*International Journal of Scientific Research and Engineering Development-– Volume 9 Issue 2, Mar-Apr 2026*
*Available at www.ijsred.com*

RESEARCH ARTICLE OPEN ACCESS

# Ransomware Attack Prediction and Detection Using Hybrid Lightgbm with Multilayer Perceptrons

Mrs. N. Rajapriya*, Ms. J. Jenefa Sharon Gold**
*(Computer Science and Engineering, Francis Xavier Engineering College, Tirunelveli
Email: rajapriyacse@francisxavier.ac.in )
** (Computer Science and Engineering, Francis Xavier Engineering College, Tirunelveli
Email: jenefasharongoldj@gmail.com )

----------**********----------

## Abstract:

Ransomware has emerged as one of the most dangerous forms of cyber threats, targeting individuals and organizations by encrypting files or restricting access to systems until a ransom is paid. Unlike traditional malware, ransomware focuses on financial extortion by denying access to critical resources. In recent years, the increase in ransomware attacks has highlighted the need for intelligent and adaptive detection systems.

*Keywords* — **Ransomware, Cybersecurity, LightGBM, Multilayer Perceptron, Machine Learning, Deep Learning**

----------**********----------

## I. INTRODUCTION

In the modern digital era, cybersecurity threats have significantly increased due to the rapid growth of internet-based systems and data-driven applications. Among these threats, ransomware has become one of the most critical and damaging forms of cyberattacks. Ransomware is a type of malicious software that encrypts user data or restricts access to systems and demands payment to restore access.

Traditional security mechanisms such as antivirus software and signature-based detection are not sufficient to handle modern ransomware attacks. These systems fail to detect new and evolving ransomware variants. Therefore, there is a need for intelligent detection systems that can identify ransomware behavior using advanced techniques.

This project focuses on developing a hybrid ransomware detection model using LightGBM and Multilayer Perceptron (MLP). By combining machine learning and deep learning approaches, the system aims to improve detection accuracy, reduce false positives, and provide a reliable cybersecurity solution.

## II. OBJECTIVES

The primary objective of this project is to design and develop an efficient and intelligent ransomware detection system using a hybrid approach that combines machine learning and deep learning techniques. The system aims to provide accurate, fast, and reliable detection of ransomware attacks while minimizing false alarms and improving adaptability to new threats.

The following specific objectives define the scope and functionality of the proposed system:

### A. Hybrid Model Development:

The main objective is to develop a hybrid model that integrates Light Gradient Boosting Machine (LightGBM) and Multilayer Perceptron (MLP). LightGBM is used for efficient feature selection and faster training, while MLP enhances deep learning

capabilities for improved classification. This combination leverages the strengths of both algorithms to achieve better performance.

**B. Accurate Ransomware Detection:**
The system aims to improve the accuracy of ransomware detection by identifying malicious behavior effectively. It focuses on reducing false positives and false negatives, ensuring that legitimate files are not misclassified while accurately detecting ransomware threats.

**C. Efficient Data Processing:**
Another objective is to preprocess and normalize the dataset to ensure high-quality input for the model. This includes handling missing values, removing noise, and applying normalization techniques to improve model performance and reliability.

**D. Automatic Feature Learning:**
The system aims to enable automatic feature extraction and learning using deep learning techniques. By using MLP, the model can identify complex patterns and relationships in the data without relying on manual feature engineering.

**E. Performance Evaluation:**
The final objective is to evaluate the performance of the proposed model using standard evaluation metrics such as accuracy, precision, recall, and F1-score. These metrics help in assessing the effectiveness and robustness of the system.

## III. MODULES AND ALGORITHMS

The proposed ransomware detection system is designed using a modular architecture to ensure efficient processing, scalability, and better performance. Each module performs a specific task in the overall workflow, similar to the structured approach followed in published research papers .

### A. Modules

**1) Data Collection Module:**
This module is responsible for collecting datasets containing both ransomware and benign samples from reliable and diverse sources. The quality and diversity of the dataset play a crucial role in determining the effectiveness of the system. The collected data may include system logs, executable files, or behavioral data, which are properly labeled for supervised learning.

**2) Data Preprocessing Module:**
In this module, the collected raw data undergoes preprocessing to make it suitable for model training. This includes handling missing values, removing noise, and eliminating irrelevant information. Normalization and feature scaling techniques are applied to ensure consistency across the dataset. These steps improve the efficiency and accuracy of the model.

**3) Data Visualization Module:**
This module focuses on analyzing the dataset using visualization techniques such as graphs, charts, and plots. Visualization helps in understanding data distribution, identifying patterns, and detecting anomalies. It also assists in making informed decisions during model development.

**4) Feature Selection Module (LightGBM):**
LightGBM is used in this module to identify and select the most relevant features from the dataset. It reduces dimensionality and eliminates redundant data, thereby improving computational efficiency. LightGBM also speeds up the training process while maintaining high performance.

**5) Classification Module (MLP):**
The Multilayer Perceptron (MLP) is used as a classification model in this system. It consists of multiple layers of neurons that process input data and learn complex patterns. The MLP classifier uses the selected features to classify the data into ransomware or benign categories with high accuracy.

**6) Evaluation Module:**
This module evaluates the performance of the system using various metrics such as accuracy, precision, recall, and F1-score. These metrics provide a comprehensive analysis of the model's effectiveness

and help in fine-tuning the system for better performance.

**B. Algorithms**

**1) LightGBM Algorithm:**
Light Gradient Boosting Machine (LightGBM) is an advanced machine learning algorithm based on decision tree learning. It is designed for high efficiency and faster training speed. LightGBM uses a leaf-wise tree growth strategy, which helps in reducing loss and improving accuracy. It is particularly useful for handling large datasets and selecting important features.

**2) Multilayer Perceptron (MLP):**

Multilayer Perceptron is a deep learning algorithm that consists of an input layer, one or more hidden layers, and an output layer. Each neuron in the network is connected with weights and uses activation functions to process data. MLP is capable of learning complex and non-linear relationships, making it highly effective for ransomware detection.

**B. Functional Modules**

**1) Feature Learning Mechanism:**

In this stage, LightGBM and MLP work together to extract and learn meaningful features from the dataset. LightGBM identifies important features, while MLP learns complex patterns from these features, improving detection capability.

**2) Classification Logic:**

The classification logic is implemented using the trained MLP model. Based on the learned patterns and extracted features, the system classifies the input data as either ransomware or benign. This ensures accurate and reliable detection.

## IV. MODULES AND ALGORITHMS

The proposed ransomware detection system follows a well-structured and systematic methodology to ensure accurate and efficient identification of ransomware attacks. The workflow is designed using a combination of machine learning and deep learning techniques, enabling the system to process large datasets and detect complex ransomware patterns effectively.

**A. Data Collection**

The first step in the methodology involves collecting a comprehensive dataset consisting of both ransomware and benign samples. These datasets are obtained from reliable sources such as cybersecurity repositories, system logs, and publicly available datasets. The collected data is carefully labeled to distinguish between malicious and non-malicious samples.
A diverse and balanced dataset is essential to ensure that the model can generalize well and accurately detect different types of ransomware attacks in real-world scenarios.

**C. Data Preprocessing**

In this stage, the collected raw data is processed to make it suitable for model training. Data preprocessing includes handling missing values, removing noise, and eliminating irrelevant or duplicate data. Normalization and feature scaling techniques are applied to bring all features into a uniform range.
This step improves data quality and ensures that the model learns effectively without being affected by inconsistencies in the dataset.

**C. Data Visualization**

Data visualization is performed to analyze the dataset and understand its structure and distribution. Graphical representations such as bar charts, histograms, and plots are used to identify patterns, trends, and anomalies in the data.
Visualization helps in gaining insights into feature importance and assists in making better decisions during model development and feature selection.

**D. Feature Selection using LightGBM**

Feature selection is carried out using the Light Gradient Boosting Machine (LightGBM) algorithm. LightGBM identifies the most relevant features from the dataset and eliminates redundant or less important features.
This reduces the dimensionality of the dataset, improves computational efficiency, and enhances model performance. Additionally, LightGBM provides faster training speed while maintaining high accuracy.

### E. Model Training and Testing

The dataset is divided into training and testing sets to evaluate the performance of the model. Typically, 80% of the data is used for training, while 20% is used for testing.
During the training phase, the model learns patterns and relationships from the data. The testing phase is used to validate the model's performance on unseen data, ensuring that it generalizes well and does not overfit.

### F. Classification using Multilayer Perceptron (MLP)

In this stage, the selected features are fed into the Multilayer Perceptron (MLP) model for classification. MLP is a deep learning model that **consists of multiple layers of interconnected neurons.**
The model learns complex and non-linear relationships in the data and classifies each input as either ransomware or benign. This improves detection accuracy and enables the system to handle sophisticated ransomware patterns.

### G. Performance Evaluation

The final stage involves evaluating the performance of the proposed system using standard evaluation metrics such as accuracy, precision, recall, and F1-score.
Accuracy measures overall correctness
Precision evaluates correctness of positive predictions
Recall measures detection capability
F1-score balances precision and recall
These metrics provide a comprehensive understanding of the system's effectiveness and reliability.

## V. EXISTING SYSTEM

Traditional ransomware detection systems primarily rely on signature-based and rule-based approaches. These methods are effective in detecting known threats but fail to identify new or unknown ransomware variants.

Signature-based detection depends on predefined patterns, making it ineffective against zero-day attacks. Rule-based systems rely on fixed conditions, which cannot adapt to evolving attack techniques. As ransomware continuously changes its behavior, these systems struggle to keep up with new threats.

Another major limitation is the high false positive rate, where legitimate files are incorrectly classified as malicious. This reduces system reliability and affects user trust. Additionally, traditional systems rely heavily on manual feature extraction, which is time-consuming and may fail to capture complex patterns in the data.

Due to these limitations, existing systems are not suitable for modern cybersecurity environments, where intelligent and adaptive solutions are required.

## VI. EXISTING SYSTEM

To overcome the limitations of traditional approaches, the proposed system introduces a hybrid ransomware detection model that combines LightGBM and Multilayer Perceptron (MLP).

The system leverages LightGBM for efficient feature selection and faster processing, while MLP is used for deep learning-based classification. This combination enhances both speed and accuracy, making the system more effective in detecting ransomware.

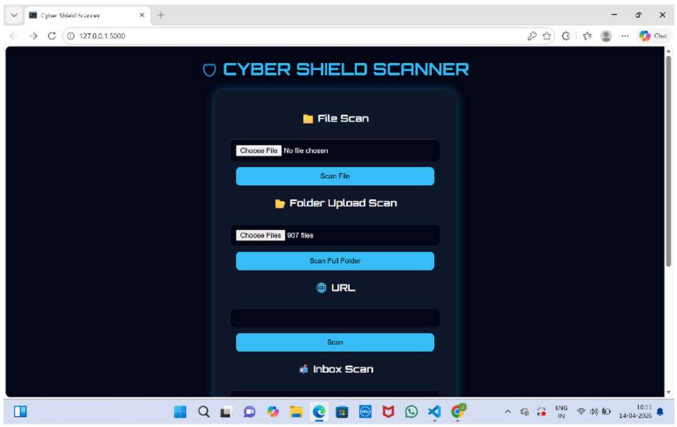
  

One of the key advantages of the proposed system is improved detection accuracy. By using advanced algorithms, the model can identify complex ransomware patterns that traditional systems fail to detect. Additionally, the system significantly reduces false positives, ensuring that legitimate files are not misclassified.

The proposed model also provides better adaptability, as it can learn and detect new ransomware variants without requiring manual updates. The automated feature learning capability further enhances efficiency by eliminating the need for manual intervention.

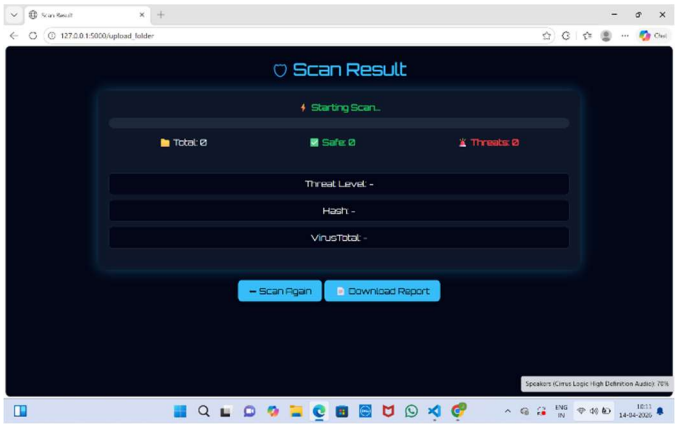
Overall, the proposed system offers a reliable, scalable, and efficient solution for ransomware detection in modern cybersecurity environments.

## VII. OUTPUT



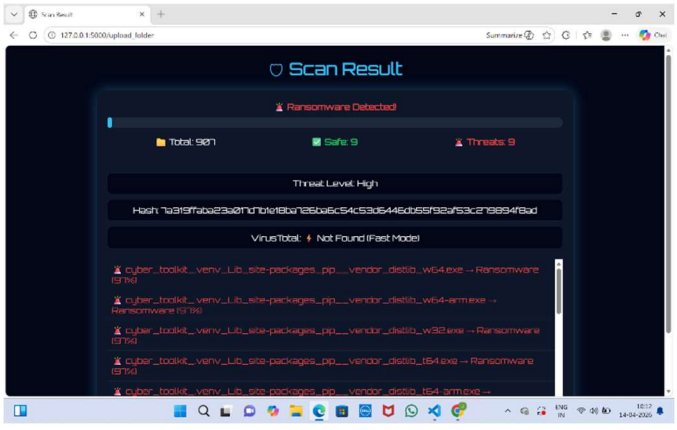

### a) User Interface for File and Folder Scanning

The first screen shows the main interface of the malware detection system, where the user can upload either a single file or multiple files (folder) for analysis. The “Choose File” option allows selecting one file, and the “Scan File” button is used to start scanning. Similarly, the folder section allows uploading multiple files and scanning them together using the “Scan Folder” button. This interface acts as the input stage where users provide files for malware analysis.




### b) Ransomware Detection Result for Folder Analysis

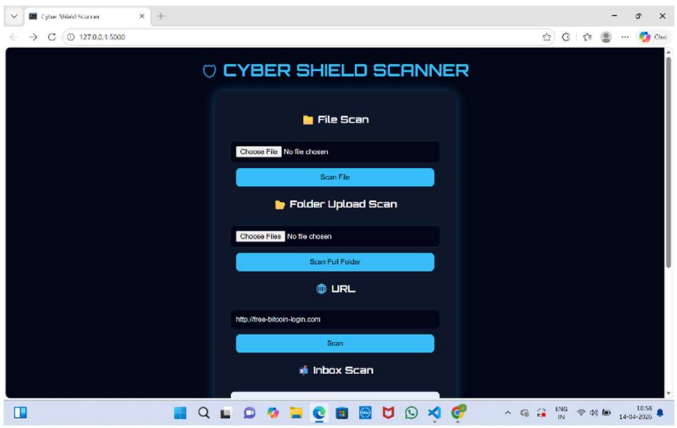
The Second screen displays the result of scanning a folder containing 907 files, out of which 9 files are identified as threats and 9 files are marked as safe.The system detects the presence of ransomware and classifies the overall threat level as **high**. Each malicious file is shown with a high confidence score of 97%, indicating strong ransomware characteristics.Additionally, a hash value is generated, and the VirusTotal status is shown as “Not Found (Fast Mode)”. The detected files are mainly executable files flagged due to suspicious behavior.This result clearly indicates that the folder contains malicious files and requires immediate attention to prevent further system compromise.




### c) Detailed Threat Information and File Analysis

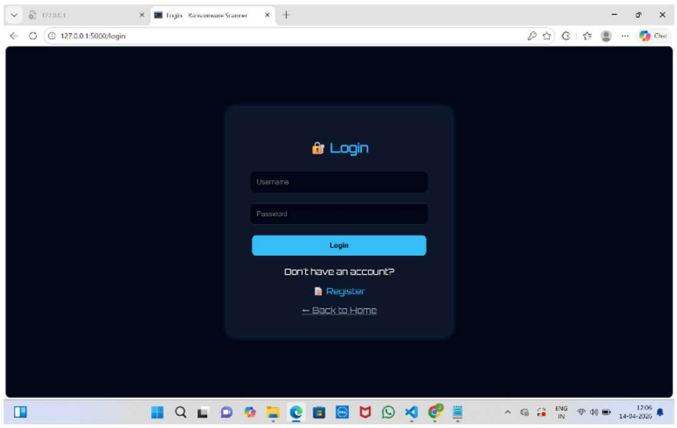
  

The third screen presents detailed information about detected threats, including threat level, hash value, and VirusTotal status. The system assigns a high threat level and generates a unique hash value for each file to enable identification and tracking.The VirusTotal status is shown as "Not Found (Fast Mode)," indicating that external verification was not performed during quick analysis. The list of detected files with high confidence values (97%) provides clear visibility into malicious components.



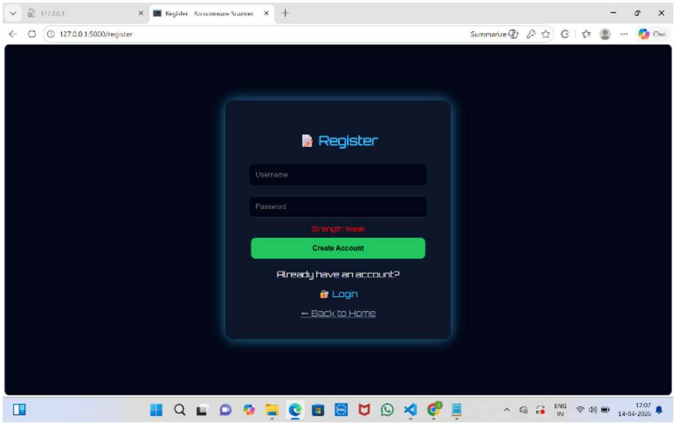

### d) URL Scanning and Suspicious Detection

The fourth screen shows the result of scanning a URL. The system identifies the entered URL as suspicious with a medium threat level.The reasons for classification includes the Presence of suspicious keywords.Absence of HTTPS protocol Newly registered or unknown domain.The system reports one threat and no safe results, indicating that the URL is potentially unsafe and should be avoided.



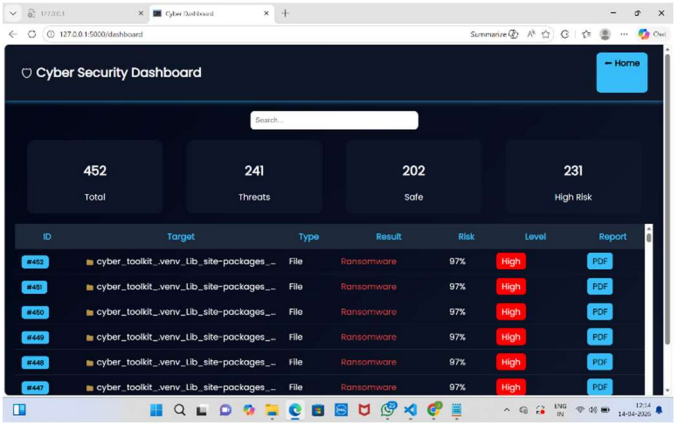

### e) User Authentication – Login System

The fifth screen illustrates the login interface, where users can securely access the system using their credentials. This feature ensures that only authorized users can utilize the malware detection functionalities.




### f) User Registration with Password Strength Indicator

The sixth screen shows the registration page, where users can create a new account. The system includes a password strength indicator, which evaluates the security level of the entered password.In the example, the password is marked as weak, encouraging users to choose stronger credentials to improve security.




### g) Cyber Security Dashboard and Report Visualization

The final screen presents the Cyber Security Dashboard, which provides a comprehensive overview of scan results. It displays statistics such as total files scanned, number of threats detected, safe files, and high-risk files. A detailed table lists each file along with its type, result, risk percentage, and threat level. The system also allows users to download reports in PDF format for further analysis and documentation.

## VIII.CONCLUSIONS

This project presents a hybrid ransomware detection system using LightGBM and Multilayer Perceptron (MLP). The proposed model effectively improves detection accuracy, reduces false positives, and enhances adaptability compared to traditional methods. By combining machine learning and deep learning techniques, the system is capable of identifying complex ransomware patterns and provides a reliable solution for modern cybersecurity challenges.

## X. ACKNOWLEDGMENT

We express our sincere gratitude to our mentor, Mrs.N. Raja priya, Assistant Professor, Department of Computer Science and Engineering, for her valuable guidance, continuous support, and encouragement throughout the development of this project. Her insights and suggestions greatly contributed to the successful completion of this research work.

## XI. FUTURE ENHANCEMENTS

- Implementing real-time malware detection for continuous system monitoring
- Integrating cloud-based security solutions for scalable and remote analysis
- Applying advanced deep learning models such as LSTM and Transformer for improved accuracy
- Expanding the dataset to enhance model performance and generalization

## REFERENCES

[1] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, 2016.

[2] L. Breiman, “Random Forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[3] G. Ke et al., “LightGBM: A Highly Efficient Gradient Boosting Decision Tree,” Advances in Neural Information Processing Systems (NeurIPS), pp. 3146–3154, 2017.

[4] Y. LeCun, Y. Bengio, and G. Hinton, “Deep Learning,” Nature, vol. 521, pp. 436–444, 2015.

[5] A. Pektaş and T. Acarman, “Deep Learning for Effective Malware Detection,” Soft Computing, vol. 24, pp. 10277–10289, 2020.

[6] S. Y. Yerima, S. Sezer, and I. Muttik, “High Accuracy Malware Detection Using Machine Learning,” IET Information Security, vol. 9, no. 6, pp. 313–320, 2015.

[7] M. Abadi et al., “TensorFlow: A System for Large-Scale Machine Learning,” OSDI, pp. 265–283, 2016.

[8] F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[9] K. Saxe and K. Berlin, “Deep Neural Network Based Malware Detection,” MALWARE Conference, 2015.

[10] J. Z. Kolter and M. A. Maloof, “Learning to Detect Malicious Executables in the Wild,” KDD, pp. 470–478, 2004.

[11] R. Richardson and M. North, “Ransomware: Evolution, Mitigation and Prevention,” International Journal of Information Security Science, vol. 6, no. 2, pp. 10–21, 2017.

[12] A. Kharraz et al., “Cutting the Gordian Knot: A Look Under the Hood of Ransomware Attacks,” International Conference on Detection of Intrusions and Malware, 2015.