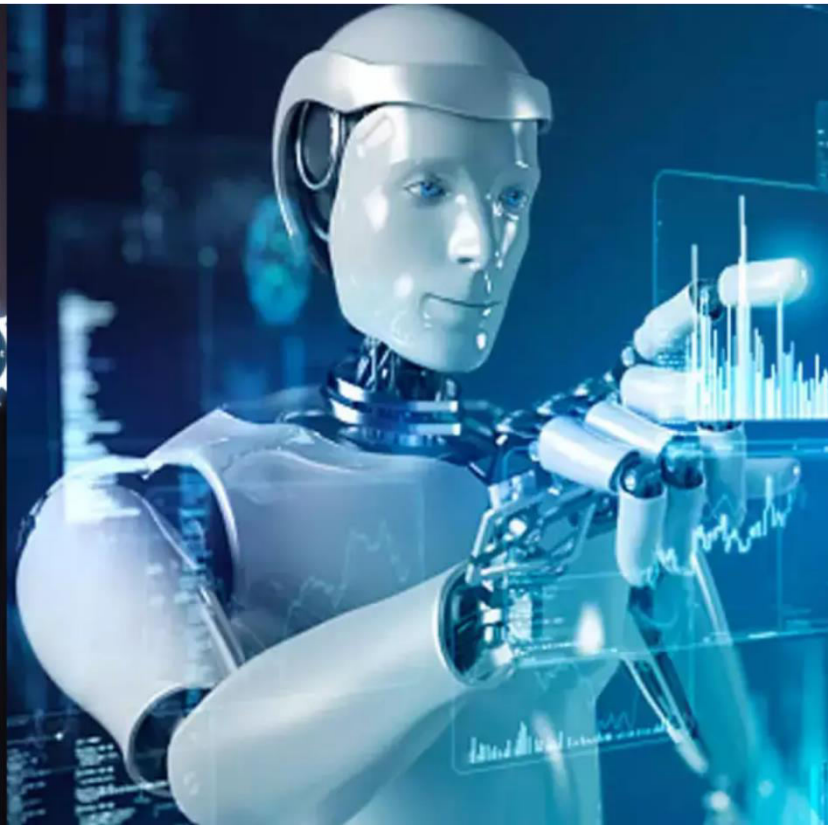
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404288|

Intelligent Pattern Recognition using Equilibrium Optimizer with Deep Learning Model for Android Malware Detection

Y. V. Abhilash[1], V. Eshwar[2], Y. Surya Mohan Reddy[3], Y. V. Abhiram[4], V. Krithika

B. Tech Students, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, India[1-4]

Assistant Professor, Department of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, India[5]

ABSTRACT: Android dominates the global smartphone industry with a market share exceeding 70%, making it a prime target for malware applications. Static malware detection techniques often fail against code obfuscation, while manipulating runtime syscall sequences remains challenging for attackers. Current syscall-based detection systems rely on numerical features like syscall frequencies and transition probability matrices, which require large datasets and are vulnerable to noise and outliers. This study proposes a novel approach that improves efficiency and accuracy using binary representations of syscall subsequences identified through the information gain method. Utilizing the CICMalDroid2020 dataset, this research trains multiple machine learning models—XGBoost, LightGBM, Random Forest—and an advanced Bi-LSTM-based deep learning model to classify 15 malware families. Results are compared to identify the best-performing model, which can be deployed for real-time malware detection.

I. INTRODUCTION

In the present scenario, Cybersecurity has become a primary area for computer scientists and network engineers that provide satisfactory solutions to numerous problems. As the outcome of the fast developments in technological growth and their essential integrations in every form of our lives, a miscellany of malware features and their expected goals have become recognized as well as studied. Among the malware diversity, the Android malware has received. attention lately which is found to occupy considerable attention globally. Android is a common operating system (OS) when compared to other that controls the OS market.

Malware attacks are familiar situations that can be recognized as an attack on Android . Numerous descriptions for malware were proposed by various researchers that depend on the damages caused. The decisive significance of the malware lies in malicious apps using different forms of encryption with the aim of gaining unauthorized access. It is aimed to perform actions that are neither authorized nor appropriate during a security breach. The three core regulations in security are namely: privacy, availability, and integrity. Malware associated with smart systems can be recognized in three contexts as threats such as objectives & attitude, sharing and infection routes, and privilege acquisition modes. Scams, junk e-mails, information thievery, and abuse of websites can be stated as hazardous aims and behavioral outlooks. Software, browsers, markets, systems, and networks can be recognized as the sharing and infected directions. Procedural exploitation and employer management like social engineering are enumerated in the privilege and acquisition means.

Android OS has been recognized as Android malware that damages or takes information in an Android-based mobile device. These are considered ransomware, adware, botnet, spyware, worms, back doors, and Trojans. Google defines malware as possibly malicious fea tures. They categorized malware as commercial and non-commercial privilege escalation, spyware, phishing, and kinds of frauds like Trojans, back doors toll fraud,and SMS fraud. To reduce the threat via Android malware, several researchers approached and developed it so far. The malware identification methods are divided into two types namely dynamic and static analysis-based classification. Dynamic analysis has the benefit that it is likely to control malicious features that employ complicated methods like code packing or encryption. Both machine learning (ML) and DL methods have shown effectiveness in the malware detection process, especially in the context of Android feature analysis and broader areas of cyber security.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404288|

DL methods are found to be highly efficient over traditional ML methods for Android malware recognition. Static analysis, based on algorithms, applies linguistic features that can be eliminated with no implementation of a feature, while dynamic analysis-based methods utilize semantic features observed when an application is executed in specific environments.

ML and non-ML approaches grapple with problems like restricted efficiency in developing malware patterns and a superior incidence of false positives. Afterward, DL approaches were determined useful, highlighting their ability to automatically extract intricate features and recognize difficult patterns, preparing them more capable of tackling the dynamic nature of Android malware. By concentrating on this feature, the research endeavors to improve the performance of Android malware detection, assisting in more

II. LITERATURE REVIEW

The authors in proposed an intelligent hyperparameter altered DL-based malware detection (IHPT-DLMD) approach. Also, the Bi-directional Long Short-Term Memory (BiLSTM) methodology has been deployed for the classification and recognition of Android malware. Finally, the glowworm swarm optimization (GSO) approach has been presented to enhance the hyperparameters of the BiLSTM method. The proposed model is assessed by a standard dataset. The authors in presented an ML-based system for Android malware recognition. Also, the devised technique makes use of the Information Gain model.

They make use of various integrations of API Calls, contextual features, and authorizations. These integrations were fed into different ML models. The authors in presented a new technique for recognizing malware in Android apps by GRU, which is a type of RNN technique. In an ML-based recognition technique that makes use of hybrid analysis-based Particle swarm optimization (PSO) and an adaptive genetic algorithm (AGA) is proposed. Primarily, feature selection (FS) was done by employing PSO to the features. Then, the demonstration of XGBoost and random forest (RF) using ML identifiers is enhanced by the AGA.

The author in projected a structure employing image-based malware depictions of Android Dalvik Executable (DEX) files. The structure utilizes the EffectiveNetB0 convolutional neural network (CNN) for feature abstraction, which was then delivered over a worldwide average pooling layer and provided in a stacking identifier. The author in projects a Rock Hyrax Swarm Optimizer with a DL-based Android malware detection (RHSODL-AMD) approach. Moreover, the Adamax enhancer with attention recurrent autoencoder (ARAE) technique has been used for Android malware recognition. In a pre-trained CNN technique in Android malware recognition is projected. An EfficientNet-B4 CNN-based method has been developed to recognize Android malware.

These features are transferred by a worldwide average pooling layer and provided as a softmax identifier. The authors presented a novel DL-based technique. Primary, it visualization a portable executable (PE) file as a colored image. Then, it identifies malware that relies on deep features employing support vector machine (SVM). The projected model integrates DL with ML methods and is observed by using benchmark datasets.

III. METHODOLOGY

Existing System

Existing systems for Android malware detection predominantly rely on static analysis, which involves examining the application's source code or binary structure to detect malicious patterns. These methods are effective against known malware but fail when the malware employs obfuscation techniques, such as encryption or polymorphism, to hide malicious intent. While dynamic analysis techniques like syscall-based detection show promise, most approaches depend on numerical features such as syscall frequencies or transition probability matrices. Model LSTM achieved accuracies between 98% and 99% for **binary classification**. However, these systems face significant challenges, including:

1. Limited ability to handle code obfuscation in static analysis.
2. High dependency on large datasets for training machine learning models.
3. Susceptibility to noise and outliers in syscall data.
4. Lack of scalability to multiclass malware classification.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404288|

Proposed system

The proposed system focuses on utilizing the CICMalDroid2020 dataset to classify Android malware into 15 families. Informative syscall subsequences are identified using the information gain method, and binary feature representation is applied to improve model efficiency. Multiple machine learning models, including XGBoost, LightGBM, and Random Forest, as well as an advanced Bi-LSTM deep learning model, are trained on this data. The models are compared based on their accuracy, precision, recall, and F1-score, with the goal of identifying the optimal solution. This approach addresses challenges of dataset imbalance, noise, and obfuscation while enhancing scalability for future use cases.

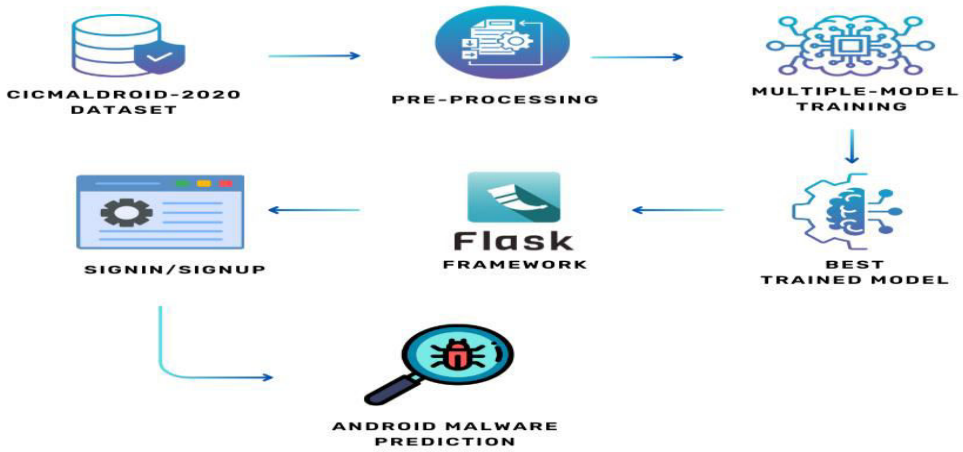


Fig 1: System Architecture

VI. IMPLEMENTATION PROCESS

1. **Dataset Collection and Pre-processing:**
   - Gather syscall data from the CICMalDroid2020 dataset.
   - Pre-process the data by cleaning, normalizing, and extracting binary feature representations from syscall subsequences.
   - Use the information gain method to identify the most informative features.

2. **Model Training and Evaluation:**
   - Train multiple machine learning models (XGBoost, LightGBM, Random Forest) and the Bi-LSTM deep learning model.
   - Evaluate the models using metrics like accuracy, precision, recall, and F1-score to identify the best-performing classifier.

3. **Web Interface:**
   - A web application is developed to provide an interactive platform for malware detection. The application features a homepage, a user authentication system with signup and signin functionality, and a prediction page where users can upload files for real-time malware detection. The backend integrates the trained models, enabling accurate and efficient predictions. The application is designed to be user-friendly and scalable, making it accessible to a wide range of users.

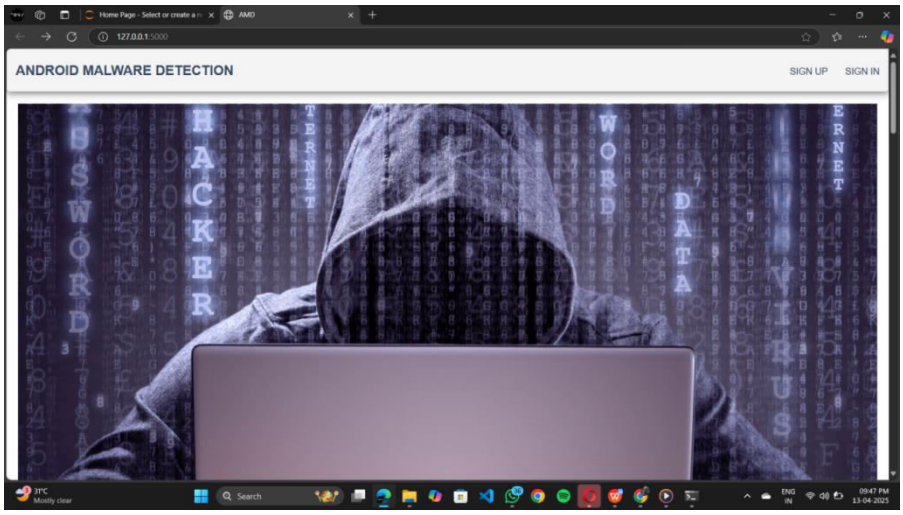
International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

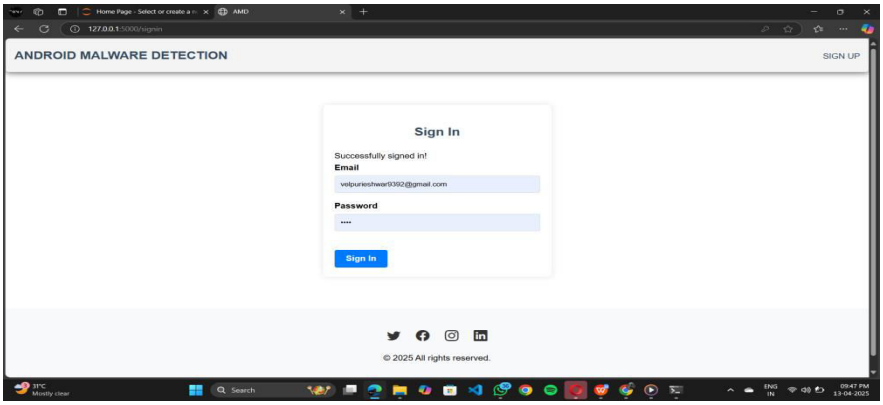
|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

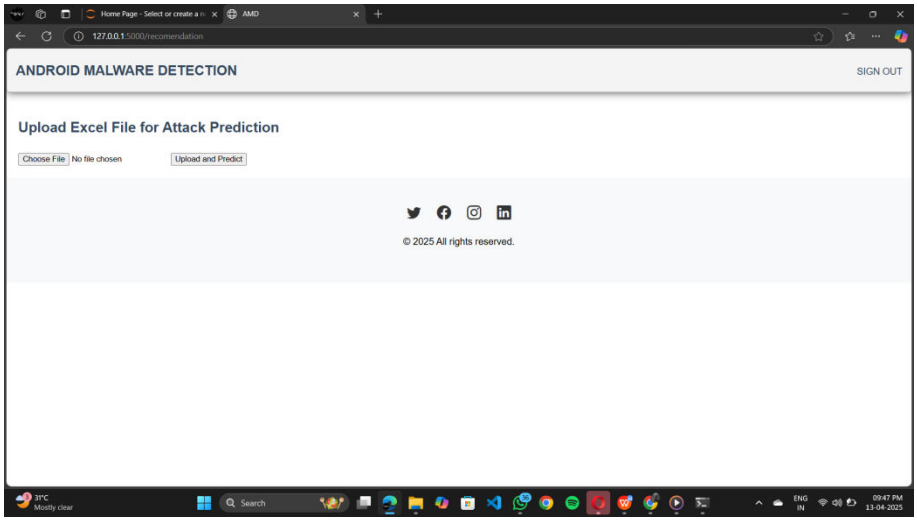
|DOI: 10.15680/IJIRSET.2025.1404288|



Homepage



User Register



Upload Excel File for Attack Prediction

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404288|

V. RESULT AND CONCLUSION

By employing advanced machine learning and deep learning models, the system achieves high accuracy in classifying malware into 15 distinct families. Analysing best model capturing the dynamic behavior of syscall data, outperforming traditional methods in handling sequential data. The results highlight the system's potential for real-world deployment, offering a scalable and robust solution for malware detection. Future enhancements will ensure that the system adapts to evolving threats, integrates advanced neural network architectures, and accommodates real-time detection in large-scale scenarios

REFERENCES

1. Gómez and A. Muñoz, ‘‘Deep learning-based attack detection and classification in Android devices,’’Electronics, vol. 12, no. 15, p. 3253, Jul. 2023.
2. Y. Zhao, L. Li, H. Wang, H. Cai, T. F. Bissyandé, J. Klein, and J. Grundy, ‘‘On the impact of sample duplication in machine-learning-based Android malware detection,’’ ACM Trans. Softw. Eng. Methodol., vol. 30, no. 3, pp. 1–38, Jul. 2021.
3. H. Wang, W. Zhang, and H. He, ‘‘You are what the permissions told me! Android malware detection based on hybrid tactics,’’ J. Inf. Secur. Appl., vol. 66, May 2022, Art. no. 103159.
4. H. Rathore, A. Nandanwar, S. K. Sahay, and M. Sewak, ‘‘Adversarial superiority in Android malware detection: Lessons from reinforcement learning based evasion attacks and defenses,’’ Forensic Sci. Int., Digit. Invest., vol. 44, Mar. 2023, Art. no. 301511.
5. V. Sihag, M. Vardhan, P. Singh, G. Choudhary, and S. Son, ‘‘De-LADY: Deep learning-based Android malware detection using Dynamic features,’’ J. Internet Serv. Inf. Secur., vol. 11, no. 2, pp. 34–45, 2021.
6. M. Ibrahim, B. Issa, and M. B. Jasser, ‘‘A method for automatic Android malware detection based on static analysis and deep learning,’’ IEEE Access, vol. 10, pp. 117334–117352, 2022.
7. Albakri, F. Alhayan, N. Alturki, S. Ahamed, and S. Shamsudheen, ‘‘Metaheuristics with deep learning model for cybersecurity and Android malware detection and classification,’’ Appl. Sci., vol. 13, no. 4, p. 2172, Feb. 2023.
8. Batouche and H. Jahankhani, ‘‘A comprehensive approach to Android malware detection using machine learning,’’ in Information Security Technologies for Controlling Pandemics, 2021, pp. 171–212.
9. P. Bhat and K. Dutta, ‘‘A multi-tiered feature selection model for Android malware detection based on feature discrimination and information gain,’’ J. King Saud Univ.-Comput. Inf. Sci., vol. 34, no. 10, pp. 9464–9477, Nov. 2022.
10. L. Hammood, I. A. Dogru, and K. Kiliç, ‘‘Machine learning-based adaptive genetic algorithm for Android malware detection in auto-driving vehicles,’’ Appl. Sci., vol. 13, no. 9, p. 5403, Apr. 2023.
11. R. Raphael and P. Mathiyalagan, ‘‘Intelligent hyperparameter-tuned deep learning-based Android malware detection and classification model,’’ J. Circuits, Syst. Comput., vol. 32, no. 11, Jul. 2023, Art. no. 2350191.
12. M. N. AlJarrah, Q. M. Yaseen, and A. M. Mustafa, ‘‘A contextaware Android malware detection approach using machine learning,’’ Information, vol. 13, no. 12, p. 563, Nov. 2022.
13. O. N. Elayan and A. M. Mustafa, ‘‘Android malware detection using deep learning,’’ Proc. Comput. Sci., vol. 184, pp. 847–852, Jan. 2021.
14. P. Yadav, N. Menon, V. Ravi, S. Vishvanathan, and T. D. Pham, ‘‘A two-stage deep learning framework for image-based Android malware detection and variant classification,’’ Comput. Intell., vol. 38, no. 5, pp. 1748–1771, Oct. 2022.
15. P. Yadav, N. Menon, V. Ravi, S. Vishvanathan, and T. D. Pham, ‘‘EfficientNet convolutional neural networks-based Android malware detection,’’ Comput. Secur., vol. 115, Apr. 2022, Art. no. 102622.
16. K. Shaukat, S. Luo, and V. Varadharajan, ‘‘A novel deep learning-based approach for malware detection,’’ Eng. Appl. Artif. Intell., vol. 122, Jun. 2023, Art. no. 106030.
17. Desnos and G. Gueguen, ‘‘Androguard documentation,’’ Tech. Rep., 2018.
18. Chen, J. Wei, and Z. Li, ‘‘Remaining useful life prediction for lithiumion batteries based on a hybrid deep learning model,’’ Processes, vol. 11, no. 8, p. 2333, Aug. 2023.
19. Zhong, G. Li, Z. Meng, H. Li, and W. He, ‘‘A self-adaptive quantum equilibrium optimizer with artificial bee colony for feature selection,’’ Comput. Biol. Med., vol. 153, Feb. 2023, Art. no. 106520. [20] Andro-AutoPsy. Accessed: Jul. 14, 2023. [Online]. Available: http://ocslab.hksecurity.net/andro-autopsy

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404288|

20. J.-W. Jang, H. Kang, J. Woo, A. Mohaisen, and H. K. Kim, "AndroAutoPsy: Anti-malware system based on similarity matching of malware and malware creator-centric information," Digit. Invest., vol. 14,pp. 17–35, Sep. 2015.
21. H. Alamro, W. Mtouaa, S. Aljameel, A. S. Salama, M. A. Hamza, and A. Y. Othman, "Automated Android malware detection using optimal ensemble learning approach for cybersecurity," IEEE Access, vol. 11, pp. 72509–72517, 2023.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details