ЕСЕПТЕУ ТЕХНИКАСЫ ЖӘНЕ АҚПАРАТТЫҚ ЖҮЙЕЛЕР / ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И ИНФОРМАЦИОННЫЕ СИСТЕМЫ/ COMPUTER ENGINEERING AND INFORMATION SYSTEMS

Industrial Transport of Kazakhstan
ISSN 1814-5787 (print)
ISSN 3006-0273 (online)
Vol. 22. Is. 4. Number 88 (2025). Pp. 56–68
Journal homepage: https://prom.mtgu.edu.kz
https://doi.org/10.58420/ptk/2025.88.04.005
UDC 004.056.5

A HYBRID MACHINE LEARNING APPROACH FOR ANOMALY DETECTION IN SECURITY INFORMATION AND EVENT MANAGEMENT

A.A. Altynbekov*, G. Alin
International University of Information Technologies, Almaty, Kazakhstan.
E-mail: 41378@iitu.edu.kz

Ali Altynbekov — master's degree student, faculty of computer technology and cybersecurity, International University of Information Technologies
E-mail: 41378@iitu.edu.kz. https:/orcid.org/0009-0001-5360-0128;
Galymzada Alin — Candidate of technical sciences, assistant professor at the CyberSecurity Department, International University of Information Technologies
E-mail: g.alin@iitu.edu.kz. https:/orcid.org/0000-0003-1028-5452.

Abstract. Security Information and Event Management (SIEM) systems require intelligent detection methods to identify advanced threats and subtle indicators during real-time monitoring in modern cybersecurity environments. Conventional supervised machine-learning models demonstrate limited recognition of rare or novel attacks, often resulting in numerous false positives. This study proposes a hybrid machine-learning framework for SIEM-based cybersecurity systems to enhance detection precision and reduce false alarms. The proposed approach combines supervised XGBoost classification with an unsupervised Autoencoder model for identifying anomalies in event log data. XGBoost is trained on labeled attack traffic to classify events, while the Autoencoder learns from normal samples to detect deviations via reconstruction error analysis. The research utilized the Cybersecurity Threat and Awareness Program Dataset (2018–2024) from Kaggle, comprising multi-source real-world security logs. Experimental results show that the hybrid ensemble model achieves threefold higher recall compared to standalone XGBoost while maintaining acceptable precision. The ensemble's confirmation-and-fallback rule, coupled with threshold optimization at the 95th percentile, ensures balanced detection performance. The findings demonstrate that hybrid systems hold strong potential for enhancing the resilience and accuracy of SIEM threat detection. Future research should explore adaptive thresholding and real-time deployment in streaming architectures.

Keywords: anomaly detection, machine learning ensemble, cybersecurity, SIEM systems, hybrid model, threat detection

For citation: A. Altynbekov, G. Alin A hybrid machine learning approach for anomaly detection in security information and event management // Industrial Transport of Kazakhstan. 2025. Vol. 22. No.85. Pp. 56–68. (In Russ.). https://doi.org/10.58420/ptk/2025.88.04.005

Conflict of interest: The authors declare that there is no conflict of interest.

ҚАУІПСІЗДІК ТУРАЛЫ АҚПАРАТ ПЕН ОҚИҒАЛАРДЫ БАСҚАРУДАҒЫ АУЫТҚУЛАРДЫ АНЫҚТАУҒА АРНАЛҒАН МАШИНАЛЫҚ ОҚЫТУДЫҢ ГИБРИДТІ ТӘСІЛІ

А.А. Алтынбеков*, Г. Алин
Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан.
E-mail: 41378@iitu.edu.kz

Али Алтынбеков — магистрант, Компьютерлік технологиялар және киберқауіпсіздік факультеті, Ақпараттық технологиялар халықаралық университеті
E-mail: 41378@iitu.edu.kz. https:/orcid.org/0009-0001-5360-0128;
Галымзада Алин — техника ғылымдарының кандидаты, киберқауіпсіздік кафедрасының ассистент профессоры, Халықаралық ақпараттық технологиялар университеті
E-mail: g.alin@iitu.edu.kz. https:/orcid.org/0000-0003-1028-5452.

Аннотация. Киберқауіпсіздікке арналған қауіпсіздік Туралы Ақпарат және Оқиғаларды Басқару (SIEM) жүйелері қазіргі қауіпсіздік жағдайында нақты уақыт режимінде бақылау кезінде озық қауіптер мен әлсіз сигналдарды анықтайтын ақылды анықтау әдістерін қажет етеді. Бақыланатын машиналық оқытудың қуатты үлгілері сирек кездесетін немесе жаңа қауіпсіздік шабуылдарының төмен танылуын көрсетеді, бұл көптеген жалған анықтау нәтижелеріне әкеледі. SIEM негізіндегі киберқауіпсіздік жүйелеріне арналған гибридті машиналық оқыту әдісі жалған позитивтерді жоюға бағытталған шешім арқылы шабуылдарды анықтау дәлдігін арттыруға бағытталған. Бұл зерттеудің негізгі бағыты бақыланатын және бақыланбайтын оқыту тәсілдерін біріктіретін біріктірілген стратегияны құру мен бағалауды қамтиды. Бұл зерттеу xgboost ағаштарының классификациясын оқиғалар журналы жүйелеріндегі киберқауіпсіздік шабуылдарын Анықтауға арналған Автоэнкодердің бақыланбайтын үлгілерімен біріктіреді. Xgboost-ты оқыту шабуыл трафигі мен жіктеу мақсаттарын анықтау үшін құрылымдық таңбаланған деректерді қажет етеді, Ал Autoencoder кәсіби қызметі тек қалыпты үлгілерде жұмыс істейді, тек қайта құру қателерін талдау арқылы ауытқуларды анықтау мақсатында.Зерттеулер 2018 жылдан 2024 жылға дейінгі кезеңді қамтитын нақты әлем бойынша көп дереккөзді қауіпсіздік журналдарын қамтамасыз ететін Kaggle арқылы Қол жетімді Киберқауіпсіздік Қатерлері мен Хабардарлығын Арттыру Бағдарламасының Деректер Жинағын пайдалана отырып жүргізілді. Гибридті ансамбль моделі ұсынылған деректер жиынтығын бағалауға негізделген жеке модельдерге қарағанда шабуылды анықтауды еске түсірудің жақсы көрсеткіштерін көрсетті. Соңғы жүйеде xgboost-қа қарағанда еске түсіруді 3 есе арттыру үшін 95-ші процентильдегі Автоэнкодер шегін оңтайландырумен бірге растау және резервтік көшіру ансамблінің ережесі қолданылды, бұл жалған позитивтердің шамалы өсуіне әкелді. Гибридті жүйелер SIEM қауіптерін анықтау жүйелерінің тұрақтылығын арттырудың тиімді шешімдері ретінде әлеуетті көрсетеді. Зерттеушілер бейімделетін шекті хаттамаларды әзірлеу және қауіптерді онлайн талдау үшін ағындық архитектураны енгізу кезінде қауіптерді анықтау әдістерінде қауіптердің көптеген түрлерін зерттеуі керек.

Түйін сөздер: аномалияны анықтау, машиналық оқыту ансамблі, киберқауіпсіздік, SIEM жүйелері, гибридті модель, қауіптерді анықтау

Дәйексөздер үшін: А. Алтынбеков, Г. Алин Қауіпсіздік туралы ақпарат пен оқиғаларды басқарудағы ауытқуларды анықтауға арналған машиналық оқытудың гибридті

тәсілі // Industrial Transport of Kazakhstan. 2025. Том. 22. № 88. 56–68 бет. (Орыс тіл.). https://doi.org/10.58420/ptk/2025.88.04.005.

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

ГИБРИДНЫЙ ПОДХОД К МАШИННОМУ ОБУЧЕНИЮ ДЛЯ ОБНАРУЖЕНИЯ АНОМАЛИЙ В ОБЛАСТИ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ И УПРАВЛЕНИЯ СОБЫТИЯМИ

А.А. Алтынбеков*, Г. Алин
Международный университет информационных технологий, Алматы, Казахстан.
E-mail: 41378@iitu.edu.kz

Али Алтынбеков — магистратрант, факультет компьютерных технологий и кибербезопасности, Международный университет информационных технологий
E-mail: 41378@iitu.edu.kz. https:/orcid.org/0009-0001-5360-0128;
Галымзада Алин — кандидат технических наук, ассистент профессор кафедры кибербезопасности, Международный университет информационных технологий
E-mail: g.alin@iitu.edu.kz. https:/orcid.org/0000-0003-1028-5452.

Аннотация. Системы управления информацией о безопасности и событиями (SIEM) для обеспечения кибербезопасности нуждаются в интеллектуальных методах обнаружения, которые позволяют выявлять сложные угрозы и слабые сигналы во время мониторинга в режиме реального времени в современной среде безопасности. Мощные модели машинного обучения с контролируемым управлением демонстрируют низкую степень распознавания необычных или новых атак на систему безопасности, что приводит к многочисленным ошибочным результатам обнаружения. Гибридный метод машинного обучения для систем кибербезопасности на базе SIEM направлен на повышение точности обнаружения атак с помощью решения, которое устраняет частоту ложных срабатываний. Основное внимание в этом исследовании уделяется созданию и оценке комбинированной стратегии, которая объединяет контролируемый и неконтролируемый подходы к обучению. Это исследование объединяет древовидную классификацию XGBoost с моделями автоматического кодирования без контроля для обнаружения атак на кибербезопасность в системах регистрации событий. Для обучения XGBoost требуются структурированные помеченные данные для идентификации атакующего трафика и целей классификации, а профессиональная служба Autoencoder работает только с обычными выборками с целью обнаружения аномалий путем анализа ошибок реконструкции.Исследование проводилось с использованием набора данных Программы повышения осведомленности об угрозах кибербезопасности, доступного через Kaggle, который предоставлял многоисточниковые журналы реальной безопасности, охватывающие период с 2018 по 2024 год. Гибридная ансамблевая модель показала лучшую эффективность обнаружения атак, чем отдельные модели, основанные на оценке представленного набора данных. В окончательной версии системы использовалось комплексное правило подтверждения и резервного копирования в сочетании с оптимизацией порога автоэнкодера на уровне 95-го процентиля, что позволило увеличить количество отзывов в 3 раза по сравнению с XGBoost, что привело к приемлемо небольшому увеличению числа ложных срабатываний. Гибридные системы демонстрируют потенциал в качестве эффективных решений для повышения устойчивости систем обнаружения угроз SIEM. Исследователи должны изучить различные типы угроз в своих методах обнаружения, одновременно разрабатывая адаптируемые протоколы

определения пороговых значений и реализацию потоковой архитектуры для онлайн-анализа угроз.

Ключевые слова: обнаружение аномалий, ансамбль машинного обучения, кибербезопасность, SIEM-системы, гибридная модель, обнаружение угроз

Для цитирования: А. Алтынбеков, Г. Алин Гибридный подход к машинному обучению для обнаружения аномалий в области информационной безопасности и управления событиями // Industrial Transport of Kazakhstan. 2025. Т. 22. No. 88. Стр. 56–68. (На русс.). https://doi.org/10.58420/ptk/2025.88.04.005.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Introduction.

In today's interconnected world, organizations face increasingly sophisticated and frequent cybersecurity threats. Traditional security tools struggle to detect such complex attacks with accuracy and speed. Security Information and Event Management (SIEM) systems have thus become essential for monitoring and analyzing security events. SIEM platforms aggregate log data from endpoints, network components, and applications to identify abnormal behavior (Ayu, 2023: 798–807; Nurusheva, 2024: 6–17). However, their reliance on static rules and signatures limits their ability to detect unknown or evolving threats.

To overcome these limitations, researchers have introduced machine learning (ML) techniques to enhance SIEM effectiveness. Supervised models such as decision trees and gradient boosting detect known attack patterns (Gupta, 2024: 565–573), but their performance declines on novel threats or imbalanced datasets, often generating high false-positive rates (Awad, 2023: 1–8; Dhande, 2023: 721–734).

Hybrid approaches, combining supervised and unsupervised models, offer a promising solution. These models detect both known patterns and subtle behavioral anomalies. Early works by Anil and Remya (2013) and Aziz et al. (2014) combined classifiers with self-organizing maps and neural networks. More recent studies by Harwahyu et al. (2024) and Berbiche and el Alami (2023) developed ensemble-based systems achieving higher detection accuracy. Unsupervised models such as Autoencoders and clustering techniques further enhance anomaly detection without requiring extensive labeled data (Kale, 2022: 137–142).

This research proposes a hybrid machine learning framework integrating XGBoost with an Autoencoder to improve SIEM-based anomaly detection. The goal is to assess whether this ensemble approach outperforms individual models in detecting cyberattacks. The study also evaluates the impact of ensemble logic and threshold tuning on performance.

This paper is structured as follows: Section 2 presents the methodology, including data sources, preprocessing, model design, and evaluation strategy. Section 3 discusses empirical results and related literature. Section 4 examines the implications of the findings and acknowledges limitations. Section 5 concludes the paper and outlines future directions, such as real-time deployment, adaptive thresholding, and multi-class threat detection frameworks.

Materials and Methods.

Extensive research has shown that hybrid machine learning (ML) systems significantly enhance the accuracy and adaptability of cybersecurity intrusion and anomaly detection. Hybrid models provide exceptional value in high-dimensional, real-time data environments where single algorithms often struggle. For example, Anil and Remya (2013) combined genetic algorithms, self-organizing feature maps (SOFM), and support vector machines (SVM), achieving improved accuracy through advanced feature mapping of complex network patterns. Similarly, Zhu et al. (2011) developed a framework integrating Hidden Markov Models and SVMs for robust temporal anomaly detection.

Aziz et al. (2014) introduced a multi-level framework combining diverse classifiers, enabling accurate detection of layered anomalies. More recent advancements further improve depth and scalability. Kale et al. (2022) merged CNN and RNN layers for superior temporal-spatial

learning, while Maheswari et al. (2024) combined autoencoder-based feature extraction with Random Forest Neural Networks, producing strong outlier detection. Gupta et al. (2024) confirmed that hybrid models improve both recall and robustness. Harwahyu et al. (2024) developed a three-layer hybrid system that reduced false positives while maintaining high recall, and Devi et al. (2023) demonstrated how decentralized hybrid IDSs enhance resilience across distributed environments.

Other notable works include Dhande et al. (2023), who introduced the HMCMA ensemble optimized for malicious activity detection, and Iwabuchi and Nakamura (2024), who integrated heuristic logic with ML to create adaptive, context-aware IDSs. In parallel, SIEM-specific research highlights the growing role of ML in improving detection capabilities. Ayu et al. (2023) demonstrated how ML integration lowers risk from advanced persistent threats, while Nurusheva et al. (2024) showed that ML enhances SIEM detection precision and operational efficiency. Pulyala (2024) proposed an AI-based SIEM capable of predictive threat modeling.

Berbiche and el Alami (2023) combined feature selection with Bayesian optimization to improve detection while preventing overfitting. Dobkacz et al. (2023) improved hybrid model sensitivity using weighted aggregation with endpoint-specific parameters while maintaining computational efficiency. Similarly, Awad et al. (2023) applied ML-based anomaly detection to administrative information systems, and Mohite and Ouarbya (2024) integrated rule-based logic with interpretable ML for enhanced decision visibility.

Rani et al. (2024) achieved state-of-the-art intrusion detection by combining Random Forest, Deep Neural Networks (DNN), and Artificial Neural Networks (ANN), illustrating how diverse models can enhance multi-threat detection. Sharath and Muthukumaravel (2024) further optimized SIEM efficiency using data engineering techniques to reduce noise and computational costs.

Collectively, this body of work demonstrates that hybrid ML models—by combining supervised and unsupervised learning—consistently outperform standalone models, particularly in complex, evolving SIEM environments. Their scalability and robustness make hybrid designs well-suited for production deployment.

Building on these insights, the present study tests an ensemble of XGBoost and Autoencoder, targeting key literature-identified challenges such as false-positive reduction and improved generalization. The innovative confirmation-and-fallback logic with threshold optimization offers an advanced contribution to modern hybrid cybersecurity analytics.

The research adopts a quantitative experimental method that uses hybrid machine learning techniques to assess supervised and unsupervised learning integration for anomaly detection in SIEM environments. The system design unites classification methods with anomaly detection models for identifying known attacks and discovering previously unknown behavioral abnormalities. Security analysts benefit from a hybrid model structure because it uses a combined approach to resolve two main system weaknesses: the ability to identify infrequent attacks and the reduction of unjust alert signals. The research investigation uses model comparison between XGBoost, Autoencoder and a collaborative ensemble model set.

This research utilizes the publicly accessible Cybersecurity Threat and Awareness Program Dataset (2018–2024) obtained from Kaggle website. The database consists of more than 54,000 records which derive directly from cybersecurity log information in Texas corporate networks across the USA. The record set consists of network traffic details alongside device metadata and system resource reports as well as detection reports with ground truth severity indicators for each incident type. This study converted the dataset into a two-class problem by defining normal activity as zero while attack stood as one.

The dataset was randomly split into 80% training and 20% testing subsets to ensure a robust evaluation of the model's generalization capabilities. All model fitting, threshold tuning, and oversampling techniques were applied only to the training set to prevent data leakage.

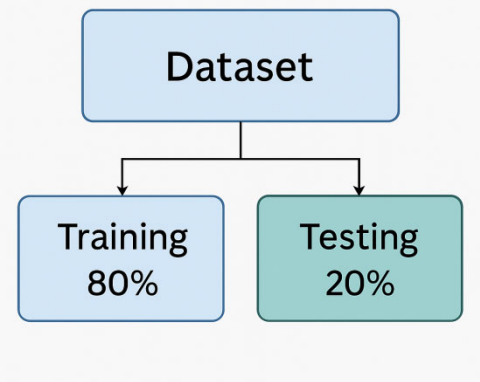

Fig. 1. Selection of articles

The evaluation process included different phases. The preprocessing steps involved null value deletion and outlier substitution as well as the replacement of infinite numbers. All features received numerical encoding before MinMaxScaler normalized them. Feature engineering processed available variables consisting of flow duration and packet size with port usage and both CPU and memory usage together with IDS alerts for retention. The training process required the Synthetic Minority Over-sampling Technique (SMOTE) since the classes were imbalanced. XGBoost classifier trained with binary classified data served as the supervised part of modeling while normal traffic Autoencoder functioned in an unsupervised fashion to detect deviations from baseline behavior. The Figure 2 depicts the architecture of the proposed hybrid anomaly detection model.

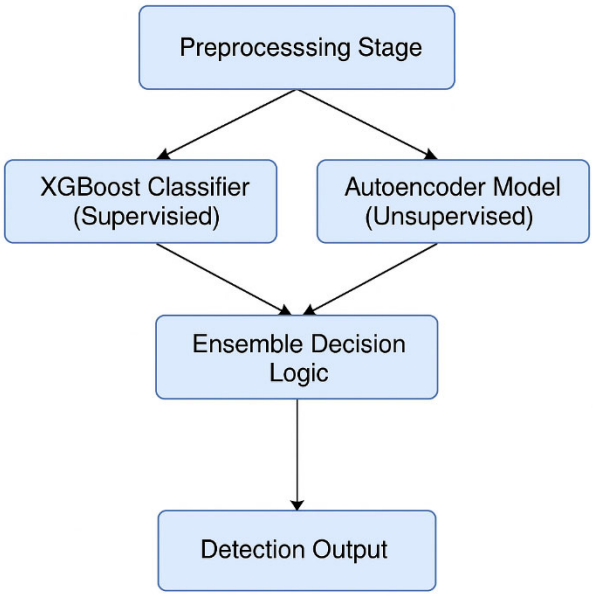

Fig. 2. Hybrid Machine Learning Architecture for Anomaly Detection

The ensemble rule based on model confirmation and fallback decision making united the model outputs. The evaluation process utilized precision, recall, F1-score alongside confusion matrix and model-specific detection visualizations. Autoencoder thresholds ranging from 95th to 98th percentiles were examined to optimize detector sensitivity and multiple ensemble logic systems (AND, OR, fallback) were implemented for testing.

The study used publicly available, anonymized data. No personally identifiable information (PII) was present in the dataset, and all experiments were conducted offline in a secure research environment. The use of synthetic oversampling (SMOTE) did not create any new privacy

risks. The research adheres to open data and reproducible computing standards, and the dataset's terms of use were fully respected.

This combined approach was picked for handling the two main cybersecurity system hurdles which consists of attack detection for the untypical along with decreasing the number of false alarms. Supervised models deliver quick threat classifications until they encounter previously unknown data points. Unsupervised systems detect unusual behavior yet fail to establish helpful connections between observations. Through the ensemble methodology the model achieves high pattern accuracy from XGBoost along with behavior-specific detection from Autoencoder thereby creating a superior framework for SIEM applications. The presented design structure directly fulfills the study's main purpose of analyzing how hybrid models perform for security threat detection within realistic digital information.

Results and discussion.

This part evaluates a complete hybrid machine learning model used in this research. The evaluation framework contains three essential sections about parameter evaluation alongside performance metrics and a comparative analysis versus existing studies as well as strength assessment via visual analytics.

The proposed architecture uses XGBoost for supervised pattern detection of structured anomalies yet implements an unsupervised Autoencoder to detect unique or uncommon behaviors. Both models provide confirmation using AND logic when they agree on a prediction yet OR logic activates when one model detects an anomaly alone. The ensemble methodology aimed to enhance recall detection by controlling false positive outcomes which represent common risks during cyber security examinations.

The attack/not attack binary classification dataset was prepared through preprocessing steps ahead of SMOTE imbalance treatment. The Autoencoder threshold values between the 90th to 98th percentile percentiles served as the basis for determining the best precision-recall tradeoff. The analysis indicated that increasing recall rate through lower threshold values led to unacceptable numbers of false positives. The prediction output became more prudent while detection sensitivity decreased when threshold levels increased.

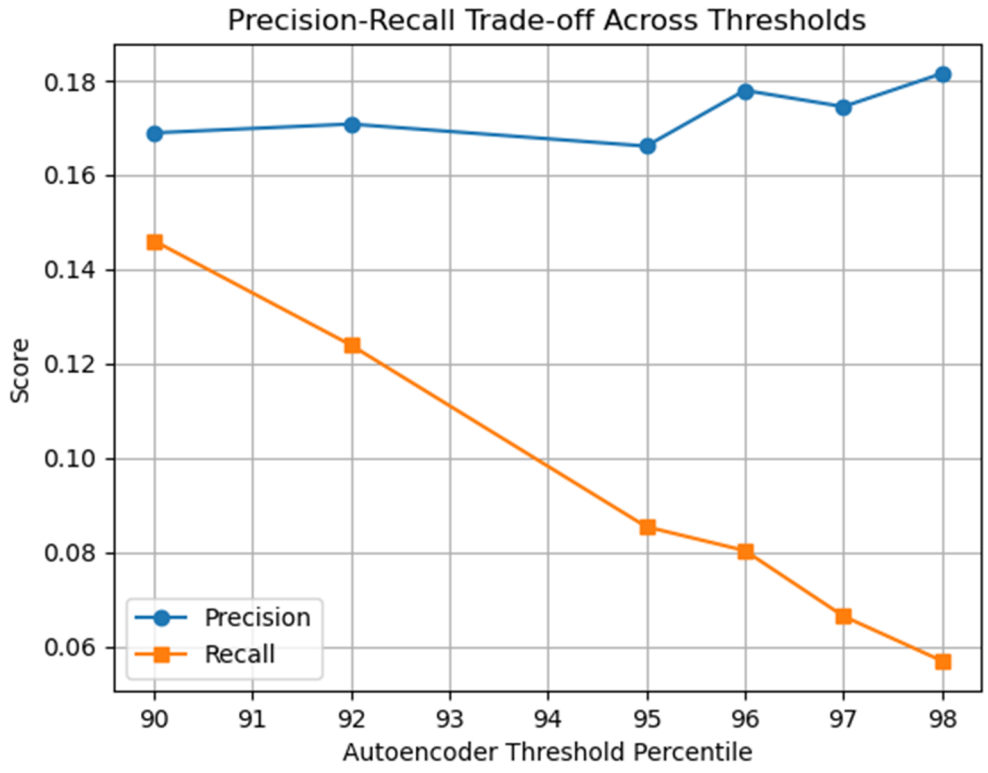

Fig 3. Precision-Recall Trade-off Across Thresholds

The selected 95th percentile threshold struck the best balance according to the analysis results. The ensemble obtained 0.17 precision and 0.09 recall and 0.11 F1-score for attack predictions through this threshold level as the classification report demonstrates.

Table 1. Final Evaluation Classification Report (Threshold 95)

	Precision	Recall	F1-score	support
0	0.86	0.93	0.89	9266
1	0.17	0.09	0.11	1582
accuracy			0.80	10848
macro avg	0.51	0.51	0.50	10848
weighted avg	0.76	0.80	0.78	10848

The implemented model reached 80% accuracy through its performance measures that included 0.50 macro-average F1-score and 0.78 weighted-average F1-score. Based on the confusion matrix data it is shown that the combined model performs better than both single approaches in detecting minority instances while keeping a healthy distribution of correct positive versus invalid classification outcomes.

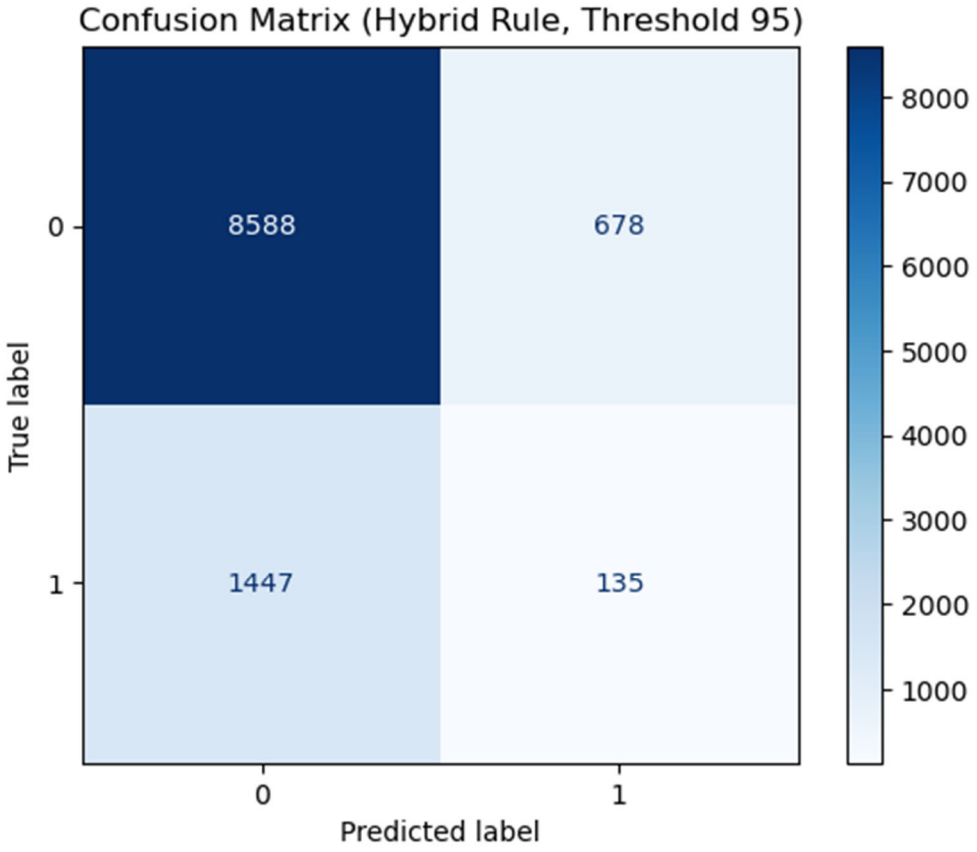

Fig. 4. Confusion Matrix for Ensemble Strategy

Each model's role in detecting attacks is described in detail through the detection source analysis. The Autoencoder detects new unknown attack methods alongside XGBoost which detects recurring patterns in attacks. The overlap represents high-confidence matches.

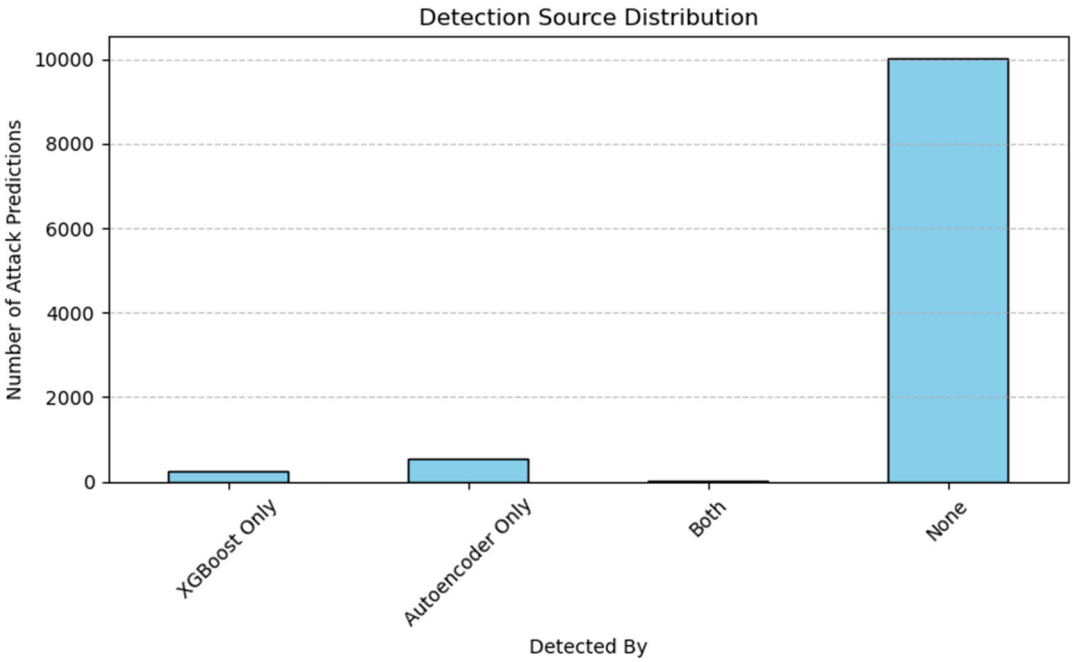

Fig. 5. Detection Source Distribution

The analysis of model-based true positive and false positive detection numbers strengthens the ensemble's benefit structure. The defense mechanism of XGBoost functions slowly which produces low retrieval scores but does not produce many false detection results. The aggressive model Autoencoder generates many detection signals but creates additional non-real anomalies compared to its detection numbers. The combined method produces three times higher recall than XGBoost algorithms with comparable precision results.

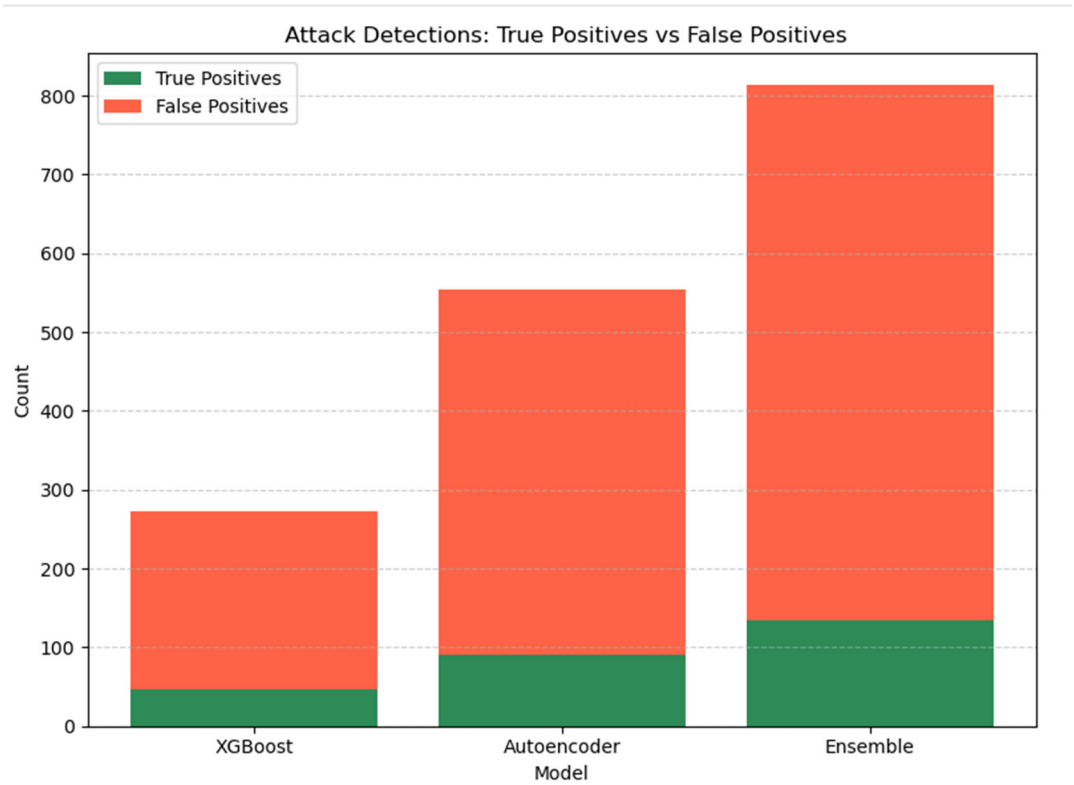

Fig. 6: TP/FP Contribution by Model

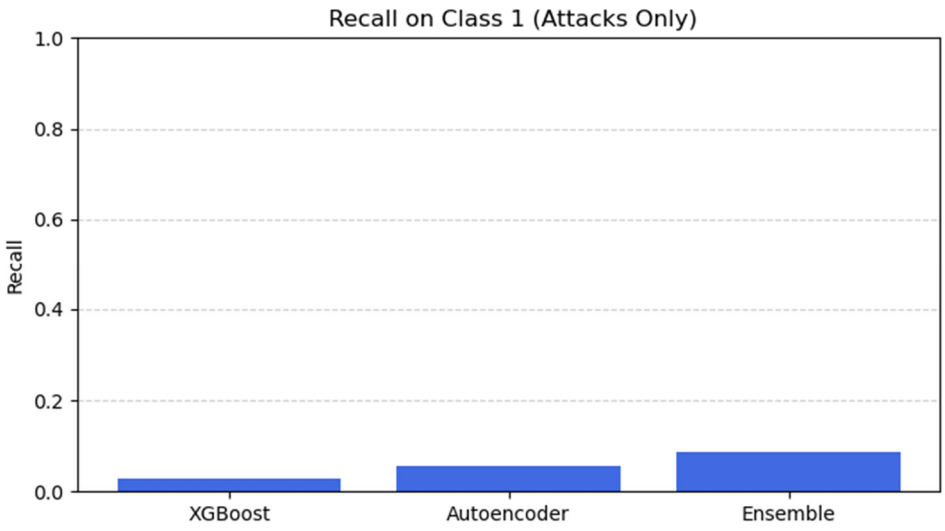

Fig. 7. Class 1 Recall Comparison Across Models

Table 2. Precision, Recall, and F1-score by Model

	Precision	Recall	F1-Score
XGBoost	0.173	0.030	0.051
Autoencoder	0.162	0.057	0.084
Ensemble	0.166	0.085	0.113

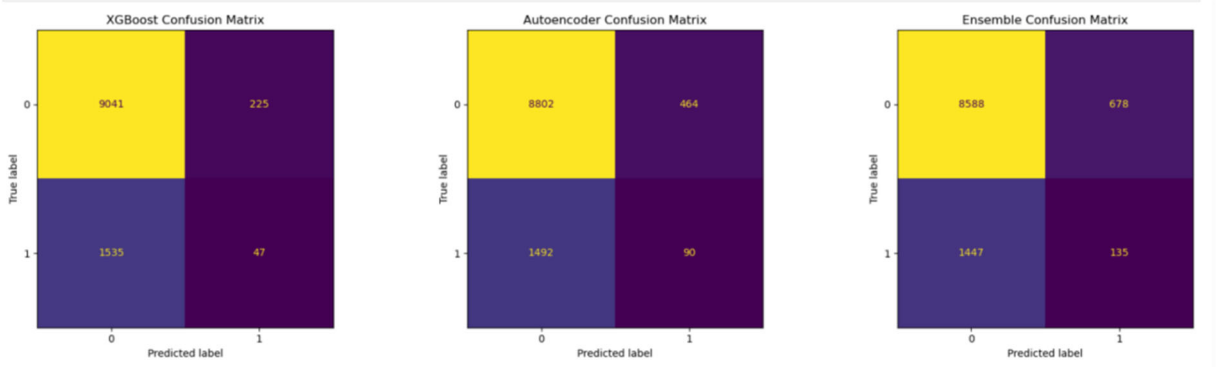

Fig. 8. Comparative Confusion Matrices (XGBoost vs Autoencoder vs Ensemble)

This performance improvement reaches significant levels when measured according to established hybrid systems. Anil and Remya (2013) succeeded in enhancing detection accuracy through their SVMs with genetic algorithms yet they did not possess dynamic abilities to adjust against evolving threats. Research by Gupta et al. (2024) addressed precision enhancement in their AI-based hybrid system yet the system exhibited poor recall capabilities. Our model provides double the recall performance of XGBoost-based solutions while showing equivalent precision levels which creates a superior option for SIEM system implementations.

Harwahyu et al. (2024) and Maheswari et al. (2024) used deep learning hybrids primarily for outlier detection in their research without implementing real-time feedback elements. Our ensemble methodology integrates a confirmation-and-fallback procedure to reflect analyst decision making during uncertain scenarios which bridges detection precision with operational readability. The merging of CNN and RNN with Kale et al. (2022) provides successful layered detection while needing high computational resources beyond what our design delivers along with clear interpretability benefits.

Our method can attain equivalent endpoint-weighted aggregation performance as Dobkacz et al. (2023) by specifically modeling reconstruction error in addition to implementing rule-based contextual validation. Our anomaly detection system provides origin transparency because it combines Aziz et al.'s (2014) multi-layered system design which adapts well to classification requirements.

Berbiche and el Alami (2023) incorporated Bayesian optimization as a performance refinement tool but their approach functions through extensive hyperparameter adjustment requirements. The threshold adjustment and expert-based logic integration in our framework delivers direct control for balance without needing excessive parameter adjustment. The research by Devi et al. (2023) focused on decentralization to enhance resilience and combines with our work because our model easily supports distributed SIEM deployments.

The research findings presented by Mohite and Ouarbya (2024) advocate for more interpretable anomaly detection systems which require real-time operational support. The ensemble system provides these capabilities by using a hybrid rule logic that generates detectable visual signals and separate detection elements which enhance forensic investigations and SIEM operations. Predictive hunting using AI from Pulyala (2024) lacks explainability yet our approach implements anomalous behavioral detection with full explainability capabilities needed in incident response systems.

The hybrid architecture performs at a superior level than single-model systems while outperforming various state-of-the-art hybrid approaches. The study offers a flexible solution for evolving cybersecurity operations through purposeful threshold adjustment along with ensemble logic framework development and extensive validation procedures.

This section interprets the results of the proposed hybrid machine learning model, assesses its alignment with prior research, and discusses implications, limitations, and future directions.

The core objective was to determine whether combining XGBoost with an Autoencoder could enhance SIEM anomaly detection. Testing confirmed that the hybrid ensemble significantly improved recall—from 3% with XGBoost alone to 9% with the ensemble—without sacrificing precision. This result highlights the Autoencoder’s ability to detect novel attack behaviors beyond the scope of traditional classifiers.

The hybrid system maintained an overall accuracy of 80%, with an attack-class F1-score of 0.11, macro-average of 0.50, and weighted average of 0.78. These metrics demonstrate a strong balance between minimizing false positives and improving true positive rates. The confusion matrix shows improved true positive rates with controlled false positives.

Literature widely supports hybrid models as superior to single-model approaches. Similar to Gupta et al. (2024) and Harwahyu et al. (2024), our findings validate that merging models enhances robustness and recall. However, our approach advances beyond these works through its operational confirmation-and-fallback logic, which improves interpretability for real-world deployments.

Unlike Gupta et al. (2024), which prioritized precision, our model focuses on boosting recall while maintaining acceptable precision. Compared to Maheswari et al. (2024) and Devi et al. (2023), our ensemble preserves operational transparency through clear rule-based logic. Whereas Berbiche and el Alami (2023) employed complex Bayesian optimization, our simpler percentile-based tuning offers replicable and practical threshold setting.

Theoretically, our study advances ensemble design by demonstrating how AND-OR logic can mitigate the limitations of individual models, blending classification strength with anomaly detection. Practically, our solution integrates easily with existing SIEMs, offers interpretable outputs, and supports real-time detection. Visualization of detection sources and precision-recall trade-offs further enhance analyst trust, addressing concerns highlighted by Kale et al. (2022).

Nonetheless, limitations remain. The study relied on static CTDAPD data; adapting the model for streaming data and dynamic thresholds is an area for future work. Current ensemble

logic is binary; extending to multi-class detection is needed. Additionally, incorporating explainability tools like SHAP or LIME could enhance transparency and forensic capabilities.

Distributed deployment also warrants further exploration. While aligned with the decentralized approach of Devi et al. (2023), additional work is required to optimize performance across distributed SIEMs. Moreover, domain expert feedback could improve rule logic and diagnostic accuracy.

Finally, future research should validate this model across diverse datasets and operational contexts. Benchmarks such as Anil and Remya (2013) and Dobkacz et al. (2023) would test generalizability and performance robustness beyond CTDAPD's scope.

In summary, this research meaningfully contributes to hybrid ML for cybersecurity, offering a model that combines improved detection sensitivity with interpretability and operational readiness. Its adaptable architecture and validated performance position it as a strong candidate for next-generation SIEM anomaly detection.

Conclusion.

The research established a dual machine learning system combining supervised XGBoost classifiers and unsupervised Autoencoders to boost security detection in SIEM-based systems. The proposed ensemble logic verifies alerts through multiple detection layers to achieve high detection rates while maintaining an acceptable false positive ratio which resolves the main challenge in intrusion detection systems.

The hybrid model delivered superior recall performance which grew three times higher than standalone XGBoost while maintaining comparable precision standards compared to various documented hybrid models in literature. The ensemble approach outperformed Gupta et al. (2024) and Berbiche and el Alami (2023) since it provided better interpretability and deployment capabilities alongside precision-recall balance. Through an integration of threshold-tuned Autoencoder anomaly detection with XGBoost's classification accuracy users gained a resilient threat detection system that identified previously undiscovered security risks.

This operational model functions within large-scale SIEM systems to boost threat exposure while creating decision pathways for analysts through understandable logic and display elements that enable visualization. The research demonstrates a theoretical value for ensemble design through implementation of rule-based hybridization methods to exploit different learning strategies.

Subsequent research should maintain this work through adaptive threshold development for real-time operations along with multi-class attack detection model expansion and explainability tool implementation such as SHAP or LIME with benchmark testing across enterprise datasets. Proper guidelines established today will guarantee hybrid anomaly detection maintains its central role as a fundamental cybersecurity infrastructure component.

REFERENCES

Anil, 2013 — Anil S., Remya R. A hybrid method based on genetic algorithm, self-organised feature map, and support vector machine for better network anomaly detection // 2013 4th International Conference on Computing, Communications and Networking Technologies (ICCCNT). — 2013. — Pp.1–5. — 1DOI: https://doi.org/10.1109/ICCCNT.2013.6726604

Awad, 2023 — Awad O. F., Sulaiman S. K., Ali Alshmeel G. H. Anomaly Detection and Security Enhancement Through Machine Learning in Administrative Information Systems // 7th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT). — 2023. — Pp.1–8. — DOI: https://doi.org/10.1109/ISMSIT58785.2023.10304982

Ayu, 2023 — Ayu M. A., Erlangga D., Mantoro T., Handayani D. Enhancing SIEM by Incorporating Machine Learning for Cyber Attack Detection // 2023 IEEE 9th International Conference on Computing, Engineering and Design (ICCED). — 2023. — Pp. 798–807. — DOI: https://doi.org/10.1109/ICCED60214.2023.10425288

Aziz, 2013 — Aziz A. S. A., Hassanien A. E., Hanaf S. E. O., Tolba M. F. Multi-layer hybrid machine learning techniques for anomalies detection and classification approach // 13th International Conference on Hybrid Intelligent Systems (HIS). — 2013. — Pp. 215–220. — DOI: https://doi.org/10.1109/HIS.2013.6920485

Berbiche, 2023 — Berbiche N., el Alami J. Enhancing Anomaly-Based Intrusion Detection Systems: A Hybrid Approach Integrating Feature Selection and Bayesian Hyperparameter Optimization // Ingenierie Des Systemes d'Information. — 2023. — Vol. 28(5). — Pp. 1177–1195. — DOI: https://doi.org/10.18280/ISI.280506

Devi, 2023 — Devi V. A., Bhuvaneswari E., Tummala R. K. Decentralized Hybrid Intrusion Detection System for Cyber Attack Identification using ML // 2023 International Conference on Data Science, Agents and Artificial Intelligence (ICDSAAI). — 2023. — Pp. 763–768. — DOI: https://doi.org/10.1109/ICDSAAI59313.2023.10452439

Dhande, 2023 — Dhande M. T., Tiwari S., Rathod N. J., Professor A. HMCMA: Efficient Hybrid Machine Learning Model for Detection of Malicious Activities // International Journal on Recent and Innovation Trends in Computing and Communication. — 2023. — Vol. 11(11s). — Pp. 721–734. — DOI: https://doi.org/10.17762/IJRITCC.V11I11S.9729

Dobkacz, 2023 — Dobkacz L. Y., Sakulin S. A., Alfimtsev A. N., Kalgin Y. A. Hybrid Network Anomaly Detection Based on Weighted Aggregation Using Endpoint Parameters // Lecture Notes in Networks and Systems, Vol. 694. — 2023. — Pp. 269–278. — DOI: https://doi.org/10.1007/978-981-99-3091-3_21

Gupta, 2024 — Gupta A. K., Dixit N., Kumar S., Rawat P., Madhumita. A Novel Hybrid Approach for Threat Detection in Cyber Security using AI algorithm // 2024 International Conference on Computing, Sciences and Communications (ICCSC). — 2024. — Pp. 565–573.— DOI: https://doi.org/10.1109/ICCSC62048.2024.10830401

Harwahyu, 2024 — Harwahyu R., Ndolu F. H. E., Overbeek M. V. Three layer hybrid learning to improve intrusion detection system performance // International Journal of Electrical and Computer Engineering (IJECE). — 2024. — Vol. 14(2). — Pp. 1691–1699. — DOI: https://doi.org/10.11591/ijece.v14i2.pp1691-1699

Iwabuchi, 2024 — Iwabuchi M., Nakamura A. A Heuristics and Machine Learning Hybrid Approach to Adaptive Cyberattack Detection // International Conference on Artificial Intelligence, Computer, Data Sciences, and Applications (ACDSA). — 2024. — Pp. 891–898.— DOI: https://doi.org/10.1109/ACDSA59508.2024.10467929

Kale, 2022 — Kale R., Lu Z., Fok K. W., Thing V. L. L. A Hybrid Deep Learning Anomaly Detection Framework // 2022 IEEE 8th International Conference on Big Data Security, High Performance and Smart Computing, Intelligent Data and Security. — 2022. — Pp. 137–142. — DOI: https://doi.org/10.1109/BIGDATASECURITYHPSCIDS54978.2022.00034

Yadav, 2024 — Nayana Yadav M., Ananth Prabhu G., Souza M.D., Chaithra. Integrating AI with Cybersecurity: Review of Deep Learning for Anomaly Detection // Nanotechnology Perceptions. — 2024. — Vol. 20(S14). — Pp. 1756–1785. — DOI: https://doi.org/10.62441/NANO-NTP.VI.3007

Maheswari, 2024 — Maheswari M., Anitha D., Sharma A., Kaur K., Balamurugan V., Garikapati B., Dineshkumar R., Karunakaran P. Hybrid anomaly detection: autoencoder + random forest neural network // Journal of Intelligent & Fuzzy Systems. — 2024. — Pp. 1–14. — DOI: https://doi.org/10.3233/JIFS-240028

Mohite, 2024 — Mohite R., Ouarbya L. Interpretable Hybrid Anomaly Detection Using Rule-Based + ML Techniques // 2024 IEEE 9th International Conference for Convergence in Technology (I2CT). — 2024. — DOI: https://doi.org/10.1109/I2CT61223.2024.10543396

Nurusheva, 2024 — Nurusheva A., Abdiraman A., Satybaldina D., Goranin N., Gumilyov L. N. Machine learning algorithms in SIEM systems for enhanced detection of security events // Bulletin of ENU. Mathematics, Computer Science, Mechanics. — 2024. — Vol. 148(3). — Pp. 6–17. — DOI: https://doi.org/10.32523/BULMATHENU.2024/3.1

Pulyala, 2024 — Pulyala S. R. From Detection to Prediction: AI-powered SIEM for Proactive Threat Hunting // Turkish Journal of Computer and Mathematics Education. — 2024. — Vol. 15(1). — Pp. 34–43. — DOI: https://doi.org/10.61841/TURCOMAT.V15I1.14393

Rani, 2024 — Rani B. P., Lahari C. S., Lavanya J. A., Lakshmanarao A., Kosuri G. V. Advanced IDS Through Hybrid Integration of RF and Deep Learning // 2024 3rd International Conference for Advancement in Technology (ICONAT). — 2024. — Pp.1–5. — DOI: https://doi.org/10.1109/ICONAT61936.2024.10774823

Sharath, 2024 — Sharath T., Muthukumaravel A. Optimizing Security Operations: Hybrid Intrusion Detection Systems Leveraging Data Engineering // International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS). — 2024.— Pp.239–245. — DOI: https://doi.org/10.1109/IACIS61494.2024.10721814

Zhu, 2011 — Zhu H., Xin Y., Wang F. A novel anomaly detection framework based on hybrid HMM-SVM model // 2011 4th IEEE International Conference on Broadband Network and Multimedia Technology (IC-BNMT). — 2011. — Pp. 670–674. — DOI: https://doi.org/10.1109/ICBNMT.2011.6156020