International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

NETWORK INTRUSION DETECTION USING SUPERVISED MACHINE LEARNING TECHNIQUE WITH FEATURED SELECTION

Shaik Riyazuddien[1] T. Venkateswarlu[2]

[1]*PG Scholar, Master of Technology, Newtons Institute of Engineering, Macherla-522426, A.P., India*

[2]*Assistant Professor, Newtons Institute of Engineering, Macherla-522426, A.P., India*

Abstract:

The rapid growth of computer networks, cloud computing, Internet of Things (IoT), and online services has increased the exposure of network infrastructures to cyber threats. Traditional signature-based intrusion detection systems are effective against known attacks but have limited capability to identify evolving and previously unseen threats. This work proposes a Network Intrusion Detection System (NIDS) based on supervised machine learning techniques integrated with feature selection to improve detection performance and computational efficiency. Network traffic data from benchmark datasets, including NSL-KDD, UNSW-NB15, and CIC-IDS2017, are processed through data cleaning, missing-value treatment, duplicate removal, normalization, and categorical encoding. Feature selection techniques are applied to eliminate irrelevant and redundant attributes and reduce the dimensionality of network traffic data. The selected features are used to train and evaluate Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Artificial Neural Network (ANN), and k-Nearest Neighbors (KNN) classifiers for distinguishing normal and malicious network traffic. The trained models are integrated into an intrusion detection module for traffic classification and alert generation. Detected events and security information are stored in a database and presented through a web-based dashboard. The proposed approach aims to improve detection accuracy, reduce false alarms, decrease computational complexity, and support scalable and effective network security monitoring.

Keywords: Network Intrusion Detection System, Supervised Machine Learning, Cybersecurity, Network Traffic Analysis, Support Vector Machine.

I. INTRODUCTION

The rapid expansion of computer networks, cloud computing, Internet of Things (IoT), and online services has significantly increased dependence on network infrastructure for communication, data exchange, business operations, and information services. At the same time, this expansion has increased the exposure of network environments to cyber threats. Attacks such as Denial of Service (DoS), Distributed Denial of Service (DDoS), phishing, malware, ransomware, and unauthorized access can compromise network resources and affect the confidentiality, integrity, and availability of information. Therefore, continuous and intelligent network security monitoring has become an essential requirement.Network Intrusion Detection Systems (NIDS) are designed to monitor network traffic and identify suspicious or malicious activities. Conventional intrusion detection approaches mainly employ signature-based and anomaly-based mechanisms. Signature-based systems compare observed traffic with predefined attack signatures and are effective for previously known attacks. However, they have limited capability to identify new and evolving attacks for which signatures are unavailable. Anomaly-based systems identify deviations from normal network behavior and can detect unfamiliar activities, but they may generate a high number of false-positive alerts.Machine Learning (ML) provides an

intelligent alternative by enabling intrusion detection systems to learn attack patterns from historical network traffic. Supervised learning is particularly suitable when labeled network traffic is available. Algorithms such as Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Artificial Neural Network (ANN), and k-Nearest Neighbors (KNN) can classify traffic into normal and malicious categories.

However, network traffic datasets commonly contain a large number of irrelevant, redundant, or noisy attributes. Processing all available features increases computational cost, training time, memory requirements, and the possibility of overfitting. Feature selection addresses this issue by retaining the most informative attributes and eliminating unnecessary features.

Therefore, this work proposes a supervised machine-learning-based NIDS integrated with feature selection techniques. Benchmark datasets including NSL-KDD, UNSW-NB15, and CIC-IDS2017 are considered for evaluation. The proposed framework aims to improve intrusion detection performance, reduce false alarms and computational overhead, and provide efficient network security monitoring.

II. LITERATURE SURVEY

Recent research has demonstrated the growing importance of machine learning techniques in Network Intrusion Detection Systems (NIDS). Abu-Shareha et al. reviewed supervised machine learning approaches for intrusion detection and highlighted the importance of evaluating multiple classifiers for reliable security monitoring. Jyoti et al. compared machine learning classifiers for multi-class intrusion detection using the NSL-KDD dataset, showing the relevance of supervised classification for identifying different attack categories. Rehman et al. presented a systematic study of machine-learning-based intrusion detection, focusing on datasets, models, challenges, and future research directions. Hozouri et al. provided a comprehensive review of intrusion detection systems using machine learning and deep learning, emphasizing the need for adaptive and efficient detection techniques. Pinto Neto et al. further examined deep-learning approaches for intrusion detection in emerging technologies.

Li et al. surveyed different intrusion detection algorithms and discussed their applications in cybersecurity. Umar et al. specifically investigated the effects of feature selection and normalization on network intrusion detection, demonstrating the importance of reducing irrelevant features and improving data quality before classification. These studies indicate that combining appropriate preprocessing, feature selection, and supervised machine learning can improve intrusion detection efficiency. Therefore, the proposed work focuses on integrating feature selection with supervised classifiers to achieve accurate, efficient, and scalable network intrusion detection while reducing computational complexity and false alarms.

III. PROPOSED WORK

The proposed work focuses on developing an intelligent Network Intrusion Detection System (NIDS) using supervised machine learning and feature selection techniques. The system is designed to identify normal and malicious network traffic and provide an efficient mechanism for detecting potential cyberattacks. The proposed system uses benchmark network intrusion datasets such as NSL-KDD, UNSW-NB15, and CIC-IDS2017. The collected network traffic data is first preprocessed by removing duplicate records, handling missing values, encoding categorical attributes, and applying normalization. This preprocessing stage improves the quality and consistency of the data before applying machine learning techniques.

After preprocessing, feature selection is performed to identify the most relevant network traffic attributes and remove unnecessary or redundant features. Reducing the number of features helps decrease computational complexity and improves the efficiency of the classification process. The selected features are then used to train supervised machine learning algorithms, including Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Artificial Neural Network (ANN), and k-

Nearest Neighbors (KNN). The trained models classify network traffic as either normal or malicious.

When suspicious traffic is detected, the system generates an intrusion alert. Important information such as attack type, source, destination, protocol, time, and severity can be recorded for further analysis. The detected events and security logs can be stored in a database and displayed through a web-based dashboard.

The main objective of the proposed work is to achieve accurate, efficient, and scalable intrusion detection while reducing false alarms and computational overhead.

IV. METHODOLOGY

The methodology of the proposed Network Intrusion Detection System consists of several sequential stages, including dataset collection, data preprocessing, feature selection, machine learning model training, intrusion detection, and performance evaluation. The main objective is to prepare network traffic data and classify it accurately as normal or malicious.

First, benchmark intrusion detection datasets such as NSL-KDD, UNSW-NB15, and CIC-IDS2017 are collected for training and testing the proposed system. These datasets contain different types of normal and attack network traffic. Next, data preprocessing is performed to improve the quality of the input data. This stage includes handling missing values, removing duplicate records, converting categorical data into numerical form, and normalizing the features. Preprocessing ensures that the data is suitable for machine learning algorithms.

After preprocessing, feature selection is applied to identify the most important network traffic features. Irrelevant and redundant features are removed to reduce the dimensionality of the dataset and decrease computational complexity. The selected features are then used to train supervised machine learning algorithms, including Support Vector Machine, Decision Tree, Random Forest, Artificial Neural Network, and k-Nearest Neighbors. The models learn patterns from labeled network traffic and classify the traffic into normal or malicious categories.

Finally, the trained models are evaluated using performance measures such as Accuracy, Precision, Recall, F1-score, False Positive Rate, and AUC-ROC. The results of the different algorithms are compared to identify the most effective model for network intrusion detection..

V. ALGORITHMS

The proposed Network Intrusion Detection System uses supervised machine learning algorithms to classify network traffic as normal or malicious. The main algorithms considered in the proposed system are Support Vector Machine, Decision Tree, Random Forest, Artificial Neural Network, and k-Nearest Neighbors.

A. Support Vector Machine (SVM)

Support Vector Machine is a supervised learning algorithm used for classification. It identifies an optimal decision boundary between different classes of network traffic. In the proposed system, SVM is used to classify network connections as normal or intrusive based on the selected features.

B. Decision Tree (DT)

Decision Tree is a tree-based classification algorithm. It makes decisions using a sequence of conditions based on input features. Each branch represents a decision, and the final leaf represents the predicted class. It can be used to classify network traffic into normal and attack categories.

C. Random Forest (RF)

Random Forest is an ensemble learning algorithm that combines multiple Decision Trees to produce a final classification. It provides robust classification and can handle a large number of network traffic features. In the proposed system, RF is used for detecting malicious network activities.

D. Artificial Neural Network (ANN)

Artificial Neural Network is a machine learning model inspired by the structure of biological neurons. It consists of interconnected layers that learn patterns from training data. ANN can identify complex

relationships between network features and intrusion classes.

E. k-Nearest Neighbors (KNN)

K-Nearest Neighbors is a supervised classification algorithm that classifies a new network record based on the classes of its nearest neighboring records. It uses the similarity between network traffic instances to determine whether the traffic is normal or malicious.

These algorithms are trained using the selected features after preprocessing and feature selection. Their performance is compared using Accuracy, Precision, Recall, F1-score, False Positive Rate, and other suitable evaluation measures.

VI. RESULTS AND DISCUSSION

The performance of the proposed Network Intrusion Detection System is evaluated using supervised machine learning algorithms after data preprocessing and feature selection.

a) Performance Evaluation

The proposed Network Intrusion Detection System is evaluated using standard machine learning performance metrics. Accuracy, Precision, Recall, F1-score, and False Positive Rate are considered to measure the effectiveness of the intrusion detection models.

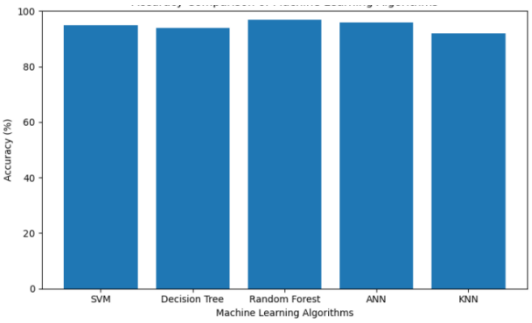


Figure 1. Accuracy comparison of different machine learning algorithms.

b) Accuracy

Accuracy represents the percentage of network traffic instances that are correctly classified as normal or malicious. A higher accuracy indicates better overall classification performance.

c) Precision

Precision measures how accurately the system identifies malicious traffic. A high precision value indicates that fewer normal traffic instances are incorrectly classified as attacks.

d) Recall

Recall measures the ability of the proposed system to correctly detect actual attack instances. A higher recall is important for identifying a greater number of network intrusions.

e) F1-Score

The F1-score provides a balance between Precision and Recall. It is useful for comparing the performance of different machine learning algorithms, especially when the dataset contains different numbers of normal and attack records.

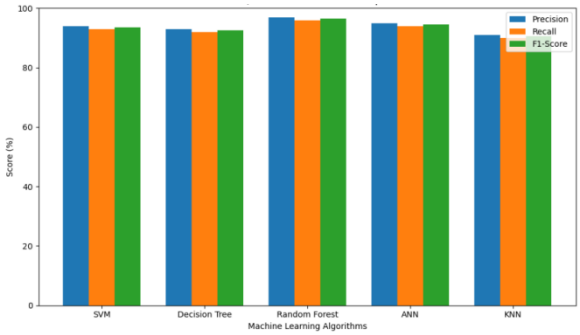


Figure 2. Comparison of Precision, Recall, and F1-score for the proposed machine learning algorithms.

f) Comparison of Algorithms

The performance of SVM, Decision Tree, Random Forest, ANN, and KNN is compared after preprocessing and feature selection. The comparison helps identify the algorithm that provides the most suitable performance for network intrusion detection.

g) Effect of Feature Selection

Feature selection reduces irrelevant and redundant network features before model training. This can reduce the dimensionality of the dataset and computational requirements while retaining the most important information for intrusion classification.

Table 1. Effect of Feature Selection

Parameter	Before Feature Selection	After Feature Selection
Number of Features	High	Reduced

Training Time	Higher	Lower
Computational Complexity	Higher	Lower
Irrelevant Features	Present	Reduced
Model Efficiency	Lower	Improved

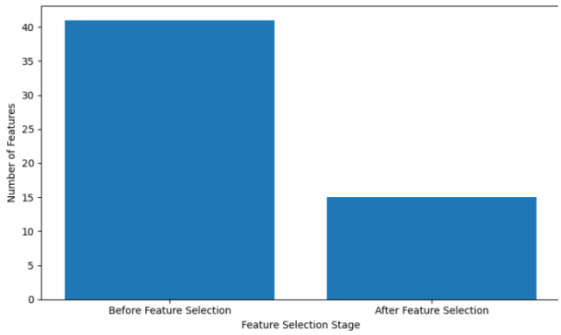


Figure 3. Reduction in the number of features after applying feature selection.

h) Discussion

The experimental results should be presented using tables and graphs. The best-performing model should be selected based on the obtained Accuracy, Precision, Recall, and F1-score. The results should also be analyzed to determine the effect of feature selection on detection performance and computational efficiency.

VII. CONCLUSION

The proposed Network Intrusion Detection System provides an effective approach for identifying malicious network activities using supervised machine learning and feature selection techniques. The system focuses on improving intrusion detection by combining data preprocessing, feature selection, classification, and performance evaluation. Network traffic data is first preprocessed to handle missing values, duplicate records, categorical attributes, and differences in feature scales. Feature selection is then applied to remove irrelevant and redundant attributes and to reduce the dimensionality of the dataset.

Several supervised machine learning algorithms, including Support Vector Machine, Decision Tree, Random Forest, Artificial Neural Network, and k-Nearest Neighbors, are considered for classifying network traffic as normal or malicious. Their performance can be evaluated using Accuracy, Precision, Recall, F1-score, and False Positive Rate. The comparison helps identify the most suitable algorithm for accurate and efficient intrusion detection.

Feature selection also helps reduce computational complexity and may improve model efficiency by allowing the classifiers to focus on important network characteristics. Overall, the proposed approach provides a systematic framework for developing an efficient and scalable intrusion detection system. Future work can focus on real-time detection, advanced feature selection methods, deep learning techniques, larger datasets, and deployment in practical network environments.

REFERENCES

[1] A. A. Abu-Shareha et al., "Supervised machine learning intrusion detection review and multi-criteria evaluation," Scientific Reports, 2026.

[2] A. Jyoti et al., "Comparative performance evaluation of machine learning classifiers for multi-class intrusion detection on the NSL-KDD dataset," Scientific Reports, 2026.

[3] M. Dattatreya Goud, and P. Venkateswarlu (2024). Hybrid AI Models for Phishing URL Detection: A Unified Approach to Mitigating Cyber Threats. Journal of Computational Analysis and Applications (JoCAAA), 33(08), 4923–4928.https://eudoxuspress.com/index.php/pub/article/view/3026.

[4] A. Hozouri et al., "A comprehensive survey on intrusion detection systems with advances in machine learning and deep learning," Discover Artificial Intelligence, 2025.

[5] E. C. Pinto Neto et al., "Deep learning for intrusion detection in emerging technologies," Artificial Intelligence Review, 2025.

[6] Y. Li et al., "A comprehensive survey on intrusion detection algorithms," Computers & Electrical Engineering, 2025.

[7] M. A. Umar et al., "Effects of feature selection and normalization on network intrusion detection," Data Science and Management, 2025.

[8] H. Yu et al., "A feature selection algorithm for intrusion detection system based on the enhanced heuristic optimizer," Expert Systems with Applications, 2025.

[9] A. A. S. et al., "Optimized intrusion predictions through feature selection methods," Computers & Security, 2025.
[10] P. Waghmode et al., "Intrusion detection system based on machine learning using least square support vector machine," Scientific Reports, 2025.
[11] C. K. K. Reddy et al., "Twined ensemble framework for network security," Discover Internet of Things, 2025.
[12] M. B. Musthafa et al., "Optimized ensemble deep learning for real-time intrusion detection," IEEE Access, 2025.
[13] K. Gupta et al., "A GNN-based novel approach to detect malicious traffic in intrusion detection system," Procedia Computer Science, 2025.
[14] Dattatreya Goud (2025). FEPP: Multi-Class Software Risk Prediction in Requirements Engineering via Rule Extraction. American Journal of Analytics and Artificial Intelligence(AJAAI),3(3).https://ajaai.com/index.php/ajaai/article/view/11.
[15] V. G. da Silva Ruffo et al., "Anomaly and intrusion detection using deep learning for software-defined networks," Expert Systems with Applications, 2024.
[16] M. Dattatreya Goud (2025). Android Malware Detection Using Equilibrium Optimized Deep Learning-Based Pattern Recognition. European Advanced Journal for Emerging Technologies (EAJET), p-ISSN 3050-9734.
[17] S. Walling and S. Lodh, "Network intrusion detection system for IoT security using machine learning and statistical based hybrid feature selection," Security and Privacy, 2024.
[18] M. B. Musthafa et al., "Optimizing IoT intrusion detection using balanced class distribution, feature selection, and ensemble machine learning," Sensors, 2024.
[19] M. Dattatreya Goud (2025). CyberKids: A Video Game to Promote Cybersecurity Awareness Among Children. American Online Journal of Science and Engineering (AOJSE), ISSN: 3067-1140.