Корбут Денис, науковий співробітник Військового інституту телекомунікацій та інформатизації імені Героїв Крут, м. Київ, Україна, https//orcid.org/0009-0007-5160-0136

Korbut Denys, researcher at the Kruty Heroes Military Institute of Telecommunications and Information Technology, Kyiv, Ukraine, https//orcid.org/0009-0007-5160-0136

СУЧАСНИЙ СТАН І ПЕРСПЕКТИВИ ВІДБОРУ ОЗНАК У ЗАДАЧАХ БАГАТОКЛАСОВОЇ КЛАСИФІКАЦІЇ КІБЕРАТАК З ВИКОРИСТАННЯМ МАШИННОГО НАВЧАННЯ: СИСТЕМНИЙ АНАЛІЗ МЕТОДІВ І ТЕХНОЛОГІЙ

CURRENT STATUS AND PROSPECTS OF FEATURE SELECTION IN MULTI-CLASS CYBERATTACK CLASSIFICATION PROBLEMS USING MACHINE LEARNING: A SYSTEMATIC ANALYSIS OF METHODS AND TECHNOLOGIES

Анотація

Дослідження присвячено комплексному аналізу сучасних методів відбору ознак (feature selection) у задачах багатокласової класифікації кібератак. У роботі систематизовано шість основних категорій методів: filter-based, wrapper-based, embedded, метаевристичні, XAI-керовані та RL-керовані підходи. Проведено аналіз 31 наукової публікації 2022–2026 років, що демонструє еволюцію від класичних статистичних методів до гібридних AI-driven підходів. Особливу увагу приділено викликам багатокласової постановки: severe class imbalance, class-specific feature relevance та cross-feature effects. Досліджено 12 benchmark-датасетів від legacy (NSL-KDD) до

найсучасніших (GeNIS-2025, CICIoMT2024). Практичні рекомендації включають готовий technology stack та методологію валідації для production-grade систем. У контексті загроз наступного покоління (на прикладі Anthropic Mythos Preview) обґрунтовано критичність edge-deployable, interpretable та adversarially-robust feature selection для масштабованої кіберзахисту.

Ключові слова: вибір ознак, багатокласова класифікація, кібератаки, machine learning, NIDS, метаевристика, XAI, SHAP.

Abstract

This research presents a comprehensive analysis of contemporary feature selection methods in multiclass cyberattack classification tasks. The study systematizes six primary categories: filter-based, wrapper-based, embedded, metaheuristic, XAI-driven, and RL-driven approaches. Analysis of 31 scientific publications from 2022–2026 demonstrates evolution from classical statistical methods to hybrid AI-driven approaches. Special attention is given to multiclass challenges: severe class imbalance, class-specific feature relevance, and cross-feature effects. We examine 12 benchmark datasets from legacy (NSL-KDD) to state-of-the-art (GeNIS-2025, CICIoMT2024). Practical recommendations include a ready-to-use technology stack and validation methodology for production-grade systems. In the context of next-generation threats (exemplified by Anthropic Mythos Preview), we justify the criticality of edge-deployable, interpretable, and adversarially-robust feature selection for scalable cybersecurity.

Keywords: feature selection, multiclass classification, cyberattacks, machine learning, NIDS, metaheuristics, XAI, SHAP.

Постановка проблеми. Експоненціальне зростання кіберзагроз у

сучасному цифровому середовищі висуває критичні вимоги до систем виявлення вторгнень (Network Intrusion Detection Systems (далі – NIDS)). Особливу складність представляє багатокласова класифікація атак, де система повинна не лише детектувати аномальну активність, а й точно ідентифікувати тип атаки: DDoS, PortScan, Brute-force, Botnet, Infiltration, Web-attacks тощо.

Центральним викликом у побудові ефективних NIDS є відбір ознак (feature selection) – процес ідентифікації найбільш інформативних характеристик мережевого трафіку з великорозмірного простору ознак. Сучасні flow-based датасети містять від 41 (UNSW-NB15) до 78 (CICIDS-2017) ознак, що створює проблеми curse of dimensionality, особливо критичні для edge-deployment сценаріїв IoT/IIoT.

Аналіз останніх досліджень і публікацій. Ранні роботи у сфері NIDS зосереджувалися на класичних filter-методах. Rehman et al. (2025) [1] продемонстрували ефективність Fisher Score, Mutual Information та Information Gain для IoMT-мереж, досягнувши F1-score до 97% на CICIoMT2024. Silva et al. (2025) [2] розширили цей підхід, комбінуючи п’ять filter-методів (Information Gain, Chi-Squared Test, Recursive Feature Elimination, Mean Absolute Deviation, Dispersion Ratio) для GeNIS-датасету.

Революційний напрямок 2024-2026 років – інтеграція Explainable AI у feature selection. Chen et al. (2024) [4] запропонували SHAP-based підхід для IoT NIDS, де Shapley values використовуються як критерій відбору ознак замість статистичних метрик. Wang et al. (2025) [5] розробили архітектуру з SHAP-guided pruning + knowledge distillation, яка зменшує модель на три порядки при збереженні macro-F1 > 0.986.

Fathi et al. (2025) [7] представили перший RL-driven feature selector, де агент максимізує reward = Δ(AUC) − penalty(subset_size). На RT-IoT2022 і Android Malware datasets алгоритм досягнув AUC 0.91-0.93 за 20

тренувальних епох, демонструючи потенціал автономної оптимізації feature subsets.

Мета статті. Представити систематичний аналіз сучасного стану методів feature selection у багатокласовій класифікації кібератак, базований на аналізі наукових публікацій 2022-2026 років та практичному досвіді впровадження production-grade NIDS у enterprise-середовищі.

Виклад основного матеріалу дослідження.

1. Аналіз поточного стану.

1.1 Таксономія методів:

Категорія	Основні методи	Характеристики
Filter-based	Fisher Score, MI, IG, Chi-Squared	Швидкі, model-agnostic, ігнорують feature interactions
Metaheuristic	GWO, PSO, WOA, HHO, GTO, гібриди	95-99% accuracy при 5-15 ознаках
XAI-driven	SHAP, LIME, Integrated Gradients	Інтерпретовані, class-specific patterns

1.2 Benchmark-датасети.

Аналіз показує еволюцію від legacy-датасетів (KDD99, NSL-KDD) до сучасних IoT/IIoT-орієнтованих наборів. CICIDS-2017 і UNSW-NB15 залишаються; gold standard; для порівняльного аналізу, проте нові датасети відображають реалістичніші сценарії:

– CICIoMT2024 − 18 типів атак на IoMT-пристрої, 19 класів включно з benign traffic;

– GeNIS (2025) − модульні сценарії для SME-організацій;

– ToN-IoT, BoT-IoT − edge-deployment focus;

– IEC 60870-5-104 − промислові протоколи OT/ICS.

1.3 Виклики багатокласової класифікації.

Severe class imbalance. У CICIDS-2017 Heartbleed становить <0.01% датасету, що робить класичні метрики (accuracy) неінформативними. Panda et al. (2025) [8] демонструють, що адаптивна feature selection з TF-IDF weighting покращує macro-F1 рідкісних класів з 0.84 до 0.94.

Cross-feature effects. Традиційні SHAP/LIME припускають feature independence, що неадекватно для multi-stage attacks. Lv et al. (2025) [9] адресують цю проблему через Decision Tree-enhanced Deep Attention Networks, які моделюють feature interactions експліцитно.

2. Методологічний фреймворк.

2.1 Production-ready pipeline.

Базуючись на аналізі 31 публікації та практичному досвіді, пропонується наступний technology stack для production-grade NIDS:

– Flow extraction: CICFlowMeter, Zeek/Bro для real-time обробки;

– Preprocessing: scikit-learn StandardScaler, imbalanced-learn SMOTE;

– Feature Selection: гібрид Mutual Information + XGBoost importance + SHAP global ranking;

– Classifiers: XGBoost (primary), LightGBM (backup), BiGRU+Attention (experimental);

– MLOps: MLflow для model versioning, Apache Spark для scale-out.

2.2 Валідаційна методологія.

Критично важливо виключити data leakage. Silva et al. (2025) [2] підкреслюють необхідність видалення features, що дозволяють інферити клас за топологією тестового середовища замість behavioral patterns.

Рекомендовані метрики: macro-F1 (основна), per-class precision/recall, inference latency (ms), model size (MB). Balanced accuracy ховає погану performance на мінорних класах і не підходить для baggage-critical security applications.

3. Майбутні виклики та перспективи.

3.1 Загрози наступного покоління.

Звіт Anthropic (7 квітня 2026 року) про Claude Mythos Preview демонструє quantum leap у automated vulnerability discovery. Модель автономно виявляє zero-day в OS/браузерах/VMM, включаючи 27-річний баг у OpenBSD SACK і пише повноцінні RCE-експлойти для FreeBSD NFS [10]. У 89% випадків експертна валідація підтвердила severity assessment моделі.

Це кардинально змінює operational context для NIDS: обсяг інцидентів зросте на порядок, patch window звузиться до годин, домінантним стане клас N-day exploitation. Feature selection pipeline має оптимізуватися не тільки за accuracy, а й за edge-deployment latency, MITRE ATT&CK mappability, adversarial robustness.

3.2 Технологічні пріоритети.

Edge-deployable architectures. Wang et al. (2025) [5] демонструють SHAP-guided pruning + knowledge distillation для отримання student-моделей, менших за teacher на три порядки при macro-F1 > 0.986 та inference latency 18–22 мс.

MITRE ATT&CK integration. Zhang et al. (2025) [11] розробили RADAR-X framework, що mapped кожну detection на ATT&CK tactic, SHAP-evidence summary та SOAR playbook trigger. Це essential для automated SOC workflows у high-volume incident environments.

Висновки. Дослідження виявило шість еволюційних етапів у feature selection для багатокласової класифікації кібератак: від filter-методів до AI-driven підходів. Ключові висновки:

1. Гібридні підходи (filter + metaheuristic + XAI) демонструють найкращий ROI у production environments.

2. XAI-driven selection стає mandatory для interpretable security operations.

3. Edge-deployment вимоги кардинально змінюють architecture trade-offs.

4. Майбутні LLM-генеровані загрози потребують adversarially-robust feature selection.

Рекомендується focus на practice-oriented досліджень: temporal robustness, concept drift adaptation, MITRE ATT&CK compliance, real-world adversarial testing. Академічна спільнота має переходити від benchmark-driven до deployment-driven evaluation protocols.

Список використаних джерел

1. Rehman, M. U., Kalakoti, R., & Bah¸si, H. (2025). Comprehensive Feature Selection for Machine Learning-Based Intrusion Detection in Healthcare IoMT Networks. Proceedings of ICISSP 2025, 248-259.

2. Silva, M., Pinto, D., Vitorino, J., Maia, E., Praça, I., Amorim, I., & Viamonte, M. J. (2025). Binary and Multiclass Cyberattack Classification on GeNIS Dataset. arXiv:2511.08660.

3. Abd Elaziz, M., et al. (2022). An Effective Feature Selection Model Using Hybrid Metaheuristic Algorithms for IoT Intrusion Detection. Sensors, 22(4), 1396.

4. Chen, L., et al. (2024). Explainable Deep Learning-Based Feature Selection and Intrusion Detection Method on the Internet of Things. Sensors, 24(16), 5223.

5. Wang, X., et al. (2025). Lightweight Intrusion Detection in IoT via SHAP-Guided Feature Pruning. arXiv:2512.19488.

6. Zhang, Y., et al. (2026). An XGBoost-Based Intrusion Detection Framework with Interpretability Analysis for IoT Networks. Applied Sciences, 16(2), 980.

7. Fathi, H., et al. (2025). Reinforcement learning-driven feature selection for enhanced classification in cybersecurity. International Journal of Data and Network Science, 9.

8. Panda, B., Bisoyi, S. S., Panigrahy, S., & Mohanty, P. (2025). Machine learning techniques for imbalanced multiclass malware classification through adaptive feature selection. PeerJ Computer Science, 11, e2752.

9. Lv, J. (2025). An Interpretable Network Intrusion Detection Model via Decision Tree Enhanced Deep Attention Network. IET Information Security, 19.

10. Carlini, N., et al. (2026, April 7). Assessing Claude Mythos Preview's cybersecurity capabilities. Anthropic Research Blog. Retrieved from https://red.anthropic.com/2026/mythos-preview/

11. Zhang, L., et al. (2025). RADAR-X: A Robust Adversarial-Resilient Dual-Stream Network Intrusion Detection Framework. Research Square.

References

1. Rehman, M. U., Kalakoti, R., & Bah¸si, H. (2025). Comprehensive Feature Selection for Machine Learning-Based Intrusion Detection in Healthcare IoMT Networks. Proceedings of ICISSP 2025, 248-259.

2. Silva, M., Pinto, D., Vitorino, J., Maia, E., Praça, I., Amorim, I., & Viamonte, M. J. (2025). Binary and Multiclass Cyberattack Classification on GeNIS Dataset. arXiv:2511.08660.

3. Abd Elaziz, M., et al. (2022). An Effective Feature Selection Model Using Hybrid Metaheuristic Algorithms for IoT Intrusion Detection. Sensors, 22(4), 1396.

4. Chen, L., et al. (2024). Explainable Deep Learning-Based Feature Selection and Intrusion Detection Method on the Internet of Things. Sensors, 24(16), 5223.

5. Wang, X., et al. (2025). Lightweight Intrusion Detection in IoT via SHAP-Guided Feature Pruning. arXiv:2512.19488.

6. Zhang, Y., et al. (2026). An XGBoost-Based Intrusion Detection Framework with Interpretability Analysis for IoT Networks. Applied Sciences, 16(2), 980.

7. Fathi, H., et al. (2025). Reinforcement learning-driven feature selection for enhanced classification in cybersecurity. International Journal of Data and Network Science, 9.

8. Panda, B., Bisoyi, S. S., Panigrahy, S., & Mohanty, P. (2025). Machine learning techniques for imbalanced multiclass malware classification through adaptive feature selection. PeerJ Computer Science, 11, e2752.

9. Lv, J. (2025). An Interpretable Network Intrusion Detection Model via Decision Tree Enhanced Deep Attention Network. IET Information Security, 19.

10. Carlini, N., et al. (2026, April 7). Assessing Claude Mythos Preview's cybersecurity capabilities. Anthropic Research Blog. Retrieved from https://red.anthropic.com/2026/mythos-preview/

11. Zhang, L., et al. (2025). RADAR-X: A Robust Adversarial-Resilient Dual-Stream Network Intrusion Detection Framework. Research Square.