися на коректну методологію навчання та перевірки, яка виключає змішування навчальної й тестової інформації та не допускає штучного завищення підсумкових метрик. По-четверте, перспективний підхід повинен забезпечувати достатню здатність до узагальнення, тобто бути орієнтованим не на відтворення локальних закономірностей конкретного набору даних, а на виявлення більш узагальнюваних ознак аномальної мережної поведінки. По-п’яте, він має бути сумісним із сучасними інтелектуальними моделями аналізу, зокрема згортковими, рекурентними або комбінованими архітектурами, щоб нове представлення трафіку могло бути безпосередньо використане в повному аналітичному конвеєрі СВВ. І, нарешті, такий підхід повинен залишатися практично придатним, тобто не втрачати можливості інтеграції в реальні системи мережного моніторингу та підтримки рішень у контурі кіберзахисту [26, 38].

Отже, проведений аналіз свідчить, що основні проблеми сучасних підходів до виявлення комп’ютерних атак зосереджені навколо чотирьох взаємопов’язаних аспектів: залежності результатів від способу представлення даних, негативного впливу дисбалансу класів на виявлення рідкісних типів атак, обмеженої здатності до узагальнення моделей і ризику перенавчання, а також завищення показників точності внаслідок некоректної експериментальної постановки. Сукупність цих обмежень показує, що подальше підвищення ефективності СВВ не може зводитися лише до вибору більш складного класифікатора. Необхідним є формалізований підхід, у якому спосіб подання мережного трафіку, організація навчання, механізми

підвищення чутливості до слабко представлених атак і коректність експериментального оцінювання розглядались би як єдине ціле. Саме ця потреба і визначатиме напрям подальшого дослідження, пов’язаного з розробленням моделі та методу виявлення КА на основі альтернативного подання мережного трафіку та його подальшого інтелектуального аналізу.

1.4 Постановка задачі дослідження

Аналіз сучасних методів і систем виявлення комп’ютерних атак у корпоративних мережах показав, що наявні сигнатурні, статистичні, машинні та глибокі підходи забезпечують автоматизацію процесу виявлення мережних загроз, однак не усувають низки суттєвих обмежень, пов’язаних із залежністю результатів від способу представлення трафіку, зниженням якості розпізнавання рідкісних типів атак в умовах дисбалансу класів, обмеженою узагальнювальною здатністю моделей за структурної варіативності даних і ризиком одержання формально високих, але методологічно завищених показників якості. Для розв’язання задачі підвищення точності систем виявлення КА у КМ необхідно розробити модель і метод виявлення КА на основі соніфікації мережного трафіку, спектрального подання даних та інтелектуального аналізу отриманих образів і вирішити такі завдання:

1. Провести аналіз методів, засобів і технологій виявлення комп’ютерних атак у корпоративних мережах, а також дослідити обмеження існуючих підходів до подання мережного трафіку та класифікації атак в умовах дисбалансу класів.

2. Розробити модель процесу виявлення комп’ютерних атак, яка поєднує етапи попередньої обробки мережного трафіку, його соніфікації, спектрального аналізу, базової класифікації, адаптивного балансування та каскадного прийняття рішень.

3. Розробити метод частотно-часового перетворення мережного трафіку, придатний для побудови структурованого двовимірного подання даних і

застосування нейромережевих моделей аналізу зображень до задачі бінарної класифікації атак.

4. Розробити метод сигнатурозбережного адаптивного балансування навчальної вибірки, який забезпечує зменшення впливу дисбалансу класів і підвищення репрезентативності міноритарних підкласів атак.

5. Розробити метод підвищення точності виявлення атак у зоні невизначеності прогнозу на основі аналізу помилок базового класифікатора та селективного перевизначення рішень за допомогою допоміжної моделі.

6. Розробити програмно-алгоритмічне забезпечення для валідації запропонованих моделей і методів та провести експериментальні дослідження їх ефективності.

Під підвищенням точності системи виявлення КА у КМ у роботі будемо розуміти не лише зростання інтегральних показників якості класифікації, а й зменшення кількості пропущених КА, підвищення чутливості до рідкісних і складних підкласів загроз, покращення збалансованості рішень між нормальним та атакувальним класами і підтвердження досягнутого ефекту в межах коректної й відтворюваної експериментальної постановки. Удосконалення способу представлення мережного трафіку, розроблення методу його інтелектуального аналізу та побудова коректної схеми навчання й оцінювання моделі розглядатимемо як основу підвищення ефективності систем виявлення КА порівняно з відомими підходами.

1.5 Висновки до першого розділу

Проведене дослідження сучасного стану проблеми виявлення КА у КМ, систем виявлення вторгнень, підходів до інтелектуального аналізу мережного трафіку та способів подання вхідних даних дало змогу зробити такі висновки:

1. Встановлено, що підвищення ефективності систем виявлення комп’ютерних атак у корпоративних мережах залишається актуальною науково-прикладною задачею, що обумовлено ускладненням мережних середовищ,

зростанням різноманітності сценаріїв КА, появою нових і модифікованих загроз та підвищенням вимог до своєчасності реагування на аномальну активність.

2. Аналіз систем виявлення вторгнень підтвердив, що сигнатурні, аномальні та гібридні СВВ мають різні переваги й обмеження, однак жоден із відомих підходів не забезпечує повного та універсального розв'язання задачі виявлення КА, особливо в умовах мінливості мережного трафіку, появи нових типів загроз і необхідності збереження високої точності розпізнавання.

3. Встановлено, що сучасні методи машинного і глибокого навчання суттєво розширюють можливості СВВ порівняно з традиційними сигнатурними та статистичними підходами, проте їх ефективність істотно залежить від способу представлення вхідних даних, структури вибірки, наявності дисбалансу класів, ризику перенавчання та здатності моделі до узагальнення в нових умовах функціонування.

4. Згідно аналізу контрольних наборів даних для задачі виявлення вторгнень встановлено доцільність використання набору NSL-KDD як основи для експериментального дослідження, оскільки він забезпечує структуроване подання мережних ознак, фіксоване розділення на навчальну та тестову частини і можливість відтворюваного порівняння результатів, однак потребує врахування історичних обмежень і нерівномірного представлення окремих класів атак.

5. Обґрунтовано доцільність розгляду соніфікації мережного трафіку як альтернативного способу подання даних у задачі виявлення КА, оскільки таке подання дає змогу враховувати часову, структурну та частотну природу мережних процесів і створює передумови для виділення закономірностей аномальної поведінки, що не завжди повно відображаються у традиційному табличному описі.

6. Встановлено, що використання соніфікації в системах виявлення КА потребує розв'язання низки невирішених питань, пов'язаних із вибором правил перетворення мережних ознак у сигналоподібне подання, збереженням інформативності нормальної та атакуючої активності, а також забезпеченням коректного й відтворюваного оцінювання побудованих моделей.

7. Встановлено та обгрунтовано наявність суперечності між потребою підвищення точності систем виявлення КА і обмеженими можливостями існуючих підходів, що переважно ґрунтуються на традиційному векторному поданні мережного трафіку та недостатньо враховують його часово-структурну природу та не забезпечують належної чутливості до слабко представлених і складних підкласів загроз.

8. Перспективним напрямом подальших досліджень відповідно до здійсненого аналізу є розроблення моделі та методу виявлення КА на основі соніфікації мережного трафіку, спектрального подання даних і сучасних методів інтелектуального аналізу, в яких одночасно враховуються інформативність представлення трафіку, вплив дисбалансу класів, коректність експериментальної постановки та практична придатність для підвищення точності СВВ у КМ.

Основні результати розділу опубліковані у працях [127–132].

РОЗДІЛ 2
МОДЕЛЬ ПРОЦЕСУ ВИЯВЛЕННЯ КОМП'ЮТЕРНИХ АТАК НА ОСНОВІ ЧАСТОТНО-ЧАСОВОГО ПРЕДСТАВЛЕННЯ МЕРЕЖНОГО ТРАФІКУ

Виявлення КА у мережному трафіку є однією з ключових задач забезпечення інформаційної безпеки КМ. Сучасні інформаційні системи функціонують в умовах зростання обсягів передачі даних, ускладнення структури мережної інфраструктури та збільшення різноманітності типів атак, що потребує постійного удосконалення методів виявлення КА.

У загальному випадку задача виявлення КА полягає у визначенні належності окремого мережного з'єднання або потоку даних до класу легітимної активності або до одного з типів атак. Формально ця задача відноситься до класу задач класифікації, проте її специфіка обумовлена низкою факторів, зокрема: високою розмірністю простору ознак; наявністю шумових та корельованих характеристик; дисбалансом класів у навчальних вибірках; складністю формалізації ознак. Вони відображають поведінкові аспекти мережної активності.

Окремим проблемним завданням є завдання щодо коректної побудови експериментальної моделі та оцінювання її якості. Високі показники точності виявлення КА не завжди супроводжуються строгим дотриманням принципів розділення навчальної та тестової вибірок або врахуванням дисбалансу класів. Тому, необхідно забезпечити дотримання принципу строгого розмежування етапів навчання та тестування, що виключає змішування даних та забезпечує коректність оцінювання узагальненої здатності моделі.

Тому, виникає необхідність формалізованого опису задачі виявлення КА, який включає визначення об'єкта дослідження, структури вхідних даних, простору ознак, цільової змінної та механізму прийняття рішення. Така формалізація є необхідною передумовою побудови узагальненої моделі процесу виявлення вторгнень та на її основі розроблення методів виявлення КА для підвищення точності, зокрема на основі частотно-часового представлення мережного трафіку.

2.1 Узагальнена модель процесу виявлення комп’ютерних атак

Область дослідження охоплює процес виявлення КА у КМ на основі аналізу мережного трафіку. У сучасних корпоративних інформаційних системах мережний трафік відображає результати взаємодії між вузлами, сервісами та користувачами, а тому містить сукупність ознак, за якими можна виявляти як нормальну активність, так і прояви зловмисного впливу. Саме тому мережний трафік розглядається як інформаційна основа для побудови системи виявлення атак, здатної автоматично приймати рішення щодо наявності чи відсутності загрози.

Мережний трафік розглядатимемо не як несистематизований потік пакетів, а як сукупність мережних з’єднань або потоків даних, для яких можуть бути сформовані структуровані описи. Кожне таке з’єднання характеризується набором параметрів, що відображають часові, протокольні, статистичні та поведінкові властивості комунікації. До них можуть належати тривалість з’єднання, обсяг переданих даних, тип протоколу, стан службових полів, інтенсивність обміну, частота звернень до сервісів та інші характеристики, які є інформативними для задачі виявлення комп’ютерних атак. Таким чином, елементарною одиницею аналізу в роботі приймається окреме мережне з’єднання, яке подане через набір ознак, що придатні для подальшої алгоритмічної обробки.

Важливою складовою області дослідження є перехід від «сирого» мережного трафіку до структурованого ознакового подання. Первинні пакети самі по собі є незручними для безпосереднього використання у класифікаційній моделі, тому потрібно виконувати етап попередньої обробки, під час якого з мережного обміну формуються інформативні характеристики з’єднань. У результаті цього повинен створюватись ознаковий опис, який відображатиме мережну подію у вигляді уніфікованого набору числових і категоріальних параметрів. Саме таке подання є базою для подальшого переходу до частотно-часового представлення мережного трафіку.

Змістовно задача дослідження полягає у побудові моделі, яка за сукупністю ознак мережного з’єднання здатна визначати його належність до нормальної

активності або до атакувального класу. У прикладному аспекті це означає розроблення підходу, що дає змогу на основі характеристик мережного трафіку своєчасно виявляти КА в корпоративному середовищі. При цьому акцент повинен бути саме на задачі виявлення факту КА, а не на детальній класифікації всіх можливих її різновидів. Такий підхід є доцільним, оскільки в умовах практичного використання першочерговим є надійне розмежування нормальної та аномальної мережної активності, тоді як подальша деталізація типу загрози може виконуватися на наступних етапах аналізу.

Особливістю досліджуваної області є складна структура мережних даних. По-перше, ознаковий простір таких даних зазвичай є багатовимірним і може містити надлишкові, корельовані або слабкоінформативні характеристики. По-друге, у наборах даних для задачі виявлення вторгнень зазвичай спостерігається виражений дисбаланс класів, тобто нормальні з’єднання та окремі масові типи атак представлені значно ширше, ніж рідкісні або складні підкласи загроз. По-третє, частина мережних подій може перебувати поблизу межі між нормальною та атакувальною поведінкою, що ускладнює прийняття однозначного рішення. Сукупність цих властивостей зумовлює потребу в розробленні спеціалізованих моделей і методів, здатних працювати в умовах структурної неоднорідності даних, дисбалансу класів і нечіткої відокремлюваності окремих спостережень.

Ще однією важливою особливістю області дослідження є необхідність урахування невизначеності прогнозу. У реальних умовах функціонування корпоративної мережі частина з’єднань може мати ознаки, що одночасно частково відповідають і нормальній активності, і потенційним атакам. У таких випадках система виявлення не повинна обмежуватися лише жорстким пороговим рішенням, а має враховувати наявність області невизначеності, у межах якої доцільним є додатковий аналіз складних прикладів. Саме тому область дослідження охоплює не лише побудову базової моделі класифікації, а й розроблення підходів до підвищення точності прийняття рішень у тих випадках, коли прогноз базової моделі не є достатньо впевненим.

Коректність дослідження в цій предметній області також передбачає дотримання принципу незалежності навчання та оцінювання. Модель повинна навчатися лише на навчальній частині даних, тоді як її ефективність має перевірятися на окремій тестовій вибірці, яка не використовується на етапі налаштування. Такий підхід є необхідним для об’єктивного оцінювання узагальнювальної здатності моделі та запобігає штучному завищенню показників точності. Отже, область дослідження включає не тільки вибір способу подання мережного трафіку і побудову класифікаційного правила, а й коректну організацію експериментальної перевірки отриманих результатів.

Таким чином, область дослідження визначається як процес аналізу мережної активності корпоративної інформаційної системи з метою виявлення КА на основі структурованого подання мережних з’єднань. Таке визначення області дослідження створює змістову основу для подальшого подання моделі.

Узагальнену модель процесу виявлення КА розглядатимемо як багатоетапний процес, що включатиме етапи попередньої обробки даних, формування простору ознак, перетворення представлення даних, навчання моделі та прийняття рішення. На відміну від окремих алгоритмічних реалізацій здійснимо акцент на концептуальну модель процесу, яка дасть змогу забезпечити можливість інтеграції різних підходів у межах єдиного формалізованого процесу.

Особливістю моделі є використання альтернативного способу представлення мережного трафіку, тобто переходу від векторного опису ознак до частотно-часової області. Це дозволяє розглядати мережне з’єднання як сигнал, характеристики якого можуть бути проаналізовані за допомогою спектральних методів. Такий підхід формує основу для побудови моделі, яка здатна враховувати внутрішні залежності між ознаками та зменшувати вплив шумових компонент.

Узагальнену модель процесу виявлення КА розглядатимемо як математично формалізовану структуру, що поєднує оператори відображення, функцію класифікації та механізм прийняття рішення.

Насамперед доцільно визначити логічну структуру процесу виявлення КА, яка задаватиме послідовність функціональних етапів обробки мережного трафіку

та взаємозв’язки між ними. Логічна структура процесу виявлення КА визначає послідовність функціональних етапів обробки мережного трафіку та взаємозв’язки між ними. Процес розглядатимемо як багаторівневу інформаційно-аналітичну структуру, що відображає перехід від «сирого» мережного трафіку до формалізованого рішення про наявність або відсутність атаки.

На першому рівні здійснюватимемо збір та агрегацію мережних даних. «Сирі» пакети будемо об’єднувати у логічно завершені з’єднання або потоки відповідно до протокольних характеристик та часових параметрів. Цей етап формує структуровану основу для здійснення подальшого аналізу.

Другий рівень відповідає за попередню обробку даних і формування простору ознак. На цьому етапі виконуватимемо виділення характеристик, що описують поведінкові та статистичні параметри з’єднання, кодування категоріальних атрибутів та нормалізація числових значень. Результатом є векторне представлення кожного мережного з’єднання.

Третій рівень пов’язаний із перетворенням представлення даних у вибрану область аналізу. У традиційних системах класифікація виконується безпосередньо у векторному просторі ознак. У розроблюваному підході передбачено додатковий етап трансформації даних у частотно-часове подання, що дає змогу враховувати структурні взаємозв’язки між ознаками у новій аналітичній області.

Четвертий рівень реалізує модель класифікації. На цьому етапі формуємо функцію прийняття рішення, параметри якої визначаються у процесі навчання. Модель отримує підготовлене представлення даних та формує прогноз щодо належності з’єднання до класу нормальної активності або атаки.

П’ятий рівень відповідає за інтерпретацію та використання результату. Отримане рішення може бути передане до підсистем реагування, журналювання або централізованого аналізу. За централізованої організації процесу передбачена можливість оновлення параметрів моделі на основі накопичених даних без зміни загальної логічної структури системи.

Таким чином, логічна структура процесу виявлення КА буде складатись з послідовних етапів збору даних, формування ознак, перетворення представлення,

класифікації та прийняття рішення. Такий підхід забезпечуватиме чітке розмежування функціональних компонентів системи та створює основу для подальшої деталізації моделі і, відповідно, підвищення точності.

Інфологічна модель процесу виявлення КА відображає структуру інформаційних потоків між її основними компонентами та визначає послідовність перетворень даних від моменту їх надходження до формування рішення. Така модель дозволяє описати процес на концептуальному рівні без прив’язки до конкретної програмної реалізації.

Початковим джерелом інформації є мережне середовище, у якому формуються пакети даних. Потік пакетів надходить до підсистеми збору та агрегування, де виконується формування логічно завершених мережних з’єднань. На цьому етапі здійснюємо структуризацію даних відповідно до протоколів і часових характеристик.

Наступний інформаційний потік спрямовується до підсистеми формування ознак. Тут кожне з’єднання перетворюємо у вектор параметрів, що описують його технічні та статистичні характеристики. Відповідні атрибути можуть включати часові інтервали, кількісні показники переданих даних, службові поля протоколів та інші характеристики. Після цього формуємо структурований масив даних, який придатний для аналізу.

У межах розроблюваного підходу передбачено додатковий етап перетворення інформації у частотно-часову область. Вектор ознак трансформується у альтернативне представлення, що дозволяє врахувати внутрішні залежності між компонентами. Таким чином, інфологічна модель включає окремий інформаційний потік, який відповідає за формування розширеного подання даних.

Далі інформація надходить до підсистеми класифікації, де на основі навченої моделі формується прогноз щодо належності з’єднання до класу нормальної активності або атаки. Результат класифікації передається до підсистеми прийняття рішення та реагування, що може включати генерацію попереджень, запис у журнал подій або передачу даних до централізованого сховища.

Важливим елементом інфологічної моделі є зворотний інформаційний потік, який забезпечує накопичення результатів класифікації та можливість подальшого оновлення параметрів моделі. У межах централізованої моделі процесу цей механізм дає змогу періодично коригувати модель без зміни загальної структури системи.

Таким чином, інфологічна модель інформаційних потоків описує процес виявлення КА послідовністю взаємопов’язаних частин, між якими циркулюють структуровані дані. Такий опис забезпечує цілісне подання процесу функціонування системи виявлення КА та створює основу для подальшої формалізації її компонентів.

Функціональна схема процесу обробки в системі виявлення КА визначає послідовність операцій, які виконуються над даними від моменту їх надходження до формування кінцевого рішення. На відміну від інфологічної моделі, що відображає рух інформаційних потоків, функціональна схема акцентує увагу на логіці виконання обчислювальних та аналітичних процедур.

Першим етапом є отримання мережного трафіку та його попередня фільтрація. На цьому рівні виконується виділення релевантних мережних подій, агрегація пакетів у з’єднання та формування структурованих записів. Метою етапу є перетворення неструктурованого потоку пакетів у впорядковану множину об’єктів аналізу.

Другий етап передбачає формування простору ознак. Кожне мережне з’єднання описується набором параметрів, що характеризують його поведінкові та протокольні властивості. Виконується кодування дискретних атрибутів і нормалізація числових значень. Результатом є уніфіковане векторне представлення даних.

Третій етап пов’язаний із трансформацією вхідного представлення у вибрану аналітичну область. Цей етап передбачає формування альтернативного частотно-часового подання мережного з’єднання. Перетворення виконується з метою отримання додаткової структурної інформації, що може бути використана моделлю класифікації.

Четвертий етап реалізує обчислення виходу моделі. На основі підготовленого представлення даних визначається прогноз належності з’єднання до одного з класів. У бінарній постановці формується оцінка ймовірності належності до класу атаки та виконується прийняття рішення відповідно до заданого порогового механізму.

П’ятий етап полягає в інтерпретації результату та передачі його до підсистеми реагування. У разі виявлення КА може бути сформовано повідомлення про інцидент або активовано відповідні заходи захисту. Результати класифікації також можуть накопичуватися для подальшого аналізу та оновлення параметрів моделі.

Таким чином, функціональна схема процесу обробки відображає логічно впорядковану послідовність дій, тобто від отримання мережних даних до формування рішення та його використання у системі виявлення КА. Такий опис створює основу для подальшої деталізації моделі класифікатора.

Система виявлення КА розглядається у контексті централізованої організації процесу аналізу, що передбачає зосередження функцій обробки, зберігання даних та оновлення моделі у єдиному аналітичному центрі. Такий підхід є характерним для корпоративних мереж, де необхідно забезпечити узгодженість політик безпеки та єдині критерії прийняття рішень.

За централізованої організації процесу мережні вузли або локальні сенсори виконують функції збору та первинної обробки трафіку. Сформовані структуровані записи передаються до центрального обчислювального модуля, у якому здійснюється формування ознак, їх перетворення у частотно-часове представлення та виконання класифікації. Таким чином, обчислювально складні процедури зосереджені в одному логічному центрі.

Централізована модель дозволяє забезпечити єдину версію алгоритму класифікації та уніфіковані параметри моделі для всієї мережі. Це спрощує контроль якості роботи системи, процедури оновлення параметрів та аналіз накопичених результатів. Наявність центрального сховища даних створює умови для повторного навчання моделі на розширених вибірках без зміни логіки

функціонування периферійних компонентів. Перевагою централізованої організації аналізу є можливість реалізації комплексного аналізу подій, що надходять з різних сегментів мережі. Агрегування інформації дозволяє враховувати взаємозв’язки між подіями та підвищувати узгодженість прийнятих рішень. Крім того, централізований підхід спрощує інтеграцію із системами журналювання, управління інцидентами та іншими компонентами. Водночас централізована модель потребує врахування навантаження на обчислювальні ресурси та канали передачі даних. Концентрація аналітичних функцій в одному центрі вимагає відповідної масштабованості та забезпечення надійності функціонування.

Таким чином, централізована модель процесу виявлення КА забезпечує єдність алгоритмічної логіки, узгодженість параметрів та можливість централізованого оновлення. Такий підхід відповідає вимогам КМ та створює основу для впровадження розроблюваних методів підвищення точності виявлення КА.

У більшості досліджень, присвячених системам виявлення вторгнень, мережний трафік розглядається у вигляді статичного вектору ознак, що безпосередньо подається на вхід класифікатора. Такий підхід є зручним з точки зору реалізації та дозволяє застосовувати широкий спектр алгоритмів машинного навчання. Водночас векторне представлення не завжди відображає внутрішні структурні залежності між ознаками та їх можливі коливальні або повторювані шаблони. Альтернативою класичному підходу є розгляд мережного з’єднання як сигналу, що може бути проаналізований у частотно-часовій області. Подібна ідея використовується в інших галузях, де складні багатовимірні дані перетворюються у сигнал із подальшим спектральним аналізом для виявлення прихованих закономірностей. Частотно-часове представлення дозволяє розглядати не лише значення окремих ознак, а й їх узгоджену структуру в межах єдиного аналітичного простору.

Частотно-часове представлення розглядатимемо як спосіб побудови альтернативного простору ознак, у якому можуть бути виявлені додаткові закономірності, які недоступні при безпосередньому аналізі початкового

векторного подання. Такий підхід створює основу для подальшого опису перетворення та формування моделі класифікації у новій області представлення.

Встановимо можливості та обмеження класичного векторного представлення мережного трафіку. Класичний підхід до побудови систем виявлення КА передбачає подання мережного з'єднання у вигляді статичного вектору ознак фіксованої розмірності. Такий підхід є зручним для реалізації та забезпечує можливість застосування широкого спектра алгоритмів машинного навчання. Водночас він має низку концептуальних обмежень, що впливають на здатність моделі виявляти складні або слабко виражені аномалії.

По-перше, векторне подання розглядає ознаки як незалежні або слабо пов'язані компоненти, тоді як у реальному мережному середовищі між параметрами з'єднання можуть існувати структурні залежності. Статичний вектор не відображає внутрішню організацію ознак і не дозволяє явно моделювати їх взаємодію в альтернативній області аналізу.

По-друге, векторне представлення не враховує можливі шаблони узгоджених змін між групами ознак. У випадках, коли аномальна поведінка проявляється не через екстремальні значення окремих параметрів, а через їх специфічну комбінацію, традиційна модель може мати обмежену чутливість.

По-третє, у високорозмірному просторі ознак виникає проблема надлишковості та корельованості параметрів. Наявність великої кількості взаємопов'язаних ознак може ускладнювати навчання моделі та знижувати її узагальнювальну здатність. За відсутності спеціальних механізмів перетворення даних модель змушена самостійно виділяти релевантні закономірності, що підвищує вимоги до обсягу навчальної вибірки.

Крім того, векторний підхід не передбачає явної інтерпретації ознак у новій аналітичній області. У той час як у частотно-часовому представленні можна аналізувати розподіл енергії сигналу або характерні спектральні компоненти, а у звичайному векторному просторі такі характеристики не визначаються без додаткових перетворень.

Таким чином, класичне векторне подання мережного трафіку, попри його універсальність і поширеність, має обмеження, пов’язані з відсутністю структурного аналізу взаємозв’язків між ознаками та складністю виявлення прихованих закономірностей. Це обґрунтовує доцільність розгляду альтернативного частотно-часового представлення, що може розширити аналітичні можливості моделі виявлення атак.

Подолання зазначених обмежень пов’язане з переходом до частотно-часового простору як розширеної області представлення мережного трафіку.

Частотно-часовий простір будемо розглядати як альтернативну та розширену область представлення мережного трафіку, що дає змогу враховувати внутрішню структуру ознак у новій аналітичній інтерпретації. На відміну від класичного векторного підходу, де кожне з’єднання подається як набір незалежних параметрів, у частотно-часовій області об’єкт аналізу інтерпретується як сигнал, характеристики якого можуть бути досліджені за допомогою спектральних методів.

Векторний опис мережного з’єднання у межах розроблюваного підходу задамо через операторне перетворення так:

$$T: X \to Z, \tag{2.1}$$

де X – простір векторів ознак мережних з’єднань; Z – частотно-часовий простір представлення; $x \in X \subseteq R^d$ – вектор ознак мережного з’єднання; R^d – d-вимірний простір дійсних чисел; d – кількість ознак, що характеризують одне мережне з’єднання.

Таке відображення, яке задане за формулою (2.1), формалізує перехід від класичного векторного подання до розширеної області аналізу, у якій можуть бути виявлені додаткові структурні закономірності між ознаками. Після введення оператора перетворення доцільно подати загальний аналітичний вигляд переходу сигналу у частотно-часову область.

Ідея полягає у тому, що компоненти вектору x можуть бути інтерпретовані як параметри сигналу, який формується відповідно до заданого правила

відображення. Отриманий сигнал буде аналізуватись за допомогою перетворення у частотну або частотно-часову область, що дасть змогу отримати двовимірне подання у вигляді спектральної структури.

Загальний аналітичний перехід сигналу у частотно-часову область задамо так:

$$Z(\omega,\tau) = \int_{-\infty}^{\infty} s(t)\, w(t-\tau)\; e^{-j\omega t}\, dt, \qquad (2.2)$$

де s(t) – сигнал, сформований на основі параметрів мережного з'єднання; w(t-τ) – віконна функція, зсунута на величину τ; τ – часовий зсув вікна; ω – кутова частота; j – уявна одиниця; t – змінна часу інтегрування; межі інтегрування задають розгляд сигналу на всій часовій осі [−∞; +∞].

Подане співвідношення задає локалізований спектральний аналіз сигналу, за якого частотні характеристики досліджуються у прив'язці до окремих часових позицій. Це створює основу для порівняння з класичним частотним представленням, у якому часову локалізацію не враховано.

Частотно-часовий простір має більшу розмірність порівняно з початковим векторним поданням, однак ця надлишковість супроводжується появою структурованої інформації. У спектральному представленні можуть проявлятися закономірності, пов'язані з узгодженими змінами ознак, які не є очевидними у вихідному просторі X.

Частотно-часовий простір не замінює початковий вектор ознак, а розширює його аналітичні можливості шляхом переходу до іншої області представлення. Таким чином, оператор T виконує функцію побудови альтернативного опису об'єкта, зберігаючи при цьому інформаційний зміст вихідних даних.

Отже, частотно-часовий простір виступає як розширена область представлення мережного трафіку, що створює передумови для побудови моделей класифікації, які здатні враховувати структурні взаємозв'язки між ознаками та потенційно підвищувати точність виявлення КА.

Використання спектрального аналізу для дослідження складних багатовимірних даних ґрунтується на припущенні, що внутрішня структура об’єкта може бути більш інформативно представлена у частотній або частотно-часовій області, ніж у вихідному просторі параметрів. У різних галузях спектральні методи застосовуються для виявлення прихованих закономірностей, періодичних компонент та аномальних коливань, які не є очевидними у часовому або просторовому поданні.

Частотне подання сигналу задамо за допомогою перетворення Фур’є так:

$$S(\omega) = \int_{-\infty}^{\infty} s(t)\, e^{-j\omega t} dt, \qquad (2.3)$$

де $s(t)$ – сигнал, сформований на основі параметрів мережного з’єднання; ω – кутова частота; j – уявна одиниця; t – змінна часу інтегрування; межі інтегрування задають аналіз сигналу на всій часовій осі [−∞; +∞].

Наведене співвідношення описує класичне спектральне подання сигналу без часової локалізації та дає змогу оцінити розподіл його енергії за частотами. Це створює передумови для переходу до формалізації входу класифікаційної моделі у розширеному просторі представлення.

У задачах аналізу реальних процесів часто використовується локалізоване перетворення, що дозволяє досліджувати зміну частотних характеристик у часі. Такий підхід реалізується за допомогою короткочасного перетворення Фур’є, яке поєднує часову та частотну локалізацію. Це створює двовимірне подання, у якому кожна точка відображає інтенсивність певної частотної компоненти у визначений момент часу.

Теоретичною передумовою застосування спектрального аналізу у задачі виявлення КА є можливість інтерпретації вектору ознак як параметрів сигналу. У такому випадку спектральне перетворення дозволяє розглядати узгоджені зміни між ознаками як гармонічні компоненти або як аномальні збурення у відповідному спектрі. Якщо аномальна поведінка мережного з’єднання проявляється у

специфічній структурі взаємозв’язків між параметрами, то вона може бути відображена у вигляді характерних частотних шаблонів.

Крім того, спектральний аналіз виконує роль певної форми перетворення простору ознак, що може сприяти зменшенню впливу шумових компонент. Перехід у частотну область дозволяє відокремлювати регулярні складові сигналу від випадкових збурень, що теоретично зменшує чутливість моделі до неінформативних змін параметрів.

Таким чином, використання спектрального аналізу у задачі виявлення КА надає можливість розкладання складної багатовимірної структури на сукупність інтерпретованих компонент. Це створює передумови для побудови моделей, що аналізують мережний трафік у розширеній частотно-часовій області та враховують структурні закономірності, які недоступні у класичному векторному поданні.

Додаткове обґрунтування цього підходу дає аналогія із задачами виявлення аномалій в інших предметних областях, де спектральні методи вже продемонстрували свою ефективність.

Застосування перетворення Фур’є у задачах аналізу мережного трафіку має аналогії з іншими галузями, де спектральний аналіз використовується для виявлення прихованих або слабко виражених аномалій. У фізичних, інженерних та геофізичних системах перехід у частотну область дозволяє ідентифікувати характерні компоненти сигналу, що відповідають певним процесам, навіть якщо у часовому поданні вони не є очевидними.

Зокрема, спектральний аналіз широко застосовувався для дослідження сейсмічних сигналів з метою виявлення аномальних коливань. У таких задачах часовий сигнал перетворюється у частотну область, де можна виділити характерні піки або зміни розподілу енергії, що відповідають певним подіям. Подібний підхід дозволяє відокремити регулярні фонові процеси від нестандартних збурень.

Аналогія з мережею полягає в тому, що мережне з’єднання, яке представлене у вигляді набору параметрів, може бути інтерпретоване як сигнал із певною внутрішньою структурою. У класичному векторному поданні аномалія проявляється як відхилення окремих значень або їх комбінацій. У частотній області

це відхилення може відповідати зміні спектрального розподілу, появі нових компонент або зсуву інтенсивності існуючих складових.

Оскільки $s(t)$ – сигнал, сформований на основі параметрів мережного з'єднання, то його частотне представлення $S(\omega)$, що визначається перетворенням Фур'є (2.3), дозволяє оцінити внесок різних частотних складових. Якщо структура взаємозв'язків між параметрами змінюється внаслідок аномальної поведінки, це може призвести до змін у спектрі сигналу. Таким чином, аномалія інтерпретується як відхилення спектральних характеристик від типового розподілу.

Крім того, застосування локалізованого спектрального аналізу дає змогу досліджувати зміну частотних компонент у межах окремих ділянок сигналу. Це особливо актуально для випадків, коли аномальна поведінка проявляється не глобально, а у певній частині структури параметрів.

Отже, аналогія використання перетворення Фур'є у задачах виявлення аномалій полягає в переході від безпосереднього аналізу вихідних значень до дослідження їх спектральної структури. Такий підхід дозволяє виявляти закономірності, що не є очевидними у початковому просторі ознак, і є основою для застосування частотно-часового представлення у системах виявлення КА.

2.2 Модель класифікації

Після визначення структури даних та обґрунтування використання частотно-часового представлення необхідно формалізувати модель класифікації, яка виконує прийняття рішення щодо належності мережного з'єднання до класу нормальної активності або атаки.

Після перетворення вихідного векторного опису мережного з'єднання вхід моделі визначимо так:

$$z = T(x), \tag{2.4}$$

де T – відображення у частотно-часову область; x – вектор ознак мережного з'єднання, причому $x \in X \subseteq R^d$; $z \in Z$; X – простір векторних описів мережних

з’єднань; Z – частотно-часовий простір представлення; d – кількість ознак одного мережного з’єднання.

Таке подання фіксує результат перетворення вихідних ознак у форму, яка придатна для подальшого класифікаційного аналізу. Після визначення вхідного об’єкта природно задати саму параметризовану функцію моделі, яка формує оцінку належності з’єднання до класу атаки.

Параметризоване відображення моделі класифікації задамо так:

$$f(z;\theta): Z \to [0,1], \tag{2.5}$$

де z – елемент частотно-часового простору представлення; θ – вектор параметрів моделі, що налаштовуються у процесі навчання; Z – простір представлення вхідних даних після перетворення; [0; 1] – інтервал значень виходу моделі, який інтерпретується як оцінка належності з’єднання до класу атаки.

Таке відображення задає неперервну оцінку, а не відразу дискретне рішення. Саме тому для переходу до кінцевого класу необхідно окремо ввести порогове правило прийняття рішення.

У бінарній постановці задачі перехід від числової оцінки до дискретної мітки задамо пороговим правилом так:

$$\hat{y} = \begin{cases} 1, & \text{якщо } f(z;\theta) \geq \tau, \\ 0, & \text{якщо } f(z;\theta) < \tau, \end{cases} \tag{2.6}$$

де $\hat{y}$=1 – рішення про належність з’єднання до класу атаки; $\hat{y}$=0 – рішення про належність з’єднання до класу нормальної активності; $f(z;\theta)$ – числова оцінка параметризованої моделі; z – елемент розширеного простору представлення; θ – вектор параметрів моделі; $\tau \in (0;1)$ – порогове значення, яке відокремлює рішення про атаку від рішення про нормальний стан; межі [0; 1] зумовлені тим, що вихід моделі інтерпретується як оцінка ймовірності.

Подане правило завершує процедуру прийняття рішення для окремого з’єднання на основі неперервного виходу моделі. Разом із тим для налаштування параметрів θ необхідно формально задати навчальну вибірку, на якій виконується навчання.

Налаштування параметрів моделі подамо як задачу мінімізації критерію якості на навчальних прикладах:

$$D_{\text{train}} = \{(z_i, y_i)\}_{i=1}^{N}. \tag{2.7}$$

де (z_i, y_i) – і-й навчальний приклад; z_i – елемент простору представлення, сформований для і-го мережного з'єднання; y_i – істинна класова мітка для і-го прикладу; і – індекс навчального прикладу, причому і=1, 2,…, N; N – загальна кількість прикладів у навчальній вибірці; $z_i \in$ Z; $y_i \in \{0,1\}$ у бінарній постановці задачі.

Таке означення задає структуру даних, на якій здійснюється налаштування параметрів моделі. Після цього можна записати функцію втрат, яка кількісно характеризує якість класифікації на навчальній вибірці.

Для бінарної класифікації використаємо логістичний критерій оптимізації, який задамо так:

$$L(\theta) = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log f(z_i;\theta) + (1-y_i)\log(1-f(z_i;\theta))]. \tag{2.8}$$

де θ – вектор параметрів моделі; N – кількість прикладів у навчальній вибірці; і – індекс навчального прикладу, причому і=1,2,…,N; y_i – істинна мітка класу і-го прикладу; z_i – вхідне представлення і-го прикладу; $f(z_i;\theta)$ – оцінка належності і-го прикладу до класу атаки; log(·) – натуральний логарифм.

Подана функція втрат кількісно відображає ступінь невідповідності між прогнозами моделі та істинними мітками класів. Тому задача навчання природно зводиться до пошуку такого набору параметрів, для якого значення цієї функції є мінімальним.

Пошук параметрів, що мінімізують критерій оптимізації, задамо так:

$$\theta^* = \arg\min_{\theta} L(\theta). \tag{2.9}$$

де $\arg\min_{\theta} L(\theta)$ – оператор пошуку такого значення θ, за якого функція втрат L(θ) набуває мінімального значення; $L(\theta)$ – критерій оптимізації, обчислений на навчальних прикладах; θ – допустимий вектор параметрів моделі.

Таке подання завершує формалізацію процесу навчання класифікаційної моделі як задачі оптимізації. Далі доцільно ще раз подати узагальнену функцію прийняття рішення як базовий елемент розроблюваної моделі.

Важливою особливістю моделі є її здатність формувати неперервну оцінку ймовірності, що створює передумови для подальшої формалізації зони невизначеності прогнозу. Наявність інтервалу значень, у якому модель демонструє знижену впевненість, може бути використана для застосування додаткових процедур прийняття рішення.

Таким чином, модель класифікації визначено параметризованою функцією у розширеному просторі представлення даних із пороговим механізмом прийняття рішення та оптимізацією на основі функції втрат. Такий формалізований опис створює основу для подальшого аналізу помилок, визначення критеріїв точності та розроблення методів підвищення ефективності системи виявлення комп'ютерних атак.

Для формалізації процесу класифікації насамперед доцільно задати узагальнену функцію прийняття рішення. У системі виявлення комп'ютерних атак функція прийняття рішення визначає правило віднесення мережного з'єднання до одного з класів на основі сформованого представлення даних. Розглянемо параметризовану модель, яка формує неперервну оцінку належності об'єкта до класу атаки з подальшим застосуванням порогового механізму.

Отримане в результаті навчання параметризоване відображення $f(z; \theta): Z \rightarrow [0; 1]$, визначене формулою (2.5), формує неперервну оцінку належності мережного з'єднання до класу КА. Перехід від цієї оцінки до дискретної мітки класу здійснюється відповідно до порогового правила, яке задане за формулою (2.6). Отже, вихід моделі та кінцеве класифікаційне рішення є взаємопов'язаними, але різними складовими процесу класифікації.

У загальному випадку функція $f(z;\theta)$ може бути реалізована логістичною регресією, багатошаровою нейронною мережею або згортковою архітектурою. Для моделей із сигмоїдним вихідним шаром зв'язок між внутрішньою дискримінантною оцінкою та ймовірнісним виходом задамо так:

$$f(z;\theta) = \sigma(g(z;\theta)) = [1 + exp(-g(z;\theta))]^{-1} \approx P(Y = 1 \mid Z = z), \qquad (2.10)$$

де z ∈ Z – представлення мережного з'єднання після застосування оператора перетворення; Z – простір сформованих представлень; θ – вектор параметрів моделі, що визначається у процесі навчання; $g(z;\theta) \in$ R – внутрішня дискримінантна оцінка моделі; σ(·) – сигмоїдна функція, причому σ: R → (0; 1); Y ∈ {0, 1} – випадкова величина, що визначає клас мережного з'єднання, де Y = 1 відповідає класу комп'ютерних атак, а Y = 0 – класу нормальної мережної активності; $P(Y = 1 \mid Z = z)$ – умовна ймовірність належності мережного з'єднання до класу атаки.

Знак наближення у формулі (2.10) підкреслює, що вихід моделі є оцінкою умовної ймовірності, сформованою за навчальними даними, а не безпосередньо заданим істинним значенням. Оцінки f(z; θ), близькі до одиниці, свідчать про високу ймовірність належності з'єднання до класу атак, тоді як значення, близькі до нуля, відповідають високій імовірності належності до класу нормальної активності.

Дискретне рішення формується шляхом порівняння оцінки $f(z;\theta)$ з пороговим значенням τ відповідно до правила, яке задане за формулою (2.6). Порогове значення може бути задане заздалегідь або визначене на валідаційній вибірці з урахуванням вимог до системи виявлення КА. Зменшення τ, як правило, підвищує чутливість моделі до КА і зменшує кількість хибнонегативних рішень, але водночас може збільшувати кількість хибнопозитивних спрацювань. Збільшення τ має протилежний вплив. Отже, вибір порога визначає співвідношення між пропущеними атаками та помилковими сигналами тривоги [20, 21].

Значення виходу моделі, розташовані поблизу порога τ, характеризуються зниженою визначеністю класифікаційного рішення. Для кількісного оцінювання

цієї властивості введемо відстань між виходом моделі та базовим пороговим значенням так:

$$\delta(z) = | f(z;\theta) - \tau |, \quad (2.11)$$

де δ(z) – абсолютна відстань між виходом моделі та порогом прийняття рішення; f(z; θ) – оцінка належності мережного з'єднання до класу атаки; τ ∈ (0; 1) – базове порогове значення; |·| – операція визначення абсолютного значення.

Чим меншим є значення δ(z), тим ближче вихід моделі розташований до межі прийняття рішення і тим нижчою є визначеність прогнозу. Навпаки, збільшення δ(z) свідчить про віддалення оцінки від порога та вищу визначеність класифікаційного рішення. Таке подання створює основу для подальшого визначення меж зони невизначеності та застосування додаткових процедур аналізу прикордонних випадків.

Узагальнено, оптимальне значення τ може визначатися на валідаційній вибірці шляхом аналізу відповідних метрик якості або з урахуванням вартості різних типів помилок. Формально оптимальний поріг задамо так:

$$\tau^* = \arg \min_{\tau \in (0,1)} R(\tau), \quad (2.12)$$

де $\arg \min_{\tau \in (0,1)}$ – оператор пошуку такого значення τз інтервалу $(0, 1)$, за якого функція $R(\tau)$набуває мінімального значення; $R(\tau)$ – функція ризику, що враховує співвідношення між помилками першого та другого роду; τ – кандидатне порогове значення; інтервал $(0, 1)$ визначається природою ймовірнісного виходу моделі.

Таке подання дозволяє розглядати вибір порога не евристично, а як окрему оптимізаційну задачу. Після цього природно перейти до випадків, коли навіть за оптимального порога вихід моделі концентрується поблизу межі прийняття рішення та потребує спеціального формалізованого опису.

Таким чином, пороговий механізм є ключовим елементом моделі класифікації, оскільки саме він визначає баланс між чутливістю та специфічністю системи. Його налаштування має здійснюватися з урахуванням вимог до інформаційної безпеки та структури даних, що аналізуються.

Окремого формалізованого опису потребують випадки, коли вихід моделі зосереджується поблизу порогового значення, тобто належить до зони невизначеності прогнозу. У межах ймовірнісної моделі класифікації вихідне значення $f(z;\theta) \in [0,1]$ відображає ступінь впевненості моделі у належності мережного з'єднання до класу атаки. Якщо значення є близьким до 0 або 1, рішення є достатньо визначеним. Проте у випадках, коли оцінка знаходиться поблизу порогового значення τ, то рівень впевненості моделі знижується. Це створює підстави для формалізації зони невизначеності прогнозу.

Параметри зони невизначеності прогнозу визначимо нерівністю так:

$$0 < \tau_{\mathrm{L}} < \tau < \tau_{\mathrm{U}} < 1. \qquad (2.13)$$

де τ_L – нижня межа інтервалу невизначеності; τ – базове порогове значення класифікації; τ_U – верхня межа інтервалу невизначеності; $\tau_L, \tau, \tau_U \in (0,1)$; нерівності $\tau_L < \tau < \tau_U$ задають розташування базового порога всередині інтервалу невизначеності; умови $0 < \tau_L$та $\tau_U < 1$ забезпечують належність усіх порогових параметрів допустимому інтервалу значень виходу моделі.

Таке співвідношення задає формальні межі області, у якій модель демонструє знижену впевненість у своєму прогнозі. На цій основі вже можна визначити саму множину об'єктів, що належать до зони невизначеності, та обґрунтувати потребу в додатковому механізмі уточнення рішень для прикордонних випадків.

Множину об'єктів, для яких вихід моделі потрапляє до зони невизначеності, визначимо так:

$$U = \{z \in Z \mid \tau_{\mathrm{L}} \le f(z;\theta) \le \tau_{\mathrm{U}}\}. \qquad (2.14)$$

де Z – простір представлення даних після перетворення початкових ознак; $f(z;\theta)$ – вихід моделі класифікації для об'єкта z за вектору параметрів θ; θ – вектор параметрів моделі; τ_L – нижня межа інтервалу невизначеності; τ_U – верхня межа інтервалу невизначеності; $\tau_L, \tau_U \in (0,1)$; умова $\tau_L \le f(z;\theta) \le \tau_U$означає, що оцінка моделі розташована всередині інтервалу невизначеності.

Таке означення формалізує підмножину прикладів, для яких базова модель не забезпечує достатньої впевненості у своєму прогнозі. Саме ця підмножина

надалі становить основу для застосування додаткових процедур аналізу або селективного перевизначення рішень.

Зону невизначеності через відстань до базового порогового значення задамо так:

$$U = \{z \in Z \mid\mid f(z;\theta) - \tau \mid \leq \delta\}, \quad (2.15)$$

де z – представлення мережного з'єднання у просторі Z; $f(z;\theta)$ – вихід моделі класифікації; τ – базове порогове значення прийняття рішення, причому $\tau \in (0,1)$; δ – параметр ширини інтервалу невизначеності, причому $\delta > 0$; $\mid f(z;\theta) - \tau \mid$ – абсолютна відстань між виходом моделі та базовим порогом; умова $\mid f(z;\theta) - \tau \mid \leq \delta$означає, що прогноз моделі є достатньо близьким до межі прийняття рішення. Це значення не є остаточною міткою класу, а є неперервною оцінкою належності з'єднання до класу атаки; саме його відстань до порогового значення τ використовується для визначення належності прогнозу до зони невизначеності.

Подане представлення є еквівалентним інтервальному опису зони невизначеності, але дає змогу безпосередньо контролювати її ширину через параметр δ. Це зручно для подальшого аналізу практичних наслідків помилкових рішень, які виникають саме поблизу порога класифікації.

Введення зони невизначеності не змінює базову структуру класифікатора, проте розширює механізм прийняття рішення, дозволяючи більш гнучко керувати співвідношенням між помилками першого та другого роду. Це необхідно для розроблення методів, що використовують цю формалізацію для підвищення точності виявлення КА у критичних хибних порогових випадках.

Важливим наслідком уведення порогового механізму та зони невизначеності є потреба у формалізованому описі помилок першого та другого роду, оскільки саме вони визначають практичну цінність класифікаційного рішення для системи інформаційної безпеки.

У задачі бінарної класифікації мережних з'єднань важливим є формалізований опис помилкових рішень, оскільки різні типи помилок мають різні наслідки для системи інформаційної безпеки. Нехай – істинна мітка класу, а –

результат класифікації, отриманий відповідно до правилай, яке задано за формулою (2.6).

Матриця невідповідностей може бути подана у вигляді множини чотирьох можливих випадків:

1) істинно негативне рішення (TN): $y = 0, \hat{y} = 0$;

2) істинно позитивне рішення (TP): $y = 1, \hat{y} = 1$;

3) хибнопозитивне рішення (FP): $y = 0, \hat{y} = 1$;

4) хибнонегативне рішення (FN): $y = 1, \hat{y} = 0$.

Помилка першого роду відповідає хибнопозитивному рішенню, коли нормальна активність класифікується як атака. Імовірність помилки першого роду визначимо як умовну ймовірність хибнопозитивного рішення так:

$$\alpha = P(\hat{Y} = 1 \mid Y = 0), \tag{2.16}$$

де $P(\hat{Y} = 1 \mid Y = 0)$ – умовна ймовірність того, що модель відносить об’єкт до класу атаки за умови, що його істинний клас є нормальним; Y – випадкова величина, що відповідає істинному класу мережного з’єднання; $\hat{Y}$ – випадкова величина, що відповідає прогнозованому класу; $Y = 0$ – подія належності з’єднання до класу нормальної активності; $\hat{Y} = 1$ – подія, за якої модель класифікує з’єднання як атаку.

Таке означення задає формальний опис хибного спрацювання системи, коли нормальна мережна активність помилково інтерпретується як КА. Для повного опису структури помилок бінарної класифікації необхідно також окремо подати імовірність пропуску атаки.

Імовірність помилки другого роду задамо так:

$$\beta = P(\hat{Y} = 0 \mid Y = 1). \tag{2.17}$$

де $P(\hat{Y} = 0 \mid Y = 1)$ – умовна ймовірність того, що модель відносить об’єкт до класу нормальної активності за умови, що його істинний клас відповідає атаці; Y – випадкова величина істинного класу; $\hat{Y}$ – випадкова величина прогнозованого

класу; $Y = 1$ – подія належності з'єднання до класу атаки; $\hat{Y} = 0$ – подія, за якої модель не виявляє атаку та класифікує її як нормальну активність.

Це співвідношення формалізує найбільш критичний для систем інформаційної безпеки тип помилки, що полягає в пропуску КА. Для кількісного аналізу всіх правильних і помилкових рішень доцільно перейти до матричного подання результатів класифікації.

Зв'язок між помилками першого та другого роду визначається вибором порогового значення τ. Зміна τ змінює співвідношення між цими ймовірностями, формуючи компроміс між чутливістю та специфічністю системи.

Таким чином, формалізація помилок першого та другого роду створює основу для подальшого визначення критеріїв точності моделі та побудови механізмів оптимізації, спрямованих на забезпечення збалансованого рівня безпеки та коректності функціонування системи виявлення КА.

2.3 Критерії оцінювання точності та узагальнювальної здатності

Базовим інструментом оцінювання точності бінарної моделі класифікації є матриця невідповідностей, яка відображає співвідношення між істинними мітками класів і прогнозами моделі.

Матриця невідповідностей є базовим інструментом оцінювання якості бінарної моделі класифікації. Вона відображає співвідношення між істинними мітками класів та прогнозами моделі й дозволяє формалізувати структуру правильних і помилкових рішень.

Нехай $y \in \{0,1\}$ – істинна мітка класу, а $\hat{y} \in \{0,1\}$ – результат класифікації. Матрицю невідповідностей для бінарної задачі класифікації подамо так:

$$\mathrm{M} = \begin{pmatrix} TN & FP \\ FN & TP \end{pmatrix}, \tag{2.18}$$

де TN – кількість істинно негативних рішень; FP – кількість хибнопозитивних рішень; FN – кількість хибнонегативних рішень; TP – кількість істинно позитивних рішень.

Таке матричне подання дає змогу компактно описати повну структуру результатів бінарної класифікації. На його основі далі можна формально визначити кожен компонент через індикаторну функцію та побудувати похідні метрики якості моделі.

Кількість істинно позитивних рішень задамо так:

$$TP = \sum_{i=1}^{N} \mathbb{I}(y_i = 1 \wedge \hat{y}_i = 1), \tag{2.19}$$

де $\sum_{i=1}^{N}$ – сумування за всіма прикладами вибірки; i – індекс прикладу, причому $i = 1,2,\dots,N$; N – загальна кількість прикладів у вибірці оцінювання; $I(\cdot)$ – індикаторна функція, що набуває значення 1, якщо логічна умова в її аргументі виконується, і 0 – в іншому випадку; y_i – істинна мітка класу i-го прикладу; $\hat{y}_i$ – прогнозована мітка класу i-го прикладу; умова $y_i = 1 \wedge \hat{y}_i = 1$ означає, що атака була правильно виявлена моделлю.

Таке означення виділяє ту частину атакувальних з'єднань, які система класифікувала коректно. Аналогічно далі визначаються правильні негативні та обидва типи помилкових рішень.

Кількість істинно негативних рішень задамо так:

$$TN = \sum_{i=1}^{N} \mathbb{I}(y_i = 0 \wedge \hat{y}_i = 0), \tag{2.20}$$

де i – індекс прикладу, причому $i = 1,2,\dots,N$; N – кількість прикладів у вибірці оцінювання; $I(\cdot)$ – індикаторна функція; y_i – істинна мітка класу i-го прикладу; $\hat{y}_i$ – прогнозована мітка класу i-го прикладу; умова $y_i = 0 \wedge \hat{y}_i = 0$ означає, що нормальна активність була правильно розпізнана моделлю.

Поданий вираз формалізує кількість коректно класифікованих нормальних з'єднань. Разом із попереднім означенням він задає сукупність правильних рішень моделі, але для повної картини необхідно також явно подати обидва типи помилок.

Кількість хибнопозитивних рішень визначимо так:

$$FP = \sum_{i=1}^{N} \mathbb{I}(y_i = 0 \wedge \hat{y}_i = 1), \quad (2.21)$$

де і – індекс прикладу, причому і=1,2,…,N; N – кількість прикладів у вибірці оцінювання; I(·) – індикаторна функція; уі – істинна мітка класу і-го прикладу; уі – прогнозована мітка класу і-го прикладу; умова уі=0∧уі=1 означає, що нормальна активність була помилково віднесена до класу КА.

Цей показник характеризує частку хибних спрацьовувань системи, тобто випадків, коли нормальну активність помилково віднесено до атакувальної. Для повної оцінки якості СВВ його необхідно аналізувати разом із часткою пропущених атак, що відповідають хибнонегативним рішенням.

Кількість хибнонегативних рішень задамо так:

$$FN = \sum_{i=1}^{N} \mathbb{I}(y_i = 1 \wedge \hat{y}_i = 0). \quad (2.22)$$

де i – індекс прикладу, причому $i = 1,2,\dots,N$; N – загальна кількість прикладів у вибірці оцінювання; $I(\cdot)$ – індикаторна функція; y_i – істинна мітка класу i-го прикладу; $\hat{y}_i$ – прогнозована мітка класу i-го прикладу; умова $y_i = 1 \wedge \hat{y}_i = 0$ означає, що КА не була виявлена моделлю та була помилково класифікована як нормальна активність.

Таке означення завершує формалізацію всіх чотирьох компонентів матриці невідповідностей. Матриця невідповідностей дозволяє отримати відомості про поведінки моделі на тестовій вибірці та є основою для обчислення похідних метрик якості. Вона також дає змогу аналізувати баланс між різними типами помилок, що є критично важливим у задачах виявлення КА.

У контексті систем інформаційної безпеки особливу увагу приділяють співвідношенню між FN та FP, оскільки пропуск атаки та хибне спрацювання мають різні наслідки для експлуатації системи. Формалізація матриці невідповідностей створює основу для подальшого визначення метрик точності, повноти та інших показників ефективності моделі. На основі матриці

невідповідностей визначаються кількісні показники якості моделі, що забезпечують комплексне оцінювання її поведінки. На основі матриці невідповідностей визначаються кількісні показники якості моделі класифікації. У задачі виявлення КА використання лише одного показника є недостатнім, оскільки різні метрики по-різному відображають поведінку моделі за умов дисбалансу класів та різної вартості помилок.

Показник загальної точності класифікації визначимо так:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \quad (2.23)$$

де TP – кількість істинно позитивних рішень; TN – кількість істинно негативних рішень; FP – кількість хибнопозитивних рішень; FN – кількість хибнонегативних рішень; чисельник $TP + TN$визначає кількість усіх правильних рішень моделі; знаменник $TP + TN + FP + FN$визначає загальну кількість прикладів, на яких виконувалося оцінювання.

Таке співвідношення задає базову узагальнену метрику якості, яка відображає частку коректно класифікованих об’єктів у вибірці. Водночас для задачі виявлення КА цього показника недостатньо, оскільки він не відображає структуру помилок і є чутливим до дисбалансу класів, тому далі логічно аналізувати також прецизійність, повноту, F1-міру, специфічність та інші похідні показники.

Точність позитивних рішень моделі визначимо часткою правильно виявлених КА серед усіх з’єднань, класифікованих як КА, так:

$$\text{Precision} = \frac{TP}{TP + FP}. \quad (2.24)$$

де TP – кількість істинно позитивних рішень, тобто атак, правильно виявлених моделлю; FP – кількість хибнопозитивних рішень, тобто нормальних з’єднань, помилково віднесених до класу атак; знаменник $TP + FP$ – загальна кількість з’єднань, які модель класифікувала як атаки.

Таке співвідношення відображає, наскільки надійними є позитивні рішення класифікатора. Водночас воно не показує, яку частку всіх реальних атак система справді виявляє, тому далі доцільно окремо ввести показник повноти.

Повноту виявлення КА або чутливість моделі задамо так:

$$\text{Recall} = \frac{TP}{TP + FN}. \tag{2.25}$$

де TP – кількість істинно позитивних рішень; FN – кількість хибнонегативних рішень, тобто атак, які модель не виявила і помилково класифікувала як нормальну активність; знаменник $TP + FN$ – загальна кількість реальних атак у вибірці оцінювання.

Дане подання характеризує здатність системи виявляти КА серед усіх справжніх атакувальних з'єднань. Оскільки для задачі виявлення КА важливими є і точність позитивних рішень, і повнота виявлення, надалі доцільно використати узагальнену міру, що враховує обидва показники одночасно.

Збалансовану оцінку якості моделі, що одночасно враховує точність і повноту, визначимо так:

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{2.26}$$

де Precision – точність позитивних рішень; Recall – повнота виявлення атак; чисельник $2 \cdot \text{Precision} \cdot \text{Recall}$задає подвоєний добуток двох базових показників; знаменник $\text{Precision} + \text{Recall}$ – сума точності та повноти.

Таке співвідношення задає гармонійне середнє між точністю та повнотою і є особливо корисним за умов дисбалансу класів. Разом із цим для задачі виявлення комп'ютерних атак важливо оцінювати не лише якість розпізнавання атак, а й здатність моделі правильно ідентифікувати нормальні з'єднання.

Специфічність моделі, яка характеризує правильність розпізнавання нормальної мережної активності, запишемо так:

$$\text{Specificity} = \frac{TN}{TN + FP}. \tag{2.27}$$

де TN – кількість істинно негативних рішень, тобто нормальних з'єднань, правильно класифікованих як нормальні; FP – кількість хибнопозитивних рішень; знаменник $TN + FP$ – загальна кількість реальних нормальних з'єднань у вибірці оцінювання.

Поданий показник відображає здатність моделі уникати хибних спрацювань щодо нормального трафіку. Після визначення базових метрик оцінювання якості доцільно перейти до формалізації структури даних, на яких здійснюється навчання та тестування моделі.

Для узагальненого аналізу здатності моделі відокремлювати класи використовується крива «чутливість–специфічність» та площа під нею. Вона відображає залежність між часткою істинно позитивних та хибнопозитивних рішень при зміні порогового значення.

Таким чином, метрики точності дозволяють комплексно оцінити якість моделі класифікації з урахуванням різних аспектів її поведінки. У задачах виявлення КА доцільно аналізувати сукупність показників, що характеризують як здатність моделі виявляти КА, так і рівень хибних спрацьовувань.

Використання лише показника загальної точності не дає підстав для об'єктивного висновку щодо ефективності системи виявлення КА.

Показник загальної точності є одним із найбільш поширених критеріїв оцінювання якості моделей класифікації. Його простота та інтуїтивна інтерпретація роблять його зручним для первинного аналізу результатів. Проте у задачі виявлення КА використання лише цього показника є недостатнім та може призводити до хибних висновків щодо ефективності моделі.

Основним обмеженням показника загальної точності є його чутливість до дисбалансу класів. У випадку, коли один із класів суттєво переважає, модель може досягати високого значення загальної точності навіть за умови незадовільного виявлення менш представленого класу. Формально показник загальної точності визначається за формулою (2.23). Якщо частка нормальних з'єднань у вибірці значно перевищує частку КА, то модель, яка переважно прогнозує нормальну

активність, може мати високе значення загальної точності, при цьому демонструючи низьку повноту для класу атак.

Другим обмеженням є відсутність інформації про структуру помилок. Значення загальної точності не дозволяє оцінити співвідношення між хибнопозитивними та хибнонегативними рішеннями. У системах інформаційної безпеки ці типи помилок мають різні наслідки, тому їх аналіз повинен виконуватися окремо.

Третім аспектом є залежність загальної точності від вибору порогового значення. Зміна порога τ може призвести до різної конфігурації помилок без істотної зміни загальної точності. Таким чином, однакове значення загальної точності може відповідати різним характеристикам поведінки моделі. Існує ризик завищеної інтерпретації результатів при використанні лише цього показника. Високе значення точності не гарантує високої здатності моделі до виявлення КА, особливо у невпевнених або рідкісних випадках.

Таким чином, для коректного оцінювання якості моделі виявлення КА необхідно використовувати сукупність метрик, що враховують структуру помилок і поведінку моделі щодо кожного класу окремо. Показник точності може застосовуватися як загальна характеристика, однак його використання без аналізу додаткових критеріїв є недостатнім для обґрунтованих висновків щодо ефективності системи.

2.4 Висновки до другого розділу

Сформовано теоретичну основу процесу виявлення комп’ютерних атак у корпоративних мережах. Побудовано узагальнену модель, що поєднує збір і попередню обробку мережного трафіку, формування вхідного представлення, класифікацію та прийняття рішення. Обґрунтовано перехід від традиційного векторного опису мережних з’єднань до частотно-часового подання, яке створює передумови для виявлення просторово-частотних закономірностей за допомогою двовимірних згорткових нейронних мереж. Формалізовано параметризовану

функцію класифікації, її ймовірнісну інтерпретацію, порогове правило прийняття рішення та зону невизначеності прогнозу. Визначено помилки першого і другого роду та систему метрик, що враховує не лише загальну точність, а й прецизійність, повноту, F1-міру, FPR і FNR. Показано, що дисбаланс класів, недостатня представленість рідкісних атак і нестабільність прикордонних прогнозів обмежують можливості базової моделі. Отримані результати обґрунтовують необхідність розроблення методів частотно-часового перетворення трафіку, сигнатурозбережного адаптивного балансування навчальної вибірки та селективного перевизначення рішень у зоні невизначеності.

Основні результати розділу опубліковані у працях [127, 130, 132].

РОЗДІЛ 3
МЕТОДИ ПІДВИЩЕННЯ ТОЧНОСТІ ВИЯВЛЕННЯ КОМП'ЮТЕРНИХ АТАК НА ОСНОВІ ЧАСТОТНО-ЧАСОВОГО ПРЕДСТАВЛЕННЯ МЕРЕЖНОГО ТРАФІКУ

3.1 Метод перетворення трафіку у хвильовий сигнал

У більшості СВКА мережний трафік подається у вигляді табличних векторів ознак, що характеризують параметри з'єднання, статистичні характеристики потоків та службові атрибути протоколів. Таке представлення є зручним для класичних алгоритмів машинного навчання, оскільки дозволяє працювати у межах багатовимірного евклідового простору ознак. Разом із тим воно не враховує можливість інтерпретації взаємозв'язків між компонентами вектору у вигляді сигналу з часовою структурою, що обмежує використання інструментів аналізу сигналів та частотно-часових методів обробки даних.

Перетворення вектору ознак у хвильову форму дозволяє розширити спосіб подання даних шляхом відображення статичного опису мережного з'єднання у динамічний сигнал із заданими параметрами. Такий підхід створює можливість переходу до частотно-часового аналізу та застосування методів обробки спектрограм, що традиційно використовуються в задачах цифрової обробки сигналів. При цьому відображення будується як формально визначений оператор, що кожному нормалізованому вектору ознак ставить у відповідність дискретний сигнал фіксованої довжини, зберігаючи однозначну відповідність між початковими значеннями компонент і параметрами сформованого сигналу.

Метод перетворення мережного трафіку у хвильовий сигнал ґрунтується на лінійному відображенні нормалізованих ознак у параметри гармонічних коливань із фіксованою тривалістю запису та сталою частотою дискретизації. Кожна ознака впливає на частоту та амплітуду відповідного часово локалізованого фрагмента сигналу, а повний запис формується шляхом послідовної сегментації часової осі.

Такий механізм не передбачає застосування нелінійних перетворень типу логарифмування або експоненційного масштабування на етапі формування сигналу, а також не використовує групування чи динамічного переставлення ознак, що забезпечує структурну відтворюваність процедури відображення.

Отриманий хвильовий сигнал розглядається як проміжне представлення мережного трафіку, яке надалі підлягає короткочасному перетворенню Фур'є для формування частотно-часової матриці. Спектрограма, що утворюється внаслідок цього перетворення, використовується як вхід для моделей глибокого навчання, орієнтованих на аналіз двовимірних структур. Таким чином, табличний простір ознак відображається у простір частотно-часових представлень, придатних для обробки згортковими нейронними мережами.

Розглядається не модель класифікації, а метод формального перетворення табличного представлення мережного трафіку у хвильову форму та подальше частотно-часове подання. Опис методу обмежується строгим визначенням правил відображення ознак у параметри сигналу, введенням відповідних операторів перетворення та аналізом їх властивостей. Питання навчання моделі класифікації та оцінювання якості виявлення атак розглядаються окремо у наступних розділах дисертації.

Метод перетворення мережного трафіку у хвильовий сигнал подамо упорядкованою послідовністю етапів, що визначають перехід від початкового опису мережного з'єднання до двовимірного частотно-часового представлення. Така структуризація забезпечить відтворюваність процедури та розмежує етапи попередньої обробки, ІКМ-перетворення та спектрального аналізу, а також дасть змогу однозначно встановити вхідні дані, операції та результат кожного кроку методу.

Етап 1. Формування початкового опису мережного з'єднання. На цьому етапі для кожного зразка мережного трафіку формуємо набір числових і категоріальних атрибутів, що характеризують параметри з'єднання, службові ознаки протоколів, статистичні характеристики потоку та інші доступні параметри. Результатом етапу

стане початковий вектор ознак, який зберігає структуру вихідного табличного опису мережного трафіку.

Етап 2. Кодування категоріальних атрибутів і узгодження простору ознак. Категоріальні параметри перетворюємо у числову форму за допомогою унітарного кодування, після чого виконуємо узгодження розмірності навчальної, валідаційної та тестової вибірок. У результаті формуємо єдиний впорядкований простір ознак сталої розмірності, придатний для подальших перетворень без втрати відповідності між компонентами вектору та їх змістом.

Етап 3. Нормалізація компонент вектору ознак. Числові компоненти приводимо до фіксованого інтервалу значень із використанням параметрів масштабування, обчислених лише на навчальній вибірці. Такий підхід унеможливлює витік інформації з валідаційної або тестової множини та забезпечує коректне подальше відображення ознак у параметри хвильового сигналу.

Етап 4. Відображення нормалізованих ознак у параметри гармонічних фрагментів. Для кожної компоненти нормалізованого вектору визначаємо амплітуду, частоту та часову позицію відповідного фрагмента сигналу. Частоту формуємо шляхом лінійного відображення значення ознаки у заданий частотний діапазон. Амплітуда відображає інтенсивність відповідної компоненти, а часову позицію визначимо фіксованим порядком ознак у векторі.

Етап 5. Побудова дискретного хвильового сигналу. На основі визначених параметрів формуємо послідовність часово локалізованих гармонічних сегментів однакової тривалості. Сукупність таких сегментів утворює дискретний сигнал фіксованої довжини за сталої частоти дискретизації, причому кожен сегмент відповідає окремій компоненті вхідного вектору ознак.

Етап 6. Формування частотно-часового подання. До отриманого дискретного сигналу застосовуємо короткочасне перетворення Фур’є, у результаті чого утворюється комплексна частотно-часова матриця. Подальше обчислення модуля спектра та логарифмічне масштабування амплітуд забезпечують побудову дійсної матриці спектрограми.

Етап 7. Підготовка спектрограми до подальшої класифікації. Сформовану спектрограму подаємо двовимірним вхідним представленням мережного з’єднання для моделей глибинного навчання, зокрема двовимірних згорткових нейронних мереж. На цьому етапі завершуємо виконання частини методу з перетворення трафіку, тоді як навчання класифікаційної моделі та експериментальне оцінювання її якості розглядатимемо окремо.

Отже, формалізовано метод побудови хвильового та частотно-часового представлення мережного трафіку. Його результатом є спектрограма мережного з’єднання, яка зберігає впорядкованість початкових ознак через часову сегментацію сигналу та відображає значення компонент у частотно-амплітудній структурі, яка придатна для подальшого нейромережевого аналізу.

Побудову методу доцільно розпочати з формалізації простору ознак мережного з’єднання та визначення вимог до його подальшого перетворення. Нехай кожен мережний сеанс або з’єднання описується множиною початкових атрибутів, що включають числові характеристики, бінарні індикатори та категоріальні параметри протоколів і служб. Після попередньої обробки дані подаються у вигляді вектора ознак фіксованої розмірності. Попередня обробка включає кодування категоріальних змінних, вирівнювання структури даних та формування єдиного впорядкованого набору компонент.

Простір ознак мережних з’єднань після попередньої обробки визначимо так:

$$X \subset \mathrm{R}^{d}, \tag{3.1}$$

де X – простір ознак мережних з’єднань після попередньої обробки; R^d – d-вимірний евклідів простір дійсних чисел; d – розмірність вектору ознак після кодування категоріальних змінних, вирівнювання структури даних і приведення всіх атрибутів до уніфікованого числового подання.

Таке співвідношення формалізує область, у якій надалі описується кожне мережне з’єднання. Після введення простору ознак необхідно явно задати його розмірність, оскільки саме вона визначає кількість компонент вхідного вектору та впливає на подальшу процедуру побудови хвильового сигналу.

Після введення простору ознак необхідно формально визначити його розмірність, оскільки саме величина d задає кількість компонент вхідного вектора і, відповідно, впливає на подальшу процедуру побудови хвильового сигналу. Такий запис дає змогу явно врахувати внесок числових ознак і категоріальних атрибутів після їх кодування в загальну структуру подання даних. Якщо початковий набір містить d_{num} числових ознак та d_{cat} категоріальних атрибутів, а кожен j-й категоріальний атрибут після one-hot кодування породжує k_j двійкових компонент, тоді розмірність простору ознак визначимо так:

$$d = d_{\text{num}} + \sum_{j=1}^{d_{\text{cat}}} k_j, \tag{3.2}$$

де d – загальна розмірність вектора ознак після попередньої обробки; d_{num} – кількість числових ознак; d_{cat} – кількість категоріальних атрибутів; k_j – кількість двійкових компонент, утворених one-hot кодуванням j-го категоріального атрибута; j – індекс категоріального атрибута, причому j = 1, 2, …, d_{cat}.

Категоріальні атрибути перетворюються у двійкове представлення методом one-hot кодування. Поданий вираз дозволяє явно врахувати внесок кожного типу ознак у загальну структуру вхідного подання. Для того щоб формально описати перетворення категоріального атрибута у двійковий вектор, доцільно окремо задати вигляд one-hot представлення.

One-hot представлення категоріального атрибута визначимо так:

$$e_c = (e_1, e_2, \dots, e_k), \tag{3.3}$$

де e_j – j-та компонента унітарного вектору; k – кількість можливих значень відповідного категоріального атрибута; j – індекс компоненти двійкового вектору, причому $j = 1,2, \dots, k$; c – конкретне значення категоріального атрибута, яке кодується.

Таке подання задає загальну структуру двійкового вектору, у якому лише одна позиція відповідає активному значенню атрибута. Після цього потрібно явно визначити правило, за яким формується кожна компонента цього вектору.

Компоненти one-hot вектору задамо таким правилом:

$$e_j = \begin{cases} 1, & \text{якщо } c = j, \\ 0, & \text{інакше.} \end{cases} \tag{3.4}$$

де c – значення категоріального атрибута, яке підлягає кодуванню; j – індекс компоненти, причому $j = 1,2,\dots,k$; k – кількість можливих значень категоріального атрибута; умова $c = j$ означає, що активною є саме j-та позиція унітарного вектору.

Таке означення формалізує перехід від категоріального значення до уніфікованого числового подання. Після кодування всіх категоріальних атрибутів і об'єднання їх із числовими ознаками можна задати загальний вигляд вектора ознак одного зразка.

Загальний вектор ознак мережного з'єднання задамо так:

$$x = (x_1, x_2, \dots, x_d), \tag{3.5}$$

де кожна компонента $x_i \in \mathrm{R}$.

Простір $X \subset R^d$ $(x \in X)$ розглядатимемо підмножиною евклідового простору зі стандартною метрикою.

Для забезпечення коректного відображення ознак у параметри хвильового сигналу виконується лінійна нормалізація кожної числової компоненти. Нехай для ознаки x_iзадані мінімальне та максимальне значення у навчальній вибірці $x_i^{\min}$ та $x_i^{\max}$, які обчислюються виключно на навчальній підвибірці з метою недопущення витоку інформації. Нормалізоване значення визначимо так:

$$\tilde{x}_i = 2 \cdot \frac{x_i - x_i^{\min}}{x_i^{\max} - x_i^{\min}} - 1. \tag{3.6}$$

де x_iу правій частині – початкове числове значення цієї ознаки до нормалізації; $x_i^{\min}$ – мінімальне значення i-ї ознаки, обчислене на навчальній вибірці; $x_i^{\max}$ – максимальне значення i-ї ознаки, обчислене на навчальній вибірці; i – індекс ознаки, причому $i = 1,2,\dots,d$; d – кількість компонент у векторі ознак; множник 2

і зсув на 1 забезпечують відображення вихідного інтервалу значень у симетричний проміжок $[-1,1]$.

Дане перетворення приводить усі координати вектору до єдиного діапазону, що є важливим для подальшого зв'язування значень ознак із параметрами гармонічних коливань. За умови коректного визначення мінімуму і максимуму для кожної координати після нормалізації виконується обмеження на значення всіх компонент.

Межі зміни нормалізованої i-ї ознаки задамо так:

$$\tilde{x}_i \in [-1,1], i = 1, \dots, d. \tag{3.7}$$

де [-1,1] – замкнений інтервал допустимих значень після нормалізації; і – індекс ознаки, причому i=1,2,…,d; d – загальна кількість ознак у векторі.

У випадку $x_i^{\max} = x_i^{\min}$ відповідна координата вважається сталою та може бути зафіксована як нульова у нормалізованому просторі.

Таким чином, кожен зразок мережного трафіку після нормалізації подамо точкою у гіперкубі так:

$$\tilde{x} \in [-1,1]^d. \tag{3.8}$$

де $[-1,1]^d$ – d-вимірний гіперкуб, у якому кожна компонента належить інтервалу $[-1,1]$; d – розмірність простору ознак.

Подане співвідношення задає компактну область, у якій надалі визначається оператор соніфікації. Після цього доцільно ще раз явно зафіксувати вигляд вхідного нормалізованого вектора як безпосереднього аргументу ІКМ-перетворення.

На цьому етапі не виконуються нелінійні перетворення типу логарифмування, експоненційного масштабування або вагового згортання груп ознак. Порядок компонент у векторі зберігається фіксованим і відповідає структурі попередньо сформованого набору ознак. Динамічне переставлення або адаптивне групування компонент не застосовується, що гарантує однозначність відповідності між координатами вектору $\tilde{x}$ та їх подальшим відображенням у часові сегменти сигналу.

Отриманий нормалізований простір ознак $[-1,1]^d$ є областю визначення оператора соніфікації, який відображає кожен вектор $\tilde{x}$у дискретний часовий сигнал фіксованої тривалості.

Після визначення простору ознак вводиться математична модель ІКМ-перетворення, яка задає правила переходу від векторного опису з'єднання до хвильової форми сигналу.

Перетворення нормалізованого вектору ознак у хвильову форму здійснюється шляхом побудови дискретного сигналу фіксованої тривалості, що складається з послідовності гармонічних фрагментів. Такий підхід відповідає моделі сегментованої імпульсно-кодової модуляції (Pulse Code Modulation, ІКМ) у широкому сенсі, де формується дискретний часовий сигнал із заданою частотою дискретизації та наперед визначеною структурою сегментації.

Нехай нормалізований зразок подамо у вигляді вектору так:

$$\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_d), \tilde{x}_i \in [-1,1], \tag{3.9}$$

де $\tilde{x}$ – нормалізований вектор ознак; $\tilde{x}_i$ – нормалізоване значення i-ї ознаки; i – індекс компоненти, причому $i = 1,2, \dots, d$; d – кількість ознак; умова $\tilde{x}_i \in [-1,1]$ означає, що кожну координату нормалізованого вектора приведено до інтервалу $[-1,1]$.

Таке подання фіксує безпосередній вхідний об'єкт для процедури ІКМ-перетворення. Для побудови дискретного хвильового сигналу тепер необхідно задати часові параметри цього сигналу та визначити загальну кількість дискретних відліків.

Загальну кількість дискретних відліків хвильового сигналу визначимо так:

$$N = f_s \cdot T. \tag{3.10}$$

де f_s – частота дискретизації сигналу, причому $f_s > 0$; T – тривалість сигналу у секундах, причому $T > 0$; добуток $f_s \cdot T$ визначає повну довжину сигналу у дискретному часі.

Таке співвідношення задає часовий масштаб хвильового подання та пов'язує параметри дискретизації з довжиною сформованого сигналу. На цій основі

здійснимо перехід до сегментації сигналу за ознаками, визначення кількості відліків на один сегмент і формування гармонічних фрагментів для кожної компоненти вектора ознак.

Сигнал поділяється на d послідовних часових сегментів, кожен із яких відповідає одній компоненті вектору ознак. Кількість дискретних відліків, що припадає на одну ознаку, визначимо так:

$$L = \left\lfloor \frac{N}{d} \right\rfloor. \tag{3.11}$$

де N – загальна кількість дискретних відліків сформованого сигналу; d – кількість ознак у нормалізованому векторі.

Таке співвідношення задає базовий часовий масштаб, у межах якого кожна окрема ознака відображається у власний фрагмент дискретного сигналу. Після визначення довжини сегмента можна формально задати інтервал індексів, який відповідає i-й ознаці.

Інтервал дискретного часу, на якому формується сегмент для i-ї ознаки, задамо так:

$$n \in \{(i-1)L, \dots, iL-1\}. \tag{3.12}$$

де i – індекс ознаки, причому $i = 1,2, \dots, d$; L – кількість відліків, що припадає на одну ознаку; множина $\{(i-1)L, \dots, iL-1\}$ задає всі значення індексу n, що належать сегменту, сформованому i-ю ознакою.

Таке подання формалізує послідовне розміщення сегментів у часовій структурі сигналу без перекриття між ними. Після цього можна встановити правило, за яким значення нормалізованої ознаки визначає частоту гармонічного коливання всередині відповідного сегмента.

Частоту гармонічного коливання для i-ї ознаки визначимо лінійним відображенням нормалізованого значення у заданий частотний діапазон так:

$$f_i = f_{\min} + \frac{\tilde{x}_i + 1}{2} \cdot (f_{\max} - f_{\min}). \tag{3.13}$$

де $f_{\min}$ – нижня межа частотного діапазону; $f_{\max}$ – верхня межа частотного діапазону; виконується умова $0 < f_{\min} < f_{\max} < \frac{f_s}{2}$; f_s – частота дискретизації сигналу; множник $\frac{\tilde{x}_i+1}{2}$ переводить нормалізоване значення з інтервалу $[-1,1]$ у проміжок $[0, 1]$.

Таке відображення забезпечує однозначну відповідність між значенням ознаки та частотою гармонічного компонента сигналу. Разом із частотою необхідно також задати амплітуду коливання, яка визначає енергетичний внесок відповідного сегмента.

Амплітуду гармонічного коливання для i-ї ознаки визначимо так:

$$A_i = \tilde{x}_i. \tag{3.14}$$

де $\tilde{x}_i$ – нормалізоване значення i-ї ознаки; i – індекс ознаки, причому $i = 1,2, \dots, d$.

Подане співвідношення задає найпростіше амплітудне відображення, за якого амплітуда безпосередньо збігається з нормалізованим значенням ознаки. Після фіксації частоти та амплітуди можна записати вираз для дискретного сигналу всередині одного сегмента.

Дискретний сигнал для сегмента, що відповідає i-й ознаці, задамо так:

$$s_i[n] = A_i \cdot \sin\left(\frac{2\pi f_i}{f_s} \cdot (n - (i-1)L)\right), n \in \{(i-1)L, \dots, iL - 1\} \tag{3.15}$$

де A_i – амплітуда гармонічного коливання для i-ї ознаки; f_i – частота гармонічного коливання для i-ї ознаки; f_s – частота дискретизації сигналу; n – індекс дискретного відліку; i – індекс ознаки, причому $i = 1,2, \dots, d$; вираз $n - (i-1)L$ задає локальний індекс усередині i-го сегмента; множина $\{(i-1)L, \dots, iL - 1\}$ визначає межі цього сегмента.

Таке подання описує один локальний гармонічний фрагмент, параметри якого повністю визначаються значенням відповідної ознаки. Наступним кроком є об’єднання всіх таких сегментів у єдиний дискретний сигнал фіксованої довжини.

Повний дискретний сигнал, сформований послідовною конкатенацією всіх сегментів, задамо так:

$$s[n] = s_i[n], \qquad (i-1) \cdot L \le n < i \cdot L, \qquad i = 1,2,\dots,d. \tag{3.16}$$

де $s_i[n]$ – сигнал i-го сегмента; n – глобальний індекс дискретного відліку; i – індекс сегмента, причому $i = 1,2,\dots,d$; L – кількість відліків в одному сегменті; d – кількість ознак. Нерівність $(i-1) \cdot L \le n < i \cdot L$ визначає часові межі i-го сегмента та однозначно пов'язує його з відповідною ознакою вхідного вектора.

Таке означення формує єдину часову структуру сигналу, у якій порядок сегментів відповідає порядку ознак у векторі. Якщо загальна довжина сигналу не ділиться націло на кількість ознак, виникає потреба окремо задати правило для залишкових відліків.

У разі наявності залишкових відліків їх заповнення нульовими значеннями визначимо так:

$$s[n] = 0, n \ge d \cdot L. \tag{3.17}$$

де n – індекс дискретного відліку; d – кількість ознак; L – кількість відліків одного сегмента; добуток $d \cdot L$ визначає сумарну кількість відліків, зайнятих усіма сегментами; умова $n \ge d \cdot L$ задає відліки, що залишилися після формування послідовності гармонічних сегментів і заповнюються нульовими значеннями.

Таке правило забезпечує фіксовану довжину сигналу навіть у випадку, коли N не кратне d. Після цього дискретний хвильовий сигнал можна подати як повну послідовність відліків на інтервалі індексів від 0 до $N-1$.

Отриманий дискретний сигнал подамо послідовністю відліків так:

$$s[n], n = 0,1,\dots,N-1, \tag{3.18}$$

де n – індекс дискретного відліку, причому $n = 0,1,\dots,N-1$; N – загальна кількість відліків сформованого сигналу.

Таке подання фіксує завершений результат ІКМ-перетворення у вигляді дискретного сигналу фіксованої довжини. У такому вигляді цей сигнал уже можна

розглядати як результат дії спеціального оператора, що відображає нормалізований вектор ознак у простір дискретних сигналів.

Слід зауважити, що на межах сегментів можливі стрибкоподібні зміни частоти та амплітуди, оскільки фази сусідніх гармонічних фрагментів не узгоджуються спеціальним чином. Це зумовлює наявність додаткових спектральних компонент, пов'язаних із дискретною сегментацією сигналу, що надалі враховується при застосуванні частотно-часового аналізу.

Оператор ІКМ-перетворення визначимо так:

$$\Phi: [-1,1]^d \to \mathrm{R}^N, \tag{3.19}$$

де $[-1;1]^d$ – область визначення оператора, тобто d-вимірний гіперкуб нормалізованих ознак; d – кількість ознак у векторі; R^N – простір дійсних векторів довжини N, у якому подається результат перетворення; N – загальна кількість відліків сигналу.

Таке відображення формалізує увесь процес переходу від нормалізованого опису мережного з'єднання до його хвильового подання.

Наступним етапом є встановлення частотно-амплітудного відображення, яке визначає відповідність між значеннями ознак і параметрами сформованого сигналу.

Сформований дискретний сигнал $s[n]$, визначений співвідношеннями (3.14)–(3.18) як результат дії оператора Ф(3.19), містить послідовність гармонічних сегментів, параметри яких однозначно визначаються нормалізованими ознаками $\tilde{x}_i$. Для переходу від часової форми подання до частотно-часової структури використовується короткочасне перетворення Фур'є (Short-Time Fourier Transform, КПФ), яке дозволяє оцінити спектральний склад сигналу на локальних інтервалах часу.

Дискретний сигнал, який надалі використовується як вхід для спектрального аналізу, задамо так:

$$s[n], n = 0,1, \dots, N-1, \tag{3.20}$$

де n – індекс дискретного відліку; N – довжина сигналу в дискретному часі; інтервал $n = 0,1, \dots, N-1$ задає повну область визначення сигналу.

Таке подання завершує формалізацію етапу побудови хвильового сигналу та вводить сигнал у вигляді, придатному для подальшого короткочасного перетворення Фур'є, побудови спектрограми та формування двовимірного представлення мережного з'єднання.

Короткочасне перетворення Фур'є для дискретного сигналу, сформованого на попередньому етапі, визначимо так:

$$Z(m,k) = \sum_{n=0}^{N-1} s[n]\, w[n - mR]\; e^{-j2\pi kn/M}, \tag{3.21}$$

де $s[n]$ – дискретний сигнал, сформований у результаті PCM-перетворення; $w[n - mR]$ – віконна функція довжини M, зсунута на mR відліків; n – індекс дискретного відліку, причому $n = 0,1, \dots, N-1$; m – індекс часової позиції або номер фрейму; k – індекс частотного відліку; R – крок зсуву вікна; M – довжина вікна; N – загальна кількість відліків сигналу; j – уявна одиниця; експоненціальний множник $e^{-j2\pi kn/M}$ задає гармонічну базисну функцію дискретного перетворення Фур'є; виконується умова $M \leq N$.

Параметри M та R є фіксованими для всіх зразків, що забезпечує сталість розмірності частотно-часового подання. За умови застосування стандартного дискретного перетворення Фур'є для кожного фрейму кількість частотних бінів визначимо так:

$$K = M, \tag{3.22}$$

де M – довжина вікна, тобто кількість дискретних відліків, що використовуються для аналізу одного фрейму.

Кількість часових фреймів, що утворюються під час ковзного аналізу сигналу, задамо так:

$$T = \left\lfloor \frac{N-M}{R} \right\rfloor + 1, \tag{3.23}$$

де N – загальна кількість відліків сигналу; M – довжина вікна; R – крок зсуву вікна; доданок 1враховує початкове положення вікна перед першим зсувом; величина $\frac{N-M}{R}$ визначає кількість допустимих зміщень вікна вздовж сигналу.

Отримане співвідношення визначає розмірність часової осі спектрального подання. Після задання розмірностей частотної та часової осей можна формально описати простір, якому належить результат короткочасного перетворення Фур'є.

Отримане частотно-часове подання сигналу визначимо як комплексну матрицю:

$$Z \in C^{K \times T}, \tag{3.24}$$

де $C^{K \times T}$ – простір комплексних матриць розміру $K \times T$; K – кількість частотних бінів; T – кількість часових фреймів.

Означенням фіксується, що результат КПФ має комплексний характер і містить одночасно амплітудну та фазову інформацію. Для подальшого аналізу в задачі побудови спектрограми доцільно перейти до модуля комплексного спектра.

Модуль комплексного спектра для кожного елемента матриці Z визначимо так:

$$| Z(m,k) | = \sqrt{(\mathrm{Re}\ Z(m,k))^2 + (\mathrm{Im}\ Z(m,k))^2}, \tag{3.25}$$

де $\mathrm{Re}\ Z(m,k)$ – дійсна частина комплексного значення $Z(m,k)$; $\mathrm{Im}\ Z(m,k)$ – уявна частина комплексного значення $Z(m,k)$; m – індекс часового фрейму, причому $m = 0,1,\dots,T-1$; k – індекс частотного біна, причому $k = 0,1,\dots,K-1$.

З метою стабілізації динамічного діапазону та зменшення впливу великих амплітуд виконується логарифмічне масштабування амплітудного спектра:

$$S(m,k) = \log(1 + | Z(m,k) |). \tag{3.26}$$

де $\log(\cdot)$ – натуральний логарифм; $| Z(m,k) |$ – модуль комплексного значення короткочасного перетворення Фур'є; доданок «1» використовується для уникнення невизначеності при нульових значеннях амплітуди; $m = 0,1,\dots,T-1$; $k = 0,1,\dots,K-1$.

У результаті формується дійсна матриця, яку вже можна інтерпретувати як спектрограму сигналу.

Простір значень спектрограми задамо так:

$$S \in R^{K \times T}, \tag{3.27}$$

де $\mathrm{R}^{K \times T}$ – простір дійсних матриць розміру $K \times T$; K – кількість частотних бінів; T – кількість часових фреймів.

Оскільки частота кожного сегмента сигналу визначається лінійним відображенням нормалізованої ознаки згідно з формулою (3.13), а амплітуда – співвідношенням згідно формули (3.14), то кожна компонента вектору $\tilde{x}$ породжує локалізований максимум енергії у відповідній області спектрограми. Гармонічна структура сегментів, яка визначена за формулою (3.14), забезпечує концентрацію спектральної енергії поблизу частоти f_i, тоді як амплітуда A_i визначає інтенсивність цього максимуму. Таким чином, формується однозначний зв'язок між значенням ознаки та її спектральним представленням.

Частотно-амплітудне відображення зберігає впорядкованість ознак, оскільки часові сегменти сигналу формуються послідовно відповідно до формул (3.12) та (3.16). Логарифмічне масштабування застосовується виключно до амплітудного спектра з метою нормалізації масштабу та не змінює відносного розташування частотних компонент у спектрограмі.

Таким чином, оператор частотно-амплітудного відображення задамо композицією двох відображень так:

$$\Psi = F_{\mathrm{STFT}} \circ \Phi, \tag{3.28}$$

де F_{STFT} – оператор короткочасного перетворення Фур'є з подальшим обчисленням логарифмованої амплітудної спектрограми; Ф – оператор ІКМ-перетворення, який переводить вектор ознак у дискретний хвильовий сигнал; символ $\circ$ означає композицію відображень, тобто послідовне застосування спочатку оператора Ф, а потім оператора FSTFT.

Отримане частотно-часове подання використовується як вхід для моделей глибокого навчання, орієнтованих на обробку двовимірних структур. Далі

необхідно формалізувати структуру отриманого тензорного представлення та його властивості з урахуванням фіксованої розмірності $K \times T$. Після задання параметрів відображення необхідно визначити часову структуру сигналу та спосіб його переходу до спектрального подання.

Часова структура сигналу, сформованого в результаті ІКМ-перетворення, визначається послідовним відображенням компонент вектору ознак у неперервні за індексом дискретні сегменти. Така побудова забезпечує однозначну відповідність між порядком ознак у вхідному векторі та їхнім положенням у часовій області.

Нехай загальна довжина сигналу становить N відліків, а кількість ознак – d. Тоді сигнал будемо розглядати як розбиття інтервалу і задамо його так:

$$T = \{0,1, \dots, N-1\}, \tag{3.29}$$

де N – загальна кількість відліків сигналу; елементи множини T задають усі допустимі значення індексу дискретного часу від першого до останнього відліку сигналу.

Часовий сегмент, що відповідає i-й ознаці вектора, визначимо так:

$$T_i = \{n \mid (i-1)L \le n < i\,L\}, i = 1, \dots, d, \tag{3.30}$$

де n – індекс дискретного відліку сигналу; i – індекс ознаки, причому $i = 1,2, \dots, d$; d – кількість ознак у векторі; L – кількість відліків, що припадає на один сегмент, зокрема $L = \lfloor N/d \rfloor$; нерівності $(i-1)L \le n < iL$задають ліву та праву межі i-го сегмента в дискретному часі.

Це створює основу для подальшого задання сигналу на кожному сегменті та обґрунтовує структурну інтерпретованість побудованого хвильового представлення.

Фрагмент сигналу, що відповідає i-й ознаці на часовому сегменті T_i, визначимо так:

$$s[n]\ |_{T_i}, \tag{3.31}$$

де $s[n]$ – повний дискретний сигнал, сформований у результаті ІКМ-перетворення; T_i – часовий сегмент, що відповідає i-й ознаці; i – індекс ознаки, причому $i = 1,2,\dots,d$; d – кількість ознак у векторі; n – індекс дискретного відліку сигналу.

Це формалізує локальний гармонічний блок сигналу, який містить інформацію лише про одну компоненту вектора ознак. Саме завдяки такому розбиттю формується кусково-періодична структура хвильового представлення.

Часову тривалість одного сегмента сигналу визначимо так:

$$\Delta t_i = \frac{L}{f_s}, \tag{3.32}$$

де L – кількість дискретних відліків, що припадає на одну ознаку; f_s – частота дискретизації сигналу, причому $f_s > 0$; i – індекс ознаки, причому $i = 1,2,\dots,d$.

З урахуванням побудови сигналу повну функцію задамо у вигляді суми з індикаторними функціями так:

$$s[n] = \sum_{i=1}^{d} s_i\,[n] \cdot \mathbf{1}_{\mathcal{T}_i}(n), \tag{3.33}$$

де $s_i[n]$ – сигнал i-го сегмента; d – кількість ознак; $\mathbf{1}_{T_i}(n)$ – індикаторна функція належності індексу n часовому сегменту T_i; n – індекс дискретного відліку; $i = 1,2,\dots,d$.

Таке представлення об’єднує всі локальні гармонічні сегменти в єдину функцію дискретного часу. Воно також чітко показує, що кожен сегмент робить внесок лише на власному часовому інтервалі.

Індикаторну функцію належності відліку nчасовому сегменту T_i визначимо так:

$$\mathbf{1}_{\mathcal{T}_i}(n) = \begin{cases} 1, & n \in \mathcal{T}_i, \\ 0, & n \notin \mathcal{T}_i. \end{cases} \tag{3.34}$$

де n – індекс дискретного відліку; T_i – часовий сегмент, що відповідає i-й ознаці; значення «1» означає належність відліку сегменту T_i, а значення 0 – його відсутність у цьому сегменті.

Подана функція формально забезпечує локалізацію кожного гармонічного компонента на власному інтервалі часу. Завдяки цьому сумарний сигнал зберігає впорядковану сегментну структуру без взаємного накладання фрагментів.

Відсутність перекриття між часовими сегментами визначимо умовою так:

$$\mathcal{T}_i \cap \mathcal{T}_j = \emptyset, i \neq j, \tag{3.35}$$

де T_j – часовий сегмент, що відповідає j-й ознаці; $\cap$ – операція перетину множин; $\emptyset$ – порожня множина; iта j – індекси ознак, причому $i, j = 1,2, \dots, d, i \neq j$.

Це є важливою властивістю для подальшого коректного спектрального аналізу. По-друге, порядок сегментів відповідає фіксованому порядку компонент вектора ознак. Динамічне переставлення сегментів або адаптивне групування не застосовується. Оскільки кожен сегмент генерується незалежно від інших, то сигнал є кусково-стаціонарним у межах окремих інтервалів $\mathcal{T}_i$. При переході між сегментами можливі стрибкоподібні зміни частоти та амплітуди, що формує характерні границі у часовій області. Саме ці межі зумовлюють появу локалізованих спектральних компонент під час застосування КПФ.

Таким чином, часову структуру сигналу можна розглядати як впорядковану послідовність гармонічних блоків однакової тривалості, кожен із яких несе інформацію про одну компоненту нормалізованого вектора ознак. Така побудова забезпечує детермінованість відображення та дозволяє зберігати структурну інтерпретованість сигналу при подальшому спектральному аналізі.

Часова структура сигналу, сформованого в результаті ІКМ-перетворення відповідно до формул (3.14)–(3.18), визначається послідовним відображенням компонент вектору ознак $\tilde{x}$(3.9) у неперервні за індексом дискретні сегменти. Така побудова забезпечує однозначну відповідність між порядком ознак у вхідному векторі та їх положенням у часовій області, що є прямим наслідком визначення інтервалів за формулою (3.12).

Оскільки кожен сегмент $s_i[n]$ (3.33) генерується незалежно від інших згідно з формулою (3.14), то сигнал є частково-стаціонарним у межах окремих інтервалів T_i. При переході між сегментами можливі стрибкоподібні зміни частоти та

амплітуди, оскільки фаза гармонічних коливань не узгоджується між сусідніми інтервалами. Це формує характерні границі у часовій області, які зумовлюють появу додаткових спектральних компонент під час застосування короткочасного перетворення Фур'є згідно формули (3.24).

Таким чином, часову структуру сигналу можна розглядати як впорядковану послідовність гармонічних блоків однакової тривалості, кожен із яких несе інформацію про одну компоненту нормалізованого вектору ознак $\tilde{x}$. Така побудова забезпечує однозначність відображення Ф(3.19) та дозволяє зберігати структурну інтерпретованість сигналу при подальшому частотно-часовому аналізі.

Завершенням процедури перетворення є формування двовимірного подання, придатного для подальшого використання у згортковій нейронній мережі.

Спектрограма $S(m,k)$, отримана в результаті застосування короткочасного перетворення Фур'є (3.24)–(3.27) до сигналу $s[n]$, сформованого оператором Ф(3.19), є двовимірною матрицею розміру $K \times T$, де K– кількість частотних бінів, а T – кількість часових фреймів, визначених на етапі формування частотно-часового подання. Для подальшого використання в архітектурах згорткових нейронних мереж (Convolutional Neural Network, CNN) необхідно формалізувати процедуру перетворення цієї матриці у тензор фіксованої структури.

Спектрограму одного зразка як двовимірну матрицю дійсних значень подамо так:

$$S \in R^{K \times T}. \tag{3.36}$$

де $R^{K \times T}$ – простір дійсних матриць розміру $K \times T$; K – кількість частотних бінів; T – кількість часових фреймів.

Таке подання задає спектрограму як структурований двовимірний масив, придатний для подальшої обробки в архітектурі згорткових нейронних мереж. Оскільки більшість реалізацій двовимірної ЗНМ працює з тензорами, наступним кроком є введення додаткового виміру каналу.

Розширення розмірності спектрограми шляхом додавання виміру каналу визначимо так:

$$\mathcal{S} = S \otimes \mathbf{1}_c, \tag{3.37}$$

де $\otimes$ – операція розширення розмірності за рахунок додаткового виміру; 1c – одиничний канал; c – кількість каналів, причому в даному випадку c=1.

Таке перетворення не змінює числовий зміст спектрограми, але приводить її до формату, сумісного з вхідними вимогами двовимірної ЗНМ. У результаті кожен зразок описується вже не матрицею, а тензором фіксованої структури.

Тензорне подання одного зразка після введення канального виміру визначимо так:

$$\mathcal{S} \in R^{K \times T \times 1}. \tag{3.38}$$

де $R^{K \times T \times 1}$ – простір дійсних тензорів розміру $K \times T \times 1$; K – кількість частотних бінів; T – кількість часових фреймів; остання розмірність «1» відповідає одному каналу. Після цього можна перейти до формування повного тензорного подання всієї навчальної вибірки.

Повне тензорне подання навчальної вибірки задамо так:

$$\mathcal{S}_{\text{train}} \in R^{N_{\text{train}} \times K \times T \times 1}. \tag{3.39}$$

де N_{train} – кількість зразків у навчальній вибірці; K – кількість частотних бінів; T – кількість часових фреймів; 1 – кількість каналів; $\mathbb{R}^{N_{\text{train}} \times K \times T \times 1}$ – простір дійсних тензорів відповідної розмірності.

Таке подання формалізує весь навчальний набір як єдиний чотиривимірний тензор, придатний для пакетного навчання двовимірної ЗНМ. Аналогічний підхід використовується і для формування тестової вибірки.

Тензорне подання тестової вибірки визначимо так:

$$\mathcal{S}_{\text{test}} \in R^{N_{\text{test}} \times K \times T \times 1}. \tag{3.40}$$

де N_{test} – кількість зразків у тестовій вибірці; K – кількість частотних бінів; T – кількість часових фреймів; 1 – кількість каналів; $\mathrm{R}^{N_{\text{test}} \times K \times T \times 1}$ – простір дійсних тензорів відповідної розмірності.

Таке представлення дозволяє трактувати спектрограму як структурований двовимірний масив, де вертикальна вісь відповідає частоті, горизонтальна – часу, а

значення елементів матриці визначають інтенсивність спектральної енергії згідно з формулою (3.27). Водночас це подання не є візуалізацією у традиційному сенсі, а числовою тензорною структурою, придатною для обчислювальної обробки.

Формування 2D-подання завершує етап перетворення простору ознак у частотно-часову область. На цьому етапі не виконується жодної класифікації або оптимізації параметрів моделі. Описана процедура визначає лише спосіб побудови вхідного тензора, який надалі використовується у моделі виявлення вторгнень, що формалізується окремо у розділі, присвяченому архітектурі та навчанню класифікатора.

Таким чином, композиція операторів нормалізації згідно з формулами (3.6)–(3.12), ІКМ-перетворення за формулою (3.19) та частотно-амплітудного відображення згідно з формулою (3.29) визначає відображення так:

$$H: [-1,1]^d \to R^{K\times T\times 1}, \tag{3.41}$$

де $[-1,1]^d$ – область визначення оператора, тобто d-вимірний простір нормалізованих векторів ознак; d – кількість ознак у векторі мережного з'єднання після попередньої обробки та нормалізації; $\mathrm{R}^{K\times T\times 1}$ – область значень оператора, тобто простір дійсних тензорів фіксованої розмірності; K – кількість частотних бінів спектрограми; T – кількість часових фреймів; остання розмірність 1 відповідає одному каналу тензорного подання.

Отримане подання слід розглядати не ізольовано, а як складову загального конвеєра системи виявлення вторгнень, у який запропонований метод інтегрується як окремий етап попередньої обробки даних.

Метод перетворення мережного трафіку у хвильову форму є окремим функціональним блоком у загальному конвеєрі системи виявлення КА (Intrusion Detection System, СВВ). Його призначення полягає у формуванні альтернативного представлення даних, яке забезпечує перехід від табличного простору ознак R^d до частотно-часової області $\mathrm{R}^{K\times T\times 1}$ відповідно до відображення, яке задано за формулою (3.41).

У запропонованому конвеєрі виявлення КА цей блок розміщується між етапом формування вектора ознак та модулем класифікації. Узагальнена структура процесу обробки зображена на рис. 3.1.

Рисунок 3.1 – Узагальнена структура процесу

Запропонований метод не змінює логіку прийняття рішення та не втручається у процес оптимізації параметрів моделі класифікації. Його функція обмежується формуванням нового представлення даних. Класифікаційна модель, її архітектура та критерії оптимізації розглядаються окремо. Інтеграція методу не потребує змін у механізмах збору трафіку або у структурі збереження даних. Вона виконується на рівні обробки вже сформованих векторів ознак x_j.

Обчислювальну складність узагальненого перетворення визначимо так:

$$O(N) + O(T \cdot M\log\ M), \tag{3.42}$$

де N – загальна кількість дискретних відліків сформованого хвильового сигналу; $O(T \cdot M\log\ M)$ – асимптотична складність етапу короткочасного перетворення Фур'є; T – кількість часових фреймів, для яких виконується спектральний аналіз; M – довжина вікна КПФ; $\log M$ – логарифмічний множник, що виникає при використанні швидкого перетворення Фур'є; добуток $T \cdot M\log\ M$відображає сумарну складність обробки всіх фреймів сигналу.

Таким чином, перетворення трафіку у хвильову форму виступає як модуль препроцесингу, що розширює простір подання даних без зміни логіки прийняття рішення у системі виявлення атак та формально інтегрується у конвеєр системи виявлення КА як окремий композиційний блок операторів. Метод може бути застосований як у централізованих, так і у розподіленому конвеєрі систем виявлення КА за умови наявності блоку попередньої нормалізації ознак.

Разом із тим застосування методу має певні обмеження, які необхідно враховувати під час його практичного використання та подальшого розвитку. Запропонований метод перетворення мережного трафіку у хвильову форму, формально визначений композицією операторів згідно формул (3.6)–(3.12), (3.19) та (3.29) і зведений до відображення H згідно формули (3.51), має певні обмеження, які необхідно враховувати під час його застосування в конвеєрі системи виявлення КА.

По-перше, метод базується на фіксованому порядку ознак у векторі $x \in \mathrm{R}^d$ (формула (3.53)) та його нормалізованому поданні $\tilde{x} \in [-1,1]^d$ (формула (3.12)). Часова структура сигналу безпосередньо залежить від цього порядку, оскільки кожна компонента займає визначений сегмент інтервалу T_i згідно з формулою (3.31). Зміна порядку ознак призводить до зміни часової сегментації та, відповідно, до зміни спектрального шаблона навіть за збереження числових значень компонент. Таким чином, відображення є інваріантним відносно значень ознак за фіксованого порядку компонент, але не є інваріантним відносно їх перестановки.

По-друге, кожна ознака має однакову часову тривалість Δt_i, визначену формулою (3.33) через співвідношення $L = \lfloor N/d \rfloor$(3.11). Така рівномірна сегментація не враховує можливу різну інформативність окремих компонент. Метод не передбачає вагового масштабування або адаптивного розподілу часових інтервалів між ознаками, що може обмежувати гнучкість подання у випадках суттєво неоднорідної структури ознак.

По-третє, частотно-амплітудне відображення є лінійним за значенням ознаки на етапі визначення частоти f_i(3.13) та амплітуди A_i(3.14). Нелінійні перетворення типу логарифмування або степеневого масштабування не застосовуються під час формування сигналу $s[n]$ (3.14)–(3.18). Це забезпечує інтерпретованість та однозначність відповідності між значенням ознаки і параметрами гармонічного сегмента, однак може обмежувати здатність до підсилення слабких варіацій ознак у часовому поданні.

По-четверте, під час переходу до спектрального подання застосовується віконна функція фіксованої довжини M та сталий крок зсуву R у визначенні КПФ

згідно з формулою (3.24). Вибір цих параметрів визначає роздільну здатність у частотній та часовій областях і безпосередньо впливає на структуру спектрограми $S \in \mathrm{R}^{K \times T}$ (формула (3.28)). Метод не включає адаптивного вибору параметрів КПФ залежно від характеристик конкретного зразка, що може обмежувати його гнучкість при обробці різнорідних типів трафіку.

По-п’яте, перетворення є повністю детермінованим і не містить стохастичних або генеративних компонентів на етапах нормалізації, ІКМ-побудови та частотно-часового аналізу. Це забезпечує відтворюваність результатів за фіксованих параметрів, проте не передбачає механізмів внутрішнього узагальнення або згладжування шуму на етапі формування сигналу.

По-шосте, розмірність спектрального представлення визначається параметрами K та T, які залежать від довжини сигналу N(3.10) та параметрів КПФ. Збільшення частотної або часової роздільної здатності призводить до зростання розмірності тензора $\mathrm{R}^{K \times T \times 1}$ (формула (3.48)) та, відповідно, до зростання обчислювальних витрат при подальшій обробці у згорткових нейронних мережах.

Таким чином, метод перетворення трафіку у хвильову форму забезпечує структуроване частотно-часове подання даних, однак його застосування обмежується фіксованістю структури ознак, лінійністю відображення, рівномірністю часової сегментації та залежністю від параметрів спектрального аналізу. Врахування зазначених обмежень є необхідною умовою коректної інтеграції методу у загальну модель системи виявлення КА та подальшої інтерпретації результатів класифікації.

Отже, розроблено та формалізовано метод перетворення мережного трафіку у хвильову форму як детерміновану процедуру відображення нормалізованого вектора ознак у частотно-часове подання. Метод побудовано як послідовність чітко визначених етапів: нормалізація ознак, формування дискретного сигналу фіксованої тривалості за принципом ІКМ-кодування, застосування короткочасного перетворення Фур’є та отримання двовимірного тензорного представлення у вигляді спектрограми. Показано, що кожна компонента вектору ознак однозначно пов’язується з відповідним часовим сегментом сигналу, а її значення визначає

параметри гармонічного коливання. Такий підхід забезпечує збереження структурної відповідності між табличним описом мережного з’єднання та його спектральним відображенням. У результаті формується двовимірна структура фіксованої розмірності, придатна для подальшої обробки згортковими нейромережевими моделями. Визначено області визначення та області значень усіх операторів перетворення, а також обґрунтовано детермінований характер процедури за фіксованих параметрів. Це гарантує відтворюваність результатів і узгодженість представлення даних у навчальному та робочому режимах системи виявлення вторгнень.

Запропонований метод не містить механізмів класифікації чи оптимізації параметрів моделі. Його функція полягає виключно у побудові альтернативного частотно-часового представлення мережного трафіку, яке використовується як вхід для моделей виявлення атак, розглянутих далі.

Таким чином, сформовано основу перетворення мережного трафіку у хвильову форму. Метод визначає базове представлення даних, на якому ґрунтуються подальші розроблені підходи до підвищення точності та збалансованості роботи системи виявлення вторгнень.

3.2 Метод сигнатурозбережного адаптивного балансування даних

У задачах інтелектуального виявлення КА однією з ключових проблем залишається суттєвий дисбаланс навчальної вибірки, що є характерним для більшості реальних та еталонних кібербезпекових наборів даних.

Кількість зразків окремих видів КА є в рази, а інколи й у десятки разів меншою за кількість нормального трафіку або за чисельність інших, більш поширених класів вторгнень. У результаті моделі машинного навчання демонструють схильність до переорієнтації на домінантні класи, що призводить до критичного зниження повноти саме для рідкісних КА. Для систем виявлення вторгнень це неприпустимо, оскільки саме рідкісні, малопредставлені або

малопомітні атаки становлять найбільшу загрозу з точки зору порушення конфіденційності, цілісності чи доступності інформаційних ресурсів.

Традиційні підходи до синтетичного доповнення даних намагаються частково розв’язати проблему дисбалансу шляхом створення додаткових зразків на основі інтерполяції між існуючими об’єктами в просторі ознак (формула (3.1)). Проте такі методи мають низку суттєвих недоліків. По-перше, вони не враховують локальну внутрішню структуру малих класів, яка визначає характерні сигнатурні шаблони поведінки атак. По-друге, вони можуть створювати синтетичні точки, що виходять за межі природного розподілу класу, формуючи нефізичні або малоймовірні комбінації ознак, які спотворюють навчальний процес. По-третє, класичні методи синтетичного доповнення зазвичай є стохастичними, що спричиняє варіативність результатів між різними запусками алгоритму та ускладнює відтворюваність експериментів і подальшу інтеграцію у промислові системи.

Таким чином, виникає потреба у методі, здатному не лише збільшувати обсяг навчальних даних, але й зберігати сигнатурні властивості рідкісних КА, контролювати відхилення синтетичних прикладів та адаптуватися до локальних структур кожного окремого класу. Такий метод має працювати на основі формально визначених, детермінованих правил формування нових зразків, відображати реальні внутрішньокласові варіації та уникати створення нефізичних шаблонів.

Ці вимоги є особливо актуальними у контексті перетворення мережного трафіку на акустичні представлення у вигляді ІКМ-сигналів та спектрограм, описаних під час формування соніфікаційного подання. Будь-які спотворення первинних ознак можуть призвести до появи нехарактерних частотних компонентів у сигналі та його спектрограмі, що безпосередньо впливатиме на якість подальшої класифікації.

Для розв’язання зазначених проблем у дисертації розроблено метод сигнатурозбережного адаптивного балансування даних, який забезпечує створення синтетичних зразків на основі локальної статистичної структури класу, враховує

його внутрішні екстремуми та гарантує контроль допустимих відхилень у просторі ознак. Метод побудований як детермінований процедурний механізм, що поєднує локальні медіанні оцінки, процентильні межі, амплітудні характеристики варіацій та сигнатурозбережні зсуви. Завдяки цьому формуються нові приклади, які не виходять за допустимі межі природної динаміки класу та зберігають його ключові поведінкові шаблони.

Запропонований підхід створює основу для формування узгоджених спектрограм у межах відображення H (формула (3.41)) та підвищення ефективності глибинних моделей у задачах виявлення вторгнень. Подальша формалізація методу сигнатурозбережного адаптивного балансування наведена далі.

У зв'язку з цим виникає задача побудови методу балансування, який не лише збільшує представлення міноритарних класів, а й зберігає їхні характерні структурні властивості.

У системах виявлення КА вихідні дані подаються у вигляді багатовимірних векторів, кожен з яких містить числові та категоріальні характеристики мережних з'єднань. Нехай повна множина спостережень позначається як X, де $x_i \in \mathrm{R}^d$ – вектор ознак, структура якого узгоджується з формалізацією простору ознак (формули (3.1)–(3.5)). Вектор ознак одного мережного з'єднання подамо у вигляді розбиття на числову та категоріальну частини так:

$$x = (x^{\text{num}}, x^{\text{cat}}), \tag{3.43}$$

де $x^{\text{num}} \in \mathrm{R}^{d_{\text{num}}}$ – підвектор числових ознак, $x^{\text{cat}} \in \{0,1\}^{d_{\text{cat}}}$ – підвектор категоріальних ознак після one-hot кодування.

Це є необхідним, бо надалі числова і категоріальна частини обробляються за різними правилами під час формування синтетичних зразків.

Множину всіх спостережень, розбиту на підмножини за класами, визначимо так:

$$X = \bigcup_{i=1}^{C} X_i, \tag{3.44}$$

де X_i – підмножина спостережень, що належать i-му класу; i – індекс класу, причому $i = 1,2, \dots, C$; C – загальна кількість класів; $X_i = \{x \in X \mid y = i\}$; y – мітка класу відповідного спостереження.

Дисбаланс між малочисельним і домінантним класами задамо співвідношенням

$$N_i \ll N_j, \tag{3.45}$$

де N_i – кількість спостережень у i-му класі; N_j – кількість спостережень у j-му класі; i – індекс малочисельного класу; j – індекс домінантного класу; символ ≪означає, що N_iє значно меншим за N_j; виконується $N_i = \mid X_i \mid$, $N_j = \mid X_j \mid$, де $\mid\cdot\mid$ – потужність множини.

Моделі машинного навчання, які навчаються на такій вибірці, у першу чергу мінімізують загальну функцію втрат, що часто призводить до орієнтації на великі класи та суттєвого зниження повноти для рідкісних атак. Наявність такого дисбалансу також спричиняє формування зміщених рішень, які не відтворюють локальну структуру малих класів і не фіксують їх характерні сигнатурні шаблони.

Традиційні методи синтетичного доповнення даних створюють нові спостереження на основі інтерполяції та випадкових варіацій у просторі R^d, однак вони не враховують внутрішню статистичну організацію малих класів. Це часто призводить до появи нефізичних або шумових точок, які не відповідають природним відхиленням у межах класу. У задачах виявлення вторгнень створення таких спотворених зразків є небажаним, оскільки після подальших перетворень у ІКМ-сигнал $s[n]$ та спектрограму $S \in \mathrm{R}^{K \times T}$ навіть невеликі викривлення у вихідних ознаках можуть спричинити появу штучних частотних компонентів. Це безпосередньо впливає на стабільність і точність глибинних моделей, що працюють із спектральними представленнями.

Відображення, що формує синтетичні приклади для i-го класу, визначимо так:

$$f_i: X_i \to \hat{X}_i, \tag{3.46}$$

де для кожного малочисельного класу X_i генерується нова множина синтетичних прикладів $\hat{X}_i \subset \mathrm{R}^d$ таким чином, щоб вони залишалися узгодженими з локальними характеристиками цього класу.

Для цього необхідно забезпечити, щоб кожне синтетичне спостереження $x' \in \hat{X}_i$ не виходило за межі природної внутрішньокласової варіативності. Нехай $N_k(x) \subset X_i$ – множина k-найближчих сусідів точки x у межах класу X_iза обраною метрикою. Локальну медіану в околі точки x визначимо так:

$$m(x) = \mathrm{median}\,(N_k(x)), \tag{3.47}$$

де $\mathrm{median}\,(\cdot)$ – оператор медіани, який у даному випадку обчислюється покомпонентно для числових ознак; $N_k(x)$ – множина k-найближчих сусідів точки x у межах класу; k – кількість найближчих сусідів, причому $k \in \mathbb{N}$, $k \geq 1$; $x \in X_i$; X_i – підмножина спостережень i-го класу.

Таке означення задає локальний центр тяжіння в оточенні точки x, який використовується як опорна статистична характеристика класу. На його основі далі можна оцінити допустиму амплітуду відхилень окремих координат від цього локального центру.

Для кожної числової координати медіанне абсолютне відхилення визначимо так:

$$\mathrm{MAD}_{\,j} = \mathrm{median}\,(|\, x_j - m_j(x)\,| : x \in N_k(x)). \tag{3.48}$$

де j – індекс числової ознаки, причому $j = 1,2,\dots,d_{\mathrm{num}}$; x_j – значення j-ї числової ознаки для точки x; $m_j(x)$ – j-та компонента вектора локальної медіани $m(x)$; $N_k(x)$ – множина k-найближчих сусідів точки x; $|\, x_j - m_j(x)\,|$ – абсолютне відхилення значення ознаки від локальної медіани; $\mathrm{median}\,(\cdot)$ – оператор медіани.

Таке подання кількісно характеризує локальну варіативність кожної числової координати в межах відповідного класу. Саме через цю величину вводимо формальну умову допустимості синтетичної точки.

Умову узгодженості синтетичної точки з локальною структурою класу задамо так:

$$| x_j' - m_j(x) | \leq \text{MAD}_{\text{lim}} \cdot \text{MAD}_j, \quad (3.49)$$

де $m_j(x)$ – j-та компонента локальної медіани; MAD_j – медіанне абсолютне відхилення для j-ї координати; MAD_{lim} – параметр, що визначає допустиму межу відхилення синтетичної точки, причому $MAD_{\text{lim}} > 0$; $j = 1,2,\dots,d_{\text{num}}$; d_{num} – кількість числових ознак.

Особливу увагу в межах постановки задачі приділено коректному формуванню як числових, так і категоріальних ознак. Для числової частини необхідно зберегти напрямок та амплітуду локальних варіацій, що задається через евклідову відстань. Задамо її так:

$$d(x) = \| x^{\text{num}} - m^{\text{num}}(x) \|_2, \quad (3.50)$$

де x_{num} – підвектор числових ознак точки x; $m_{\text{num}}(x)$ – числова частина вектора локальної медіани; $\|\cdot\|_2$ – евклідова норма; $x \in X_i$; X_i – підмножина спостережень i-го класу.

На його основі далі введемо адаптивний коефіцієнт зміщення, який визначає масштаб формування синтетичного прикладу, так:

$$s(x) = \alpha \cdot d(x)^{\beta}, \quad (3.51)$$

де α – масштабний параметр, причому $\alpha > 0$; $d(x)$ – евклідова відстань між числовою частиною точки та її локальною медіаною; β – показниковий параметр, який керує нелінійністю зміщення, причому $\beta \geq 0$.

Формування категоріальних ознак повинно бути детермінованим та узгодженим із локальними закономірностями. Це досягається за рахунок вибору найбільш типового значення (моди) серед множини $N_k(x)$ для відповідної категоріальної ознаки, що забезпечує збереження структурної узгодженості one-hot подання.

Отже, задача полягає у створенні механізму синтетичного доповнення даних, який не лише збільшує чисельність малопредставлених класів, а й забезпечує збереження їх внутрішньої структури, сигнатурних шаблонів та статистичних залежностей між ознаками. Метод повинен формувати синтетичні зразки у межах

природної варіативності класу, працювати детерміновано та бути сумісним із подальшими етапами формування ІКМ-сигналів і спектрограм, визначеними під час побудови соніфікаційного представлення. Далі буде наведено математичну модель, що формалізує правила побудови таких синтетичних зразків і механізми контролю їх узгодженості з оригінальною структурою даних.

Для формального опису запропонованого підходу введено математичну модель сигнатурозбережного адаптивного балансування навчальної вибірки. Математична модель методу сигнатурозбережного адаптивного балансування даних базується на припущенні, що сигнатурні властивості кожного класу КА можуть бути описані його статистичною структурою у просторі числових ознак. На відміну від підходів, що формують синтетичні зразки шляхом випадкових інтерполяцій, у даній моделі використовується аргументований статистичний опис кожного класу на основі навчальної вибірки.

Нехай навчальна множина згрупована за типами атак так:

$$X = \cup_{i=1}^{C} X_i, \tag{3.52}$$

де X – навчальна множина; X_i – підмножина зразків i-го класу; i – індекс класу, причому $i = 1,2, \dots, C$; C – загальна кількість класів. Оператор об’єднання означає, що навчальна множина складається з класових підмножин X_i.

Таке подання задає структуру навчальної множини у вигляді об’єднання класових підмножин і створює основу для побудови окремої статистичної моделі для кожного класу. Після цього доцільно ще раз явно зафіксувати внутрішню структуру вектору ознак окремого зразка.

Структуру вектору ознак кожного зразка визначимо так:

$$x = (x^{num}, x^{cat}), \tag{3.53}$$

де x_{num} – числова частина вектору ознак; x_{cat} – категоріальна частина вектору ознак; $x_{\text{num}} \in \mathbb{R}^{d_{\text{num}}}$; d_{num} – кількість числових ознак; $x_{\text{cat}} \in \{0,1\}^{d_{\text{cat}}}$; d_{cat} – кількість компонент категоріальної частини після one-hot кодування; $x \in \mathbb{R}^d$, де $d = d_{\text{num}} + d_{\text{cat}}$.

Поданий запис за формулою (3.53) завершує формалізацію структури окремого зразка в задачі сигнатурозбережного адаптивного балансування. Надалі саме на основі числової частини x_{num} оцінюються глобальні статистичні характеристики класу, а категоріальна частина x_{cat}обробляється за детермінованими правилами збереження локальної структурної узгодженості.

У межах запропонованого методу процентильними межами вважатимемо нижню та верхню емпіричні квантильні границі розподілу кожної числової ознаки в межах відповідного класу або підкласу КА. Вони не є експертно заданими чи довільно підібраними порогами, а обчислюються за навчальною вибіркою. Для j-ї числової ознаки i-го класу визначаються межі q_{ij}^{low} та q_{ij}^{high}, що відповідають значенням функції емпіричного квантила на рівнях 0,05 та 0,95 відповідно. Отже, $q_{ij}^{\text{low}} = Q_{0,05}$ і $q_{ij}^{\text{high}} = Q_{0,95}$. Такий вибір формує діапазон між 5-м і 95-м процентилями та відсікає крайні, потенційно шумові значення. Для ознак із наближеним до нормального розподілом цей інтервал можна інтерпретувати як наближений діапазон допустимої внутрішньокласової варіативності. Формально процентильний інтервал задамо так:

$$I_{ij} = \left[q_{ij}^{\text{low}}, q_{ij}^{\text{high}}\right], \tag{3.54}$$

де I_{ij} – допустимий процентильний інтервал значень j-ї числової ознаки для i-го класу; q_{ij}^{low} та q_{ij}^{high} – відповідно нижня і верхня процентильні межі цієї ознаки; i – індекс класу, причому $i = 1,2,\dots,C$; j – індекс числової ознаки, причому $j = 1,2,\dots,d_{\text{num}}$; C – загальна кількість класів; d_{num} – кількість числових ознак. Інтервал I_{ij} формалізує глобально допустимий діапазон зміни ознаки в межах конкретного класу. Таким чином, для кожного класу формується словник статистичних сигнатур (рис. 3.2).

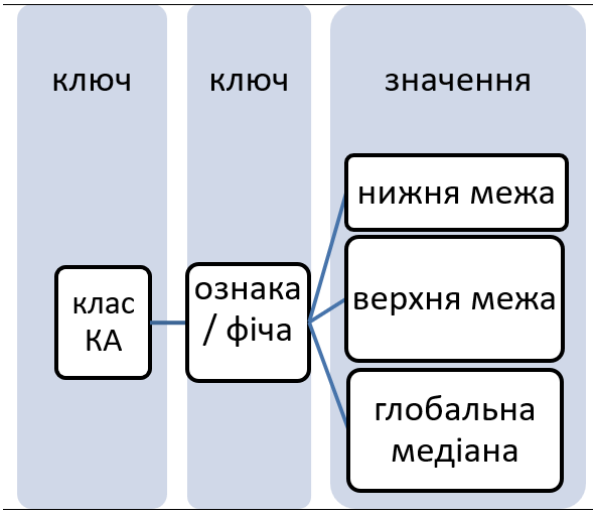

Рисунок 3.2 – Структура варіацій числових ознак конкретного типу КА

Ця структура відображає характерні діапазони варіацій числових ознак конкретного типу атаки та використовується як глобальне обмеження при синтезі нових зразків (рис. 3.3).

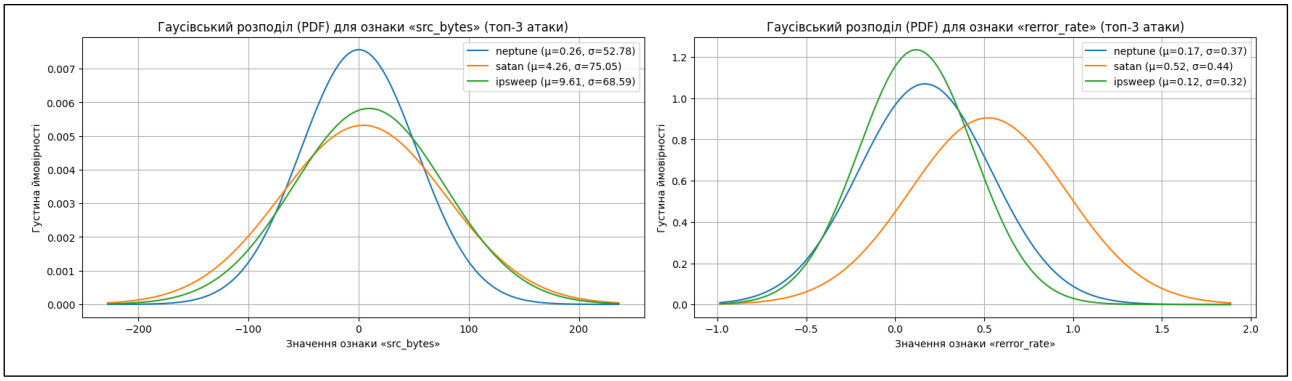

Рисунок 3.3 – Діапазони варіацій числових ознак «src_bytes» та «rerror_rate» для конкретних типів КА

На рисунку 3.3 подано спосіб формування статистичних сигнатур для окремих типів КА на прикладі двох числових ознак – «src_bytes» та «rerror_rate». У лівій частині наведено гаусівські апроксимації розподілу ознаки «src_bytes» для типів атак neptune, satan та ipsweep, у правій частині – відповідні апроксимації для ознаки «rerror_rate». Горизонтальна вісь відображає значення ознаки, вертикальна

– густину ймовірності, а окремі криві характеризують відмінність статистичних профілів різних типів атак.

Зміщення центрів кривих і різна ширина розподілів свідчать, що навіть для однакової числової ознаки допустимі інтервали значень залежать від конкретного підкласу атаки. Тому процентильні межі, визначені за навчальними прикладами відповідного підкласу, використовуються не як універсальні пороги для всієї вибірки, а як класоспецифічні обмеження внутрішньокласової варіативності.

У запропонованому методі це має безпосереднє алгоритмічне значення. Під час синтезу нового зразка його числові компоненти обмежуються межами статистичного профілю того підкласу, до якого належить вихідний приклад. Такий підхід дає змогу зберігати характерні ознаки атак і зменшувати ймовірність появи синтетичних зразків, які формально отримані інтерполяцією, але виходять за межі реалістичної структури мережного трафіку.

Локальна структура класу. Для кожного вихідного зразка $x \in X_i$ додатково визначимо локальну множину найближчих сусідів так:

$$N_k(x) = \{x^{(1)}, x^{(2)}, \dots, x^{(k)}\} \subset X_i, \tag{3.55}$$

де x – вихідний зразок, для якого формується локальне оточення; $x_1, x_2, \dots, x_k$ – сусідні зразки, що належать тому самому класу; k – кількість найближчих сусідів, причому $k \in \mathbb{N}$, $k \geq 1$; X_i – підмножина зразків i-го класу; i – індекс класу.

Пошук сусідів виконується у просторі числових ознак за евклідовою метрикою, що забезпечує коректне відображення локальної геометрії класу.

Локальну медіану в околі точки x задамо так:

$$m^{loc}(x) = median(N_k(x)), \tag{3.56}$$

де $\mathrm{median}\,(\cdot)$ – оператор медіани, який для числових ознак обчислюється покомпонентно; $N_k(x)$ – множина k-найближчих сусідів точки x; x – вихідний зразок.

Локальна медіана відображає найбільш типове значення ознак у безпосередньому оточенні зразка.

Таким чином, у моделі одночасно використовуються дві характеристики:

1) глобальна медіана класу m^{glob}, що визначає сигнатурний центр класу;

2) локальна медіана $m^{loc}(x)$, що відображає конкретну підструктуру всередині класу.

Глобальну амплітуду допустимих варіацій j-ї числової ознаки для i-го класу визначимо так:

$$A_{ij}^{glob} = q_{ij}^{high} - q_{ij}^{low}, \tag{3.57}$$

де q_{ij}^{high} – верхня процентильна межа ознаки; q_{ij}^{low} – нижня процентильна межа ознаки; i – індекс класу; j – індекс числової ознаки.

Таке співвідношення задає природний статистичний діапазон зміни ознаки для конкретного класу. Воно надалі використовується як масштабуючий множник при формуванні синтетичних координат.

Визначення масштабу зсуву. Відстань від зразка до його локальної медіани визначимо так:

$$d(x) = \| x^{num} - m^{loc}(x) \|_2, \tag{3.58}$$

де x_{num} – підвектор числових ознак зразка x; $m_{\text{loc}}(x)$ – вектор локальної медіани; $\|\cdot\|_2$ – евклідова норма.

Адаптивний коефіцієнт зміщення визначимо так:

$$s(x) = \alpha \cdot d(x)^{\beta}, \tag{3.59}$$

де α – параметр масштабу зміщення, причому $\alpha > 0$; $d(x)$ – відстань від зразка до локальної медіани; β – показниковий параметр нелінійності, причому $\beta \geq 1$.

Напрямок зсуву з урахуванням глобальної сигнатури. Для кожної координати jнапрямок модифікації визначоми з урахуванням як локальної, так і глобальної тенденції так:

$$\delta_j(x) = x_j - m_{ij}^{glob}. \tag{3.60}$$

де x_j – значення j-ї числової ознаки для зразка x; m_{ij}^{glob} – глобальна медіана j-ї числової ознаки для i-го класу; i – індекс класу; j – індекс числової ознаки.

Значення величини $\delta_j(x)$ визначає, у який бік відносно глобальної сигнатури класу розташований зразок. Таким чином, синтетичні точки формуються у напрямку, узгодженому з типовою структурою всього класу, а не лише локального околу.

Формування синтетичних числових координат. Синтетичне значення ознаки задамо так:

$$x_j' = x_j + sign(\delta_j(x)) \cdot s(x) \cdot A_{ij}^{glob}, \tag{3.61}$$

де x_j – початкове значення цієї ознаки; sign $(\delta_j(x))$ – знак відхилення $\delta_j(x)$, який задає напрямок зміщення; $s(x)$ – адаптивний коефіцієнт зміщення; A_{ij}^{glob} – глобальна амплітуда допустимих варіацій j-ї ознаки для i-го класу; i – індекс класу; j – індекс числової ознаки.

Після цього виконується жорстке обмеження. Умову статистичної допустимості синтетичного значення ознаки визначимо так:

$$q_{ij}^{low} \leq x_j' \leq q_{ij}^{high}, \tag{3.62}$$

де q_{ij}^{low} та q_{ij}^{high} – відповідно нижня і верхня процентильні межі j-ї числової ознаки для i-го класу; $x_j{}'$ – синтетичне значення j-ї числової ознаки; i – індекс класу; j – індекс числової ознаки.

Нерівність гарантує перебування синтетичного значення у статистично допустимому інтервалі відповідного класу.

Такий механізм забезпечує збереження глобальної сигнатури КА, контроль амплітуди зсуву, відсутність нефізичних або статистично необґрунтованих значень.

Категоріальні ознаки. Категоріальні координати не підлягають гаусівській апроксимації. Для них використаємо детермінований принцип вибору моди у локальному околі так:

$$x_j^{cat\prime} = mode(\{y_j^{cat} : y \in N_k(x)\}). \tag{3.63}$$

де $\mathrm{mode}\,(\cdot)$ – оператор моди, що повертає найчастіше значення в заданій множині; $y_{j,\mathrm{cat}}$ – значення j-ї категоріальної ознаки для сусіднього зразка y; $N_k(x)$ – множина k-найближчих сусідів точки x; j – індекс категоріальної ознаки.

Це забезпечує збереження типової категоріальної структури без внесення стохастичних змін.

Підсилення екстремумів. Для відтворення характерних пікових шаблонів класу допускається згладжене зміщення синтетичного значення у напрямку глобального екстремуму класу. Задамо його так:

$$x'_j \leftarrow x'_j + \gamma(ext_{ij}^{glob} - x'_j), \tag{3.64}$$

де x'_j – поточне синтетичне значення j-ї числової ознаки після основного етапу формування; $\leftarrow$ – оператор оновлення значення; γ – коефіцієнт інтенсивності підсилення, причому $\gamma \in [0,1]$; $\mathrm{ext}_{ij}^{\mathrm{glob}}$ – глобальний екстремум j-ї числової ознаки для i-го класу; i – індекс класу; j – індекс числової ознаки.

Таким чином, модель поєднує: глобальний статистичний опис класу через гаусівську апроксимацію та процентильні межі, локальну геометрію підкласів, адаптивне масштабування зсуву, детерміноване формування категоріальних ознак.

Усі параметри моделі – k, α, β, γ – мають чітку інтерпретацію та забезпечують контроль над масштабом і характером синтезу. Запропонована математична модель дозволяє формувати синтетичні зразки, що зберігають сигнатурні властивості конкретних типів КА, не порушують їх статистичної структури та забезпечують прогнозовану поведінку при подальшому перетворенні у ІКМ-сигнали та спектрограми. Це створює стабільну основу для навчання глибинних моделей у задачах виявлення вторгнень. На основі побудованої математичної моделі визначається послідовність алгоритмічних кроків реалізації методу.

Алгоритм сигнатурозбережного адаптивного балансування реалізує сформульовану математичну модель та призначений для детермінованого формування синтетичних зразків малопредставлених класів КА із збереженням їх внутрішньокласової структури. На відміну від стохастичних або інтерполяційних підходів, у яких нові спостереження залежать від випадкових параметрів і можуть

виходити за межі природної варіативності, запропонована процедура ґрунтується на поєднанні глобальних статистичних сигнатур класу та локальної геометрії найближчого оточення вихідного зразка. Це забезпечує відтворюваність синтезу за фіксованих параметрів методу, а також узгодженість синтетичних даних із розподілом відповідного класу.

На початковому етапі алгоритму використаємо навчальну вибірку, у межах якої дані групуються за підкласами КА. Для кожного підкласу формується глобальна статистична сигнатура числових ознак, що відображає типові значення та межі допустимих відхилень, характерних саме для даного типу атаки. Така сигнатура розглядається як формалізований опис внутрішньокласової поведінки і застосовується як обмеження, яке запобігає виникненню синтетичних значень, несумісних із природною структурою підкласу. Категоріальні ознаки на цьому етапі не моделюються розподілами, а обробляються окремим детермінованим правилом, що виключає появу нефізичних комбінацій категорій.

Далі синтез виконується для кожного вихідного зразка малопредставленого підкласу з урахуванням локальної структури. У просторі числових ознак формується локальний окол шляхом відбору k найближчих сусідів, після чого обчислюється локальна медіана як робастна оцінка центральної тенденції у найближчому оточенні. Локальна медіана визначає типове значення ознак у підкластері, до якого належить вихідний зразок, тоді як глобальна медіана та глобальні межі, отримані на першому етапі, задають загальну сигнатурну «рамку» підкласу. Саме поєднання локальної та глобальної інформації забезпечує узгодженість синтетичних прикладів як із підструктурами підкласу, так і з його загальним статистичним профілем.

На основі локального околу визначається масштаб адаптивного зміщення, який залежить від відстані вихідного зразка до локальної медіани. Напрямок корекції числових координат встановлюється таким чином, щоб синтетичний зразок формувався у напрямку, узгодженому з типовою поведінкою підкласу: локальна медіана задає «найближчу норму» підкласу у безпосередньому оточенні, а глобальна медіана використовується як орієнтир для розрізнення характерних

сигнатурних тенденцій від випадкових локальних флуктуацій. Завдяки цьому зсув не є довільним і не зводиться до інтерполяції між двома точками, а відповідає контрольованій модифікації ознак, пов’язаній зі структурою класу.

Після визначення масштабу та напрямку виконується побудова числової частини синтетичного зразка шляхом детермінованого зміщення координат із урахуванням амплітуди допустимих варіацій, характерної для відповідної ознаки. Сформовані значення підлягають перевірці узгодженості. По-перше, контролюється віддалення від локальної медіани за робастними межами, що відповідають допустимій внутрішньокласовій варіативності. По-друге, забезпечується належність до глобально допустимого інтервалу підкласу, сформованого на етапі побудови сигнатур. У випадку порушення обмежень виконується корекція значень або відбраковування синтетичного кандидата. Такий контроль зменшує ризик появи нефізичних числових поєднань, які в подальшому могли б породжувати штучні спектральні компоненти під час перетворення у ІКМ-сигнали та спектрограми.

Категоріальні ознаки синтетичного зразка формуються після завершення побудови числової частини. Для кожної категоріальної координати обирається найбільш типове значення в межах локального околу, що забезпечує відповідність синтетичної комбінації категорій реально спостережуваним шаблонам підкласу та усуває залежність від випадкового вибору. За потреби, для числових координат додатково застосовується сигнатурозбережне підсилення характерних екстремумів, яке використовується для відтворення локальних пікових структур підкласу, важливих для формування відтворюваних спектральних «відбитків» у подальшому частотно-часовому представленні.

Завершальним етапом є включення синтетичного зразка до множини даних відповідного підкласу та повторення процедури до досягнення цільового обсягу доповнення. Оскільки всі компоненти алгоритму визначені детерміновано за фіксованих параметрів, отримана збалансована вибірка є відтворюваною, а її структура – керованою. Це забезпечує прогнозованість навчання моделей на

спектральних представленнях та зменшує ризик того, що синтетичні приклади спотворять сигнатурні властивості рідкісних атак.

Для узагальненого відображення місця та логіки роботи методу в межах підготовки даних подамо його у вигляді текстової блок-схеми (рис. 3.4).

Таким чином, наведений опис алгоритму реалізації методу подає його як послідовну процедуру синтезу, у якій поєднуються локальні робастні оцінки та глобальні сигнатурні межі підкласів. Це дозволяє забезпечити структурну коректність синтетичних даних і підготувати вибірку, сумісну з подальшим соніфікаційним перетворенням та спектральним аналізом у складі конвеєра СВВ.

Для обґрунтування доцільності запропонованого підходу необхідно зіставити його з існуючими методами балансування навчальних вибірок.

Проблема дисбалансу класів у задачах виявлення КА традиційно вирішується за допомогою методів випадкового перевибірювання, інтерполяційних алгоритмів типу SMOTE та його модифікацій, а також генеративних моделей. Більшість із цих підходів орієнтована на вирівнювання чисельності класів, проте не враховує специфіку сигнатурної структури КА та особливості подальшого перетворення ознак у спектральне подання.

Запропонований сигнатурозбережний адаптивний метод відрізняється тим, що розглядає підклас КА як статистично структуровану множину, для якої попередньо формується глобальна сигнатура числових ознак. Синтетичні зразки генеруються не шляхом випадкової інтерполяції між довільними сусідами, а через контрольоване зміщення в межах статистично обґрунтованих інтервалів класу з урахуванням як глобальної, так і локальної структури. Такий підхід орієнтований не лише на збільшення кількості прикладів, а й на збереження аномальної геометрії класу, що є критично важливим при подальшій соніфікації та формуванні спектрограм.

У класичному SMOTE нові зразки формуються як лінійні комбінації між випадково обраним прикладом та одним із його сусідів. Це забезпечує простоту реалізації та відносно невисоку обчислювальну складність, однак призводить до появи точок, які можуть виходити за межі природної сигнатурної області підкласу.

Крім того, метод не враховує глобальні статистичні характеристики класу та не контролює відповідність синтетичних значень типовим діапазонам варіацій.

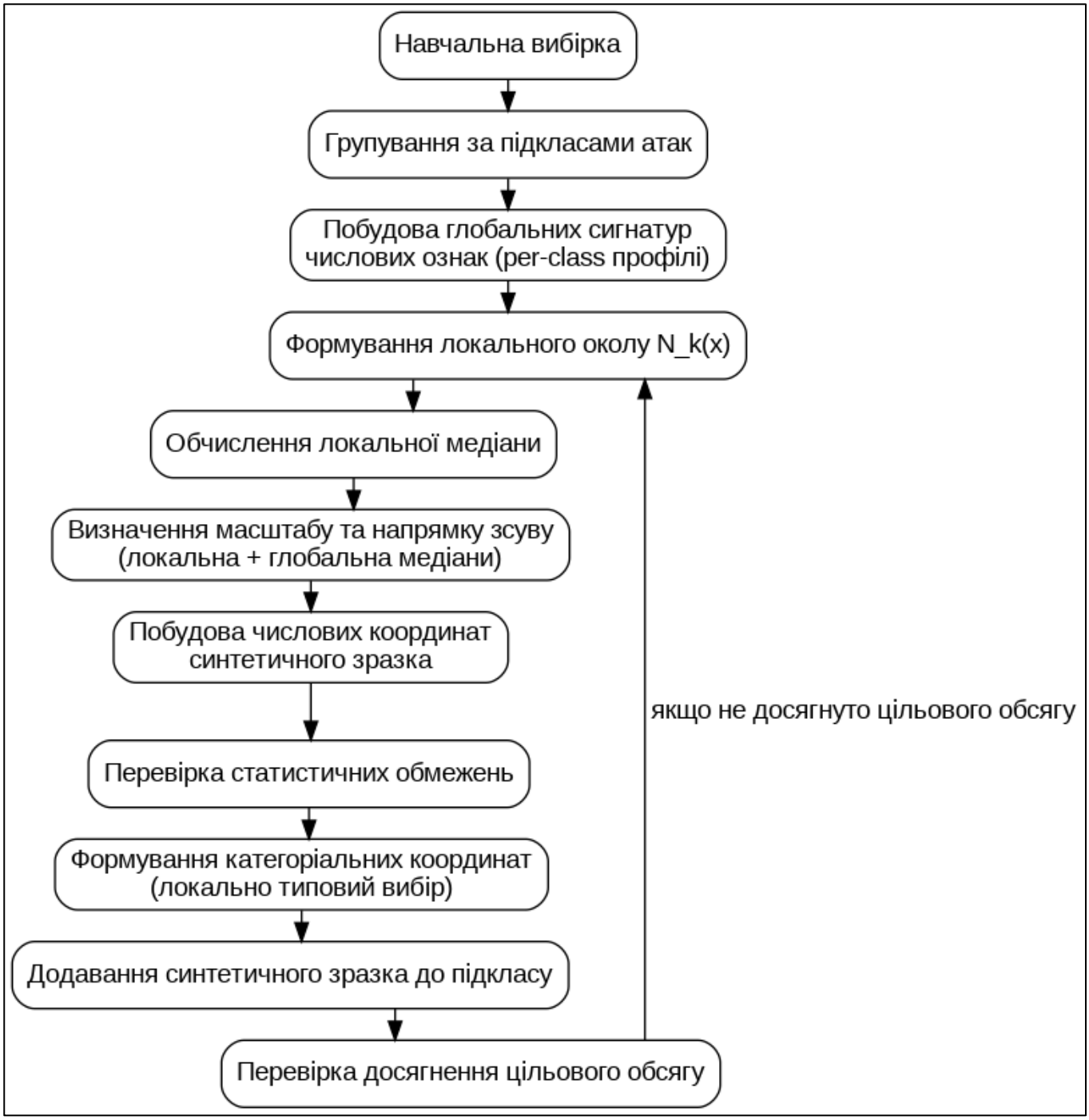

Рисунок 3.4 – Схема функціонування методу

Випадковий надсемплінг, у свою чергу, не змінює структуру даних, а лише дублює існуючі приклади. Такий підхід не створює нової інформації та може призводити до перенавчання моделі на повторюваних зразках.

Запропонований метод поєднує переваги обох підходів. З однієї сторони, він створює нові приклади, а з іншої – обмежує їх формування статистично обґрунтованими межами та враховує внутрішню структуру класу. Порівняльна характеристика методів балансування подана в табл. 3.1.

Представлене порівняння демонструє, що основною відмінністю запропонованого методу є наявність глобального статистичного контролю та поєднання його з локальною структурою підкласу, що забезпечує сигнатурозбережний характер синтезу.

Таблиця 3.1 – Порівняльна таблиця методів балансування

Характеристика	Випадковий надсемплінг	SMOTE	Сигнатурозбережний метод
Принцип формування	Дублювання зразків	Лінійна інтерполяція між сусідами	Детерміноване зміщення в межах глобальної та локальної сигнатури
Відтворюваність результату	Висока за фіксованих параметрів	Залежить від випадкових виборів	Висока (детермінованість за фіксованих параметрів)
Збереження сигнатури класу	Повне (але без варіацій)	Часткове, можливі спотворення	Контрольоване та статистично обґрунтоване
Контроль допустимих меж ознак	Відсутній	Відсутній	Явно враховується через глобальні межі
Ризик перенавчання	Високий	Помірний	Знижений за рахунок контрольованої варіативності
Обчислювальна складність	Низька	Помірна	Помірна (потребує попередньої статистичної оцінки)
Сумісність із соніфікацією	Нейтральна	Можливі спектральні артефакти	Орієнтований на стабільність спектрального подання

Для наочного аналізу розглянемо три ілюстрації, зображені на рис. 3.5-3.7.

На рис. 3.5 представлено нормалізований вектор числових ознак реального зразка підкласу «warezmaster». Значення атрибутів подано в інтервалі нормалізації [-1; 1]. Відображено характерну сигнатурну структуру зразка, зокрема наявність виражених пікових компонент окремих ознак та стабільних від'ємних значень більшості інших координат. Така конфігурація формує специфічний профіль підкласу, який надалі використовується як базова сигнатура для порівняння із синтетично згенерованими прикладами.

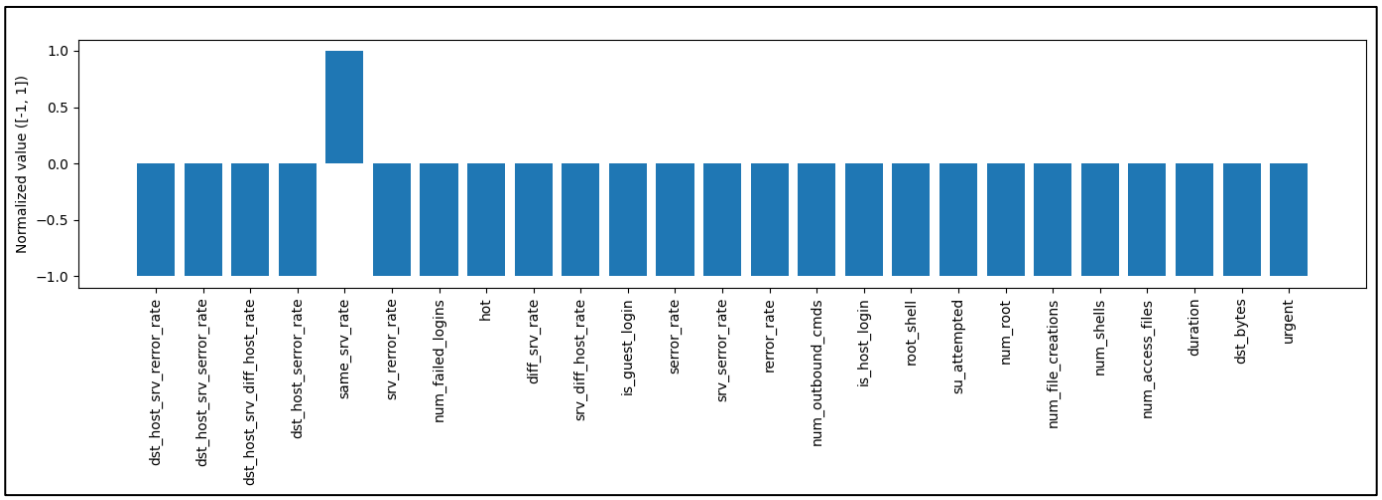

Рисунок 3.5 - Приклад реального зразка з навчального набору даних (class = warezmaster)

На рис. 3.6 відображено синтетичний запис, сформований сигнатурозбережним адаптивним методом на основі реального зразка, наведеного на рис. 3.6. Збережено загальну форму сигнатури, співвідношення амплітуд між атрибутами та характерні пікові структури. Модифікації числових координат здійснено в межах статистично допустимих інтервалів підкласу, що забезпечило узгодженість синтетичного зразка з глобальною та локальною структурою класу. Геометрія вектору ознак залишилася стабільною, а відхилення мають контрольований характер.

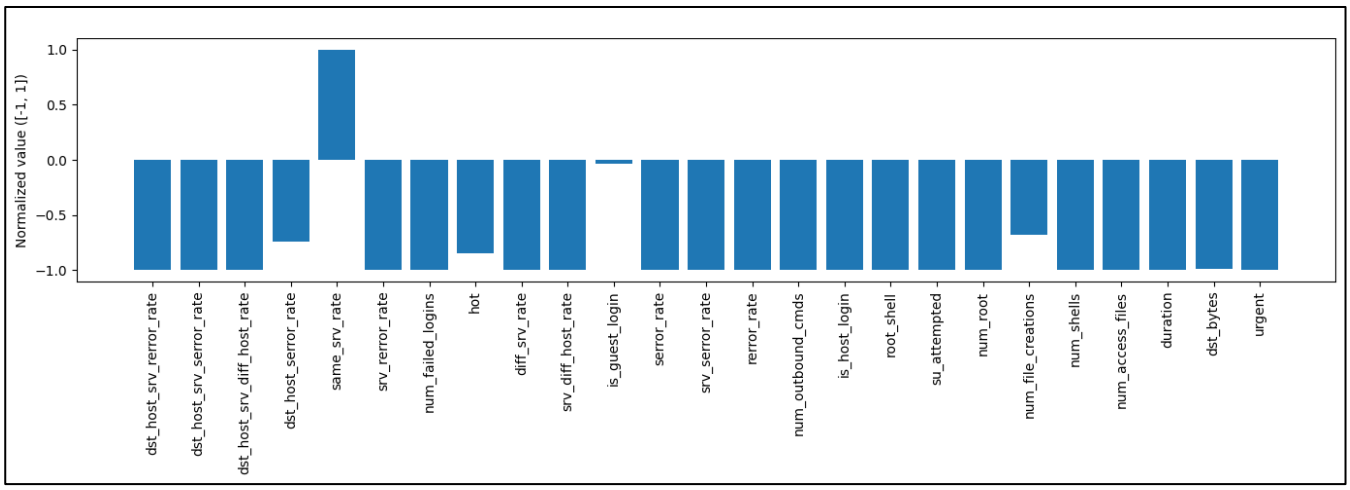

Рисунок 3.6 - Синтетичний зразок, сформований запропонованим методом

На рис. 3.7 представлено синтетичний запис, який отриманий за допомогою інтерполяційного методу SMOTE. Спостерігається часткове згладжування окремих атрибутів та відсутність чітко вираженого контролю глобальних сигнатурних меж підкласу. Зміни координат мають інтерполяційний характер і не пов’язані з глобальною статистичною структурою класу. Порівняння з рисунком 3.8 демонструє, що запропонований метод забезпечує більш детерміноване та сигнатурозбережне формування синтетичних прикладів, тоді як інтерполяційний підхід створює варіації, що менш жорстко прив’язані до характерної структури підкласу.

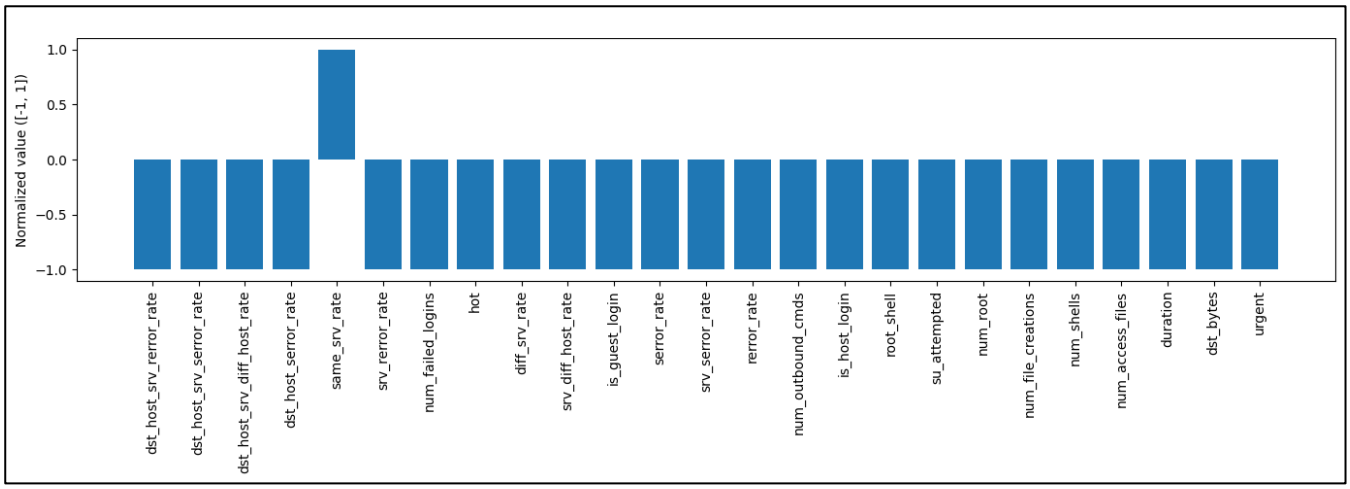

Рисунок 3.7 - Синтетичний зразок, сформований методом SMOTE

Таким чином, проведене порівняння свідчить, що запропонований сигнатурозбережний адаптивний метод орієнтований не лише на вирівнювання чисельності класів, а й на структурну коректність синтетичних даних. Це є принциповою відмінністю від існуючих рішень і визначає доцільність його використання в задачах, де важливе збереження аномальної геометрії підкласів атак та стабільність подальшого спектрального представлення в межах соніфікаційного конвеєра систем виявлення КА. Практичне значення розробленого методу визначається також можливістю його інтеграції в загальний соніфікаційний конвеєр системи виявлення КА.

Метод сигнатурозбережного адаптивного балансування даних інтегрується у соніфікаційний конвеєр системи виявлення КА на етапі формування навчальної вибірки, тобто до побудови ІКМ-сигналів і спектрограм. Такий підхід обумовлений тим, що соніфікація та подальше частотно-часове перетворення є чутливими до викривлень вихідних ознак, тобто означає, що синтетичні зразки, які сформовані без контролю внутрішньокласової структури, можуть породжувати нехарактерні спектральні компоненти, що погіршує узагальнювальну здатність згорткових моделей. Відтак балансування доцільно виконувати в табличному просторі ознак, забезпечуючи структурну коректність даних до їх акустичного відображення.

У загальному конвеєрі систем виявлення КА соніфікаційний конвеєр розглядається як базова послідовність обробки, що включає формування вектору ознак, попередню обробку, перетворення у ІКМ-подання, побудову спектрограм і класифікацію двовимірної ЗНМ. Запропонований метод інтегрується як окремий модуль підготовки даних, який активується лише у навчальному режимі та не змінює структуру базового конвеєра. Його роль полягає у цільовому доповненні малопредставлених класів атак синтетичними прикладами, сформованими на основі статистично узгоджених сигнатур класу та контрольованих відхилень. Це дозволяє зменшити ризик «нефізичних» синтетичних точок і, як наслідок, підвищити узгодженість спектральних представлень, на яких навчається класифікатор.

Модуль балансування реалізується як процедурний етап із чітко визначеною послідовністю дій: виділення класів, що потребують підсилення; оцінювання сигнатурних меж для числових ознак у межах кожного класу; побудова синтетичних зразків із урахуванням локальної структури та типових значень; детерміноване формування категоріальних компонент; контроль допустимих відхилень і включення синтетики до навчального набору. Тому на вхід соніфікаційного блоку надходить навчальна вибірка, збалансована з урахуванням сигнатурних властивостей атак, що забезпечує більш прогнозоване навчання та зменшення деградації спектральних шаблонів.

У робочому режимі систем виявлення КА (інференс) модуль балансування не застосовується, оскільки його призначенням є підготовка навчальних даних, а не модифікація реального потоку трафіку. Таким чином, інтеграція методу (рис. 3.8) зберігає незмінною базову логіку обробки мережних записів, водночас забезпечуючи якіснішу структуру навчального набору для рідкісних КА та зменшуючи негативний вплив дисбалансу класів на соніфікаційний підхід.

Таке розташування методу (рис. 3.8), тобто базова послідовність обробки та дотичний модуль підготовки навчальної вибірки (застосовується лише на етапі навчання), забезпечує оптимальне поєднання його функціональних властивостей зі структурними вимогами соніфікаційного підходу і гарантує, що синтетично доповнені дані будуть коректно відображені у спектральній формі.

Метод має певні обмеження, які повинні бути враховані під час його застосування у різних умовах навчання та експлуатації моделі. Вони пов’язані із природою рідкісних класів та поведінкою локальних статистик. Основною передумовою його коректної роботи є достатня репрезентативність локальних околів. Якщо рідкісний клас містить надто малу кількість спостережень або має фрагментовану структуру з декількома віддаленими підкластерами, то локальна медіана, процентильні межі та медіанне абсолютне відхилення можуть неточно відображати внутрішню варіативність. У таких випадках синтетичні зразки можуть бути частково зміщеними або надмірно узагальненими.

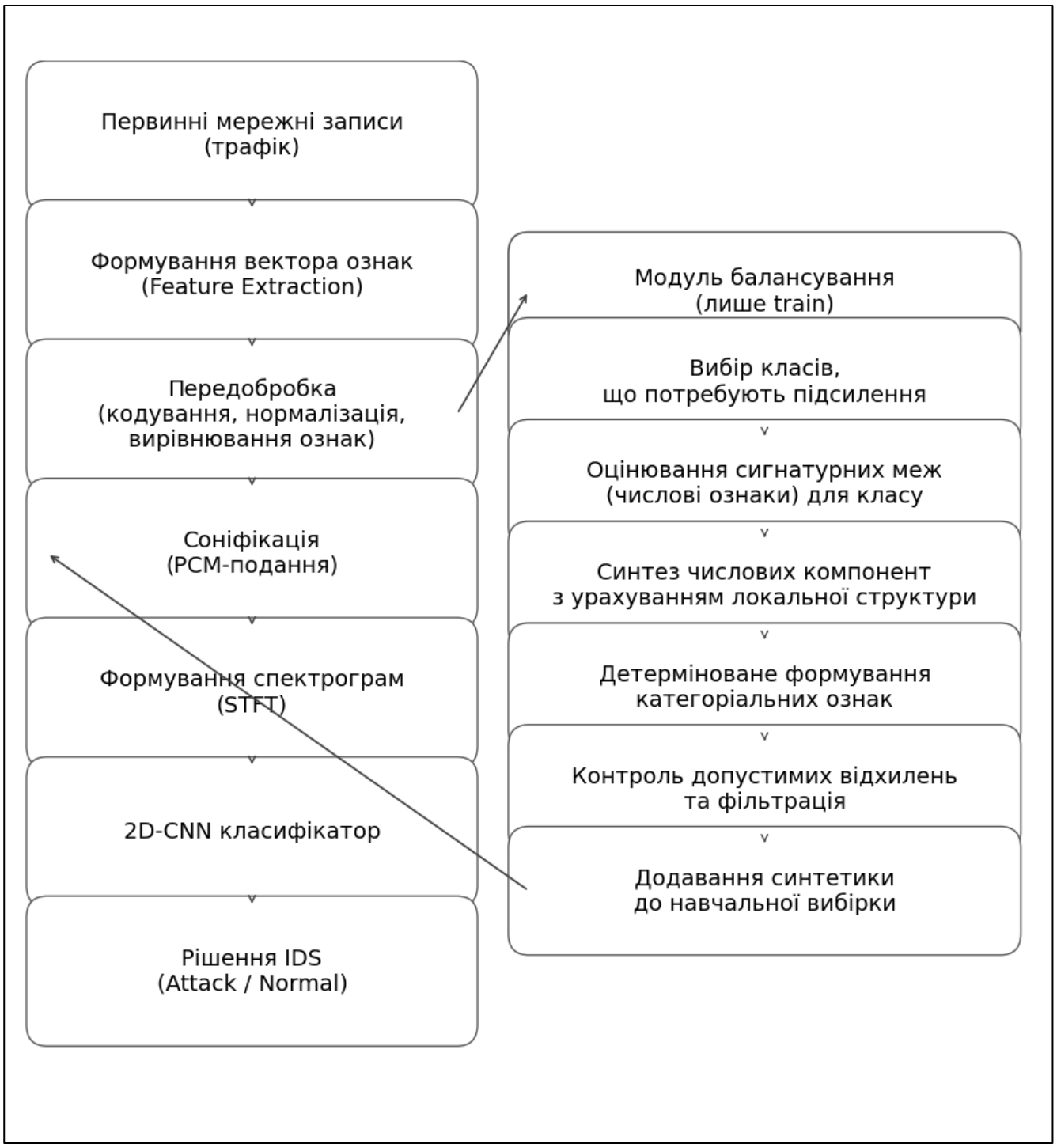

Рисунок 3.8 - Інтеграція методу сигнатурозбережного адаптивного балансування у конвеєр

При малому параметрі локального околу дані стають чутливими до випадкових шумових значень, тоді як надмірно великі околи згладжують критичні шаблони рідкісних атак. Таким чином, метод потребує підбору параметрів відповідно до густини та структури класу. Обмеження також проявляються у сфері

категоріальних ознак: використання моди забезпечує збереження типових значень, але водночас зменшує різноманітність і може приглушувати менш частотні, але потенційно значущі категоріальні варіації.

Метод не орієнтований на створення абсолютно нових поведінкових структур класу, а лише на відтворення тих, що вже присутні у даних. Унаслідок цього він ефективний для компактних і локально узгоджених класів, але менш придатний для випадків, коли рідкісні атаки демонструють широке внутрішнє різноманіття або коли їх спектральні характеристики сильно залежать від дрібних числових варіацій. Разом ці особливості визначають робочі межі методу та підкреслюють необхідність урахування структури вихідної вибірки під час його застосування.

Отже, метод сигнатурозбережного адаптивного балансування дає змогу зменшити негативний вплив дисбалансу класів без втрати характерних властивостей міноритарних підкласів і створює основу для подальшого підсилення точності класифікації.

Розглянутий метод сигнатурозбережного адаптивного балансування даних забезпечує формування збалансованої вибірки для малопредставлених класів завдяки детермінованому моделюванню локальної структури кожного класу. Використання медіанних та процентильних характеристик дозволяє уникати спотворень, характерних для інтерполяційних та стохастичних методів, водночас забезпечуючи збереження внутрішніх сигнатур КА, які мають ключове значення при подальшому перетворенні даних у ІКМ-сигнали та спектрограми. Завдяки адаптивному контролю зміщень, обмеженню варіацій за медіанним абсолютним відхиленням (MAD) та спрямованому формуванню числових і категоріальних ознак синтетичні зразки залишаються узгодженими зі статистичною природою класу.

Метод інтегрується у соніфікаційний конвеєр систем виявлення КА на етапі формування навчальної вибірки, забезпечуючи узгодженість спектральних представлень та підвищуючи якість класифікації рідкісних типів вторгнень. Детермінований характер алгоритму робить результати синтезу відтворюваними,

що є важливою властивістю для промислового застосування та порівнянності експериментів. Попри певні обмеження, пов’язані з розміром вибірки та структурною варіативністю окремих класів, метод демонструє високу ефективність у задачах, де збереження локальних сигнатур та узгодженості спектральної форми є критично важливими.

3.3 Метод підвищення точності виявлення комп’ютерних атак у зонах невизначеності на основі аналізу помилок та селективного перевизначення рішень

Викладені вище положення задають вихідний контекст для подальшого уточнення процесу виявлення КА: табличні характеристики мережного трафіку розглядаються як джерело частотно-часового подання, а нейромережна модель використовується для первинної класифікації мережних з’єднань. У межах цього підрозділу такі положення розглядаються не як завершена архітектурна схема, а як методологічна основа для розроблення методу підвищення точності виявлення КА у випадках, коли рішення базової моделі є помилковим або недостатньо визначеним.

Разом із тим аналіз задачі бінарної класифікації виявляє низку обмежень, що стримують подальше підвищення точності. По-перше, навіть за умов попереднього балансування вибірки спостерігається нерівномірність якості розпізнавання різних підкласів атак. Окремі типи вторгнень характеризуються складною внутрішньою структурою або багатомодальним розподілом, що призводить до зростання частоти помилок саме для цих підкласів. За відсутності механізму цілеспрямованої корекції таких помилок модель зберігає локальні зони зниженої достовірності класифікації.

По-друге, у просторі прогнозів формуються область невизначеності, для яких рішення моделі є нестабільним і чутливим до незначних змін ознак. Такі області характеризуються підвищеним ризиком хибних рішень і фактично утворюють зону невизначеності, яка не враховується у класичній одноетапній схемі класифікації з фіксованим порогом.

Розроблюваний метод підвищення точності виявлення КА поєднує два взаємодоповнювальні механізми. Перший механізм базується на аналізі помилок базової моделі та адаптивному формуванні розширеної навчальної вибірки для проблемних підкласів КА. Другий механізм передбачає введення зони невизначеності прогнозу та селективне перевизначення рішень із використанням допоміжної моделі. Такий підхід дозволяє впливати як на структуру навчальних даних, так і на процедуру прийняття рішення, не змінюючи базову архітектуру перетворення ознак.

Метод розроблено з дотриманням принципу відсутності витоку інформації. Усі етапи адаптації та налаштування виконуються виключно на навчальних і валідаційних даних, тоді як тестова вибірка використовується лише для остаточного оцінювання. Розглянемо формалізацію задачі підвищення точності, опис механізму адаптивного розширення вибірки, введення зони невизначеності та побудову правила селективної корекції рішень у межах соніфікаційного конвеєра систем виявлення КА.

У зв’язку з цим постає задача підвищення точності бінарної класифікації в тих областях простору прогнозів, де базова модель демонструє найбільшу кількість помилок або невпевнені рішення.

Розглядаємо задачу бінарної класифікації мережного трафіку в межах системи виявлення КА, у якій кожен зразок описується вектором ознак та відповідною міткою класу, що відображає належність до нормального або атакуючого трафіку. Базова модель формує для кожного зразка оцінку апостеріорної ймовірності належності до класу КА, після чого остаточне рішення приймається шляхом порівняння цієї оцінки з фіксованим порогом.

У класичній постановці задача навчання зводиться до мінімізації середнього значення функції втрат на навчальній вибірці. Проте така глобальна оптимізація не гарантує рівномірної якості розпізнавання для всіх типів атак. На практиці спостерігається нерівномірний розподіл помилок між підкласами атак, що зумовлено як різною представленістю цих підкласів у вибірці, так і складністю

їхньої внутрішньої структури. У результаті навіть за прийнятного загального рівня точності модель може демонструвати знижену ефективність для окремих типів КА.

Особливо критичними для системи виявлення КА є хибнонегативні рішення, що відповідають пропущеним атакам. Водночас аналіз помилок лише на глобальному рівні не дозволяє визначити, які саме підкласи формують основний внесок у загальну похибку. Тому доцільним є перехід від узагальненої оцінки якості до структурованого аналізу помилок із урахуванням типів КА.

Нехай множина атакуючих зразків поділяється на підкласи, кожен з яких відповідає окремому типу вторгнення. Для кожного підкласу може бути визначено кількість та відносну частоту помилкових рішень базової моделі на навчальній і валідаційній вибірках. Порівняння цих показників дозволяє виявити підкласи, для яких модель демонструє знижений рівень узагальнювальної здатності. Саме такі підкласи розглядаються як проблемні та потребують цілеспрямованої корекції структури навчальних даних.

Окрім структурної нерівномірності помилок, у просторі прогнозів формуються порогові області, у яких оцінка ймовірності наближається до порогового значення прийняття рішення. Для таких зразків характерна знижена впевненість моделі та підвищена чутливість до незначних змін ознак або параметрів навчання. Вказана підмножина утворює зону невизначеності прогнозу, в межах якої стандартне одноетапне порогове правило може бути недостатнім для забезпечення стабільності рішень.

Таким чином, задача підвищення точності бінарної класифікації в системі виявлення КА формулюється як розроблення методу, що забезпечує зменшення загальної кількості помилкових рішень шляхом одночасного впливу на дві складові. По-перше, адаптивну зміну структури навчальної вибірки з урахуванням розподілу помилок між підкласами атак. По-друге, модифікацію процедури прийняття рішення для зразків, що належать до зони невизначеності прогнозу.

Реалізація такого підходу передбачає формалізований аналіз помилок базової моделі та подальшу інтеграцію механізмів адаптації вибірки і селективного перевизначення рішень у межах єдиного алгоритму, що розглядається далі.

Для розв’язання поставленої задачі вводиться математична модель підвищення точності, яка заснована на аналізі помилок базового класифікатора та формалізації зони невизначеності прогнозу.

Запропонований підхід розглядається як надбудова над базовою бінарною моделлю виявлення КА, яка формує для кожного зразка оцінку ймовірності належності до класу КА. Математична модель методу включає два взаємопов’язані компоненти: формалізацію та виявлення зони невизначеності прогнозу за результатами роботи базової моделі без використання тестових даних; побудову допоміжного бінарного класифікатора, навчання якого сфокусовано на зразках із цієї зони та підсилено прикладами типів атак, що найчастіше помиляються у зонах низької впевненості.

1. Базова модель і розподіл прогнозів.

Множину даних для побудови механізму уточнення рішень задаємо так:

$$D_{\text{adapt}} = D_{\text{train}} \cup D_{\text{val}}, \tag{3.65}$$

де D_{train} – навчальна підвибірка; D_{val} – валідаційна підвибірка; $\cup$ – операція об’єднання множин.

Базову модель у межах цього етапу подамо як відображення:

$$f_{\text{base}}: R^d \to [0,1], \tag{3.66}$$

де R^d – d-вимірний простір ознак; d – кількість ознак одного зразка; [0; 1] – інтервал значень виходу моделі, який інтерпретується як оцінка належності до класу КА.

Числову оцінку базової моделі для окремого зразка запишемо так:

$$p_i^{(\text{base})} = f_{\text{base}}(x_i), \tag{3.67}$$

де x_i – вектор ознак i-го зразка; f_{base} – базова класифікаційна модель; x_i – вектор ознак i-го зразка; i – індекс зразка, причому $i = 1,2,\dots,N$; N – кількість зразків у розглядуваній множині даних.

Таке співвідношення формалізує числову оцінку, яка далі трактується як апостеріорна ймовірність належності зразка до класу КА. Для переходу від

неперервної оцінки до дискретного рішення необхідно ввести порогове правило класифікації.

Дискретну мітку базової моделі для окремого зразка визначимо так:

$$\hat{y}_i^{(\text{base})} = \begin{cases} 1, & p_i^{(\text{base})} \geq \theta, \\ 0, & p_i^{(\text{base})} < \theta, \end{cases} \tag{3.68}$$

де θ – поріг прийняття рішення; $\theta \in (0,1)$; $\hat{y}_i^{\text{base}} = 1$ відповідає віднесенню зразка до класу атаки; $\hat{y}_i^{\text{base}} = 0$ – до класу нормальної активності.

Таке правило завершує етап базової класифікації та створює основу для аналізу нестабільних рішень.

У межах моделі методу використовується не тестова, а саме множина D_{adapt}, оскільки визначення параметрів механізму уточнення на тестових даних призводить до витоку інформації та некоректної оцінки узагальнювальної здатності. Така постановка відповідає реальному режиму експлуатації систем виявлення КА, у якому параметри механізму прийняття рішення не можуть підлаштовуватися під майбутній трафік.

2. Визначення τ-зони невизначеності без витоку інформації.

Зона невизначеності вводиться як інтервал значень прогнозу базової моделі, у межах якого рішення є найбільш нестабільним і має підвищену частоту помилок. Параметри τ-зони невизначеності задамо умовою так:

$$0 \leq \tau_L < \theta < \tau_U \leq 1, \tag{3.69}$$

де τ_L – нижня межа зони невизначеності; τ_U – верхня межа зони невизначеності; θ – базовий поріг прийняття рішення; $\tau_L, \theta, \tau_U \in [0,1]$; нерівності $\tau_L < \theta < \tau_U$означають, що базовий поріг лежить усередині інтервалу невизначеності.

Таке співвідношення формалізує область прогнозів, у якій рішення базової моделі є найменш стійкими.

Тоді множину індексів зразків, що належать до зони невизначеності, задамо так:

$$U = \left\{ i \mid \tau_L \le p_i^{(\text{base})} \le \tau_U \right\}, \tag{3.70}$$

де i – індекс зразка; p_i^{base} – прогноз базової моделі для i-го зразка; τ_L і τ_U – нижня та верхня межі зони невизначеності.

Ключовою вимогою є формування інтервалу $[\tau_L, \tau_U]$ на основі емпіричного розподілу прогнозів та помилок базової моделі на D_{adapt}. Практично це означає, що межі інтервалу обираються так, щоб зона була достатньо широкою для забезпечення репрезентативності навчальних даних у області нестабільних рішень. Занадто вузька зона (наприклад, симетричний інтервал навколо θіз малою шириною) призводить до дефіциту прикладів і знижує здатність допоміжної моделі навчитися відокремлювати складні випадки. Натомість помірно широка зона (наприклад, інтерквартильний діапазон відносно «центру» області з максимальною концентрацією помилок) забезпечує компроміс між локалізацією невизначеності та достатнім обсягом даних для подальшого навчання.

Частку зразків, що потрапляють до зони невизначеності, визначимо так:

$$\eta = \frac{| U |}{N_{\text{test}}}, \tag{3.71}$$

де $| U |$ – потужність множини U, тобто кількість індексів, що належать зоні невизначеності; N_{test} – розмір вибірки, на якій оцінюється ця частка.

Співвідношення (3.71) характеризує частку випадків, у яких у робочому режимі активується механізм уточнення рішення допоміжною моделлю.

3. Виявлення помилково складних типів КА у τ-зоні.

Оскільки зона невизначеності описує випадки підвищеної похибки, саме в ній концентрується суттєва частка хибних рішень. Індикатор помилки базової моделі для i-го зразка задамо так:

$$e_i = \begin{cases} 1, & y_i \ne \hat{y}_i^{(\text{base})}, \\ 0, & y_i = \hat{y}_i^{(\text{base})}. \end{cases} \tag{3.72}$$

де y_i – істинна мітка класу i-го зразка; $\hat{y}_i^{\text{base}}$ – рішення базової моделі; $i = 1,2, \dots, N$; значення $e_i = 1$ означає наявність помилки, а $e_i = 0$ – правильну класифікацію.

Нехай множина атакуючих зразків розбивається на підкласи C_k (типи КА), $k = 1, \dots, K$. Тоді для кожного типу атаки визначимо кількість помилок базової моделі в межах τ-зони так:

$$E_k^{(U)} = \sum_{i:\, x_i \in C_k,\ i \in U} e_i, \tag{3.73}$$

де C_k у зоні невизначеності; C_k – k-й підклас атак; k – індекс підкласу, причому $k = 1,2, \dots, K$; K – кількість підкласів атак; $x_i \in C_k$означає, що i-й зразок належить підкласу C_k; $i \in U$означає, що цей зразок потрапляє до зони невизначеності; e_i – індикатор помилки базової моделі.

Сумування дає змогу оцінити внесок кожного підкласу КА у загальну кількість помилок у найскладнішій області прогнозів. На цій основі далі визначається відносна частота помилок для кожного підкласу. Відносну частоту помилок у τ-зоні для підкласу C_k визначимо так:

$$\rho_k^{(U)} = \frac{E_k^{(U)}}{N_k^{(U)}}, N_k^{(U)} = |\ \{\ i \mid x_i \in C_k,\ i \in U\ \}\ |. \tag{3.74}$$

де ρ_k^U – відносна частота помилок базової моделі для підкласу C_k у зоні невизначеності; E_k^U – кількість помилок у підкласі C_kв межах зони U; N_k^U – кількість зразків підкласу C_k, які потрапили до зони невизначеності; $|\cdot|$ – потужність множини; $k = 1,2, \dots, K$.

На основі цих показників сформуємо множину пріоритетних (проблемних) типів КА так:

$$C^* = \{\ C_k \mid \rho_k^{(U)} \geq \gamma,\ N_k^{(U)} \geq N_{\min}\ \}, \tag{3.75}$$

де C^* – множина пріоритетних або проблемних підкласів КА; C_k – k-й підклас КА; ρ_k^U – відносна частота помилок у зоні невизначеності; γ – поріг складності підкласу; N_k^U – кількість зразків підкласу C_kу зоні невизначеності; $N_{\min}$ – мінімально допустима кількість зразків, необхідна для статистично стійкої оцінки.

4. Помилково-орієнтоване підсилення даних у τ-зоні.

На відміну від глобального балансування, у межах методу виконується локальне підсилення саме тих типів КА, які формують основний внесок у помилки в τ-зоні. Підмножину даних, що належить до τ-зони та використовується для локального підсилення, задамо так:

$$D_U = \{(x_i, y_i) \in D_{\text{train}} \mid i \in U\}. \quad (3.76)$$

де (x_i, y_i) – i-й навчальний приклад; D_{train} – навчальна вибірка; $i \in U$ означає, що прогноз базової моделі для цього прикладу належить інтервалу $[\tau_L, \tau_U]$.

Цільовий розмір підкласу C_k у τ-зоні визначимо так:

$$N_{k,U}^* = \min(\alpha_k N_{k,U}, N_{\max}), \quad (3.77)$$

де $N_{k,U}^*$ – цільовий розмір підкласу C_k у зоні невизначеності після підсилення; $N_{k,U}$ – поточна кількість прикладів підкласу C_k у множині D_U; α_k – коефіцієнт підсилення для підкласу C_k, причому $\alpha_k \geq 1$; $N_{\max}$ – максимальне допустиме обмеження на розмір підкласу в зоні; $\min(\cdot,\cdot)$ – оператор вибору меншого з двох значень.

Далі потрібно явно пов'язати коефіцієнт підсилення з рівнем складності відповідного підкласу. Коефіцієнт α_k задається залежно від складності підкласу в τ-зоні, зокрема визначимо його за правилом так:

$$\alpha_k = 1 + \beta \rho_k^{(U)}, \quad (3.78)$$

де β – параметр інтенсивності підсилення, причому $\beta > 0$; ρ_k^U – відносна частота помилок базової моделі для підкласу C_k у зоні невизначеності; $k = 1,2,\dots,K$.

Множину синтетичних зразків для підкласу C_k позначимо $S_{k,U}$, а її потужність визначимо так:

$$|S_{k,U}| = N_{k,U}^* - N_{k,U}. \quad (3.79)$$

де $N_{k,U}^*$ – цільовий розмір підкласу після підсилення; $N_{k,U}$ – поточний розмір підкласу в зоні невизначеності.

Разом із тим підсилення КА не повинно руйнувати глобальне співвідношення між атакувальним і нормальним трафіком у зоні.

Оскільки в τ-зоні важливо не лише підсилити атаки, а й зберегти коректне співвідношення з нормальним трафіком, то введемо обмеження на глобальну частку КА у зоні:

$$\sum_{k\geq 1} N^*_{k,U} \leq \lambda\, N_{0,U}, \tag{3.80}$$

де $N^*_{k,U}$ – цільовий розмір k-го атакувального підкласу в зоні невизначеності після підсилення; сума $\sum_{k\geq 1}$ виконується за всіма підкласами атак, тобто за $k = 1,2,\dots,K$; $N_{0,U}$ – кількість прикладів нормального трафіку в зоні невизначеності; λ – допустимий коефіцієнт дисбалансу, причому $\lambda > 0$.

Таким чином сформуємо адаптовану вибірку для навчання допоміжного класифікатора так:

$$D^{(U)}_{\text{boost}} = D_U \cup \bigcup_{C_k \in C^*} S_{k,U}, \tag{3.81}$$

де D_U – вихідна підмножина навчальних даних, що потрапляє до τ-зони; C^* – множина пріоритетних підкласів КА; $S_{k,U}$ – множина синтетичних зразків для підкласу C_k; $\bigcup_{C_k \in C^*}$ – об'єднання синтетичних множин для всіх проблемних підкласів.

Таке подання формує спеціалізовану навчальну вибірку, зосереджену на складних прикордонних випадках і проблемних типах КА. На цій вибірці далі навчається допоміжний бінарний класифікатор.

5. Допоміжний бінарний класифікатор для τ-зони.

Допоміжну модель у зоні невизначеності подамо як відображення:

$$f_{\text{boost}}\colon \mathrm{R}^d \to [0,1], \tag{3.82}$$

де R^d – d-вимірний простір ознак; d – кількість ознак одного зразка; $[0, 1]$ – інтервал значень виходу моделі.

Числову оцінку допоміжної моделі для окремого зразка визначимо так:

$$p_i^{(\text{boost})} = f_{\text{boost}}(x_i). \tag{3.83}$$

де x_i – вектор ознак і-го зразка; i=1,2,…,N.

Модель f_{boost} навчається на вибірці $D_{\text{boost}}^{(U)}$, тобто на прикладах τ-зони з підсиленими проблемними типами атак. Така постановка забезпечує спеціалізацію допоміжної моделі на хибних порогових випадках та підвищує чутливість до типів атак, що найчастіше помиляються саме в області невизначеності.

6. Селективне перевизначення рішень у τ-зоні.

Остаточне рішення методу формується як композиція базової та допоміжної моделей. Для зразків поза τ-зоною зберігається рішення базової моделі, яке задамо так:

$$\hat{y}_i^{(\text{final})} = \hat{y}_i^{(\text{base})}, i \notin U. \tag{3.84}$$

де $\hat{y}_i^{\text{base}}$ – рішення базової моделі; U – множина індексів зразків у зоні невизначеності; умова $i \notin U$ означає, що прогноз зразка не належить області нестабільних рішень.

Для зразків у τ-зоні застосуємо правило селективного перевизначення з урахуванням порогів упевненості допоміжної моделі:

$$\hat{y}_i^{(\text{final})} = \begin{cases} 1, & p_i^{(\text{boost})} \geq \theta_U, \\ 0, & p_i^{(\text{boost})} \leq \theta_L, \\ \hat{y}_i^{(\text{base})}, & \theta_L < p_i^{(\text{boost})} < \theta_U, \end{cases} \quad i \in U. \tag{3.85}$$

де $\hat{y}_i^{\text{final}}$ – остаточне рішення для i-го зразка; p_i^{boost} – прогноз допоміжної моделі; θ_L – нижній поріг упевненості; θ_U – верхній поріг упевненості; $\theta_L, \theta_U \in [0,1]$; виконується умова $0 \leq \theta_L < \theta < \theta_U \leq 1$; $\hat{y}_i^{\text{base}}$ – рішення базової моделі; $i \in U$ означає, що зразок належить зоні невизначеності.

Таке правило реалізує механізм порогового керування за рівнем упевненості: допоміжна модель змінює рішення базової лише за умови достатньої впевненості, а в неоднозначних випадках зберігається базовий прогноз.

Для контролю ступеня втручання введемо індикатор активації перевизначення так:

$$r_i = \begin{cases} 1, & i \in U \wedge \left(p_i^{(\text{boost})} \leq \theta_L \vee p_i^{(\text{boost})} \geq \theta_U\right), \\ 0, & \text{в інших випадках}, \end{cases} \tag{3.86}$$

де r_i – індикатор того, що для i-го зразка було активовано механізм перевизначення; U – множина індексів зразків у зоні невизначеності; p_i^{boost} – прогноз допоміжної моделі; θ_Lі θ_U – нижній та верхній пороги впевненості; ∧ – логічна операція «і»; ∨ – логічна операція «або».

Таке означення дає змогу кількісно контролювати, для яких саме прикладів допоміжна модель реально втручається у базове рішення. На цій основі далі вводимо інтегральну частку перевизначених рішень.

Частку перевизначених рішень у вибірці визначимо так:

$$\zeta = \frac{1}{N}\sum_{i=1}^{N} r_i\,, \tag{3.87}$$

де N – загальна кількість зразків у вибірці, на якій оцінюється втручання допоміжної моделі; ri – індикатор факту перевизначення для і-го зразка; i=1N – сумування за всіма зразками вибірки; i=1,2,…,N.

Таке співвідношення завершує формалізацію механізму селективного підвищення точності в зоні невизначеності та дає інтегральну оцінку інтенсивності втручання допоміжної моделі. У сукупності формули цього блоку описують повний контур методу, тобто від побудови базового прогнозу та виділення τ-зони до локального помилково-орієнтованого підсилення і перевизначення рішень із пороговим керуванням за рівнем упевненості.

Узагальнено, запропонована математична модель формалізує послідовність кроків методу, наведену на рис. 3.11.

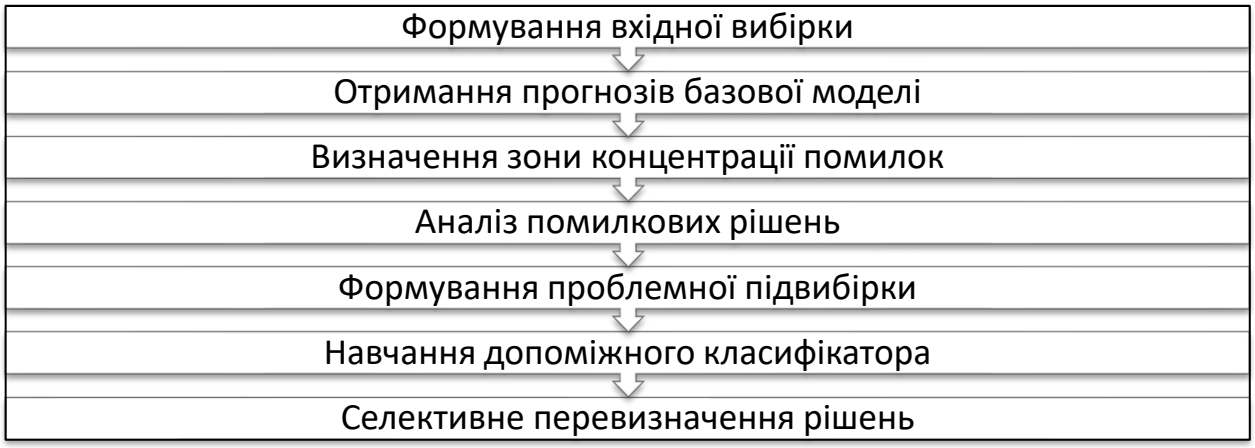

Рисунок 3.11 – Послідовність методу підвищення точності

На рисунку 3.11 зображена алгоритмічна послідовність реалізації методу підвищення точності. Блоки схеми відповідають таким крокам методу: 1) підготовка даних і навчання базової моделі; 2) отримання прогнозів базової моделі на навчальній та валідаційній підвибірках; 3) аналіз помилок і визначення проблемних типів КА; 4) формування τ-зони невизначеності; 5) побудова адаптованої навчальної множини з помилково-орієнтованим підсиленням; 6) навчання допоміжної моделі; 7) селективне перевизначення рішень, що належать до τ-зони та задовольняють умови порогової впевненості; 8) формування остаточного рішення та оцінювання результату.

Отже, схема з рисунку 3.11 конкретизує порядок переходу від первинного прогнозу базової моделі до остаточного рішення каскадної схеми. Така інтерпретація пов’язує графічні блоки з формалізованими нижче етапами методу й показує, що втручання допоміжної моделі є локальним та виконується лише для зразків із підвищеною невизначеністю прогнозу. Це забезпечує локалізоване підвищення точності в найбільш нестабільній області прогнозів без зміни базового соніфікаційного перетворення та без використання тестових даних для налаштування методу. Побудована модель конкретизується у вигляді алгоритму функціонування методу, який визначає порядок селективного перевизначення рішень.

Запропонований метод підвищення точності виявлення КА реалізується як послідовність взаємопов’язаних етапів, що охоплюють побудову базової моделі, аналіз її помилок, локальну адаптацію навчальних даних, навчання допоміжної моделі та селективне перевизначення рішень у зоні невизначеності прогнозу. Усі етапи виконуються з обов’язковим дотриманням розділення навчальної, валідаційної та тестової вибірок, що унеможливлює витік інформації та забезпечує коректність експериментальної оцінки.

На першому етапі здійснюється підготовка даних і навчання базової моделі класифікації. Формуються навчальна та валідаційна підвибірки, після чого виконується оптимізація параметрів моделі на навчальних даних. Після завершення навчання базова модель застосовується до об’єднаної множини навчальних і валідаційних зразків з метою отримання прогнозів і подальшого аналізу структури помилок. Тестова вибірка на цьому етапі не використовується.

Другий етап полягає в аналізі помилок базової моделі на навчальній та валідаційній підвибірках. Для кожного типу атак визначається кількість та відносна частота хибних рішень, зокрема прогнозів в зоні невизначеності. На основі цього аналізу формується множина проблемних типів КА, які демонструють підвищену складність для базової моделі. Саме ці типи КА підлягають подальшому цілеспрямованому підсиленню.

Третій етап передбачає визначення τ-зони невизначеності прогнозу. Межі цієї зони встановлюються на основі розподілу прогнозів базової моделі та концентрації помилок у області нестабільних рішень. Визначення інтервалу виконується виключно за результатами роботи моделі на навчальній та валідаційній вибірках. Таким чином, зона невизначеності відображає реальні області нестабільності прийняття рішення, не використовуючи інформацію з тестових даних. Одночасно контролюється частка зразків, що потрапляють до цієї зони, щоб забезпечити стабільність та обмежити надмірне втручання допоміжної моделі.

Четвертий етап полягає у виконанні помилково-орієнтованого підсилення навчальної вибірки. Для проблемних типів КА, визначених на попередньому етапі,

формується розширене представлення в межах τ-зони. Синтетичні зразки генеруються з урахуванням локальної структури ознак, що забезпечує збереження сигнатурних властивостей КА. Одночасно контролюється співвідношення між атакуючим та нормальним трафіком у зоні невизначеності, щоб уникнути штучного домінування атак у адаптованій вибірці. У результаті формується спеціалізована навчальна множина, орієнтована саме випадки з низькою впевненістю.

П’ятий етап передбачає навчання допоміжної моделі на адаптованій вибірці τ-зони. Структура моделі допоміжної моделі може бути ідентичною базовій, що дозволяє ізолювати вплив саме зміни структури даних. Допоміжна модель набуває підвищеної чутливості до тих типів КА, які формують основний внесок у помилки базової моделі в області невизначеності.

Шостий етап реалізується на стадії експлуатації системи. Для кожного зразка тестової або реальної робочої вибірки спочатку формується прогноз базової моделі. Якщо прогноз не належить до τ-зони, остаточне рішення приймається без змін. Якщо ж зразок потрапляє до зони невизначеності, то додатково обчислюється прогноз допоміжної моделі. Перевизначення рішення відбувається лише за умови достатньої впевненості допоміжної моделі, що реалізує механізм порогового керування за рівнем упевненості. У випадках, коли допоміжна модель не демонструє достатньої впевненості, зберігається початкове рішення базової моделі.

З алгоритмічної точки зору метод поєднує два взаємодоповнюючі процеси: адаптивну зміну структури навчальних даних; локалізовану корекцію рішень у області невизначеності прогнозів. Додаткові обчислювальні витрати пов’язані з навчанням другої моделі та її використанням лише для частини зразків, що належать до τ-зони. Оскільки допоміжна модель застосовується вибірково, зростання складності на етапі прогнозування є пропорційним частці нестабільних випадків.

Таким чином, алгоритм функціонування методу забезпечує послідовну інтеграцію аналізу помилок, локального балансування та селективного перевизначення рішень у межах соніфікаційного конвеєра систем виявлення КА.

Отримана процедура є структуровано визначеною, відтворюваною та придатною до практичної реалізації без модифікації базових етапів перетворення ознак.

У структурі загального соніфікаційного конвеєра систем виявлення КА цей метод виконує роль додаткового коригувального механізму, що застосовується до прогнозів базової моделі.

Запропонований метод підвищення точності розроблено з урахуванням його безпосередньої інтеграції у вже сформований соніфікаційний конвеєр систем виявлення атак. Конвеєр передбачає послідовне перетворення табличних характеристик мережного трафіку у частотно-часове представлення з подальшою класифікацією за допомогою згорткової нейромережної моделі. Запропонований метод не змінює структуру цього перетворення, а функціонує як надбудова на рівні навчання та прийняття рішення.

На рівні обробки ознак усі етапи залишаються незмінними. Вхідні мережні записи проходять нормалізацію, соніфікацію, формування спектрограми та подання у форматі, придатному для двовимірної ЗНМ класифікатора. Базова і допоміжна моделі використовують однаковий механізм перетворення ознак, що гарантує єдність простору представлення даних. Відмінність між ними визначається виключно структурою навчальних вибірок і відповідними параметрами, отриманими в процесі навчання.

Інтеграція методу здійснюється на двох рівнях. Перший рівень це рівень навчання. Після побудови базової моделі на стандартній навчальній вибірці виконується аналіз її помилок на навчальній та валідаційній підвибірках. На основі цього аналізу формується τ-зона невизначеності та виконується локальне помилково-орієнтоване підсилення прикордонних прикладів. Сформована адаптована вибірка використовується для навчання допоміжної моделі. При цьому жодних змін у процедурі соніфікації або спектрального перетворення не вноситься, а модифікується лише розподіл прикладів у навчальному просторі.

Другий рівень це рівень інференсу. У процесі експлуатації кожен зразок проходить стандартний соніфікаційний pipeline і спочатку класифікується базовою моделлю. Якщо отриманий прогноз належить до області високої впевненості,

рішення приймається без додаткової обробки. Якщо ж прогноз потрапляє до τ-зони невизначеності, активується допоміжна модель. Остаточне рішення формується за правилом селективного перевизначення, яке дозволяє змінювати прогноз базової моделі лише у випадках достатньої впевненості допоміжного класифікатора.

Спрощена схема інтеграції методу підвищення точності зображена на рис. 3.12.

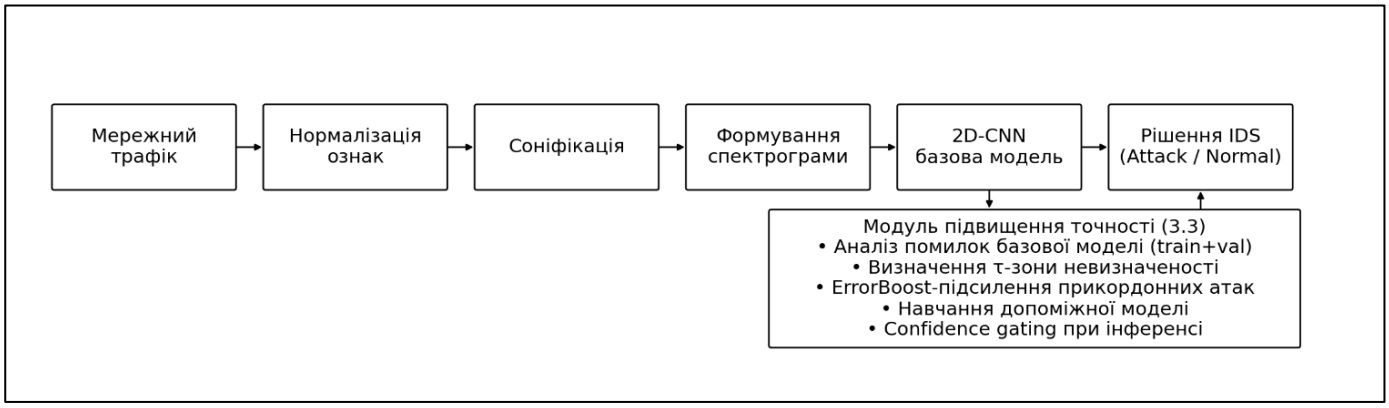

Рисунок 3.12 – Спрощена схема інтеграції методу підвищення точності

З архітектурної точки зору (рис. 3.12) метод не потребує введення нових джерел даних, додаткових сенсорів або змін у механізмах збору мережної інформації. Його реалізація здійснюється на програмному рівні в межах блоку навчання та блоку прийняття рішення. Базовий соніфікаційний конвеєр зберігає свою структуру, а модуль підвищення точності функціонує як розширення логіки класифікації.

Важливою властивістю інтеграції є збереження коректності експериментальної перевірки. Усі параметри, пов’язані з визначенням τ-зони та навчанням допоміжної моделі, встановлюються виключно на основі навчальної та валідаційної вибірок. Тестові дані використовуються лише для фінального оцінювання, що унеможливлює витік інформації та забезпечує об’єктивність отриманих результатів.

Таким чином, метод підвищення точності інтегрується у соніфікаційний конвеєр систем виявлення КА як модуль адаптивної корекції структури навчальних даних і локалізованого уточнення рішень у прикордонних областях простору

прогнозів. Його застосування не потребує структурної перебудови системи та може бути реалізоване в межах існуючої архітектури без зміни базових етапів перетворення ознак.

Разом із тим ефективність запропонованого підходу залежить від низки умов, тому необхідно окремо розглянути його обмеження, чутливість до параметрів та межі коректного застосування в різних сценаріях.

Запропонований метод підвищення точності виявлення КА базується на двох ключових механізмах: локальній адаптації навчальної вибірки; селективному перевизначенні рішень у зоні невизначеності прогнозу. Незважаючи на формалізованість і відтворюваність процедури, його застосування пов’язане з низкою методологічних, статистичних та обчислювальних обмежень, які визначають межі коректного використання методу.

Перше обмеження стосується якості аналізу помилок базової моделі. Формування множини проблемних типів КА базується на емпіричних оцінках частоти помилок на навчальній та валідаційній підвибірках. Якщо ці вибірки є нерепрезентативними або містять значну частку шумових прикладів, то оцінки складності окремих типів КА можуть бути нестабільними. У такому випадку можливе некоректне визначення підкласів для підсилення, що знижує ефективність адаптації. Для зменшення цього ризику доцільно встановлювати мінімальний поріг кількості прикладів для кожного підкласу, нижче якого адаптація не виконується.

Друге обмеження пов’язане з масштабом помилково-орієнтованого підсилення. Надмірне підсилення окремих типів атак може призвести до зміни геометрії розподілу даних у просторі ознак і формування штучних кластерів, які не відображають реальні закономірності мережного трафіку. У крайніх випадках домінування синтетичних прикладів знижує узагальнювальну здатність моделі та підвищує ризик перенавчання. Тому частка синтетичних даних має обмежуватися, а співвідношення між атакуючим і нормальним трафіком — контролюватися під час формування навчальної множини.

Третє обмеження стосується вибору меж τ-зони невизначеності. Занадто вузький інтервал призводить до незначної кількості прикордонних прикладів і,

відповідно, обмеженого впливу допоміжної моделі. Надмірно широка зона, навпаки, розширює область втручання та фактично перетворює допоміжну модель на альтернативний глобальний класифікатор. У такому випадку втрачається локалізований характер методу, а також зростають обчислювальні витрати. Тому інтервал має охоплювати саме області з підвищеною концентрацією помилок, не виходячи за межі прикордонної області.

Четверте обмеження пов’язане з параметрами селективного перевизначення рішень. Якщо правило порогового керування за рівнем упевненості налаштоване занадто агресивно, допоміжна модель починає заміщувати базову для значної частини прикладів, що наближає метод до каскадної або ансамблевої схеми. Якщо ж умови перевизначення є надто жорсткими, допоміжна модель використовується епізодично, і ефект локальної корекції стає мінімальним. Оптимальний режим передбачає обмежену, але стабільну частку перевизначених рішень.

П’яте обмеження стосується узгодженості базової та допоміжної моделей. Оскільки вони навчаються на різних емпіричних розподілах, можливе зростання розбіжностей у прогнозах. Значні систематичні відмінності між оцінками можуть свідчити про надмірну зміну структури навчальних даних або про невдалий вибір параметрів τ-зони. У таких випадках необхідним є коригування коефіцієнтів підсилення та меж прикордонної області.

З обчислювальної точки зору метод передбачає додаткове навчання допоміжної моделі, що збільшує витрати на етапі навчання. Водночас на етапі інференсу допоміжний класифікатор застосовується лише до зразків, що належать до τ-зони, тому зростання складності є пропорційним її розміру і не змінює асимптотичної структури алгоритму. За умови помірної ширини τ-зони метод зберігає прийнятний рівень обчислювальних витрат.

Узагальнено обмеження та властивості методу наведено в табл. 3.2.

Таким чином, якість роботи методу визначається збалансованим вибором параметрів адаптації та меж прикордонної області, а також репрезентативністю навчальних даних. За дотримання зазначених умов метод зберігає локалізований характер втручання у процес класифікації, не потребує структурної модифікації

соніфікаційного конвеєра систем виявлення КА та забезпечує контрольоване зростання обчислювальних витрат.

Таблиця 3.2 – Обмеження та властивості методу підвищення точності

Група властивостей	Характеристика	Потенційний ризик	Умови коректного застосування
Статистичні	Нерепрезентативність навчальних даних	Ненадійне визначення проблемних атак	Мінімальний поріг кількості прикладів, очищення шумових даних
Структурні	Надмірне помилково-орієнтоване підсилення	Домінування синтетичних зразків, перенавчання	Обмеження частки синтетики, контроль співвідношення з нормальними прикладами
Параметричні	Невдалий вибір τ-зони	Або відсутність ефекту, або глобальне заміщення базової моделі	Вибір інтервалу за концентрацією помилок на навчальній та валідаційній вибірках
Логіка прийняття рішення	Невідповідні пороги керування за рівнем упевненості	Надмірне або недостатнє перевизначення	Контроль частки перевизначених рішень
Узгодженість моделей	Сильна розбіжність прогнозів	Нестабільність композиційного рішення	Помірні коефіцієнти підсилення, калібрування порогів
Обчислювальні	Додаткове навчання другої моделі	Зростання часу навчання	Локальне застосування допоміжної моделі лише в τ-зоні

3.4 Висновки до третього розділу

Розроблено взаємопов’язані методи підвищення ефективності виявлення комп’ютерних атак. Метод частотно-часового перетворення забезпечує послідовне нормування ознак, їх кодування у хвильовий сигнал, ІКМ-подання та формування спектрограм за допомогою КПФ. Метод SPAS виконує адаптивне розширення навчальної вибірки в межах 5-го і 95-го процентилів ознак відповідного підкласу,

завдяки чому зберігаються характерні сигнатурні властивості атак. Для складних випадків класифікації сформовано помилково-орієнтоване підсилення проблемних підкласів і τ-зону невизначеності, у межах якої рішення базової моделі може бути селективно уточнене допоміжним класифікатором. Запропонований каскад не використовує тестові дані під час навчання та налаштування, що зберігає незалежність експериментальної оцінки. Розроблені методи створюють цілісний механізм підвищення повноти виявлення рідкісних атак, зменшення кількості хибнонегативних рішень і контрольованого уточнення прогнозів без повної заміни базової моделі.

Основні результати розділу опубліковані у працях [128, 129, 131].

РОЗДІЛ 4
МЕТОДИКА ЕКСПЕРИМЕНТАЛЬНОГО ДОСЛІДЖЕННЯ ТА ОЦІНЮВАННЯ ЕФЕКТИВНОСТІ МОДЕЛІ Й МЕТОДІВ ВИЯВЛЕННЯ КОМП'ЮТЕРНИХ АТАК

4.1 Організація експериментального дослідження

Експериментальне дослідження розроблених методів виконано на наборі даних NSL-KDD, який є однією з найбільш поширених тестових вибірок для задач виявлення КА. Вибір цього набору даних зумовлений кількома причинами. По-перше, NSL-KDD є модифікованою та вдосконаленою версією набору KDD Cup'99, у якій зменшено надлишковість повторюваних записів, що дає змогу отримати більш коректні результати оцінювання моделей. По-друге, набір містить як нормальні мережні з'єднання, так і широкий спектр атак різних типів, що забезпечує можливість дослідження роботи системи в умовах неоднорідного та дисбалансного розподілу класів. По-третє, структура NSL-KDD добре підходить для перевірки як базового соніфікаційного підходу, так і спеціалізованих методів балансування та підвищення точності у зонах невизначеності прогнозу.

У межах проведеного дослідження використано стандартний поділ набору NSL-KDD на навчальну та тестову вибірки. Навчальна вибірка містить 125973 записи, тоді як тестова – 22544 записи. Загальний обсяг використаних даних становить 148517 записів. Така структура дозволяє провести коректне навчання моделей на навчальній частині та незалежне оцінювання якості на тестовій вибірці.

Аналіз наведених на рис. 4.1 даних свідчить, що навчальна вибірка за обсягом істотно перевищує тестову. Це є характерною особливістю набору NSL-KDD і забезпечує достатню кількість зразків для формування моделей машинного навчання. Водночас сама по собі кількісна перевага навчальної вибірки не гарантує однакової представленості всіх типів КА, що зумовлює потребу в окремому аналізі структури класів і дисбалансу.

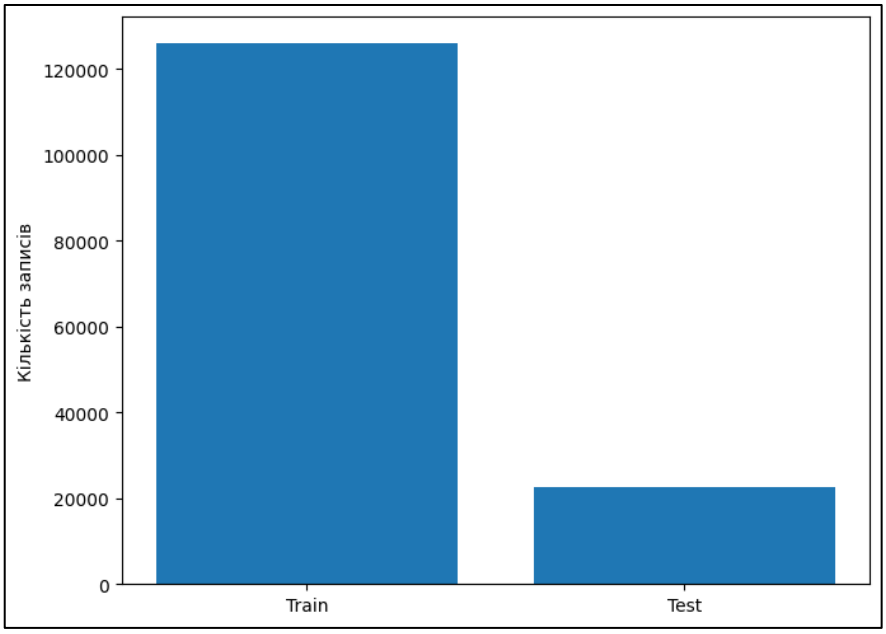

Рисунок 4.1 – Кількість записів у навчальній та тестовій вибірках NSL-KDD

Наступним важливим аспектом організації експериментального дослідження є аналіз розподілу основних класів normal та attack у навчальній і тестовій вибірках. У навчальній вибірці міститься 67343 нормальні записи, що становить 53,46 %, і 58630 записів КА, що становить 46,54 %. У тестовій вибірці, навпаки, кількість КА переважає, бо наявні 12833 записи КА (56,92 %) проти 9711 нормальних записів (43,08 %). Співвідношення основних класів у вибірках зображено діаграмою на рис. 4.2.

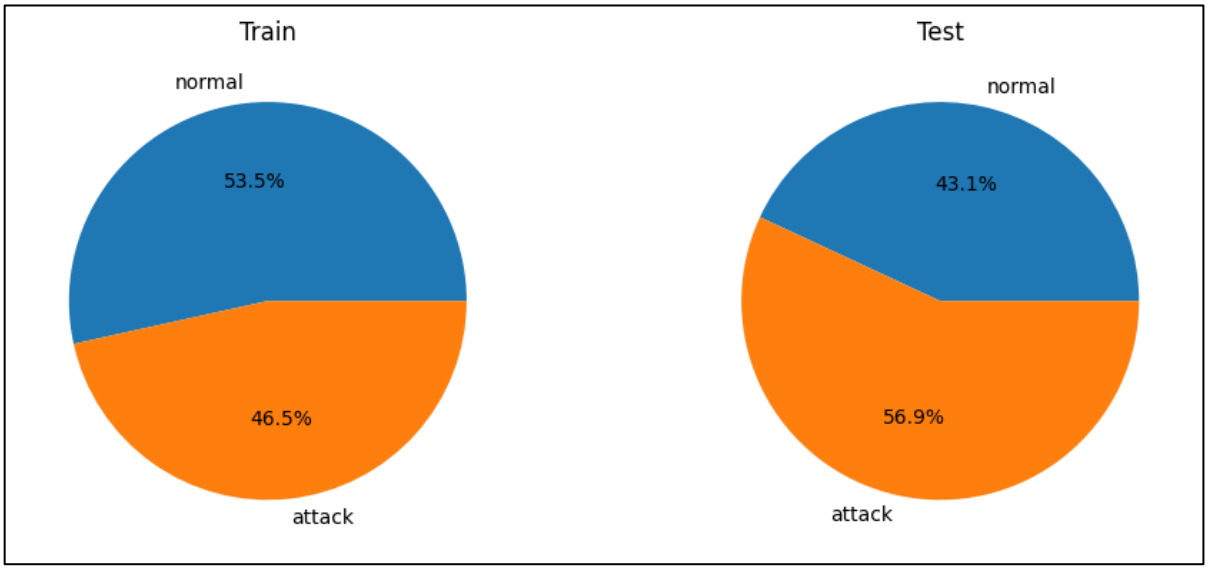

Рисунок 4.2 – Співвідношення нормальної активності та атак у навчальній і тестовій вибірках

Отримані результати показують, що структура навчальної та тестової вибірок не є однаковою. Якщо в навчальній вибірці частка нормального трафіку дещо переважає, то в тестовій, навпаки, більшу частку становлять КА. Така невідповідність є важливою з точки зору оцінювання моделі, оскільки в реальному експерименті система повинна зберігати здатність до узагальнення за умов, коли співвідношення класів у тестових даних відрізняється від навчальних.

Окрему увагу в межах експериментального дослідження слід приділити тому факту, що частина типів КА відсутня у навчальній вибірці, але присутня у тестовій. До таких атак належать: apache2, httptunnel, mailbomb, mscan, named, processtable, ps, saint, sendmail, snmpgetattack, snmpguess, sqlattack, udpstorm, worm, xlock, xsnoop, xterm. Наявність таких типів КА у тестовій вибірці ускладнює задачу класифікації, оскільки модель не має можливості безпосередньо навчитися на відповідних зразках під час тренування. Тестова вибірка NSL-KDD є складнішою за навчальну не лише через зміну часток класів, а й через наявність нових типів атак. Це підвищує вимоги до узагальнювальної здатності моделі та додатково обґрунтовує доцільність застосування спеціалізованих методів підвищення точності розпізнавання.

Ще однією суттєвою характеристикою набору NSL-KDD є виражений дисбаланс між окремими типами КА. У навчальній вибірці найбільшу кількість записів має КА neptune – 41214 записів, що становить 70,30 % усіх КА. Значно менше представлені типи satan (3633, або 6,20 %), ipsweep (3599, або 6,14 %), portsweep (2931, або 5,00 %), smurf (2646, або 4,51 %) та nmap (1493, або 2,55 %). Таким чином, уже в межах навчальної вибірки спостерігається значна нерівномірність розподілу між підкласами КА.

Як видно з рис. 4.3, домінування КА neptune є дуже значним. Це означає, що без застосування додаткових методів балансування модель може виявитися надмірно орієнтованою саме на наймасовіші типи КА, тоді як рідкісніші підкласи розпізнаватимуться істотно гірше.

У тестовій вибірці також спостерігається виражений дисбаланс, однак її структура відрізняється від навчальної. Найбільш поширеною КА знову є neptune

– 4657 записів, або 36,29 % усіх КА. Далі йдуть guess_passwd (1231, або 9,59 %), mscan (996, або 7,76 %), warezmaster (944, або 7,36 %), apache2 (737, або 5,74 %) та satan (735, або 5,73 %).

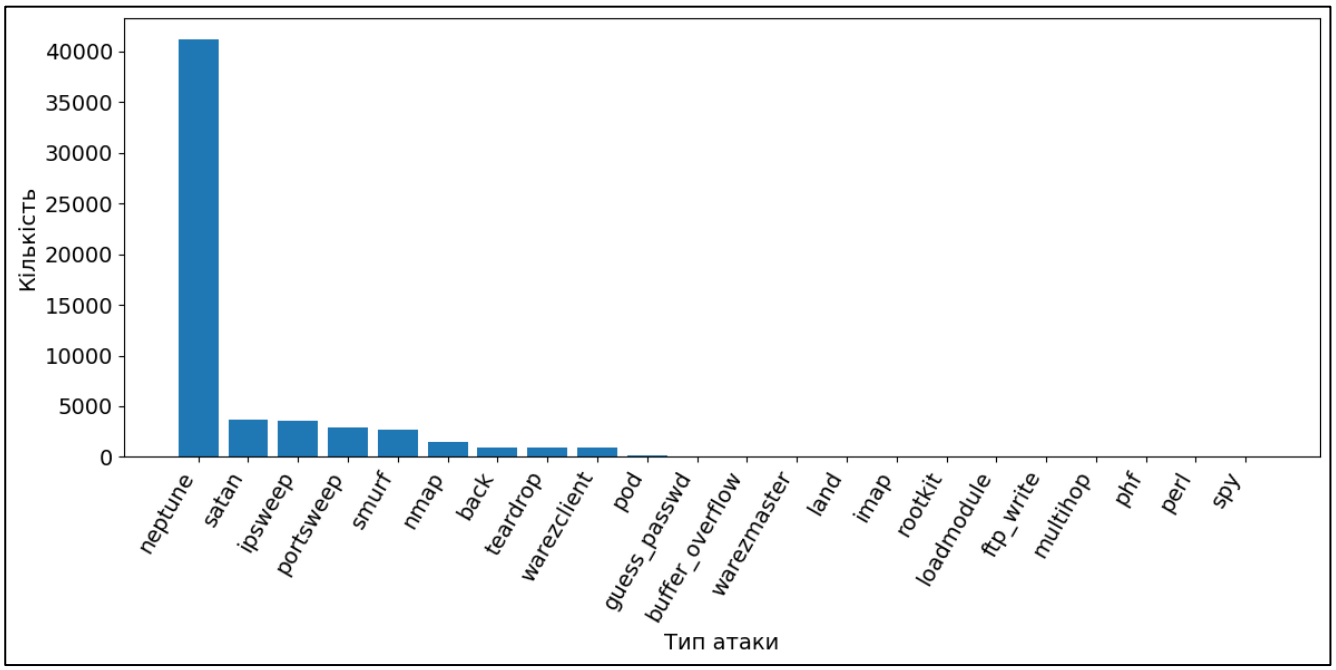

Рисунок 4.3 – Розподіл типів КА у навчальній вибірці

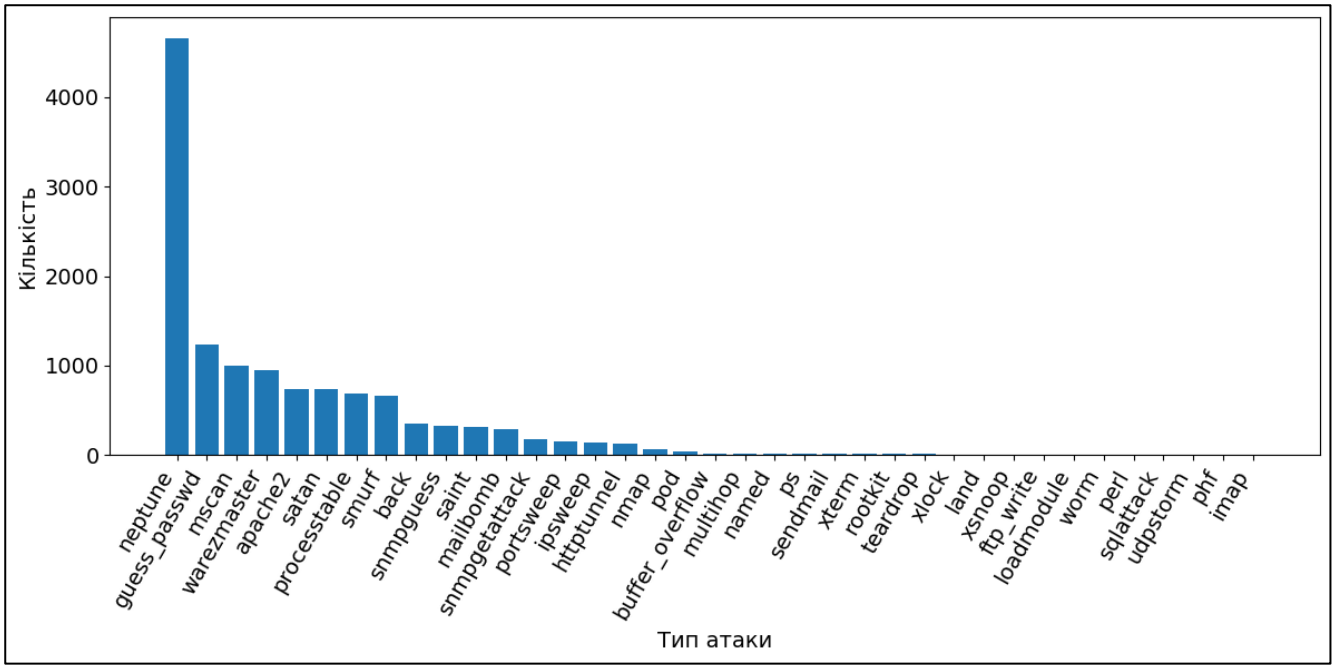

Рисунок 4.4 – Розподіл типів КА у тестовій вибірці

Тому не лише співвідношення normal / attack, але й внутрішня структура підкласів КА у навчальній та тестовій вибірках істотно відрізняються. Наприклад,

у тестовій вибірці серед найпоширеніших КА з’являються guess_passwd, mscan, warezmaster, apache2, які або не були домінуючими у навчальній вибірці, або взагалі були відсутні в ній. Це створює додаткові труднощі для побудови ефективної моделі виявлення КА.

Для додаткового унаочнення структури домінантних підкласів КА в додатку Г наведено детальніше подання найпоширеніших типів атак у навчальній та тестовій вибірках NSL-KDD (рисунки Г.1–Г.2).

Аналіз повного розподілу підтверджує наявність рідкісних КА, кількість зразків яких є дуже малою порівняно з домінуючими підкласами. Саме ця умова є однією з ключових причин зниження якості класифікації в задачах виявлення вторгнень і обґрунтовує необхідність застосування методів сигнатурозбережного балансування та додаткових механізмів підвищення точності.

Отже, організація експериментального дослідження базується на використанні набору даних NSL-KDD, який характеризується достатнім загальним обсягом записів, наявністю як нормального трафіку, так і багатьох типів КА, відмінністю структури навчальної та тестової вибірок, а також вираженим дисбалансом підкласів КА. Виявлені особливості набору даних безпосередньо впливають на складність задачі класифікації та визначають необхідність розроблення і дослідження спеціалізованих методів підвищення ефективності виявлення КА на основі соніфікації мережного трафіку.

4.2 Дослідження базового соніфікаційного підходу та порівняння з векторною моделлю

На цьому етапі виконано експериментальну перевірку базового підходу до подання мережного трафіку у вигляді аудіосигналів (соніфікація) з подальшим перетворенням у спектрограми та застосуванням згорткової нейромережі (двовимірної ЗНМ). Метою цього етапу не є підвищення точності розпізнавання КА з будь-якими витратами, а формування та обґрунтування альтернативного каналу подання даних, який може бути використаний у подальшому для інтеграції

методів балансування та підвищення точності в зоні невизначеності прогнозу. Отже, на цьому етапі перевірятимемо право на існування соніфікаційного подання як повноцінного варіанту представлення мережних з'єднань для виявлення КА.

Експерименти проведено на наборі даних NSL-KDD, структуру якого наведено раніше. Для даного етапу використано всі наявні записи: навчальна вибірка містить 125973 з'єднання; тестова – 22544 з'єднання. Мітки класів зведено до бінарної постановки, тобто normal відповідає класу «0», а всі типи КА – класу «1».

Попередня обробка вхідних даних виконувалася однаково для обох моделей. Категоріальні ознаки protocol_type, service, flag перетворено за допомогою one-hot кодування, після чого застосовано нормалізацію MinMaxScaler до діапазону $[-1; 1]$. У результаті розмірність векторного представлення одного мережного запису становить 122 ознаки.

Експериментальне дослідження базового підходу доцільно розпочати з опису формування соніфікованого подання мережного трафіку та побудови спектрограм, які використовуються як вхідні дані для моделі.

Соніфікаційний канал подання даних реалізовано як перетворення нормованого вектору ознак у ІКМ-сигнал фіксованої довжини (0,4 с за частоти дискретизації 2000 Гц). Кожна ознака впливає на параметри синусоїдального фрагмента (частота та амплітуда), а отриманий аудіосигнал подається на перетворення короткочасного Фур'є, результатом якого є спектрограма, що використовується як вхідний тензор для двовимірної ЗНМ. У виконаному експерименті сформовано спектрограми розміру 129×8 (та додатковим каналом 1), тобто вхід до згорткової мережі має форму 129×8×1.

На рис. 4.7 зображено приклад хвильового представлення та спектрограми. Додатковий приклад нормального запису після соніфікації та порівняння подібності спектрограм однотипних атак наведено на рис. Г.4–Г.5.

Після визначення способу подання даних необхідно задати параметри моделей і умови їх навчання, що забезпечує коректність подальшого порівняння

результатів. Для порівняння ефективності соніфікаційного представлення використано дві моделі:

1) векторна модель (MLP) без соніфікації, яка працює безпосередньо з one-hot кодуванням і нормалізованими ознаками;

2) соніфікована модель (двовимірної ЗНМ), яка працює зі спектрограмами, отриманими після ІКМ+КПФ.

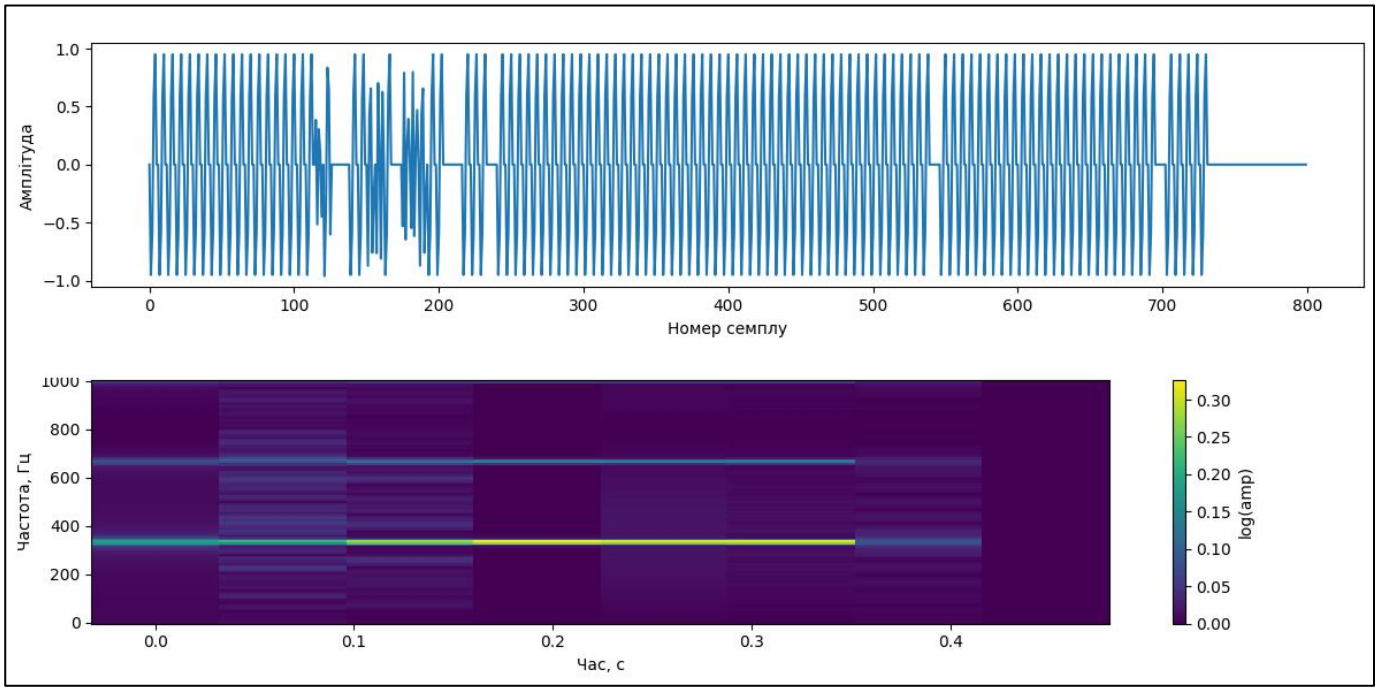

Рисунок 4.7 – Запис атаки після соніфікації: хвильове представлення та спектрограма

Основні параметри моделей наведено в табл. 4.6.

На цій основі виконується порівняння результатів векторної та соніфікованої моделей, що дає змогу оцінити вплив частотно-часового подання на якість класифікації.

Оцінювання якості виконано на тестовій вибірці NSL-KDD за метриками точності, прецизійності, повноти та F1-міри, а також додатково обчислено частку хибнопозитивних рішень (FPR) та частку хибнонегативних рішень (FNR) як важливі характеристики для СВВ. Результати соніфікованої двовимірної ЗНМ наведено у табл. 4.7, а порівняння з векторною моделлю – у табл. 4.8.

Таблиця 4.6 – Параметри моделей та спосіб подання вхідних даних

Модель	Подання вхідних даних	Розмір входу	Структура моделі	Кількість епох	Розмір пакета	Швидкість навчання	Оптимізатор	Функція втрат
Векторна модель	One-Hot + MinMaxScaler(-1,1)	122	Dense(128)–Dense(64)–Dense(1)	5	128	0,0005	Adam	binary_crossentropy
Соніфікована модель	One-Hot + MinMaxScaler(-1,1) + ІКМ + КПФ	129×8×1	Conv2D(16)–MaxPool–Conv2D(32)–GlobalMaxPool–Dense(32)–Dense(1)	5	128	0,0005	Adam	binary_crossentropy

Для деталізації результатів базового порівняння у додатку Г подано матрицю невідповідностей соніфікованої моделі, графіки її навчання, розподіл правильних і помилкових рішень за інтервалами прогнозованої ймовірності, а також аналогічні матеріали для векторної моделі (рисунки Г.6–Г.7, Г.9–Г.12; таблиця Г.1).

Таблиця 4.7 – Результати соніфікованої моделі (двовимірної ЗНМ) на тестовій вибірці

Модель	Точність	Прецизійність	Повнота	F1-міра	FPR	FNR
Соніфікована двовимірна ЗНМ	0,7741	0,9346	0,6486	0,7658	0,0599	0,3514

Таблиця 4.8 – Порівняння метрик векторної та соніфікованої моделей

Модель	Точність	Прецизійність	Повнота	F1-міра	FPR	FNR
Векторна	0,7766	0,9249	0,6613	0,7712	0,0710	0,3387
Соніфікована	0,7741	0,9346	0,6486	0,7658	0,0599	0,3514

З табл. 4.8 видно, що обидві моделі демонструють порівнянний рівень точності в умовах однакової постановки бінарної класифікації нормальної активності та атаки. При цьому соніфікована модель має дещо менший показник точності порівняно з векторною (0,7741 проти 0,7766), однак демонструє нижчу частку хибнопозитивних рішень (FPR) (0,0599 проти 0,0710), тобто рідше помилково класифікує нормальний трафік як атакувальний. Водночас частка хибнонегативних рішень (FNR) соніфікованої моделі є дещо вищим (0,3514 проти 0,3387), що вказує на актуальність подальших методів підвищення повноти розпізнавання атак у складних областях простору ознак.

Таким чином, отримані результати підтверджують, що соніфікаційний канал подання даних є конкурентним щодо векторного подання в задачі виявлення КА і може розглядатися як повноцінна основа для подальших експериментів із балансуванням та селективним підвищенням точності.

Окремо оцінено витрати часу на підготовку даних, навчання та інференс для векторної та соніфікованої моделей. Результати наведено в табл. 4.9 і проілюстровано на рис. 4.11. Графічне зіставлення часу навчання моделей додатково наведено на рисунку Г.8.

Таблиця 4.9 – Порівняння витрат часу для векторної та соніфікованої моделей

Модель	Час підготовки даних, с	Час навчання, с	Час інференсу на тестовій вибірці, с	Сумарний час, с
Векторна	0,00	25,75	2,42	28,17
Соніфікована	524,47	221,03	14,51	760,00

Згідно даних з табл. 4.9, соніфікаційний підхід потребує додаткового часу на генерацію ІКМ-сигналів і спектрограм (524,47 с), а також має більший час навчання моделі (221,03 с) і інференсу на тестовій вибірці (14,51 с) порівняно з векторною моделлю. Таким чином, соніфікований канал подання є більш ресурсомістким, однак забезпечує перехід до двовимірного простору представлення, у якому можуть бути ефективно виявлені локальні й структурні шаблони засобами згорткових мереж. Саме тому на наступних етапах доцільно розглядати його як базову платформу для інтеграції спеціалізованих методів (зокрема балансування та селективного перевизначення рішень у зоні невизначеності).

Отримані результати (рис. 4.16) підтверджують працездатність базового соніфікаційного підходу та дають підстави перейти до дослідження спеціалізованих методів підвищення точності виявлення атак.

Проведене експериментальне дослідження підтвердило, що соніфікаційний канал подання мережного трафіку (ІКМ+КПФ) дозволяє сформувати структуроване двовимірне представлення, придатне для аналізу двовимірної ЗНМ, і забезпечує зіставний із векторним поданням рівень якості в задачі бінарного виявлення атак. Отримані результати обґрунтовують доцільність подальшого використання соніфікаційної моделі як базової для експериментів із балансуванням навчальної вибірки та підвищенням точності в зоні невизначеності прогнозу.

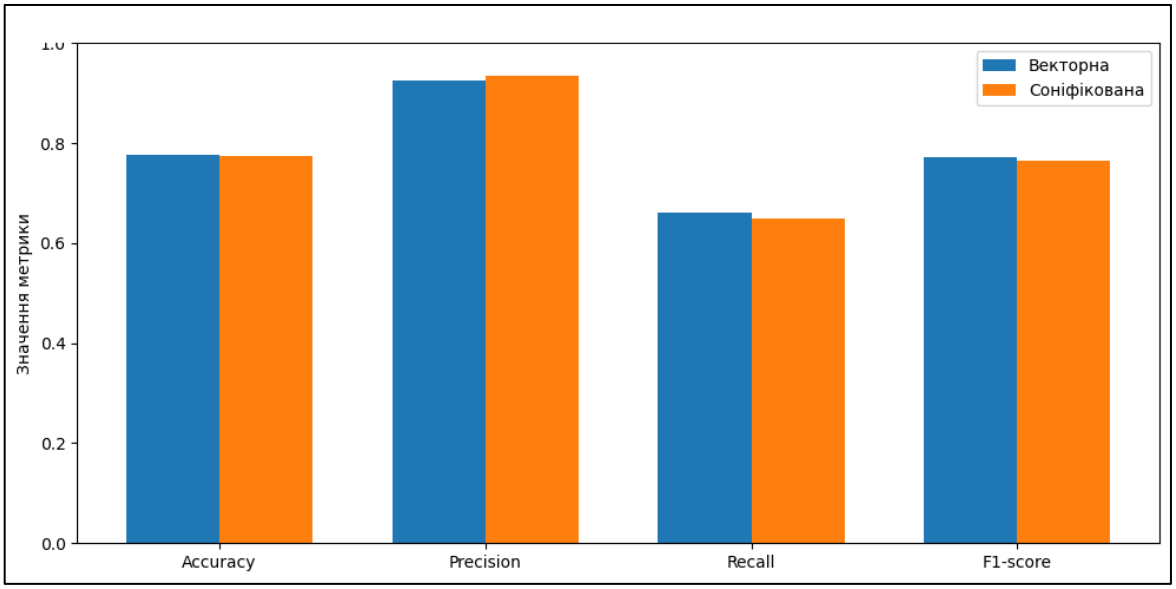

Рисунок 4.16 – Порівняння основних метрик моделей

4.3 Дослідження методу сигнатурозбережного балансування навчальної вибірки (SPAS) та порівняння з SMOTENC

Аналіз набору NSL-KDD показав, що він має виражений дисбаланс не лише між класами normal/attack, а й усередині класу атак. Частина підтипів представлена десятками тисяч записів (наприклад, neptune), тоді як інші підкласи містять сотні або навіть одиничні зразки. У таких умовах стандартне навчання нейромережевого класифікатора призводить до зміщення оптимізації функції втрат у бік домінуючих атак, що підвищує точність на масових підкласах, але погіршує повноту виявлення рідкісних атак (зростання FN). З практичної точки зору для систем виявлення КА критичним є саме зменшення кількості пропущених КА, тому балансування навчальної вибірки розглядається як один із базових інструментів підвищення чутливості системи.

Метою цього етапу є експериментальне дослідження впливу балансування на якість соніфікованої моделі (ІКМ, КПФ, двовимірної ЗНМ) та порівняння трьох режимів навчання:

1) базовий режим (без балансування) – навчання на початковій навчальній частині вибірки;

2) SMOTENC – балансування класичним методом синтетичного надсемплінгу з підтримкою категоріальних ознак;

3) SPAS – запропонований метод сигнатурозбережного адаптивного формування синтетичних зразків, який обмежує їх у межах допустимих профілів КА на основі процентильних меж, щоб уникнути генерації некоректних точок у просторі ознак.

Для забезпечення порівнюваності експерименту в усіх режимах застосовано однаковий соніфікаційний конвеєр, однакову архітектуру двовимірної ЗНМ та однакові параметри навчання (5 епох, розмір пакета 128, Adam, швидкість навчання 0,0005). Балансування застосовувалося лише до навчальної частини вибірки, тоді як валідаційна та тестова множини залишалися незмінними.

Подальше вдосконалення моделі потребує врахування дисбалансу підтипів атак, який істотно впливає на повноту виявлення рідкісних і складних для класифікації прикладів. Для кількісного підтвердження дисбалансу підтипів КА у навчальній вибірці побудовано розподіл частот КА до застосування балансування.

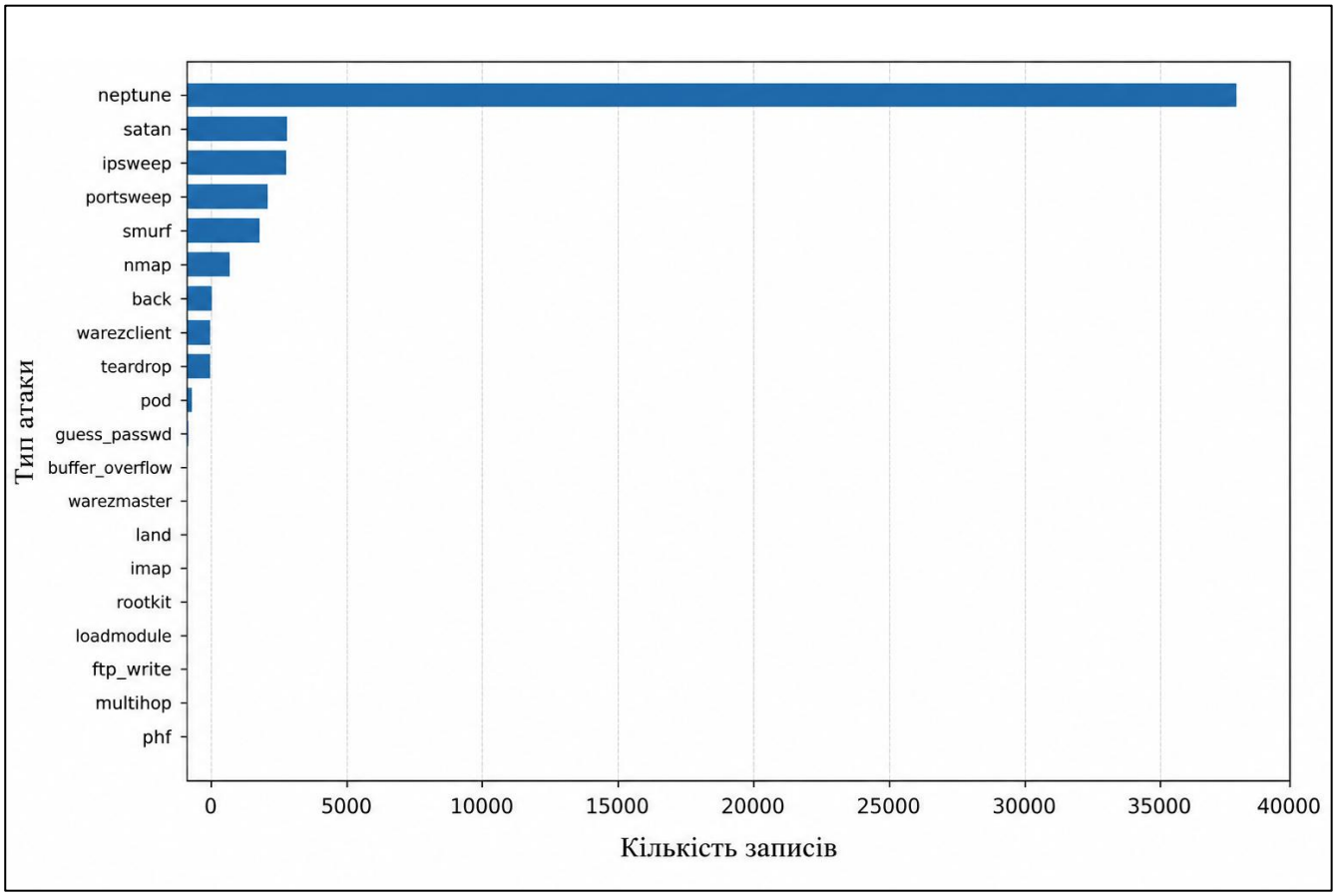

Рисунок 4.13 – Найпоширеніші типи КА у навчальній вибірці (до балансування)

З рис. 4.13 встановлено домінування КА neptune та значне відставання інших підтипів. Такий дисбаланс формує ситуацію, коли модель під час навчання розглядає значно більше прикладів одного підкласу та істотно менше – інших, що без спеціальних заходів призводить до зниження повноти виявлення рідкісних атак. Відтак задача балансування формулюється як збільшення представленості міноритарних підкласів атак у навчальній вибірці без зміни тестової вибірки та без розмивання статистичних характеристик КА. З огляду на це формується план балансування навчальної вибірки та визначаються цільові підкласи КА, для яких

така процедура є найбільш доцільною. Додаткову наочність щодо зміни розподілу найпоширеніших типів атак до та після застосування SPAS забезпечують рисунки Г.13–Г.14.

Балансування виконувалося за попередньо сформованим планом надсемплінгу: для обраних підтипів атак задається цільова кількість зразків, визначена множником із додатковим обмеженням на мінімальну кількість і верхньою межею, щоб не перевищувати масштаб найбільшої КА. Такий підхід дає змогу збільшити представленість рідкісних класів до порівнюваних обсягів, не перетворюючи задачу на повне вирівнювання.

Таблиця 4.11 – План балансування для цільових КА (фрагмент)

№	Тип КА	Поточна кількість	Цільова кількість	Множник
1	back	846	4230	5,00
2	guess_passwd	50	1000	20,00
3	buffer_overflow	25	1000	40,00
4	warezmaster	17	1000	58,82
5	rootkit	10	1000	100,00
6	loadmodule	9	1000	111,11
7	multihop	7	1000	142,86

План балансування є спільною основою для порівняння SMOTENC та SPAS: обидва методи намагаються досягти співставних цільових обсягів, але відрізняються способом генерації синтетичних зразків і контролем їх відповідності.

Реалізація цього етапу спирається на метод SPAS, який передбачає сигнатурозбережне розширення навчальних даних із контролем допустимих меж модифікації ознак.

Ключова відмінність SPAS від класичного SMOTE-підходу полягає в тому, що SPAS формує синтетичні зразки в межах допустимої області значень ознак, характерної для конкретного підкласу КА. Методологічно такі межі формуються як процентильні інтервали: для кожної числової ознаки на множині реальних

прикладів відповідного підкласу обчислюються нижня та верхня межі q_low і q_high. У проведеному експерименті використано інтервал q = (0,05; 0,95), тобто значення синтетичного зразка після інтерполяції обмежується діапазоном між 5-м і 95-м процентилями. Це зменшує ризик появи синтетичних точок, які формально збільшують вибірку, але виходять за межі характерного профілю атаки. Для категоріальних ознак protocol_type, service та flag застосовується копіювання значень з одного з батьківських прикладів, що дає змогу зберігати структурні сигнатури підкласу КА. Для ілюстрації процентильних меж на рис. 4.14 наведено приклад для атаки guess_passwd за вісьмома обраними ознаками.

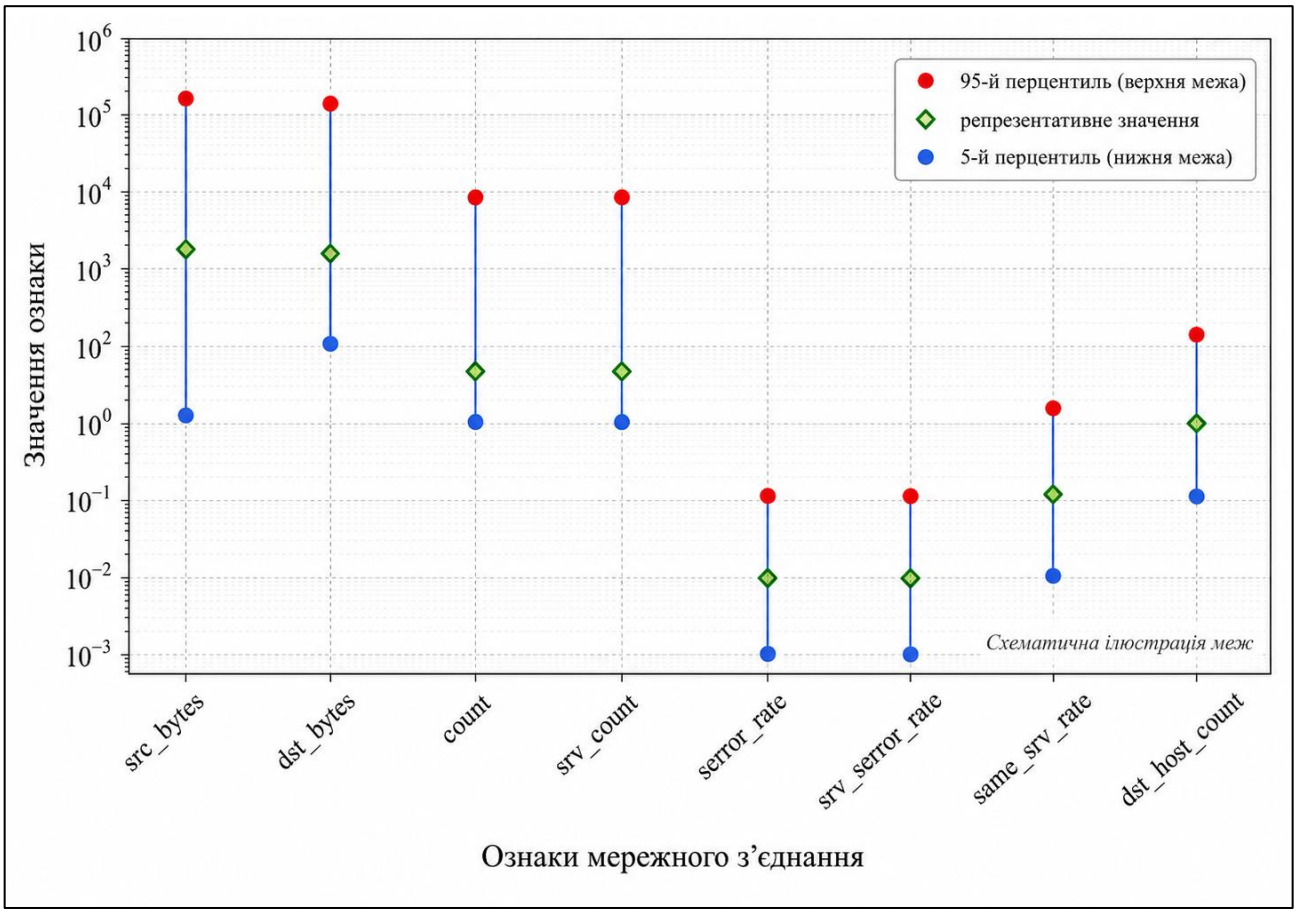

Рисунок 4.14 – Процентильні межі для КА guess_passwd на прикладі 8-ми ознак

З табл. 4.12 встановлено, що для частини ознак інтервали є дуже вузькими, наприклад для dst_bytes (кількість байтів, переданих від призначення до джерела) або same_srv_rate. Це свідчить про наявність стабільного профілю відповідної КА.

Саме такі випадки є критичними для сигнатурозбережного балансування: неконтрольований синтез може породжувати значення, не властиві цьому підкласу атаки, тоді як SPAS утримує синтетичні зразки в характерному діапазоні. Отже, SPAS не лише збільшує кількість прикладів міноритарних КА, а й зменшує ризик появи статистично некоректних синтетичних зразків, які можуть погіршувати узагальнювальну здатність моделі.

Таблиця 4.12 – Процентильні межі для атаки guess_passwd на прикладі 8-ми ознак (SPAS)

Ознака	нижня межа	верхня межа
src_bytes	125,00	126,00
dst_bytes (кількість байтів, переданих від призначення до джерела)	179,00	179,00
count	1,00	2,00
srv_count	1,00	2,00
serror_rate	0,00	0,275
srv_serror_rate	0,00	0,275
same_srv_rate	1,00	1,00
dst_host_count	2,45	49,55

Ефективність запропонованого підходу оцінюється у порівняльному експерименті з базовою моделлю та методом SMOTENC.

Експеримент виконано у трьох режимах (базова модель / SMOTENC / SPAS) на однаковій тестовій множині. Основні результати наведено в табл. 4.13, а компоненти матриці невідповідності – у табл. 4.14.

Таблиця 4.13 – Порівняння якості соніфікованої моделі без балансування, з SMOTENC та з SPAS

Режим	Точність	Прецизійність	Повнота	F1-міра	FPR	FNR	TN	FP	FN	TP
Базовий режим	0,7566	0,9615	0,5963	0,7361	0,0315	0,4037	9405	306	5181	7652
SMOTENC	0,7949	0,9784	0,6541	0,7840	0,0191	0,3459	9526	185	4439	8394
SPAS	0,8457	0,9225	0,7958	0,8545	0,0884	0,2042	8853	858	2621	10212

Таблиця 4.14 – Порівняння матриць невідповідностей для трьох режимів

Режим	TN	FP	FN	TP
Базовий режим	9405	306	5181	7652
SMOTENC	9526	185	4439	8394
SPAS	8853	858	2621	10212

Ключові спостереження за результатами табл. 4.13–4.14:

1) базовий режим (без балансування) забезпечує високу прецизійність (0,9615), але низьку повноту (0,5963), що означає те, що модель рідко помилково сигналізує КА (низьку кількість хибнопозитивних рішень (FP)), проте пропускає значну частину КА (FN=5181) і це є небажаним для систем виявлення КА;

2) SMOTENC демонструє збалансованіший компроміс, при якому повнота зростає до 0,6541, кількість хибнонегативних рішень (FN) зменшується до 4439 і при цьому прецизійність залишається дуже високою (0,9784), а кількість хибнопозитивних рішень (FP) зменшується до 185, то такий ефект може бути наслідком синтетики або зміни внутрішньої межі класифікації після балансування;

3) SPAS показує найбільший приріст повноти до 0,7958 та найменшу кількість хибнонегативних рішень (FN) (2621), тобто забезпечує суттєве підвищення повноти виявлення атак і при цьому разом із цим спостерігається

збільшення кількості хибнопозитивних рішень (FP) (858), що відображається в зростанні частки хибнопозитивних рішень (FPR) (0,0884), отже, SPAS підсилює чутливість детектора, але може збільшувати кількість помилкових спрацювань.

Проведене дослідження демонструє, що збалансування вибірки здатне суттєво знизити кількість хибнонегативних рішень (FN), але виникає побічний ефект зростання FP. Саме тому далі вводиться механізм підвищення точності в зоні невизначеності прогнозу та селективного перевизначення рішень, який повинен утримувати переваги SPAS щодо повноти, зменшуючи кількість хибнопозитивних рішень (FP).

Поряд із метриками якості необхідно врахувати обчислювальні витрати та вплив кожного варіанта балансування на загальний конвеєр обробки даних. Оскільки соніфікаційний підхід передбачає генерацію спектрограм короткочасного перетворення Фур'є, обчислювальні витрати мають значення для відтворюваності експерименту. У табл. 4.15 наведено тривалість формування спектрограм та навчання моделей у трьох режимах.

Таблиця 4.15 – Порівняння обчислювальних витрат формування спектрограм та навчання

Режим	Кількість навчальних зразків	Час КПФ (навч./валід./тест.), с	Час навчання, с
Базовий режим (без балансування)	113375	545,22	324,76
SMOTENC	125627	350,09	357,93
SPAS	125627	350,64	309,24

З табл. 4.15 встановлено, що використання балансування збільшує обсяг навчальної вибірки (113375 → 125627). Водночас у виконаному експерименті час КПФ для SMOTENC та SPAS є меншим, ніж у базовому режимі. Це може пояснюватися особливостями виконання обчислень у конкретному середовищі

(кешування, стабілізація продуктивності під час повторних викликів, варіації навантаження). Загалом ці дані підтверджують, що балансування не призводить до критичного зростання часу навчання та може бути використане у межах запропонованого конвеєра.

Отже, застосування балансування SPAS підтверджує доцільність сигнатурозбережного підходу як засобу підвищення якості виявлення проблемних підкласів атак без істотного погіршення загальної якості моделі.

4.4 Дослідження методу підвищення точності в зоні невизначеності прогнозу

Наступний етап експериментального дослідження пов’язаний з аналізом прогнозів моделі, навченої на збалансован методом SPASій вибірці, та формуванням паспортів прогнозів для подальшого виявлення проблемних зон.

Як базову модель першого рівня використано соніфіковану двовимірної ЗНМ, навчення якої проводиться на навчальній множині, сформованій методом SPAS. На початковому кроці формується план балансування та виконується генерація синтетичних зразків атак, після чого кількість навчальних прикладів становить 125627.

Після навчення базової моделі виконано оцінювання на тестовій вибірці. Базова модель забезпечила такі результати: точність = 0,8212, прецизійність = 0,9331, повнота = 0,7389, F1-міра = 0,8247.

Для подальших етапів каскадної схеми для навчальної та валідаційної множин сформовано паспорти прогнозів, які містять істинні мітки, бінарні рішення базової моделі та прогнозовані ймовірності. Ці паспорти використовуються:

а) для аналізу помилок та вибору проблемних підкласів КА;

б) для автоматизованого підбору параметрів τ-зони та порогів селективного перевизначення.

На основі цього аналізу визначаються проблемні підкласи КА і формується вибірка з помилково-орієнтованим підсиленням для додаткового навчання допоміжної моделі.

Проблемні підкласи атак визначаються за кількістю помилок базової моделі на навчальній та валідаційній вибірках. Відбір таких підкласів дозволяє цілеспрямовано підсилити навчальну множину саме там, де базова модель демонструє найбільшу кількість помилкових або прикордонних рішень.

На основі відібраних проблемних підкласів сформовано план помилково-орієнтованого підсилення (рис. 4.15, рис. 4.16). Їхня представленість збільшується шляхом додаткового генерування синтетичних зразків методом SPAS. У результаті обсяг навчальної множини для додаткової моделі істотно зростає та становить 469050 прикладів.

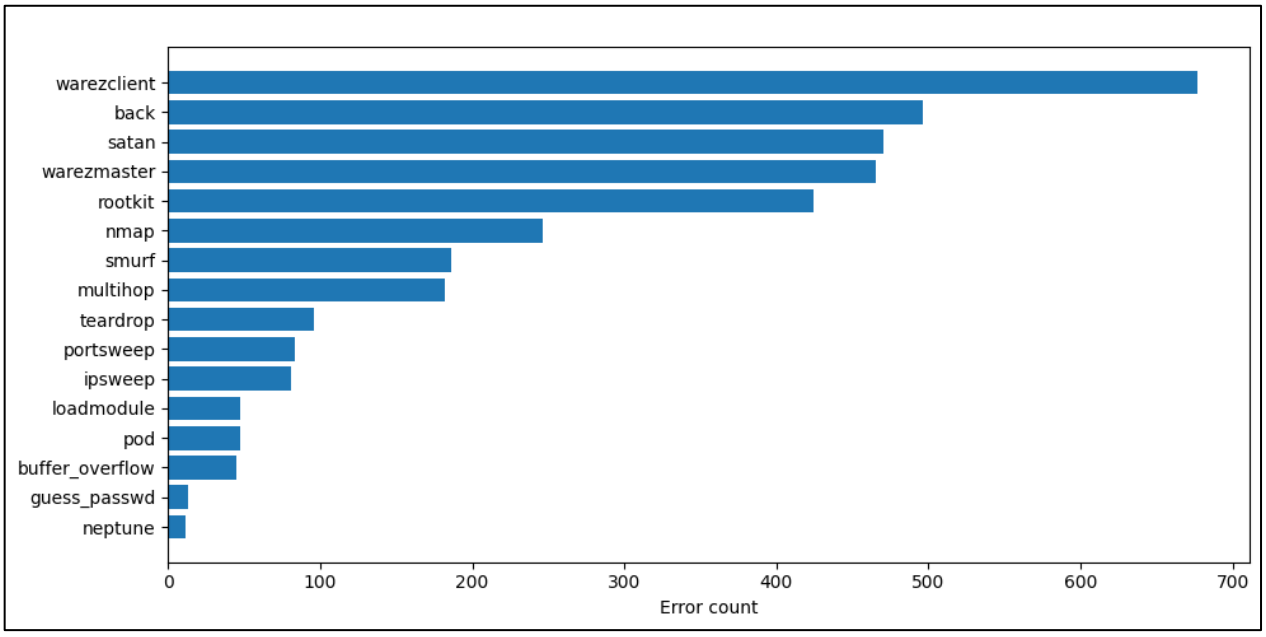

Рисунок 4.15 – Найбільш помилкові підкласи КА для базової моделі (навчальна та валідаційна вибірки)

Додаткова модель у запропонованій схемі не є альтернативою базовій моделі. Вона розглядається як спеціалізований класифікатор, орієнтований на зменшення помилок у підкласах, де базова модель має найбільшу частку неправильних рішень. Сформована вибірка використовується для побудови допоміжної моделі, призначеної для уточнення рішень у найбільш складних випадках класифікації.

Додаткова модель навчалася на вибірці з помилково-орієнтованим підсиленням (SPAS + цільове підсилення проблемних підкласів). На тестовій

множині додаткова модель досягла: Точність = 0,8418, Прецизійність = 0,8827, Повнота = 0,8328, F1-міра = 0,8570.

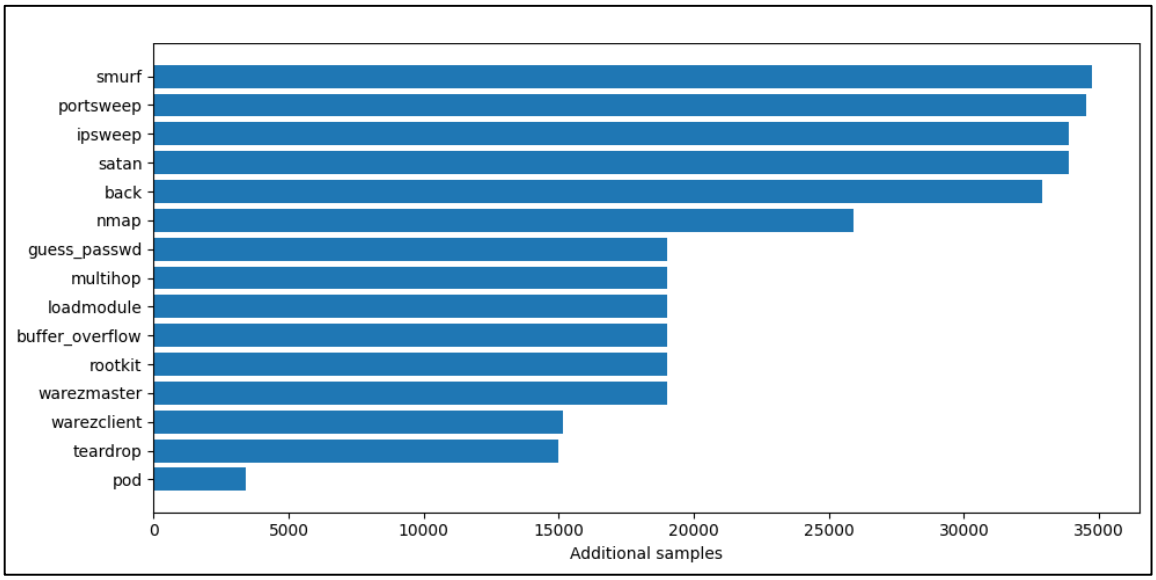

Рисунок 4.16 – Збільшення кількості навчальних зразків після застосування помилково-орієнтоване підсилення

Спостерігається суттєве підвищення повноти порівняно з базовою моделлю, однак одночасно збільшується кількість FP (1420 проти 680 у базової), що відображається на зниженні прецизійності. Це є типовою ситуацією для моделей, навчання яких значною мірою орієнтоване на збільшення чутливості до атак. З огляду на це додаткова модель у подальшому використовується не для повної заміни базової моделі, а для селективного уточнення рішень у зоні невизначеності. Матрицю невідповідностей додаткової моделі підсилення наведено на рис. Г.15.

Подальше підвищення точності реалізується через визначення τ-зони невизначеності та селективне перевизначення рішень у межах цієї області.

Зона невизначеності прогнозу τ визначається за ймовірністю базової моделі p_1. Записи, для яких значення p_1 знаходиться в інтервалі $[\tau_L; \tau_U]$, вважаються такими, для яких базова модель має підвищений ризик помилки. Для цих записів виконується отримання прогнозу додаткової моделі p_2.

Параметри τ_L, τ_U, а також пороги селективного перевизначення визначаються автоматизовано на валідаційній множині шляхом перебору

параметрів і мінімізації кількості помилок у τ-зоні (табл. 4.18, рис. 4.17). У результаті обрано: $\tau_L = 0{,}40$, $\tau_U = 0{,}60$, а також пороги впевненості додаткової моделі: $\text{thr}_{lo} = 0{,}50$, $\text{thr}_{hi} = 0{,}70$.

Таблиця 4.18 – Параметри τ-зони та порогів селективного перевизначення, обрані на валідаційній множині

Параметр	Значення
Нижня межа τ-зони, τ_L	0,40
Верхня межа τ-зони, τ_U	0,60
Нижній поріг впевненості додаткової моделі, thr_{lo}	0,50
Верхній поріг впевненості додаткової моделі, thr_{hi}	0,70
Розмір τ-зони на валідації, записів	325
Кількість помилок у τ-зоні до перевизначення, записів	145
Кількість помилок у τ-зоні після перевизначення, записів	118
Частка записів у τ-зоні, для яких застосовано перевизначення	0,8615

Таким чином, додаткова модель використовується лише за умови достатньо високої впевненості, що зменшує ризик неконтрольованого збільшення кількості хибнопозитивних рішень (FP).

Ефективність запропонованої каскадної схеми оцінюється за результатами змішаної каскадної моделі та аналізом приросту ключових метрик якості.

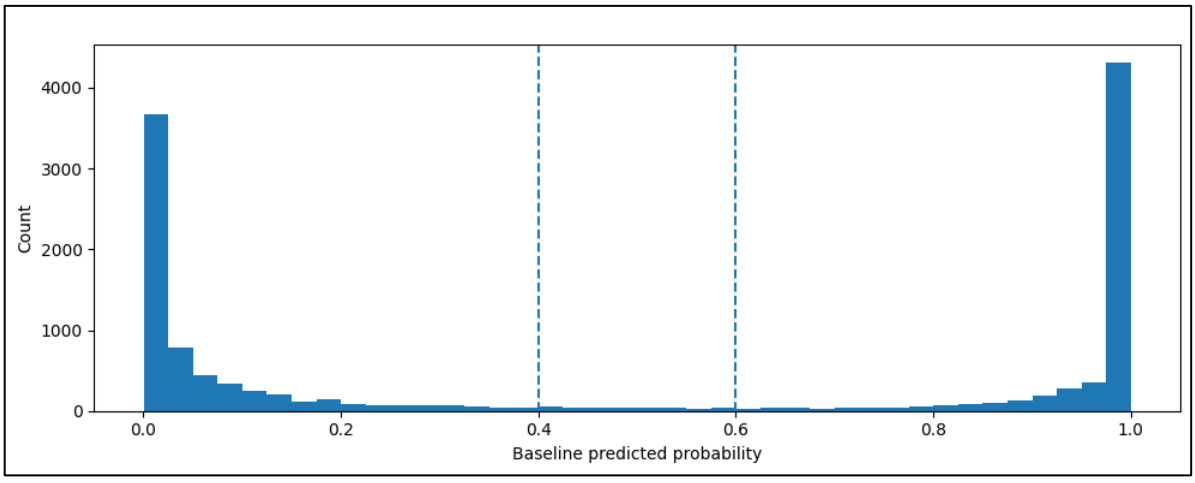

Рисунок 4.17 – Розподіл ймовірностей базової моделі на валідаційній множині та виділена τ-зона

Після підбору параметрів на валідації каскадна схема застосована до тестової множини. Розмір τ-зони на тесті становить 1466 записів, тобто лише частина прикладів підлягає уточненню рішення додатковою моделлю. Це пояснює відносно помірний приріст агрегованих метрик.

Результати каскадної схеми на тестовій множині: точність = 0,8287, прецизійність = 0,9293, повнота = 0,7567, F1-міра = 0,8342.

Порівняння базової моделі та каскадної схеми показує збільшення повноти з 0,7389 до 0,7567 та зростання F1-міри з 0,8247 до 0,8342. Приріст F1-міри становить приблизно 0,95 відсотка, що є прийнятним результатом, оскільки каскадний механізм впливає лише на підмножину записів, відібраних за критерієм невизначеності. При цьому прецизійність зберігається на високому рівні, що свідчить про контрольованість селективного перевизначення. Додаткові результати для змішаної каскадної схеми та підкласів атак, що формують складні випадки у τ-зоні, наведено на рисунках Г.3, Г.16–Г.17.

Додатково оцінено розподіл підкласів КА (рис. 4.18), які найчастіше потрапляють у τ-зону на валідаційній та тестовій множинах.

Ці результати підтверджують, що τ-зона концентрує найбільш складні для класифікації ситуації, зокрема підкласи атак, які вже ідентифіковані як проблемні за результатами аналізу помилок.

Отримані результати свідчать про доцільність каскадного уточнення рішень у τ-зоні як засобу контрольованого підвищення точності виявлення атак без повної заміни базового класифікатора.

Окремо наведено переліки підкласів атак (рис. 4.19), для яких каскадна схема забезпечує найбільший приріст повноти відносно базової моделі.

Запропонована каскадна схема підвищення точності, що поєднує балансування SPAS, помилково-орієнтоване підсилення проблемних підкласів КА, визначення τ-зони та селективне перевизначення рішень, продемонструвала помірний, але послідовний приріст агрегованих метрик на тестовій множині. Зокрема, значення F1-міри збільшилося приблизно на 1 % при збереженні високої точності та підвищенні повноти виявлення атак. Отримані результати

підтверджують доцільність використання каскадного механізму як засобу контрольованого уточнення рішень для складних прикладів без повної заміни базової моделі.

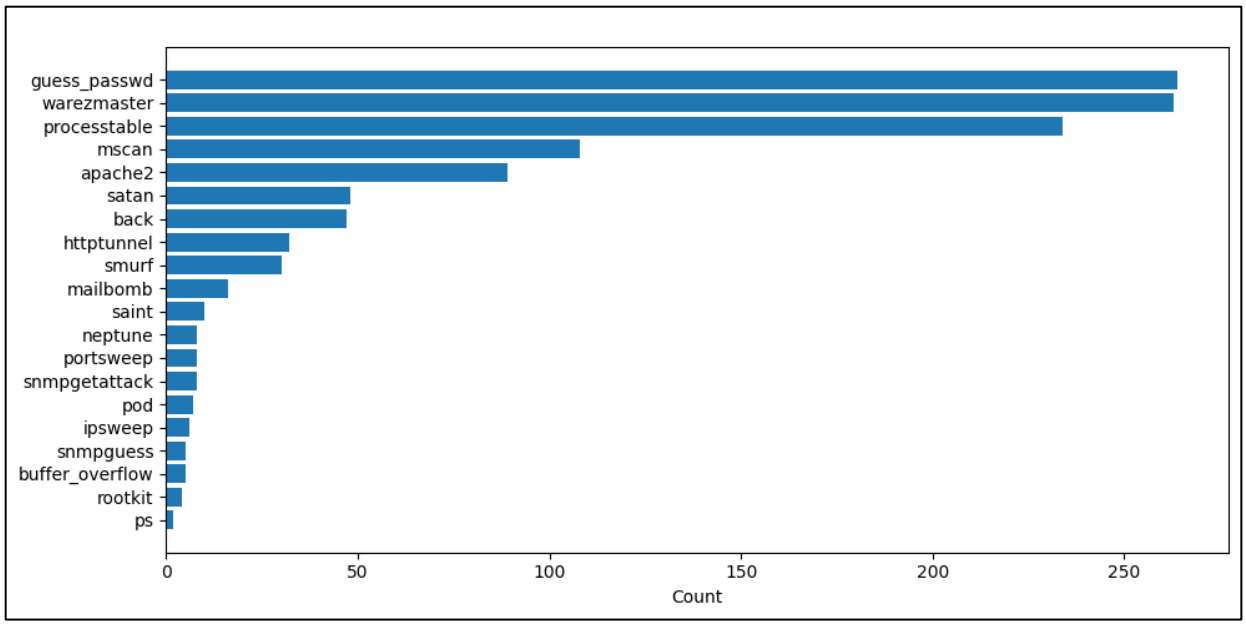

Рисунок 4.18 – Найпоширеніші підкласи атак у τ-зоні на тестовій множині

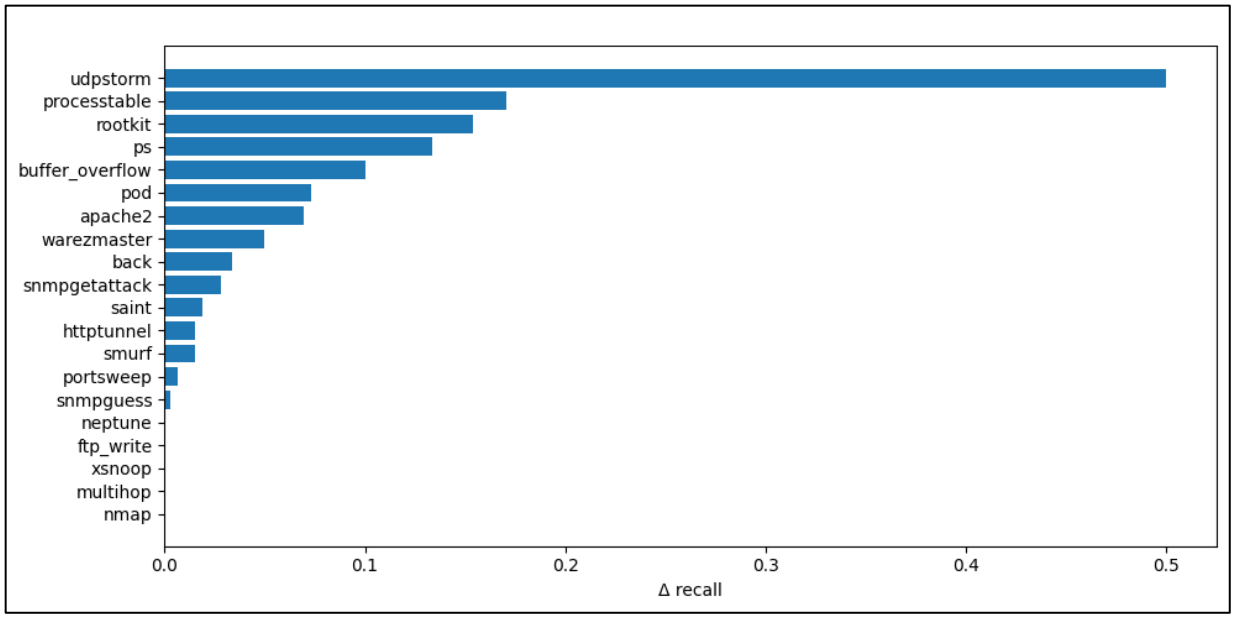

Рисунок 4.19 – Топ підкласів атак за приростом повноти: каскадна схема проти базової моделі

4.4 Висновки до четвертого розділу

Виконано експериментальне дослідження ефективності запропонованого соніфікаційного підходу до подання мережного трафіку та методів підвищення якості виявлення атак на основі балансування навчальної вибірки і селективного уточнення рішень у зоні невизначеності прогнозу. Об’єктом експериментального дослідження використано набір даних NSL-KDD, який характеризується відмінностями структури навчальної та тестової вибірок і вираженою нерівномірністю розподілу підкласів атак. Встановлені властивості набору даних зумовили доцільність перевірки методів, орієнтованих на роботу в умовах дисбалансу та наявності складних для розпізнавання підкласів.

На етапі формування базового конвеєра підтверджено працездатність соніфікаційного подання як альтернативного каналу представлення мережних з’єднань для задачі виявлення вторгнень. Порівняння векторної моделі та соніфікованої двовимірної ЗНМ показало близькі значення узагальнювальних метрик на тестовій вибірці, що підтвердило можливість використання спектрального представлення ІКМ+КПФ як основи для подальших експериментів із балансуванням та каскадним прийняттям рішень. Таким чином, соніфікаційний конвеєр розглядається як коректний спосіб подання даних, придатний для застосування нейромережевих методів аналізу зображень до спектрограм мережного трафіку.

Під час дослідження балансування встановлено суттєвий вплив структури навчальної вибірки на повноту виявлення КА. Класичний підхід SMOTENC забезпечив покращення метрик порівняно з навчанням без балансування, тоді як запропонований сигнатурозбережний метод SPAS продемонстрував найбільший приріст показників, насамперед за рахунок підвищення повноти розпізнавання атак і зменшення кількості пропущених КА. Результати підтвердили, що SPAS є ефективним інструментом формування навчальної множини для соніфікованої моделі за рахунок генерації синтетичних зразків у межах характерних діапазонів ознак для відповідних підкласів КА.

Досліджено каскадний механізм підвищення точності у зоні невизначеності прогнозу, який поєднує базову модель, додаткову модель, навчальну множину, підсилену за підходом помилково-орієнтованого підсилення із використанням SPAS, та механізм селективного перевизначення рішень у τ-зоні. Отриманий приріст агрегованих метрик виявився помірним і становив близько одного відсотка за показником F1-міри відносно базової моделі на SPAS-вибірці, бо каскадна корекція застосовується лише до підмножини записів, які належать до зони невизначеності прогнозу. Водночас експериментально підтверджено принципову доцільність каскадного уточнення, оскільки воно забезпечує відтворюваний додатковий внесок у якість детектування без зміни основного конвеєра соніфікації.

Таким чином, запропонований підхід до подання мережного трафіку на основі соніфікації є працездатною альтернативою традиційному векторному поданню, метод SPAS забезпечує суттєве покращення якості детектування КА в умовах дисбалансу підкласів, а каскадний механізм помилково-орієнтованого підсилення із селективним перевизначенням рішень у τ-зоні забезпечує додаткове покращення показників якості, що є обґрунтованим результатом експериментального дослідження.

Основні результати розділу опубліковані у працях [127–132].

ВИСНОВКИ

У результаті виконання дисертаційного дослідження було розв’язано актуальну науково-прикладну задачу підвищення точності та збалансованості виявлення комп’ютерних атак у корпоративних мережах на основі частотно-часового представлення мережного трафіку, адаптивного балансування навчальної вибірки та селективного перевизначення рішень у зоні невизначеності прогнозу.

У роботі отримано наступні наукові та практичні результати.

1. Проведено аналіз сучасних методів, засобів і технологій виявлення комп’ютерних атак у корпоративних мережах, а також досліджено обмеження існуючих підходів до подання мережного трафіку та класифікації атак в умовах дисбалансу класів. Встановлено, що традиційне векторне подання ознак не завжди забезпечує достатню інформативність для моделей глибинного навчання, а відомі процедури балансування навчальних вибірок не повною мірою враховують структурні особливості міноритарних підкласів атак. Показано, що перспективним напрямом підвищення ефективності систем виявлення КА є перехід до частотно-часового подання мережного трафіку, доповненого спеціалізованими механізмами адаптивного балансування та корекції рішень у складних випадках класифікації.

2. Розроблено модель процесу виявлення КА, яка поєднує етапи попередньої обробки мережного трафіку, його соніфікації, спектрального аналізу, базової класифікації, адаптивного балансування навчальної вибірки та каскадного прийняття рішень. Особливістю моделі є узгодження процесів навчання і функціонування систем виявлення КА у межах єдиного соніфікаційного конвеєра, що забезпечує формування інформативного двовимірного подання даних, придатного для застосування згорткових нейронних мереж, та створює основу для підвищення якості класифікації в умовах дисбалансу класів.

3. Розроблено метод частотно-часового перетворення мережного трафіку, придатний для побудови структурованого двовимірного подання даних і застосування нейромережевих моделей аналізу зображень до задачі бінарної класифікації атак. Метод ґрунтується на інтерпретації багатовимірного простору

параметрів мережного з’єднання як параметрів сигналу, подальшому ІКМ-перетворенні та формуванні спектрограм на основі короткочасного перетворення Фур’є. Це дало змогу перейти від табличного подання мережних ознак до спектральних образів, які краще відображають приховані просторово-частотні закономірності мережної активності та підвищують інформативність вхідних даних для моделі класифікації.

За результатами експерименту соніфікована двовимірна ЗНМ забезпечила конкурентні показники порівняно з векторною моделлю, підтвердивши працездатність частотно-часового подання як альтернативного способу формування інформативних вхідних даних.

4. Розроблено метод сигнатурозбережного адаптивного балансування навчальної вибірки, який забезпечує зменшення впливу дисбалансу класів і підвищення репрезентативності міноритарних підкласів КА. На відміну від стандартних процедур надсемплінгу та підвибірки, запропонований метод орієнтований на збереження характерних структурних властивостей проблемних підкласів та врахування помилок базової моделі при формуванні розширеної навчальної множини. Це дало змогу посилити представленість рідкісних КА у навчальному процесі без істотного руйнування їх сигнатурних шаблонів і створило передумови для підвищення повноти їх виявлення. Застосування SPAS підвищило повноту виявлення атак до 79,58 %, тоді як для SMOTENC вона становила 65,41 %; кількість хибнонегативних рішень зменшилася до 2621 проти 4439 для SMOTENC.

5. Розроблено метод підвищення точності виявлення КА у зоні невизначеності прогнозу на основі аналізу помилок базового класифікатора та селективного перевизначення рішень за допомогою допоміжної моделі. Метод базується на формалізації τ-інтервалу невизначеності, виділенні проблемних прикладів, для яких базова модель формує прикордонні або помилкові прогнози, та побудові додаткової моделі, навченої на складних випадках класифікації. Запропонований підхід дав змогу зменшити кількість хибнонегативних рішень без порушення принципу незалежності тестової вибірки та підвищити якість прийняття рішень у складних випадках прогнозування.

Каскадна схема підвищила повноту з 73,89 % до 75,67 %, а F1-міру – з 82,47 % до 83,42 %, тобто забезпечила приріст F1-міри приблизно на 0,95 відсоткового пункта при збереженні високої прецизійності 92,93 %.

6. Розроблено програмно-алгоритмічне забезпечення для валідації запропонованих моделей і методів та проведено експериментальні дослідження їх ефективності на наборі даних NSL-KDD. Експериментальна перевірка підтвердила доцільність використання соніфікованого подання мережного трафіку для задачі виявлення вторгнень, а також ефективність поєднання базової двовимірної ЗНМ-моделі з методом SPAS і селективним перевизначенням рішень у τ-зоні. Отримані результати показали, що запропонований підхід забезпечує більш якісне виявлення комп’ютерних атак порівняно з векторною моделлю без соніфікації та базовою соніфікованою моделлю без спеціалізованих механізмів балансування й обробки невизначеності, що підтверджує практичну придатність розроблених рішень для створення та модернізації систем моніторингу безпеки КМ.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Mbona I., Eloff J.H.P. Detecting Zero-Day Intrusion Attacks Using Semi-Supervised Machine Learning Approaches. *IEEE Access*. 2022. Vol. 10. P. 69822–69838. URL: https://doi.org/10.1109/ACCESS.2022.3187116 (дата звернення: 09.03.2026).

2. Aldweesh A., Derhab A., Emam A.Z. Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues. *Knowl.-Based Syst*. 2020. Vol. 189. Art. 105124. URL: https://doi.org/10.1016/j.knosys.2019.105124 (дата звернення: 09.03.2026).

3. Olouhal O.U., Yange T.S., Okerekel G.E., Bakpol F.S. Cutting Edge Trends in Deception Based Intrusion Detection Systems–A Survey. *J. Inf. Secur*. 2021. Vol. 12. P. 250–269. URL: https://doi.org/10.4236/jis.2021.124014 (дата звернення: 09.03.2026).

4. Asharf J., Moustafa N., Khurshid H., Debie E., Haider W., Wahab A. A Review of Intrusion Detection Systems Using Machine and Deep Learning in Internet of Things: Challenges, Solutions and Future Directions. *Electronics*. 2020. Vol. 9. Art. 1177. URL: https://www.mdpi.com/2079-9292/9/7/1177 (дата звернення: 09.03.2026).

5. Shitharth S., Kshirsagar P.R., Balachandran P.K., Alyoubi K.H., Khadidos A.O. An Innovative Perceptual Pigeon Galvanized Optimization (PPGO) Based Likelihood Naïve Bayes (LNB) Classification Approach for Network Intrusion Detection System. *IEEE Access*. 2022. Vol. 10. P. 46424–46441. URL: https://doi.org/10.1109/ACCESS.2022.3171660 (дата звернення: 09.03.2026).

6. Prashanth S.K., Shitharth S., Praveen Kumar B., Subedha V., Sangeetha K. Optimal Feature Selection Based on Evolutionary Algorithm for Intrusion Detection. *SN Comput. Sci*. 2022. Vol. 3. Art. 439. URL: https://doi.org/10.1007/s42979-022-01325-4 (дата звернення: 09.03.2026).

7. Sheikh M.S., Peng Y. Procedures, Criteria, and Machine Learning Techniques for Network Traffic Classification: A Survey. *IEEE Access*. 2022. Vol. 10. P. 61135–61158. URL: https://doi.org/10.1109/ACCESS.2022.3181135 (дата звернення: 09.03.2026).

8. Lin Z., Shi Y., Xue Z. IDSGAN: Generative adversarial networks for attack generation against intrusion detection. arXiv. 2021. URL: https://arxiv.org/abs/1809.02077 (дата звернення: 09.03.2026).

9. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative adversarial networks. *Commun. ACM*. 2020. Vol. 63, no. 11. P. 139–144. URL: https://doi.org/10.1145/3422622 (дата звернення: 09.03.2026).

10. Abdelmoumin G., Whitaker J., Rawat D.B., Rahman A. A Survey on Data-Driven Learning for Intelligent Network Intrusion Detection Systems. *Electronics*. 2022. Vol. 11. Art. 213. URL: https://doi.org/10.3390/electronics11020213 (дата звернення: 09.03.2026).

11. Zhu Y., Cui L., Ding Z., Li L., Liu Y., Hao Z. Black box attack and network intrusion detection using machine learning for malicious traffic. *Comput. Secur*. 2022. Vol. 123. Art. 102922. URL: https://doi.org/10.1016/j.cose.2022.102922 (дата звернення: 09.03.2026).

12. Balyan A.K., Ahuja S., Lilhore U.K., Sharma S.K., Manoharan P., Algarni A.D., Elmannai H., Raahemifar K. A Hybrid Intrusion Detection Model Using EGA-PSO and Improved Random Forest Method. *Sensors*. 2022. Vol. 22. Art. 5986. URL: https://doi.org/10.3390/s22165986 (дата звернення: 09.03.2026).

13. Shahriar M.D., Haque N.I., Rahman M.A., Alonso M. G-IDS: Generative Adversarial Networks Assisted Intrusion Detection System. In: *Proceedings of the 2020 IEEE 44th Annual Computers, Software and Applications Conference (COMPSAC)*. Madrid, Spain, 2020. URL: https://doi.org/10.1109/COMPSAC48688.2020.0-218 (дата звернення: 09.03.2026).

14. Vaccari I., Carlevaro A., Narteni S., Cambiaso E., Mongelli M. eXplainable and Reliable Against Adversarial Machine Learning in Data Analytics. *IEEE Access*. 2022. Vol. 10. P. 83949–83970. URL: https://doi.org/10.1109/ACCESS.2022.3197299 (дата звернення: 09.03.2026).

15. Fasci L.S., Fisichella G.L., Qian C. Disarming visualization-based approaches in malware detection systems. *Comput. Secur*. 2023. Vol. 126. Art. 103062. URL: https://doi.org/10.1016/j.cose.2022.103062 (дата звернення: 09.03.2026).

16. TensorFlow. 2022. URL: https://www.tensorflow.org/ (дата звернення: 09.03.2026).

17. Keras. 2022. URL: https://keras.io/about/ (дата звернення: 09.03.2026).

18. Azeroual O., Nikiforova A. Apache Spark and MLlib-Based Intrusion Detection System or How the Big Data Technologies Can Secure the Data. *Information*. 2022. Vol. 13. Art. 58. URL: https://doi.org/10.3390/info13020058 (дата звернення: 09.03.2026).

19. Chahar R., Kaur D. A systematic review of the machine learning algorithms for the computational analysis in different domains. *International Journal of Advanced Technology and Engineering Exploration*. 2020. Vol. 7. P. 147–164. URL: https://doi.org/10.19101/IJATEE.2020.762212 (дата звернення: 09.03.2026).

20. Ahmad Z., Shahid Khan A., Shiang C., Ahmad F. Network intrusion detection system: A systematic study of machine learning and deep learning approaches. *Transactions on Emerging Telecommunications Technologies.* 2021. Vol. 32. Art. e4150. URL: https://doi.org/10.1002/ett.4150 (дата звернення: 09.03.2026).

21. Al Jallad K., Aljnidi M., Desouki M.S. Anomaly detection optimization using big data and deep learning to reduce false-positive. *Journal of Big Data*. 2020. Vol. 7. Art. 68. URL: https://doi.org/10.1186/s40537-020-00347-w (дата звернення: 09.03.2026).

22. Alzubaidi L., Zhang J., Humaidi A.J., Al-Dujaili A., Duan Y., Al-Shamma O., Farhan L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*. 2021. Vol. 8. Art. 53. URL: https://doi.org/10.1186/s40537-021-00444-8 (дата звернення: 09.03.2026).

23. Sarker I.H. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Computer Science*. 2021. Vol. 2. Art. 420. URL: https://doi.org/10.1007/s42979-021-00815-1 (дата звернення: 09.03.2026).

24. Kocher G., Kumar G. Machine learning and deep learning methods for intrusion detection systems: Recent developments and challenges. *Soft Computing.* 2021. Vol. 25. P. 9731–9763. URL: https://doi.org/10.1007/s00500-021-05893-0 (дата звернення: 09.03.2026).

25. Alzaqebah A., Aljarah I., Al-Kadi O., Damaševičius R. A Modified Grey Wolf Optimization Algorithm for an Intrusion Detection System. *Mathematics.* 2022. Vol. 10. Art. 999. URL: https://doi.org/10.3390/math10070999 (дата звернення: 09.03.2026).

26. Ali M.H., Jaber M.M., Abd S.K., Rehman A., Awan M.J., Damaševičius R., Bahaj S.A. Threat Analysis and Distributed Denial of Service (DDoS) Attack Recognition in the Internet of Things (IoT). *Electronics.* 2022. Vol. 11. Art. 494. URL: https://doi.org/10.3390/electronics11030494 (дата звернення: 09.03.2026).

27. Fu Y., Du Y., Cao Z., Li Q., Xiang W. A Deep Learning Model for Network Intrusion Detection with Imbalanced Data. *Electronics.* 2022. Vol. 11. Art. 898. URL: https://doi.org/10.3390/electronics11060898 (дата звернення: 09.03.2026).

28. Mahalakshmi G., Uma E., Aroosiya M., Vinitha M. Intrusion Detection System Using Convolutional Neural Network on UNSW NB15 Dataset. *Advances in Parallel Computing.* 2021. Vol. 40. P. 1–8. URL: https://doi.org/10.3233/APC210017 (дата звернення: 09.03.2026).

29. Li G., Jung J.J. Deep learning for anomaly detection in multivariate time series: Approaches, applications, and challenges. *Information Fusion.* 2023. Vol. 91. P. 93–102. URL: https://doi.org/10.1016/j.inffus.2022.10.001 (дата звернення: 09.03.2026).

30. Landauer M., Onder S., Skopik F., Wurzenberger M. Deep learning for anomaly detection in log data: A survey. *Machine Learning with Applications.* 2023. Vol. 12. Art. 100470. URL: https://doi.org/10.1016/j.mlwa.2023.100470 (дата звернення: 09.03.2026).

31. Gupta C., Gill N.S., Maheshwary P., Pandit S.V., Gulia P., Pareek P.K. An Ensemble Machine Learning Method for Analyzing Various Medical Datasets. *Journal of Intelligent Systems and Internet of Things.* 2024. Vol. 13. P. 177–195. URL:

https://scholar.google.com/scholar_lookup?title=An+Ensemble+Machine+Learning+Method+for+Analyzing+Various+Medical+Datasets&author=Gupta,+C.&author=Gill,+N.S.&author=Maheshwary,+P.&author=Pandit,+S.V.&author=Gulia,+P.&author=Pareek,+P.K.&publication_year=2024&journal=J.+Intell.+Syst.+Internet+Things&volume=13&pages=177%E2%80%93195 (дата звернення: 09.03.2026).

32. Duong H.T., Le V.T., Hoang V.T. Deep learning-based anomaly detection in video surveillance: A survey. *Sensors.* 2023. Vol. 23. Art. 5024. URL: https://doi.org/10.3390/s23105024 (дата звернення: 09.03.2026).

33. Abusitta A., de Carvalho G.H., Wahab O.A., Halabi T., Fung B.C., Al Mamoori S. Deep learning-enabled anomaly detection for IoT systems. *Internet of Things.* 2023. Vol. 21. Art. 100656. URL: https://doi.org/10.1016/j.iot.2022.100656 (дата звернення: 09.03.2026).

34. Pareek P.K., Siddhanti P., Anupkant S., Husain S.O., Bhuvaneshwarri I. Ensemble Based Machine Learning Classifier for Sybil Attack Detection in Vehicular AD-HOC Networks (VANETS). In: *Proceedings of the 2024 Second International Conference on Data Science and Information System (ICDSIS).* Hassan, India, 2024. P. 1–4. URL: https://scholar.google.com/scholar_lookup?title=Ensemble+Based+Machine+Learning+Classifier+for+Sybil+Attack+Detection+in+Vehicular+AD-HOC+Networks+(VANETS)&conference=Proceedings+of+the+2024+Second+International+Conference+on+Data+Science+and+Information+System+(ICDSIS)&author=Pareek,+P.K.&author=Siddhanti,+P.&author=Anupkant,+S.&author=Husain,+S.O.&author=Bhuvaneshwarri,+I.&publication_year=2024&pages=1%E2%80%934 (дата звернення: 09.03.2026).

35. Gómez Á.L.P., Maimó L.F., Celdrán A.H., Clemente F.J.G. SUSAN: A Deep Learning based anomaly detection framework for sustainable industry. Sustainable Computing: Informatics and Systems. 2023. Vol. 37. Art. 100842. URL: https://doi.org/10.1016/j.suscom.2022.100842 (дата звернення: 09.03.2026).

36. Berroukham A., Housni K., Lahraichi M., Boulfrifi I. Deep learning-based methods for anomaly detection in video surveillance: A review. *Bulletin of Electrical*

Engineering and Informatics. 2023. Vol. 12. P. 314–327. URL: https://doi.org/10.11591/eei.v12i1.4185 (дата звернення: 09.03.2026).

37. Zipfel J., Verworner F., Fischer M., Wieland U., Kraus M., Zschech P. Anomaly detection for industrial quality assurance: A comparative evaluation of unsupervised deep learning models. *Computers & Industrial Engineering.* 2023. Vol. 177. Art. 109045. URL: https://doi.org/10.1016/j.cie.2023.109045 (дата звернення: 09.03.2026).

38. Pillai S.E.V.S., Polimetla K., Prakash C.S., Pareek P.K., Zanke P. Risk Prediction and Management for Effective Cyber Security Using Weighted Fuzzy C Means Clustering. In: *Proceedings of the 2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE).* Ballari, India, 2024. P. 1–4. URL: https://scholar.google.com/scholar_lookup?title=Risk+Prediction+and+Management+for+Effective+Cyber+Security+Using+Weighted+Fuzzy+C+Means+Clustering&conference=Proceedings+of+the+2024+Third+International+Conference+on+Distributed+Computing+and+Electrical+Circuits+and+Electronics+(ICDCECE)&author=Pillai,+S.E.V.S.&author=Polimetla,+K.&author=Prakash,+C.S.&author=Pareek,+P.K.&author=Zanke,+P.&publication_year=2024&pages=1%E2%80%934 (дата звернення: 09.03.2026).

39. Wang X., Wang Y., Javaheri Z., Almutairi L., Moghadamnejad N., Younes O.S. Federated deep learning for anomaly detection in the internet of things. *Computers & Electrical Engineering.* 2023. Vol. 108. Art. 108651. URL: https://doi.org/10.1016/j.compeleceng.2023.108651 (дата звернення: 09.03.2026).

40. Bovenzi G., Aceto G., Ciuonzo D., Montieri A., Persico V., Pescapé A. Network anomaly detection methods in IoT environments via deep learning: A fair comparison of performance and robustness. *Computers & Security.* 2023. Vol. 128. Art. 103167. URL: https://doi.org/10.1016/j.cose.2023.103167 (дата звернення: 09.03.2026).

41. Pazho A.D., Noghre G.A., Purkayastha A.A., Vempati J., Martin O., Tabkhi H. A survey of graph-based deep learning for anomaly detection in distributed systems.

IEEE Transactions on Knowledge and Data Engineering. 2023. Vol. 36. P. 1–20. URL: https://doi.org/10.1109/TKDE.2023.3282898 (дата звернення: 09.03.2026).

42. Alloqmani A., Abushark Y.B., Khan A.I. Anomaly detection of breast cancer using deep learning. *Arabian Journal for Science and Engineering.* 2023. Vol. 48. P. 10977–11002. URL: https://doi.org/10.1007/s13369-023-07786-1 (дата звернення: 09.03.2026).

43. He Z., Chen Y., Zhang H., Zhang D. WKN-OC: A new deep learning method for anomaly detection in intelligent vehicles. *IEEE Transactions on Intelligent Vehicles.* 2023. Vol. 8. P. 2162–2172. URL: https://doi.org/10.1109/TIV.2022.3216401 (дата звернення: 09.03.2026).

44. Inuwa M.M., Das R. A comparative analysis of various machine learning methods for anomaly detection in cyber attacks on IoT networks. *Internet of Things.* 2024. Vol. 26. Art. 101162. URL: https://doi.org/10.1016/j.iot.2024.101162 (дата звернення: 09.03.2026).

45. Altulaihan E., Almaiah M.A., Aljughaiman A. Anomaly Detection IDS for Detecting DoS Attacks in IoT Networks Based on Machine Learning Algorithms. *Sensors.* 2024. Vol. 24. Art. 713. URL: https://doi.org/10.3390/s24030713 (дата звернення: 09.03.2026).

46. Xin Q., Xu Z., Guo L., Zhao F., Wu B. IoT Traffic Classification and Anomaly Detection Method based on Deep Autoencoders. *Applied and Computational Engineering.* 2024. Vol. 69. P. 64–70. URL: https://doi.org/10.54254/2755-2721/69/2024M19629 (дата звернення: 09.03.2026).

47. Bhatia M.P.S., Sangwan S.R. Soft computing for anomaly detection and prediction to mitigate IoT-based real-time abuse. *Personal and Ubiquitous Computing.* 2024. Vol. 28. P. 123–133. URL: https://doi.org/10.1007/s00779-022-01714-9 (дата звернення: 09.03.2026).

48. Khan M.M., Alkhathami M. Anomaly detection in IoT-based healthcare: Machine learning for enhanced security. *Scientific Reports.* 2024. Vol. 14. Art. 5872. URL: https://doi.org/10.1038/s41598-024-56205-x (дата звернення: 09.03.2026).

49. Murugesh C., Murugan S. Evolutionary Optimization with Variational Auto encoder-based Denial of Service Attack Detection and Classification in Wireless Sensor Networks. In: *Proceedings of the 2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS).* Trichy, India, 2022. P. 994–1000. URL: https://scholar.google.com/scholar_lookup?title=Evolutionary+Optimization+with+Variational+Auto+encoder-based+Denial+of+Service+Attack+Detection+and+Classification+in+Wireless+Sensor+Networks&conference=Proceedings+of+the+2022+International+Conference+on+Augmented+Intelligence+and+Sustainable+Systems+(ICAISS)&author=Murugesh,+C.&author=Murugan,+S.&publication_year=2022&pages=994%E2%80%931000 (дата звернення: 09.03.2026).

50. Thirumalraj A., Asha V., Kavin B.P. An Improved Hunter-Prey Optimizer-Based DenseNet Model for Classification of Hyper-Spectral Images. *In: AI and IoT-Based Technologies for Precision Medicine. Hershey, PA*, 2023. P. 76–96. URL: https://scholar.google.com/scholar_lookup?title=An+Improved+Hunter-Prey+Optimizer-Based+DenseNet+Model+for+Classification+of+Hyper-Spectral+Images&author=Thirumalraj,+A.&author=Asha,+V.&author=Kavin,+B.P.&publication_year=2023&pages=76%E2%80%9396 (дата звернення: 09.03.2026).

51. Thirumalraj A., Anusuya V.S., Manjunatha B. Detection of Ephemeral Sand River Flow Using Hybrid Sandpiper Optimization-Based CNN Model. In: *Innovations in Machine Learning and IoT for Water Management.* Hershey, PA, 2024. P. 195–214. URL: https://scholar.google.com/scholar_lookup?title=Detection+of+Ephemeral+Sand+River+Flow+Using+Hybrid+Sandpiper+Optimization-Based+CNN+Model&author=Thirumalraj,+A.&author=Anusuya,+V.S.&author=Manjunatha,+B.&publication_year=2024&pages=195%E2%80%93214 (дата звернення: 09.03.2026).

52. Elmasry W., Akbulut A., Zaim A.H. Evolving deep learning architectures for network intrusion detection using a double PSO metaheuristic. *Computer Networks.*

2020. Vol. 168. Art. 107042. URL: https://doi.org/10.1016/j.comnet.2019.107042 (дата звернення: 09.03.2026).

53. Muhammad G., Hossain M.S., Garg S. Stacked autoencoder-based intrusion detection system to combat financial fraudulent. *IEEE Internet of Things Journal*. 2020. URL: https://doi.org/10.1109/JIOT.2020.3041184 (дата звернення: 09.03.2026).

54. Tang C., Luktarhan N., Zhao Y. SAAE-DNN: Deep Learning Method on Intrusion Detection. *Symmetry*. 2020. Vol. 12, no. 10. Art. 1695. URL: https://doi.org/10.3390/sym12101695 (дата звернення: 09.03.2026).

55. Alladi T., Chamola V., Sikdar B., Choo K. Consumer IoT: Security Vulnerability Case Studies and Solutions. *IEEE Consumer Electronics Magazine*. 2020. Vol. 9, no. 2. P. 17–25. URL: https://doi.org/10.1109/MCE.2019.2953740 (дата звернення: 09.03.2026).

56. Kilincer I., Ertam F., Sengur A. Machine Learning Methods for Cyber Security Intrusion Detection: Datasets and Comparative Study. *Computer Networks*. 2021. Vol. 188. Art. 107840. URL: https://doi.org/10.1016/j.comnet.2021.107840 (дата звернення: 09.03.2026).

57. Rodríguez E., Otero B., Gutiérrez N., Canal R. A Survey of Deep Learning Techniques for Cybersecurity in Mobile Networks. *IEEE Communications Surveys & Tutorials*. 2021. Vol. 23, no. 3. P. 1920–1955. URL: https://doi.org/10.1109/COMST.2021.3086296 (дата звернення: 09.03.2026).

58. Masum M., Shahriar H. TL-NID: Deep Neural Network with Transfer Learning for Network Intrusion Detection. In: *Proceedings of the 15th International Conference for Internet Technology and Secured Transactions (ICITST)*. London, UK, 2020. P. 1–7. URL: https://doi.org/10.23919/ICITST51030.2020.9351317 (дата звернення: 09.03.2026).

59. Li X., Hu Z., Xu M., Wang Y., Ma J. Transfer Learning Based Intrusion Detection Scheme for Internet of Vehicles. *Information Sciences*. 2021. Vol. 547. P. 119–135. URL: https://doi.org/10.1016/j.ins.2020.05.130 (дата звернення: 09.03.2026).

60. Mehedi S.T., Anwar A., Rahman Z., Ahmed K. Deep Transfer Learning Based Intrusion Detection System for Electric Vehicular Networks. *Sensors.* 2021. Vol. 21. Art. 4736. URL: https://doi.org/10.3390/s21144736 (дата звернення: 09.03.2026).

61. Kang H., Kwak B., Lee Y.H., Lee H., Lee H., Kim H.K. Car Hacking: Attack and Defense Challenge 2020 Dataset. *IEEE Dataport.* 2021. URL: https://ieee-dataport.org/open-access/car-hacking-attack-defense-challenge-2020-dataset (дата звернення: 09.03.2026).

62. Fan Y., Li Y., Zhan M., Cui H., Zhang Y. IoTDefender: A Federated Transfer Learning Intrusion Detection Framework for 5G IoT. In: *Proceedings of the 2020 IEEE 14th International Conference on Big Data Science and Engineering (BigDataSE).* 2020. P. 88–95. URL: https://doi.org/10.1109/BigDataSE50710.2020.00020 (дата звернення: 09.03.2026).

63. Idrissi I., Azizi M., Moussaoui O. Accelerating the Update of a DL-Based IDS for IoT Using Deep Transfer Learning. *Indonesian Journal of Electrical Engineering and Computer Science.* 2021. Vol. 23, no. 2. P. 1059–1067. URL: https://doi.org/10.11591/ijeecs.v23.i2.pp1059-1067 (дата звернення: 09.03.2026).

64. Guan J., Cai J., Bai H. Deep Transfer Learning-Based Network Traffic Classification for Scarce Dataset in 5G IoT Systems. *International Journal of Machine Learning and Cybernetics.* 2021. Vol. 12. P. 3351–3365. URL: https://doi.org/10.1007/s13042-021-01415-4 (дата звернення: 09.03.2026).

65. Kolesnikov A., Beyer L., Zhai X., Puigcerver J., Yung J., Gelly S., Houlsby N. Big Transfer (BiT): General Visual Representation Learning. In: *Proceedings of the European Conference on Computer Vision (ECCV).* Glasgow, UK, 2020. P. 491–507. URL: https://doi.org/10.1007/978-3-030-58558-7_29 (дата звернення: 09.03.2026).

66. Mehedi S.T., Anwar A., Rahman Z., Ahmed K., Islam R. Dependable Intrusion Detection System for IoT: A Deep Transfer Learning-Based Approach. *IEEE Transactions on Industrial Informatics.* 2022. URL: https://doi.org/10.1109/TII.2022.3164770 (дата звернення: 09.03.2026).

67. Zhuang F., Qi Z., Duan K., Xi D., Zhu Y., Zhu H., Xiong H., He Q. A Comprehensive Survey on Transfer Learning. *Proceedings of the IEEE.* 2021. Vol. 109,

no. 1. P. 43–76. URL: https://doi.org/10.1109/JPROC.2020.3004555 (дата звернення: 09.03.2026).

68. Abdulrahman A.A., Ibrahem M.K. Toward Constructing a Balanced Intrusion Detection Dataset Based on CICIDS2017. *Samarra J. Pure Appl. Sci.* 2020. Vol. 2, no. 3. P. 132–142. URL: https://iasj.rdd.edu.iq/journals/uploads/2024/12/25/b570f3d9e7891d3d8207cf7de5f01a86.pdf (дата звернення: 09.03.2026).

69. Wotawa F., Muhlburger H. On the Effects of Data Sampling for Deep Learning on Highly Imbalanced Data from SCADA Power Grid Substation Networks for Intrusion Detection. In: *Proceedings of the 2021 IEEE 21st International Conference on Software Quality, Reliability and Security (QRS).* Haikou, China, 2021. P. 864–872. URL: https://doi.org/10.1109/QRS54544.2021.00095 (дата звернення: 09.03.2026).

70. Fundin A. Generating Datasets Through the Introduction of an Attack Agent in a SCADA Testbed: A methodology of creating datasets for intrusion detection research in a SCADA system using IEC-60870-5-104. *Master's Thesis.* Linköping University. Linköping, Sweden, 2021. URL: http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-175911 (дата звернення: 09.03.2026).

71. Suaboot J., Fahad A., Tari Z., Grundy J., Mahmood A.N., Almalawi A., Zomaya A.Y., Drira K. A Taxonomy of Supervised Learning for IDSs in SCADA Environments. *ACM Comput. Surv.* 2020. Vol. 53, no. 2. P. 1–37. URL: https://doi.org/10.1145/3379499 (дата звернення: 09.03.2026).

72. Kulkarni A., Chong D., Batarseh F.A. Foundations of Data Imbalance and Solutions for a Data Democracy. In: Data Democracy: *At the Nexus of Artificial Intelligence, Software Development, and Knowledge Engineering.* Amsterdam, 2020. P. 83–106. URL: https://doi.org/10.1016/B978-0-12-818366-3.00005-8 (дата звернення: 09.03.2026).

73. Peterson J.M., Leevy J.L., Khoshgoftaar T.M. A Review and Analysis of the Bot-IoT Dataset. In: *Proceedings of the 2021 IEEE 15th International Conference on Service-Oriented System Engineering (SOSE).* Oxford, UK, 2021. P. 20–27. URL: https://doi.org/10.1109/SOSE52839.2021.00007 (дата звернення: 09.03.2026).

74. Jiang K., Wang W., Wang A., Wu H. Network Intrusion Detection Combined Hybrid Sampling with Deep Hierarchical Network. *IEEE Access.* 2020. Vol. 8. P. 32464–32476. URL: https://doi.org/10.1109/ACCESS.2020.2973730 (дата звернення: 09.03.2026).

75. Aziz M.N., Ahmad T. Clustering Under-Sampling Data for Improving the Performance of Intrusion Detection System. *Journal of Engineering Science and Technology.* 2021. Vol. 16, no. 2. P. 1342–1355. URL: https://jestec.taylors.edu.my/Vol%2016%20issue%202%20April%202021/16_2_32.pdf (дата звернення: 09.03.2026).

76. Wang Z., Jiang D., Huo L., Yang W. An efficient network intrusion detection approach based on deep learning. *Wireless Networks.* 2021. P. 1–14. URL: https://doi.org/10.1007/s11276-021-02698-9 (дата звернення: 09.03.2026).

77. Azzaoui H., Boukhamla A.Z.E., Arroyo D., Bensayah A. Developing new deep-learning model to enhance network intrusion classification. *Evolutionary Systems.* 2021. Vol. 13. P. 17–25. URL: https://doi.org/10.1007/s12530-020-09364-z (дата звернення: 09.03.2026).

78. Aouedi O., Piamrat K., Parrein B. Ensemble-Based Deep Learning Model for Network Traffic Classification. *IEEE Transactions on Network and Service Management.* 2022. Vol. 19. P. 4124–4135. URL: https://doi.org/10.1109/TNSM.2022.3193748 (дата звернення: 09.03.2026).

79. Balamurugan N.M., Adimoolam M., Alsharif M.H., Uthansakul P. A Novel Method for Improved Network Traffic Prediction Using Enhanced Deep Reinforcement Learning Algorithm. *Sensors.* 2022. Vol. 22. Art. 5006. URL: https://doi.org/10.3390/s22135006 (дата звернення: 09.03.2026).

80. Driss M., Aljehani A., Boulila W., Ghandorh H., Al-Sarem M. Servicing Your Requirements: An FCA and RCA-Driven Approach for Semantic Web Services Composition. *IEEE Access.* 2020. Vol. 8. P. 59326–59339. URL: https://doi.org/10.1109/ACCESS.2020.2982592 (дата звернення: 09.03.2026).

81. Driss M., Hasan D., Boulila W., Ahmad J. Microservices in IoT Security: Current Solutions, Research Challenges, and Future Directions. *Procedia Computer*

Science. 2021. Vol. 192. P. 2385–2395. URL: https://doi.org/10.1016/j.procs.2021.09.007 (дата звернення: 09.03.2026).

82. Izadi A., Ahmadi S., Rajabzadeh M. Network Traffic Classification Using Deep Learning Networks and Bayesian Data Fusion. *Journal of Network and Systems Management.* 2022. Vol. 30. Art. 25. URL: https://doi.org/10.1007/s10922-021-09639-z (дата звернення: 09.03.2026).

83. Kumar S.A.P., Suresh A., Anand S.R., Chokkanathan K., Vijayasarathy M. Hybridization of Mean Shift Clustering and Deep Packet Inspected Classification for Network Traffic Analysis. *Wireless Personal Communications.* 2021. Vol. 127. P. 217–233. URL: https://doi.org/10.1007/s11277-021-08208-6 (дата звернення: 09.03.2026).

84. Fox G., Boppana R.V. Detection of Malicious Network Flows with Low Preprocessing Overhead. *Network.* 2022. Vol. 2. P. 628–642. URL: https://doi.org/10.3390/network2040036 (дата звернення: 09.03.2026).

85. Qin J., Han X., Wang C., Hu Q., Jiang B., Zhang C., Lu Z. Network Traffic Classification Based on SD Sampling and Hierarchical Ensemble Learning. *Security and Communication Networks.* 2023. Vol. 2023. Art. 4374385. URL: https://doi.org/10.1155/2023/4374385 (дата звернення: 09.03.2026).

86. Rodríguez M., Alesanco Á., Mehavilla L., García J. Evaluation of Machine Learning Techniques for Traffic Flow-Based Intrusion Detection. *Sensors.* 2022. Vol. 22. Art. 9326. URL: https://doi.org/10.3390/s22239326 (дата звернення: 09.03.2026).

87. Malik A., de Frein R., Al-Zeyadi M., Andreu-Perez J. Intelligent SDN Traffic Classification Using Deep Learning: Deep-SDN. In: *Proceedings of the 2020 2nd International Conference on Computer Communication and the Internet (ICCCI).* Nagoya, Japan, 2020. P. 184–189. URL: https://doi.org/10.1109/iccci49374.2020.9145971 (дата звернення: 09.03.2026).

88. Abbasi M., Shahraki A., Taherkordi A. Deep Learning for Network Traffic Monitoring and Analysis (NTMA): A Survey. *Computer Communications.* 2021. Vol. 170. P. 19–41. URL: https://doi.org/10.1016/j.comcom.2021.01.021 (дата звернення: 09.03.2026).

89. Ben Atitallah S., Driss M., Boulila W., Koubaa A., Ben Ghézala H. Fusion of Convolutional Neural Networks Based on Dempster–Shafer Theory for Automatic Pneumonia Detection from Chest X-ray Images. *International Journal of Imaging Systems and Technology*. 2021. Vol. 32. P. 658–672. URL: https://doi.org/10.1002/ima.22653 (дата звернення: 09.03.2026).

90. Ben Atitallah S., Driss M., Boulila W., Ben Ghézala H. Randomly Initialized Convolutional Neural Network for the Recognition of COVID-19 Using X-ray Images. *International Journal of Imaging Systems and Technology*. 2021. Vol. 32. P. 55–73. URL: https://doi.org/10.1002/ima.22654 (дата звернення: 09.03.2026).

91. He H., Huang G., Zhang B., Zheng Z. Research on DoS Traffic Detection Model Based on Random Forest and Multilayer Perceptron. *Security and Communication Networks*. 2022. Vol. 2022. Art. 2076987. URL: https://doi.org/10.1155/2022/2076987 (дата звернення: 09.03.2026).

92. Ponmalar P.S., Kamboj A., Shete V., Harikrishnan R. Machine Learning Based Network Traffic Predictive Analysis. *Review of Computer Engineering Research*. 2022. Vol. 9. P. 96–108. URL: https://doi.org/10.18488/76.v9i2.3065 (дата звернення: 09.03.2026).

93. Hardegen C., Pfulb B., Rieger S., Gepperth A. Predicting Network Flow Characteristics Using Deep Learning and Real-World Network Traffic. *IEEE Transactions on Network and Service Management*. 2020. Vol. 17. P. 2662–2676. URL: https://doi.org/10.1109/TNSM.2020.3025131 (дата звернення: 09.03.2026).

94. Bolakhrif A., Ozger M., Sandberg D., Cavdar C. AI-Assisted Network Traffic Prediction Without Warm-Up Periods. In: *Proceedings of the 2022 IEEE 95th Vehicular Technology Conference (VTC2022-Spring)*. Helsinki, Finland, 2022. P. 1–6. URL: https://doi.org/10.1109/vtc2022-spring54318.2022.9860997 (дата звернення: 09.03.2026).

95. Adeke J.M., Chen J., Zhang L., Mensah R.N.K., Tong K. An Efficient Approach Based on Parameter Optimization for Network Traffic Classification Using Machine Learning. In: *Proceedings of the 2020 7th International Conference on*

Dependable Systems and Their Applications (DSA). Xi'an, China, 2020. P. 99–107. URL: https://doi.org/10.1109/dsa51864.2020.00021 (дата звернення: 09.03.2026).

96. Aouedi O., Piamrat K. F-BIDS: Federated-Blending Based Intrusion Detection System. *Pervasive and Mobile Computing.* 2023. Vol. 89. Art. 101750. URL: https://doi.org/10.1016/j.pmcj.2023.101750 (дата звернення: 09.03.2026).

97. Xie Q., Guo T., Chen Y., Xiao Y., Wang X., Zhao B.Y. Deep Graph Convolutional Networks for Incident-Driven Traffic Speed Prediction. In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management.* Online, 2020. URL: https://doi.org/10.1145/3340531.3411873 (дата звернення: 09.03.2026).

98. Han Z., Guan J., Yao Y., Yao S. Adaptive Convolutional Neural Network Structure for Network Traffic Classification. In: *Proceedings of the 2021 IEEE 27th International Conference on Parallel and Distributed Systems (ICPADS).* Beijing, China, 2021. P. 249–256. URL: https://doi.org/10.1109/HORA52670.2021.9461273 (дата звернення: 09.03.2026).

99. Hammad M., Hewahi N., Elmedany W. T-SNERF: A Novel High Accuracy Machine Learning Approach for Intrusion Detection Systems. *IET Information Security.* 2021. Vol. 15. P. 178–190. URL: https://doi.org/10.1049/ise2.12020 (дата звернення: 09.03.2026).

100. Sarker I.H., Kayes A.S.M., Badsha S., Alqahtani H., Watters P., Ng A. Cybersecurity Data Science: An Overview from Machine Learning Perspective. *Journal of Big Data.* 2020. Vol. 7. Art. 41. URL: https://doi.org/10.1049/ise2.12020 (дата звернення: 09.03.2026).

101. Jamalipour A., Murali S. A Taxonomy of Machine-Learning-Based Intrusion Detection Systems for the Internet of Things: A Survey. *IEEE Internet of Things Journal.* 2021. Vol. 9. P. 9444–9466. URL: https://doi.org/10.1109/JIOT.2021.3126811 (дата звернення: 09.03.2026).

102. Hwang R.-H., Peng M.-C., Huang C.-W., Lin P.-C., Nguyen V.-L. An Unsupervised Deep Learning Model for Early Network Traffic Anomaly Detection. *IEEE*

Access. 2020. Vol. 8. P. 30387–30399. URL: https://doi.org/10.1109/ACCESS.2020.2973023 (дата звернення: 09.03.2026).

103. Altunay H.C., Albayrak Z., Özalp A.N., Çakmak M. Analysis of Anomaly Detection Approaches Performed Through Deep Learning Methods in SCADA Systems. In: *Proceedings of the 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA).* Ankara, Turkey, 2021. P. 1–6. URL: https://doi.org/10.1109/HORA52670.2021.9461273 (дата звернення: 09.03.2026).

104. Balla A., Habaebi M.H., Elsheikh E.A., Islam M.R., Suliman F.M. The Effect of Dataset Imbalance on the Performance of SCADA Intrusion Detection Systems. *Sensors.* 2023. Vol. 23, no. 2. Art. 758. URL: https://doi.org/10.3390/s23020758 (дата звернення: 09.03.2026).

105. Nedeljkovic D., Jakovljevic Z. CNN Based Method for the Development of Cyber-Attacks Detection Algorithms in Industrial Control Systems. *Computers & Security.* 2022. Vol. 114. Art. 102585. URL: https://doi.org/10.1016/j.cose.2021.102585 (дата звернення: 09.03.2026).

106. Qian J., Du X., Chen B., Qu B., Zeng K., Liu J. Cyber-Physical Integrated Intrusion Detection Scheme in SCADA System of Process Manufacturing Industry. *IEEE Access.* 2020. Vol. 8. P. 147471–147481. URL: https://doi.org/10.1109/ACCESS.2020.3015900 (дата звернення: 09.03.2026).

107. Jamoos M., Mora A.M., AlKhanafseh M., Surakhi O. A New Data-Balancing Approach Based on Generative Adversarial Network for Network Intrusion Detection System. *Electronics.* 2023. Vol. 12, no. 13. Art. 2851. URL: https://doi.org/10.3390/electronics12132851 (дата звернення: 09.03.2026).

108. Al-Asiri M., El-Alfy E.-S.M. On Using Physical Based Intrusion Detection in SCADA Systems. *Procedia Computer Science.* 2020. Vol. 170. P. 34–42. URL: https://doi.org/10.1016/j.procs.2020.03.007 (дата звернення: 09.03.2026).

109. Ding P., Li J., Wen M., Wang L., Li H. Efficient BiSRU Combined with Feature Dimensionality Reduction for Abnormal Traffic Detection. *IEEE Access.* 2020. Vol. 8. P. 164414–164427. URL: https://doi.org/10.1109/ACCESS.2020.3022355 (дата звернення: 09.03.2026).

110. Mubarak S., Habaebi M.H., Islam M.R., Balla A., Tahir M., Elsheikh E.A.A., Suliman F.M. Industrial Datasets with ICS Testbed and Attack Detection Using Machine Learning Techniques. *Intelligent Automation & Soft Computing.* 2022. Vol. 31, no. 3. P. 1345–1360. URL: https://doi.org/10.32604/iasc.2022.020801 (дата звернення: 09.03.2026).

111. Mubarak S., Habaebi M.H., Islam M.R., Abdul Rahman F.D., Tahir M. Anomaly Detection in ICS Datasets with Machine Learning Algorithms. *Computer Systems Science and Engineering.* 2021. Vol. 37, no. 1. P. 33–46. URL: https://doi.org/10.32604/csse.2021.014384 (дата звернення: 09.03.2026).

112. Liao X., Li K., Zhu X., Liu K.J.R. Robust Detection of Image Operator Chain with Two-Stream Convolutional Neural Network. *IEEE Journal of Selected Topics in Signal Processing.* 2020. Vol. 14, no. 5. P. 955–968. URL: https://doi.org/10.1109/JSTSP.2020.3002391 (дата звернення: 09.03.2026).

113. Alotaibi A., Rassam M.A. Enhancing the Sustainability of Deep-Learning-Based Network Intrusion Detection Classifiers against Adversarial Attacks. *Sustainability.* 2023. Vol. 15, no. 12. Art. 9801. URL: https://doi.org/10.3390/su15129801 (дата звернення: 09.03.2026).

114. Mari A.-G., Zinca D., Dobrota V. Development of a Machine-Learning Intrusion Detection System and Testing of Its Performance Using a Generative Adversarial Network. *Sensors.* 2023. Vol. 23, no. 3. Art. 1315. URL: https://doi.org/10.3390/s23031315 (дата звернення: 09.03.2026).

115. Vo H.V., Du H.P., Nguyen H.N. APELID: Enhancing Real-Time Intrusion Detection with Augmented WGAN and Parallel Ensemble Learning. *Computers & Security.* 2024. Vol. 136. Art. 103567. URL: https://doi.org/10.1016/j.cose.2023.103567 (дата звернення: 09.03.2026).

116. Ling J., Zhu Z., Luo Y., Wang H. An Intrusion Detection Method for Industrial Control Systems Based on Bidirectional Simple Recurrent Unit. *Computers & Electrical Engineering.* 2021. Vol. 91. Art. 107049. URL: https://doi.org/10.1016/j.compeleceng.2021.107049 (дата звернення: 09.03.2026).

117. Ayodele B., Buttigieg V. SDN as a Defence Mechanism: A Comprehensive Survey. *International Journal of Information Security.* 2024. Vol. 23. P. 141–185. URL: https://doi.org/10.1007/s10207-023-00764-1 (дата звернення: 09.03.2026).

118. Khan S. Detection of DoS and DDoS Attacks on 5G Network Slices Using Deep Learning Approach. *Ph.D. Thesis.* University of Regina. Regina, SK, Canada, 2023. URL: https://scholar.google.com/scholar_lookup?title=Detection+of+DoS+and+DDoS+Attacks+on+5G+Network+Slices+Using+Deep+Learning+Approach&author=Khan,+S.&publication_year=2023 (дата звернення: 09.03.2026).

119. Yungaicela-Naula N.M., Vargas-Rosales C., Pérez-Díaz J.A. SDN/NFV-Based Framework for Autonomous Defense against Slow-Rate DDoS Attacks by Using Reinforcement Learning. *Future Generation Computer Systems.* 2023. Vol. 149. P. 637–649. URL: https://doi.org/10.1016/j.future.2023.08.007 (дата звернення: 09.03.2026).

120. Shoaib F., Chow Y.-W., Vlahu-Gjorgievska E., Nguyen C. Mitigating Timing Side-Channel Attacks in Software-Defined Networks: Detection and Response. *Telecom.* 2023. Vol. 4. P. 877–900. URL: https://doi.org/10.3390/telecom4040038 (дата звернення: 09.03.2026).

121. Sheibani M., Konur S., Awan I. DDoS Attack Detection and Mitigation in Software-Defined Networking-Based 5G Mobile Networks with Multiple Controllers. In: *Proceedings of the 9th International Conference on Future Internet of Things and Cloud (FiCloud).* Rome, Italy, 2022. P. 32–39. URL: https://scholar.google.com/scholar_lookup?title=DDoS+Attack+Detection+and+Mitigation+in+Software-Defined+Networking-Based+5G+Mobile+Networks+with+Multiple+Controllers&conference=Proceedings+of+the+9th+International+Conference+on+Future+Internet+of+Things+and+Cloud+(FiCloud)&author=Sheibani,+M.&author=Konur,+S.&author=Awan,+I.&publication_year=2022&pages=32%E2%80%9339 (дата звернення: 09.03.2026).

122. Yungaicela-Naula N.M., Vargas-Rosales C., Perez-Diaz J.A. SDN-Based Architecture for Transport and Application Layer DDoS Attack Detection by Using

Machine and Deep Learning. *IEEE Access.* 2021. Vol. 9. P. 108495–108512. URL: https://doi.org/10.1109/ACCESS.2021.3101650 (дата звернення: 09.03.2026).

123. Sun P., Guo Z., Wang G., Lan J., Hu Y. MARVEL: Enabling Controller Load Balancing in Software-Defined Networks with Multi-Agent Reinforcement Learning. *Computer Networks.* 2020. Vol. 177. Art. 107230. URL: https://doi.org/10.1016/j.comnet.2020.107230 (дата звернення: 09.03.2026).

124. Zaman S., Tauqeer H., Ahmad W., Shah S.M.A., Ilyas M. Implementation of Intrusion Detection System in the Internet of Things: A Survey. In: *Proceedings of the 2020 IEEE 23rd International Multitopic Conference (INMIC).* Bahawalpur, Pakistan, 2020. P. 1–6. URL: https://doi.org/10.1109/INMIC50486.2020.9318047 (дата звернення: 09.03.2026).

125. Sharma P. Critical Review of Various Intrusion Detection Techniques for Internet of Things. In: *Proceedings of the 2nd International Conference on Data, Engineering and Applications (IDEA).* Bhopal, India, 2020. P. 1–6. URL: https://doi.org/10.1109/IDEA49133.2020.9170732 (дата звернення: 09.03.2026).

126. Rahim R., Ahanger A.S., Khan S.M., Ma F. Analysis of IDS using Feature Selection Approach on NSL-KDD Dataset. In: *Proceedings of the SCRS Conference on Intelligent Systems.* Bangalore, India, 2022. P. 475–481. URL: https://doi.org/10.52458/978-93-91842-08-6-45 (дата звернення: 09.03.2026).

127. Семенюк Б., Каштальян А. Виявлення комп’ютерних атак із використанням 2D-CNN та соніфікації мережного трафіку. Information Technology: *Computer Science, Software Engineering and Cyber Security.* 2025. С. 193–201. DOI: https://doi.org/10.32782/IT/2025-1-26

128. Семенюк Б. В., Савенко Б. О. Метод виявлення комп’ютерних атак на основі адаптивності та балансування навчальної вибірки. *Вісник Черкаського державного технологічного університету*. 2026. № 1. С. 34-43. DOI: https://doi.org/10.62660/bcstu/1.2026.3

129. Семенюк Б. В., Савенко Б. О. Метод підвищення точності системи виявлення атак у зонах невизначеності на основі аналізу помилок та селективного

перевизначення рішень. *Системні технології*. 2026. № 2. С. 99-110. DOI: https://doi.org/10.34185/1562-9945-2-163-2026-09

130. Семенюк Б., Корецька Л. Особливості проєктування та дослідження системи виявлення вторгнень на основі соніфікації мережевого трафіку. *Measuring and Computing Devices in Technological Processes.* 2026. № 1. С. 246–254. DOI: https://doi.org/10.31891/2219-9365-2026-85-31

131. Семенюк Б., Стецюк Ю. Дослідження впливу синтетичного балансування навчальної вибірки на точність виявлення комп'ютерних атак. *Measuring and Computing Devices in Technological Processes*. 2026. № 2. С. 410–417. DOI: https://doi.org/10.31891/2219-9365-2026-86-48

132. Semeniuk B., Kashtalian A., Martiniuk D., Drozd A., Abdel-Badeeh M. Salem. Detection of Computer Attacks Based on Sonification of Network Traffic. Intelitsis'25: *The 6th International Workshop on Intelligent Information Technologies & Systems of Information Security*, April 04, 2025, Khmelnytskyi, Ukraine. Pp. 252–263. URL: https://ceur-ws.org/Vol-3963/paper21.pdf

133. Свідоцтво про реєстрацію авторського права на твір (програму) № 145847 Україна. Комп'ютерна програма «Sonified IDS» / Семенюк Б.В., Нічепорук А.О., Савенко О.С. Дата реєстрації 24.04.2026. URL: https://sis.nipo.gov.ua/uk/search/detail/1925939/

ДОДАТКИ

ДОДАТОК А

Список публікацій здобувача

Наукові праці, в яких опубліковані основні наукові результати дисертації

1. Семенюк Б., Каштальян А. Виявлення комп'ютерних атак із використанням 2D-CNN та соніфікації мережного трафіку. *Information Technology: Computer Science, Software Engineering and Cyber Security*. 2025. С. 193–201. DOI: https://doi.org/10.32782/IT/2025-1-26

2. Семенюк Б. В., Савенко Б. О. Метод виявлення комп'ютерних атак на основі адаптивності та балансування навчальної вибірки. *Вісник Черкаського державного технологічного університету*. 2026. № 1. С. 34-43. DOI: https://doi.org/10.62660/bcstu/1.2026.3

3. Семенюк Б. В., Савенко Б. О. Метод підвищення точності системи виявлення атак у зонах невизначеності на основі аналізу помилок та селективного перевизначення рішень. *Системні технології*. 2026. № 2. С. 99-110. DOI: https://doi.org/10.34185/1562-9945-2-163-2026-09

4. Семенюк Б., Корецька Л. Особливості проєктування та дослідження системи виявлення вторгнень на основі соніфікації мережевого трафіку. *Measuring and Computing Devices in Technological Processes*. 2026. № 1. С. 246–254. DOI: https://doi.org/10.31891/2219-9365-2026-85-31

5. Семенюк Б., Стецюк Ю. Дослідження впливу синтетичного балансування навчальної вибірки на точність виявлення комп'ютерних атак. *Measuring and Computing Devices in Technological Processes*. 2026. № 2. С. 410–417. DOI: https://doi.org/10.31891/2219-9365-2026-86-48

Праці, які засвідчують апробацію матеріалів дисертації

6. Semeniuk B., Kashtalian A., Martiniuk D., Drozd A., Abdel-Badeeh M. Salem. Detection of computer attacks based on sonification of network traffic. *Intelitsis'25: The 6th International Workshop on Intelligent Information Technologies & Systems of*

Information Security, April 04, 2025, Khmelnytskyi, Ukraine. Pp. 252-263. URL: https://ceur-ws.org/Vol-3963/paper21.pdf *(Scopus)*

Публікації, які додатково відображають наукові результати дисертації

7. Свідоцтво про реєстрацію авторського права на твір (програму) № 145847 Україна. Комп'ютерна програма «Sonified IDS» / Семенюк Б.В., Нічепорук А.О., Савенко О.С. Дата реєстрації 24.04.2026. URL: https://sis.nipo.gov.ua/uk/search/detail/1925939/

ДОДАТОК Б

Акти впровадження

«Затверджую»
Директор ТОВ «ІТТ»
Сімогук В.С.
«__» червня 2026 р.

АКТ
про впровадження результатів дисертаційної роботи
Семенюка Богдана Васильовича за темою
«Моделі та методи виявлення комп'ютерних атак в корпоративних мережах на основі частотно-часового представлення мережного трафіку»

Результати дисертаційної роботи аспіранта кафедри комп'ютерної інженерії та інформаційних систем Хмельницького національного університету Семенюка Б.В. впроваджені в ТОВ «ІТТ».

У процесі впровадження методів та засобів виявлення і запобігання витокам конфіденційної інформації в корпоративній мережі ТОВ «ІТТ». були використані результати, отримані Семенюком Б.В. особисто:

1) розроблена модель процесу виявлення комп'ютерних атак на основі соніфікації мережного трафіку, особливістю якої є інтеграція звукового перетворення багатовимірного простору ознак, спектрального аналізу, адаптивного балансування навчальної вибірки та каскадного механізму прийняття рішень у межах узагальненої моделі процесу виявлення атак, що дає змогу узгодити етапи навчання і функціонування моделі та підвищити збалансованість класифікації в умовах дисбалансу класів;

2) розроблений метод частотно-часового перетворення мережного трафіку, який, на відміну від відомих векторних підходів до представлення ознак, передбачає інтерпретацію багатовимірного простору параметрів з'єднання як параметрів сигналу з лінійним частотно-амплітудним відображенням і подальшим спектральним аналізом, що дає змогу формувати структуроване 2D-подання трафіку та підвищити інформативність вхідних даних для задачі класифікації атак.

Упровадження зазначених результатів дозволило виявляти комп'ютерні атаки на основі соніфікації мережного трафіку. Розроблена модель процесу виявлення комп'ютерних атак у корпоративних мережах та методи частотно-часового перетворення мережного трафіку, сигнатурозбережного адаптивного балансування навчальної вибірки і селективного підвищення точності в зоні невизначеності прогнозу дають змогу формувати інформативне спектральне подання мережних з'єднань, застосовувати двовимірні згорткові нейронні мережі для автоматизованого виявлення атак та зменшувати негативний вплив дисбалансу класів на якість класифікації на 3-8 %.

Цей акт не є підставою для фінансових розрахунків.

Заступник директора Іванов О.В.

Системний адміністратор Веремеєнко В.А.

ДОВІДКА

про впровадження результатів дисертаційної роботи
аспіранта кафедри комп'ютерної інженерії та інформаційних систем
Хмельницького національного університету Семенюка Богдана Васильовича
за темою «Моделі та методи виявлення комп'ютерних атак в корпоративних мережах на основі частотно-часового представлення мережного трафіку»

Результати дисертаційної роботи аспіранта кафедри комп'ютерної інженерії та інформаційних систем Хмельницького національного університету Семенюка Б.В. впроваджені в ПП «Авіві».

У процесі впровадження методів та засобів виявлення і запобігання витокам конфіденційної інформації в корпоративній мережі ПП «Авіві». були використані результати, одержані Семенюком Б.В. особисто:

1) розроблено метод сигнатурозбережного адаптивного балансування навчальної вибірки, який, на відміну від відомих процедур надсемплінгу та підвибірки, характеризується збереженням структурних характеристик міноритарного класу та врахуванням помилок базової моделі під час формування розширеної навчальної множини, що дає змогу зменшити вплив дисбалансу класів на функцію ризику, підвищити повноту виявлення атак і збалансованість класифікації;

2) розроблено метод підвищення точності виявлення атак у зоні невизначеності прогнозу, який, на відміну від відомих порогових схем прийняття рішень, передбачає формалізацію τ-інтервалу невизначеності та селективне перевизначення рішень на основі допоміжної моделі, навченої на помилкових прикладах базового класифікатора, що дає змогу зменшити кількість хибнонегативних рішень без порушення принципу незалежності тестової вибірки.

Упровадження зазначених результатів в ПП «Авіві» дозволило підвищити рівень виявлення комп'ютерних атак в корпоративній мережі на основі частотно-часового представлення мережного трафіку на 7-11 %.

Ця довідка не є підставою для фінансових розрахунків.

29.06.2026 р.

Директор ПП «Авіві» _____ В. В. Аскеров

«Затверджую»
В.о. ректора Хмельницького
національного університету
Віктор ЛОПАТОВСЬКИЙ
«26» червня 2026 р.

АКТ

про впровадження в освітній процес Хмельницького національного університету результатів дисертаційної роботи аспіранта кафедри комп'ютерної інженерії та інформаційних систем Семенюка Богдана Васильовича

«Моделі та методи виявлення комп'ютерних атак в корпоративних мережах на основі частотно-часового представлення мережного трафіку»

Ми, комісія в складі: декана факультету інформаційних технологій, професора кафедри комп'ютерної інженерії та інформаційних систем, д.т.н., професора Говорущенко Т. О. (голова комісії), завідувача кафедри комп'ютерної інженерії та інформаційних систем, д.ф., доцента Павлової О. О., доцента кафедри комп'ютерної інженерії та інформаційних систем, к.т.н., доцента Капустян М. В. склала акт про те, що результати дисертаційної роботи Семенюка Б. впроваджені та використовуються в освітньому процесі Хмельницького національного університету при викладанні дисциплін на кафедрі комп'ютерної інженерії та інформаційних систем для здобувачів спеціальності F7 Комп'ютерна інженерія, зокрема в курсах «Безпека та захист комп'ютерних систем», «Моделювання та методи оптимізації в наукових та експериментальних дослідженнях», «Методології забезпечення якості, надійності, гарантоздатності та безпеки комп'ютерних систем та мереж».

При викладанні цих дисциплін викладачами кафедр використовувалися наступні матеріали досліджень, отримані Семенюком Б. особисто:

1) розроблено модель процесу виявлення комп'ютерних атак на основі соніфікації мережного трафіку, особливістю якої є інтеграція звукового перетворення багатовимірного простору ознак, спектрального аналізу, адаптивного балансування навчальної вибірки та каскадного механізму прийняття рішень у межах узагальненої моделі процесу виявлення атак, що дає змогу узгодити етапи навчання і функціонування моделі та підвищити збалансованість класифікації в умовах дисбалансу класів;

2) розроблено метод частотно-часового перетворення мережного трафіку, який, на відміну від відомих векторних підходів до представлення ознак, передбачає інтерпретацію багатовимірного простору параметрів

з'єднання як параметрів сигналу з лінійним частотно-амплітудним відображенням і подальшим спектральним аналізом, що дає змогу формувати структуроване 2D-подання трафіку та підвищити інформативність вхідних даних для задачі класифікації атак;

3) розроблено метод сигнатурозбережного адаптивного балансування навчальної вибірки, який, на відміну від відомих процедур надсемплінгу та підвибірки, характеризується збереженням структурних характеристик міноритарного класу та врахуванням помилок базової моделі під час формування розширеної навчальної множини, що дає змогу зменшити вплив дисбалансу класів на функцію ризику, підвищити повноту виявлення атак і збалансованість класифікації;

4) розроблено метод підвищення точності виявлення атак у зоні невизначеності прогнозу, який, на відміну від відомих порогових схем прийняття рішень, передбачає формалізацію τ-інтервалу невизначеності та селективне перевизначення рішень на основі допоміжної моделі, навченої на помилкових прикладах базового класифікатора, що дає змогу зменшити кількість хибнонегативних рішень без порушення принципу незалежності тестової вибірки.

Розроблена модель процесу виявлення комп'ютерних атак у корпоративних мережах та методи частотно-часового перетворення мережного трафіку, сигнатурозбережного адаптивного балансування навчальної вибірки і селективного підвищення точності в зоні невизначеності прогнозу дають змогу формувати інформативне спектральне подання мережних з'єднань, застосовувати двовимірні згорткові нейронні мережі для автоматизованого виявлення атак та зменшувати негативний вплив дисбалансу класів на якість класифікації.

Отримані матеріали досліджень дозволили розробити лабораторні практикуми з використанням архітектури засобів виявлення комп'ютерних атак у корпоративних мережах та методів частотно-часового перетворення мережного трафіку.

Говорущенко Т. О.
Павлова О. О.
Капустян М. В.

ДОДАТОК В

Результати експериментів

Таблиця В.1 – Порівняння метрик векторної та соніфікованої моделей

Модель	Точність	Прецизійність	Повнота	F1-міра	FPR	FNR
Векторна	0,7766	0,9249	0,6613	0,7712	0,0710	0,3387
Соніфікована	0,7741	0,9346	0,6486	0,7658	0,0599	0,3514

Таблиця В.2 – Результати соніфікованої моделі без балансування, з SMOTENC та з SPAS

Режим	Точність	Прецизійність	Повнота	F1-міра	FPR	FNR	TN	FP	FN	TP
Базовий режим (без балансування)	0,7566	0,9615	0,5963	0,7361	0,0315	0,4037	9405	306	5181	7652
SMOTENC	0,7949	0,9784	0,6541	0,7840	0,0191	0,3459	9526	185	4439	8394
SPAS	0,8457	0,9225	0,7958	0,8545	0,0884	0,2042	8853	858	2621	10212

Таблиця В.3 – Компоненти матриць невідповідностей для трьох режимів балансування

Режим	TN	FP	FN	TP
Базовий режим (без балансування)	9405	306	5181	7652
SMOTENC	9526	185	4439	8394
SPAS	8853	858	2621	10212

Таблиця В.4 – Порівняння обчислювальних витрат формування спектрограм та навчання

Режим	Кількість навчальних зразків	Час STFT (навч./валід./тест.), с	Час навчання, с
Базовий режим (без балансування)	113375	545,22	324,76
SMOTENC	125627	350,09	357,93
SPAS	125627	350,64	309,24

Таблиця В.5 – Параметри τ-зони та порогів селективного перевизначення, обрані на валідаційній множині

Параметр	Значення
Нижня межа τ-зони	0,40
Верхня межа τ-зони	0,60
Нижній поріг впевненості додаткової моделі	0,50
Верхній поріг впевненості додаткової моделі	0,70
Розмір τ-зони на валідації, записів	325
Кількість помилок у τ-зоні до перевизначення, записів	145
Кількість помилок у τ-зоні після перевизначення, записів	118
Частка записів у τ-зоні, для яких застосовано перевизначення	0,8615

Таблиця В.6 – Порівняння базової SPAS-моделі та каскадної змішаної каскадної моделі на тестовій множині

Модель	Точність	Прецизійність	Повнота	F1-міра
Базова модель (SPAS)	0,8212	0,9331	0,7389	0,8247
Змішана каскадна модель	0,8287	0,9293	0,7567	0,8342

Таблиця В.7 – Додаткові характеристики застосування каскадної схеми

Показник	Значення
Розмір τ-зони на валідації, записів	325
Розмір τ-зони на тестовій множині, записів	1466
Кількість помилок у τ-зоні до перевизначення на валідації, записів	145
Кількість помилок у τ-зоні після перевизначення на валідації, записів	118
Точність каскадної схеми	0,8287
Прецизійність каскадної схеми	0,9293
Повнота каскадної схеми	0,7567
F1-міра каскадної схеми	0,8342

ДОДАТОК Г

Додаткові результати

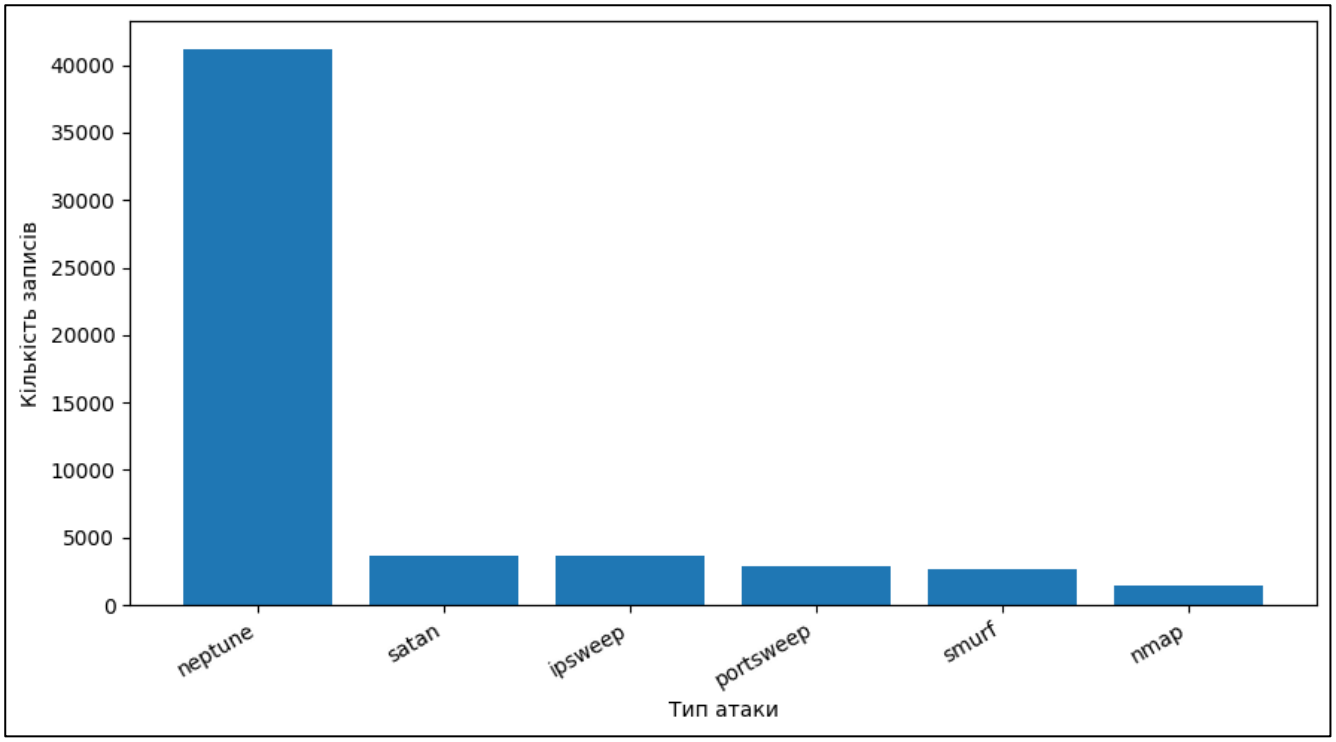

Рисунок Г.1 – Найпоширеніші типи атак у навчальній вибірці NSL-KDD

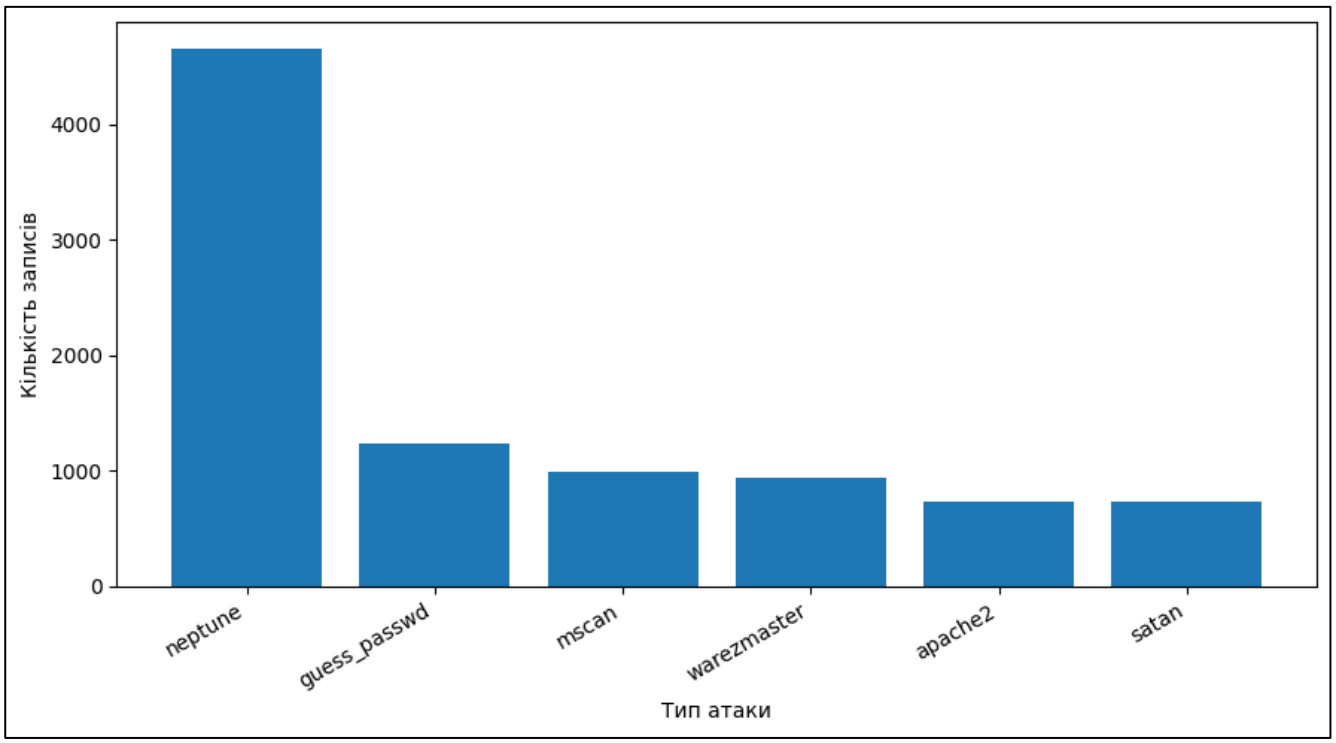

Рисунок Г.2 – Найпоширеніші типи атак у тестовій вибірці NSL-KDD

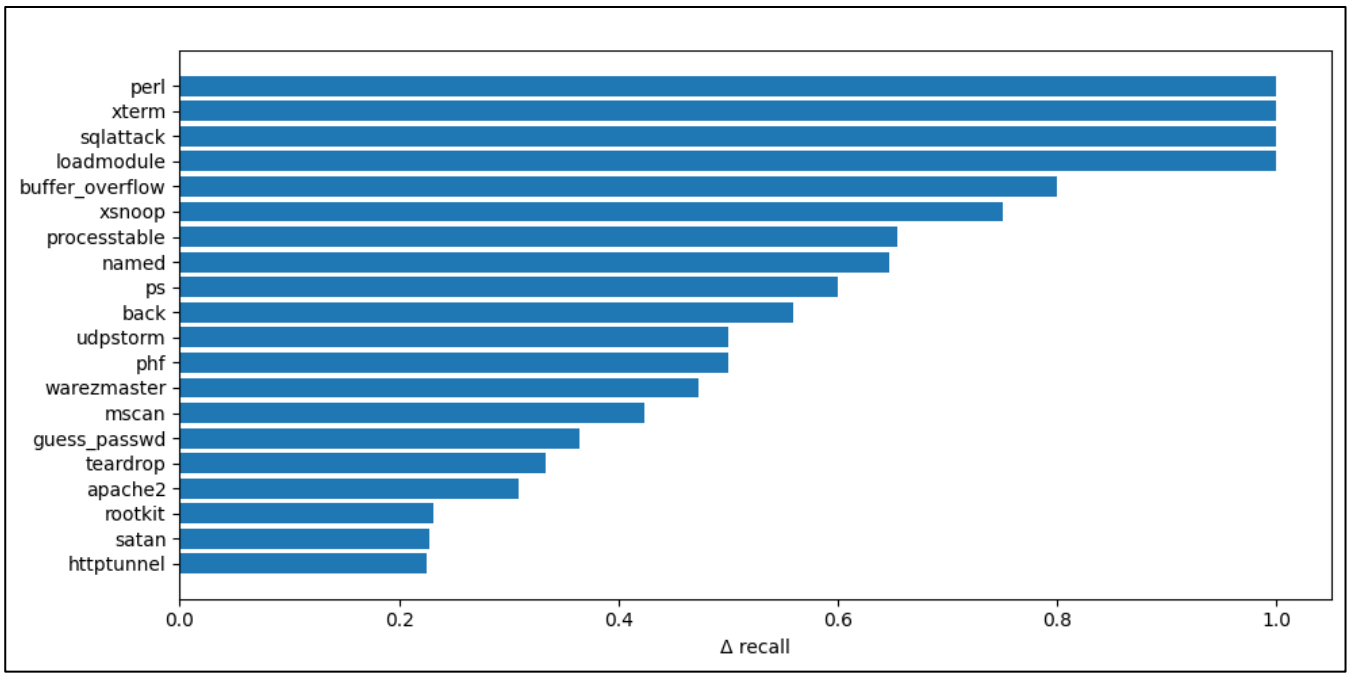

Рисунок Г.3 – Топ-20 типів атак у прирості точності

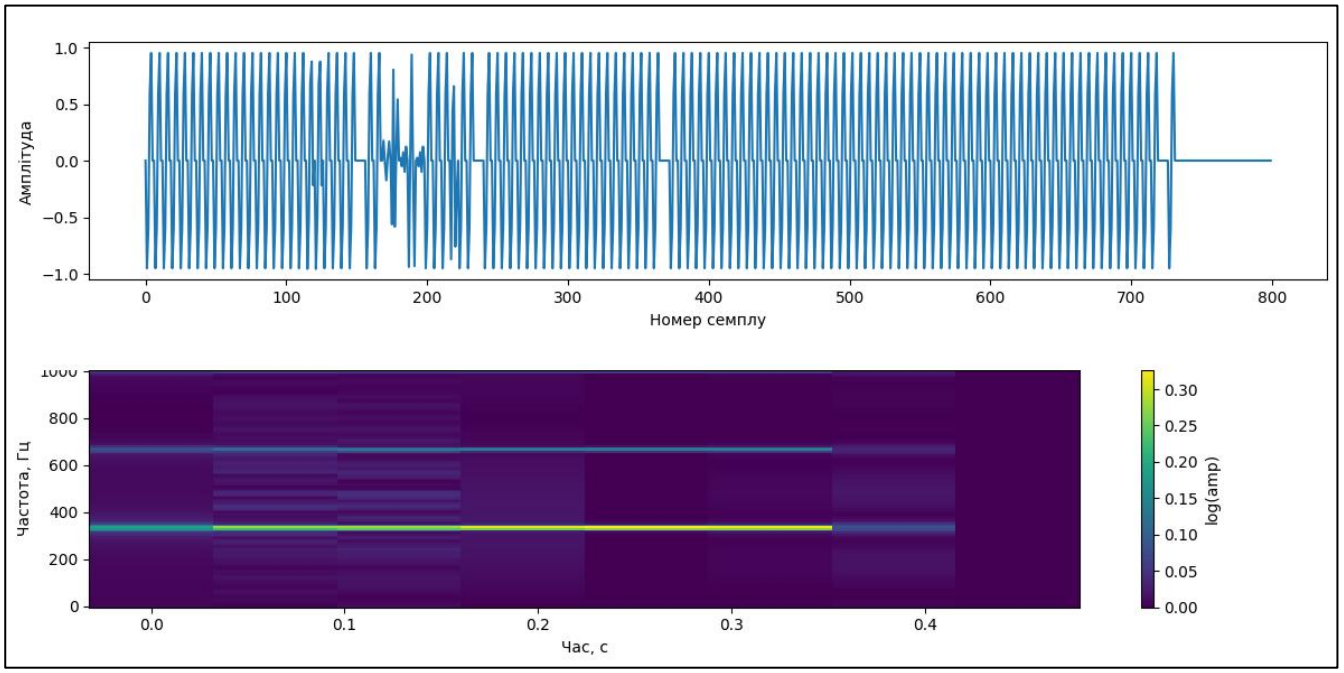

Рисунок Г.4 – Нормальний запис після соніфікації: хвильове представлення та спектрограма

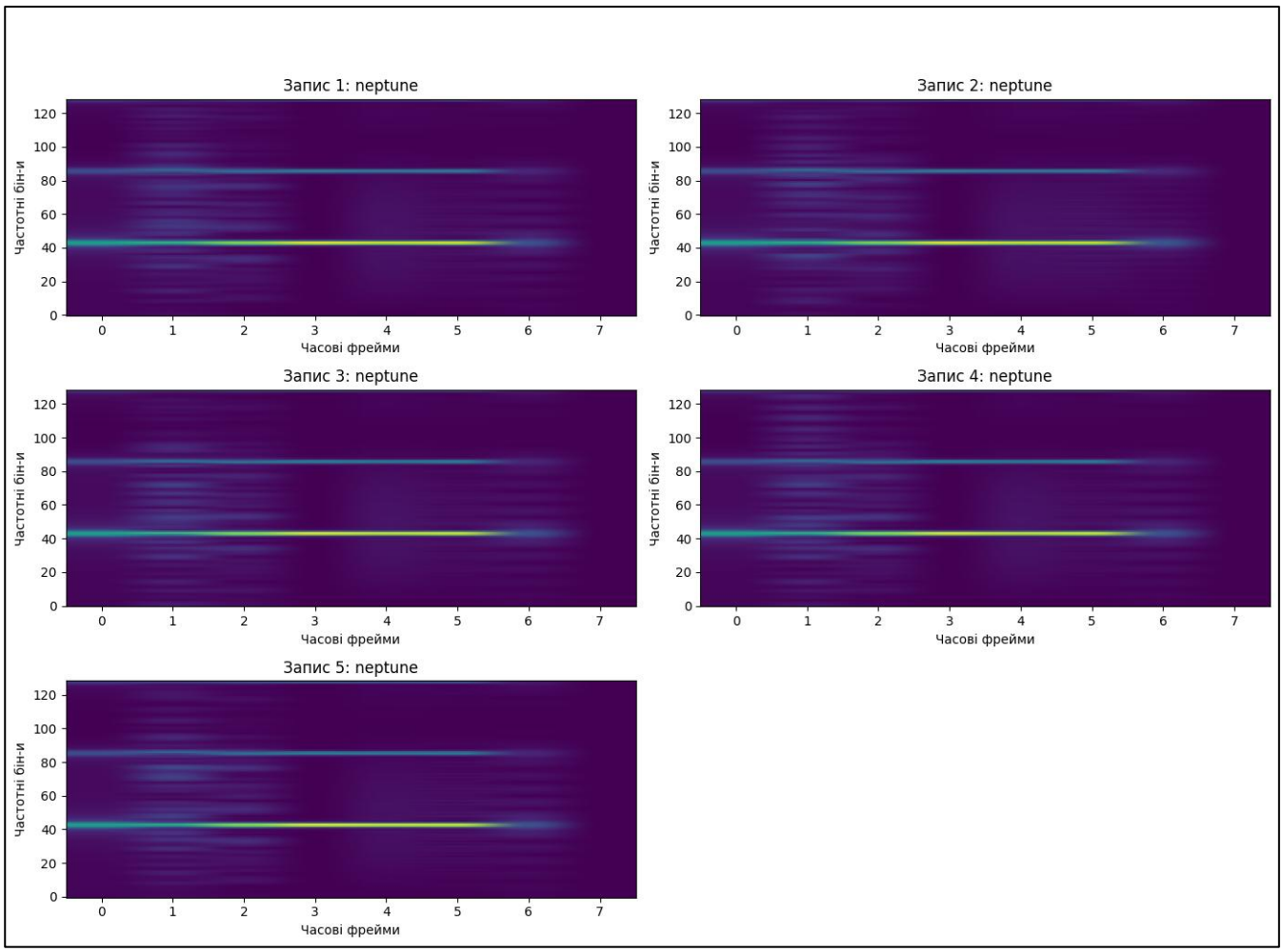

Рисунок Г.5 – Подібність спектрограм для 5 записів одного типу атаки

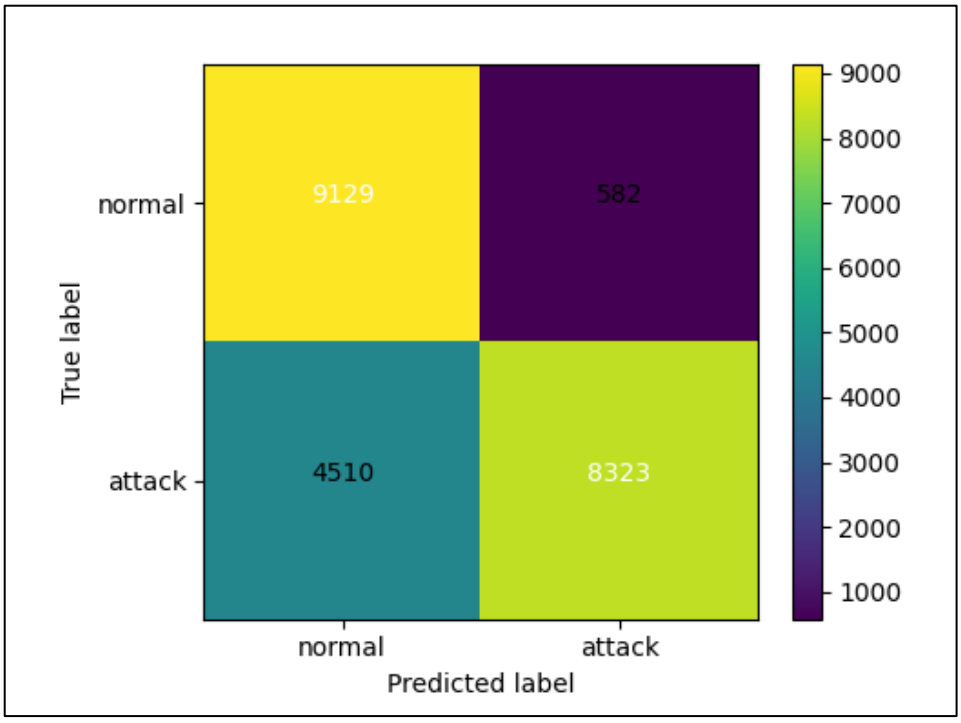

Рисунок Г.6 – Матриця невідповідності для соніфікованої моделі

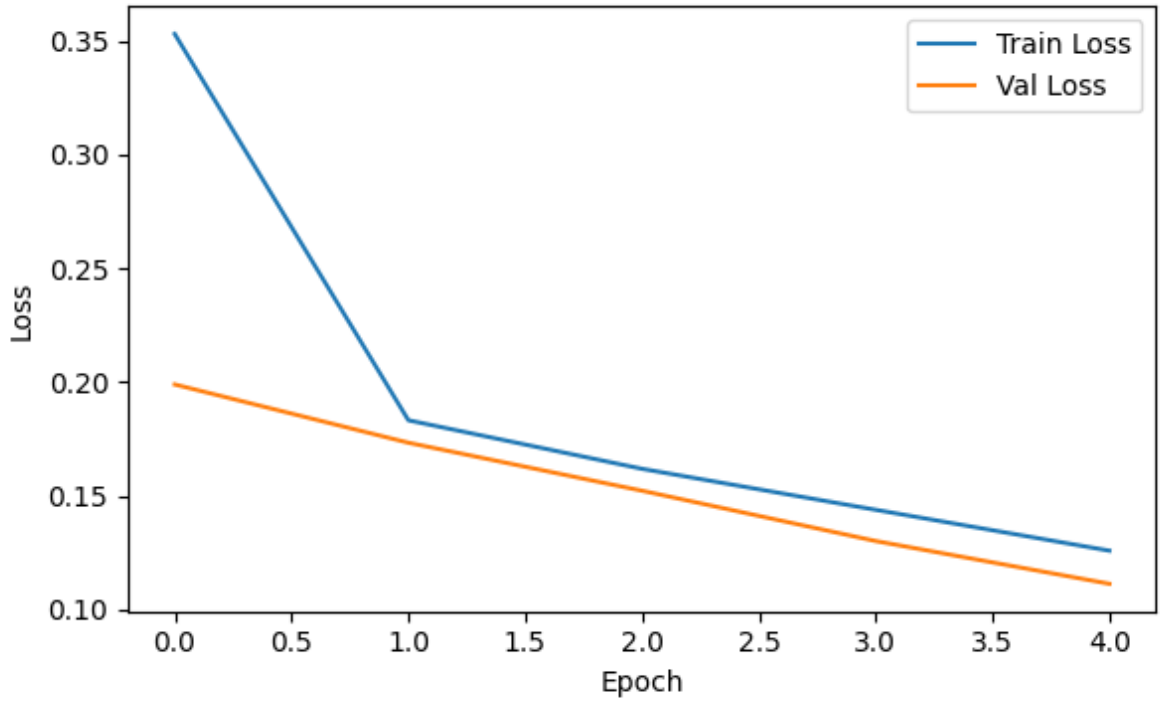

Рисунок Г.7 – Динаміка функції втрат базової соніфікованої моделі

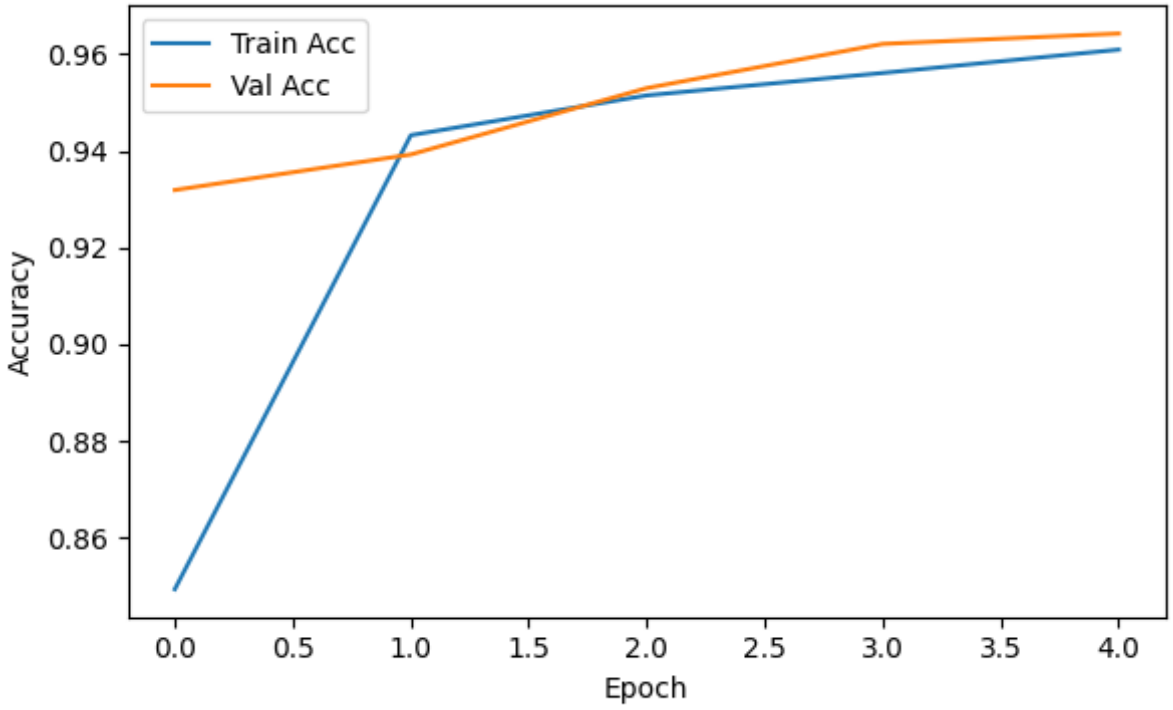

Рисунок Г.8 – Динаміка точності базової соніфікованої моделі

Таблиця Г.1 – Розподіл правильних і помилкових рішень моделей за інтервалами прогнозованої ймовірності

Модель	Інтервал ймовірності	Кількість прогнозів	Правильні	Помилкові
Векторна	[0,0; 0,1)	12256	8818	3438
Векторна	[0,1; 0,2)	495	133	362
Векторна	[0,2; 0,3)	374	43	331
Векторна	[0,3; 0,4)	122	17	105
Векторна	[0,4; 0,5)	122	11	111
Векторна	[0,5; 0,6)	134	123	11
Векторна	[0,6; 0,7)	105	95	10
Векторна	[0,7; 0,8)	179	174	5
Векторна	[0,8; 0,9)	385	376	9
Векторна	[0,9; 1,0]	8372	7718	654
Соніфікована	[0,0; 0,1)	10479	8368	2111
Соніфікована	[0,1; 0,2)	1663	372	1291
Соніфікована	[0,2; 0,3)	638	163	475
Соніфікована	[0,3; 0,4)	473	109	364
Соніфікована	[0,4; 0,5)	386	117	269
Соніфікована	[0,5; 0,6)	492	305	187
Соніфікована	[0,6; 0,7)	387	238	149
Соніфікована	[0,7; 0,8)	375	275	100
Соніфікована	[0,8; 0,9)	783	718	65
Соніфікована	[0,9; 1,0]	6868	6787	81

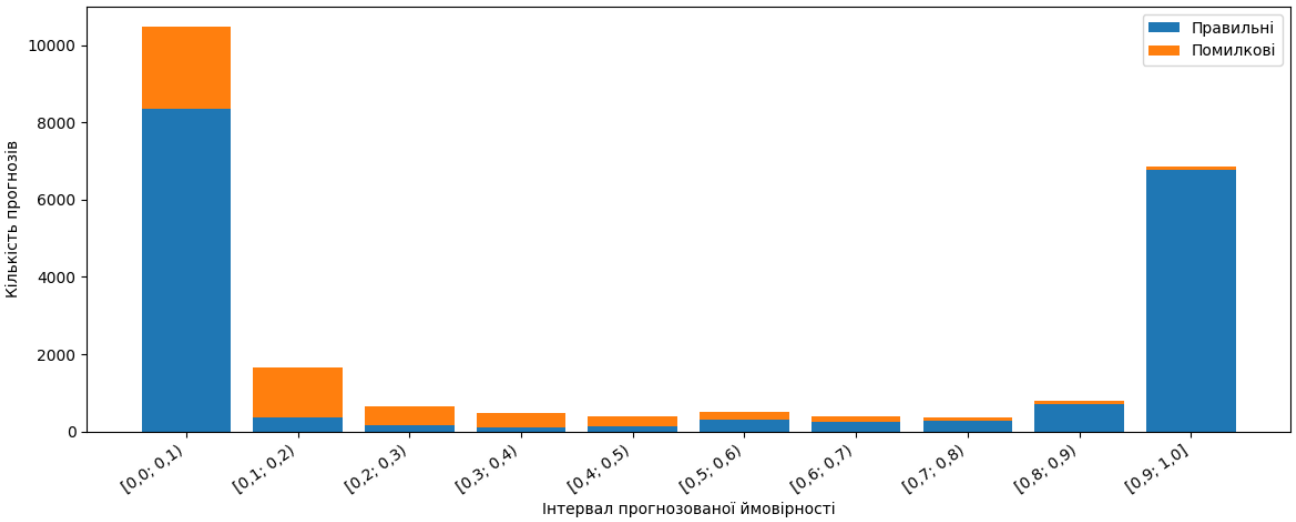

Рисунок Г.9 – Розподіл правильних і помилкових рішень соніфікованої моделі за рівнем впевненості

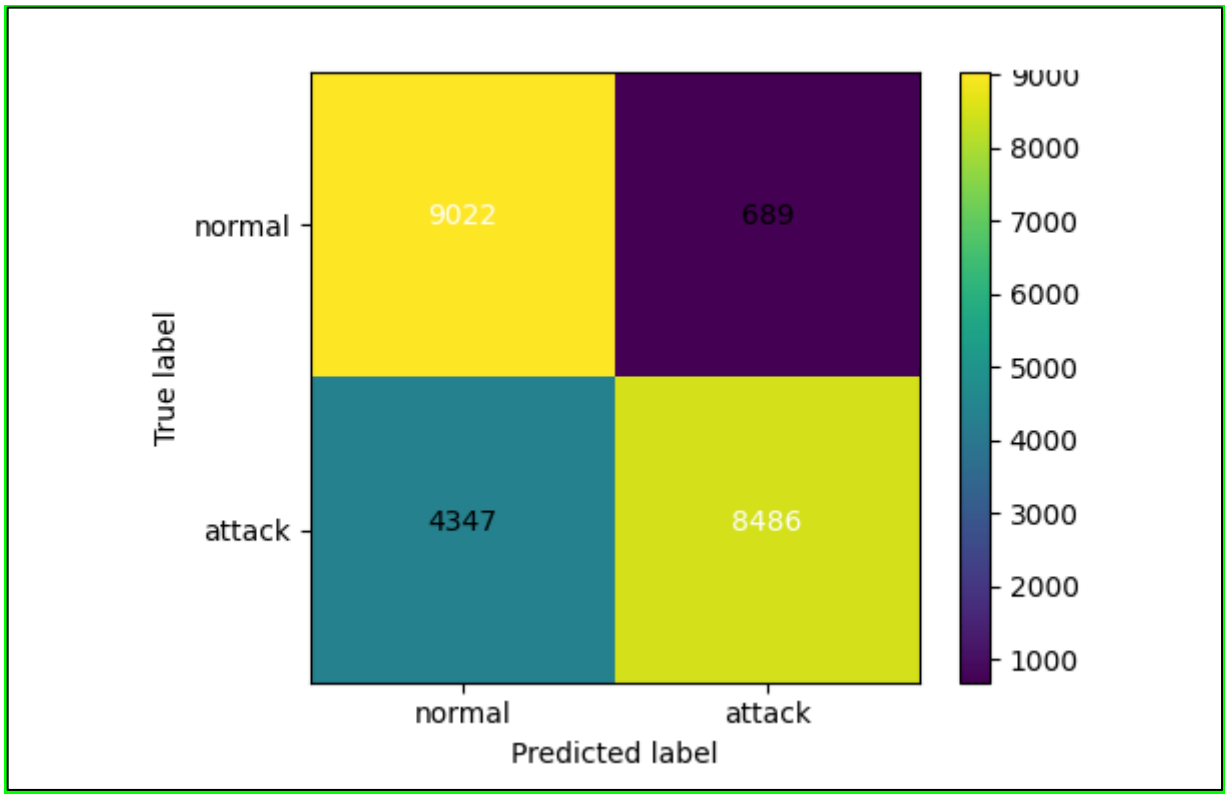

Рисунок Г.10 – Матриця невідповідностей векторної моделі

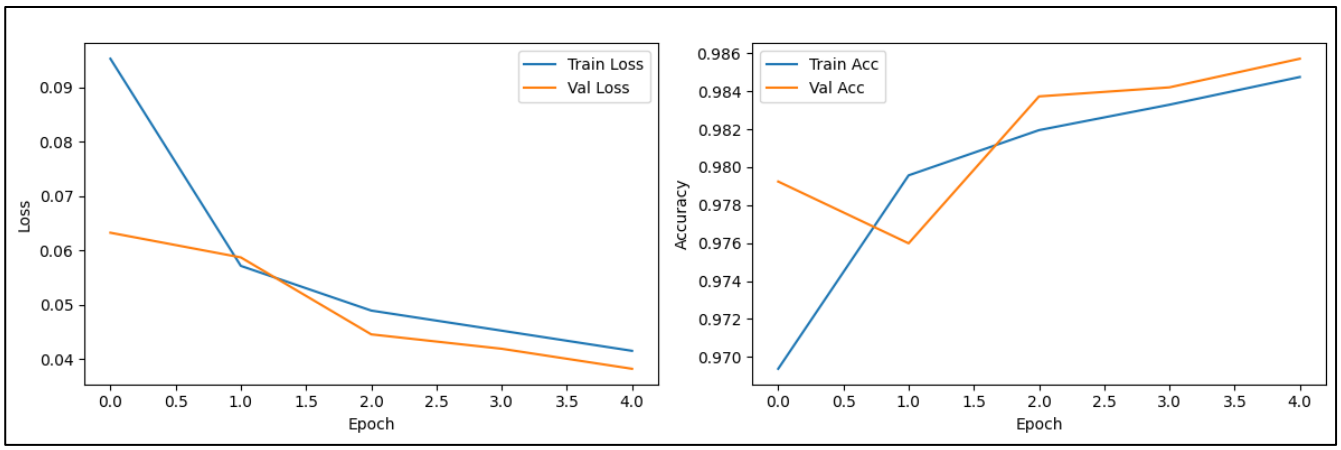

Рисунок Г.11 – Динаміка функції втрат та точності векторної моделі

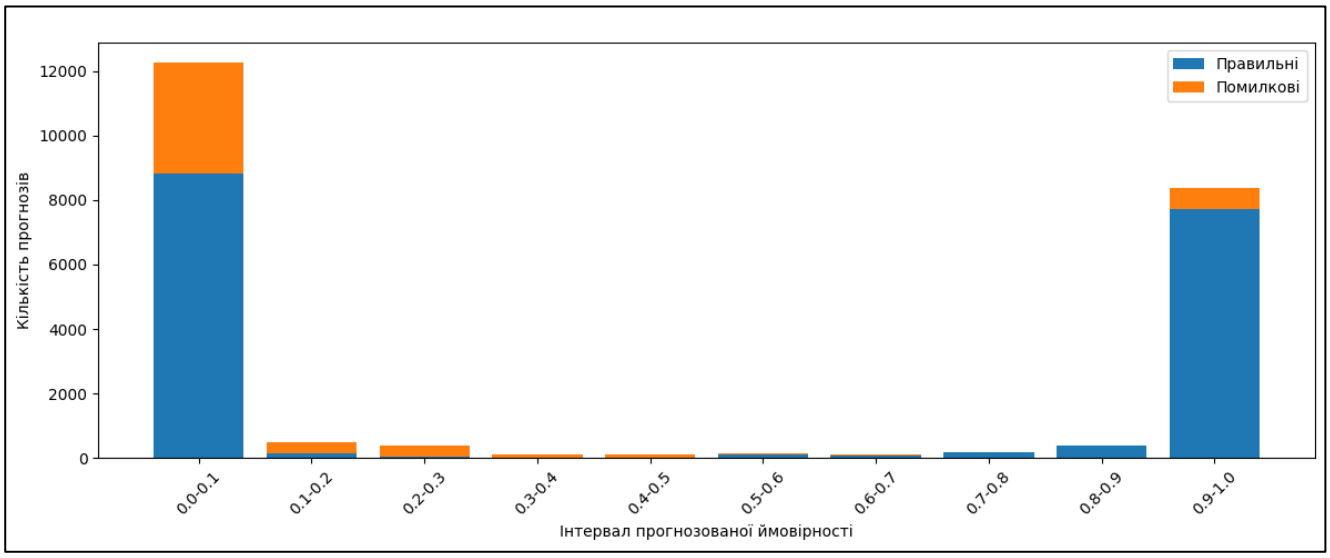

Рисунок Г.12 – Розподіл правильних і помилкових рішень векторної моделі за рівнем впевненості

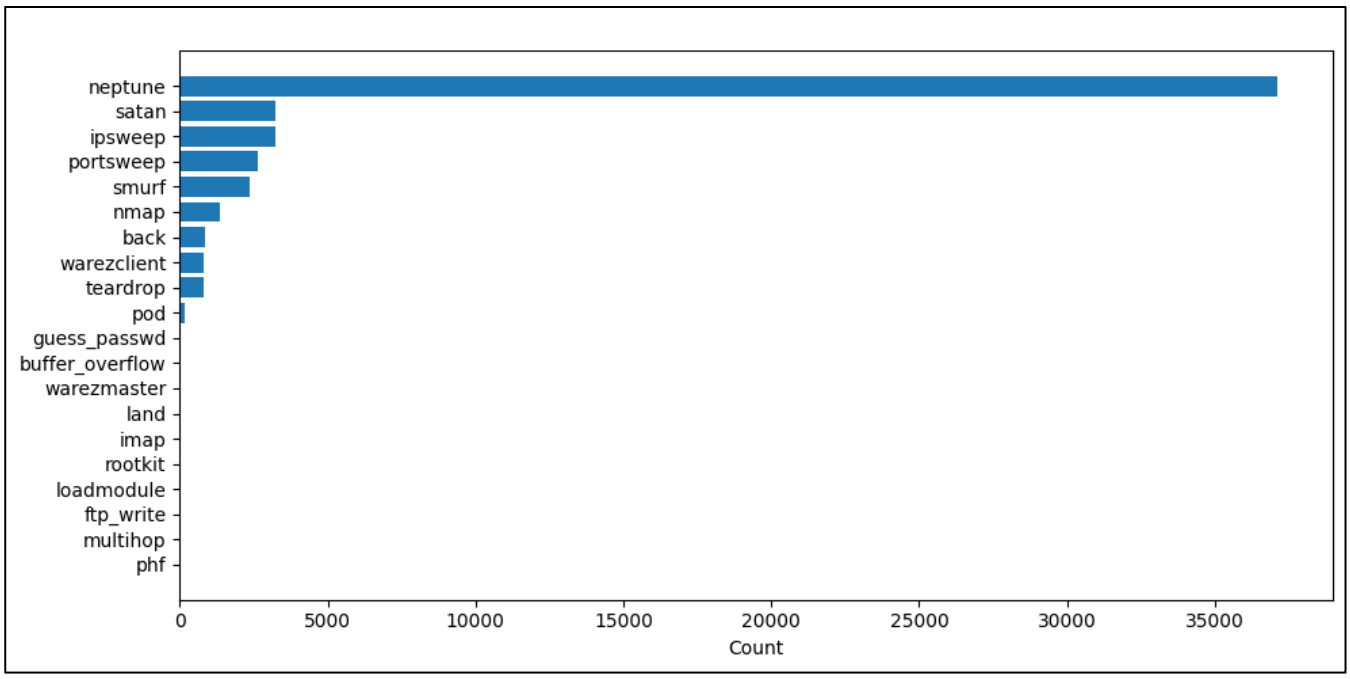

Рисунок Г.13 – Найпоширеніші типи атак у навчальній вибірці до балансування SPAS

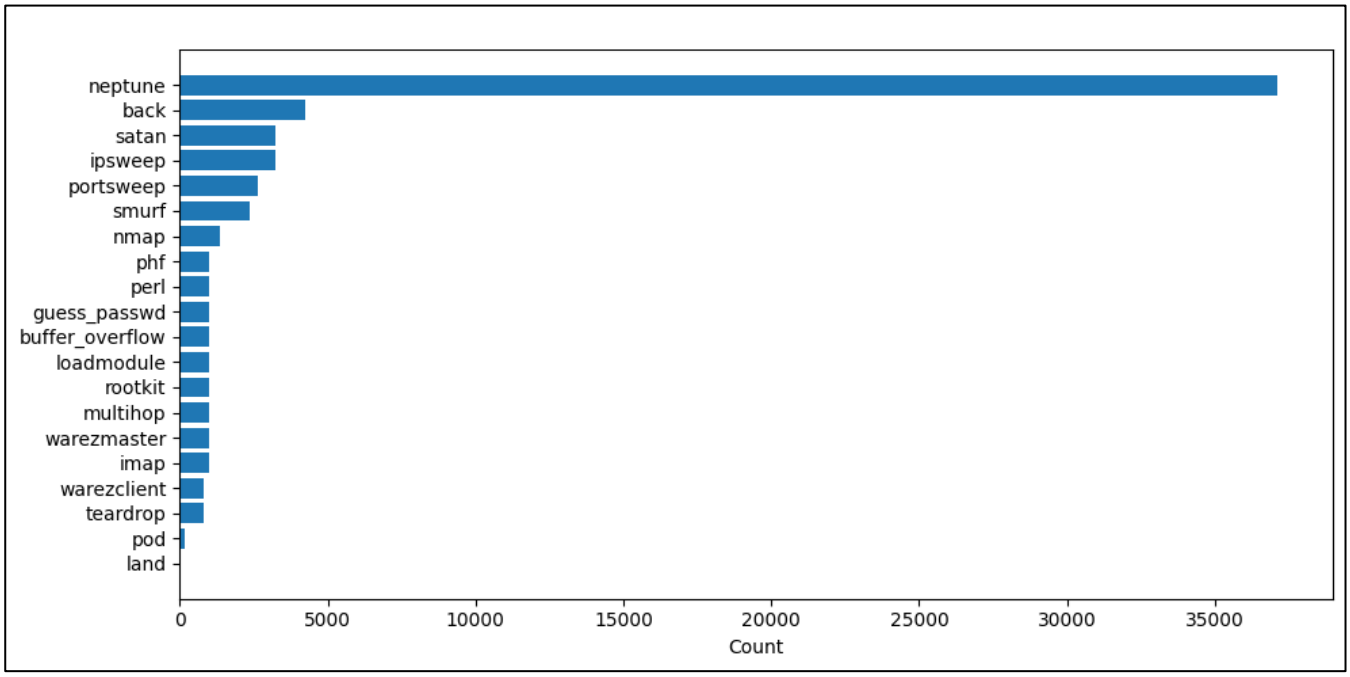

Рисунок Г.14 – Найпоширеніші типи атак у навчальній вибірці після балансування SPAS

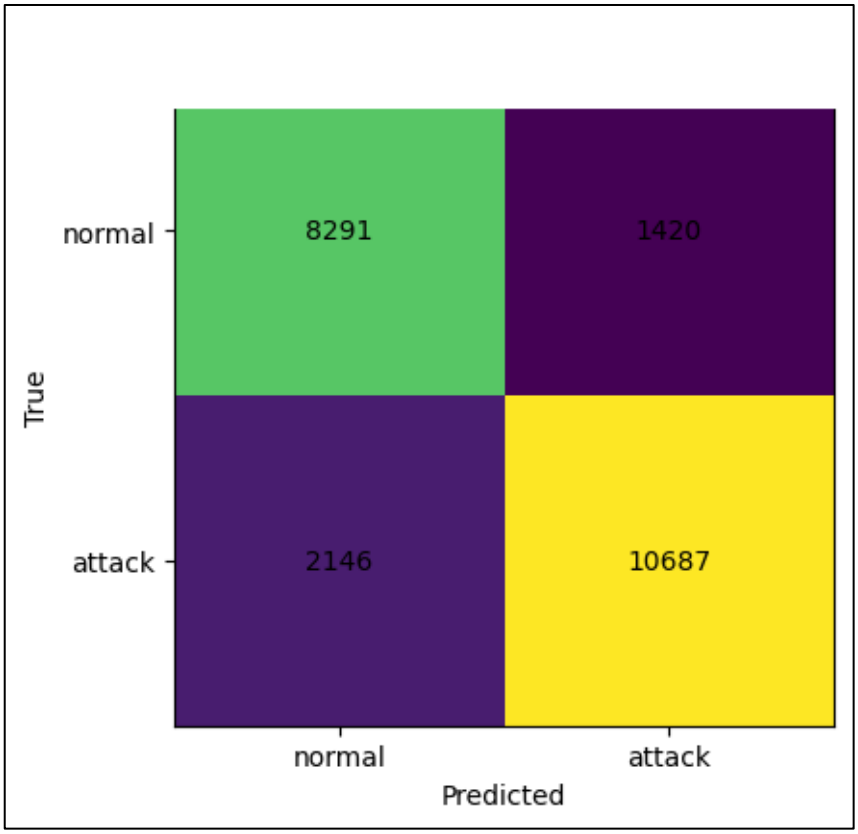

Рисунок Г.15 – Матриця невідповідностей додаткової моделі підсилення

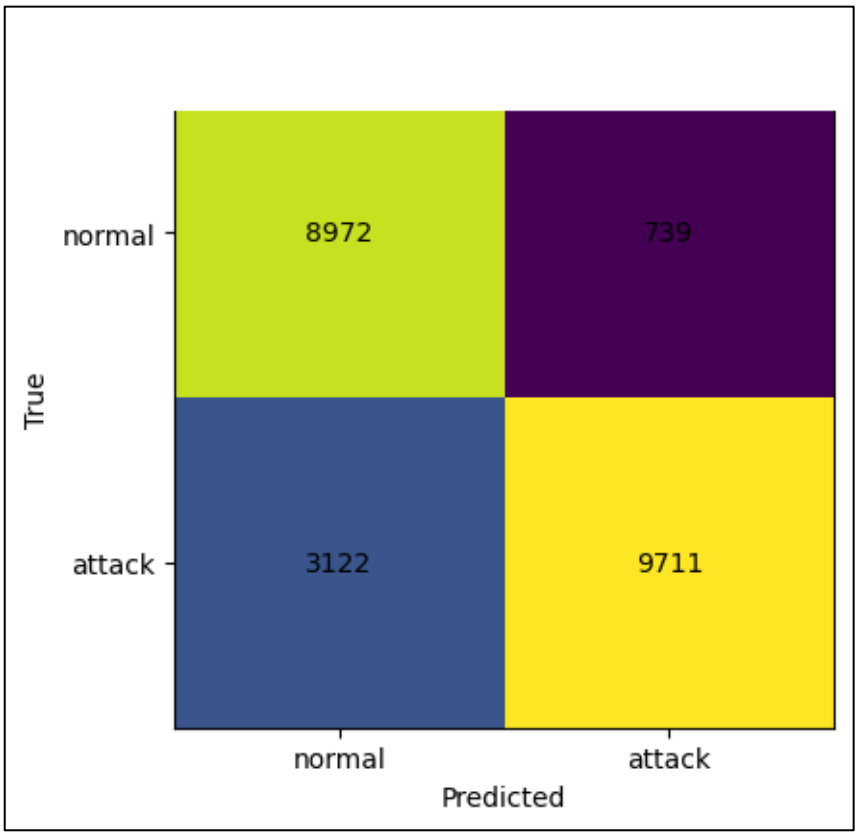

Рисунок Г.16 – Матриця невідповідностей змішаної каскадної схеми

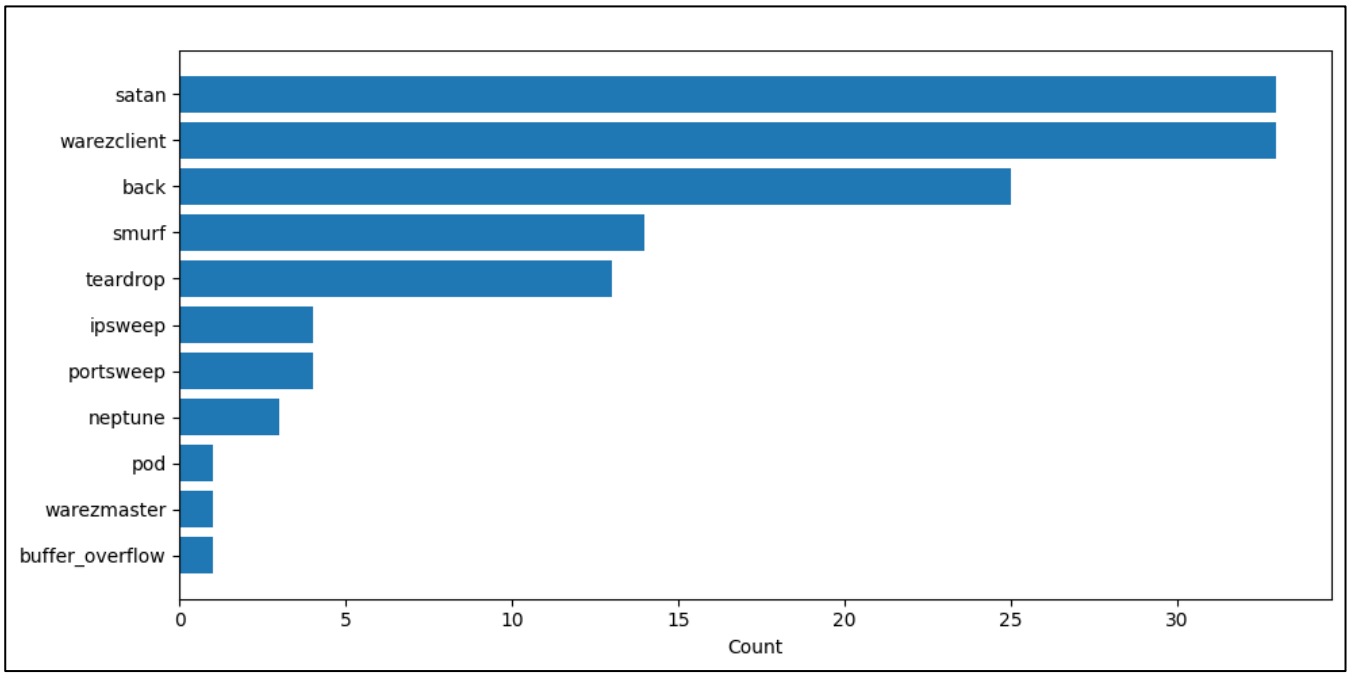

Рисунок Г.17 – Найпоширеніші підкласи атак у τ-зоні на валідаційній множині

ДОДАТОК Д

Вихідний програмний код

Вихідний програмний код, використаний під час проведення експериментальних досліджень, розміщено у відкритому репозиторії GitHub:

Semeniuk B. SPAS: Signature-Preserving Adaptive Sampling for Sonification-Based Intrusion Detection. GitHub. URL: https://github.com/Faludore/SPAS (дата звернення: 10.07.2026).

На рисунку Д.1 наведено структуру репозиторію з вихідним програмним кодом і файлами, необхідними для відтворення експериментальних досліджень.

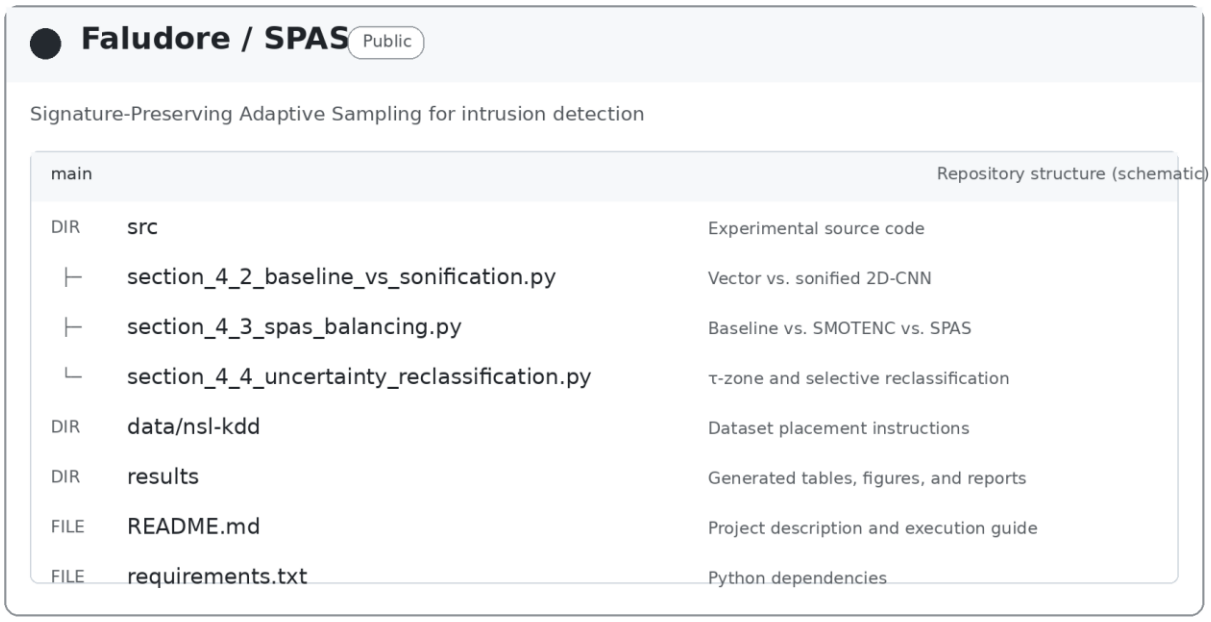

Рисунок Д.1 – Структура репозиторію GitHub із вихідним програмним кодом

Репозиторій об’єднує програмні компоненти, що реалізують запропоновані в дисертації модель і методи виявлення комп’ютерних атак. Програмний код забезпечує підготовку набору даних NSL-KDD, перетворення ознак мережних з’єднань у хвильовий сигнал, формування частотно-часового подання за допомогою короткочасного перетворення Фур’є, навчання двовимірної згорткової нейронної мережі, балансування навчальної вибірки методом SPAS, помилково-

орієнтоване підсилення ErrorBoost і селективне перевизначення рішень у зоні невизначеності прогнозу.

Структура репозиторію є такою:

– *src/section_4_2_baseline_vs_sonification.py* – порівняння базової векторної моделі та соніфікованої моделі з використанням PCM, STFT і 2D-CNN;

– *src/section_4_3_spas_balancing.py* – експериментальне порівняння початкової вибірки, SMOTENC і запропонованого методу сигнатурозбережного адаптивного балансування SPAS;

– *src/section_4_4_uncertainty_reclassification.py* – реалізація каскадної схеми SPAS, ErrorBoost, формування τ-зони невизначеності, навчання допоміжної моделі та селективного перевизначення рішень;

– *data/nsl-kdd/README.md* – інструкція щодо розміщення файлів KDDTrain+.txt і KDDTest+.txt;

– *results/* – каталог для збереження таблиць, рисунків, конфігурацій і підсумкових звітів експериментів;

– *README.md* – опис призначення репозиторію, структури програмного забезпечення, залежностей і порядку відтворення експериментів;

– *requirements.txt* – перелік бібліотек Python, необхідних для запуску програмного коду.

Для запуску програмного забезпечення рекомендовано використовувати Python 3.10 або 3.11 і встановити залежності командою pip install -r requirements.txt. Файли набору даних не включено до репозиторію; їх потрібно розмістити в каталозі data/nsl-kdd або вказати шлях до них за допомогою змінної середовища NSL_KDD_DIR. Результати виконання програм за замовчуванням зберігаються в каталозі results.