Ransomware Threat Detection and Mitigation Using Machine Learning Models

Prof. Faiz Aman[1], Aishwarya U [2], Archana K R [3], Keerthana A S[4], Spoorthy S G [5]
[1]*Assistant Professor, Department of Computer Science and Engineering, Bahubali College of Engineering, Shravanabelagola*
[2,3,4,5] *UG Student, Department of Computer Science and Engineering Bahubali College of Engineering, Shravanabelagola*

***Abstract*—Ransomware attacks are leading to massive financial losses and interruptions in operations across the world. Conventional signature models are useless against new strains and zero-day attacks. In this paper, we have discussed the development of an intelligent machine learning model capable of real-time ransomware attack identification and mitigation. Our model uses an ensemble model consisting of 40% of the random forests model, 40% of the XG Boost model, and 20% of the neural network model for identifying behavioral patterns in PDF, docx, and JSON files. In our experimental results on the CIC-Evasive-PDFMal2022 dataset, we achieved an accuracy of 99.10%, precision of 98.72%, and a recall of 99.48%. We have wrapped our machine learning model in a Gradio framework for real-time identification and recommendation of the ransomware attack mitigation process in personal as well as professional settings.**

***Index Terms*—Adversarial Machine Learning, Cybersecurity, Ensemble Learning, Explainable AI, Malware Detection, Ransomware Threat Detection, Real-time Detection Systems, XGBoost**

I. INTRODUCTION

The rapid expansion in digital connectivity, cloud computing, and remote work environments has drastically expanded the modern digital attack surface. Consequently, ransomware has evolved into one of the most serious cybersecurity threats that organizations, governments, and individuals around the world are facing, resulting in billions of dollars in economic losses annually. Sophisticated ransomware attacks encrypt critical data or lock entire systems, with demands for ransom payments, all too often ending in irreversible data loss when the target is unable or unwilling to give in to the demands. Conventional signature-based and rule-based mechanisms have lost their effectiveness in dealing with recent ransomware variants that deploy obfuscation techniques, polymorphic code, and zero-day exploits to bypass detection.

Machine learning represents a paradigm shift in ransomware defense, whereby the extraction of behavioral patterns from data replaces reliance upon predefined signatures. Machine learning models analyze file metadata, document structure, encryption indicators, and anomalous activity patterns to provide real-time detection of both known and previously unseen variants of ransomware.

This paper proposes a broad intelligent ransomware threat detection and mitigation system based on the weighted ensemble of machine learning classifiers comprising Random Forest, XGBoost, and Neural Networks. The proposed framework extracts over 50 salient behavioral and structural features from PDF, DOCX, and JSON files and applies optimized preprocessing techniques to further enhance the detection accuracy. This work aims at collecting and preprocessing ransomware datasets, extracting meaningful behavioral features, training and optimizing multiple classifiers, designing an effective ensemble strategy, creating a real-time user interface for threat detection, and evaluating system performance against existing approaches. This work has demonstrated the effectiveness of using ensemble machine learning techniques to strengthen modern cybersecurity defenses against such evolving ransomware threats.

II. METHODOLOGY

A. System Overview

The proposed system is an end-to-end machine learning approach to detecting and mitigating ransomware by examining files that are stored in PDF, DOCX, and JSON format. Here, it uses a CIC-Evasive-PDFMal2022 dataset comprising 10,025 labeled benign or infected files to learn patterns to discriminate between normal and infected files. Here, each file is broken down to isolate 85 static features, which are then processed and fed to three machine learning models, namely Random Forest, XGBoost, and a Neural Network, whose outputs are accumulated through a weighted voting mechanism. Finally, a classification result identifying a file as benign or malicious, along with a level of confidence, is provided to users via a Gradio-enabled interactive web interface hosted on Hugging Face Spaces.

B. System Architecture

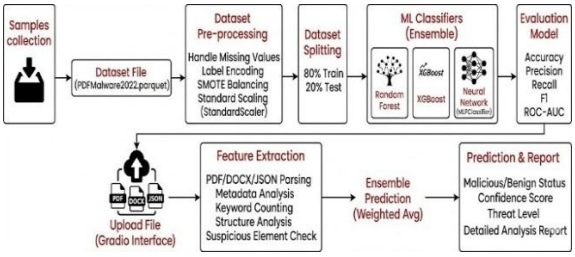

Fig. 1: The architecture of the system and flow of data in the proposed ransomware detection framework. The modular pipeline starts from dataset collection and preprocessing and passes through ensemble model training and evaluation. In its deployment, it processes user-uploaded PDF, DOCX, and JSON files to produce real-time ransomware predictions with corresponding threat levels and mitigation recommendations [1].

C. Dataset and Feature Extraction

This work utilizes the CIC-Evasive-PDFMal2022 dataset from the Canadian Institute for Cybersecurity, UNB, comprising 10,025 labeled samples (5,557 malicious and 4,468 benign) gathered using Contagio and VirusTotal. This dataset has evasive ransomware variations which can bypass traditional signature-based detection. Preprocessing: stratified 80:20 train-test splitting, imputing missing values, SMOTE oversampling on a train fold for class balancing, and StandardScaler normalization. Model development is done using 5-fold cross-validation in order to assess generalization. Feature extraction was carried out for PDF, DOCX, and JSON file formats by using format-specific parsers: PyPDF2 for PDFs, python-docx for DOCX, and regex-based analysis for JSON. Each of the samples generates an 85-dimensional feature vector comprising 12 general file attributes such as file size, metadata, and page count; 25 structural descriptors such as object/stream count, xref tables, and encryption flags; and 48 behavior and security-related indicators such as embedded scripts, auto-execution triggers, obfuscation markers, and temporal anomalies. Thus, this comprehensive representation will capture content and structural anomalies across formats and support robust ransomware classification.

D. Models of Machine Learning and Ensemble Strategy

The proposed system classifies a file as benign or malicious by using the weighted ensemble of three supervised learning models: Random Forest, XGBoost, and a feed-forward Neural Network, based on 85 extracted features[4]. The Random Forest classifier uses 100 decision trees with controlled depth (max_depth = 15) and minimum sample splits (min_samples_split = 10) to model non-linear feature interactions while preventing overfitting. Gradient boosting is used in XGBoost with regularization parameters (alpha = 1.0, lambda = 1.0) and a maximum tree depth of 6. A learning rate of 0.1 is used to ensure controlled training and generalization, making it effective to capture subtle ransomware patterns.

Neural Network-128 → 6-4 → 3-2 neurons, ReLU activation in the hidden layers, and sigmoid for the output when dealing with binary classification. To avoid overfitting, Dropout regularization is used with p = 0.3. The model is optimized using Adam with a learning rate of 0.001, whereas Early Stopping stops training after 100 epochs with no improvement.

Each model outputs a probability score for the malicious class during the inference process. Those scores are combined using the weighted ensemble:

$$P_{ensemble} = 0.4 \times P_{RF} + 0.4 \times P_{XGB} + 0.2 \times P_{NN}$$

If the $P_{ensemble} \geq 0.5$, the file is classified as malicious; otherwise, it is marked as benign. The

weights were selected based on individual model validation performance, so higher accuracy gives higher influence. This approach would couple the strengths of tree-based models, which provide good interpretability and robustness, with neural networks that have a good capacity for adaptive nonlinear pattern learning, thus improving overall accuracy, reducing false positives, and achieving more stable predictions against various ransomware variants.

E. Performance Evaluation and Results

In the proposed system, its effectiveness is checked using several performance metrics-accuracy, precision, recall, F1-score, and ROC-AUC First, all the models were trained on 80% of the dataset and then evaluated on the held-out 20% test set to check the robustness and generalization by using 5-fold cross-validation during training.

The confusion matrix in the following figure shows high counts of true positives and true negatives and a very limited number of misclassifications. The performance of the ensemble model is 99.15% compared to a Random Forest classifier with an accuracy of 99.25%, XGBoost with an accuracy of 99.10%, and Neural Network with an accuracy of 98.05%. The ROC curve depicts an AUC value of 0.9995, which shows excellent discrimination between benign and malicious files [1]. In general, feature importance analysis resulted in the identification of key factors such as StartXref, MetadataSize, JavaScript, and PdfSize, which are strongly contributing to ransomware detection. These results confirm the robustness and interpretability of the proposed approach.

$$\text{Accuracy: } ACC = \frac{TP+TN}{TP+FP+TN+FN}$$

$$\text{Precision: } P = \frac{TP}{TP+FP}$$

$$\text{Recall: } R = \frac{TP}{TP+FN}$$

$$\text{F1-score: } F1 = 2 \times \frac{P \times R}{P+R}$$

$$\text{FPR: } FPR = \frac{FP}{FP+TN}$$

$$\text{AUC: } AUC = \int_0^1 TPR(FPR)\, dFPR$$

Fig. 2. Definitions of the evaluation metrics used to measure model performance.

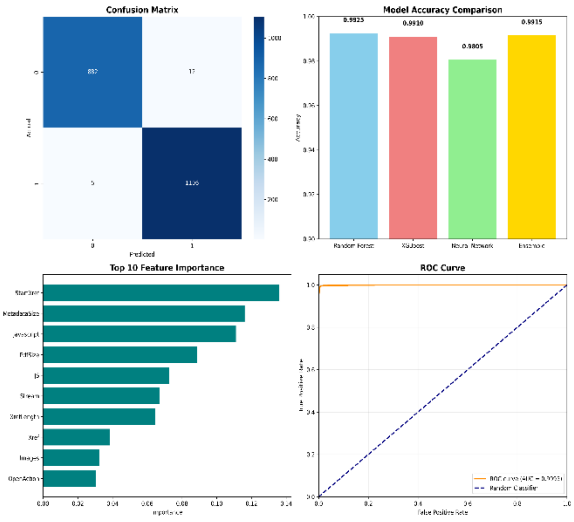

Fig. 3: Confusion matrix, model accuracy comparison, top-10 feature importance, and ROC curve for the proposed ransomware detection system

F. Real-Time Deployment Workflow

End-to-end workflow of user interaction and system response through the Gradio web interface.

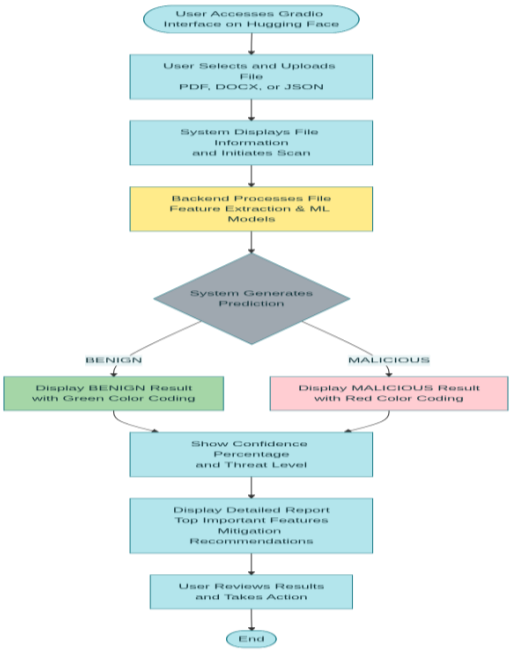

Fig. 4: End-to-end workflow of user interaction and system response through the Gradio web interface

The proposed ransomware detection solution is implemented as a web application with a Gradio interface running on a Hugging Face Space, which

allows real-time analysis of files. The user is required to upload the document files, which can be of PDF, DOCX, and JSON formats, through the web interface. The uploaded document is validated and submitted to the backend pipeline for analysis based on which the threat level, either SAFE, LOW, MEDIUM, HIGH, and CRITICAL, is assigned to the benign and malicious documents, respectively.

It shows the result of classification, the level of confidence in the prediction, and the top 10 most significant contributing features to make things clear to the user. In a malicious classification, it gives recommendations to mitigate the attack, including quarantining the files, launching a full system scan, or alerting the administration to make appropriate and correct decisions.

III. RESULTS

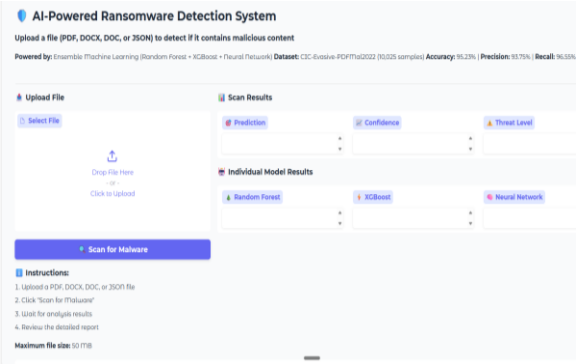

Fig. 5. Gradio-based user interface for uploading files and displaying ransomware detection results.

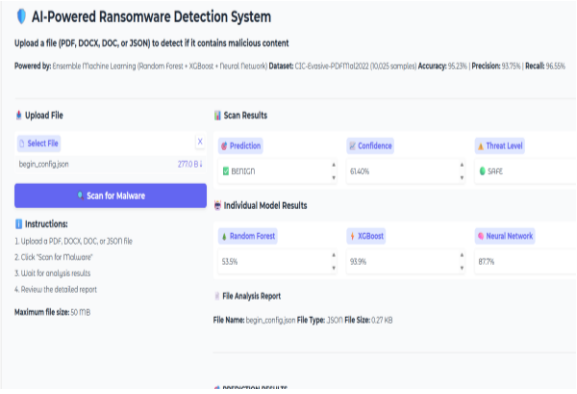

Fig. 8. Output of the ransomware detection system showing a benign classification for a JSON file with corresponding confidence score and individual model predictions.

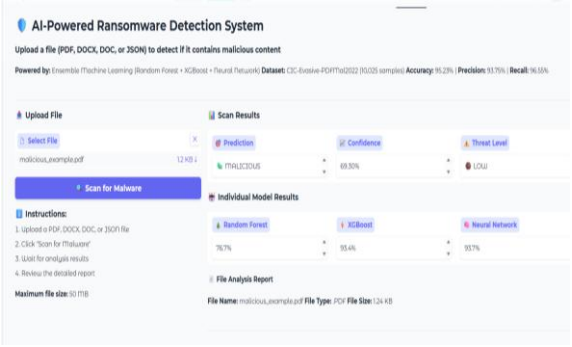

Fig. 6. Real-time scan output showing detection verdict, confidence score, threat level, and individual model predictions for an uploaded malicious file.

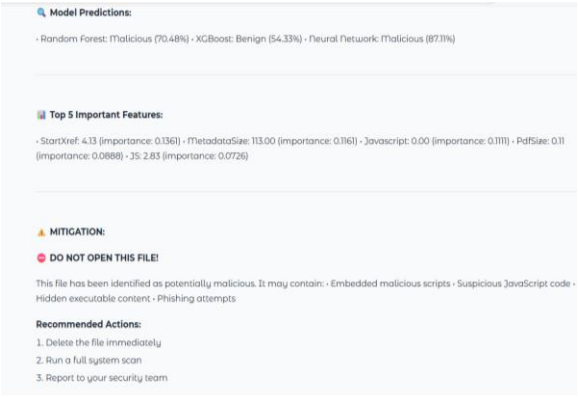

Fig. 7. Model predictions, top features, and mitigation recommendations displayed for a malicious file.

IV. CONCLUSION

The proposed work introduces a machine learning-based ransomware threat detection and remediation framework that concentrates on the analysis of PDF, DOCX, and JSON data to detect malicious patterns and categorize the file as malicious or benign. The proposed scheme utilizes the weighted combination strategy of the Random Forest classifier, the XGBoost classifier, and the Deep Neural Network to increase the detection rate and provide improved resilience compared to standalone methods. The feature importance method enhances the interpretability of the results by emphasizing the key attributes that signify malicious patterns.

Experimental results indicate the strength of performance achieved in terms of accuracy, precision, recall, and ROC-AUC metrics along with a low false-positive rate, establishing the system's efficacy for

real-time execution. The addition of an actionable mitigation module (file isolation mechanism, system scan suggestions, and incident reporting) and Gradio-based web interface execution on the Hugging Face platform increases real-time response efficiency in the real world.

In summary, the experimental results confirm the validity of using ensemble machine learning methods as an effective approach toward scalable ransomware mitigation. Future studies are planned on kernel-level data, networking patterns, adversarial attacks, as well as learning algorithms aimed at enhancing the accuracy of zero-day or evasive ransomware variant classification.

REFERENCES

[1] Author Names: Malak Aljabri, Fahd Alhaidari Wasmiyah Alqahtani, Aminah Albuainain, Jana Alshaya, "Ransomware detection based on machine learning using memory features**".** In *"Egyptian Informatics Journal"***,** pp. no. 8, vol. 25, 2024.

[2] Author Names: Daryle Smith, Sajad Khorsandroo, And Kaushik Roy, **"**Machine Learning Algorithms and Frameworks in Ransomware Detection". In *"IEE Access",* pp. no. 14, volume 10,2022.

[3] Author Names: Jannatul Ferdous, Rafiqul Islam, Arash Mahboubi, and Md Zahidul Islam, "Ransomware Detection Using Machine Learning Techniques". In *"RESEARCHER CAB"*, pp. no. 20, vol. 3,2024.

[4] Author Names: B. Idoko, F. Ogwueleka, and S. Bassey, "Systematic Literature Review on Malware Detection and Machine Learning Algorithms: Identifying Gaps for Possible Remedies," International Journal of Computers, pp. no 179–189, vol. 10, 2025.

[5] Author Names: Jamil Ispahany, Md Rafiqul Islam, Md Zahidul Islam, And M. Arif Khan, **"**Ransomware detection using machine learning: A review, research limitations and future directions"**.** In **"***IEEE Access"*, pp. no. 29, vol. 4,2016.

[6] Author Names: Salwa Razaulla Amjad Gawanmeh, Claude Fachkha, Christine Markarian, "The Age of Ransomware: A Survey on the Evolution, Taxonomy, and Research Directions". In *"IEEE Access"*, pp. no. 26, vol. 11,2023.

[7] Author Names: Amjad Alraizza 1, and Abdulmohsen Algarni, "Ransomware Detection Using Machine Learning: A Survey**".** In **"***MDPI Access"*, pp. no. 24.

[8] Author Names: Adrian Brodzik1, Tomasz Malec-Kruszy´ nski1, "Ransomware Detection Using Machine Learning in the Linux Kernel**".** In *"arXiv"***,** pp. no. 8, vol. 1,2024.

[9] Author Names: Amardeep Singh, Zohaib Mushtaq, "Enhancing Ransomware Attack Detection Using Transfer

[10] Learning and Deep Learning Ensemble Models on Cloud-Encrypted Data". In *"MDPI Access"*, pp. no. 31, vol.12,2023.

[11] Author Names: Amardeep Singh, Zohaib Mushtaq, "Ransomware Threat Mitigation Through Network Traffic Analysis and Machine Learning Techniques". In *"MDPI Access"*, pp. no. 12.

[12] Author Names: A. Singh and Z. Mushtaq, "Ransomware Threat Mitigation Through Network Traffic Analysis and Machine Learning Techniques," MDPI Access, pp. 12.

[13] Author Names: A. Kapoor, A. Gupta, R. Gupta, S. Tanwar, G. Sharma, and I. E. Davidson, "Ransomware Detection, Avoidance, and Mitigation Scheme: A Review and Future Directions," Sustainability, pp. 8.

[14] Author Names: J. S. Sankar, M. Lohitha, B. Yugesha, B. Y. Chowdary, and K. P. Raj, "Ransomware Detection using Machine Learning," IJARSCT, pp. 218–224.

[15] Author Names: F. R. Alzaabi and A. Mehmood, "A Review of Recent Advances, Challenges, and Opportunities in Malicious Insider Threat Detection Using Machine Learning Methods," IEEE Access, pp. 1–18, 2024.