WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Comparative Analysis of Deep Learning-Based Approaches for Detecting Malwares in Android

Malavika N[1], Prof. Dilna P M[2]

[1]Department of Computer Science and Engineering Government Engineering College, Wayanad Kerala, India malavikan878@gmail.com
[2]Department of Computer Science and Engineering Government Engineering College, Wayanad Kerala, India dilnapm@gecwyd.ac.in
DOI : https://doi.org/10.55248/gengpi.6.0325.1111

ABSTRACT—

The use of smartphones has grown rapidly in recent years. This has led to an increase in their popularity as targets for attackers. The field of malware detection is a never-ending competition between attackers and anti-malware developers. According to the latest report published by the computer security company a new malware can be found in online repositories every 8s. So it is important to protect various devices and networks from the malware attack. The increasing sophistication of mobile threats and the adaptability of ML make it a promising solution for Android malware detection. It requires ongoing research and a robust dataset for model training, making it a challenging but impactful field with significant potential for innovation and security enhancement. The wide range of capabilities offered by smartphones and the rising number of activities carried out by their users, including social networking, online banking, and gaming, has given rise to very serious concerns about device security and personal privacy. Since Android is an open-source platform, it is easy for malware developer to launch their attacks and develop Android malware apps that cause severe harm. The project aims to detect malware using a hybrid model by using various deep learning algorithms so thereby performance can be increased. It also uses attention mechansim, sentimental analysis based on the reviews of the customer. The dataset of the project is taken from the Kaggle. Data Sources of Android apps are collected from app stores and third-party sources. Malware samples are gathered from malware repositories such as VirusShare, AndroZoo, DREBIN Dataset, AMD(Android Malware Dataset)..

Index Terms - Deep Learning, Sentiment Analysis, attention mechanism

I. Introduction

As Android is the most widely used mobile operating system globally, it has become a prime target for malicious attackers. Although various detection techniques exist for identifying Android malware, traditional methods such as signature-based and heuristic-based approaches are insufficient to address the challenges posed by modern malware. Research has shown that machine learning classifiers can effectively analyze app permissions to distinguish between malicious and benign Android applications. Several machine learning techniques leverage permission-based features to create models for malware detection on Android devices. However, the effectiveness of these methods heavily depends on the quality of the raw or feature datasets used. A significant challenge in Android malware research remains the limited avail ability of comprehensive and up-to-date raw malware datasets Android's widespread use has made it a prime target for malware attacks, creating significant challenges for cybersecurity. Traditional detection methods, such as signature-based and heuristic approaches, often fail to address the complexities of modern malware. As a result, machine learning techniques have emerged as an adaptive and scalable solution for detecting Android malware. Current models typically rely on analyzing features such as permissions, API calls, system behaviors, and network activities extracted from Android applica tions. Static analysis models examine an app's code and metadata without execution, using machine learning algorithms like Random Forest and Support Vector Machines to classify applications. However, these models are vulnerable to obfuscation techniques used by advanced malware. Dynamic analysis models, on the other hand, observe runtime behavior in con trolled environments, utilizing methods like Recurrent Neural Networks and Long Short-Term Memorynetworks to analyze sequential execution data, though this approach can be computationally intensive and slower. Hybrid models combine static and dynamic features to harness the strengths of both, often employing ensemble learning or advanced deep learning architectures such as Convolutional Neural Networks integrated with RNNs for enhanced accuracy and robustness.Despite their advantages, machine learning-based detection systems face challenges, including the labor-intensive process of feature engineering, the adaptability of modern malware, reliance on outdated or imbalanced datasets, and the risk of false positives and negatives. While these models show promise, ongoing advancements in feature extraction, dataset quality, and hybrid methodologies are essential to address these limitations and keep up with the evolving malware landscape. The introduction provides an overview of the key challenges in Android malware attack, highlights the limitations of existing solutions, and explains the motivation behind adopting a hybrid approach. By combining the unique strengths of these two deep learning techniques, the proposed system offers a promising solution for mitigating the impact of malware attacks, contributing to enhanced security and uninterrupted service availability

II. LITERATURE REVIEW

[1] This paper presents an in-depth review of Android malware detection systems that leverage machine learning techniques to address the growing threat of malicious applications. It explores various ML approaches, including supervised, unsupervised, and ensemble methods, focusing on their use in feature extraction, classification, and anomaly detection. The review covers feature analysis techniques, such as static, dynamic, and hybrid methods, which serve as critical inputs for distinguishing between benign and malicious apps. Additionally, the paper discusses evaluation metrics like accuracy, precision, recall, and AUC-ROC, which are used to measure the performance of these models. It highlights key challenges, such as the evolving nature of malware, limitations in available datasets, and the computational demands of real- time detection. The paper also outlines future directions, emphasizing the need for improved feature engineering, real-time capabilities, and advanced ML techniques like transfer learning and federated learning. This review aims to provide valuable insights into ML-based Android malware detection systems and their potential for enhancing mobile security.

[2] This paper reviews various Android malware detection approaches that utilize machine learning techniques to combat the increasing threat of malicious applications. It provides an overview of ML-based methods used for analyzing and classifying malware, highlighting static, dynamic, and hybrid analysis techniques as the primary methods for extracting features from Android applications. The paper discusses the application of supervised, unsupervised, and ensemble learning algorithms in malware detection and their effectiveness in identifying malicious behavior. It also evaluates common datasets, performance metrics such as accuracy, precision, and recall, and the challenges associated with evolving malware, adversarial attacks,and dataset limitations. The paper emphasizes the need for robust, adaptive, and scalable detection systems, offering insights into future directions like integrating hybrid approaches, using advanced ML techniques, and addressing real-time detection requirements to improve Android security

[3] This paper focuses on identifying significant Android permissions that play a critical role in machine-learning-based malware detection. Android permissions, which regulate app access to sensitive resources, are a key feature used to distinguish between benign and malicious applications. The study emphasizes how the selection of relevant and significant permissions can improve the performance and accuracy of malware detection models by reducing noise and computational complexity. It explores various feature selection techniques to identify the most impactful permissions and integrates these into machine learning frameworks for effective classification. Additionally, the paper evaluates the impact of these permissions on detection accuracy and discusses challenges such as permission misuse and evasion tactics by malware. The findings highlight the importance of permission analysis in building efficient and reliable ML-based Android malware detection systems, offering insights for researchers to enhance detection capabilities.

[4] This paper proposes a novel machine learning approach for Android malware detection that leverages the co-existence of features to improve accuracy and robustness. Unlike traditional methods that analyze individual features in isolation, this approach focuses on identifying and exploiting the relationships and patterns between multiple features, such as permissions, API calls, and behavioral attributes, to detect malicious applications more effectively. By modeling these co-existing features, the method enhances the representation of app behavior, enabling better differentiation between benign and malicious apps. The paper demonstrates the approach's effectiveness through experimental evaluations, showing improvements in detection performance and resilience against evasion techniques. This innovative framework offers a promising direction for enhancing the accuracy and reliability of ML-based Android malware detection systems.

[5] This paper introduces a static analysis framework designed for generating permission-based datasets and detecting Android malware using machine learning techniques. The framework focuses on extracting permissions from Android application packages (APKs) to create a structured dataset for training ML models. By relying on static analysis, the framework avoids the runtime overhead of dynamic analysis while capturing key features that indicate malicious behavior. The study evaluates various ML algorithms on the generated dataset to classify apps as benign or malicious, highlighting the role of permission features in improving detection accuracy. Additionally, the paper discusses the advantages of static analysis, such as scalability and speed, while addressing challenges like obfuscated code and evolving malware tactics. This framework demonstrates the potential of permission-based static analysis combined with ML to enhance Android malware detection efficiency.

[6] This paper presents an intelligent pattern recognition framework for Android malware detection that combines the Equilibrium Optimizer (EO) algorithm with a deep learning model. The approach leverages the EO, a nature-inspired optimization technique, to select optimal features from Android applications, improving the efficiency and accuracy of the detection process. By focusing on relevant features, the proposed method reduces computational complexity while maintaining high performance. The selected features are then fed into a deep learning model,which learns complex patterns and behaviors to effectively classify applications as benign or malicious. Experimental results demonstrate the superiority of this integrated approach in terms of accuracy, precision, and robustness compared to traditional methods. The framework highlights the potential of combining advanced optimization algorithms with deep learning for intelligent and scalable Android malware detection.

[7] This paper introduces EvadeDroid, a practical evasion attack targeting machine learning-based black-box Android malware detection systems. The study highlights vulnerabilities in ML based detection models, demonstrating how attackers can manipulate malicious applications to bypass detection while retaining their harmful functionality. EvadeDroid employs techniques such as injecting benign features or obfuscating malicious ones to evade detection in a black-box setting, where the internal workings of the detection system are unknown. The paper evaluates the attack's effectiveness across multiple detection frameworks and demonstrates significant reductions in detection accuracy. Additionally, it discusses the

implications of such evasion attacks on the reliability of ML-based systems and emphasizes the need for more ro bust and resilient malware detection mechanisms. This work provides insights into the security challenges faced by Android malware detection models and advocates for the development of defenses against adversarial attacks.

[8] The paper presents PermDroid, a framework developed for Android malware detection using a novel feature selection approach combined with machine learning techniques. The framework focuses on improving detection accuracy by selecting the most relevant permissions from Android applications, which are critical indicators of malicious behavior. By utilizing an efficient feature selection method, PermDroid reduces the complexity of the dataset while maintaining high detection performance. The selected features are then fed into machine learning models, such as decision trees or support vector machines, to classify apps as benign or malicious. The paper demonstrates the effectiveness of PermDroid in identifying Android malware with reduced false positives and improved computational efficiency. It highlights the importance of permission-based analysis in malware detection and offers a promising solution for scalable,real-time Android security.

[9] This paper presents an Android malware detection system that leverages machine learning techniques to classify Android applications as benign or malicious. The system uses a combination of static and dynamic analysis to extract relevant features from Android applications, such as permissions, API calls, and system behaviors. These features are then processed and used to train various machine learning models, including decision trees, support vector machines, and random forests, to detect malicious behavior with high accuracy. It emphasizes the effectiveness of ML in automating the detection process and improving the scalability of malware detection systems. It also discusses challenges such as dataset imbalance, evolving malware tactics, and the trade-off between detection accuracy and computational efficiency. The system demonstrates results in terms of detection rate and false positive reduction, making it a viable solution for enhancing Android security.

[10] The review explores various frameworks for Android malware detection and identification that leverage machine learning (ML) and deep learning (DL) techniques, highlighting their effectiveness and challenges. It begins by addressing the growing threat of Android malware due to the open nature of the Android ecosystem and the limitations of traditional detection methods, such as signature-based and heuristic approaches. The paper categorizes frameworks into static analysis, which examines code, permissions, and application metadata, and dynamic analysis, focusing on runtime behaviors like network activity and system calls. ML techniques, including Random Forest, Support Vector Machines (SVM), and clustering methods, are discussed alongside DL models like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks, showcasing their potential in identifying complex patterns. Evaluation metrics such as accuracy, precision, recall, and F1 score are analyzed, with an emphasis on balancing detection rates and minimizing false positives. The review also highlights challenges such as the computational demands of DL models, vulnerabilities to adversarial attacks, and the need for diverse datasets. A comparative analysis of existing frameworks is provided, identifying their strengths, weaknesses, and practical applications. Finally, the paper suggests future directions, including federated learning for privacy-preserving detection, explainable AI for model interpretability, and lightweight DL models optimized for mobile devices.

III. PROPOSED MODEL

The proposed model using deep learning is to create a sophisticated and accurate system capable of identifying ma- licious applications on the Android platform. This involves developing a robust detection model that leverages advanced deep learning techniques to improve classification accuracy and adapt to evolving malware threats. A key focus is on extracting comprehensive features from Android applications, including permissions, API calls, network activity, and be- havioral patterns, to enhance the model's ability to detect malware effectively. By integrating both static code analysis and dynamic behavior monitoring, the system aims to improve its detection capabilities across pre-installation and runtime scenarios. The project also emphasizes scalability and effi- ciency, ensuring the system can handle large datasets and provide real-time detection for immediate threat mitigation. Efforts are directed toward reducing false positives and false negatives to ensure reliability and trustworthiness. Advanced deep learning architectures, such as Convolutional Neural Net- works (CNNs) and Recurrent Neural Networks (RNNs), are utilized to automatically learn and identify complex patterns in malware data. Additionally, explainable AI mechanisms are incorporated to make the system's decisions transparent and understandable for users and analysts. By utilizing diverse and enriched datasets for training and evaluation, the system aims to remain effective against diverse malware strains.

IV. CONCLUSION

This proposed model presents a robust and efficient system for detecting malicious Android applications using a com- bination of static, dynamic, and permission-based analysis techniques, integrated with machine learning for accurate classification. The architecture ensures comprehensive examination of APK files by leveraging static analysis for code and metadata inspection, dynamic analysis for runtime behavior monitoring, and permission analysis for identifying potential over-privileged or risky permissions. The integration of these features into a unified feature vector enables a pre-trained machine learning model to classify applications effectively as either benign or potentially malicious. The system's modu- lar design allows for scalability and adaptability, making it suitable for real-world use cases in the cybersecurity domain. By incorporating threat analysis for flagged applications, the project not only identifies potential malware but also pro- vides deeper insights into their behaviors and risks, aiding in proactive threat mitigation. Overall, this project demonstrates a highly automated, reliable, and intelligent solution for ad- dressing the growing challenges posed by Android malware, contributing significantly to mobile application security.

References

[1] Ali Muzaffar, Hani Ragab Hassen, Michael A Lones, and Hind Zantout. An in-depth review of machine learning based android malware detec- tion. *Computers & Security*, 121:102833, 2022.

[2] Jin Li, Lichao Sun, Qiben Yan, Zhiqiang Li, Witawas Srisa-An, and Heng Ye. Significant permission identification for machine-learning-based android malware detection. *IEEE Transactions on Industrial Informatics*, 14(7):3216–3225, 2018.

[3] Kaijun Liu, Shengwei Xu, Guoai Xu, Miao Zhang, Dawei Sun, and Haifeng Liu. A review of android malware detection approaches based on machine learning. *IEEE access*, 8:124579–124607, 2020.

[4] Esraa Odat and Qussai M Yaseen. A novel machine learning approach for android malware detection based on the co-existence of features. *IEEE Access*, 11:15471–15484, 2023.

[5] Amarjyoti Pathak, Th Shanta Kumar, and Utpal Barman. Static analysis framework for permission-based dataset generation and android malware detection using machine learning. *EURASIP Journal on Information Security*, 2024(1):33, 2024.

[6] Mohammed Maray, Mashael Maashi, Haya Mesfer Alshahrani, Sumayh S Aljameel, Sitelbanat Abdelbagi, and Ahmed S Salama. Intelligent pattern recognition using equilibrium optimizer with deep learning model for android malware detection. *IEEE Access*, 2024.

[7] Hamid Bostani and Veelasha Moonsamy. Evadedroid: A practical evasion attack on machine learning for black-box android malware detection. *Computers & Security*, 139:103676, 2024.

[8] Arvind Mahindru, Himani Arora, Abhinav Kumar, Sachin Kumar Gupta, Shubham Mahajan, Seifedine Kadry, and Jungeun Kim. Permdroid a framework developed using proposed feature selection approach and machine learning techniques for android malware detection. *Scientific Reports*, 14(1):10724, 2024.

[9] Amanpreet Kaur, Sangeeta Lal, Shruti Goel, Mrinal Pandey, and Astha Agarwal. Android malware detection system using machine learning. In *Proceedings of the 2024 Sixteenth International Conference on Contemporary Computing*, pages 186–191, 2024.

[10] Santosh K Smmarwar, Govind P Gupta, and Sanjay Kumar. Android malware detection and identification frameworks by leveraging the machine and deep learning techniques: A comprehensive review. *Telematics and Informatics Reports*, page 100130, 2024.