RESEARCH PAPER

PDF Malware Detection Using CNN Based Machine Learning

M.Varatharaj [1*], Darshan K [2*], Ajay R [3*] , Dharnis K [4*]

[1*] Associative Professor, Department of Computer Science and Engineering, V.S.B College of Engineering Technical campus, kinathukadavu, Coimbatore, Tamilnadu, India.

[2*] UG Scholar, Department of Computer Science and Engineering, V.S.B College of Engineering Technical campus, kinathukadavu, Coimbatore, Tamilnadu, India

[3*] UG Scholar, Department of Computer Science and Engineering, V.S.B College of Engineering Technical campus, kinathukadavu, Coimbatore, Tamilnadu, India.

[4*] UG Scholar, Department of Computer Science and Engineering, V.S.B College of Engineering Technical campus, kinathukadavu, Coimbatore, Tamilnadu, India.

Abstract-- Portable Document Format (PDF) files represent a popular choice for sharing digital documents because of their flexibility and platform independence. Besides their extensive use, PDF files have also become a favorite medium for malware distribution as hackers embed malicious scripts and exploit vulnerabilities in PDF readers. Conventional malware detection methods mostly depend on signature, based or rule, based approaches, which fail to detect new and zero, day attacks. To mitigate these restrictions, this paper introduces an automated PDF malware detection framework utilizing Convolutional Neural Networks (CNNs) where the proposed model not only looks at PDF files structurally but also at the byte, level allowing it to learn complex patterns automatically without the need for manual feature extraction. Hence, the model becomes capable of identifying obfuscated and even new malware samples effectively. In addition, to resolve the issue of deep learning models being black boxes, the paper details further incorporation of explainability analysis via SHAP to understand the model's prediction mechanism and to identify the features contributing most to classification decisions. Experimental results demonstrate that the newly proposed system performs very well in detecting malware while at the same time allowing for interpretability. The overall framework thus offers a promising and transparent method for the improvement of security measures against the ever, changing threats of PDF malware.

Keywords-- *PDF Malware Detection, Convolutional Neural Network, Deep Learning, Explainable Artificial Intelligence, SHAP, Cybersecurity.*

How to cite this article: Varatharaj M, Darshan K, Ajay R, Dharnis K. PDF malware detection using CNN based machine learning. Int J Drug Deliv Technol. 2026;16(74s): 2063-2070. DOI: 10.25258/ijddt.16.74s.167

INTRODUCTION

Portable Document Format (PDF) files are one of the most popular and widely used file formats when compared to other file formats in communication and information sharing on the Internet in terms of portability and reduced size of the files, and the files are able to maintain the format of the documents on any platform and operating system, including the hardware and software combinations [1], [2]. The PDF files have become highly popular in terms of email communication, research publications, businesses, governments, and e-commerce due to the features and benefits of the files. However, the rising popularity of PDF files has also encouraged malicious users to exploit the format for their malicious purposes. In fact, malicious users often exploit the PDF format for spreading malware by embedding malicious scripts, JavaScript codes, and objectsin the structure of the PDF file [5], [22]. When a malicious PDF file is executed by the user, malicious scripts may execute in the background without the knowledge of the user, which may result in unauthorized access of the system, data theft, and even system compromise [7], [29].

The security threats associated with the PDF format underscore the need for developing an efficient mechanism to detect malicious PDF files. Traditionally, various mechanisms have been proposed for malware detection in PDF files. The most frequently used mechanisms for malware detection are signature-based detection and rule-based detection. These detection mechanisms are used to identify malicious PDF files by comparing them with previously identified malicious signatures and rule sets, respectively [4], [30]. Although signature-based detection and rule-based detection are effective detection mechanisms for identifying previously identified malware, they are unable to identify newly emerging malware and zero-day attacks that lack a signature [10], [33].

The malware authors are changing their malicious code to evade detection by traditional detection mechanisms; thus, traditional detection mechanisms are no longer reliable for detection. In order to address the aforementioned shortcomings of the aforementioned techniques, researchers have proposed various machine learning-based malware detection techniques that consider the structural and behavioral properties of PDF files [11], [23]. Machine learning-based malware detection techniques have been proven to be effective in improving the accuracy of malware detection compared to the aforementioned techniques [3], [8].

However, most of the proposed machine learning-based malware detection techniques heavily rely on manual feature extraction and engineering, which is considered time-consuming and requires domain expertise [31]. In recent years, various deep learning techniques have been proposed and have shown promise in the domain of malware detection based on their ability to learn from the data in an efficient manner. Among these techniques, Convolutional Neural Networks have shown promising results in the analysis of the byte-level representations of files in the classification of malware. These techniques

have shown promising results in the identification of patterns that may indicate the presence of malware, thereby eliminating the need for manually engineered features. Various techniques have also been proposed in the visualization of malware in the form of images. Although these models were successful in their task, they are classified as black-box systems. This means that it is difficult to understand the decision-making processes used by these models. This could be a problem for cybersecurity experts who must be able to understand why a certain file is classified as malicious. In order to address the aforementioned problem, a set of techniques called Explainable Artificial Intelligence (XAI) has been proposed to improve the interpretability of machine learning and deep learning techniques [32].

In particular, it has been shown that the SHAP method has been successful in interpreting the results of machine learning and deep learning techniques by calculating the feature contribution to the final classification result [10]. Recent research has shown that the application of deep learning techniques for malware detection can lead to better results in terms of accuracy and interpretability [18], [19], [20]. Apart from the use of deep learning techniques, various research has been conducted on the use of hybrid and hierarchical machine learning techniques for improving malware detection and cybersecurity systems [16], [21].

These techniques use various types of machine learning techniques to improve the efficiency and flexibility of the malware detection system. Moreover, modern cybersecurity systems emphasize the importance of developing efficient malware detection systems that can handle large amounts of data and detect complex types of attacks in real-time systems [29].Inspired by the challenges and importance of the topic, the proposed research aims to develop an efficient AI-based system that can be used to detect malicious PDF documents using Convolutional Neural Network and Explainable AI techniques. The proposed system will be efficient in improving the accuracy of the malware detection system. The main contributions of the proposed research are discussed as follows: A CNN-based malware detection model with the capability to learn the characteristics of PDF files at the byte and structural levels for effective malware detection and threat identification [12], [26].

The rest of the paper is structured as follows. Section II presents an in-depth discussion on the related work in PDF malware detection and the role of machine learning in the security system. Section III presents the methodology and system architecture. Section IV presents the experimental results and performance evaluation. Finally, Section V presents the conclusion of the paper and the future research direction.

I. Related Works

Portable Document Format (PDF) files in digital communication and sharing of documents, attackers have begun using PDF files as a means of spreading malware. Consequently, a lot of research has been conducted on the detection of malicious PDF files using various security mechanisms. Most of the initial research work used signature-based detection mechanisms, which areimplemented in most of the traditional antivirus systems. Signature-based mechanisms work on the principle of comparing the signatures of incoming files with known signatures of malicious files stored in a database. If a match is found, the file is marked as malicious. Signature-based mechanisms have proven effective in detecting known malicious attacks and are popular in commercial products [1],[30]. However, these mechanisms have various disadvantages in dealing with new types of attacks, i.e., zero-day attacks, where the signature of the malware is not known [4], [33].

In order to avoid these drawbacks, static analysis-based methods for detecting malicious PDF files have been proposed. The static analysis-based method for detecting malicious PDF files involves the analysis of the internal structure of the PDF files. In the research proposed in [1], [17], the analysis of the structural features and metadata information related to the PDF files has been proposed as a method for detecting malicious PDF files. Moreover, in the research proposed in [2], the analysis of the hierarchical structure of the documents has been proposed as a method for detecting malicious PDF files. In other research, the analysis of the JavaScript code of the PDF files has been proposed as a method for detecting malicious PDF files. Although the analysis of the PDF files using the static analysis-based method provides a clear idea about the internal PDF file behavior, the method has some drawbacks. In order to improve the performance of the detection of the system, various research studies have explored the use of machine learning-based detection mechanisms.

Machine learning-based detection mechanisms use the previously labeled data sets to learn the pattern and classify the PDF files as malicious or benign based on the features of the files. Various research studies have used machine learning algorithms such as Support Vector Machines, Random Forest, and Decision Trees to identify the presence of malware in the PDF files [23],[31]. The proposed detection mechanisms analyze the structural features, behavioral features, and statistical features of the PDF files. Various research studies have also explored the use of classification mechanisms to identify the presence of encrypted malware traffic and network-based malware behavior using machine learning-based detection mechanisms [8], [27]. Although the proposed detection mechanisms using machine learning algorithms improve the performance of the detection of the system compared to the traditional mechanisms, the performance of the detection of the system depends on the features of the files. In recent times, deep learning techniques have attracted considerable attention in the field of malware detection due to their ability to automatically learn complex patterns from raw data [9], [28].

Among various techniques of deep learning, Convolutional Neural Network (CNN) has proven its effectiveness in the analysis of malware using the byte-level representations of files. Saxe and Berlin presented the use of deep neural networks for the detection of malware using binary program features [3]. Further research was conducted on the use of CNN-based architectures for the analysis of malware

patterns and detection of malware [11], [26]. Additionally, the use of visual representations of malware was also presented by transforming binary files into images and using deep learning techniques for the detection of malware [13], [24]. Some research has also been conducted on hybrid and advanced deep learning models that can be utilized to improve the malware detection results. For example, hybrid CNN-based models and lightweight deep learning models have been proposed in order to improve the malware detection results in terms of classification accuracy while keeping the models computationally efficient [15], [21].

Moreover, malware detection models based on machine learning that can be utilized to identify malicious PDF files have also been proposed in order to identify suspicious PDF structures and behaviors [12], [25]. Some research has also been conducted on proposing cybersecurity frameworks and models that can be utilized to identify different types of cyber threats in distributed environments [16], [29]. Most of the deep learning models behave as a black box, and it becomes hard for the analyst to understand the reasoning behind the classification results. This may reduce the trust of cybersecurity professionals in automated systems. To overcome these problems, Explainable Artificial Intelligence (XAI) methods have been proposed, which help in improving the interpretability of machine learning models [32]. Among these methods, SHAP (SHapley Additive explanations) has been used as a popular method to explain the results of machine learning models by measuring the contribution of each feature in the classification results [10]. Recently, various studies have shown the effectiveness of using explainability methods in improving the reliability of deep learning-based malware detection systems [18], [19].

II. Proposed Methodology

A. Overview

The proposed methodology suggests an AI-based model for detecting malicious PDF documents using Convolutional Neural Networks (CNN) in association with Explainable Artificial Intelligence (XAI). The idea behind this model is to automatically detect harmful PDF documents based on structural features of the documents. The traditional methods of detecting malware usually involve signatures, but these methods are not effective in dealing with new malware.

B. System Architecture

The architecture of the proposed system is illustrated in Fig. 1**.** The system consists of several components including PDF input module, preprocessing module, feature extraction module, CNN classification model, and explainability module**.**

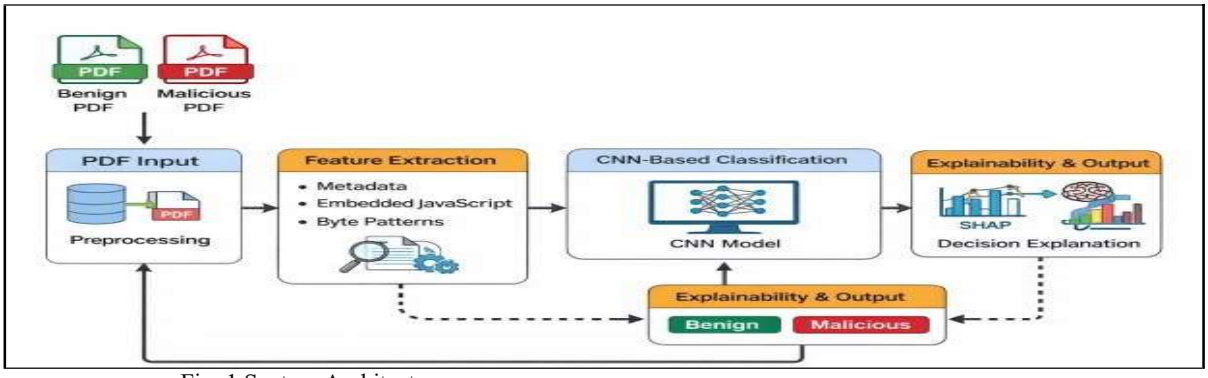


Fig. 1 System Architecture

1. PDF Input Module
   - The system can handle PDF files that users can upload or can be retrieved from a dataset. The files can be both harmless and malware.
2. Preprocessing Module
   - In this module, the data is preprocessed by cleaning up the data from the PDF files in a manner that is conducive for analysis. Some of the operations performed in this module include removing unwanted noise from the data and converting PDF data for analysis.
3. Feature Extraction Module
   Important features from the PDF files are extracted in this module. The features include:
   - Metadata information
   - JavaScript code embedded in the PDF files
4. CNN-Based Classification Module
   - The extracted features from the PDF files are used in this module for analysis by a Convolutional Neural Network (CNN). The CNN model learns hidden patterns in
5. Explainability Module (SHAP)

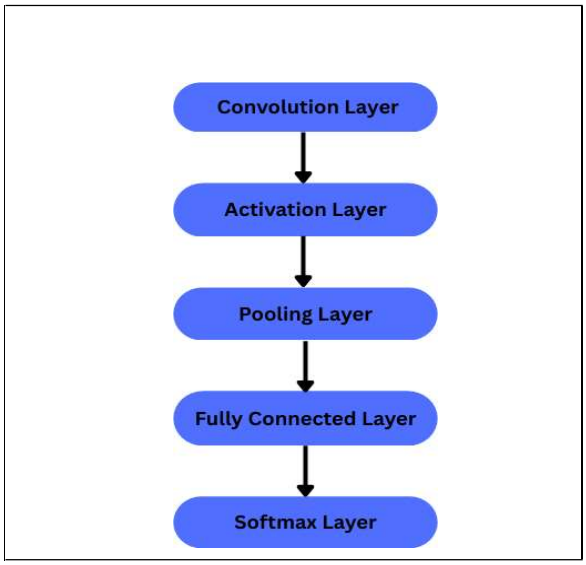
- In this module, a black-box model is used for addressing the black-box problem in CNN models. The model is based on SHAP for feature contribution analysis.

B. CNN-Based Malware Detection Model

The CNN model is used for learning patterns from the features of the PDF files in Fig. 2. The CNN architecture comprises different layers of processing for the data.

1. Convolution Layer
   - Used for extracting local patterns from feature maps.
2. Activation Layer
   - Adds non-linearity to the CNN model.
3. Pooling Layer
   - Dimensionality reduction.
4. Fully Connected Layer
   - Combines all learned features for classification.
5. Softmax Layer
   - For probability generation for the classification.

the process of feature extraction on the structural features and the content features of the PDF file.



The proposed system uses a Convolutional Neural Network (CNN) to learn complex patterns from PDF files. During training, the CNN automatically extracts important structural and byte-level features from the input data.

The convolution operation used for feature extraction is defined as:

$$F(x) = \max(0,x) \quad (1)$$

This function enables the CNN to learn non-linear relationships within the features.

Classification errors are minimized by utilizing the Cross-Entropy Loss function:

$$L = -\sum (\text{Actual Value} \times \log(\text{Predicted Value})) \quad (2)$$

The weights are updated by utilizing the Gradient Descent optimization:

$$W = W - \alpha \nabla L \quad (3)$$

ALGORITHM 1.

Fig. 2.Malware Detection Model

E. Explainable Decision Support using SHAP

In order to provide a more transparent decision-making process, a SHAP approach is incorporated into the proposed system.

ALGORITHM 2.

1. Collect the dataset containing benign and malicious PDF files.
2. Assign labels to each PDF file:
   - Benign PDF file: y=0
   - Malicious PDF file: y=1
3. Pre-process the PDF file.
4. Extract important features from the PDF such as:
   - Metadata information
5. Convert the extracted features into numerical feature vectors suitable for CNN input.
6. Define CNN architecture.
7. Split the dataset into the training and testing sets.
8. Train the CNN model using the training dataset.
9. Compute the network weights by using the back propagation algorithm.
10. Compute the accuracy of the malware detection system by using the testing dataset.
11. Apply SHAP Explainable AI.
12. Store the trained CNN model.

D. Feature Extraction and Data Representation

Feature extraction is a critical process in the development of a malware detection system. The system is able to carry out Receive the input PDF File.

1. Perform Pre-processing on the input PDF file.
2. Extract structural features as well as byte-level features from input files.
3. Transform PDF file features into CNN format feature vector.
4. Load the trained CNN Model.
5. Input feature vector into trained CNN model.
6. Calculate probability score for malware detection
7. Apply the classification rule
8. Display the classification result.
9. Generate a security alert if malicious content is detected.

III. PERFORMANCE ANALYSIS

A. Detection Accuracy Analysis

Detection accuracy measures how correctly the proposed system classifies PDF files as benign or malicious. The accuracy is calculated using the following equation:

$$\text{Accuracy}=\frac{TP+TN}{FP+FNTP+TN}$$

In Fig. 3, the x-axis represents the number of training iterations and the y-axis represents the detection accuracy of the CNN model.

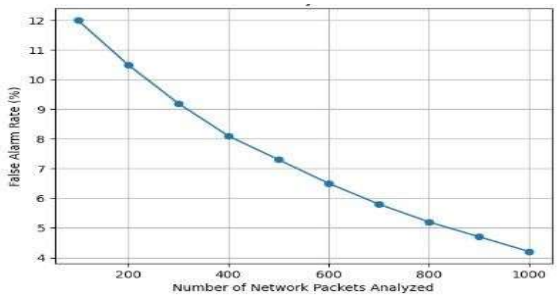


Fig. 3.Detection Accuracy Analysis

B. Detection Time Analysis

The term "detection time" refers to the time taken by the system to examine a PDF file and identify whether it is a malware or a normal PDF file. The significance of this parameter is that it can help prevent malware from executing and causing damage to the system. The proposed CNN-based PDF detection system works by examining the PDF file by extracting key features from it and classifying it using a CNN model. The results obtained from the Fig. 4 Indicate that the proposed system can successfully identify a malware PDF file in a short period of time. As the number of PDF files increases, the detection time is seen to increase slightly by the proposed system. This shows that the proposed approach can be used for malware detection applications.

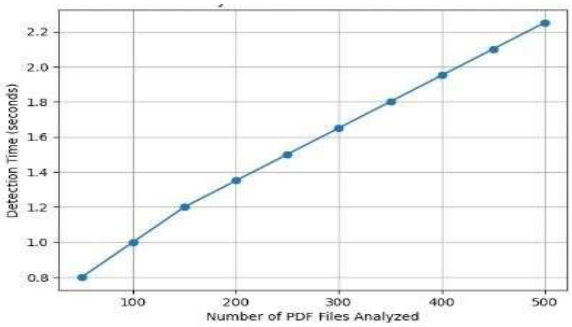


Fig.4. Detection Time Analysis

C. Scalability Analysis

Scalability is the capacity that the proposed system has in terms of maintaining efficient performance as the size of the data set increases. In the proposed malicious PDF detection system, the scalability is determined based on the capacity that the proposed system has in terms of efficiently performing when the number of PDF files increases. As the number of PDF files increases, the system is still able to efficiently perform the operation on the PDF files while maintaining efficient performance with minimal reduction in the efficiency of the detection system. Fig. 5 proves that the proposed system can be used in the real world since it can efficiently perform the operation in the real world.

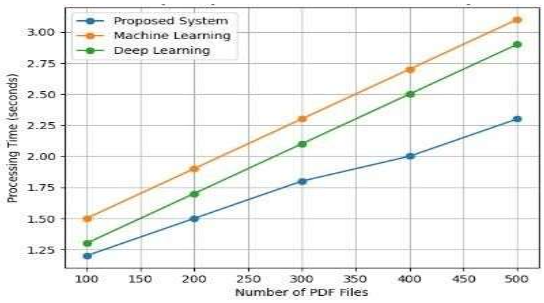


Fig.5.Scalabiity Analysis

This Table I compare different malware detection techniques based on performance metrics such as accuracy, speed, explainability, and scalability. The proposed CNN + SHAP model achieves high accuracy and scalability while also providing improved explainability compared to traditional and deep learning method

Table I Comparison of Malware Detection Methods

Method	Detection Accuracy	Detection Speed	Explainability	Scalability
Signature-based Detection	Low	Fast	Low	Low
Machine-learning method	Medium	Medium	Medium	Medium
Proposed CNN+SHAP Model	High	Fast	High	High

IV. Conclusion

In this research, the authors have presented a framework for the detection of malicious PDF documents using the AI technique. In the proposed system, the CNN technique has been employed for the analysis of the byte-level and structural features of the PDF files in order to detect the malicious content effectively. This technique minimizes the dependency on the feature engineering of the data and maximizes the potential of detecting complex types of malware. To minimize the lack of transparency in the deep learning technique, which is commonly observed in the AI technique, the XAI technique has been employed in the proposed system using the SHAP technique. This technique provides significant explanations regarding the predictions made by the system and the contribution of the different features in the classification of the data. Overall, it is believed that the proposed framework can contribute to strengthening cybersecurity by facilitating early detection of malicious PDF files, as well as obtaining better insight into the decision-making process of the model. For future works, it is suggested that the system can be extended by using larger datasets, deeper learning, and real-time deployment to improve detection accuracy.

REFERENCES

[1] M. Smutz and A. Stavrou, "Malicious PDF detection using metadata and structural features," Proc. 28th Annual Computer Security Applications Conference, 2012, pp. 239–248.

[2] N. Šrndić and P. Laskov, "Detection of malicious PDF files based on hierarchical document structure," Proc. Network and Distributed System Security Symposium (NDSS), 2013.

[3] J. Saxe and K. Berlin, "Deep neural network based malware detection using two dimensional binary program features," Proc. Int. Conf. on Malicious and Unwanted Software, 2015.

[4] Y. Ye, T. Li, D. Adjeroh, and S. Iyengar, "A survey on malware detection using data mining techniques," ACM Computing Surveys, vol. 50, no. 3, pp. 1–40, 2017.

[5] M. Z. Rafique and J. Caballero, "Firma: Malware clustering and network signature generation with mixed network behaviors," Proc. RAID Conference, 2013.

[6] Mary, P.A., Maalini, D., Manivannan, K., Umamaheswari, K., Balaji, S. and Deepak, R., "A Secure Aware Routing Protocol for Efficient and Reliable FPANETS," in 2024 First International Conference on Software, Systems and Information Technology (SSITCON) (pp. 1-6). IEEE, 2024.

[7] A. Kolosnjaji, A. Zarras, G. Webster, and C. Eckert, "Deep learning for classification of malware system call sequences," Australasian Joint Conference on Artificial Intelligence, 2016.

[8] B. Anderson and D. McGrew, "Machine learning for encrypted malware traffic classification," Proc. ACM Workshop on Artificial Intelligence and Security, 2016.

[9] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[10] S. Lundberg and S. Lee, "A unified approach to interpreting model predictions," Advances in Neural Information Processing Systems, 2017.

[11] Pathak, D., Gowda, D., Manivannan, K., Aghav, S., Srinivas, V. and Gireesh, N., "Advanced Machine Learning Approaches to Evaluate User Feedback on Virtual Assistants for System Optimization," in 2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS) (pp. 1140-1147). IEEE, 2024

[12] R. Fettaya and Y. Mansour, "Detecting malicious PDF using convolutional neural networks," 2020.

[13] L. Nataraj, S. Karthikeyan, G. Jacob, and B. Manjunath, "Malware images: Visualization and automatic classification," Proc. ACM Symposium on Visualization for Cyber Security, 2011.

[14] M. Abadi et al., "TensorFlow: A system for large-scale machine learning," Proc. USENIX Symposium on Operating Systems Design and Implementation, 2016.

[15] S. U. Jan, S. Ahmed, V. Shakhov, and I. Koo, "Toward a lightweight deep learning-based malware detection system," IEEE Access, vol. 7, pp. 1–12, 2019.

[16] Manivannan, K., Jayanthi, S., Kavitha, T., Geetha, M., Mary, P.A. and Nelson, S., "Towards Adaptive and Scalable DDoS Detection in Distributed System using Tuned Hierarchical Machine Learning Techniques," in 2025 International Conference on Computing and Communications (COMPUTINGCON) (pp. 1-6). IEEE, 2025.

[17] Y. Yamaguchi, Y. Miyake, and H. Nakano, "Detecting malicious PDF documents using machine learning techniques," International Journal of Information Security, 2016.

[18] Jino, R.F., Paulsamy, A.M., Shanmugam, G. and Vishwakarma, R.K., "Enhancing network security: a deep learning-based method to detect and diminish attacks,"International Journal of Electronic Security and Digital Forensics, 2025.

[19] J. Jayapradha et al., "Deep learning combined with a technique for detecting viruses in PDFs and URLs," Journal of Neonatal Surgery, 2025.

[20] P. Reddy et al., "Explainable machine learning models for detecting malware in PDF files," International Journal of Computational Learning & Intelligence, 2025.

[21] M. Hashmi, "Malware detection and classification using hybrid CNN and machine learning," International Journal of Intelligent Systems and Applications in Engineering, 2023.

[22] T. M. Mohammed et al., "HAPSSA: Holistic approach to PDF malware detection using signal and statistical analysis," 2021.M. Kaif et al., "PDF malware detection using machine learning," International Journal for Research in Applied Science and Engineering Technology, 2025.

[23] A.Almusawi, "Visual malware detection by deep learning techniques," Journal of Millimeterwave Communication, 2021.

[24] X. Wang et al., "Machine learning-based detection method for malicious PDF files," Applied Soft Computing, 2026.

[25] D. Sahu and S. Tripathy, "CNN-based malware detection using image analysis framework," Springer Lecture Notes in Networks and Systems, 2025.

[26] A. Jain et al., "Stacked autoencoder based neural network for identifying malicious traffic," International Journal of Intelligent Systems and Applications in Engineering, 2023.

[27] Y. Ye et al., "Intelligent malware detection using deep learning techniques," IEEE Security & Privacy, 2018.

[28] Sathyanarayanan, M., Manivannan, K., Mithin, K., Manoj, G., Mugesh, M. and Kishore, N., "Design and Implementation of a Cybersecurity Framework for Encrypted File Management System," in 2025 6th International Conference for Emerging Technology (INCET) (pp. 1-6). IEEE, 2025.

[29] A. Shabtai et al., "Machine learning approaches for malware detection," Information Security Technical Report, 2012.

[30] S. Garcia et al., "An empirical comparison of machine learning models for malware detection," Computers & Security, 2014.

[31] H. Saidane et al., "Explainable AI techniques for cybersecurity applications," IEEE Access, 2022.

[32] R. Doriguzzi-Corin et al., "Deep learning approaches for network attack detection," IEEE Transactions on Network and Service Management, 2020.