Journal of Information Systems Engineering and Management
2025, 10(62s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Investigating the Effectiveness of Supervised and Deep Learning Models: An Assessment for Binary Intrusion Detection in Structured Network Traffic

Abdullah Albalawi[1],
[1]*Department of Computer Science, College of Computing and Information Technology, Shaqra University, Shaqra, Saudi Arabia.*
Email:aalbalawi@su.edu.sa

ARTICLE INFO

Received: 18 Dec 2024

Revised: 10 Feb 2025

Accepted: 28 Feb 2025

ABSTRACT

Introduction: The growing complexity and frequency of cyber threats necessitate more adaptive and accurate intrusion detection systems (IDS). Traditional rule-based systems often fail to detect emerging cyberattacks patterns, underscoring the importance of data-driven approaches. Machine-learning methods, particularly deep learning and supervised learning, often promising capabilities for detecting anomalies in complex network environments. This study addresses the rising need for robust IDS models capable of distinguishing between benign and malicious traffic with high precision.

Objectives: The primary objective of this research is to perform a comparative assessment of four supervised learning models, including Multi-Layer Perceptron, Bagging Tree Ensemble, CatBoost, and One-Dimensional Convolutional Neural Network, for binary intrusion detection. The study aims to examine their performance in terms of accuracy, precision, recall, and F1-score, thereby identifying the most effective classifier for structured network traffic data.

Methods: A labeled network traffic dataset encompassing 4,000 records and 25 attributes was used. The dataset included both categorical and continuous variables. There was a class imbalance, with most of the instances being benign network traffic. Preprocessing was done by applying one-hot encoding to categorical features and z-score normalization to numerical features. All models were trained based on an 80:20 stratified train-test split and evaluated by traditional performance measures.

Results: CatBoost exhibited the highest overall performance among the models, attaining an accuracy of 0.89, a precision of 0.99, and an F1-score of 0.86. The Bagging Tree model attained perfect precision (1.00). The 1D CNN model demonstrated competitive recall and F1-score. The MLP had poor recall due to the increased number of false negatives. The results of the study have shown the efficiency of ensemble and deep learning approaches for binary intrusion classification. while also revealing the inherent precision-recall trade-offs among different models.

Conclusions: The study underscores the effectiveness of CatBoost and other ensemble-based approaches for intrusion-detection in structured network traffic datasets. It also emphasizes the importance of selecting models based on organizational risk tolerance. The research provides a practical performance benchmark to guide future IDS model selection and optimization in cybersecurity applications.

Keywords: Intrusion Detection System (IDS); Supervised Learning; CatBoost; Ensemble Methods; Deep Learning; Network Security

INTRODUCTION

The growing interconnectivity of digital systems has significantly expanded the attack surface of contemporary networks, rendering them vulnerable to a variety of cyber threats, including denial-of-service (DoS) attacks, brute-force invasions, and malware infections.[1] Conventional intrusion detection systems (IDS), which predominantly depend on static rules and signature-based detection, generally fail to recognize new or evolving attack patterns.[2]

Journal of Information Systems Engineering and Management
2025, 10(62s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

This has resulted in an increasing interest in machine learning (ML)–based intrusion detection methods, which possess the capability to learn from data and adapt to emerging threats. [3,4]

Supervised learning methodologies have been extensively utilized in this context, encompassing conventional models such as Support Vector Machines (SVM), Random Forests (RF), and Multi-Layer Perceptrons (MLP), as well as contemporary implementations involving Convolutional Neural Networks (CNNs) and ensemble tree-based models such as Bagging and Boosting techniques.[5–7] These algorithms have demonstrated promising efficacy in identifying intrusions on benchmark datasets. Nonetheless, practical challenges persist, notably the problem of class imbalance, which impacts detection sensitivity, alongside the necessity for models that balance accuracy, recall, and computational efficiency.[8]

Although several studies have investigated intrusion detection using extensive legacy datasets like NSL-KDD[9] and CICIDS2017[10], there is a paucity of research assessing various supervised classifiers using compact, contemporary datasets that more accurately represent limited operational contexts. Moreover, much less attention is usually paid to the analyses of confusion matrices and the performance trade-offs across models, although they are relevant to real-time deployment decisions.

The latest research on both machine learning (ML) and deep learning (DL) has significantly promoted the performance of IDS by enhancing its accuracy and decreasing the number of false alarms. Multiple studies have shown that supervised and ensemble methods surpass traditional models in identifying cybersecurity risks, particularly in scenarios with class imbalance and intricate feature spaces.

1D Convolutional Neural Networks (1D-CNN) have proven to be particularly adept in capturing temporal relationships in network traffic data. Setiawan et al. (2024)[11] employed a 1D-CNN on the NF-UQ-NIDS-v2 dataset, which yielded a classification accuracy of 94% with a higher performance in detecting DDoS, DoS, bot, and scanning attacks. Nonetheless, performance deteriorated for infiltration and worm-based attacks. Chen et al. (2023)[12] enhanced this using an AdaBoost-CNN hybrid, resulting in superior classification accuracy and resilience against various attack types. Additionally, Benaddi et al. (2023)[13] combined CNN with LSTM for IoT contexts (Bot-IoT dataset), attaining a 99.20% accuracy and 0.80% false alarm rate. Designs of similar architecture for SDNIoT networks yielded accuracies of 99.80% (Thandalam & Anithaashri, 2023)[14], that conform the adaptability of CNN to various cybersecurity contexts.

The methodologies of ensemble learning, namely Bagging, Boosting, and Stacking, have been found performing especially well in IDS. A Bagging and Partial Decision Tree ensemble attained an accuracy of 99.7166% during cross-validation (Gaikwad & Thool, 2015)[15]. A hybrid model integrating AdaBoost, Bagging, and classifiers like SVM, RF, and KNN achieved an accuracy of 99.7%, 0.053 FNR, and 0.004 FAR on the CICIDS2017 dataset (Mhawi et al., 2022)[16]. In the IoMT environments of hospitals, Stacking reached the highest accuracy at 98.88%, outperforming Bagging at 97.83% and Boosting at 88.68% (Alsolami et al., 2024)[17]. Similarly, Ibrahim and Al-Wadi (2024)[18] revealed that weighted voting ensembles that combine Logistic Regression, AdaBoost, and XGBoost showed an accuracy of 99.60%, hence proving the efficiency of hybrid meta-learners.

The ensembles of Gradient Boosting Decision Tree (GBDT), especially CatBoost, has shown robustness in handling imbalanced datasets. Louk and Tama (2022)[19] concluded that CatBoost showed greater accuracy and stability. Subsequently, Louk and Tama (2023)[20] presented a dual ensemble (Bagging + GBDT), that is even more effective for detection. Du et al. (2023)[21] suggested a two-stage ensemble of CNN and CatBoost, resulting in significant enhancements in classification accuracy. Yilmaz and Bardak (2022)[22] found that XGBoost regularly surpassed other models in F1 score across multiple datasets.

One comparative study, Belouch and El Hadaj (2017)[23] showed that the performance of stacking ensembles outperformed boosting or bagging alone. Tama and Rhee (2017)[24] corroborated this by assessing five ensemble strategies, noting that stacking and boosting yielded better classification accuracy and reduced false positive rates in comparison to bagging and rotation forests.

Advanced feature selection methods like mRMR and hybrid encodings have been employed to diminish dimensionality and enhance detection performance [25,26]. Techniques like SMOTE, hybrid sampling, and ranked feature bagging were essential for analyzing imbalanced data. Azhagiri et al. (2024)[27] achieved 99.71% accuracy using Bagging with minimal features, while Pham et al. (2018)[28] and Zhang (2007)[29] validated the effectiveness of Bagging

with improved feature sets. Meanwhile, Hou et al. (2022)[30] revealed that semi-supervised learning employing Self-Training Mixup Decision Trees (STM-DT) produced elevated macro F1 scores, rendering them suitable for scenarios with scarce labeled data.

Random Forest has proven itself as a highly efficient classifier by achieving an accuracy of 99.886% [31] and AUC-ROC of 0.98 in tests for malware detection[32]. Decision Trees, especially with the use of the CART model, were highly effective at multiclass classification tasks with an average macro F1-score of 0.96878 [33]. At the same time, SVM and Naïve Bayes achieved inconsistent performance depending on feature dimensionality and dataset [34,35].

The NSL-KDD and CICIDS-2017 datasets have remained the benchmark to evaluate IDS models until this day[26,36]. The CICIoTDataset2023 has played a significant role in research on IoT-specific IDS, with Random Forest achieving an accuracy of 98.41%[35]. The utilization of CatBoost and LightGBM in Bakhareva et al. (2019)[37] showed strong performance in both binary and multiclass scenarios.

Studies have emphasized the role of confusion matrices and precision–recall curves in performance evaluation. Manai et al. (2024)[38] showed that confusion matrix-based analysis improved model precision and recall. Hasanin et al. (2019)[39] and Zuech et al. (2021)[40] showed precision values over 97% with CNN-based classifiers, with recall rates also reaching 96%.

The issue of labeling scarcity has been addressed through the utilization of GANs and semi-supervised learning. Kumar and Sinha (2023)[41] employed Wasserstein Conditional GANs (WCGAN) in conjunction with XGBoost to address imbalance issues. Likewise, Hakim et al. (2025)[42] employed SMOTE and resampling techniques to enhance the detection of exfiltration attacks.

Multi-label intrusion detection has gained attention with two-stage model fusion methodologies providing improved categorization of intricate attack vectors [43]. Rajput and Upadhyay (2025)[44] also constructed hybrid models that surpassed individual classifiers in identifying DDoS attacks, attaining an accuracy of 99.65%.

This research study proposes a comparison of four machine learning models for binary intrusion detection against the identified challenges. These models include Multi-Layer Perceptron (MLP), Bagging Tree ensemble, CatBoost classifier, and a one-dimensional Convolutional Neural Network (1D CNN). The basis of the analysis is a labeled dataset consisting of 4,000 records of network sessions with 25 categorical and continuous features. The methodology also included standardized preprocessing, stratified sampling, and uniform evaluation through accuracy, precision, recall, F1-score, and AUC metrics. This is aimed at assessing the performance of these models and analyzing their strengths and weaknesses in differentiating benign and malicious traffic.

This research work provides a structured and reproducible pipeline for the study of binary intrusion detection, putting an emphasis on practical insights such as error patterns based on the confusion matrix and recall limitations of the models that can be developed for real-world security systems.

OBJECTIVES

The primary objective of this research is to perform a comparative assessment of four supervised learning models, including Multi-Layer Perceptron, Bagging Tree Ensemble, CatBoost, and One-Dimensional Convolutional Neural Network, for binary intrusion detection. The study aims to examine their performance in terms of accuracy, precision, recall, and F1-score, thereby identifying the most effective classifier for structured network traffic data.

METHODS

This study utilizes a dataset comprising 4,000 network traffic records and 25 variables, which include both categorical and continuous features pertinent to intrusion detection. Continuous features comprise packet length (20 to 1,500 bytes), payload size (0 to 1,200 bytes), and anomaly score (0.0 to 1.0). Categorical attributes include protocol type (e.g., TCP, UDP, ICMP), attack type, and severity level. The target variable “attack type” includes seven classes: Benign, DoS, DDoS, Brute Force, Port Scan, Botnet, and Ransomware. There are three categories in the severity level: Low, Medium, and High. In this dataset, the instances are labeled for supervised learning, which is an imbalanced class problem, having approximately 55% benign and 45% malicious traffic instances.

Journal of Information Systems Engineering and Management
2025, 10(62s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The dataset was imported into MATLAB and preprocessed by removing rows containing missing values. A one-hot encoding for protocol type, encryption status, and browser type was carried out, while the target variable "attack detected" was designated as the response label Y. Five numeric features were standardized using z-score normalization; these included network packet size, session duration, IP reputation score, login attempts, and failed logins. Finally, the feature matrix X consisted of the normalized numeric variables combined with the encoded categorical variables.

The stratified sampling of the dataset provided an 80:20 ratio for training and testing. The subsets, including XTrain, YTrain, XTest, YTest, were utilized to train a Multi-Layer Perceptron (MLP) classifier, which was configured with two hidden layers containing 128 and 64 neurons, respectively. The MLP was trained using the scaled conjugate gradient algorithm.

An ensemble classifier (based on the bagging technique) was trained using 100 decision trees. Each was limited to a maximum of 30 splits. The model was used for binary classification. It was evaluated using the test dataset by generating both predicted class labels and class probabilities.

An AdaBoostM1 ensemble classifier was also trained, incorporating categorical predictors explicitly. The Session Id and target variable were omitted from the feature set. The training involved 100 boosting iterations utilizing decision tree learners on an 80:20 train-test split. The categorical characteristics of protocol type, encryption status, and browser type were preserved throughout the training process.

A one-dimensional Convolutional Neural Network (1D CNN) was utilized for the purpose of binary classification. The dataset instances were transformed into one-dimensional sequences of features. Then it was fed into the model as an ordered signal. The network architecture comprised an input layer for sequences, followed by two one-dimensional convolutional layers containing 32 and 64 filters, respectively. Each convolutional layer was succeeded by batch normalization and a rectified linear unit (ReLU) activation function. A global average pooling layer was employed to diminish spatial dimensions prior to transmitting the output to a fully connected layer consisting of two neurons. A softmax layer was subsequently followed by a classification layer. The model was trained with the Adam optimizer for 30 epochs, utilizing a mini-batch size of 64. It was validated on the test set during training.

Accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC) were calculated for all trained models to facilitate comparative analysis.

RESULTS

The confusion matrices depict the allocation of accurate and inaccurate predictions among benign (class 0) and malicious (class 1) instances. The CatBoost model recorded the highest counts of true positive (514) and true negative (837), alongside a minimal number of false positives (6), demonstrating robust discrimination capability. The Bagging Tree and 1D CNN models also demonstrated effective performance, particularly in reducing false negatives. Conversely, the MLP model exhibited the greatest number of false negatives (190), thereby diminishing its recall performance. These visualizations offer insight into error patterns specific to models, hence complementing metric-based evaluations.

Journal of Information Systems Engineering and Management
2025, 10(62s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

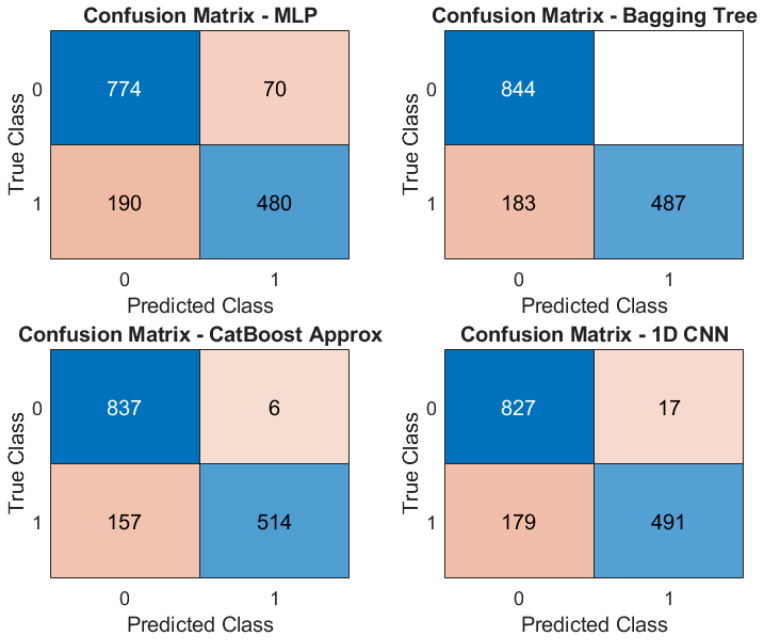

Figure 1 Confusion matrices showing true vs. predicted classifications for MLP, Bagging Tree, CatBoost, and 1D CNN models on the intrusion detection test set.

Figure 2 and Table 1 present a comparative analysis of four classification models based on Accuracy, Precision, Recall, F1-Score, and Area Under the ROC Curve (AUC). CatBoost demonstrated the highest accuracy (0.89), while also exhibiting a robust balance in precision (0.99), recall (0.77), and F1-score (0.86). The Bagging Tree model gave perfect precision of 1.00 and an AUC value of 0.87, competitive with other models. The 1D CNN model demonstrated high precision (0.97) alongside balanced recall and F1-score values. The MLP model exhibited moderate performance across all metrics, recording the lowest recall (0.72) within the group.

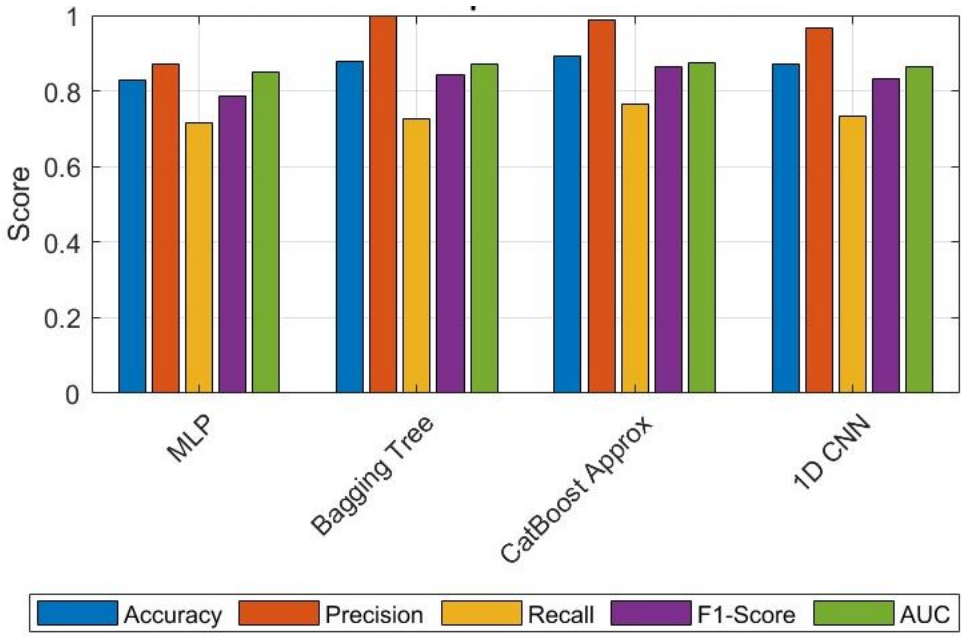

Figure 2 Comparative performance of MLP, Bagging Tree, CatBoost, and 1D CNN models across five evaluation metrics: Accuracy, Precision, Recall, F1-Score, and AUC.

Journal of Information Systems Engineering and Management
2025, 10(62s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 1 Bar chart comparing the classification performance of MLP, Bagging Tree, CatBoost, and 1D CNN models using Accuracy, Precision, Recall, F1-Score, and AUC metrics.

Model	Accuracy	Precision	Recall	F1_Score	AUC
MLP	0.83	0.87	0.72	0.79	0.85
Bagging Tree	0.88	1.00	0.73	0.84	0.87
CatBoost	0.89	0.99	0.77	0.86	0.87
1D CNN	0.87	0.97	0.73	0.83	0.86

DISCUSSION

This study has done binary intrusion detection using a structured network traffic dataset consisting of 4,000 labeled records. Four supervised machine learning models were employed: Multi-Layer Perceptron (MLP), Bagging Tree Ensemble, CatBoost, and a one-dimensional Convolutional Neural Network (1D CNN). The primary aim was to benchmark these models using a unified evaluation framework based on precision, recall, F1-score, AUC, and confusion matrix analysis to inform their applicability in real-world intrusion detection systems (IDS).

CatBoost exhibited superior performance among the tested models, attaining the highest accuracy (0.89), precision (0.99), and F1-score (0.86) across the evaluated metrics. These findings align with recent literature that highlights the effectiveness of gradient-boosted decision tree algorithms in managing structured tabular data, especially in contexts involving imbalanced class distributions and intricate nonlinear feature interactions.[6,7] The Bagging Tree model demonstrated a marginally lower recall but attained perfect precision (1.00). It reflects its conservative approach toward identifying benign traffic while keeping high accuracy in detecting confirmed intrusions. The 1D CNN showed strong predictive performance, achieving a precision of 0.97 and a recall of 0.73. This indicates its efficacy in extracting hierarchical feature patterns from flattened tabular input.

Conversely, the MLP model demonstrated a higher incidence of false negatives, as indicated by its lower recall (0.72) and F1-score (0.79). This suggests that, though architectural flexibility, the MLP was less effective in distinguishing malicious traffic from benign sessions using the current feature configuration and given dataset size. This might be explained by the fact that the model is sensitive to overfitting with limited-scale datasets, especially when complex temporal or sequence patterns are not present in the data.

Analysis of the confusion matrix corroborates these observations. CatBoost had better true positive and true negative counts, while at the same time keeping minimal false positives. On the other hand, MLP exhibited the highest rate of false negatives. This difference highlights how different model inductive biases impact the sensitivity and specificity trade-offs. This is an important factor in designing IDSs because false negatives or missed attacks could be much more costly than false positives or false alerts.

This study differs from previous research that assesses models using legacy datasets like NSL-KDD or artificial class balance simulations by maintaining the original class distribution (55% benign, 45% malicious), thereby providing a more accurate performance evaluation. All models were evaluated using stratified sampling and z-score normalized inputs, which guaranteed comparability and reproducibility of the experimental conditions.

Ensemble models, which consist of algorithms such as CatBoost and Bagging Tree, optimally balance interpretability, precision, and low variance. Hence, these are suitable for real-time IDS applications that prioritize reliability and minimum false alarm rates. Deep learning models, such as 1D CNN, can function as viable alternatives in contexts where computational resources are ample and the learning of temporal signals is required. In this study, the MLP's limited performance indicates that further feature engineering, architecture optimization, or expansion of training data is necessary to improve its effectiveness.

This study provides a concise and thorough benchmark for supervised binary intrusion detection utilizing structured features. It underscores the comparative advantage of tree-based ensembles versus convolutional models and emphasizes the impact of model selection on false negative rates, along with establishing a reproducible framework

for future research in IDS systems, particularly aimed at achieving lightweight and high-accuracy classification in constrained dataset environments.

In conclusion, this study conducted a comparative evaluation of four supervised learning models, including MLP, Bagging Tree, CatBoost, and 1D CNN, for binary intrusion detection utilizing a structured dataset comprising 4,000 network traffic records. The study demonstrated the ranking of CatBoost as the best model in terms of achieving the highest accuracy and F1-score, closely followed by the Bagging Tree and 1D CNN, through standardized preprocessing, stratified sampling and strict metric-based evaluation. The MLP model exhibited constraints in recall, suggesting an increased propensity for false negatives. The examination of the confusion matrix presented significant trade-offs between sensitivity and specificity within the models, which imposes careful model selection in real-world IDS deployment. Ensemble models presented strong, interpretable results, while deep learning models showed competitive performance given limited inputs. This work pinpoints a benchmarking framework for IDS design on grounds of realistic imbalanced data distributions and reproducible methods. In addition, it forms the basis for future research in feature optimization, model explainability, and real-time deployment within cybersecurity contexts.

ACKNOWLEDGEMENT

We would like to extend our sincere gratitude to Shaqra University for their unwavering support throughout the research and preparation of this publication. The resources and academic environment provided by the university have played an integral role in shaping the outcome of this work. We are thankful for the opportunity to contribute to the scholarly community, and we recognize the invaluable contribution of Shaqra University in making this endeavor possible.

REFRENCES

[1] Lekkala, S. & Gurijala, P. Building Blocks of Network Security. in *Security and Privacy for Modern Networks: Strategies and Insights for Safeguarding Digital Infrastructures* (eds. Lekkala, S. & Gurijala, P.) **13–21** (Apress, Berkeley, CA, 2024). doi:10.1007/979-8-8688-0823-4_2.

[2] Ahmed, U. *et al.* Signature-based intrusion detection using machine learning and deep learning approaches empowered with fuzzy clustering. *Sci. Rep.* **15**, 1726 (2025).

[3] Ashiku, L. & Dagli, C. Network Intrusion Detection System using Deep Learning. *Procedia Comput. Sci.* **185**, 239–247 (2021).

[4] Qazi, E. U. H., Faheem, M. H. & Zia, T. HDLNIDS: Hybrid Deep-Learning-Based Network Intrusion Detection System. *Appl. Sci.* **13**, 4921 (2023).

[5] Shah, S. A. R. & Issac, B. Performance comparison of intrusion detection systems and application of machine learning to Snort system. *Future Gener. Comput. Syst.* **80**, 157–170 (2018).

[6] Gamage, S. & Samarabandu, J. Deep learning methods in network intrusion detection: A survey and an objective comparison. *J. Netw. Comput. Appl.* **169**, 102767 (2020).

[7] Hidayat, I., Ali, M. Z. & Arshad, A. Machine Learning-Based Intrusion Detection System: An Experimental Comparison. *J. Comput. Cogn. Eng.* **2**, 88–97 (2023).

[8] Kilincer, I. F., Ertam, F. & Sengur, A. Machine learning methods for cyber security intrusion detection: Datasets and comparative study. *Comput. Netw.* **188**, 107840 (2021).

[9] Zaib, M. H. NSL-KDD. https://www.kaggle.com/datasets/hassan06/nslkdd (2019).

[10] D'hooge, L. CIC-IDS2017. (2022).

[11] Setiawan, A., A. ,. Widodo, A. M. ,. Firmansyah, G. ,. Wisnujati. Network Intrusion Detection Using 1D Convolutional Neural Networks. *Proc. - ICE3IS 2024* (2024).

[12] Chen, H., You, G. & Shiue, Y. Application of an improved convolutional neural network-based method in network intrusion detection. in *2023 3rd International Conference on Intelligent Communications and Computing (ICC 2023)* (IEEE, 2023).

[13] Benaddi, H., Jouhari, M., Ibrahimi, K., Benslimane, A. & Amhoud, E. Improvement of Anomaly Detection System in the IoT Networks using CNN-LSTM Approach. in *GLOBECOM 2023-2023 IEEE Global Communications Conference* 3771–3776 (IEEE, 2023).

[14] Thandalam, C. & Anithaashri, T. Enhancing Network Security in SDNIoT Environments Through CNN-Based Attack Detection. in *International Conference on Self Sustainable Artificial Intelligence Systems, ICSSAS 2023 - Proceedings* (2023).
[15] Gaikwad, D. & Thool, R. Intrusion detection system using bagging with partial decision treebase classifier. *Procedia Comput. Sci.* **49**, 92–98 (2015).
[16] Mhawi, D., Aldallal, A. & Hassan, S. Advanced feature-selection-based hybrid ensemble learning algorithms for network intrusion detection systems. *Symmetry* **14**, 1461 (2022).
[17] Alsolami, T., Alsharif, B. & Ilyas, M. Enhancing cybersecurity in healthcare: Evaluating ensemble learning models for intrusion detection in the internet of medical things. *Sensors* **24**, 5937 (2024).
[18] Ibrahim, M. & Al-Wadi, A. Enhancing IoMT network security using ensemble learning-based intrusion detection systems. *J. Eng. Res.* (2024).
[19] Louk, M. & Tama, B. Revisiting gradient boosting-based approaches for learning imbalanced data: A case of anomaly detection on power grids. *Big Data Cogn. Comput.* **6**, 41 (2022).
[20] Louk, M. H. L. & Tama, B. A. Dual-IDS: A bagging-based gradient boosting decision tree model for network anomaly intrusion detection system. *Expert Syst. Appl.* **213**, 119030 (2023).
[21] Du, R., Li, R. & Zhang, Z. Ensemble Two Stage Machine Learning for Network Abnormal Detection. in *Proceedings of the 2023 15th International Conference on Machine Learning and Computing* 97–102 (2023).
[22] Yilmaz, M. & Bardak, B. An explainable anomaly detection benchmark of gradient boosting algorithms for network intrusion detection systems. in *2022 Innovations in Intelligent Systems and Applications Conference (ASYU)* 1–6 (IEEE, 2022).
[23] Belouch, M. & El Hadaj, S. Comparison of ensemble learning methods applied to network intrusion detection. in *ACM International Conference Proceeding Series* (2017).
[24] Tama, B. A. & Rhee, K.-H. Performance evaluation of intrusion detection system using classifier ensembles. *Int. J. Internet Protoc. Technol.* **10**, 23–31 (2017).
[25] Zhang, Y. & Wang, Z. Feature Engineering and Model Optimization Based Classification Method for Network Intrusion Detection. *Appl. Sci. Switz.* (2023).
[26] Rana, N., Alshehri, H., Abdali, M. A. & Madkhali, W. A. Optimized Intrusion Detection System for Attack Classification Using Machine Learning and Deep Learning Techniques. in *2024 5th International Conference on Intelligent Data Science Technologies and Applications (IDSTA 2024)* (2024).
[27] Azhagiri, M., Rajesh, A., Karthik, S. & Raja, K. An intrusion detection system using ranked feature bagging. *Int. J. Inf. Technol.* **16**, 1213–1219 (2024).
[28] Pham, N. T., Foo, E., Suriadi, S., Jeffrey, H. & Lahza, H. F. M. Improving performance of intrusion detection system using ensemble methods and feature selection. in *Proceedings of the Australasian computer science week multiconference* 1–6 (2018).
[29] Zhang, L. A Method for improving the stability of feature selection algorithm. in *Third international conference on natural computation (ICNC 2007)* vol. 1 715–717 (IEEE, 2007).
[30] Hou, Y., Teo, S. G., Chen, Z. & Truong-Huu, T. Handling Labeled Data Insufficiency: Semi-supervised Learning with Self-Training Mixup Decision Tree for Classification of Network Attacking Traffic. *IEEE Trans. Dependable Secure Comput.* (2022).
[31] Tripathi, K. & Das, S. Enhancing network security through machine learning: A comparative study of classification algorithms. *Adv. Electron. Comput. Phys. Chem. Sci.* (2025).
[32] Manzano, C., Meneses, C., Leger, P. & Fukuda, H. An Empirical Evaluation of Supervised Learning Methods for Network Malware Identification Based on Feature Selection. *Complexity* (2022).
[33] Bacevicius, M. & Paulauskaite-Taraseviciene, A. Machine Learning Algorithms for Raw and Unbalanced Intrusion Detection Data in a Multi-Class Classification Problem. *Appl. Sci.* **13**, 7328 (2023).
[34] Messaoud, M. CLASSIFICATION OF NETWORK TRAFFIC USING MACHINE LEARNING MODELS ON THE NETML DATASET. *Int. J. Comput. Netw. Commun.* (2025).
[35] McNair, A., Precious-Esue, D., Newson, S. & Rahimi, N. Enhancing IoT Network Defense: A Comparative Study of Machine Learning Algorithms for Attack Classification. in *Communications in Computer and Information Science* (2025).

[36] Krivchenkov, A., Misnevs, B. & Grakovski, A. Using Machine Learning for DoS Attacks Diagnostics. in *Lecture Notes in Networks and Systems* (2021).
[37] Bakhareva, N., Shukhman, A., Matveev, A. & Legashev, L. Attack Detection in Enterprise Networks by Machine Learning Methods. in *Proceedings - 2019 International Russian Automation Conference, RusAutoCon 2019* (2019).
[38] Manai, E., Mejri, M. & Fattahi, J. Confusion Matrix Explainability to Improve Model Performance: Application to Network Intrusion Detection. in *10th 2024 International Conference on Control, Decision and Information Technologies, CoDIT 2024* (2024).
[39] Hasanin, T., Khoshgoftaar, T. M. & Leevy, J. L. A comparison of performance metrics with severely imbalanced network security big data. in *Proceedings - 2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science, IRI 2019* (2019).
[40] Zuech, R., Hancock, J. & Khoshgoftaar, T. M. Detecting web attacks using random undersampling and ensemble learners. *J. Big Data* (2021).
[41] Kumar, V. & Sinha, D. Synthetic attack data generation model applying generative adversarial network for intrusion detection. *Comput. Secur.* (2023).
[42] Hakim, A. R., Ramli, K., Salman, M. & Agustina, E. R. Improving model performance for predicting exfiltration attacks through resampling strategies. *IIUM Eng. J.* (2025).
[43] Huang, Y., Gou, J., Fan, Z. & Zhuang, Y. A multi-label network attack detection approach based on two-stage model fusion. *J. Inf. Secur. Appl.* (2024).
[44] Rajput, D. S. & Upadhyay, A. K. Enhancing Network Security: An Ensemble of Machine Learning Model for Detection of Distributed Denial of Service Attack. in *Communications in Computer and Information Science* (2025).