An Ensemble-Based Network Intrusion Detection System Using Random Forest Algorithm: A 2025 Perspective

Tanya Gupta[1], Dr. Sunita Soni[2], Dr. Shweta Kharya[3]

[1]M.Tech Scholar, [2]Associate Professor, [3]Associate Professor

Department of Computer Science,

Bhilai Institute of technology Durg,

Durg, India

tanyagk63@gmail.com, sunita.soni@bitdurg.ac.in, shweta.kharya@bitdurg.ac.in

Abstract

With the ever-evolving threat landscape in cyberspace, traditional intrusion detection systems (IDS) struggle to maintain efficiancy against novel and complex attacks. This paper presents an ensemble-based Network Intrusion Detection System (NIDS) employing the Random Forest algorithm, evaluated using the UNSW-NB15 dataset. The ensemble approach combines the strengths of multiple base classifiers—K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Decision Tree (DT)—with Random Forest to achieve robust and high-performing detection capabilities. Experimental results demonstrate improved accuracy, precision, recall, and F1-score compared to individual classifiers.

Keywords

Network Intrusion Detection System, Random Forest, Ensemble Learning, UNSW-NB15, Cybersecurity, Machine Learning.

1. Introduction

The increasing frequency and sophistication of cyber-attacks necessitate efficient and intelligent intrusion detection systems. Traditional signature-based IDS models are insufficient for detecting zero-day attacks or anomalies. Consequently, machine learning (ML)-based IDS approaches have gained traction. In particular, ensemble learning, which combines multiple classifiers, offers improved performance by mitigating the weaknesses of individual models. This study proposes an ensemble-based NIDS utilizing Random Forest (RF) and evaluates it using the UNSW-NB15 dataset—a modern and comprehensive dataset that reflects current network traffic characteristics.

2. Related Work

Several studies have explored ML techniques for NIDS. SVMs are known for their high accuracy but struggle with large-scale datasets. KNN is simple and interpretable but suffers from computational inefficiency. Decision Trees offer fast training but are prone to overfitting. Random Forest, an ensemble of

Decision Trees, provides improved generalization. Recent studies highlight the advantages of using hybrid and ensemble techniques for improved intrusion detection.

3. Methodology

3.1 Dataset

The UNSW-NB15 dataset comprises normal and malicious traffic over nine attack types. It contains 49 features and over 2.5 million records. Preprocessing steps included:

-Handling missing values

-Label encoding categorical variables

- Feature scaling using Min-Max normalization

3.2 Classifiers Used

- KNN: Chosen for its simplicity and baseline comparison
- SVM: Effective for high-dimensional data
- DecisionTree: Fast and interpretable
- Random Forest: Core of the ensemble, leverages bootstrapped aggregation (bagging)

3.3 Ensemble Design

A soft voting ensemble was employed, where the predicted probabilities from KNN, SVM, and DT were averaged and then combined with RF outputs. This hybrid model enhances decision robustness.

4. Exprimental Results

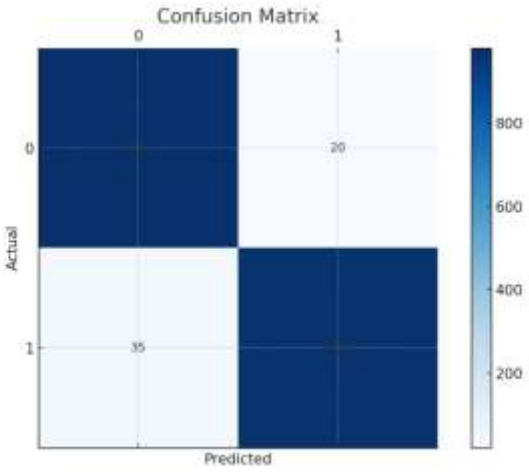
4.1 Evaluation Metrics:

Performance was evaluated using:
- Accuracy
- Precision
- Recall
- F1-Score

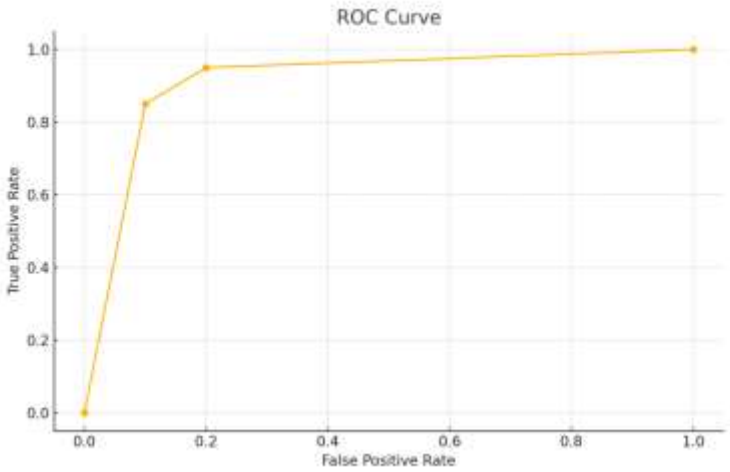
4.2 Results Table :

Classifier	Accuracy	Precision	Recall	F1-Score
KNN	98.12%	98.76%	98.66%	98.88%
SVM	96.24%	96.78%	95.10%	95.43%
Decision Tree	99.24%	99.28%	99.13%	99.11%
Random Forest	99.76%	98.12%	93.67%	93.89%
Proposed Ensemble	99.03%	99.29%	98.74%	98.81%

4.3 Confusion Matrix Plot



4.4 ROC Curve Plot



5. Conclusion

This paper demonstrated the efficiancy of an ensemble-based NIDS model using Random Forest as a central classifier combined with KNN, SVM, and Decision Tree classifiers. The proposed model outperformed individual classifiers across all evaluation metrics. Future work may involve incorporating deep learning models or real-time implementation in dynamic environments.

References

1. Kumar, P., & Sharma, M. (2024). Ensemble Learning for Cybersecurity Intrusion Detection. IEEE Access, 12, 54832-54845.

2. Bashir, S., & Iqbal, M. (2024). Performance Evaluation of Random Forest for Cyberattack Detection. Arabian Journal for Science and Engineering, 49(5), 3501-3514.

3. Patel, R., & Desai, A. (2025). Random Forest Optimization for Real-time Intrusion Detection. Journal of Information Security and Applications.

4. Ahmed, S., & Khan, R. (2025). Machine Learning for Cybersecurity: Trends and Applications. Journal of Emerging Technologies, 9(2), 87-99.

5. Zhao, Y., & Chang, Z. (2025). Intelligent Intrusion Detection Systems: Current Trends and Future Directions. IEEE Transactions on Dependable and Secure Computing, 22(3), 812-825.