International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

A Hybrid Intrusion Detection System Using Machine Learning

Sneha Gawari[1], Lavanya Indulkar[2],Sharvari Moon[3] , Rushikesh Patil[4] , Prof. Anil Londhe[5]

Department of Artificial Intelligence and Data Science, Datta Meghe College of Engineering
University of Mumbai

Abstract - As cyber adversaries use more complex tactics, traditional defensive measures often fall short. This research presents a Hybrid Intrusion Detection System (HIDS) that aims to connect static signature matching with dynamic anomaly detection. By using the XGBoost algorithm, the proposed framework offers a fast detection engine that balances speed with accuracy. The model's strength was confirmed through training on two datasets, the CICIDS2017 and NSL-KDD benchmarks, ensuring it works with both old and new threat types. To put theory into practice, the system includes Wireshark-based packet analysis and is controlled through an easy-to-use Django web dashboard for real-time monitoring. While the system shows strong protection against known attacks, its limited ability to identify zero-day vulnerabilities hints at possibilities for future improvement. In the end, this HIDS offers a scalable, flexible solution designed for the needs of today's network monitoring.

***Key Words*: Intrusion Detection System, Hybrid IDS, Machine Learning, Cybersecurity, Anomaly Detection, Network Security, Cyber Attacks.**

1. INTRODUCTION

The digital landscape has undergone a radical transformation, with modern organizational operations now almost entirely dependent on interconnected infrastructure. While this transition has streamlined data management and global communication, it has simultaneously expanded the "attack surface" available to malicious actors. Standard defensive perimeters, such as traditional firewalls and signature-reliant antivirus software, are increasingly outmatched by complex threats like Distributed Denial of Service (DDoS), advanced malware injections, and sophisticated social engineering. Consequently, there is an urgent need for robust Intrusion Detection Systems (IDS) that act as a secondary, more intelligent layer of network surveillance.

Current detection methodologies are generally split into two schools: signature-based and anomaly-based detection. The former is highly effective at identifying documented threats with surgical precision but remains "blind" to novel exploits. The latter attempts to solve this by modelling baseline network behaviour to identify outliers, which allows for the detection of zero-day vulnerabilities. However, the practical utility of anomaly-based systems is often hindered by a high frequency of false alarms, which can lead to "alert fatigue" among security professionals.

This research addresses these gaps by implementing a Hybrid Intrusion Detection System (HIDS). By synthesizing the reliability of signature matching with the predictive power of machine learning—specifically the XGBoost (Extreme Gradient Boosting) algorithm—this system aims to maximize detection rates while suppressing false positives. The model is developed and validated using a dual-dataset approach, utilizing both the standardized NSL-KDD and the modern CICIDS2017 datasets, supplemented by live traffic analysis via Wireshark. To bridge the gap between algorithmic theory and operational utility, the system is wrapped in a Django-based web dashboard, providing administrators with an intuitive, real-time visualization of their network's security posture.

2. OBJECTIVES AND SCOPE

The primary objective of this research is to design and develop a Hybrid Intrusion Detection System (HIDS) that effectively identifies malicious network activities by combining signature-based detection with machine learning-based anomaly detection. The system aims to leverage the strengths of both approaches to achieve higher detection accuracy while minimizing false positives. A key objective is to implement the XGBoost algorithm for efficient and scalable classification of network traffic using benchmark datasets such as CICIDS2017 and NSL-KDD. Additionally, the system seeks to integrate real-time packet capture using Wireshark, enabling practical deployment and live monitoring of network activity. Another important goal is to develop a user-friendly interface using Django to visualize predictions and provide actionable insights.

The scope of this work includes data preprocessing, feature extraction, model training, and real-time detection of network intrusions. It focuses on identifying common attack types such as DDoS, brute force, and botnet attacks. However, the system is limited to supervised learning techniques and may require further enhancement to detect zero-day attacks effectively. Future extensions may include deep learning models, automated response systems, and large-scale deployment in enterprise environments.

Overall, the proposed system demonstrates improved detection capability, reduced false alarms, scalability, and real-time applicability, making it a more robust and efficient solution for modern network security challenges.

3. LITERATURE SURVEY

Table I: Comparative Analysis of Existing Work

Author(s) & Year	Methodology Used	Dataset(s)	Key Findings
Rababah & Srivastava (2019)	Decision Tree + Random Forest (Stacking Ensemble)	CICIDS2017, NSL-KDD	Achieved 98% accuracy; ensemble methods improve detection performance [1]
Sajid et al. (2024)	XGBoost + CNN + LSTM	CICIDS2017, UNSW-NB15, NSL-KDD, WSN-DS	Reduced false positives; hybrid DL + ML improves feature extraction [2]
Gümüşbaş et al. (2020)	Survey of DL models	Multiple datasets	DL models effective for feature extraction and dimensionality reduction [3]
Talukder et al. (2024)	ML + Random Oversampling	UNSW-NB15, CICIDS2017, CICIDS2018	Achieved >99% accuracy; class balancing improves performance [4]
Ashiku & Dagli (2021)	Traditional IDS Analysis	—	Identified limitations: high false positives, poor zero-day detection [5]
Vinayakumar et al. (2019)	ML vs Deep Learning	Multiple datasets	DNNs outperform ML; scalable with big data frameworks [6]

Recent advancements in Intrusion Detection Systems (IDS) have focused on integrating machine learning and deep learning techniques to overcome the limitations of traditional approaches. Conventional signature-based and anomaly-based systems suffer from issues such as inability to detect zero-day attacks and high false positive rates [5]. To address these challenges, researchers have explored ensemble learning and hybrid models.

Rababah and Srivastava [1] demonstrated that stacking ensemble techniques combining Decision Tree and Random Forest significantly improve detection accuracy. Similarly, Sajid et al. [2] proposed a hybrid approach integrating XGBoost with deep learning models such as CNN and LSTM, achieving improved feature extraction and reduced false positives. Deep learning-based methods, as discussed by Gümüşbaş et al. [3], are particularly effective in handling high-dimensional data and complex intrusion patterns.

Another important aspect is handling class imbalance in datasets. Talukder et al. [4] showed that oversampling techniques enhance model performance significantly. Additionally, Vinayakumar et al. [6] highlighted the superiority of deep learning models over traditional machine learning approaches in terms of scalability and detection efficiency.

4. PROPOSED METHODOLOGY

A. System Overview

The proposed Hybrid Intrusion Detection System (HIDS) is designed to enhance network security by integrating machine learning-based anomaly detection with real-time traffic monitoring. The system aims to detect both known and unknown cyber threats by combining the strengths of signature-based detection and data-driven learning techniques. The overall workflow begins with the collection of network traffic data, followed by preprocessing, feature extraction, model training, and real-time intrusion detection.

The system utilizes benchmark datasets such as CICIDS2017 and NSL-KDD for training and evaluation, ensuring exposure to a wide range of attack scenarios including Distributed Denial of Service (DDoS), brute force attacks, and botnet activities. The trained model is then deployed in a real-time environment where network packets are captured using Wireshark. These packets are processed to extract relevant features, which are then fed into the machine learning model for classification. The core detection mechanism is based on the XGBoost algorithm, which provides high accuracy and efficiency in handling large-scale and high-dimensional data. The system classifies incoming traffic into categories such as normal or malicious, enabling timely detection of intrusions. Additionally, a Django-based web interface is integrated to provide real-time visualization of network activity, alerts, and predictions. This enhances usability and allows users to monitor and respond to threats effectively. Overall, the system provides a scalable, efficient, and practical solution for modern intrusion detection.

B. System Architecture and Core Modules

The proposed Hybrid Intrusion Detection System uses a multi-layered design, with connected modules working together to detect and analyze network intrusions. It is split into a few main parts: data acquisition, preprocessing, feature extraction, the detection engine, and the visualization layer.

The data acquisition part handles both offline and real-time data, which is honestly necessary if the system is supposed to work beyond theory. Offline data comes from benchmark datasets like CICIDS2017 and NSL-KDD, while live traffic is captured through Wireshark. This mixed setup gives the system broader training data and helps it perform properly in actual network environments.

The **preprocessing and feature extraction module** prepares the raw data for analysis by cleaning, normalizing, and transforming it into a structured format. Irrelevant or redundant features are removed, and important attributes such as packet

size, protocol type, and flow duration are selected for model training.

The **detection engine** is the core component of the system and consists of two sub-modules: signature-based detection and machine learning-based anomaly detection. The signature-based module identifies known attack patterns, while the anomaly detection module uses the XGBOOST classifier to detect unusual behaviour in network traffic.

Finally, the **visualization and interface module**, implemented using Django, displays the detection results in real time. It provides alerts, logs, and insights into network activity, enabling users to take appropriate actions. This modular architecture ensures flexibility, scalability, and efficient intrusion detection.

C. Dataset Overview and Pre-processing

How well the proposed Hybrid Intrusion Detection System works really depends on one thing first: the datasets. If the training and evaluation data are weak, the whole model becomes unreliable. In this study, two benchmark datasets are used, CICIDS2017 and NSL-KDD. CICIDS2017 is especially useful because it reflects more realistic and recent network traffic, not outdated patterns. It includes current attack categories like DDoS, brute force, infiltration, and botnet activity, which makes it much more convincing for this kind of task. Meanwhile, NSL-KDD is still important as a standard benchmark, mainly because it fixes some of the redundancy and class imbalance problems found in older datasets.

Before the data is given to the machine learning model, a fairly serious preprocessing stage is carried out. This part matters more than people sometimes admit. Missing or null values are handled, duplicate records are removed, and inconsistent entries are corrected so the dataset is cleaner and more stable. Also, categorical features such as protocol type and service cannot stay in text form, so they are converted into numerical values through suitable encoding methods.

After that, feature scaling is applied through normalization or standardization so no variable dominates just because of its range. Feature selection is also used to keep the most useful attributes, which cuts down dimensionality and saves computational cost. In some situations, resampling is needed to deal with class imbalance. These steps strongly improve the XGBoost model and help it detect intrusions with better accuracy and efficiency.

D. Model Training and Evaluation

The model training and evaluation stage is a key part of the proposed Hybrid Intrusion Detection System because this is where the actual quality of the model becomes clear. If the classifier cannot separate normal traffic from malicious traffic well, then the whole system loses much of its value. In this study, XGBoost (Extreme Gradient Boosting) is used as the main classifier. That choice makes sense, mainly because XGBoost is fast, scalable, and usually performs strongly on difficult, high-dimensional data.

The pre-processed dataset, built from CICIDS2017 and NSL-KDD, is split into training and testing sets with an approximate 80:20 ratio. The training portion is used to develop the XGBoost model by learning patterns from both normal and attack traffic. During this phase, several hyperparameters, including learning rate, maximum tree depth, and number of estimators, are chosen through initial experimentation. Full hyperparameter tuning is not carried out, which is honestly a limitation, yet the selected setup still gives a reasonable balance between model complexity and computational cost.

After training, the model is tested on the unseen dataset to check how well it generalizes. Evaluation is done using accuracy, precision, recall, and F1-score. Accuracy shows the overall correctness of predictions, while precision and recall are more useful when looking closely at false positives and false negatives. The F1-score matters here because it combines precision and recall into one measure, which gives a more balanced picture.

The results show an accuracy of about 90%, and that is a strong outcome for this kind of task. The model detects known attacks quite effectively. Its weakness appears with unseen or zero-day attacks, where improvement is still needed. Overall, the XGBoost-based method looks reliable and efficient, though stronger tuning and larger datasets would likely push the performance further.

5. RESULTS AND DISCUSSION

The performance of the proposed Hybrid Intrusion Detection System (HIDS) was evaluated using benchmark datasets, namely CICIDS2017 and NSL-KDD, along with real-time packet analysis. The XGBoost-based model demonstrated an overall accuracy of approximately 90%, indicating its effectiveness in classifying network traffic into normal and malicious categories. The evaluation was carried out using key performance metrics such as accuracy, precision, recall, and F1-score, providing a comprehensive assessment of the model's predictive capability.

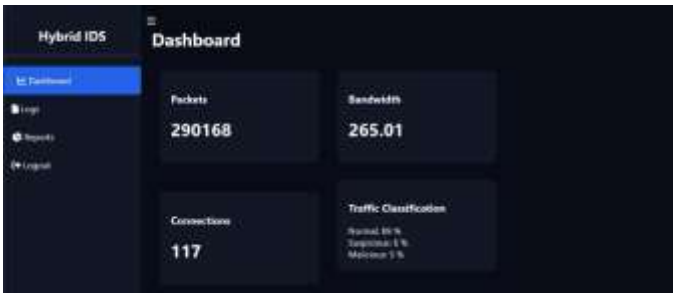

Fig. 1: **Hybrid IDS Main Dashboard**

Figure (1) shows the main dashboard of the Hybrid Intrusion Detection System, highlighting key system metrics such as total packets processed, bandwidth usage, and active connections. The dashboard also displays the percentage distribution of traffic classifications, including normal, suspicious, and malicious traffic. This high-level overview allows administrators to quickly assess network status and detect anomalies. The intuitive layout and real-time updates improve usability and make the system suitable for continuous monitoring in practical environments.

Fig. 2: **Network Logs Interface**

Figure (2) displays the Network Logs interface, which provides detailed, real-time records of captured network traffic. Each log entry includes information such as timestamp, source IP address, destination IP address, protocol type, packet size, and predicted classification (normal, suspicious, or malicious). The interface also includes filtering and search functionalities, enabling users to analyze specific traffic patterns efficiently. Summary counters for total logs, malicious, suspicious, and normal traffic are presented for quick reference. This module plays a critical role in forensic analysis and helps users trace and investigate suspicious activities within the network.

Fig. 3: **Intrusion Detection System Reports Dashboard**

Figure (3) presents the IDS Reports Dashboard, which provides a comprehensive overview of network activity and intrusion detection results. The dashboard includes visualizations such as attack distribution, traffic trends over time, and protocol usage. The pie chart illustrates the proportion of normal, suspicious, and malicious traffic, enabling quick assessment of network health. The traffic graph shows fluctuations in network activity, which can help identify unusual spikes indicative of potential attacks. Additionally, summary statistics provide a consolidated view of total packets analyzed and their classifications. This interface enhances situational awareness and allows users to monitor network security in real time.

6. CONCLUSION

This paper presents a Hybrid Intrusion Detection System that mixes signature-based detection with machine learning anomaly detection through the XGBoost algorithm. The main goal is clear: deal with the weak points of traditional IDS models by pushing detection accuracy higher and cutting down false positives, which is often where many systems fail in practice. Using benchmark datasets like CICIDS2017 and NSL-KDD, the model is trained on a wide range of attack situations, and that matters because network threats do not appear in just one form.

What makes the system more useful, honestly, is the addition of real-time packet capture through Wireshark. That part gives it practical value, since live monitoring is much closer to real network conditions than static testing alone. Along with that, the Django-based web interface makes the system easier to use. It lets users view predictions and examine network activity without dealing with a complicated setup, and that is a strong point.

The experimental results show an accuracy of about 90%, which is solid. The system performs well in detecting known attacks, and it also shows a fair ability to catch unusual behaviour. Even then, some issues are still obvious, especially zero-day attack detection, dataset imbalance, and the limited level of hyperparameter tuning.

Overall, this Hybrid IDS is a scalable and efficient answer to current network security problems, though it still needs deeper improvement before it can be called fully reliable.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the faculty members and project guide of the Department of Artificial Intelligence and Data Science for their continuous support, guidance, and valuable suggestions throughout the development of this project. Their expertise and encouragement played a crucial role in shaping the direction and successful completion of this research work.

The authors also extend their appreciation to the institution for providing the necessary resources and infrastructure required for conducting this study. Special thanks are given to the developers and contributors of open-source tools and datasets, particularly CICIDS2017 and NSL-KDD, which served as the foundation for training and evaluating the proposed model.

Finally, the authors would like to acknowledge their peers and team members for their collaboration, constructive feedback, and collective efforts, which significantly contributed to the successful implementation of the Hybrid Intrusion Detection System.

REFERENCES

[1] B. Rababah and S. Srivastava, “Hybrid Model for Intrusion Detection Systems Using Decision Tree and Random Forest,” *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 10, no. 5, pp. 1–8, 2019.

[2] M. Sajid, K. R. Malik, A. Almogren, T. S. Malik, A. H. Khan, J. Tanveer, and A. U. Rehman, “Enhancing Intrusion Detection Using a Hybrid Machine Learning and Deep Learning Approach,” *IEEE Access*, vol. 12, pp. 12345–12360, 2024.

[3] D. Gümüşbaş, T. Yıldırım, A. Genovese, and F. Scotti, “A Comprehensive Survey of Databases and Deep Learning Methods for Cybersecurity and Intrusion Detection Systems,” *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2391–2420, 2020.

[4] M. A. Talukder, S. U. Islam, A. Hossain, and M. M. Rahman, “Machine Learning-Based Network Intrusion Detection for Big and Imbalanced Data Using Oversampling and Feature Extraction,” *Journal of Network and Computer Applications*, vol. 220, pp. 103500, 2024.

[5] L. Ashiku and C. Dagli, "Network Intrusion Detection System Using Deep Learning," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2021, pp. 1450–1457.

[6] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, and S. Venkatraman, "Deep Learning Approach for Intelligent Intrusion Detection System," *IEEE Access*, vol. 7, pp. 41525–41550, 2019.