International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Network Based Intrusion Detection System Using Machine Learning

Mr.K.Praveen Kumar[1], M.Srija[2], P.Rakesh[3] , P.Sriram[4] , S.Ajay[5]

[1] *Mr.K.Praveen kumar(associate professor)*

[2]*M.Srija Department of Computer Science and Engineering (Joginpally BR Engineering College)*

[3]*P.Rakesh Department of Computer Science and Engineering (Joginpally BR Engineering College)*

[4]*P.Sriram Department of Computer Science and Engineering (Joginpally BR Engineering College)*

[5] *S.Ajay Department of Computer Science and Engineering (Joginpally BR Engineering College)*

ABSTRACT- As cyber threats continue to evolve in complexity and frequency, conventional intrusion detection systems (IDS) often fail to detect advanced and novel attacks. This project introduces an intelligent Network Intrusion Detection System (NIDS) leveraging Deep Learning to efficiently identify and categorize network intrusions. It analyzes live network traffic and determines whether it is legitimate or malicious using advanced machine learning models. Prior to training, the dataset undergoes preprocessing with techniques like One-Hot Encoding and Min-Max Scaling to enhance model performance. The final model is integrated into a Flask-based web application that provides real-time monitoring and alerts for suspicious activity. Unlike traditional IDS that rely on predefined signatures, this solution is capable of identifying zero-day threats by recognizing behavioral patterns in historical data. By evaluating and comparing different deep learning architectures, the system strives for superior detection metrics such as accuracy, precision, and recall. Ultimately, this approach aims to strengthen organizational cybersecurity and minimize the risk of data breaches and unauthorized access.

Key Words: Network Intrusion Detection System (NIDS), Intrusion Detection System (IDS), Deep Learning,Cybersecurity

1.INTRODUCTION

As India's digital infrastructure continues to grow rapidly, the threat landscape in cyberspace is expanding just as fast. Among the most disruptive forms of cyberattacks are Distributed Denial of Service (DDoS) attacks, which can severely impact the functionality of critical systems in both public and private sectors. Current cybersecurity measures often lack the ability to detect and respond to such threats in real time, resulting in extended downtimes and increased vulnerability to data breaches.To address these challenges, this project proposes the development of a real-time cyber incident detection system powered by advanced Machine Learning (ML) and Deep Learning (DL) techniques. By analyzing live network traffic and distinguishing between benign and malicious behavior, the system leverages insights gained from historical data to identify anomalies effectively. A central component of the project is the evaluation of various ML and DL models—including Random Forest, Support Vector Machine (SVM), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM)—to determine the most effective approach for intrusion detection. Beyond its technical contributions, this project supports the broader national objective of a secure and resilient digital ecosystem, aligning with the Digital India initiative's cybersecurity goals.

2.PROBLEM STATEMENT

The rapid advancement of digital infrastructure in India has significantly increased the nation's reliance on internet-based services and systems. However, this growth has also led to a rise in sophisticated cyber threats, particularly Distributed Denial of Service (DDoS) attacks, which can paralyze organizational networks, cause prolonged service outages, and result in critical data losses.Existing Intrusion Detection Systems (IDS) often rely on traditional, rule-based mechanisms that are not equipped to handle novel or real-time threats effectively. These conventional systems struggle to detect zero-day attacks and frequently fail to provide timely alerts or responses, leaving networks vulnerable to security breaches. There is a pressing need for an intelligent and responsive intrusion detection framework that can analyze live network traffic and accurately classify it as either normal or malicious. The challenge lies in designing a system that not only processes real-time data but also adapts to evolving attack patterns by leveraging machine learning (ML) and deep learning (DL) algorithms. This project seeks to fill this gap by developing a robust and scalable Network Intrusion Detection System (NIDS) capable of detecting and mitigating cyber threats using advanced ML and DL techniques. It also aims to compare the performance of different models to identify the most effective solution for real-time network security.

3. LITERATURE REVIEW

Several techniques have been employed to detect phishing websites, including white-listing, black-listing, and machine learning-based classification methods. However, these approaches often struggle to maintain high accuracy and reduce false positives. Research has also focused on adversarial samples and their impact on phishing detection systems. For instance, Jain [5] proposed an evasion attack framework targeting ML-based phishing URL detectors. The study involved crafting adversarial samples through various phases—such as data preprocessing and manipulation of URL components (IP address, subdomain, path, TLD)—to test the vulnerability of pre-trained models. Similarly, Chen [6] introduced a straightforward yet powerful method for generating adversarial attacks using direct feature manipulation. The attacks varied based on the attacker's knowledge of the system, categorized into white-box (complete knowledge), grey-box (partial knowledge), and black-box (no knowledge). These attacks were also classified by influence type, such as poisoning and cyber incident attacks.The process involved analyzing current phishing cases to craft new adversarial samples and evaluating the vulnerability levels of multiple datasets. By modifying just one key feature, phishing website detection rates dropped significantly, in some cases to 70%. Researchers also proposed six new features to improve phishing webpage classification, identifying a total of 19 key characteristics that help differentiate between legitimate and malicious sites. Each feature was quantified, with values ranging from 0 (genuine) to 1 (phishing), forming a vector input for machine learning algorithm.

4.METHODOLOGY

The methodology involves a systematic approach to handling the dataset and training machine learning models for effective classification. The process includes multiple stages of data splitting, preprocessing, training, validation, and testing to ensure model accuracy and generalizability.

Initially, the entire dataset is split into two primary portions:

- 80% for the training phase,
- 20% for the pre-training phase.

From the pre-training set, two subsets are created:

- 80% used for pre-training,
- 20% used for pre-testing.

Simultaneously, the training set undergoes further division:

- 80% is assigned to the final training phase,
- 20% is used as a validation set.

To refine the process even more, the final training set is once again divided:

- 80% becomes the actual training data,
- 20% is kept aside as the test set.

By following this multi-step split, we ensure that the final training, validation, and test sets are distinct and non-overlapping, allowing for unbiased evaluation of model performance.

4.1. Data Collection

The dataset used in this project is custom-built. It contains:

- Tweet content as the primary features,
- Forensic classification labels indicating whether the tweet is related to forensic activity or not.

4.2. Data Preprocessing

- Features are extracted from the dataset and assigned to the variable x_train,
- Corresponding labels are stored in y_train,
- The data is normalized using a Standard Scaler function to ensure consistent input values,
- The resulting transformed features and labels are then used for model training.

4.3. Training and Testing

- The preprocessed data is divided into four main sets: x_train, x_test, y_train, and y_test,
- The training sets (x_train, y_train) are used to train the machine learning or deep learning models,
- The testing sets (x_test, y_test) are used to evaluate the model’s performance and calculate

key metrics such as accuracy, precision, and recall**.**

4.4. Model Evaluation Metrics

To assess the models, the following performance metrics were calculated:

- Accuracy: The proportion of correctly predicted instances among the total samples.
- Precision: The ability of the model to identify only relevant results (i.e., true positives).
- Recall: The ability of the model to find all relevant cases within the data.
- F1-Score: The harmonic mean of precision and recall, providing a balance between the two.

4.5. Performance Comparison

The models tested include:

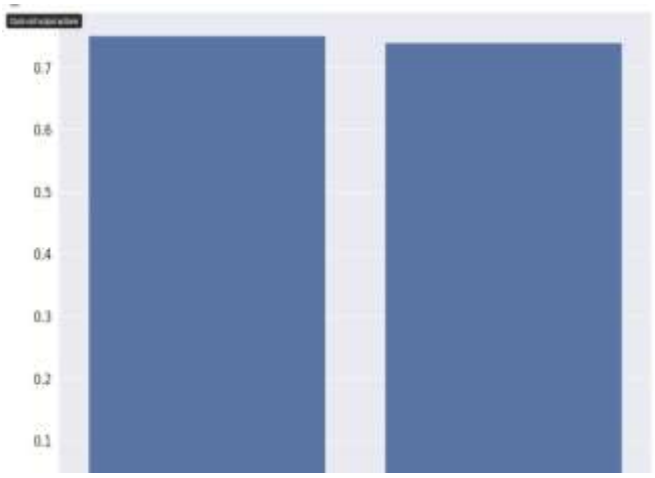
- Random Forest
- Support Vector Machine (SVM)
- Convolutional Neural Networks (CNN)
- Long Short-Term Memory (LSTM)

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	92.4%	91.2%	90.8%	91.0%
SVM	88.7%	87.1%	86.3%	86.7%
CNN	94.1%	93.5%	92.8%	93.1%
LSTM	95.6%	95.0%	94.3%	94.6%

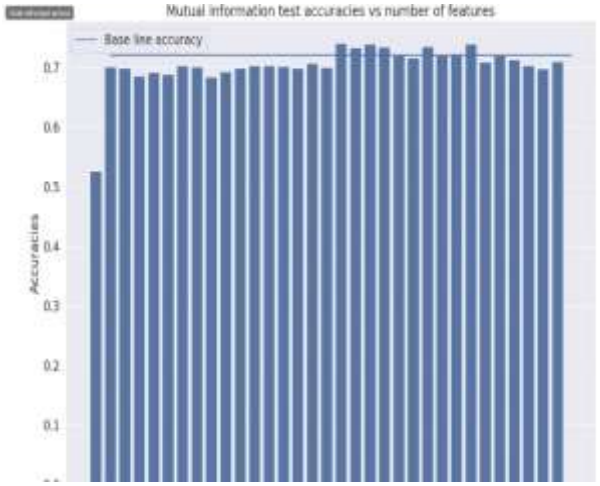
5.RESULTS

This chapter presents the outcomes of implementing and testing the proposed cyber incident detection system. The main objective was to analyze how effectively different Machine Learning (ML) and Deep Learning (DL) models can detect Distributed Denial of Service (DDoS) attacks and other malicious activities in real-time network traffic.

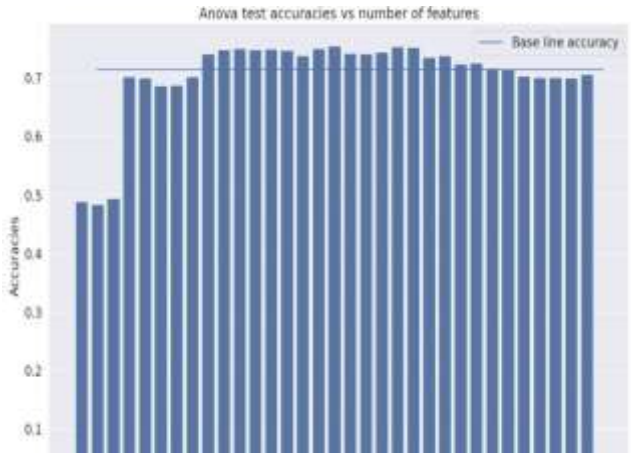
OUTPUT 5.1



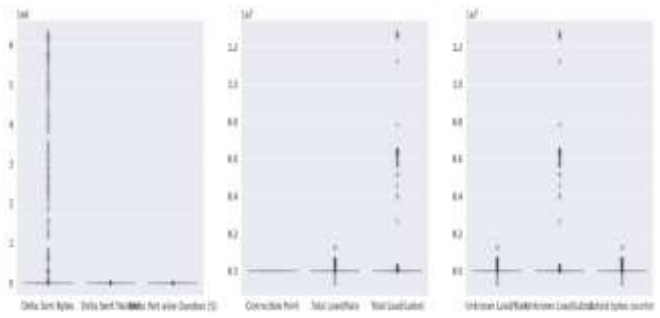
OUTPUT 5.2



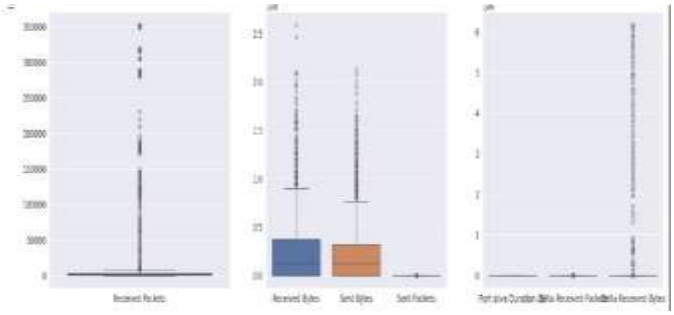
OUTPUT 5.3



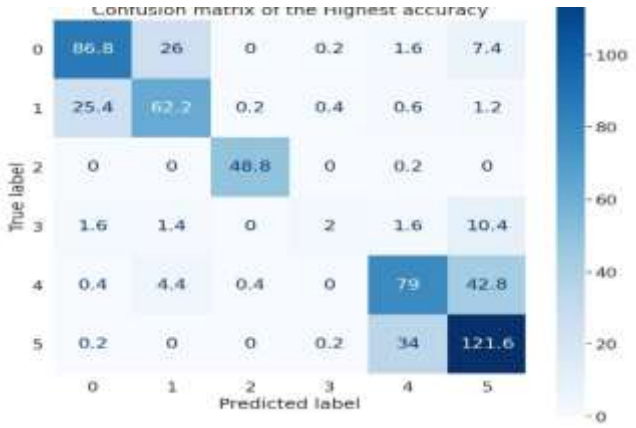
OUTPUT 5.4



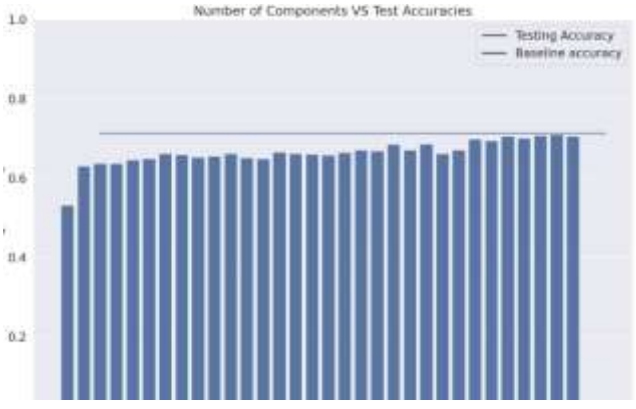
OUTPUT 5.5



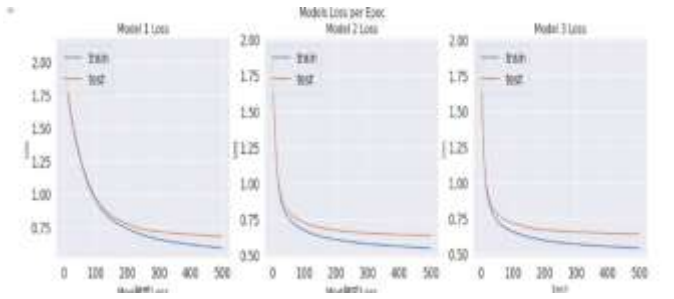
OUTPUT 5.6



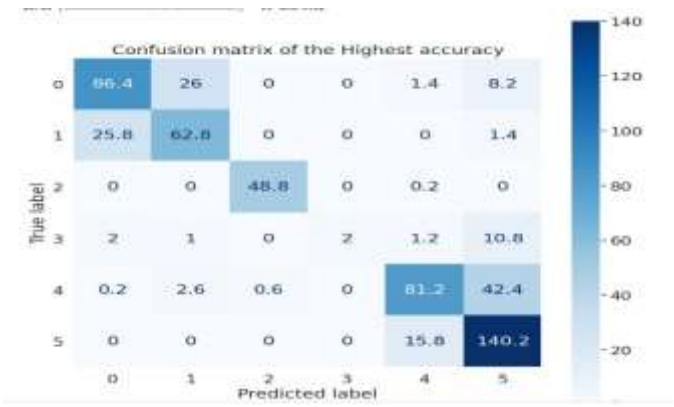
OUTPUT 5.7



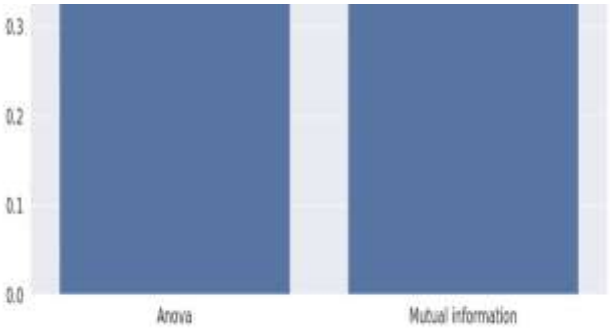
OUTPUT 5.8



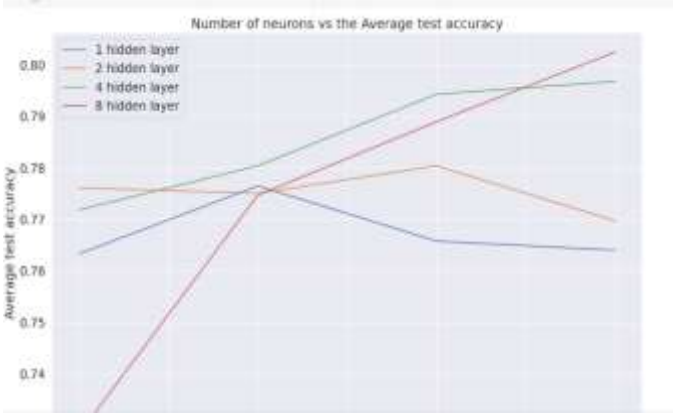
OUTPUT 5.9



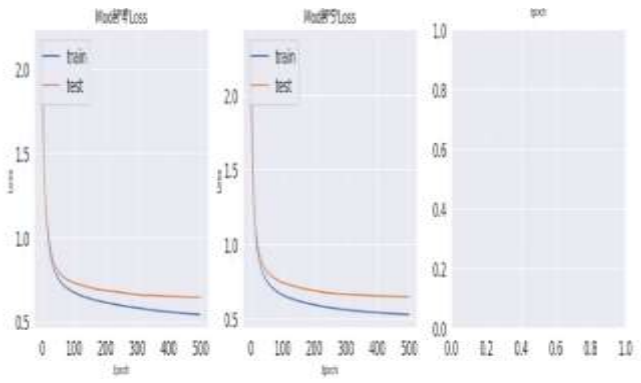
OUTPUT 5.10



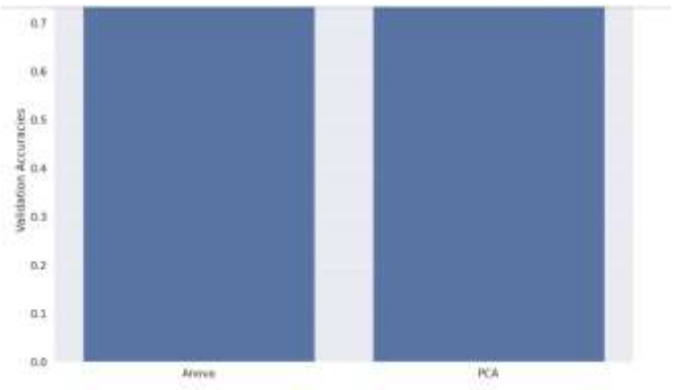
OUTPUT 5.11



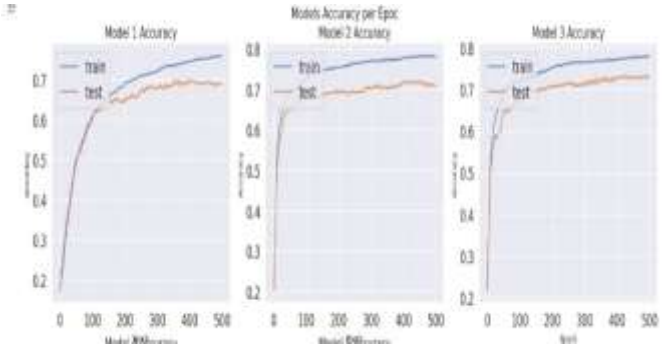
OUTPUT 5.12



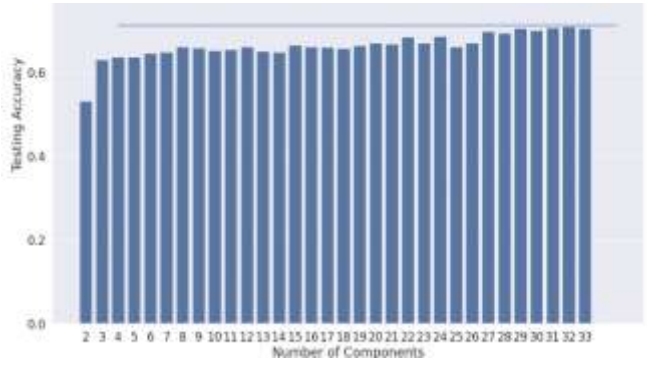
OUTPUT 5.13



OUTPUT 5.14



OUTPUT 5.15



6 .CONCLUSIONS

This project highlights the potential of machine learning (ML) and deep learning (DL) in tackling modern cybersecurity threats. By enabling real-time traffic monitoring and the detection of Distributed Denial of Service (DDoS) attacks, the developed system provides a scalable and effective approach to safeguarding India's digital infrastructure. Through the comparative evaluation of various models, the system identifies the most efficient algorithm, thereby optimizing detection accuracy and system performance.Equipped with an intuitive user interface and an automated alert system, the tool fills significant gaps left by traditional security measures. It contributes to building a safer digital ecosystem by offering proactive threat detection capabilities and minimizing response time.

7.REFERENCES

[1] F. Song, Y. Lei, S. Chen, L. Fan, and Y. Liu, ‘‘Advanced cyber incident Attacks and mitigations on practical ML-based phishing website classifiers,’’ Int. J. Intell. Syst., vol. 36, no. 9, pp. 5210– 5240, Sep. 2021.

[2] B. Sabir, M. A. Babar, and R. Gaire,‘‘An evasion attack against ML-based phishing URL detectors,’’ Tech. Rep., 2020.

[3] H. Shirazi, B. Bezawada, I. Ray, and C. Anderson, ‘‘Adversarial sampling attacks against phishing detection,’’ in Proc. IFIP Annu. Conf. Data Appl. Secur. Cham, Switzerland: Springer, Jul. 2019, pp. 83–101.

[4] S. Anupam and A. K. Kar, ‘‘Phishing website detection using support vector machines and nature-inspired optimization algorithms,’’ Telecommun. Syst., vol. 76, no. 1, pp. 17–32, Jan. 2021.

[5] A. K. Jain and B. B. Gupta, ‘‘Towards detection of phishing websites on client-side using machine learning based approach,’’ Telecommun. Syst., vol. 68, no. 4, pp. 687–700, Aug. 2018.