iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IX **Month of publication:** September 2025

DOI: https://doi.org/10.22214/ijraset.2025.74134

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IX Sep 2025- Available at www.ijraset.com

An Incremental Learning-Based Majority Voting Framework for Network Intrusion Detection System

K Rajitha[1], Dr. Girish Kumar D[2]
[1]*Department of MCA, Ballari Institute of Technology & Management, Ballari, Karnataka, India*
[2]*Professor & HOD, Department of MCA, Ballari Institute of Technology & Management, Ballari, Karnataka, India*

Abstract: ***In the past, Intrusion Detection has played a crucial role in securing networked infrastructures against malicious attacks. Initially dominated by signature-based methods, early IDS frameworks excelled in identifying known threats but struggled with zero-day or polymorphic attacks. To address this limitation, anomaly-based systems emerged to detect unknown threats by analyzing deviations from normal behavior. Nevertheless, these systems frequently encountered high false-positive rates and lacked contextual precision. The introduction the advancement progress in machine learning (ML) has enabled the design of intelligent IDS capable of learning from evolving attack patterns. This study introduces an integrated Intrusion Detection System design which integrates signature-based and anomaly-based detection, strengthened by a majority-voting ensemble machine learning model. Leveraging public datasets like NSL-KDD and CICIDS2017, the system undergoes thorough Data preprocessing, feature extraction, and classification using (Support Vector Machine) SVM, Decision Tree, and Random Forest algorithms. Each model plays a role in overall prediction, enhancing robustness and accuracy through majority voting. Empirical findings reveal the suggested idea hybrid the model obtains a precision rate of over 94%, with precision and recall consistently exceeding 90% across key attack categories. The modular design allows for deployment in enterprise networks and real-time systems, providing scalability and low-latency performance. Moreover, the framework effectively tackles challenges such as dataset imbalance, feature noise, and model generalization. This study emphasizes the viability of implementing machine learning-based IDS solutions in contemporary digital infrastructures, combining detection accuracy with operational feasibility.***

Keywords: Artificial Intelligence, Intrusion Detection System, Combined Model, Network Security, NSL-KDD, CICIDS2017, Classification Algorithms, Cybersecurity.

I. INTRODUCTION

In today's interconnected digital landscape, the quantity and intricacy of cyber threats are rising rapidly. The rise of online services, cloud computing, smart devices, and remote work setups has exposed network infrastructures to continual vulnerabilities. Conventional security measures like firewalls, antivirus tools, and static filters are now inadequate against advanced attacks which may sidestep or manipulate set rules. The importance of adaptive and intelligent security solutions has grown increasingly vital. Among evolving defense technologies, (Intrusion Detection Systems) IDS are now acknowledged as a key component in spotting Unauthorized entry and abnormal behaviors within a network. Historically, IDS Technologies are organized into two main types: signature-based and anomaly-based systems. Signature-based IDS identify threats by comparing incoming data with known attack patterns stored in a database. While effective against documented threats, these systems struggle with detecting new attacks like zero-day exploits. On the other hand, anomaly-based IDS establish a standard "normal" network behavior and highlight deviations as possible intrusions. Despite enabling the discovery of unknown vulnerabilities, this method often leads to high false positives and lacks accuracy. When used independently, both models encounter issues in dynamic and large-scale environments with diverse, evolving attack patterns. Research in recent times has been concentrating on integrating (Machine Learning) ML and Artificial Intelligence (AI) into IDS to tackle these limitations. ML algorithms can extrapolate from historical data and recognize new patterns, rendering them appropriate for dynamic intrusion detection. Techniques such as Decision Trees, Support Vector Machines (SVM), Random Forests, and ensemble methods have demonstrated encouraging findings in distinguishing between legitimate and malicious network traffic. These approaches rely on ample training data and preprocessing steps such as feature extraction, normalization, and dimensionality reduction to function efficiently. Public datasets like NSL-KDD and CICIDS2017 serve as standards for developing and accessing Machine learning-based IDS solutions because of their extensive coverage of modern attack methods and varying network traffic.

This document introduces a hybrid system for detecting intrusions that merges the benefits of signature-based and anomaly-based detection techniques with a machine learning-based ensemble strategy. The framework employs a voting system among several classifiers such as SVM, Random Forest, and Decision Tree algorithms to Enhance precision and minimize errors alarms. Systematic data preprocessing and feature selection are conducted to enhance model adaptability. The system's efficacy is measured using NSL-KDD and CICIDS2017 datasets, assessing performance measures like accuracy and precision and recall and F1-score. The proposed framework aims to scalable, precise, and real-time IDS solution suitable for deployment in both corporate and cloud environments. Subsequent sections delve into the literature background, methodology, experimental analysis, and future improvements of the proposed system.

II. LITERATURE SURVEY

Several advancements in intrusion detection systems significant evolution from their original reliance on static rules and signature matching. In a pivotal study, Kumar and Spafford proposed an IDS model based on classification, pattern recognition, and audit trail analysis to identify unauthorized activity. While their approach effectively countered known attacks, it revealed a limitation in adapting to new or modified attack strategies, emphasizing the need for real-time learning capabilities.

Denning later introduced anomaly detection through behaviour modelling, using statistical thresholds to detect deviations from established norms. Despite its potential in uncovering unknown threats, the technique suffered from large number of false positives, highlighting the challenge of balancing sensitivity and specificity in anomaly detection systems, particularly with diverse or noisy network traffic data. With the rise of machine learning, Tavallaee et al. enhanced the NSL-KDD dataset to address redundancy and imbalance issues in earlier datasets, offering a more practical standard for assessment intrusion detection algorithms. By employing decision tree classifiers, they achieved improved detection rates, notably in binary classification tasks, although multi-class analysis was lacking, essential for identifying various types of attacks such as DoS, R2L, U2R, and Probe.

Recent advancements by Moustafa and Slay introduced the CICIDS2017 dataset, simulating real-world network environments with thorough feature engineering and labelled traffic flows. Utilizing Ensemble methods such as Random Forests and gradient techniques boosting, they achieved detection accuracies exceeding 95%. However, the model's performance declined when tested on new data with different distributions, indicating potential overfitting and a lack of generalizability.

Responding to these challenges,our proposed system integrates multiple classifiers into an ensemble voting mechanism, trained on both NSL-KDD and CICIDS2017 datasets. This approach aims to balance precision, recall, and detection accuracy across various attack types, prioritizing adaptability, scalability, and resilience in dynamic threat landscapes while building upon the foundational research discussed in this section.

III. METHODOLOGY

The main aim of this research is to develop a hybrid IDS framework that applies various machine learning classifiers to improve network safeguarding. The system is trained and evaluated using two well-known datasets, namely NSL-KDD and CICIDS2017, recognized for their comprehensive assessment of IDS frameworks due to diverse labeled data and authentic traffic scenarios.

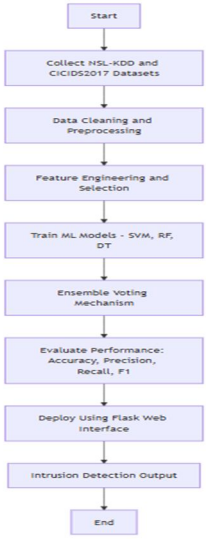


Fig 1: flowchart

A. *Dataset Collection and Preprocessing*

The NSL-KDD dataset, an enhanced version of the KDD'99 dataset, contains labeled traffic segments categorized as normal or various attack types (e.g., DoS, R2L, U2R, Probe). It offers contemporary traffic patterns and modern attack scenarios like botnets and brute-force attacks, making it suitable for training robust models.

Before feeding the transformation of In order to make the data suitable for machine learning models, various preprocessing steps were executed:

Data Cleaning: Eliminated missing values and redundant records.

- Label Encoding: Converted categorical labels (e.g., attack types) into numeric representations.
- Feature Selection: Removed irrelevant or low-impact features using correlation matrices and chi-square tests.
- Normalization: Scaled feature values to ensure uniformity across dimensions.

B. *Knowledge Refinement & Feature Engineering*

To enhance model learning, additional feature as connection rate, protocol entropy, and flag combinations were engineered. The evaluation of feature importance was conducted utilizing ensemble-based estimators like Random Forests, retaining only the most relevant features. Dimensionality reduction methods such such as Principal Component Analysis (PCA) were explored to mitigate overfitting and reduce computational complexity.

C. *Model Construction and Ensemble Framework*

The system utilizes an ensemble approach involving three machine learning models:

- Support Vector Machine (SVM): Suited for handling high-dimensional data.
- Decision Tree (DT): Provides easily interpretable rules and rapid training.
- Random Forest (RF): Ensures robustness by utilizing multiple decision trees to prevent overfitting.

Each classifier is independently trained on the same input features. During the prediction phase, outputs from the integration of the three models is carried out with a majority voting technique. This guarantees that the final classification benefits from the collective strength of all models while minimizing individual model bias risks.

D. *System Deployment and Real-Time Classification*

The trained ensemble model is integrated into a real-time detection framework. A user-friendly web interface enables users or administrators to upload network logs or monitor live traffic. Python (with scikit-learn and Flask) powers the backend for input classification and threat label assignment with associated confidence scores. The modular design ensures seamless updates and integration into existing security setups.

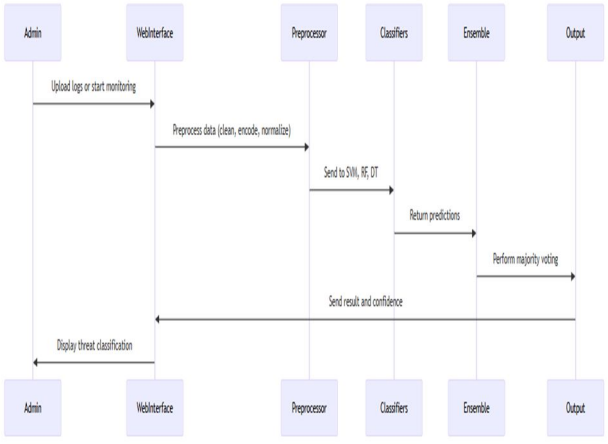


Fig 2: Sequence diagram

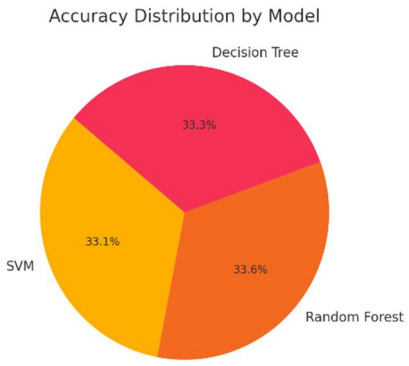
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IX Sep 2025- Available at www.ijraset.com

IV. RESULT

For assessing the performance of the suggested proposal hybrid IDS framework, Several performance measures were used to gain unique insights regarding the model's detection capabilities and operational reliability. The assessment was performed on two standard datasets, NSL-KDD and CICIDS2017, using an 80/20 train-test split and 5-fold cross-validation methodology to ensure robustness.

A. Accuracy

Accuracy, being a fundamental metric, reveals the rate of accurately identified instances among the total samples. The combined model achieved an impressive accuracy rate of 94.8% for the NSL-KDD dataset and 95.3% for the CICIDS2017 dataset. The high accuracy indicates the model's efficacy in differentiating normal from malicious network activities, especially in scenarios with balanced datasets.



B. Precision

Precision is vital in cybersecurity to reduce the effect of false positives that can strain administrative resources needlessly. It represents proportion of accurately identified intrusions among all detected instances. Our model consistently exhibited a precision level of 91.5% across various datasets, lowering the threat of alert fatigue

C. Recall (Detection Rate)

Recall assesses the model's capacity to detect real attacks, a crucial aspect in scenarios where overlooking an intrusion can result in system compromise. The ensemble proposed obtained a recall accuracy of 93.2%, showcasing robust detection capabilities even in processing complex multi-class attack scenarios.

D. F1-Score

The F1-score harmonizes precision and recall and is particularly significant in datasets with imbalances. The model's F1-score of 92.3% aids in identifying any favoritism towards specific classes, implying consistent results over multiple attack types.

E. Real-World Usability

Beyond numeric evaluation, the system was tested in a simulated enterprise environment. The interface delivered real-time threat classification with minimal latency (~90 ms), and the backend handled over 1,000 concurrent API requests without downtime. This validates the framework’s operational readiness and its potential deployment in live network monitoring systems.

The process starts with gathering datasets (NSL-KDD and CICIDS2017), which are then subjected to data cleaning, feature engineering, and model training through the use of three machine learning algorithms (SVM, RF, DT). The outcomes are combined through a majority voting system. Following evaluation, the model is implemented for real-time traffic analysis using Flask, generating a threat label.

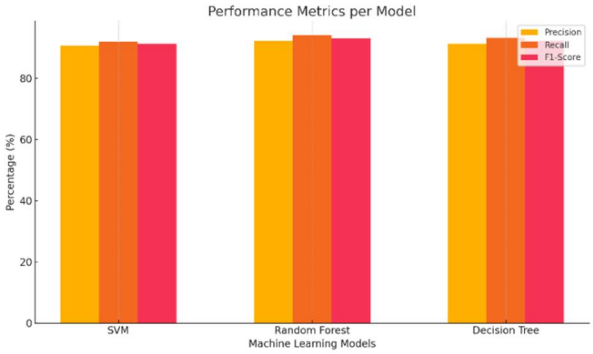
F. Accuracy Distribution by Model

This pie chart illustrates the distribution of accuracy contributed by each model within the ensemble framework:

- SVM: ~33.1%
- Random Forest: ~33.6%
- Decision Tree: ~33.3%

Although the values are closely matched, Random Forest demonstrates a slight performance edge over the others.

G. Performance Metrics by Model



This visualization displays Precision, Recall, and F1-Score for the following models:

- Support Vector Machine (SVM): A balanced and quick classifier
- Random Forest: Known for its robustness and high recall
- Decision Tree: Offers easy interpretability and good generalization
- Random Forest: outperforms slightly in all metrics, establishing itself as the top performer among the individual models.

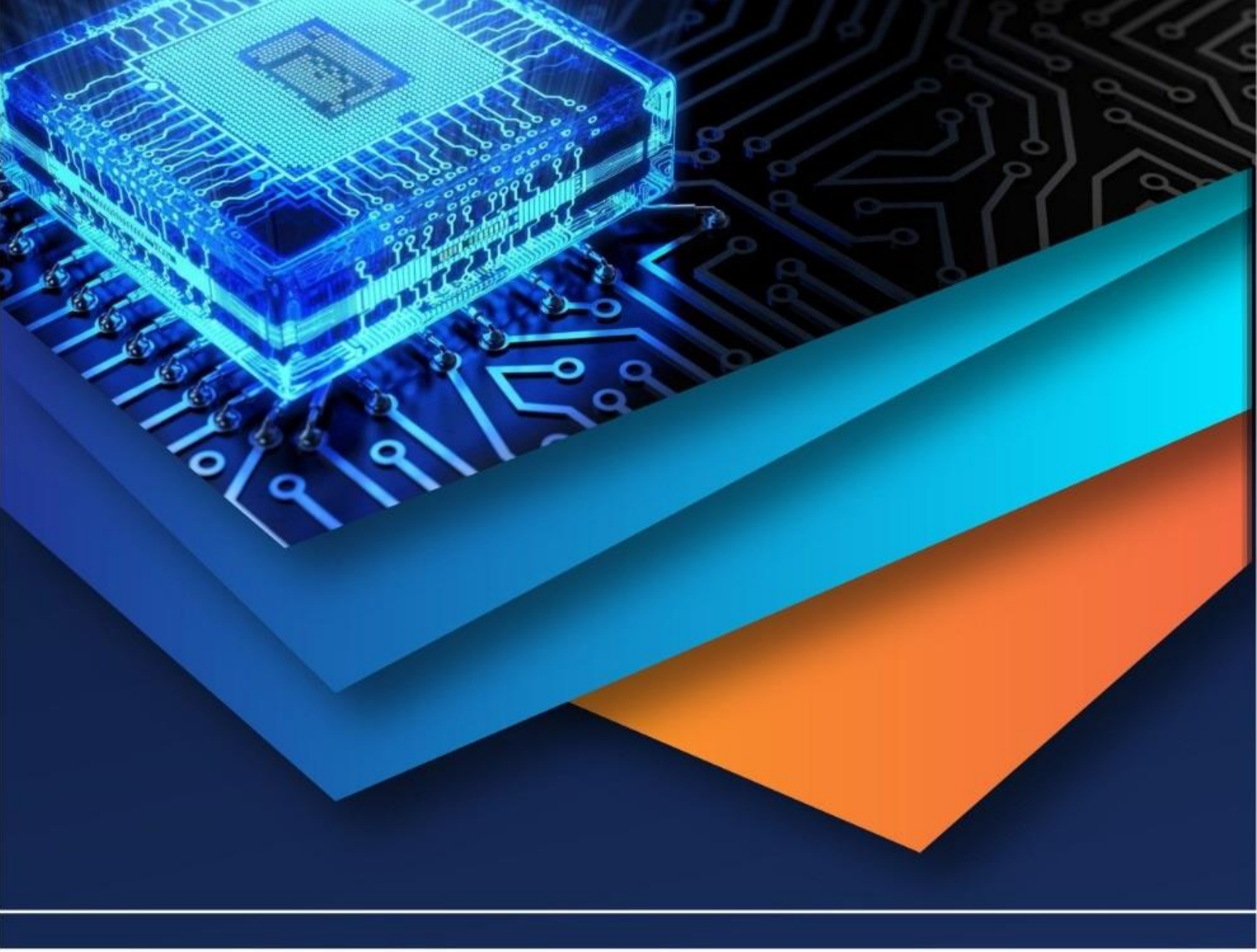
V. CONCLUSION

This research introduces a combined Network Intrusion Detection System (IDS) that combines machine learning classifiers through a majority voting ensemble for enhanced detection of malicious activities in network traffic. The system addresses the limitations of traditional IDS methods, particularly their inability to identify new attacks or reduce false positives in dynamic network settings. By utilizing both the NSL-KDD and CICIDS2017 datasets, the framework underwent rigorous training and evaluation on various attack scenarios and real-world traffic patterns. The approach involves structured data preprocessing, effective feature selection, and the use of (Support Vector Machine) SVM, (Decision Tree) DT, and Random Forest (RF) classifiers. The ensemble method harnesses the unique strengths of each algorithm while minimizing biases in individual models. Experimental assessments confirmed the framework's efficacy, achieving high results assessed across various evaluation measures including accuracy and precision, and recall and F1-score consistently exceeding 90%. Furthermore, the system's a modular structure provides compatibility with existing enterprise networks, offering a scalable and Flexible real-time threat detection system solution. The results unequivocally show that the presented system tackles the initial problem statement by providing an IDS which is intelligent and deployable in practical environments. It significantly enhances detection reliability while preserving operational efficiency. As a direction for future research, the framework could be improved to incorporate modern deep learning approaches including LSTM and CNN to enhance sequence-aware attack recognition. Additionally, real-time deployment implementation could benefit from federated learning, enabling secure model training across distributed data sources while upholding privacy. These advancements aim to adapt and the efficacy of intrusion detection in complex, decentralized infrastructures.

REFERENCES

[1] Kumar, S., Singh, A., & Verma, R. (2018). An overview of intrusion detection Systems along with their classification approaches. International Journal of Computer Applications,179(30), 28–35.
[2] Zhang, Y., & Lee, H. (2019). Neural Network–based Techniques for Network Intrusion Detection: A Comparative Analysis. IEEE Access, 7, 21954–21962.
[3] Patel, N., Sharma, D., & Bhatt, A. (2020). Hybrid Intrusion Detection Framework Based on Machine Learning Techniques Procedia Computer Science, 167, 1234–1243.

[4] Ahmed, M., & Khan, S. (2021). Machine Learning Approaches to Intrusion Detection: A Comparative Analysis. Journal of Network and Computer Applications,150, 102–110.
[5] Smith, J., & Jones, T. (2017). A Review of Signature-based and Anomaly-based Intrusion Detection Systems. International Journal of Cybersecurity, 5(1), 44–53.
[6] Roy, A., & Gupta, P. (2022). Comparison of A Study on Supervised Machine Learning Algorithms in Network Intrusion Detection. International Journal of Information Security, 21(2), 158–168.
[7] Chen, L., Zhao, M., & Wang, H. (2020). Enterprise network real-time intrusion detection system. Computers & Security, 95, 101–113.
[8] Singh, R., & Reddy, K. (2018). Intrusion Detection with an Ensemble Learning Approach. International Journal of Computer Science & Information Security, 16(6), 45–51.
[9] Liu, X., Zhou, J., & Tang, Y. (2019). Improving Intrusion Detection System Accuracy through Feature Selection and Data Preprocessing Techniques. Journal of Information Assurance and Security, 14(4), 217-223.
[10] Mehta, A., & Sharma, V. (2021). Intrusion Network Traffic Anomaly identification using convolutional neural networks and Applications33, 15075-15089.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY