Benha Journal of Engineering Science and Technology

BJEST (2025) Vol. 2 Issue (1) pp: 117-125
DOI
https://bjest.journals.ekb.eg/

Advances in Ensemble Machine Learning for Network Intrusion Detection Systems: A Comprehensive Review

Mohmed A. Salama [a*], Radwa M. Tawfeek [a], Sara Hamdy [a], and Omar M. Salim [a]

[a]Electrical Engineering Department, Benha Faculty of Engineering, Benha University, Benha, Egypt
*Corresponding Author E-mail: mohamed.salama@bhit.bu.edu.eg

ABSTRACT

As cyber threats grow increasingly sophisticated, robust network security demands adaptive intrusion detection systems (IDS). Traditional machine learning-based IDS often struggle with high false alarm rates and poor generalization to emerging attacks, while deep learning-based IDS offer high detection accuracy but require significant computational resources. Ensemble learning techniques provide an effective balance between efficiency and accuracy, improving detection through model diversity and decision aggregation. This review explores ensemble-based intrusion detection systems, emphasizing diverse aggregation techniques, including homogeneous and heterogeneous ensemble methods. It provides an in-depth analysis of feature selection strategies, data balancing techniques, and classification models, offering a comparative assessment across benchmark datasets. Additionally, the study highlights key challenges and outlines future research directions to advance ensemble learning in network intrusion detection.

KEYWORDS

Ensemble Learning, NIDS, Feature Selection, Machine Learning, Data Imbalance.

1. INTRODUCTION

With the rapid expansion of digital infrastructures, cybercrimes have escalated in both frequency and sophistication, posing significant threats to individuals, organizations, and governments. Malicious actors continuously develop advanced attack strategies, including malware propagation, distributed denial-of-service (DDoS) attacks, phishing schemes, and zero-day exploits, leading to financial losses and compromised data integrity. To counteract these threats, cybersecurity solutions must evolve to ensure robust network protection, with intrusion detection systems (IDS) playing a critical role in identifying and mitigating cyber threats before substantial damage occurs [1, 2].

Among IDS solutions, network intrusion detection systems (NIDS) have emerged as a fundamental component in network security, enabling the detection of anomalous activities by monitoring network traffic in real time. NIDS are categorized into signature-based and anomaly-based detection systems. Signature-based NIDS rely on predefined attack patterns to detect known threats but struggle with identifying novel and zero-day attacks. In contrast, anomaly-based NIDS utilize statistical models and machine learning algorithms to detect deviations from normal behavior, enhancing their ability to identify previously unseen threats[3]. However, despite their advantages, anomaly-based systems face challenges such as high false positive rates and complexity[4].

The integration of artificial intelligence (AI) and machine learning (ML) techniques into NIDS has significantly improved their detection capabilities. Machine learning models, including support vector machines (SVM), decision trees (DT), k-nearest neighbors (kNN), and random forests (RF), have been widely adopted to classify network traffic efficiently. These models enhance detection accuracy, automate feature extraction, and reduce human intervention in cybersecurity systems [5]. However, NIDS based on ML algorithms face challenges such as data imbalance, high computational costs, and poor adaptability to evolving attack strategies, limiting their real-world deployment in dynamic network environments [6].

Traditional ML-based NIDS exhibit several limitations that hinder their efficiency in detecting sophisticated cyber threats. Many models rely on manually engineered features, making them susceptible to adversarial evasion techniques and requiring frequent updates to maintain effectiveness. Additionally, ML models often suffer from high false positive rates and difficulty in distinguishing between benign anomalies and actual intrusions, leading to alert fatigue and reduced operational efficiency [7].

Deep learning (DL) has emerged as a promising alternative for intrusion detection, leveraging architectures such as convolutional neural networks (CNN), recurrent neural networks (RNN), and long short-term memory (LSTM) networks. These models excel at automatic feature extraction and learning hierarchical representations of network traffic, enabling them to detect complex attack patterns with higher accuracy [8]. However, deep learning-based NIDS present notable challenges, including high computational demands, increased training time, and vulnerability to adversarial attacks. Furthermore, their deployment in real-time intrusion detection remains constrained by hardware limitations and the need for large, labelled datasets.

To address these challenges, ensemble learning techniques have been introduced as a more efficient and robust solution for NIDS. Ensemble classifiers combine multiple weak learners to improve detection accuracy, enhance model generalization, and reduce false positive rates. Methods such as bagging, boosting, and stacking have been widely explored, offering a balance between computational efficiency and detection performance. Compared to traditional ML models, ensemble classifiers provide superior adaptability to evolving threats by leveraging diverse decision-making processes. Unlike deep learning models, ensemble approaches require less computational power and storage, making them more suitable for real-time intrusion detection in resource-constrained environments [9].

This review aims to analyze and compare ensemble-based NIDS, highlighting their advancements, advantages, and potential research directions in securing modern network infrastructures. The remaining sections of the article are structured as follows: Section 2 explain background related NIDS based AI such benchmark datasets and evaluation matrices. Section 3 discusses the most recent studies of NIDS that rely on ensemble machine learning in their structure.

Finally, Section 4 provides conclusion and future approaches of NIDS research.

2. BACKGROUND

To train and evaluate these systems, researchers rely on benchmark datasets that reflect real-world network activities. Additionally, the use of ensemble learning techniques has emerged as a promising approach to enhance the accuracy and reliability of intrusion detection models. This section provides an overview of popular intrusion detection datasets, discusses key ensemble learning methodologies, and outlines commonly used evaluation metrics for assessing NIDS performance.

2.1. Intrusion Detection Datasets

Intrusion detection systems (IDS) use benchmark datasets to train and evaluate their performance in detecting cyber threats. These datasets simulate real-world network traffic and contain labelled instances of normal and malicious activities. Effective IDS models require comprehensive datasets that cover diverse attack scenarios, realistic traffic patterns, and class balance. However, many datasets suffer from issues such as imbalanced attack distributions, redundant records, and lack of real-time traffic characteristics, which can hinder their effectiveness in practical deployments

2.1.1 NSL-KDD Dataset

The NSL-KDD dataset as depicted in **Figure *1*** is an improved version of the KDDCup99 dataset, designed to overcome issues related to redundant and duplicated records that affected classifier performance. It consists of 125,973 training instances and 22,544 test instances, covering four main attack categories: Denial-of-Service (DoS), Probe, Remote-to-Local (R2L), and User-to-Root (U2R) attacks. While it reduces redundancy and ensures a more balanced class distribution, NSL-KDD remains outdated and lacks representation of modern network threats

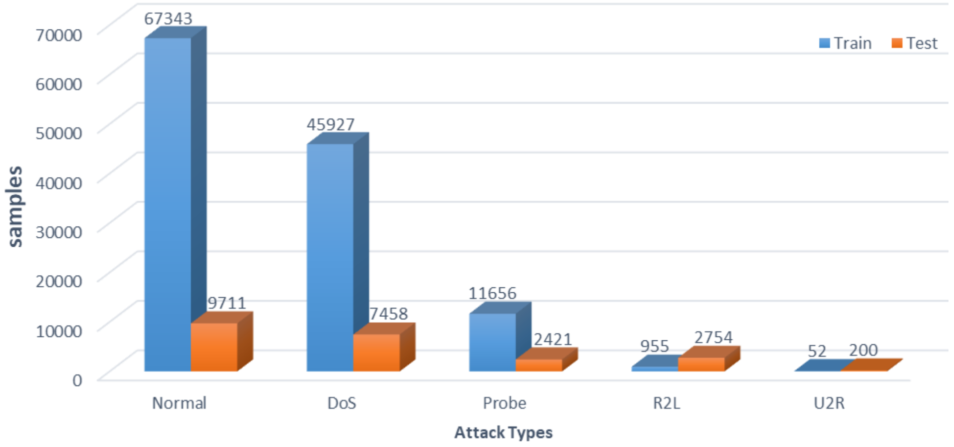

Figure 1 Distribution of NSL-KDD dataset

2.1.2 UNSW-NB15 Dataset

To address the limitations of older datasets, UNSW-NB15 [10] was introduced, providing a more realistic representation of contemporary network traffic. It consists of 82,332 records, with nine attack categories as listed in **Table** ***1***, including Backdoors, Exploits, Fuzzers, DoS, and Reconnaissance attacks. UNSW-NB15 integrates modern network features such as flow-based characteristics and hybrid attack simulations, making it a more comprehensive benchmark for evaluating IDS models. However, its class imbalance poses challenges for training machine learning models, often requiring data balancing techniques.

2.1.3 CICIDS2017 Dataset

Developed by the Canadian Institute for Cybersecurity, CICIDS2017 is a large-scale dataset that captures realistic network traffic, including both normal and attack scenarios. It contains 2,273,097 samples, with 15 attack types, including Botnets, DDoS, Brute Force, and Web Attacks. The dataset provides detailed flow-based features, making it highly suitable for deep learning-based intrusion detection. However, the high dimensionality and computational requirements of CICIDS2017 present challenges for real-time deployment.

Table 1 Distribution of UNSW-NB 15 dataset

Class	Training set	Weight (%)	Testing set	Weight (%)
Normal	56,000	31.94%	37,000	44.94%
Generic	40,000	22.81%	18,871	22.92%
Exploits	33,393	19.04%	11,132	13.52%
Fuzzers	18,184	10.37%	6,062	7.36%
DoS	12,264	6.99%	4,089	4.97%
Reconnaissance	10,491	5.98%	3,496	4.25%
Analysis	2,000	1.14%	677	0.82%
Backdoors	1,746	0.99%	583	0.71%
shellcode	1,133	0.65%	378	0.46%
Worms	130	0.07%	44	0.05%

2.2 *Ensemble Techniques in Intrusion Detection*

They enhance IDS performance by combining multiple base classifiers to improve accuracy, robustness, and generalization. Unlike single classifiers, ensemble methods reduce variance and mitigate model biases, making them well-suited for handling the complexities of network intrusion detection. The three primary ensemble techniques used in IDS are bagging, boosting, and stacking.

2.2.1 Bagging

Bagging (Bootstrap Aggregating) is an ensemble method that trains multiple instances of the same base classifier on different random subsets of the training data and averages their predictions to improve stability and accuracy [11]. A widely used bagging-based classifier in NIDS is the Random Forest (RF) algorithm, which constructs multiple decision trees and aggregates their predictions. The final classification is determined by majority voting where represents individual decision trees and is the number of trees. RF offers high detection accuracy, robustness against overfitting, and interpretability, making it a preferred choice for intrusion detection tasks [5].

$$P(y) = \frac{1}{n}\sum_{i=1}^{n} h_i(x) \quad (1)$$

Where $h_i(x)$ represents individual decision trees, and n is the number of trees.

2.2.2 Stacking

Stacking combines multiple base classifiers using a meta-classifier, allowing diverse models to contribute to final predictions. The meta-classifier, trained on the outputs of base learners, refines final predictions by learning optimal combinations. Unlike bagging and boosting, stacking enables heterogeneous models such as SVM, Decision Trees, to leverage complementary strengths, enhancing detection performance and generalization.

$$P(y) = g\big(h_i(x), h_2(x), \dots, h_n(x)\big) \quad (2)$$

Where $h_i(x)$ are individual classifiers and g is the meta-classifier

2.3 Evaluation Metrics for NIDS

To assess IDS performance, various evaluation metrics are employed, ensuring reliable detection of cyber threats. The most widely used metrics are included in Table 2

Table 2 Evaluation Metric for IDS Classification Task

Evaluation Metric	Definition	Mathematical equation
Accuracy (ACC)	Measures overall correctness of predictions	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision (P)	Evaluates how many predicted intrusions are actual attacks	$\frac{TP}{TP+FP}$
Recall (R)	Measures the proportion of actual attacks correctly identified	$\frac{TP}{TP+FN}$
F1-Score (F1)	Harmonic mean of precision and recall, balancing false positives and false negatives	$2 * \frac{P * R}{P+R}$
False alarm rate (FAR)	Indicates the proportion of benign instances misclassified as intrusions	$\frac{FP}{TN+FP}$

3. LITERATURE SURVEY

This section reviews recent studies on ensemble-based intrusion detection systems, focusing on their architectural designs, datasets, and evaluation methodologies. It highlights how various ensemble techniques address common challenges such as data imbalance, feature selection, and real-time detection. The analysis offers insights into current trends and emerging directions in the field.

Thockchom et al. [12] proposed a stacking ensemble-based intrusion detection system integrating Gaussian Naïve Bayes, Decision Tree, and Logistic Regression, with Stochastic Gradient Descent as the meta-classifier. The model, evaluated on KDD Cup 1999, UNSW-NB15, and CIC-IDS2017 datasets, demonstrated superior accuracy and reduced false positives compared to individual classifiers. Preprocessing operations integrated with feature selection using the Chi-square test enhanced classification performance, particularly in handling class imbalance. Compared to existing ensemble approaches, the proposed model achieved higher detection rates across multiple attack categories.

Dual-IDS [13], a bagging-based gradient boosting ensemble that integrates GBM, LightGBM, CatBoost, and XGBoost to enhance intrusion detection .NSL-KDD, UNSW-NB15, and HIKARI-2021 datasets are used, the proposed system achieved 94.66% accuracy, 92.94% recall, and a significantly reduced false positive rate of 1.3%, outperforming traditional ensemble methods. The results demonstrate the effectiveness of hybrid boosting and bagging techniques in improving detection robustness and minimizing false alarms.

Authors in [14] introduced an ensemble-based IDS using Balanced Bagging, XGBoost, and Random Forest with Hellinger distance to enhance detection accuracy on the UNSW-NB15 dataset. The model applies Elastic Net and Sequential Feature Selection for dimensionality reduction and utilizes majority voting for final predictions. Results show improved classification performance, particularly in detecting minority attack classes, outperforming traditional methods in both binary and multiclass classification.

Another study [15] suggested an intrusion detection system utilizing Bagging with Partial Decision Trees, enhanced by Genetic Algorithm for optimal feature selection. The model demonstrated improved classification accuracy and reduced false positives compared to J48, Random Forest, and Naïve Bayes, evaluated on the KDD Cup 1999 dataset. The integration of Bagging enhanced model robustness, particularly in detecting minority attack classes. Their findings suggest that feature selection via Genetic Algorithm significantly refines IDS performance.

In the study [16], the authors introduced an ensemble intrusion detection system based on M-AdaBoost-A, integrating the AUC metric to enhance classification in imbalanced datasets. Two variants were proposed: M-AdaBoost-A-SMV, employing majority voting for efficiency, and M-AdaBoost-A-PSO, leveraging Particle Swarm Optimization for optimal classifier weighting. The models demonstrated superior detection rates, particularly for minority attack classes, outperforming SAMME and AdaBoost.M1 on AWID and NSL-KDD datasets.

DIS-IoT, a deep learning-based stacking ensemble for IoT intrusion detection, integrating MLP, DNN, CNN, and LSTM models with Dempster–Shafer theory for decision fusion is proposed in [17]. ToN_IoT, CICIDS2017, and SWaT datasets used for training the suggested architecture, the model outperformed traditional deep learning and ensemble methods, achieving higher accuracy and lower false positive rates. The results highlight the effectiveness of deep feature extraction and uncertainty-aware decision fusion in enhancing network security.

Louk and Tama [18] combined a PSO-driven feature selection and hybrid ensemble model combining Bagging and Gradient Boosting Machine (GBM) for anomaly-based intrusion detection. Evaluated on NSL-KDD, UNSW-NB15, and CICIDS-2017 datasets, the model achieved 99.98% accuracy, 99.99% recall, and an AUC of 1.00, outperforming traditional ensemble methods. By integrating PSO for feature selection and majority voting across multiple GBM models, the approach significantly reduces false positives while improving classification robustness

Moving to a Double-Layered Hybrid Approach (DLHA) [19], integrating Naïve Bayes for high-frequency attacks and SVM for rare attack detection, optimized using PCA-based feature selection. the model achieved 96.67% detection for R2L and 100% for U2R, outperforming traditional ensemble methods. By leveraging specialized classifiers per attack category, DLHA significantly improves anomaly detection while maintaining computational efficiency.

A two-stage intrusion detection system (IDS) for Software-Defined IoT (SD-IoT) has proposed in [20], integrating an improved Firefly Algorithm with differential evolution for feature selection and a weighted voting ensemble of C4.5, MLP, and IBL for classification. Trained on NSL-KDD and UNSW-NB15 datasets, the model achieved 99.00% accuracy and 0.81% FPR on NSL-KDD, and 88.46% accuracy on UNSW-NB15, outperforming traditional ensemble approaches. The results highlight the effectiveness of hybrid optimization and ensemble learning in reducing false positives while improving detection rates.

Authors in [21] suggested an intrusion detection system (IDS) integrating Adaptive Synthetic (ADASYN) oversampling with LightGBM to enhance classification performance and computational efficiency. NSL-KDD, UNSW-NB15, and CICIDS2017 datasets have been used for evaluation, the architecture achieved 99.91% accuracy and 99.87% F1-score on CICIDS2017, outperforming traditional ensemble methods. By leveraging gradient-based optimization and adaptive resampling, the approach significantly improves minority attack detection while maintaining low false positive rates.

Ahmed et al. [22] integrated a two-stage intrusion detection system (IDS) integrating Auto-Encoder (AE) for feature selection and LSTM for sequential attack classification. the model achieved 89.0% accuracy, 88.0% detection rate, and an F1-score of 0.91 on NSL-KDD, outperforming conventional machine learning and deep learning-based IDS approaches. By leveraging dimensionality reduction and temporal pattern recognition, the proposed AE-LSTM framework enhances anomaly detection while maintaining a low false alarm rate (11.0%).

6. CONCLUSION AND FUTURE DIRECTIONS

This paper presents a comprehensive review of ensemble machine learning techniques for network intrusion detection systems (NIDS), highlighting their strengths in overcoming the limitations of traditional machine learning and deep learning approaches. By examining various ensemble techniques such as bagging, boosting, and stacking. The study emphasizes their ability to enhance detection accuracy, reduce false positives, and improve adaptability to evolving cyber threats. It also provides a comparative analysis of feature selection strategies, data balancing techniques, and classifier performance across benchmark datasets. Moving forward, future research should focus on optimizing real-time detection by developing lightweight, low-latency ensemble models suitable for high-speed networks. In addition, Federated ensemble learning presents a promising direction by enabling collaborative intrusion detection across distributed environments while preserving data privacy. Another crucial research point regarding enhancing IDS based on AI, automated model tuning using AutoML can streamline the selection of optimal hyperparameters, reducing the dependency on manual configuration. Finally, the integration of explainable AI techniques will be crucial for enhancing the interpretability of ensemble decisions, thereby increasing transparency and fostering trust in intrusion detection outcomes.

REFERENCES

[1] Cebula, J.J. and L.R. Young, A taxonomy of operational cyber security risks. Software Engineering Institute, 2010.

[2] Thompson, E.C., Cybersecurity incident response: How to contain, eradicate, and recover from incidents. 2018: Apress.

[3] Uikey, R. and M. Gyanchandani. Survey on classification techniques applied to intrusion detection system and its comparative analysis. in 2019 International Conference on Communication and Electronics Systems (ICCES). 2019. IEEE.

[4] Ahmad, Z., et al., Network intrusion detection system: A systematic study of machine learning and deep learning approaches. Transactions on Emerging Telecommunications Technologies, 2021. 32(1): p. e4150.

[5] Chebrolu, S., A. Abraham, and J.P. Thomas, Feature deduction and ensemble design of intrusion detection systems. Computers & security, 2005. 24(4): p. 295-307.

[6] Kumar, V., et al., An integrated rule based intrusion detection system: analysis on UNSW-NB15 data set and the real time online dataset. Cluster Computing, 2020. 23: p. 1397-1418.

[7] Tama, B.A. and K.-H. Rhee, An in-depth experimental study of anomaly detection using gradient boosted machine. Neural Computing and Applications, 2019. 31: p. 955-965.

[8] Shone, N., et al., A deep learning approach to network intrusion detection. IEEE transactions on emerging topics in computational intelligence, 2018. 2(1): p. 41-50.

[9] Barnard, P., N. Marchetti, and L.A. DaSilva, Robust network intrusion detection through explainable artificial intelligence (XAI). IEEE Networking Letters, 2022. 4(3): p. 167-171.

[10] Moustafa, N. and J. Slay. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). in 2015 military communications and information systems conference

(MilCIS). 2015. IEEE.

[11] Breiman, L., Bagging predictors. Machine learning, 1996. 24(2): p. 123-140.

[12] Thockchom, N., M.M. Singh, and U. Nandi, A novel ensemble learning-based model for network intrusion detection. Complex & Intelligent Systems, 2023. 9(5): p. 5693-5714.

[13] Louk, M.H.L. and B.A. Tama, Dual-IDS: A bagging-based gradient boosting decision tree model for network anomaly intrusion detection system. Expert Systems with Applications, 2023. 213: p. 119030.

[14] Das, A. and B. Sunitha, Anomaly-based network intrusion detection using ensemble machine learning approach. International Journal of Advanced Computer Science and Applications, 2022. 13(2).

[15] Gaikwad, D. and R.C. Thool, Intrusion detection system using bagging with partial decision treebase classifier. Procedia Computer Science, 2015. 49: p. 92-98.

[16] Zhou, Y., T.A. Mazzuchi, and S. Sarkani, M-AdaBoost-A based ensemble system for network intrusion detection. Expert Systems with Applications, 2020. 162: p. 113864.

[17] Lazzarini, R., H. Tianfield, and V. Charissis, A stacking ensemble of deep learning models for IoT intrusion detection. Knowledge-Based Systems, 2023. 279: p. 110941.

[18] Louk, M.H.L. and B.A. Tama, PSO-driven feature selection and hybrid ensemble for network anomaly detection. Big Data and Cognitive Computing, 2022. 6(4): p. 137.

[19] Wisanwanichthan, T. and M. Thammawichai, A double-layered hybrid approach for network intrusion detection system using combined naive bayes and SVM. Ieee Access, 2021. 9: p. 138432-138450.

[20] Tian, Q., et al., A two-stage intrusion detection approach for software-defined IoT networks. Soft Computing, 2021. 25: p. 10935-10951.

[21] Liu, J., Y. Gao, and F. Hu, A fast network intrusion detection system using adaptive synthetic oversampling and LightGBM. Computers & Security, 2021. 106: p. 102289.

[22] Mushtaq, E., et al., A two-stage intrusion detection system with auto-encoder and LSTMs. Applied Soft Computing, 2022.