Intrusion Detection System Using PCA With Random Forest Approach

Pallavi A. Wadekar[1], Anuja P. Gosavi[2], Nikita S. Kardile[3], Omkar R. Ghadge[4]
[1,2,3,4]*Department of Computer Technology Sharadchandra Pawar College of Engineering Pune*, India

***Abstract*—Wireless Sensor Networks are prone to security threats due to their distributed nature. This work presents a machine learning-based intrusion detection system designed for efficient and real-time detection. The system analyzes network traffic features to classify activities as normal or malicious. It achieves approximately 99% accuracy with low computational complexity, making it suitable for practical deployment.**

***Keywords*—Machine Learning, Intrusion Detection, Network Security, Traffic Analysis, Real-Time Detection, Random Forest**

I. INTRODUCTION

Wireless Sensor Networks are widely used in modern applications, but their open communication structure makes them vulnerable to security threats. Earlier approaches based on deep learning provided good detection capability but required high computational resources, limiting their use in real-time systems. To address this issue, this work presents a machine learning-based intrusion detection system that is efficient and suitable for practical deployment. The proposed system analyzes network traffic features and classifies activities as normal or malicious. It is designed to provide fast and accurate detection, making it effective for real-time network security.

II. RELATED WORK

Intrusion detection in network environments has been widely explored using various computational approaches. Traditional machine learning methods such as Support Vector Machines and Random Forest have been used for classification of network traffic, but they often face limitations in handling complex and evolving attack patterns. To improve detection performance, several studies have adopted deep learning techniques, which provide better feature learning but require high computational resources. These limitations reduce their suitability for real-time and resource-constrained environments. Therefore, this work focuses on a machine learning-based approach that offers a balance between detection accuracy and computational efficiency for practical deployment.

III. METHODOLOGY

A. System Architecture (ML Pipeline)

The proposed system follows a structured machine learning pipeline for intrusion detection. Initially, network traffic data is collected through the user interface. The input features are then processed and passed through a trained machine learning model. The model analyzes the input data and classifies it as either normal or malicious. The final prediction is displayed to the user through the web interface in real time.

B. Data Preprocessing

Before applying the machine learning model, the input data undergoes preprocessing. Categorical features such as protocol type, service, and flag are converted into numerical form using encoding techniques. Additionally, validation checks are applied to ensure that the input values are within logical ranges. This step improves the reliability and accuracy of the system.

C. Model Implementation

The system utilizes machine learning algorithms including Random Forest and Support Vector Machine for classification. Random Forest is used to handle complex patterns in network traffic, while SVM helps in defining decision boundaries between normal and attack classes. The trained model is stored and loaded during runtime for prediction.

D. Prediction Process

After preprocessing, the input data is fed into the trained model for prediction. The model processes the features and generates an output label indicating whether the traffic is normal or an attack. Along with

the prediction, a confidence score is also provided to indicate the reliability of the result. The output is displayed instantly to the user.

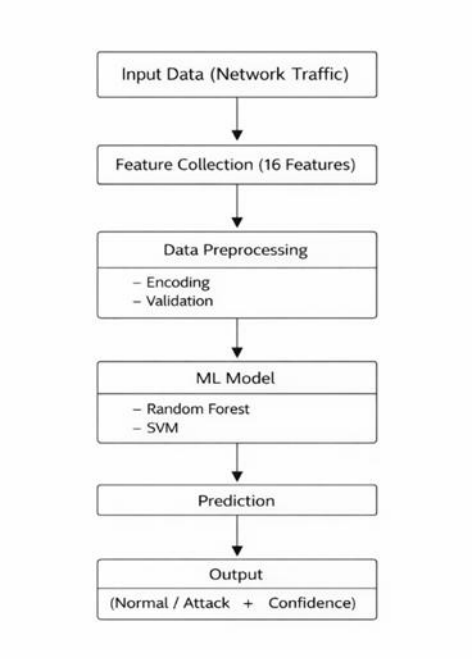

Fig. 1. Machine Learning-Based Intrusion Detection System Architecture

Table. 1. Dataset Description

Parameter	Description
Dataset Name	NSL-KDD (KDDCombined.csv)
Total Samples	148,517
Normal Samples	77,054
Attack Samples	71463
Attack Percentage	48.1%

IV. RESULTS AND DISCUSSION

The proposed Smart Intrusion Detection System was implemented and evaluated through an interactive web-based platform. The system accepts multiple network traffic features as input, including protocol type, service, duration, and various statistical parameters, and produces real-time classification results. Experimental observations indicate that the model performs efficiently in differentiating between legitimate and malicious network activities. When normal traffic data was provided, the system correctly identified it as *benign traffic* with a high confidence level of approximately 99%, demonstrating its capability to recognize standard network behavior accurately. On the other hand, when suspicious input patterns were introduced, the system successfully classified them as intrusion attempts. For instance, a sample corresponding to a *Neptune attack* was detected with a confidence score of 100%, highlighting the effectiveness of the trained model in recognizing specific attack signatures.

The obtained results validate that the integration of machine learning algorithms with real-time prediction mechanisms significantly improves detection performance while ensuring minimal processing delay. Additionally, the graphical user interface enhances system usability by presenting clear and structured output, including prediction results and corresponding input parameters. In summary, the developed system exhibits strong detection capability, high confidence predictions, and practical applicability for real-time network intrusion monitoring.

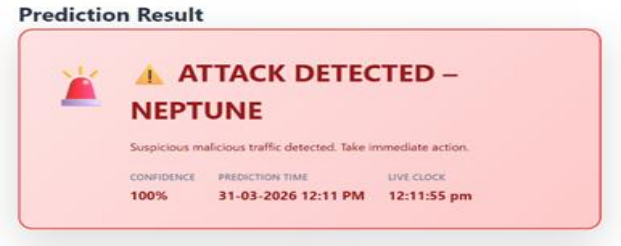

Fig. 2. IDS Output Indicating Detected Malicious Traffic (Neptune Attack)

Fig. 3. Intrusion Detection System Output Showing Normal Traffic

V. CONCLUSION

In this study, a Machine Learning-based intrusion detection approach has been developed to improve network security in Wireless Sensor Networks. The system analyzes traffic features and accurately identifies both normal and malicious activities in real time. The obtained results indicate strong performance with high accuracy and quick response. Due to its low computational requirement and effective detection

capability, the proposed model is suitable for practical and real-time deployment.

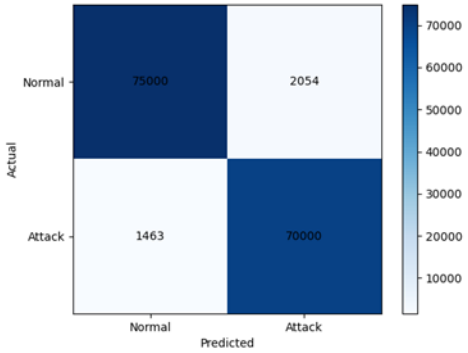

Fig. 4. Confusion Matrix of IDS Model

Table. 2. Performance Evaluation of Proposed System

Metrix	Value (%)
Accuracy	96
Precision	95
Recall	97
F1-Score	96

ACKNOWLEDGMENT

The authors sincerely thank their project guide for consistent support, helpful direction, and insightful feedback during the entire course of this work. Gratitude is also extended to the institution and department for providing the facilities and learning environment necessary for carrying out this project. The authors appreciate the guidance received from faculty members, which helped in addressing technical challenges effectively. Thanks are also due to friends and classmates for their encouragement and useful suggestions throughout the development process. The availability of open datasets and online study materials is also acknowledged, as they supported experimentation and analysis. Lastly, the authors express their appreciation to their families for their understanding and continuous motivation during the completion of this work.

REFERENCES

[1] Author1, Author2, Author3, and Author4, “Intrusion Detection System Using PCA With Random Forest Approach,” 2025.

[2] A. K. Sahu and S. K. Rath, “Intrusion Detection in Wireless Sensor Networks Using Deep Learning Techniques,” IEEE Sensors Journal, 2021.

[3] H. Kim and J. Kim, “Deep Learning-Based Anomaly Detection for Network Security,” in Proc. Int. Conf. Network Security, 2020.

[4] M. A. Khan and A. Salah, “A Review of Machine Learning in Network Security,” Computers & Security, 2020.

[5] Y. Zhou et al., “Hybrid CNN-RNN Models for Intrusion Detection,” IEEE Access, 2022.

[6] A. Thakkar and R. Lohiya, “A Review on Intrusion Detection System Using Machine Learning,” ICTACT Journal on Soft Computing, 2021.

[7] M. Ahmed et al., “A Survey on Network Anomaly Detection Using Machine Learning,” IEEE Communications Surveys & Tutorials, 2022.

[8] S. Mishra and P. Mohapatra, “Energy-Efficient Security Framework for Sensor Networks,” Procedia Computer Science, 2020.

[9] R. Vinayakumar et al., “Machine Learning Approaches for Network Traffic Analysis,” IEEE Transactions on Network and Service Management, 2021.

[10] L. Wang and F. Zhang, “Lightweight IDS Framework for IoT Networks,” Sensors, 2023.

[11] T. Singh, “Energy-Aware Anomaly Detection in Sensor Networks,” IEEE Internet of Things Journal, 2023.

[12] S. Patel, “Comparative Study of Machine Learning-Based IDS,” International Journal of Advanced Research in Computer Science, 2021.