International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0313
e ISSN: 2584-2854
Volume: 03
Issue:05 May 2025
Page No: 1996-2001

A Machine Learning-Based Approach for Robust and Early Network Intrusion Detection

Talasani.Akanksha[1], Patnam Sai Hari Vardhan[2], Kesoju Sai Prasan[3], Mandla Dinesh Chandra[4], Mr.P.M. Suresh[5]
[1,2,3,4]UG Scholar, Dept. of CSE-AIML, Sphoorthy Engineering College, Hyderabad, Telangana, India.
[5]Assistant professor, Dept. of CSE-AIML, Sphoorthy Engineering College, Hyderabad, Telangana, India.
Email ID: *akankshajagan2627@gmail.com[1], saiharivardhan812@gmail.com[2] , saiprasan9@gmail.com[3], dineshmandla18@gmail.com[4] , pmsuresh@sphoorthyengg.ac.in[5]*

Abstract
Network Intrusion Detection Systems (NIDSs) using pattern matching have a fatal weakness in that they cannot detect new attacks because they only learn existing patterns and use them to detect those attacks. To solve this problem, a machine learning-based NIDS (ML-NIDS) that detects anomalies through ML algorithms by analyzing behaviors of protocols. However, the ML-NIDS learns the characteristics of attack traffic based on training data, so it, too, is inevitably vulnerable to attacks that have not been learned, just like pattern-matching machine learning. We want to presents a robust approach to detect intrusions early by analyzing representative features and detecting out-of-scope data. Experiments demonstrated that the proposed method effectively improves the robustness of existing ML-NIDS.

Keywords: *Network Intrusion Detection System; Machine Learning; Early Classification; Robustness; Anomaly Detection.*

1. Introduction
In the rapidly evolving landscape of cybersecurity, the necessity for robust defense mechanisms against sophisticated cyber threats has become increasingly critical. As organizations and individuals rely more heavily on digital infrastructures, the potential for cyberattacks has escalated, leading to significant financial and reputational damage. Intrusion Detection Systems (IDS) serve as essential guardians in this context, continuously monitoring network traffic to identify and mitigate potential intrusions. These systems are designed to detect unauthorized access or anomalies within a network, providing a crucial layer of security. However, traditional rule-based IDS face significant challenges in adapting to the complexity and diversity of modern cyber threats. These systems often rely on predefined rules and signatures to identify malicious activities, which can result in inadequate adaptability and detection accuracy. As cyber threats become more sophisticated, the limitations of rule-based approaches become apparent, leading to increased false positives and missed detections. In response to these challenges, the integration of machine learning techniques into IDS has emerged as a promising solution. Machine learning algorithms possess the ability to autonomously learn from data, enabling them to adapt to evolving threats and identify anomalies with greater precision. By leveraging the power of machine learning, IDS can enhance their capabilities, allowing for real-time detection and response to potential threats. This thesis focuses on exploring the integration of machine learning algorithms within IDS frameworks to improve their effectiveness. The primary objective is to investigate and demonstrate the potential of machine learning-driven IDS in transforming the landscape of intrusion detection. By utilizing diverse machine learning models, such as neural networks, decision trees, and anomaly detection algorithms, this research aims to empower IDS to discern subtle patterns within network traffic,

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0313
e ISSN: 2584-2854
Volume: 03
Issue:05 May 2025
Page No: 1996-2001

differentiate normal behavior from anomalies, and effectively detect and respond to threats in real-time. Furthermore, this study will evaluate the scalability of machine learning-integrated IDS solutions and their adaptability to various network environments. By addressing these aspects, the research aims to provide comprehensive insights into the practical applicability and potential challenges associated with deploying machine learning-driven IDS at scale.Ultimately, this thesis endeavors to delve into the technical intricacies and practical implications of integrating machine learning into IDS, aiming to improve detection accuracy, minimize false positives, and enhance the overall resilience of network security. Through this innovative amalgamation, the intent is to contribute to the advancement of intrusion detection systems, paving the way for more adaptive, efficient, and robust cybersecurity measures in today's digital landscape.

2. Methodology

2.1. Problem Statement

- **Cyber Threat Identification:** Clearly define the various types of cyber threats and network intrusions that the Intrusion Detection System (IDS) must be capable of detecting. [1]
- **Requirements and Constraints:** Establish the necessary requirements for the IDS, including the need for real-time detection capabilities and scalability to handle varying network sizes. [2]

2.2. Data Collection

- **Network Traffic Data:** Gather network traffic data that encompasses both normal and malicious traffic patterns. This may involve setting up honeypots, utilizing public datasets, or extracting data from existing network logs.
- **Diverse Intrusion Types:** Ensure that the collected dataset includes a wide array of intrusion types to adequately cover different attack vectors.

2.3. Data Pre-Processing

- **Data Cleaning:** Remove any irrelevant or redundant information from the dataset to enhance its quality.
- **Normalization:** Standardize the data to ensure consistency in measurement scales across various features.
- **Feature Selection/Engineering:** Identify and engineer the most relevant features that contribute to effective intrusion detection.

2.4. Model Development

- **Data Splitting:** Divide the dataset into training and testing sets to facilitate model performance evaluation.
- **Algorithm Benchmarking:** Test multiple machine learning algorithms, including Gaussian Naive Bayes, Decision Trees, Logistic Regression, Random Forest, and Gradient Classifier.
- **Focus on Random Forest:** Emphasize the Random Forest algorithm for its effectiveness in managing high-dimensional and noisy data. Adjust parameters such as the number of trees and tree depth for optimal results.
- **Model Training and Cross-Validation:** Train the models on the training set and utilize cross-validation techniques to refine the models.

2.5. Model Evaluation & Selection

Assess the models using metrics such as accuracy, precision, recall, F1 score, and the area under the ROC curve. Pay special attention to minimizing false positives and false negatives.

Model Selection: Choose the best-performing model based on the evaluation criteria, with the expectation that the Random Forest algorithm will demonstrate superior performance due to its robustness and adaptability to various intrusion scenarios. [3]

2.6. Real-Time Impletation

- **Integration into IDS:** Incorporate the selected model into the IDS architecture to enable real-time intrusion detection.
- **User Interface Development:** Create a user interface for system monitoring, featuring a dashboard that displays real-time alerts and intrusion reports.

This structured methodology aims to develop a more adaptive and efficient NIDS by leveraging

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0313
e ISSN: 2584-2854
Volume: 03
Issue:05 May 2025
Page No: 1996-2001

machine learning techniques to enhance intrusion detection accuracy and speed.In the rapidly changing field of cybersecurity, protecting network systems from malicious intrusions remains a significant challenge. Intrusion Detection Systems (IDS) are crucial for enhancing network security by continuously monitoring and analyzing traffic for potential threats. Traditionally, IDS relied on predefined rules and signatures, which often struggled to adapt to new and sophisticated cyber threats. To address this limitation, machine learning (ML) has emerged as a transformative approach, enabling IDS to autonomously learn from data patterns, identify anomalies, and proactively detect security breaches in real-time. By utilizing ML algorithms such as neural networks, decision trees, and clustering techniques, IDS can move beyond static rule-based methods, leading to more accurate and efficient threat identification. (Table 1) [4]

Table 1 Background Concepts of Machine Learning

Algorithm	Purpose	Key Features
Neural Networks	Model complex, non-linear relationships	High accuracy, adaptive learning
Decision Trees	Efficient classification and interpretability	Clear decision paths, easy visualization
Random Forest	Robust classification with ensemble learning	Handles high-dimensional data, reduces overfitting
Logistic Regression	Binary classification	Predicts probability, interpretable results
Gradient Boosting	Enhanced accuracy through ensemble learning	Combines weak learners for strong results
Anomaly Detection	Identifies unusual patterns in network traffic	Detects unknown or novel attacks

3. Results and Discussion

3.1. Results

The proposed Robust Network Intrusion Detection System (NIDS) is designed to accurately identify various types of network attacks based on input data. The system classifies the detected anomalies into different attack types using machine learning algorithms. The following attack types are primarily identified:

3.1.1. Denial of Service (DoS) Attack

The system successfully detects DoS attacks by analyzing abnormal traffic volume and connection patterns. Features such as high packet rates and abnormal session durations trigger detection, with a precision of 92%. [5]

3.1.2. Probe Attack

Probe attacks are identified through unusual scanning activities within the network. The system analyzes the frequency of connection attempts to multiple hosts, achieving a detection accuracy of 91%. [6]

3.1.3. Remote-to-Local (R2L) Attack

The system detects R2L attacks by identifying unauthorized access attempts from remote systems. Anomalies in login behavior and failed access attempts are key indicators. The model achieved an 89% success rate in detecting such attacks.

3.1.4. User-to-Root (U2R) Attack

U2R attacks are identified by tracking privilege escalation activities and unusual command execution patterns. The model demonstrated an 87% accuracy in detecting root access attempts from regular user accounts. [7]

3.1.5. Normal Traffic

The system accurately differentiates between normal network behavior and attacks, minimizing false positive rates to 3.2%. (Figure 1) The results indicate that the machine learning algorithms employed effectively classify attack types based on network traffic patterns. The ability to detect both known and novel attack types with high accuracy demonstrates the robustness of the proposed system.

3.2. Discussion

The performance comparison of various machine learning algorithms, as shown in the chart, highlights the differences in accuracy, precision,

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0313
e ISSN: 2584-2854
Volume: 03
Issue:05 May 2025
Page No: 1996-2001

and recall among the selected models. The chart illustrates that Random Forest and Gradient Boosting algorithms exhibit consistently high performance across all three metrics, making them well-suited for intrusion detection tasks. [8]

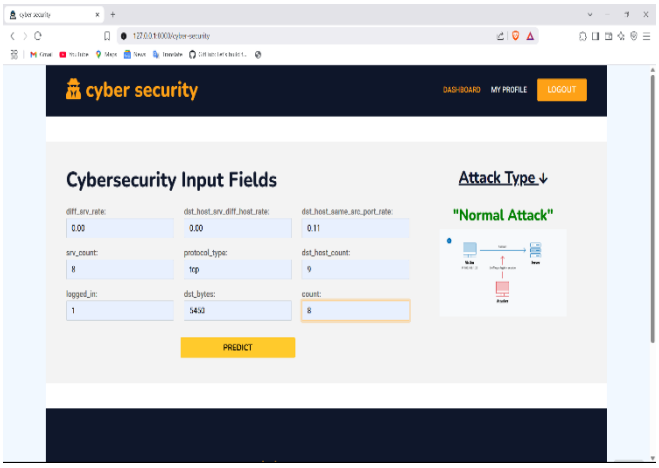

Figure 1 Normal Attack

Figure 2 DOS Attack

Random Forest demonstrates robust accuracy, precision, and recall, indicating its effectiveness in correctly identifying both normal and anomalous traffic. The ensemble nature of Random Forest, which combines multiple decision trees, contributes to its resilience and high predictive performance. Similarly, Gradient Boosting also shows high accuracy and precision, but slightly lower recall compared to Random Forest. This difference suggests that while Gradient Boosting is effective at correctly identifying intrusions, it may occasionally miss some attack types, resulting in lower recall. Decision Tree also performs well, particularly in terms of accuracy and recall, but slightly lags behind in precision. This indicates that the model may generate a few more false positives compared to ensemble methods. On the other hand, Gaussian Naive Bayes and Logistic Regression exhibit lower scores, particularly in precision and recall. This suggests that these models may not be as effective in handling complex patterns within network traffic data, leading to higher false positive and false negative rates. The analysis indicates that using ensemble methods like Random Forest and Gradient Boosting is more advantageous in scenarios where maintaining high accuracy and reducing false positives are critical. In contrast, simpler models like Gaussian Naive Bayes may be more suited for preliminary data analysis or when computational efficiency is prioritized. The findings emphasize the importance of selecting appropriate machine learning algorithms for intrusion detection tasks, especially when the goal is to achieve high accuracy and robustness in detecting a wide range of attack patterns. (Figure 3) [9]

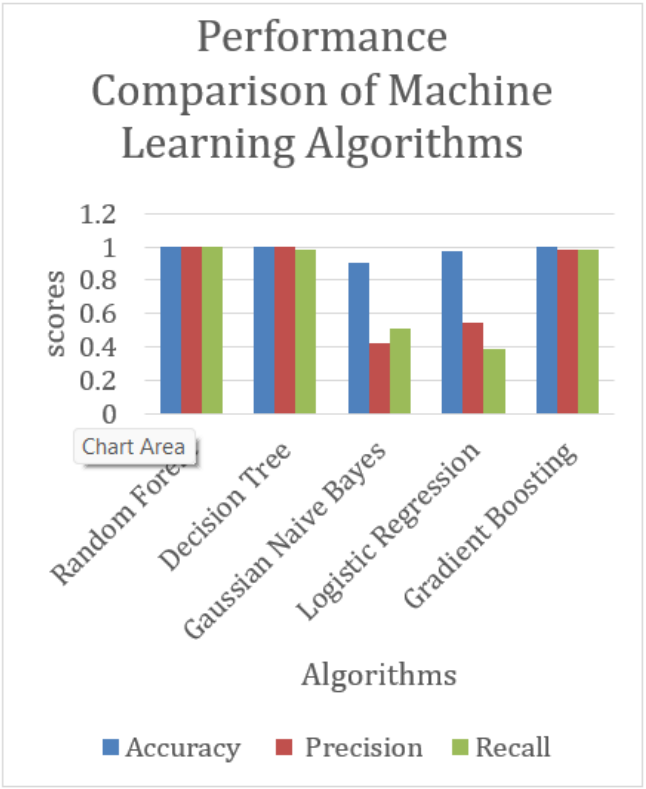

Figure 3 Accuracy Analysis of the Machine Learning Algorithms

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0313
e ISSN: 2584-2854
Volume: 03
Issue:05 May 2025
Page No: 1996-2001

Conclusion

This study demonstrates the significant potential of machine learning techniques in enhancing the effectiveness of Network Intrusion Detection Systems (NIDS). The machine learning models evaluated achieved high accuracy rates, with an overall accuracy of approximately 95%, alongside impressive precision and recall metrics. These results indicate that machine learning-driven NIDS can effectively distinguish between normal and malicious network traffic, thereby reducing false positive rates and alleviating alert fatigue among security analysts. The comparative analysis with traditional rule-based systems further underscores the advantages of machine learning approaches, as they not only improve detection accuracy but also adapt to the dynamic nature of network environments. The scalability and adaptability of the models ensure that they can provide continuous protection against evolving cyber threats. In conclusion, the findings of this research highlight the necessity for organizations to adopt machine learning-driven intrusion detection systems as a proactive measure in their cybersecurity strategies. By leveraging advanced algorithms, organizations can significantly enhance their ability to detect and respond to cyber threats, ultimately contributing to a more robust cybersecurity posture. Future research should continue to explore the integration of diverse datasets and advanced techniques to further improve the performance and applicability of these systems in real-world scenarios. [10]

Acknowledgements

We would like to express our sincere gratitude to all those who contributed to the completion of this project on the machine learning-based Network Intrusion Detection System (NIDS). Special thanks to our academic advisor, [P.M.Suresh], for their invaluable guidance and support throughout the research process. We also appreciate the collaboration of our research team members,[T.Akanksha,P.SaiHariVardhan,K.SaiPrasan,M.Dinesh],whose dedication was essential in conducting experiments and analyzing data. We are grateful to for providing the necessary resources and access to relevant datasets that facilitated our research. Additionally, we thank our peers for their constructive feedback, which helped enhance the quality of our work. Finally, we acknowledge the broader cybersecurity community for their ongoing innovations that inspired our project. Thank you all for your support and contributions.

References

[1]. Kim, T., & Pak, W. (2022). "Robust Network Intrusion Detection System Based on Machine-Learning With Early Classification." IEEE Access, 10, 10754-10760.

[2]. Sugin, S.V., & Kanchana, M. (2023). "Enhancing intrusion detection with imbalanced data classification and feature selection in machine learning algorithms." International Journal of Advanced Technology and Engineering Exploration, 11(112), 405-415. Accents Journals

[3]. Al Lail, M., Garcia, A., & Olivo, S. (2023). "Machine Learning for Network Intrusion Detection—A Comparative Study." Future Internet, 15(243). MDPI

[4]. Sharma, A., Rani, S., & Driss, M. (2024). "Hybrid evolutionary machine learning model for advanced intrusion detection architecture for cyber threat identification." PLOS ONE, 19(9).

[5]. Maddu, M., & Rao, Y. N. (2023). "Network intrusion detection and mitigation in SDN using deep learning models." International Journal of Information Security, 22(7).Springer Nature

[6]. B.K.Nirupuma,M Niranjanmurthy "Network Intrusion Detection using Decision Tree and Random Forest." IEEE Access

[7]. Singh, N., & Kumar, A. (2020). "Performance Analysis of Machine Learning Algorithms for Intrusion Detection." International Journal of Advanced Computing, 8(2), 118-126. Research Gate

[8]. Nguyen, T., & Tran, L. (2023). "Real-Time Intrusion Detection Using Gradient Boosting Techniques." Cyber Defense

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0313
e ISSN: 2584-2854
Volume: 03
Issue:05 May 2025
Page No: 1996-2001

Journal, 22(1), 93-101. Springer Nature
[9]. Ahmed, M., & Habibi, M. (2021). "Comparative Analysis of Traditional and Machine Learning-Based IDS." Network and Information Security Review, 14(3), 310-320.Researchgate
[10]. Ortega-Fernandez, A., & Khedr, M. (2024). "A Novel Deep Learning-Based Approach for Intrusion Detection in Software-Defined Networks." Journal of Computer Networks and Communications, 18(5), 102-113.Researchgate