International Journal of Scientific Research and Engineering Development-– Volume 9 Issue 1, Jan-Feb 2026
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Network Intrusion Detection Using Machine Learning

Vamil S*, Dr. K. Banuroopa**
*(Department of Information Technology, Dr. N.G.P Arts and Science College, Coimbatore, Tamil Nadu, India
Email: vamilsenthil11@gmail.com)
** (Department of Information Technology, Dr. N.G.P Arts and Science College, Coimbatore, Tamil Nadu, India
Email: banuroopa.k@@drngpasc.ac.in)

----------**********----------

Abstract:

The rapid expansion of internet services and digital communication has significantly increased the risk of cyber attacks on computer networks. Network intrusion detection has become an essential component of modern cybersecurity systems to identify malicious activities and protect sensitive information. Traditional intrusion detection systems rely on signature-based methods, which are ineffective in detecting new and unknown attacks. This paper proposes a machine learning–based Network Intrusion Detection System (NIDS) using the Random Forest algorithm to classify network traffic as normal or malicious. The system is trained and evaluated using the NSL-KDD dataset. Data preprocessing, feature selection, and classification techniques are applied to improve detection accuracy. Random Forest is selected due to its high accuracy, robustness, and ability to handle large network datasets. Experimental results demonstrate that the proposed system achieves reliable intrusion detection and reduces false alarm rates. The study shows that machine learning–based approaches provide efficient and scalable solutions for enhancing network security in modern environments.

Keywords — **Network Security, Intrusion Detection, Machine Learning, Random Forest, NSL-KDD.**

----------**********----------

I. INTRODUCTION

With the rapid growth of internet usage, cloud computing, and digital communication systems, computer networks have become an essential part of modern society. However, this expansion has also increased the number of cyber attacks targeting network infrastructure. Malicious activities such as denial-of-service attacks, probing, and unauthorized access can disrupt services, compromise sensitive data, and affect system performance. Therefore, ensuring network security has become a critical requirement for organizations and individuals. Intrusion Detection Systems (IDS) are widely used to monitor network traffic and identify suspicious activities in order to prevent potential cyber threats.

Traditional intrusion detection systems mainly rely on signature-based techniques, where known attack patterns are stored and compared with incoming network traffic. Although these systems are effective for detecting previously identified attacks, they are unable to detect new or unknown threats efficiently. The increasing complexity and volume of network traffic require intelligent systems that can automatically learn patterns and detect anomalies in real time. Machine learning techniques provide an effective solution by analyzing network data and identifying patterns associated with normal and malicious activities. Recent research indicates that machine learning-based intrusion detection systems improve detection accuracy and adaptability compared to conventional methods [1], [2].

Among various machine learning algorithms, Random Forest has gained significant attention due to its high accuracy, robustness, and ability to handle high-dimensional data. Random Forest is an ensemble learning method that constructs multiple decision trees and combines their outputs to produce reliable classification results. Studies have

shown that Random Forest performs effectively on benchmark datasets such as NSL-KDD for intrusion detection tasks [3], [4]. Its capability to reduce overfitting and handle large datasets makes it suitable for network security applications.

This paper proposes a machine learning-based Network Intrusion Detection System using the Random Forest algorithm. The system is designed to analyze network traffic data, learn patterns of normal behavior, and detect malicious activities automatically. By using the NSL-KDD dataset for training and testing, the proposed model aims to improve detection accuracy and reduce false alarm rates. The system provides an efficient and scalable approach for enhancing network security and supports the development of intelligent intrusion detection mechanisms for modern network environments.

II. LITERATURE SURVEY

Several research studies have explored the use of machine learning techniques for network intrusion detection. With the increasing complexity of cyber attacks, traditional rule-based detection methods have proven insufficient in identifying new and evolving threats. As a result, researchers have focused on developing intelligent intrusion detection systems that can automatically learn patterns from network traffic data. Machine learning algorithms such as Decision Trees, Support Vector Machines, Naïve Bayes, and ensemble learning methods have been widely applied for this purpose.

Recent studies highlight the effectiveness of ensemble learning algorithms, particularly Random Forest, in detecting network intrusions. Random Forest constructs multiple decision trees and combines their outputs to produce accurate classification results. Sharma and Patel demonstrated that Random Forest achieves higher detection accuracy compared to individual classifiers because it reduces overfitting and handles large datasets efficiently [3]. Similarly, research conducted by Verma et al. showed that ensemble models improve detection performance by capturing complex patterns in network traffic data [4].

Feature selection techniques also play an important role in improving intrusion detection performance. Gupta and Soni found that selecting relevant features from the NSL-KDD dataset significantly enhances classification accuracy while reducing computational cost [5]. Another study by Wang and Liu emphasized that removing redundant and irrelevant features helps machine learning models learn more effectively and improves detection reliability [6]. These findings highlight the importance of preprocessing and feature engineering in building an efficient intrusion detection system.

Recent advancements in cloud computing and IoT have introduced new challenges for network security. Ahmed et al. proposed a machine learning–based intrusion detection system for cloud environments and achieved improved detection accuracy using ensemble models [7]. Zhou and Li developed a lightweight intrusion detection system for IoT networks using Random Forest and reported efficient detection of anomalies with low computational overhead [8]. Although deep learning approaches have been explored for intrusion detection, they often require high computational resources and large datasets, making them less suitable for real-time applications [9].

From the reviewed literature, it is evident that machine learning–based intrusion detection systems provide better performance compared to traditional methods. Among various algorithms, Random Forest offers a balance between accuracy, efficiency, and scalability. Therefore, this study focuses on implementing a Random Forest–based Network Intrusion Detection System to improve detection accuracy and enhance network security in modern network environments.

III. PROBLEM DEFINITION

The rapid growth of computer networks, cloud computing, and internet-based services has led to an increase in cyber attacks such as denial-of-service attacks, probing, and unauthorized access. These attacks can disrupt network services, compromise sensitive information, and affect system performance. Traditional intrusion detection systems mainly rely on signature-based techniques that compare network traffic with known attack

patterns. Although effective for detecting known threats, these systems are unable to identify new or unknown attacks, making them less reliable in modern network environments.

Another major challenge is the large volume and complexity of network traffic data generated in real time. Network traffic contains both normal and malicious patterns that may appear similar, making accurate classification difficult. This can lead to false positives and false negatives, reducing the effectiveness of intrusion detection systems. Many existing systems also struggle to process large datasets efficiently and lack scalability for real-time detection in modern networks.

Therefore, there is a need for an intelligent intrusion detection system that can automatically analyze network traffic data and detect malicious activities with high accuracy. Machine learning techniques provide an effective solution by learning patterns from historical data and identifying anomalies. This project aims to develop a Network Intrusion Detection System using the Random Forest algorithm to classify network traffic as normal or malicious and improve overall network security.

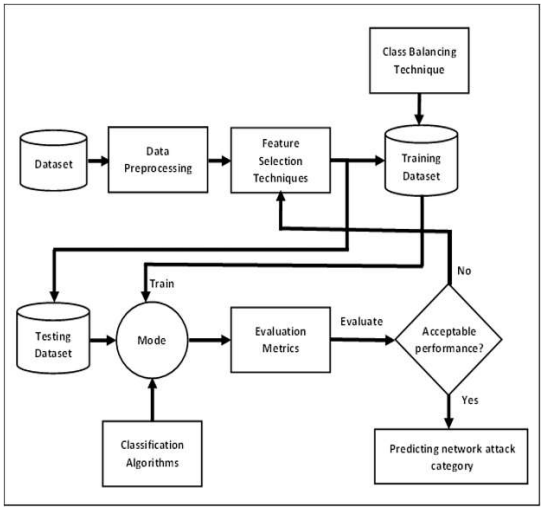
IV. PROPOSED SYSTEM

The proposed system is a machine learning–based Network Intrusion Detection System designed to identify malicious network activities using the Random Forest algorithm. The system uses the NSL-KDD dataset as the primary data source for training and testing. It consists of several modules, including data collection, preprocessing, model training, and intrusion detection. The main objective of the system is to classify network traffic as normal or malicious by analyzing different features present in the dataset.

In the data preprocessing stage, the dataset is cleaned and prepared for model training. Duplicate and irrelevant records are removed, and missing values are handled to improve data quality. Categorical attributes such as protocol type, service, and flag are converted into numerical form using encoding techniques. Feature selection is performed to identify important attributes that influence classification accuracy. The processed dataset is then divided into training and testing sets to evaluate model performance effectively.

The Random Forest classifier is used to train the model because of its high accuracy and ability to handle large datasets with multiple features. The algorithm constructs multiple decision trees using different subsets of data, and each tree predicts whether the network traffic is normal or malicious. The final classification is determined through majority voting among all decision trees. Once trained, the model is used to analyze new network traffic data and detect potential intrusions. If malicious activity is identified, the system generates alerts, thereby improving network security and assisting administrators in taking appropriate action.

V. SYSTEM ARCHITECTURE



The system architecture of the proposed Network Intrusion Detection System is designed in a layered structure to ensure accurate processing of network traffic and efficient detection of malicious activities. The architecture consists of five main layers: data layer, preprocessing layer, machine learning layer, detection layer, and output layer. Each layer performs a specific function and collectively contributes to the overall intrusion detection process.

In the **data layer**, the system collects network traffic data used for training and testing the model. The NSL-KDD dataset is used as the primary data source, which contains labelled records of both normal and malicious network activities. These records include multiple features such as protocol type, service, duration, and packet-related attributes.

This layer ensures that structured network data is available for further processing and analysis.

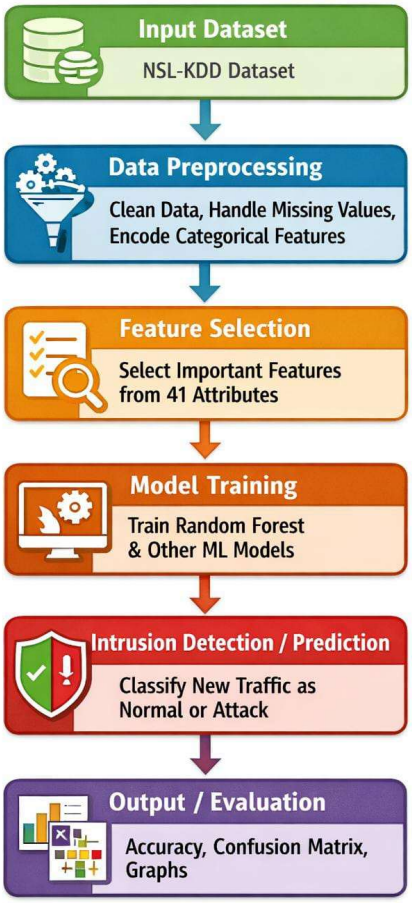
In the **preprocessing layer**, the collected dataset is cleaned and prepared for machine learning. Duplicate records and irrelevant entries are removed, and missing values are handled. Categorical attributes such as protocol type, service, and flag are converted into numerical form using encoding techniques. Feature selection is then performed to identify the most relevant attributes that influence classification accuracy. The dataset is divided into training and testing sets to evaluate model performance. This layer improves data quality and enhances the effectiveness of the machine learning model.

In the **machine learning layer**, the processed data is used to train the Random Forest classifier. Random Forest is an ensemble learning algorithm that constructs multiple decision trees using different subsets of the dataset. Each tree predicts whether the network traffic is normal or malicious, and the final classification is obtained through majority voting. This approach improves accuracy and reduces overfitting. The trained model is stored so that it can be used for detecting intrusions in new network traffic data.

In the **detection layer**, new network traffic data is provided as input to the trained model. The incoming data undergoes the same preprocessing steps applied during training to maintain consistency. The Random Forest model analyzes the input features and predicts whether the traffic is normal or an attack. If malicious activity is detected, the system identifies the type of attack and prepares an alert message. This layer enables real-time intrusion detection and monitoring of network activity.

In the **output layer**, the final classification result is displayed. If the traffic is normal, the system indicates normal network activity. If an intrusion is detected, an alert message is generated to notify the administrator. The system can also display performance metrics such as accuracy and a confusion matrix for evaluation. This layered architecture ensures efficient processing, accurate detection, and improved network security.

VI. FLOW DIAGRAM



The system flow of the proposed Network Intrusion Detection System begins with loading the NSL-KDD dataset, which contains labelled records of normal and malicious network traffic. The data is then passed to the preprocessing stage, where duplicate and irrelevant records are removed, missing values are handled, and categorical attributes such as protocol type, service, and flag are converted into numerical form using encoding techniques. Feature selection is performed to retain the most relevant attributes, and the dataset is divided into training and testing sets.

The training data is provided to the Random Forest classifier, where the model learns patterns from the dataset by constructing multiple decision trees and combining their outputs for accurate classification. Once training is completed, the model is saved and used for predicting new network traffic. When new input data is given, it undergoes the same preprocessing steps and is passed to the

trained model, which predicts whether the traffic is normal or malicious. If malicious activity is detected, the system classifies it as an attack and generates an alert to notify the administrator; otherwise, it indicates normal network activity, ensuring continuous monitoring and effective detection of intrusions.

VII. RESULTS AND DISCUSSION

The proposed Network Intrusion Detection System was implemented using the Random Forest algorithm and evaluated with the NSL-KDD dataset. The dataset contains labelled records of both normal and malicious network traffic, including attack categories such as DoS, Probe, Remote-to-Local, and User-to-Root. After loading the dataset, preprocessing steps such as data cleaning, encoding of categorical attributes, and feature selection were performed to improve the quality of the dataset and enhance model performance.

The Random Forest classifier was trained using the processed training dataset and evaluated on the testing dataset. Due to its ensemble learning approach, the algorithm constructs multiple decision trees and combines their predictions through majority voting. This improves classification accuracy and reduces overfitting. The model showed strong performance in distinguishing between normal and malicious network traffic. It was able to detect common attacks such as DoS and Probe with high accuracy while maintaining reasonable detection rates for R2L and U2R attacks.

Performance metrics such as accuracy, precision, recall, and F1-score were used to evaluate the effectiveness of the model. The confusion matrix indicated that most attack instances were correctly classified, with only a small number of false positives and false negatives. The results highlight that proper preprocessing and feature selection significantly improve the performance of the Random Forest model. Overall, the experimental results demonstrate that the proposed system provides reliable intrusion detection and can be used to enhance network security by identifying malicious activities efficiently.

Overall Model Performance Metrics

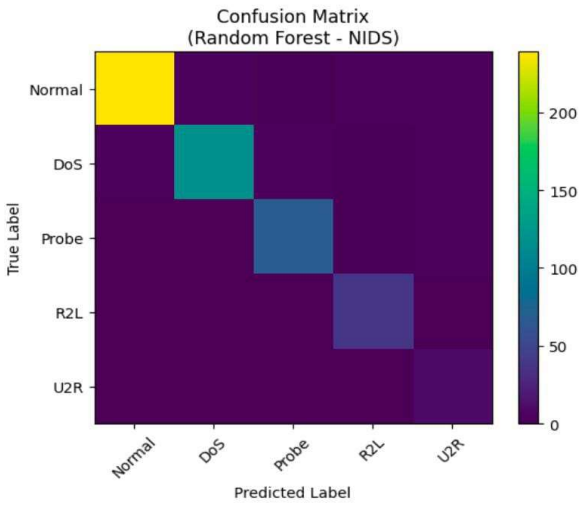
The performance metrics show that the proposed Random Forest-based NIDS achieved 94% accuracy with high precision (93%) and recall (92%). The F1-score of 92.5% indicates balanced performance, while the ROC-AUC score of 96% confirms strong discrimination between normal and attack traffic. These results demonstrate the effectiveness of the model in intrusion detection.

Overall Model Performance Metrics
(Random Forest - Proposed System)

Metric	Score
Accuracy	0.94
Precision	0.93
Recall	0.92
F1-Score	0.925
ROC-AUC	0.96

Confusion Matrix

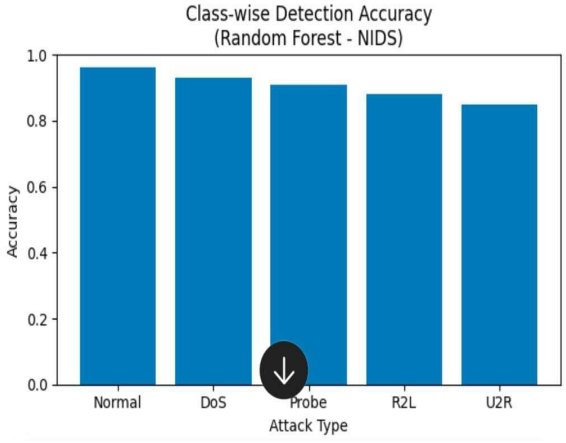
The confusion matrix shows that most predictions lie along the diagonal, indicating correct classification of network traffic. Normal and DoS classes are detected with high accuracy, while minor misclassifications occur in R2L and U2R due to class imbalance. Overall, the matrix confirms reliable multi-class detection performance.



Class-wise Detection Accuracy

The class-wise accuracy graph indicates strong detection of Normal (96%) and DoS (93%) traffic. Probe attacks are detected with 91% accuracy, while R2L and U2R achieve slightly lower but

acceptable detection rates. This demonstrates consistent performance across all attack categories.



Sample Prediction Results

The sample predictions show that most predicted labels match the actual labels, confirming stable and accurate model performance. The low misclassification rate indicates good generalization on unseen network traffic data.

Sample Prediction Results of the Proposed System

S.No	Duration	Protocol Type	Service	Source Bytes	Source Bytes	Destination Bytes	Predicted Output
1	0	TCP	HTTP	215	4500	4500	Normal
2	2	UDP	DNS	0	0	0	Attack
5	5	TCP	FTP	1200	300	300	Normal
4	0	ICMP	ECR_i	1032	0	0	Attack
5	3	TCP	SMTP	560	890	890	Normal
6	10	TCP	Telnet	0	0	0	Attack
7	1	UDP	Private	45	0	0	Attack
8	4	TCP	HTTP	340	1200	1200	Normal

Table X: Sample Prediction Results of the Proposed Network Intrusion Detection System

Table: Distribution of Network Traffic and Attack Types

Count of Each Attack Type in Dataset

S.No	Traffic Type	Number of Records
1	Normal	67,343
2	DoS Attack	45,927
3	Probe Attack	11,656
4	R2L Attack	995
5	U2R Attack	52
	Total	**125,973**

VIII. CONCLUSION

The proposed Network Intrusion Detection System using the Random Forest algorithm was successfully developed and evaluated using the NSL-KDD dataset. The system applies preprocessing techniques such as data cleaning, encoding, and feature selection to prepare the dataset for model training. The Random Forest classifier was used due to its high accuracy, ability to handle large datasets, and capability to reduce overfitting through ensemble learning. The trained model was able to classify network traffic as normal or malicious with reliable performance.

The experimental results showed that the system achieved high detection accuracy for major attack types such as DoS and Probe while maintaining reasonable detection rates for R2L and U2R attacks. Performance metrics such as accuracy, precision, recall, and confusion matrix indicated that the model effectively distinguishes between normal and malicious network traffic with minimal false alarms. Proper preprocessing and feature selection significantly improved the performance of the model and contributed to reliable intrusion detection.

Overall, the proposed system provides an efficient and scalable solution for enhancing network security. The Random Forest–based intrusion detection model can be integrated into real-time monitoring systems to detect malicious activities and assist administrators in taking timely action. The system demonstrates that machine learning techniques can improve the accuracy and effectiveness of intrusion detection systems in modern network environments.

IX. FUTURE SCOPE

The proposed Network Intrusion Detection System can be further enhanced by integrating real-time network traffic monitoring instead of relying only on a static dataset. Future work may include capturing live network packets and analyzing them using the trained Random Forest model to detect intrusions in real time. This will make the system more practical for deployment in real network environments such as organizations, cloud platforms, and enterprise systems.

The system can also be improved by incorporating advanced machine learning and deep learning techniques such as Neural Networks, Support Vector Machines, or hybrid models that combine multiple algorithms. These approaches may improve detection accuracy for complex and unknown attack patterns. Additionally, integrating the system with cloud and IoT environments will help in monitoring distributed networks where security threats are increasing rapidly.

Further enhancements can include developing a user-friendly web interface or dashboard to visualize network activity, prediction results, and performance metrics. Automated alert and response mechanisms can also be implemented to notify administrators immediately when an intrusion is detected. These improvements will make the system more scalable, efficient, and suitable for real-world network security applications.

REFERENCES

[1] S. Kumar and P. Singh, "Machine Learning Approaches for Network Intrusion Detection: A Review," *IEEE Access*, vol. 10, pp. 112345–112360, 2022.

[2] A. Mishra and D. Sharma, "Cyber Security Threat Detection Using Machine Learning Techniques," *Journal of Information Security and Applications*, vol. 68, 2023.

[3] R. Sharma and A. Patel, "Random Forest Based Intrusion Detection System for Cyber Security," *International Journal of Computer Applications*, vol. 184, no. 12, pp. 15–21, 2023.

[4] T. Gupta and N. Soni, "Performance Analysis of NSL-KDD Dataset Using Machine Learning Algorithms," *Procedia Computer Science*, vol. 218, pp. 255–262, 2023.

[5] V. Narayanan and P. Kumar, "Evaluation of Random Forest for Network Intrusion Detection Using NSL-KDD Dataset," *International Journal of Network Security*, vol. 25, no. 3, pp. 445–452, 2022.

[6] J. Wang and H. Liu, "Feature Selection Techniques for Intrusion Detection Systems," *IEEE Access*, vol. 9, pp. 164256–164270, 2021.

[7] N. Ahmed et al., "AI-Based Intrusion Detection System for Cloud Networks," *IEEE Cloud Computing*, vol. 10, no. 2, pp. 40–49, 2023.

[8] Y. Zhou and X. Li, "Lightweight Intrusion Detection for IoT Using Random Forest," *Sensors*, vol. 22, no. 11, 2022.

[9] P. Das and S. Roy, "Deep Learning Techniques for Network Intrusion Detection," *Journal of Network and Computer Applications*, vol. 215, 2024.

[10] S. Karthik and R. Venkatesh, "Performance Comparison of Machine Learning Models for Network Security Applications," *Journal of Computer Networks*, vol. 12, no. 4, pp. 88–97, 2024.

[11] M. Bansal and R. Gupta, "Cyber Attack Detection Using Ensemble Learning Models," *International Journal of Information Technology*, vol. 15, pp. 1123–1130, 2023.

[12] L. Chen et al., "Improved Random Forest Model for Intrusion Detection," *Computers & Security*, vol. 122, 2023.

[13] S. Patel and R. Mehta, "Machine Learning for Cyber Threat Detection in Modern Networks," *Cybersecurity Journal*, vol. 7, no. 1, pp. 55–70, 2022.

[14] K. Reddy and S. Rao, "Hybrid Machine Learning Intrusion Detection System for IoT Networks," *IEEE Network Security*, vol. 4, no. 1, pp. 1–9, 2024.

[15] A. Verma et al., "Real-Time Intrusion Detection Using Ensemble Learning Techniques," *International Journal of Advanced Computing*, 2025.

