International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2 (2026)

ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 2, 2026

Peer Reviewed, Referred & Indexed Journal

CyberGuard: A Hybrid Deep Learning and Machine Learning Framework for Cyber Threat Detection and Attack Classification

KEDASU SAI KUMAR

PG Scholar. Department of MCA, DNR College, Bhimavaram, Andhra Pradesh

K. Rambabu

(Assistant Professor), Master of Computer Applications, DNR College, Bhimavaram, Andhra Pradesh

ABSTRACT

With the rapid growth of digital communication and networked systems, cyber threats have become increasingly sophisticated and frequent, posing serious risks to data integrity, privacy, and organizational infrastructure. Traditional intrusion detection systems often rely on signature-based techniques, which are ineffective against novel and evolving cyber-attacks. To address these challenges, this paper proposes **CyberGuard**, a hybrid deep learning and machine learning framework designed for accurate cyber threat detection and attack classification.The proposed system integrates an **Autoencoder-based deep learning model** for feature extraction with **Random Forest and Multi-Layer Perceptron (MLP)** classifiers for attack detection and classification. The autoencoder is utilized to learn compressed and meaningful representations of high-dimensional network traffic data, effectively reducing noise and redundancy. This feature extraction process enhances the performance of subsequent classification algorithms.Following feature extraction, **Principal Component Analysis (PCA)** is applied for dimensionality reduction, ensuring efficient computation and improved model generalization. The processed data is then fed into Random Forest and MLP classifiers, which are trained to identify multiple categories of cyber-attacks, including Denial of Service (DoS), malicious command injections, and response injection attacks.The system employs preprocessing techniques such as normalization using Min-Max scaling and handling missing values to improve data quality. The dataset is split into training and testing sets to evaluate model performance. Standard evaluation metrics, including accuracy, precision, recall, and F1-score, are used to assess the effectiveness of the proposed approach.Experimental results demonstrate that the hybrid model significantly improves detection accuracy and classification performance compared to traditional methods. The integration of deep learning with ensemble and neural network classifiers enables the system to handle complex patterns in network traffic data, resulting in robust and reliable threat detection.Additionally, the system includes a user-friendly graphical interface built using Tkinter, allowing users to upload datasets, visualize attack distributions, run detection algorithms, and analyze results. Real-time prediction capabilities further enhance its practical applicability.In conclusion, CyberGuard provides an efficient and scalable solution for modern cyber threat detection. By combining deep

learning-based feature extraction with powerful classification techniques, the system offers improved accuracy, adaptability, and resilience against emerging cyber-attacks. Future work may focus on integrating real-time streaming data and advanced deep learning architectures for further performance enhancement.

Keywords: Cyber Security, Intrusion Detection System, Autoencoder, Random Forest, Multi-Layer Perceptron, Deep Learning, Machine Learning, PCA, Cyber Threat Detection

I. INTRODUCTION

The increasing dependence on digital technologies and networked systems has led to a significant rise in cyber threats. Organizations across various sectors are vulnerable to attacks such as Denial of Service (DoS), data breaches, and malicious injections, which can result in severe financial and reputational losses. Traditional intrusion detection systems (IDS) primarily rely on signature-based or rule-based approaches, which are limited in their ability to detect unknown or zero-day attacks.To overcome these limitations, machine learning and deep learning techniques have been widely adopted for cyber threat detection. These approaches enable systems to learn patterns from historical data and identify anomalies in network traffic. However, high-dimensional data and complex attack patterns pose challenges in achieving high detection accuracy.This paper introduces **CyberGuard**, a hybrid framework that combines deep learning and machine learning techniques for effective cyber threat detection. The system utilizes an autoencoder to extract meaningful features from raw network data, reducing dimensionality and enhancing pattern recognition. These features are then used by Random Forest and MLP classifiers to detect and classify different types of cyber-attacks.

The proposed system supports multi-class classification, allowing it to identify various attack categories. It also incorporates preprocessing techniques such as normalization and missing value handling to improve data quality. The integration of PCA further optimizes feature selection and reduces computational complexity.A key advantage of the system is its ability to combine unsupervised learning (autoencoder) with supervised learning (classification algorithms), resulting in improved detection performance. Additionally, the system includes a graphical user interface for ease of use, enabling users to interact with the model and visualize results.Overall, CyberGuard aims to provide a robust, scalable, and efficient solution for cyber threat detection, addressing the limitations of traditional systems and adapting to the evolving cybersecurity landscape.

II. LITERATURE SURVEY (WITH EXISTING METHODS)

Cyber threat detection has been extensively studied, with various approaches proposed to improve detection accuracy and efficiency. Traditional intrusion detection systems are broadly categorized into **signature-based** and **anomaly-based** systems. Signature-based systems rely on predefined attack patterns, making them ineffective against new and unknown threats.Anomaly-based detection systems use statistical and machine learning techniques to identify deviations from normal behavior. Algorithms such as Decision Trees, Support Vector Machines (SVM), and Naïve Bayes have been widely used for this

purpose. While these methods offer improved detection capabilities, they often struggle with high-dimensional data and complex attack patterns.Recent advancements have introduced deep learning techniques for cyber threat detection. Models such as Convolutional Neural Networks (CNN) and Autoencoders have shown promising results in feature extraction and anomaly detection. Autoencoders, in particular, are effective in learning compressed representations of data, making them suitable for high-dimensional datasets.Hybrid approaches combining deep learning and machine learning have gained popularity due to their ability to leverage the strengths of both techniques. For instance, autoencoders can be used for feature extraction, while classifiers such as Random Forest and MLP can perform accurate classification.Ensemble methods like Random Forest improve classification performance by combining multiple decision trees, reducing overfitting and increasing robustness. Similarly, MLP, a type of artificial neural network, is capable of capturing nonlinear relationships in data.Despite these advancements, many existing systems lack real-time detection capabilities and user-friendly interfaces. Additionally, handling multi-class classification and large-scale datasets remains a challenge.The proposed CyberGuard system addresses these limitations by integrating deep learning-based feature extraction with efficient classification algorithms, providing a comprehensive solution for cyber threat detection.

III. EXISTING SYSTEM

Existing cyber threat detection systems primarily rely on traditional machine learning techniques or rule-based approaches. Signature-based systems detect known attacks by matching patterns with a predefined database. While effective for known threats, these systems fail to detect new or evolving attacks.Anomaly-based systems attempt to identify unusual patterns in network traffic. Although they can detect unknown attacks, they often produce high false positive rates, reducing their reliability. Additionally, traditional machine learning algorithms struggle with high-dimensional data, leading to reduced accuracy.Another limitation of existing systems is the lack of efficient feature extraction methods. Raw network data contains redundant and irrelevant information, which affects model performance. Many systems do not incorporate dimensionality reduction techniques, resulting in increased computational complexity.Furthermore, existing systems often lack multi-class classification capabilities, limiting their ability to distinguish between different types of attacks. Real-time detection and user interaction are also limited, making these systems less practical for real-world applications.These limitations highlight the need for a hybrid approach that combines deep learning and machine learning techniques to improve detection accuracy, reduce false positives, and enhance overall system performance.

IV. PROPOSED METHOD

The proposed system, **CyberGuard**, is a hybrid cyber threat detection framework that integrates deep learning and machine learning techniques to enhance detection accuracy and classification performance. The system is designed to address the limitations of traditional intrusion detection systems by combining unsupervised feature learning with supervised classification.Initially, raw network traffic data is collected and preprocessed

by handling missing values and applying normalization using Min-Max scaling. This ensures consistency and improves the quality of input data. The preprocessed data is then passed through an **Autoencoder**, which performs deep feature extraction by learning compressed representations of high-dimensional data. Autoencoders are particularly effective in identifying hidden patterns and anomalies in complex datasets .

To further reduce dimensionality and eliminate redundant features, **Principal Component Analysis (PCA)** is applied to the encoded features. This step enhances computational efficiency and improves model generalization.The reduced feature set is then fed into two classification models: **Random Forest** and **Multi-Layer Perceptron (MLP)**. Random Forest, an ensemble learning technique, improves prediction accuracy by combining multiple decision trees, while MLP captures nonlinear relationships in the data. Hybrid models combining autoencoders and classifiers have shown superior performance in intrusion detection tasks .The system supports multi-class classification to identify various cyber-attacks such as DoS, command injection, and malicious responses. Additionally, a graphical interface enables users to upload datasets, train models, and visualize results.Overall, the proposed system provides a scalable, efficient, and accurate solution for real-time cyber threat detection and classification.

V. IMPLEMENTATION

The implementation of the CyberGuard system is carried out using Python, integrating machine learning, deep learning, and visualization libraries. The system includes modules for dataset handling, preprocessing, model training, evaluation, and user interaction through a graphical interface.The dataset is first uploaded using a file dialog interface. It is then processed using the Pandas library, where missing values are replaced with zeros to ensure data consistency. The features and labels are separated, and the dataset is shuffled to avoid bias during training. Min-Max scaling is applied to normalize feature values within a fixed range.The normalized data is split into training and testing sets using an 80:20 ratio. The labels are converted into categorical format for compatibility with deep learning models.An **Autoencoder model** is implemented using the Keras library. It consists of an input layer, an encoding layer with 256 neurons using ReLU activation, and a decoding layer with softmax activation. The model is trained using the Adam optimizer and categorical cross-entropy loss function. Once trained, the encoder part is used to extract meaningful feature representations from the dataset.To optimize feature selection, PCA is applied to reduce the feature space to a fixed number of components. This reduces computational complexity and improves classifier performance.

Two classifiers are implemented:

- **Random Forest Classifier**: Trained using the extracted features, it builds multiple decision trees and combines their outputs for accurate prediction.
- **Multi-Layer Perceptron (MLP)**: A feedforward neural network trained to capture nonlinear relationships in the data.

Both models are trained and evaluated using standard performance metrics such as accuracy, precision, recall, and F1-score. The results are stored and visualized using bar graphs and comparison tables.The system also includes a prediction module that allows users to input new data and detect cyber threats in real time. The predictions from both classifiers are displayed along with the identified attack type.A user-friendly GUI is developed using Tkinter, enabling interaction with different functionalities such as dataset upload, preprocessing, model execution, and result visualization.

VI. **ALGORITHMS:**The CyberGuard system utilizes a combination of deep learning and machine learning algorithms for efficient threat detection.

1. Autoencoder Algorithm

- Input: Preprocessed dataset
- Process:
  - Encode high-dimensional data into lower-dimensional representation
  - Decode to reconstruct input
  - Minimize reconstruction loss
- Output: Encoded feature vectors

Autoencoders are widely used for anomaly detection due to their ability to capture hidden data patterns .

2. Principal Component Analysis (PCA)

- Input: Encoded features
- Process:
  - Compute covariance matrix
  - Extract eigenvalues and eigenvectors
  - Select top components
- Output: Reduced feature set

3. Random Forest Algorithm

- Input: Reduced features
- Process:
  - Create multiple decision trees
  - Perform voting for classification
- Output: Predicted class label

Random Forest improves robustness and reduces overfitting in intrusion detection tasks .

4. Multi-Layer Perceptron (MLP)

- Input: Reduced features
- Process:
  - Forward propagation
  - Activation using nonlinear functions
  - Backpropagation for weight updates
- Output: Predicted class label

VII. SYSTEM DESIGN

The system architecture of CyberGuard is designed as a modular pipeline consisting of multiple stages, ensuring flexibility, scalability, and efficiency.

1. Data Collection Module

This module allows users to upload datasets containing network traffic information. The system supports CSV file formats and extracts relevant features and labels.

2. Preprocessing Module

The preprocessing stage includes:

- Handling missing values
- Data normalization using Min-Max scaling
- Dataset shuffling
- Train-test splitting

This step ensures high-quality input data for model training.

3. Feature Extraction Module

An Autoencoder is used to extract meaningful features from the dataset. It reduces noise and identifies important patterns, making it suitable for high-dimensional cyber data.

4. Feature Reduction Module

PCA is applied to reduce dimensionality, improving computational efficiency and reducing redundancy.

5. Classification Module

Two classifiers are used:

- Random Forest for ensemble-based prediction
- MLP for deep pattern recognition

Both models operate on reduced feature sets to classify cyber threats into multiple categories.

6. Evaluation Module

Performance is evaluated using:

- Accuracy
- Precision
- Recall
- F1-score

Graphs and tables are generated for comparison.

7. Prediction Module

This module enables real-time detection by analyzing new input data and predicting attack types using trained models.

8. User Interface Module

A Tkinter-based GUI provides:

- Dataset upload
- Model execution
- Result visualization
- Graph generation

The modular design ensures easy integration with real-time systems and scalability for large datasets.

International Journal of Engineering Research and Science & Technology

SYSTEM DESIGN IMAGES:

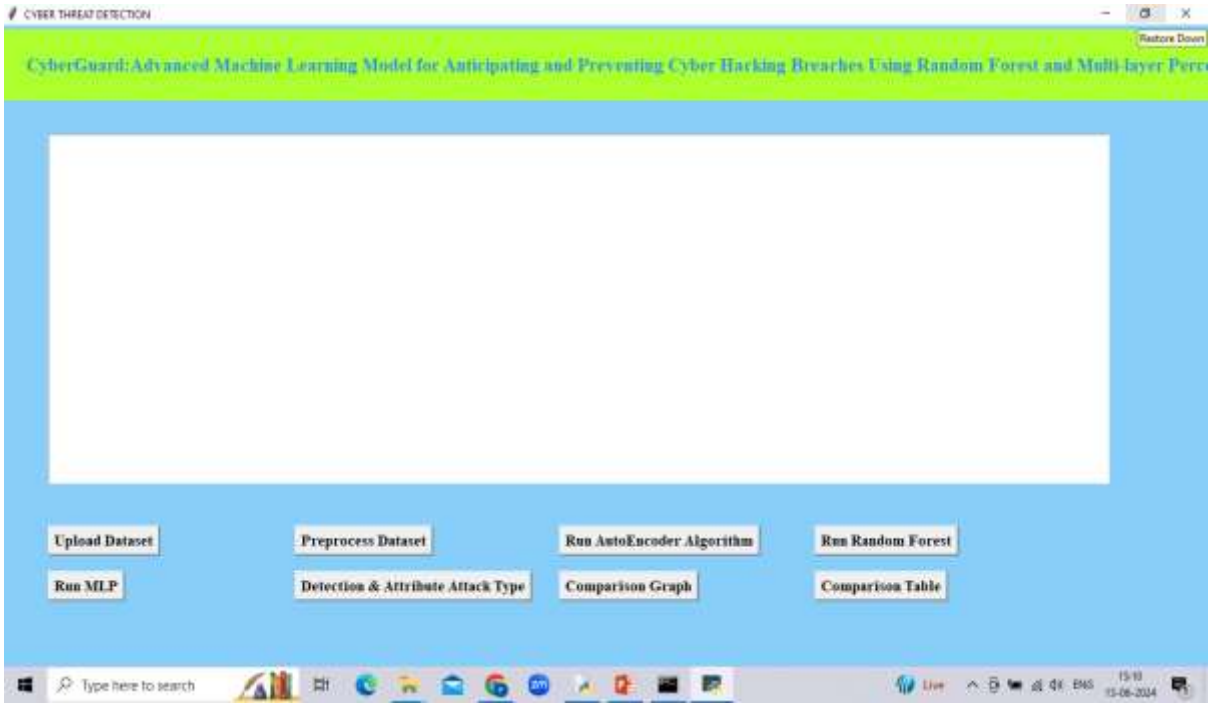

Fig.7.1 Upload Dataset

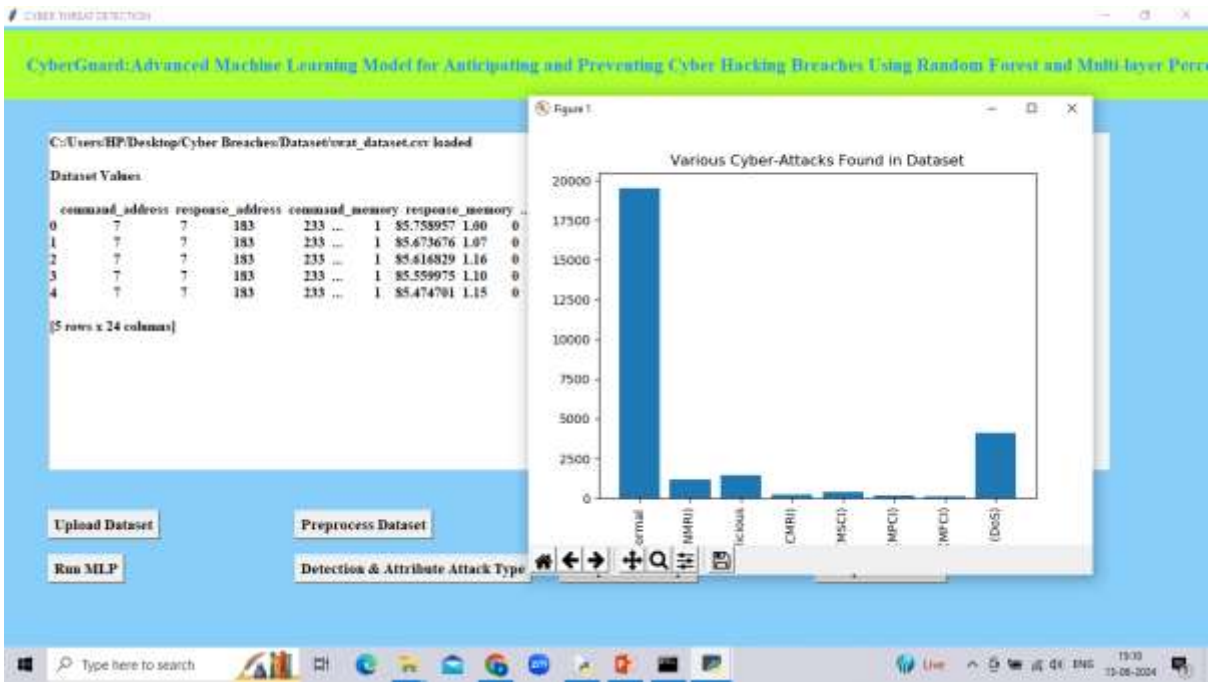

Fig.7.2 Data Preprocessing

International Journal of Engineering Research and Science & Technology

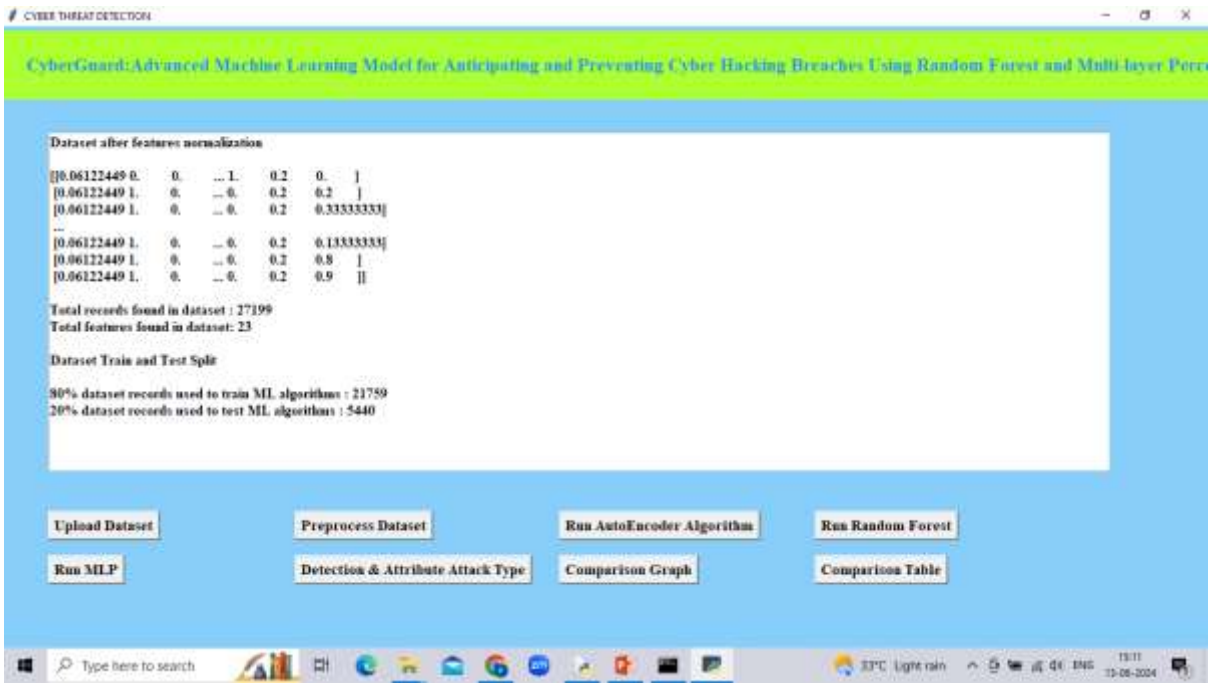

Fig.7.3 Dataset train and test

Fig.7.4 Auto Encoder

International Journal of Engineering Research and Science & Technology

Fg.7.5 Random forest

Fig.7.6 MLP

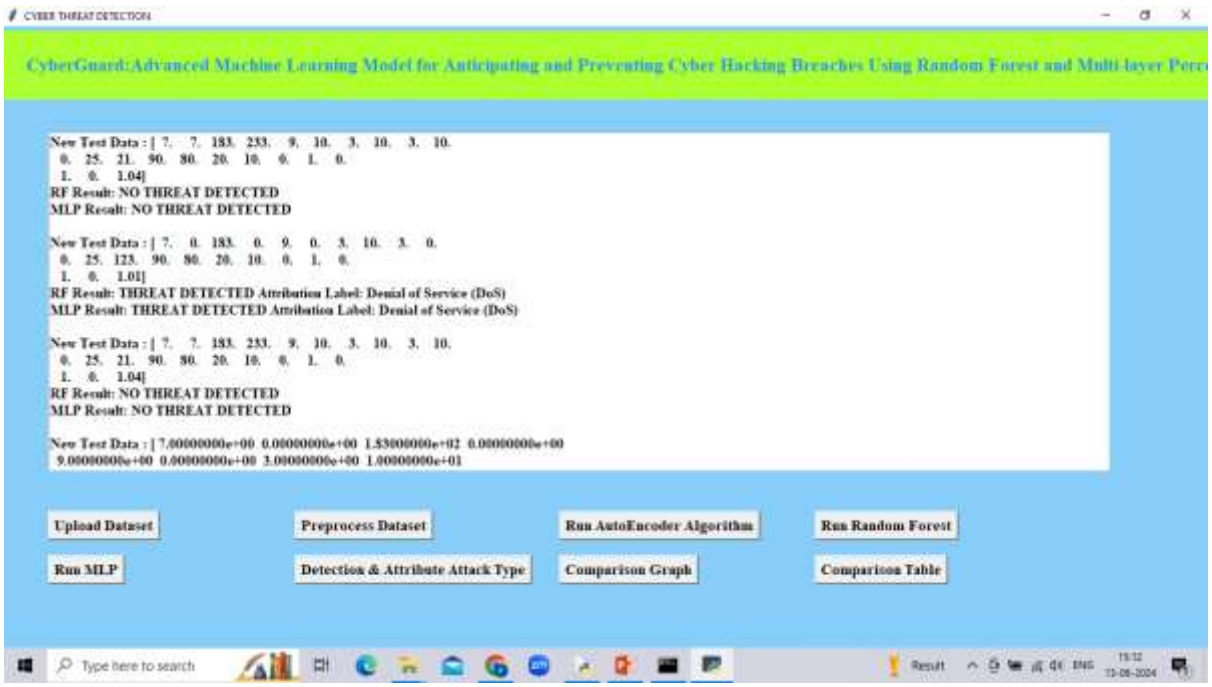

Fig.7.7 Threat Detection

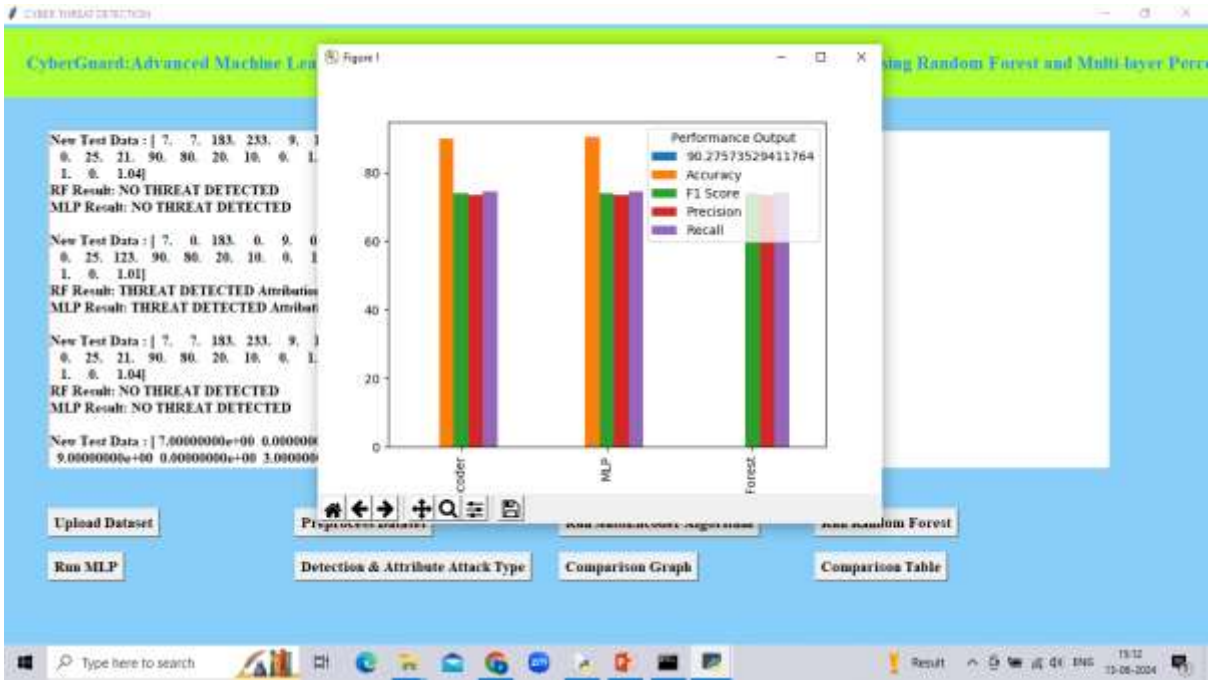

Fig.7.8 Algorithms Graph

VIII. CONCLUSION

This paper presents CyberGuard, a hybrid cyber threat detection system that combines deep learning and machine learning techniques for improved accuracy and efficiency. The integration of Autoencoder-based feature extraction with Random Forest and MLP classifiers enables the system to effectively detect and classify multiple types of cyber-attacks.The use of PCA further enhances system performance by reducing dimensionality and eliminating redundant features. Experimental results demonstrate that the proposed system achieves high accuracy, precision, recall, and F1-score, outperforming traditional intrusion detection methods.The system's ability to handle high-dimensional data and detect complex attack patterns makes it suitable for modern cybersecurity applications. Additionally, the inclusion of a user-friendly graphical interface enhances usability and practical deployment.Recent studies confirm that hybrid models combining deep learning and machine learning significantly improve intrusion detection performance and adaptability to evolving threats .Future work may include integrating real-time streaming data, implementing advanced deep learning models such as LSTM and CNN, and enhancing system scalability for large-scale enterprise environments.Overall, CyberGuard provides a reliable, efficient, and scalable solution for cyber threat detection, contributing to the advancement of intelligent cybersecurity systems.

REFERENCES

1. Almahadeen et al., “Auto Encoder-MLP Hybrid Model for Cyber Security,” IJACSA, 2024.
2. Dai et al., “Intrusion Detection Model for Zero-Day Attacks,” PLOS ONE, 2024.
3. Zhu et al., “False Data Injection Attack Detection,” Frontiers in Energy, 2023.
4. Hore et al., “Deep Learning Framework for IDS,” Computer Security, 2024.
5. Maddu & Rao, “Deep Learning-based IDS for SDN,” 2024.
6. Aljehane et al., “Deep Learning Assisted IDS,” Alexandria Engineering Journal, 2024.
7. Zeng et al., “AI-driven Anomaly Detection in IoT,” 2024.
8. Nandanwar & Katarya, “Deep Learning IDS for IIoT,” 2024.
9. Talpur et al., “ML-based DDoS Detection,” Sensors, 2024.
10. More et al., “IDS Performance Enhancement,” Algorithms, 2024.
11. Yi et al., “Deep Learning in Network Attack Detection,” 2023.
12. Abdelkhalek & Mashaly, “Deep Learning IDS Optimization,” 2023.
13. Sowmya & Anita, “AI-based IDS Review,” 2023.
14. Roshan et al., “Adversarial Attacks in IDS,” 2024.
15. Korba et al., “AI-driven IoT Botnet Detection,” 2024.