International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 06 June 2025, pp: 01-09 www.ijaem.net ISSN: 2395-5252

Multi-View Feature Fusion for Effective Malware Classification Using Deep Learning

1st M.D.Shelar, 2nd Aishwarya.C.Ayare, 3rd Trushna.S.Bagade, 4th Shivashree.P.Nimbalkar, 5th Muskan.A.Mujawar

Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering and Technology, Baramati, Baramati (Pune), India

department of Computer Engineering, VidyaPratishthan's Kamalnayan Bajaj Institute of Engineering and Technology, Baramati, Baramati (Pune), India

Department of Computer Engineering, VidyaPratishthan'sKamalnayan Bajaj Institute of Engineering and Technology, Baramati, Baramati (Pune), India

Department of Computer Engineering, VidyaPratishthan's Kamalnayan Bajaj Institute of Engineering and Technology, Baramati, Baramati (Pune), India

Department of Computer Engineering, VidyaPratishthan's Kamalnayan Bajaj Institute of Engineering and Technology, Baramati, Baramati (Pune), India

Date of Submission: 28-05-2025 Date of Acceptance: 08-06-2025

ABSTRACT—The rapid increase in global malware infections has necessitated the development of robust malware detection systems to mitigate threats, such as ransomware and crypto-miners, that aim for financial gain. Deep learning-based Convolutional Neural Network (CNN) model for classifying malware in Portable Executable (PE) binary files using a fusion feature set approach. An extensive evaluation of various deep learning architectures and machine learning classifiers, including Support Vector Machines (SVM), was conducted across multi-aspect feature sets encompassing static, dynamic, and image-based features, ultimately selecting the CNN model for optimal performance. The model achieved an accuracy using fusion feature sets, demonstrating its robustness and generalizability on unseen malware datasets.

Index Terms—Deep learning, malware detection, Malicious File
Detection, Multi-View Learning, Static and Dynamic Analysis

I. INTRODUCTION

The growing complexity and diversity of malware types and attack methods have made malware detection a fundamental component of cybersecurity. Malware, or malicious software, is designed to infiltrate a computer without the user's consent or awareness, posing significant risks to data security, network integrity, and system functionality. Malware can serve various malicious purposes, such as gaining unauthorized access to private networks, stealing sensitive information, hijacking system resources, or disrupting computing and communication processes. This project seeks to address the challenges posed by such sophisticated threats through a hybrid detection approach that leverages both Convolutional Neural Networks (CNN) and Support Vector Machines (SVM) to analyze malware from multiple data perspectives. CNN is utilized for imagebased malware detection, enabling the system to recognize visual patterns that distinguish malicious from benign files by converting malware binaries into grayscale images. SVM, on the other hand, is applied to textual inputs, such as API call sequences and PE header details, allowing it to detect structural and behavioral patterns within the malware that are not visually identifiable. By combining CNN and SVM in a unified model, this approach captures both static and dynamic features of malware, thereby enhancing detection accuracy across a variety of data formats. The objective is to develop a reliable and robust detection system that can efficiently identify and classify malware patterns.

II. LITERATURE SURVEY

1. A Malware Detection Approach Based on Deep Learning and Memory Forensics

The paper through Shuhui Zhang, Changdong Hu, Lianhai Wang, Miodrag J. Mihaljevic, ShujiangXu, and TianLan, titled ”A Malware Detection Approach Based on Deep Learning and Memory Forensics”, introduces a deep masteringbased totally solution for malware detection, specifically targeting the project of fileless malware. By utilizing convolutional neural networks (CNNs) and reminiscence forensics, this method successfully detects malware that resides only in reminiscence, making traditional signature-primarily based detection methods insufficient. The authors compiled a dataset of Portable Executable (PE) memory fragments and trained a CNN version to classify benign and malicious fragments, attaining an accuracy of ninety seven.FortyeightThis model, designed for performance and precision, demonstrates good sized ability in coping with dynamic and fileless malware threats, improving detection costs beyond traditional system getting to know strategies.

2. Malicious Detection

In their study, XianchunZheng and HuiLi address the growing challenge of detecting malicious encrypted traffic in the context of widespread encryption with advanced security skills and Internet expansion and causing the Using deep learning techniques, research focuses on data preprocessing and model selection. Their proposed model, MET-FMF, shows that malicious activity in encrypted networks can be detected accurately. However, scalability is a concern, as the performance of METFMF needs further validation in varying network sizes. Furthermore, the processing of encrypted vehicles must comply with privacy laws to avoid disclosure of sensitive information, emphasizing data privacy as a key consideration. 3.Android malware detection

In their paper, Zaiteg Singh, Farman Ali, Deepak Thakur, Tanya Gera, Babar Shah, and Tamer Abuhed propose Android malware detection by converting malware samples into grayscale images and using feature fusion technique This approach yield CNNs of handcrafted features (GLCM, GIST, LBP) to increase classification accuracy It combines with features known from Overall, research no reliability of DREBIN dataset limits its scope, and manual feature methods may miss some bad tendencies, making the system effective in collecting new malignancies is uncertain.

4. MalJPEG: Machine Learning-Based Solution for the Detection of Malicious JPEG Images

The paper ”MalJPEG: Machine Learning-Based Solution for the Detection of Malicious JPEG Images”byAviad Cohen, NirNissim, and Yuval Elovici presents the first method for detecting malicious JPEG images using machine-learning JPEG image consumption role, relating to how popular, presumably for And for security reasons, factors leading to malware attacks. Previous work on malware detection focuses mostly on executable files, PDFs, and Office documents, while most of the detected malicious images are limited to steganography The authors use MalJPEG, an example resulting in ten discriminating factors removed from JPEG file systems. When more than 156,000 images were tested, MalJPEG achieved a high detection accuracy (AUC of 0.997) and 95.1true positive rate using the LightGBM classification Proven limitations include scalability challenges and the need for the model adapts to evolving image-based malwarethreats.

5. A Multi-Visual Feature Fusion Approach to Effective Malware Classification Using Deep Learning

RajasekharChaganti, Vinayakumar Ravi, and Tuan D. In a paper titled ”A Multi-Visual Feature Fusion Approach to Effective Malware Classification Using Deep Learning” by

Pham, a deep learning-based framework a combines multiple visual features fusion for malware classification . It addresses the limitations of traditional methods for bug detection by extracting static, dynamic, and image-based feature sets from portable executable (PE) files, signature-based, inferencebased literature survey in this research using the Convolutional Neural Network (CNN) model , from sandboxing techniques, to state-of-the-art feature driven machine learning (ML) and deep learning (DL) techniques so fusion to improve accuracy, focusing on the design of malware detection and detection methods strength. Previous studies tend to focus on single features, whereas this study aims to increase visibility by combining multiple features The results show improved classification accuracy, especially in of malware complexity, which demonstrates the ability of hybrids to adapt to malware threats 6.MalFSM: Feature Subset Selection Method for Malware Family Classification

The paper ”MalFSM: Feature Subset Selection Method for Malware Family Classification” presents a framework to improve malware classification by reducing feature dimensionality MalFSM selects 16 major opcode features and together with a large file, line count metadata , creates an 18- feature subset . Tested on

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 06 June 2025, pp: 01-09 www.ijaem.net ISSN: 2395-5252

Microsoft Malware Classification dataset, MalFSM achieves high accuracy (upto 98.6reduces classification time and resources compared to other methods, making it an effective method for malware family classification.

III. PROPOSED SYSTEM

A. Detailed Proposed System

The proposed system introduces a multi-view feature fusion framework for malware classification using deep learning, particularly Convolutional Neural Networks (CNNs), complemented by Support Vector Machines (SVMs). The framework is designed to analyze Portable Executable (PE) binary files from various perspectives, incorporating static features (like PE section and import table data), dynamic features (such as API call sequences), and image-based representations of binaries. This multi-faceted approach allows the model to learn rich patterns that differentiate malicious from benign files with high accuracy.

The system demonstrates strong generalization on unseen data, confirming its capability to adapt to evolving malware threats. Through fusion of diverse feature types, the system reduces false positives and negatives significantly compared to single-view or conventional models. It supports efficient deployment in real-world cybersecurity environments and forms a foundation for future enhancements involving additional feature types or more complex deep learning architectures. This ensures scalability and relevance in dynamic threat landscapes. The fused feature vector is passed through the trained model, and the system classifies the input file as either ”malware” or ”benign”. The decision is made based on the learned patterns from the training phase. The model’s robustness ensures low false positives/negatives and strong performance even on unseen or evolving malware types.

B. Software Requirements

- Programming Languages

1) Python for deep learning model development
2) JavaScript , HTML and CSS for client-side interactivity and visualization.

- Libraries & Frameworks

1) TensorFlow / Keras (or PyTorch):Provides high level APIs For model architecture, training loops and evaluation metrics.
2) scikit-learn: For implementing and training machine learning models (SVM etc.).
3) Pandas and NumPy: For data manipulation, preprocessing, and analysis.
4) Matplotlib: Helps in understanding model behavior and results.

- Database

1) SQLite: For managing user queries, storing processed data, and saving detection results.

- Development Tools

1) Integrated Development Environment (IDE) such as Spyder.

C Hardware Requirements

- Hardware

1) Processor: Multi-core CPU
2) Memory: Minimum 8 GB RAM for basic deployments; 16 GB or more recommended
3) Storage: SSD with at least 256 GB of storage space
4) Network: High-speed internet connection

- Development and Testing Environment

1) A standard workstation or laptop capable of running Python and associated libraries.
2) GPU (optional): For training more complex models

IV. METHODOLOGY

A. Structure

The architecture of the system is multi-layered and consists of four main parts:

1.Data Preparation Layer : This layer includes the Dataset and Pre-processing modules. The dataset provides the raw data, which is then cleaned and normalized during preprocessing. This ensures that the data is consistent and ready for further analysis.

2. Feature Engineering Layer: In this layer, the system performs Feature Extraction and Feature Fusion. Feature extraction involves identifying and collecting relevant characteristics from the data, such as static features (metadata, file properties) and image features (visual representations). Feature fusion then combines these different types of features into a single, comprehensive set, enhancing the model’s ability to distinguish between benign and malicious files.
3. Model Training Layer: This part consists of the Algorithms Used and Train Model modules. Various machine learning algorithms—such as Convolutional Neural Networks (CNN), Support Vector Machines (SVM), Random Forest, and Deep Neural Networks (DNN)—are applied to the fused features. The model is trained to recognize patterns that indicate whether a file is malicious or benign.
4. Deployment and Decision Layer: The final layer includes the Deploy Trained Model and the Malicious Files Decision Module. The trained model is deployed to analyze new input

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 06 June 2025, pp: 01-09 www.ijaem.net ISSN: 2395-5252

files in real time. When a file is submitted, the system evaluates it and makes a decision: if the file is malicious, it is flagged; if benign, it is allowed to pass.

B.System Architecture

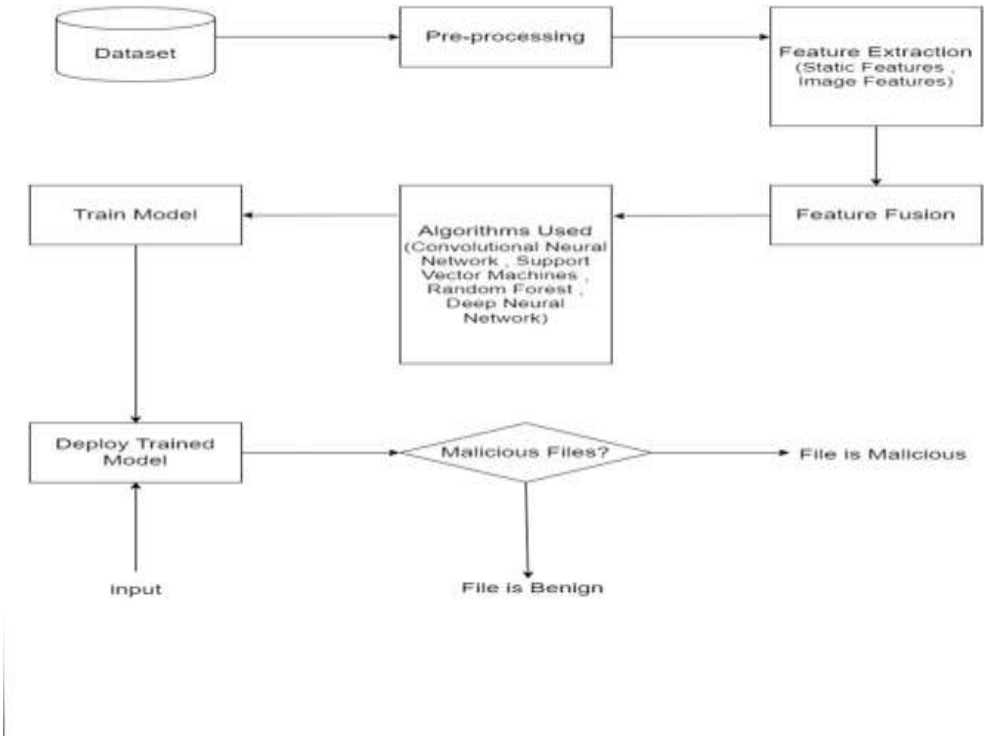

Fig. 1. System Architecture

C. Modules of the Project in Detail

1. User Registration Module: The user registration module allows new users to create an account by providing personal details such as full name, address, email, phone number, gender, age, and password. It implements input validation to ensure data integrity, including checks for valid email formats, password strength, and unique usernames. Upon successful validation, the user information is securely stored in the evaluation.db SQLite database.
2. User Login Module: The user login module enables existing users to log in using their credentials. It verifies user credentials against the stored data in the database. Upon successful login, users are granted access to the main application interface using the subprocess module. This module also provides an option for new users to navigate to the registration module, ensuring a seamless user experience.
3. Graphical User Interface (GUI): The GUI utilizes Tkinter to create user-friendly interfaces for both registration and login processes. It incorporates background images and styled widgets to enhance user experience. The interface adjusts to different screen resolutions for optimal display, ensuring accessibility across various devices.
4. Main Application Interface :The main application interface serves as the central hub of the application, providing access to the core features post-login. While the specific functionalities of this module are not detailed in the provided snippets, it likely includes features related to malware detection, analysis, and reporting, serving as the primary interface for users to interact with the system.
5. Database Management: The database management module employs SQLite for lightweight and efficient data storage. It handles the creation of tables, insertion of new records, and retrieval of existing data. To ensure security, sensitive information, such as passwords, is stored securely, adhering to best practices for data protection.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 06 June 2025, pp: 01-09 www.ijaem.net ISSN: 2395-5252

V. RESULT

Fig. 2. User Page

Fig. 3. Registration Page

Fig. 4. Login Page

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 06 June 2025, pp: 01-09 www.ijaem.net ISSN: 2395-5252

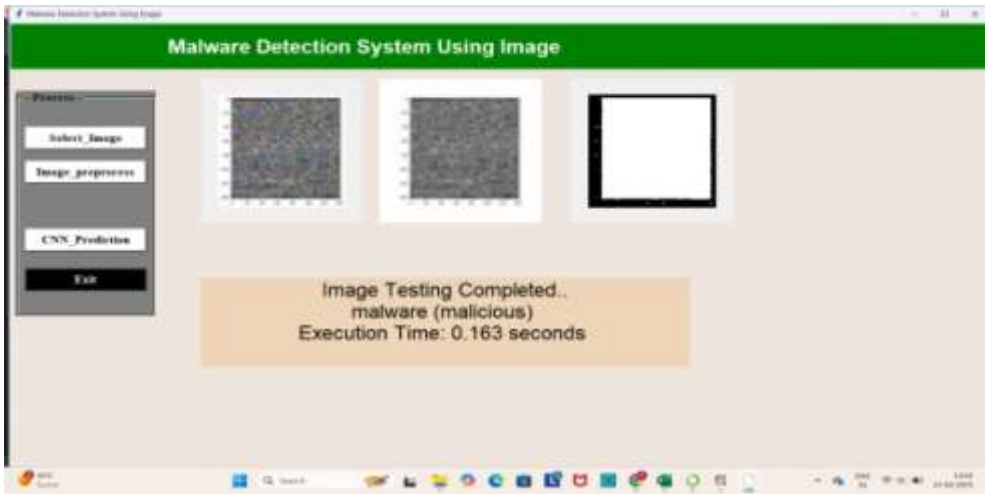

Fig. 5. Malicious Image

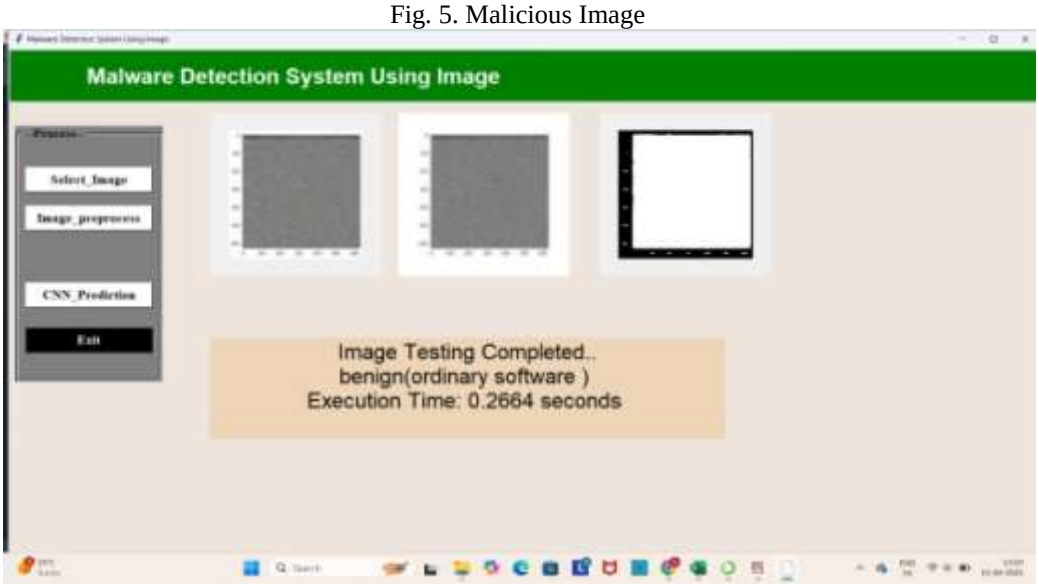

Fig. 6. CNN Bening

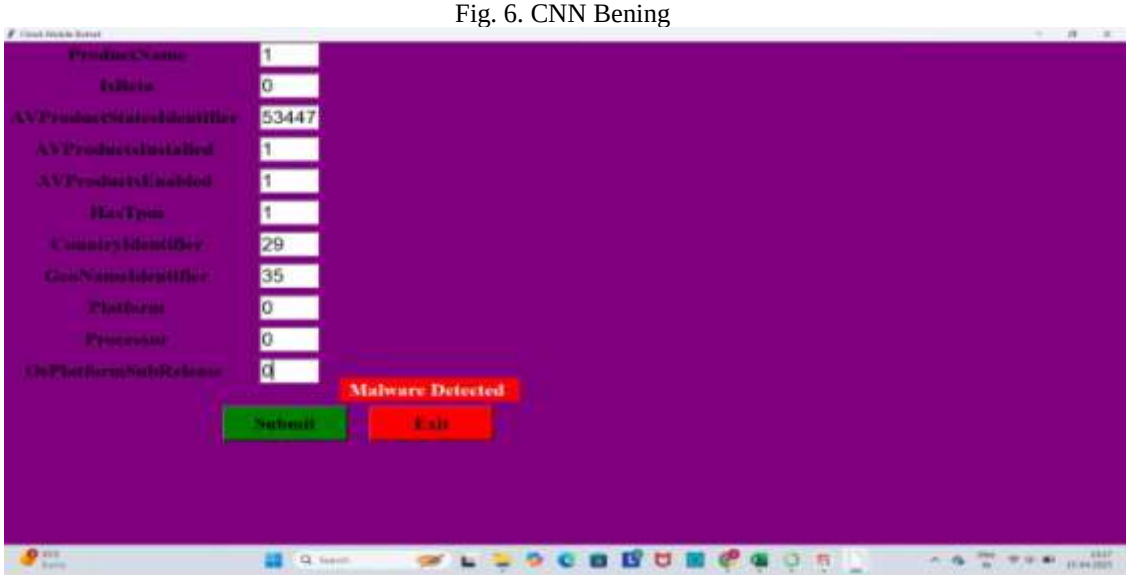

Fig. 7. malware detected

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 06 June 2025, pp: 01-09 www.ijaem.net ISSN: 2395-5252

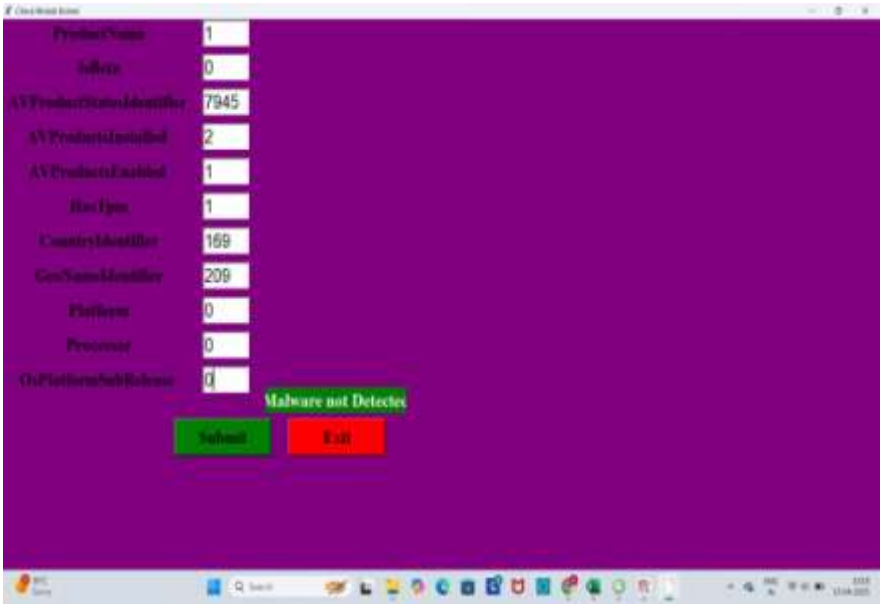

Fig. 8. Malware not detected

Performance Metrics

The CNN and the Support Vector Machine (SVM) were the two main classifiers used to test the system. Performance metrics seen during testing are compiled in the following table.

Model	Accuracy	Precision	Recall	F1-Score	Time (ms)
SVM	96	96	95	95	3-10
CNN	95	94	95	94	1-10

As seen in the table, CNN is powerful for malware images and learns from visual clues.Svm is fast and accurate for structured text features.Use CNN for complex,hidden malware and SVM for quick feature based detection.

Comparison of Model Accuracies

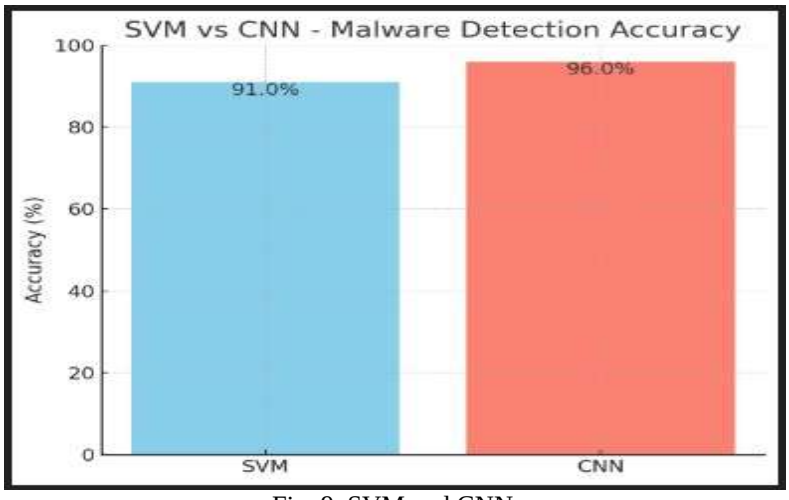

Fig. 9. SVM and CNN

VI. FUTURE WORK

While the proposed multi-view feature fusion approach using CNN and SVM has shown promising results in improving malware classification accuracy, there remains ample scope for enhancement. Future work can focus on expanding the dataset to include more diverse and sophisticated malware samples, particularly those from emerging malware families, to further improve the model's generalizability. Additionally, incorporating advanced deep learning architectures such as hybrid neural networks or transformer-based models could potentially enhance the system's ability to detect complex and evolving threats. Real-time detection capabilities can also be explored to enable immediate response to malware activity in dynamic environments. Furthermore, integrating explainable AI techniques would make the system's decisionmaking more transparent, aiding cybersecurity professionals in understanding detection outcomes. Lastly, optimizing the model for deployment in resource-constrained environments, such as IoT devices or edge computing platforms, would increase its practical applicability in real-world cybersecurity scenarios.

VII. CONCLUSION

In conclusion, this research advances the field of malware detection by demonstrating that multi-view deep learning—rooted in comprehensive feature fusion—offers a scalable, accurate, and adaptable framework for combating sophisticated cyber threats. By integrating the strengths of CNNs and SVMs while addressing their respective limitations, the study provides a blueprint for next-generation detection systems capable of meeting the dynamic challenges of cybersecurity. Future efforts should focus on refining real-time deployment, expanding threat coverage, and ensuring ethical alignment with privacy and transparency standards.

The methodology highlights the importance of data preprocessing (e.g., balancing datasets, standard scaling) and model optimization to mitigate overfitting and enhance generalizability. Results reveal that feature fusion not only compensates for the limitations of individual analysis techniques (e.g., static analysis failing against packed malware) but also achieves a balanced precision-recall trade-off, reducing false positives and negatives—a critical requirement for real-world cybersecurity applications. Furthermore, the model's strong performance on unseen data validates its adaptability to emerging threats, addressing a key gap in traditional signature-based detection systems.

In summary, this study presents a comprehensive and innovative approach to malware detection by integrating multiview feature fusion with advanced deep learning techniques. By combining static, dynamic, and image-based features extracted from Portable Executable (PE) files, the proposed methodology captures a broader and more detailed representation of malware behavior and structure. The Convolutional Neural Network (CNN), particularly when paired with the "Mergeall" fusion strategy, demonstrates superior performance in detecting and classifying malware with high accuracy, minimizing false positives and negatives. The model's robustness is further highlighted by its generalizability across unseen datasets, indicating real-world applicability in dynamic cybersecurity environments. The incorporation of Support Vector Machine (SVM) for textual analysis adds an extra layer of precision, making the system versatile and adaptive. This dual-model framework not only elevates detection rates but also addresses the limitations of single-perspective models by leveraging the unique strengths of each technique. The results validate the potential of hybrid, deep-learning-driven solutions to revolutionize malware classification, paving the way for more intelligent, responsive, and resource-efficient cybersecurity systems. Future research can build upon this foundation by exploring more sophisticated fusion strategies and emerging neural architectures to further enhance malware detection capabilities in an increasingly complex threat landscape.

REFERENCES

[1] Manoj D. Shelar, S. Srinivasa Rao ,"Malware detection of executable file", International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN: 2278-3075, Volume-9 Issue-3, January 2020

[2] Gavrilut D., Cimpoesu M., Anton D., Ciortuz L., "Malware PredictionUsing Machine Learning", International Multiconference on Computer Science and Information Technology, 2009.

[3] Baset, M. "Machine Learning For Malware Detection", 2016.

[4] Rhode, M., Burnap, P., and Jones, K. (2018). Early-stage malware prediction using recurrent neural networks. Computers and Security, 77 , 578-594.

[5] Yeo, M., Koo, Y., Yoon, Y., Hwang, T., Ryu, J., Song, J., Park, C., "Flow-based malware detection using convolutional neural network", 2018 International Conference on Information Networking, 2018.

[6] Rhode, M., Burnap, P., Jones, K., "Early-stage malware prediction using recurrent neural networks", computers security, 2018.
[7] AviadCohen,NirNissim,YuvalElovici(2019). " MalJPEG: Mchine Learning Based Solution For the Detection of Malitious JPEG Images."
[8] S. Rezaei and X. Liu, "Deep learning for encrypted traffic classification:An overview," IEEE Commun. Mag., vol. 57, no. 5, pp. 76 – 81, May 2019
[9] Venkatraman, S., Alazab, M., Vinayakumar, R. (2019). A hybrid deep learning image-based analysis for effective malware detection. Journal of Information Security and Applications, 47, 377-389.
[10] Nisa, M., Shah, J. H., Kanwal, S., Raza, M., Khan, M. A., Damaˇseviˇcius, R., and Blaˇzauskas, T. (2020). Hybrid malware classification method using segmentationbased fractal texture analysis and deep convolution neural network features. Applied Sciences, 10(14) , 4966.
[11] Kyadige, A., Rudd, E. M., and Berlin, K. (2020, May). Learning from Context: A Multi-View Deep Learning Architecture for Malware Detection. In 2020 IEEE Security and Privacy Workshops (SPW) (pp. 1 7). IEEE.
[12] Aslan, O. A., Samet, R. (2020). A comprehensive¨ review on malware detection approaches. IEEE Access, 8, 6249-6271.
[13] Abusitta, A., Li, M. Q., Fung, B. C. (2021). Malware classification and composition analysis: A survey of recent developments. Journal of Information Security and Applications, 59, 102828.
[14] Manoj D. Shelar, S. Srinivasa Rao ," "Deep malware hunter based unrivaled malware detection schema thru cache retrospective empiricism", Concurrency and Computation: Practice and Experience, WILEY, Volume34, Issue19 ,30 August 2022
[15] RajeshkarChaganti, VinayaKumar Ravi, A. F., Tuan D. Pham,December(2022). Deep Learning based model of Multi View Feature Fusion for Effective Malware Classification.
[16] Chaganti, R., Ravi, V., Pham, T. D. (2022). Deep 1095 Learning based Cross Architecture Internet of Things malware Detection and Classification.
[17] Manoj D. Shelar S. Srinivasa Rao," "Enhanced capsule network-based executable files malware detection and classification—deep learning approach", Concurrency and Computation: Practice and Experience, WILEY, Volume34, Issue19 ,30 August 2022, Volume36, Issue4 ,15 February 2024