https://doi.org/10.65542/djei. v2i2.49
Dasinya Journal for Engineering and Informatics. **2026**, *2*, *16*.

Article

A Hybrid Machine Learning Approach for Enhancing Intrusion Detection Systems Using CICIDS2017 Dataset

Ara Zozan Miran [1,2] *, **Govand Salih Kadir** [3]

[1] Department of Information Technology, Technical College of Duhok, Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq, ara.miran@auas.edu.krd
[2] Department of Information Technology, Technical College of Informatics- Akre, Akre University for Applied Science, ara.miran@auas.edu.krd
[3] School of Science and Engineering, University of Kurdistan Hewlêr (UKH), Erbil City, Iraq, g.kadir@ukh.edu.krd
* Correspondence: ara.miran@auas.edu.krd

Abstract

Traditional Intrusion Detection Systems (IDSs) are often ineffective because they rely on signature-based methods, resulting in high false-positive rates. With the expansion of new Artificial Intelligence (AI), especially Machine Learning (ML) algorithms, used in IDS systems, it is possible to achieve high performance in detecting anomalous and known threats. For that reason, this study aims to use Machine Learning (ML) algorithms to learn different patterns of known and anomalous threats across the Data Link (Layer 2) and Network (Layer 3) characteristics of the OSI model, as a basis for adaptive systems for IDS frameworks. Two supervised models (Random Forest, XGBoost) and two unsupervised models (Isolation Forest, One-Class SVM) were compared on the CICIDS2017 online dataset. The Results indicate that XGBoost achieved 99.3% accuracy on known threats, while One-Class SVM achieved 92.1% accuracy for unknown threats. Later, model performance was evaluated using standard classification metrics with a paired t-test. The findings of this study show the importance of combining supervised and unsupervised ML algorithms as a hybrid system to detect, classify, and learn from known and anomalous threats on different layers of network traffic. As a result, the findings can serve as a base for the adaptive IDS systems that can learn features and achieve higher performance.

Keywords: Intrusion Detection System (IDS); Machine Learning (ML); Supervised Learning; Unsupervised Learning; Anomaly Detection.

1. Introduction

With the rise in technology and its positive impact on our daily tasks, it also poses significant cybersecurity risks and cybercrimes. To minimize these risks, one security measure is the use of Intrusion Detection Systems (IDSs). This system monitors network traffic to detect known threats; however, it is limited by high false-positive rates and low true-negative rates [1].

Intrusion Detection Systems (IDSs) are integral to information security principles, as they are used to build network security architectures [4]. IDS serves as a critical frontline defense in cybersecurity architecture, providing a systematic

Received: March 17, 2026; Revised: April 18, 2026; Accepted: April 18, 2026; Published: May 08, 2026; Available online: May 08, 2026.

method for tracking and analyzing network and system activity to detect policy breaches and malicious activity [15]. These systems provide confidentiality, integrity, and availability for networked systems by preventing unauthorized access to the network [16]. Artificial Intelligence (AI), particularly machine learning (ML), will help the IDS system detect attacks more effectively and learn from attack patterns to adapt to unknown threats [1].

Supervised and unsupervised ML algorithms are used to detect and classify known and unknown threats across the OSI model. To compare ML algorithms and evaluate them using performance metrics such as accuracy, precision, recall, and F1-score. First, a testable dataset is required to obtain performance results therefore, the online CICIDS2017 dataset has been used to train ML algorithms.

A variety of algorithms, such as Decision trees, Random Forests, Support Vector Machines, and Ensemble methods, have been used in supervised learning, with relatively successful results in IDS [18]. Ensemble methods have demonstrated higher predictive performance by combining multiple base learners. Based on preliminary experimental results, the supervised model XGBoost was superior to the other models, achieving higher accuracy (99.3%), a higher F1-score (99.4%), and a better balance between precision and recall (99.3% and 99.5%). In unsupervised learning, anomaly-detection models, including One-Class SVM and the Isolation Forest algorithm, were evaluated to assess their ability to detect unusual or zero-day threats. One- Class SVM shows the highest result due to the expertise of measuring separation between abnormal traffic and unsupervised learning detection of attack compared with supervised models trained on a formed pattern, which shows higher results than Isolation Forest in terms of accuracy (92.1%), and higher F1-score (70.5%) and greater results in precision and recall (87.0%-59.2%). Unlike previous studies that focus only on a single network layer, this research focuses on combining two layers, specifically layers 2 and 3 of the OSI model. This approach provides a more comprehensive understanding of network behavior across multiple layers than traditional single-layer IDS studies. As the main contribution of this research is summarized in:

- A comparative evaluation of supervised (XGBoost, Random Forest) and unsupervised (One-Class SVM, Isolation Forest) machine learning algorithms using the CICIDS2017 dataset.
- Focusing on analyzing the network features of Layers 2 and 3 of the OSI model and highlighting their role in the early stages of intrusion detection.
- A hybrid IDS perspective that demonstrates how combining both approaches can support future adaptive security systems.

2. Related Work

Due to the rapid evolution of cyber threats, research on Intrusion Detection Systems (IDSs) has increased lately, and IDSs constitute the frontline defense that can benefit from knowledge gained from machine learning algorithms [14]. For that reason, there is an increasing body of research on machine learning and deep learning approaches to improve the detection of complex and emerging attacks, to increase accuracy, flexibility, and robustness [5]. From 2020 to 2026, studies have highlighted the success of supervised and unsupervised techniques, with algorithms such as Support Vector Machines (SVMs), Random Forests (RFs), XGBoost, and deep learning frameworks, including Long Short-Term Memory (LSTM) networks, achieving high levels of accuracy and robustness [12]. This research builds on these findings by combining supervised and unsupervised learning to strengthen multi-layer intrusion detection performance as a foundation for adaptive IDS systems. The CICIDS2017 dataset has been widely used because it provides extensive coverage of real-world network traffic and attack patterns, and it serves as a valuable benchmark for evaluating the generalization of different IDS models [11].

In a review paper [2], this author used the snowballing technique to review and analyze 49 research papers on ML algorithms for evaluating and enhancing IDS systems. The results show that the Random Forest algorithm can achieve up to 96%. The paper's most significant finding is that the algorithm achieves low false rates and high detection rates compared with other algorithms. Moreover, the research [3] developed an early IDS control network that combines a honeypot with the model-free reinforcement learning algorithm, State-Action-Reward-State-Action (SARSA). Two agents were used: an environmental agent that intentionally reduces the classification agent's reward to encourage complexity, and a classification agent that predicts the attack type. The framework was tested and compared to man-in-the-middle and DDoS attacks using real-world devices. Traditional supervised models (Support Vector Machine and

Random Forest) and other deep Reinforcement Learning approaches (Deep Q-Network, Double DQN, Actor-Critic, and Policy Gradient) were also evaluated on the benchmark. Although this research was limited to two attack types, it could be expanded to include meta-heuristic algorithms and additional Reinforcement Learning to improve scalability and robustness.

Additionally, [4] Proposed AI-enhanced IDS for honeypots that combine machine learning, Large Language Models (LLMs), and game theory to build deception systems capable of adapting to attacker behavior. The approach showed a 20% increase in engagement time and a 35% improvement in detection rate compared to traditional honeypots. However, it remains limited by training complexity and data quality issues. Future work can focus on automating model tuning and improving LLM interpretability for security contexts. Finally, [5] proposed an optimized machine learning model for intrusion detection that combines the Rao optimization algorithm with Support Vector Machine (SVM) and other ML methods for better feature selection and classification. They tested their model on the KDDCup99 and CICIDS 2017 datasets, achieving accuracies of 100% and 97.6%, respectively. The high results showed that ML algorithms can achieve high classification rates between attack and BENIGN on selected datasets.

Recent research also focused on deep learning concepts, particularly Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNNs), and hybrid deep learning frameworks [2]. Although those models achieve almost 99% accuracy on the CICIDS 2017 datasets, researchers faced challenges at the time, as the models required more resources, longer training times, and large labeled datasets. At the same time, machine learning models can achieve competitive performance while maintaining lower computational complexity. Within that, the Previous work indicates that combining supervised and unsupervised machine learning algorithms in a hybrid system provides a foundation for an adaptive IDS workflow.

3. Methodology

The methodology section consists of a dataset description, a data preprocessing stage, and the dimensionality reduction techniques used to compare the machine learning algorithms. This section describes the overall structure of the proposed IDS workflow as shown in Figure 1. The flowchart diagrams illustrate the proposed IDS methodology. The process began with dataset description and preprocessing, followed by feature extraction focusing on layers 2 and 3, and then Dimensionality reduction using Principal Component Analysis (PCA) to improve efficiency. The processed data was then used to train both supervised and unsupervised machine learning models, with performance evaluation using standard classification metrics.

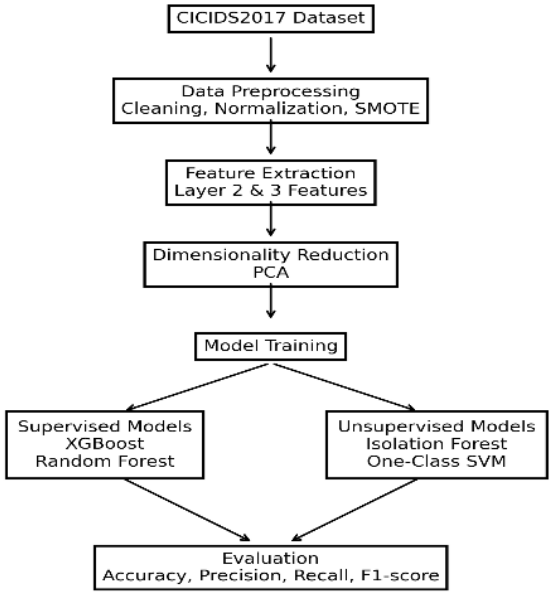

Figure 1. Flowchart Diagram of IDS Workflow

3.1. Dataset Description

The CICIDS2017 dataset is used to train and test the ML algorithms. This dataset covers a wide range of attacks across the OSI model's layers, though our focus is on Layer 2 and Layer 3 traffic features. These layers provide helpful information about packet characteristics and network traffic patterns, supporting the identification of unusual activity before data reaches the highest possible level of encryption at the application layer, based on features such as MAC addresses, IP headers, TTL (Time to Live) values, and protocol types. A total of 1,668,530 records were used in the experimental setup: 1,402,023 for BENIGN flows, and 266,507 for attack flows. Then, accuracy, precision, recall, and F1-score were compared for the selected ML algorithms.

3.2. Data Preprocessing

The proposed method uses a systematic workflow that includes preprocessing, feature extraction, model training, and evaluation, forming a foundation for future integration with adaptive IDS Frameworks. It is based on the CICIDS2017 dataset, which contains both BENIGIN and malicious network flows. The following steps of data preprocessing include cleaning incomplete records, applying Min-Max normalization, encoding categorical variables (e.g., protocol types), and using SMOTE to balance the data and reduce class imbalance. The data were then split into 60% for training, 20% for validation, and 20% for testing, with balanced evaluation of both normal and attack traffic. This split will ensure that sufficient data is available for testing the ML algorithms. The model configuration was then applied and optimized using a grid search. A total of 1,668,530 records were used, consisting of 938,547 for training, 312,850 for validation, and 417,133 for testing. This split is used to support model generalization for both known and unknown attacks on the network system.

3.3. Dimensionality Reduction

Dimensionality reduction was done by Principal Component Analysis (PCA). This technique is used to reduce the number of features in a dataset while preserving most of the original information [3]. The four selected ML algorithms, including Random Forest, XGBoost, Isolation Forest, and One-Class SVM, were compared using hyperparameters optimized via grid search and evaluated using accuracy, precision, recall, and F1-score to create a scalable, adaptive IDS Framework. The PCA visualizations are presented in the experimental section.

3.4. The selected Machine Learning (ML) Algorithms

Machine learning algorithms are classified as supervised, unsupervised, or hybrid [17]. Those algorithms are trained on datasets and improved by learning from patterns [6]. For that reason, machine learning algorithms are suitable for IDS detection of known and unknown threats, for reducing the number of attacks on systems. The rest of this research paper describes each selected algorithm in detail, along with its results.

3.4.1. Supervised Machine Learning Algorithms

Among the supervised machine learning algorithms in this research, two of the most effective models for IDS tasks have been discussed: Random Forest and XGBoost, supported by recent research demonstrating strong performance in accuracy, recall, and robustness across datasets such as CICIDS2017 and NSL-KDD. [4]Random Forest is an ensemble of decision trees built via bagging, in which each tree is trained on a random subset of data [21]. The second supervised algorithm used is XGBoost, which employs gradient boosting, including second-order objective functions, limited-data awareness, and regularization, to achieve high accuracy and efficiency [22]. Those selected algorithms are among the highest-accuracy and are suitable as the basis for the adaptive IDS system, as shown in the experimental results section.

3.4.2. Unsupervised Machine Learning Algorithms

In addition to the supervised algorithm, this research will compare two unsupervised machine learning algorithms with comparable capabilities for detecting and classifying threats across the two layers of the OSI model. These algorithms have been selected for their effectiveness, yielding high accuracy and a low error ratio. The first algorithm is Isolation Forest, which was selected for its ability to process data in tree format using randomly selected features [8]. Changes are less likely in samples farther down the tree, because they will require even more cuts to be isolated. Also,

the presence of the samples on shorter branches of the tree indicates abnormality. The second algorithm is the one-class SVM, which identifies anomalous samples by separating normal from abnormal samples using a hyperplane [20].

3.5. Experimental Setup

The Machine Learning models were implemented in Python using the scikit-learn and XGBoost libraries, and execution was performed on a workstation with an Intel Core i7 processor, 64 GB of RAM, and an NVIDIA RTX-series GPU. This combination will facilitate smoother data preprocessing and more efficient ML model training.

4. Experiments and Results

The experimental section presents the ML model's results, including several evaluation metrics, statistical validation, and confusion matrices for the selected models. Finally, PCA visualization is provided to show feature-space separation and dimensionality reduction, to understand the underlying data structure and the models' effectiveness in separating different classes.

4.1. Evaluation Metrics and Statistical Validation

The evaluation of results used standard classification metrics, including accuracy, precision, recall, and F1-score, to assess reliability and minimize bias in the datasets. To show performance differences, paired t-tests ($p < 0.05$) with 95% confidence intervals were used; these tests compare model performance using the same dataset. The assumptions of the tests were satisfied because the models were evaluated under the same conditions using consistent data splits. Table 1 shows that the four algorithms, including XGBoost, achieved the highest accuracy and F1-score. Additional evaluations, including ROC-AUC, precision-recall curves, and false-positive rate (FPR), are also required to assess metrics that will be considered for future evaluation.

Table 1. The Comparison Table of the Selected Four ML Algorithms.

Model	Type	Accuracy	Precision	Recall	F1-Score	p-value
Random Forest	**Supervised**	**98.8% (98.3–99.2)**	**99.2% (98.8–99.6)**	**98.8% (98.4–99.1)**	**99.0% (98.7–99.3)**	**0.042**
XGBoost	**Supervised**	**99.3% (98.9–99.6)**	**99.3% (98.9–99.6)**	**99.5% (99.2–99.7)**	**99.4% (99.1–99.7)**	**—**
Isolation Forest	**Unsupervised**	**91.6% (90.9–92.3)**	**83.1% (81.7–84.5)**	**59.7% (57.2–62.1)**	**69.5% (67.9–71.2)**	**<0.001**
One-Class SVM	**Unsupervised**	**92.1% (91.4–92.8)**	**87.0% (85.8–88.2)**	**59.2% (57.0–61.5)**	**70.5% (68.2–72.7)**	**<0.001**

4.2. Supervised Model Performance

The random forest model achieved 98.8% accuracy in classifying both normal and anomalous attack traffic, resulting in only a few misclassifications. Figure 2 indicates that this model has a strong capability to distinguish between BENIGN and malicious activities, supporting its role as a reliable supervised model.

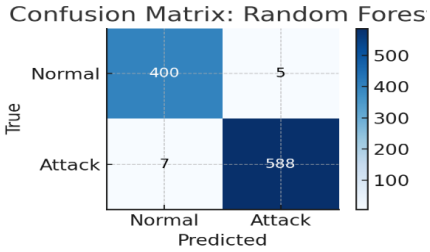

Figure 2. Confusion Matrix of Random Forest.

Figure 3 illustrates the confusion matrix for XGBoost, showing fewer misclassifications than the Random Forest on the same dataset when identifying threats on layers 2 and 3 of the OSI model. This shows that XGBoost achieves slightly better results, with 99.3% accuracy, indicating that it improves the enhancement and regularization processes and enhances its ability to handle imbalanced and complex data.

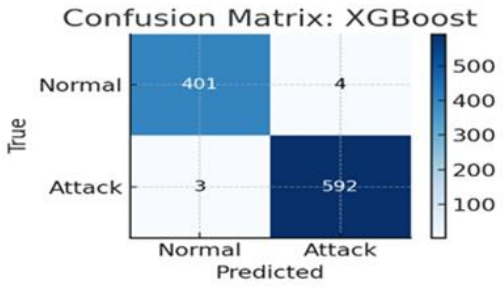

Figure 3. Confusion Matrix of XGBoost.

4.3. Unsupervised Model Performance

To train the Isolation Forest model, the label data were removed, indicating that the model is fully unsupervised and focuses solely on learning the pattern of BENIGN traffic from both Layer 2 and Layer 3 features of the OSI model. This will help identify anomalies and previously unseen patterns in the dataset that supervised models may have missed. Figure 4 shows that the model achieved an accuracy of 91.6%, which is substantially lower than supervised algorithms. Demonstrating that while Isolation Forest can detect anomalous behavior without labeling data, it has the weakness of a limited capability to distinguish between BENIGN and malicious traffic, which leads to higher misclassification compared to other models.

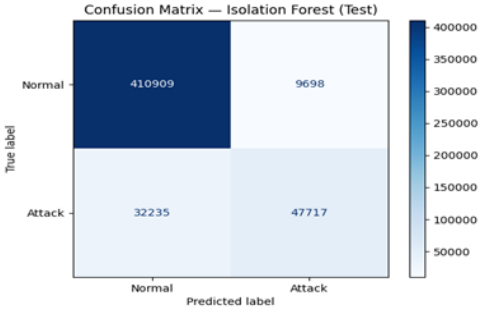

Figure 4. Confusion Matrix of Isolation Forest.

Figure 5 shows the confusion matrix for the One-Class SVM. The algorithm correctly classified most regular traffic (30,000 true negatives) and only a small fraction of the same attacks (507 false positives). 3,406 attacks were identified as true positives, and 2,344 were mistakenly identified as regular traffic (false negatives). The metrics show an accuracy of 92.1%, a precision of 87.0%, and an F1 Score of 70.5%. This algorithm can better separate the attack and BENIGN classes, indicating that it achieves higher results than the isolation forest.

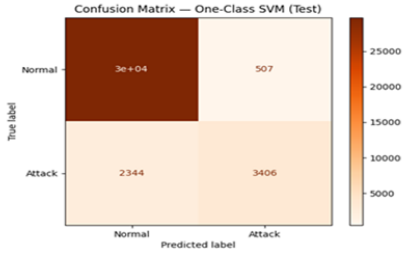

Figure 5. Confusion Matrix of One Class- SVM.

4.4. Principal Component Analysis (PCA) Test

The Principal Component Analysis (PCA) was used to reduce computational complexity, remove noise, and identify the most informative patterns. This research used the PCA as a dimensionality reduction method that preserves the most significant information by transforming correlated variables into fewer uncorrelated principal components. In the intrusion detection setting, PCA helps visualize attack and regular traffic and serves as a preprocessing step to improve ML model performance by eliminating redundant or irrelevant features [27].

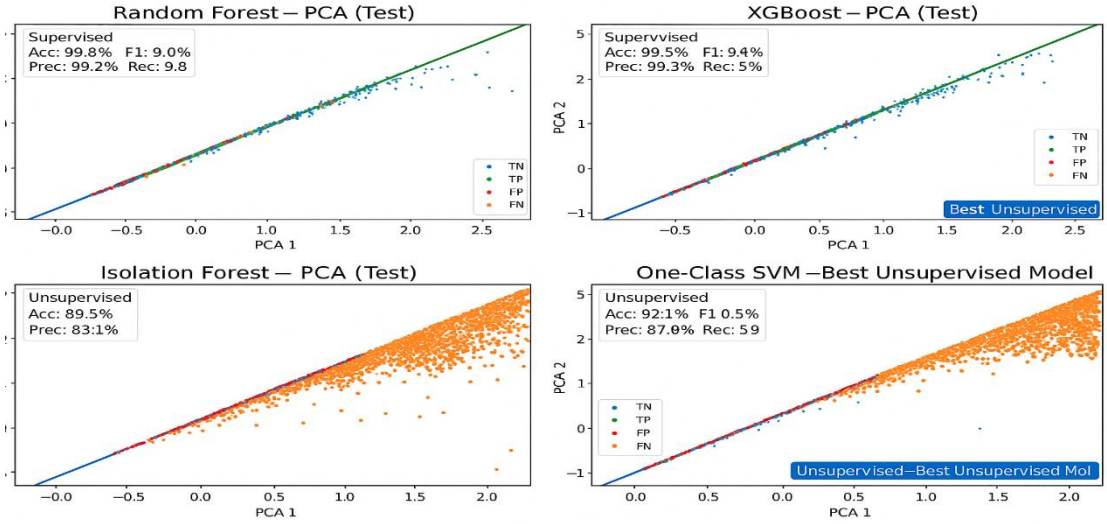

Figure 6. (PCA) of ML Models.

Figure 6 shows the predictions of Layer 2 and 3 traffic features using PCA under the four Machine Learning algorithms: Random Forest, XGBoost, Isolation Forest, and One-Class SVM. The points are samples from the test set, and the classification results are shown as (blue) true negatives, (green) true positives, (orange) false positives, and (red) false negatives. The supervised models (top row) are more clustered and better separated between BENIGN and malicious traffic; specifically, XGBoost establishes the most precise boundary, resulting in minimal misclassifications. This proves that it is more effective with (99.3% accuracy and 99.4% F1-score). Also, Random Forest is accurate with (98.8%), but with more isolated results. The unsupervised models (bottom row) have higher error rates, with the Isolation Forest showing more false positives and false negatives (higher error rates). On the other hand, as shown, the One-Class SVM achieves higher precision (87.0%) and an F1 score (70.5%) than the Isolation Forest. Overall, the results indicate that among supervised algorithms, XGBoost achieves the highest detection of known threats. As the algorithm can learn from complex patterns in the dataset and work well with tabular datasets like the one used, it can easily distinguish between known and anomalous features. On the other hand, among unsupervised algorithms, One-Class SVM achieves the highest detection performance against anomalous threats. However, overall performance remains lower than that of supervised models, as unsupervised models are trained only on normal data and lack access to labeled attack data. This highlights the importance of combining supervised and unsupervised models to achieve a better balanced and improved accuracy in the IDS Framework, as supervised models focus on accuracy and unsupervised models detect anomalies or previously unseen behavior without labeled data.

5. Conclusions

This research evaluated and tested the performance of four Machine Learning algorithms: two supervised (XGBoost and Random Forest) and two unsupervised (One-Class SVM and Isolation Forest). These algorithms aim to detect known and unknown threats using the CICIDS2017 dataset. The two selected supervised models were XGBoost and Random Forest. XGBoost achieved higher accuracy (99.3%) with balanced precision, F1 Score, and recall. On the other hand, the two selected unsupervised models were One-Class SVM and Isolation Forest, as One-Class SVM achieved higher accuracy, F1-score, and AUC (92.1%). This result underscores the importance of this algorithm for unknown attacks in the absence of labels. The Isolation Forest offers scalability and higher recall results (59.7%). Those

algorithms were selected to detect attacks at the second and third layers of the OSI model, as both XGBoost and One-Class SVM showed the greatest impact on hybrid IDS systems.

From the practical viewpoint, the use of those models provides an effective and understandable solution for real-time data across different environments. The findings highlight the impact of using both supervised and unsupervised models, especially when analyzing multiple layers of the network.

Future work will use the results of the two best-performing algorithms to detect and learn from known and unknown threats and then adapt them to IDS frameworks and adaptive honeypot systems to detect such threats and enable dynamic cyber defense. As this research focused only on the CICIDS2017 dataset to evaluate network performance, generalization to other datasets has not been evaluated. For that, cross-dataset validation using benchmarks such as UNSW-NB15 and CICDDoS2019 will be evaluated in the future, using k-fold cross-validation to improve reliability.

Author Contributions: This research was contributed by both authors, Ara and Govand. As in all research parts, including writing, visualization, conceptualization, analysis, and resource collection, the first author, Ara, conducted all work under the supervision of the second author, Govand. All authors have read and agreed to the published version of the manuscript".

Funding: This research received no external funding.

Acknowledgments: The authors have reviewed and edited the output and take full responsibility for the content of this publication."

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	**Artificial Intelligence**
CNNs	**Convolutional Neural Networks**
DDOS	**Distributed Denial of Service**
FPR	**False-Positive Rate**
IDS	**Intrusion Detection System**
IP	**Internet Protocol**
LSTM	**Long Short-Term Memory**
MAC	**Media Access Control**
ML	**Machine Learning**
OSI	**Open Systems Interconnection**
RF	**Random Forest**
SARSA	**State-Action-Reward-State-Action**
SVM	**Support Vector Machines**

References

1. Ahmad, Z.; Khan, A.S.; Shiang, C.W.; Abdullah, J.; Ahmad, F. Network intrusion detection system: A systematic study of machine learning and deep learning approaches. Trans. Emerg. Telecommun. Technol. 2021, 32, e4150. https://doi.org/10.1002/ett.4150.
2. Alhayali, R.A.I.; Aljanabi, M.; Ali, A.H.; Mohammed, M.A.; Sutikno, T. Optimized machine learning algorithm for intrusion detection. Indones. J. Electr. Eng. Comput. Sci. 2021, 24, 590–599. https://doi.org/10.11591/ijeecs.v24.i1.pp590-599
3. Chua, T.H.; Salam, I. Evaluation of machine learning algorithms in network-based intrusion detection using progressive dataset. Symmetry 2023, 15, 1251. https://doi.org/10.3390/sym15061251
4. Crespo-Martínez, I.S.; Campazas-Vega, A.; Guerrero-Higueras, Á.M.; Riego-DelCastillo, V.; Álvarez-Aparicio, C.; Fernández-Llamas, C. SQL injection attack detection in network flow data. Comput. Secur. 2023, 127, 103093. https://doi.org/10.1016/j.cose.2023.103093

5. Ileberi, E.; Sun, Y.; Wang, Z. Performance evaluation of machine learning methods for credit card fraud detection using SMOTE and AdaBoost. IEEE Access 2021, 9, 165286–165294. https://doi.org/10.1109/ACCESS.2021.3134330
6. Khan, M.M. Developing an AI-powered intrusion detection system for cloud infrastructure. J. Artif. Intell. Mach. Learn. Data Sci. 2024, 2, 1074–1080. https://doi.org/10.51219/JAIMLD/mohammed-mustafa-khan/255
7. Lew, J.; Shah, D.A.; Pati, S.; Cattell, S.; Zhang, M.; Sandhupatla, A.; Ng, C.; Goli, N.; Sinclair, M.D.; Rogers, T.G.; Aamodt, T.M. Analyzing machine learning workloads using a detailed GPU simulator. In Proceedings of the 2019 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS); IEEE: Madison, WI, USA, 2019; pp. 151–152. https://doi.org/10.1109/ISPASS.2019.00028
8. Lifandali, O.; Abghour, N.; Chiba, Z. Feature selection using a combination of ant colony optimization and random forest algorithms applied to an isolation forest-based intrusion detection system. Procedia Comput. Sci. 2023, 220, 796–805. https://doi.org/10.1016/j.procs.2023.03.106
9. Mahbooba, B.; Timilsina, M.; Sahal, R.; Serrano, M. Explainable artificial intelligence (XAI) to enhance trust management in intrusion detection systems using a decision tree model. Complexity 2021, 2021, 6634811. https://doi.org/10.1155/2021/6634811
10. Maseer, Z.K.; Yusof, R.; Bahaman, N.; Mostafa, S.A.; Foozy, C.F.M. Benchmarking of machine learning for anomaly-based intrusion detection systems in the CICIDS2017 dataset. IEEE Access 2021, 9, 22351–22370. https://doi.org/10.1109/ACCESS.2021.3056614
11. Mhamdi, L.; Isa, M.M. Securing SDN: Hybrid autoencoder-random forest for intrusion detection and attack mitigation. J. Netw. Comput. Appl. 2024, 225, 103868. https://doi.org/10.1016/j.jnca.2024.103868
12. Ozkan-Okay, M.; Samet, R.; Aslan, O.; Gupta, D. A comprehensive systematic literature review on intrusion detection systems. IEEE Access 2021, 9, 157727–157760. https://doi.org/10.1109/ACCESS.2021.3129336
13. Parveen Sultana, H.; Shrivastava, N.; Dominic, D.D.; Nalini, N.; Balajee, J.M. Comparison of machine learning algorithms to build an optimized network intrusion detection system. J. Comput. Theor. Nanosci. 2019, 16, 2541–2549. https://doi.org/10.1166/jctn.2019.7929
14. Sadiq, S.; Eesa, A.S. Optimization algorithms for intrusion detection system: A review. Int. J. Res. Granthaalayah 2020, 8, 217–225. https://doi.org/10.29121/granthaalayah.v8.i8.2020.1031
15. Silivery, A.K.; Rao Kovvur, R.M.; Solleti, R.; Kumar, L.S.; Madhu, B. A model for multi-attack classification to improve intrusion detection performance using deep learning approaches. Meas. Sens. 2023, 30, 100924. https://doi.org/10.1016/j.measen.2023.100924
16. Singh Chinthalapudi, S. Detecting and mitigating SQL injection in .NET applications using AI-based anomaly detection. Int. J. Innov. Sci. Res. Technol. 2025, 10, 2582–2595. https://doi.org/10.38124/ijisrt/25mar1676
17. Smith, J.; Kevin, E. AI-powered intrusion detection systems for next-generation cloud. ResearchGate 2025. Available online: https://www.researchgate.net/publication/390448273
18. Sulaiman, N.S.; Nasir, A.; Othman, W.R.W.; Wahab, S.F.A.; Aziz, N.S.; Yacob, A.; Samsudin, N. Intrusion detection system techniques: A review. J. Phys. Conf. Ser. 2021, 1874, 012042. https://doi.org/10.1088/1742-6596/1874/1/012042
19. Wang, B.X.; Chen, J.L.; Yu, C.L. An AI-powered network threat detection system. IEEE Access 2022, 10, 54029–54037. https://doi.org/10.1109/ACCESS.2022.3175886
20. Wang, C.; Sun, Y.; Lv, S.; Wang, C.; Liu, H.; Wang, B. Intrusion detection system based on one-class support vector machine and Gaussian mixture model. Electronics 2023, 12, 930. https://doi.org/10.3390/electronics12040930
21. Waskle, S.; Parashar, L.; Singh, U. Intrusion detection system using PCA with random forest approach. In Proceedings of the 2020 IEEE International Conference on Electronics and Sustainable Communication Systems (ICESC); IEEE, 2020. https://doi.org/10.1109/ICESC48915.2020.9155656
22. Xu, W.; Fan, Y. Intrusion detection systems based on logarithmic autoencoder and XGBoost. Secur. Commun. Netw. 2022, 2022, 9068724. https://doi.org/10.1155/2022/9068724
23. Al Hasan, R.A.; Hamza, E.K. An improved intrusion detection system using machine learning with singular value decomposition and principal component analysis. Int. J. Intell. Eng. Syst. 2023, 16, 25–38. https://doi.org/10.22266/ijies2023.0831.03
24. Lanz, S.; Pignol, S.L.R.; Schmitt, P.; Wang, H.; Papaioannou, M.; Choudhary, G.; Dragoni, N. Optimizing Internet of Things honeypots with machine learning: A review. Appl. Sci. 2025, 15, 5251. https://doi.org/10.3390/app15105251
25. Onyebueke, A.E.; David, A.; Munu, S. Network intrusion detection system using XGBoost and random forest algorithms. Asian J. Pure Appl. Math. 2023, 5, 1–?. https://doi.org/10.54254/2753-8818/31/20241171

26. Pashaei, A.; Akbari, M.E.; Zolfy Lighvan, M.; Charmin, A. Early intrusion detection system using a honeypot for industrial control networks. Results Eng. 2022, 16, 100576. https://doi.org/10.1016/j.rineng.2022.100576
27. Saranya, T.; Sridevi, S.; Deisy, C.; Chung, T.D.; Khan, M.K.A.A. Performance analysis of machine learning algorithms in intrusion detection systems: A review. Procedia Compute. Sci. 2020, 171, 1251–1260. https://doi.org/10.1016/j.procs.2020.04.133