International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

AUTOMATED ANDROID MALWARE DETECTION USING OPTIMAL ENSEMBLE LEARNING APPROACH FOR CYBER SECURITY

Dr .D.Kalyankumar[1], Buchi Bhagyasri[2]
Pulipati Joshu Swaraj[3], Marri Ravi [4] , Bhavirisetti Naveen[5]
Associate Professor, Department of CSE-Cyber Security, Chalapathi institute of technology, Mothadaka, Guntur, Andhra Pradesh, India-522016.
[2,3,4,5] UG Scholar, Department of CSE-Cyber Security, Chalapathi institute of technology, Mothadaka, Guntur, Andhra Pradesh, India-522016.

ABSTRACT
The rapid expansion of Android devices has increased exposure to evolving cybersecurity threats, particularly malware. Traditional single-model detection techniques often struggle to maintain high accuracy due to the diverse and sophisticated nature of modern attacks. This study proposes an optimal ensemble learning approach for automated Android malware detection, integrating multiple machine learning classifiers to enhance robustness, precision, and adaptability. By combining features extracted from application permissions, API calls, behavioral patterns, and static code attributes, the ensemble model mitigates individual classifier limitations and improves overall detection performance. Experimental results demonstrate that the proposed framework outperforms conventional standalone models in terms of accuracy, recall, and false-positive reduction. This approach offers a scalable and reliable solution for strengthening mobile cybersecurity and protecting users against emerging Android malware threats.

KEYWORDS: Android Malware, Cyber Security, Ensemble Learning, Machine Learning, Automated Detection, Mobile Security, Threat Analysis

1. INTRODUCTION

The rapid growth of Android-based devices has significantly transformed the mobile computing landscape, making it the most widely used mobile operating system worldwide. However, this widespread adoption has also made Android a prime target for cybercriminals who exploit vulnerabilities to deploy malicious applications. Android malware can lead to severe consequences such as data theft, financial loss, unauthorized access, and privacy breaches. Traditional malware detection techniques, including signature-based methods, are no longer sufficient to combat modern sophisticated attacks, especially zero-day threats.

To address these challenges, machine learning-based approaches have gained popularity due to their ability to identify unknown and evolving malware patterns. Among these, ensemble learning techniques have proven to be highly effective by combining multiple classifiers to improve detection accuracy and robustness. An optimal ensemble learning approach leverages the strengths of various algorithms while minimizing their weaknesses, leading to enhanced performance in malware classification.

This research proposes an automated Android malware detection system using an optimal ensemble learning model. The system analyzes application features such as permissions, API calls, and behavioral patterns to classify apps as benign or malicious. By integrating multiple machine learning models like Random Forest, Support Vector Machine, and Gradient Boosting, the proposed approach achieves higher detection accuracy compared to single-model systems.

Furthermore, the system is designed to operate efficiently in real-time environments, making it suitable for deployment in mobile devices or cloud-based security frameworks. The ultimate goal is to provide a reliable, scalable, and intelligent solution to safeguard Android users from emerging cyber threats while maintaining system performance and usability.

2. LITERATURE SURVEY

Numerous studies have explored Android malware detection using both static and dynamic analysis techniques. Static analysis involves examining application code, permissions, and manifest files without executing the application. Researchers have used features like API calls, permissions, and opcode sequences to train machine learning models such as Decision Trees, Naive Bayes, and Support Vector Machines. While static analysis is efficient, it is often vulnerable to code obfuscation techniques used by malware developers.

Dynamic analysis, on the other hand, observes application behavior during runtime, including system calls, network traffic, and user interactions. Although it provides deeper insights into application behavior, it is resource-intensive and time-consuming. Hybrid approaches combining static and dynamic analysis have

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

been proposed to overcome these limitations, but they increase system complexity.

Recent advancements have focused on deep learning models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) for malware detection. These models automatically extract complex patterns but require large datasets and high computational power. Ensemble learning methods have emerged as a promising solution by combining multiple classifiers to enhance detection performance.

Studies have shown that ensemble techniques like Bagging, Boosting, and Stacking outperform individual classifiers in terms of accuracy, precision, and recall. For example, Random Forest and Gradient Boosting models have demonstrated high effectiveness in identifying malware patterns. However, selecting the optimal combination of models remains a challenge.

The proposed system builds upon existing research by introducing an optimal ensemble framework that intelligently selects and combines multiple classifiers. This approach aims to maximize detection accuracy while minimizing false positives and computational overhead, making it suitable for real-world deployment in Android security systems.

3. EXISTING SYSTEM

The existing Android malware detection systems primarily rely on traditional techniques such as signature-based detection and single machine learning classifiers. Signature-based methods identify malware by comparing application code with a database of known malicious signatures. While this approach is fast and efficient, it fails to detect new or unknown malware variants, commonly referred to as zero-day attacks. Additionally, maintaining and updating signature databases is a continuous challenge.

Machine learning-based systems have been introduced to overcome these limitations. These systems typically use a single classifier trained on features extracted from Android applications, such as permissions, API calls, and network behavior. Commonly used classifiers include Decision Trees, Support Vector Machines, and K-Nearest Neighbors. Although these methods improve detection rates compared to traditional approaches, they often suffer from issues such as overfitting, limited generalization, and reduced accuracy when dealing with diverse datasets.

Another limitation of existing systems is their inability to handle imbalanced datasets effectively. Malware datasets often contain significantly fewer malicious samples compared to benign ones, leading to biased models that favor the majority class. Furthermore, many systems rely solely on static or dynamic analysis, which limits their effectiveness in detecting sophisticated malware that uses evasion techniques such as code obfuscation and encryption.

Scalability and real-time performance are also concerns in current systems. Dynamic analysis-based approaches require extensive computational resources and time, making them unsuitable for real-time applications. Additionally, many existing solutions do not provide adaptability to evolving malware patterns.

These limitations highlight the need for a more robust, scalable, and accurate detection system. The proposed ensemble learning-based approach addresses these challenges by combining multiple classifiers to improve detection performance and ensure better generalization across diverse datasets.

4. PROPOSED SYSTEM

The proposed system introduces an automated Android malware detection framework using an optimal ensemble learning approach. The system is designed to improve detection accuracy, reduce false positives, and enhance robustness against evolving malware threats. It integrates multiple machine learning models to create a powerful hybrid classifier capable of identifying both known and unknown malware.

The system begins with data collection, where a dataset of Android applications, including both benign and malicious samples, is gathered. Feature extraction is performed to obtain relevant attributes such as permissions, API calls, intent filters, and system calls. These features are then preprocessed to remove noise and normalize the data for better model performance.

The core component of the system is the ensemble learning model. Multiple classifiers, such as Random Forest, Support Vector Machine, and Gradient Boosting, are trained individually on the dataset. An optimization technique is applied to select the best combination of models based on performance metrics like accuracy, precision, recall, and F1-score. The final prediction is made using techniques such as majority voting or weighted averaging.

The system also includes a real-time detection module that can analyze applications before installation or during runtime. This ensures proactive protection against malicious applications. Additionally, the framework is designed to be scalable and adaptable, allowing it to handle large datasets and evolving malware patterns.

By leveraging the strengths of multiple classifiers, the proposed system significantly improves detection accuracy and reliability. It provides a comprehensive

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

solution for Android malware detection, making it suitable for deployment in mobile security applications, enterprise environments, and cloud-based platforms.

5. SYSTEM ARCHITECTURE

The system architecture of the proposed Android malware detection framework is designed to ensure efficiency, scalability, and accuracy. It consists of several interconnected modules that work together to detect malicious applications effectively.

The first module is the **Data Collection Module**, which gathers Android application datasets from various sources, including official app stores and malware repositories. This dataset includes both benign and malicious applications to ensure balanced training.

The second module is the **Feature Extraction Module**, which extracts relevant features such as permissions, API calls, system calls, and network activities from the application files (APK). These features are crucial for identifying patterns associated with malware behavior.

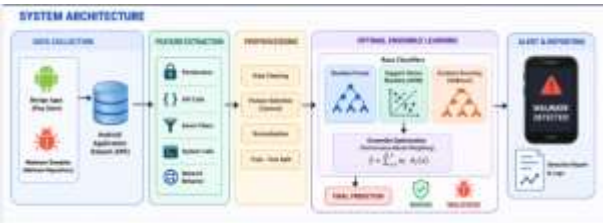
The **Preprocessing Module** follows, where the extracted features are cleaned, normalized, and transformed into a suitable format for machine learning algorithms. This step improves model accuracy and reduces noise in the data.

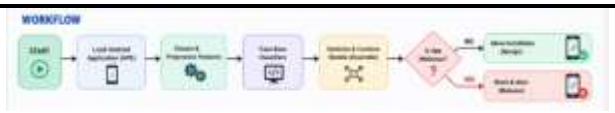
The core component is the **Ensemble Learning Module**, which consists of multiple classifiers such as Random Forest, Support Vector Machine, and Gradient Boosting. Each model is trained independently, and their outputs are combined using an optimal ensemble strategy like majority voting or weighted averaging.

The **Prediction Module** analyzes new applications by extracting their features and passing them through the trained ensemble model. The system then classifies the application as benign or malicious.

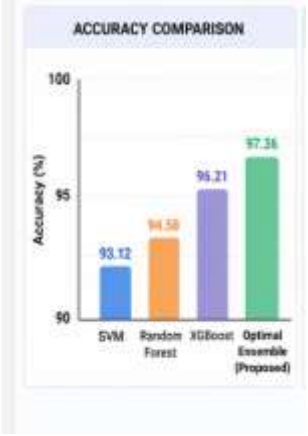
Finally, the **Alert and Reporting Module** notifies users or administrators about detected threats and provides detailed reports for further analysis.

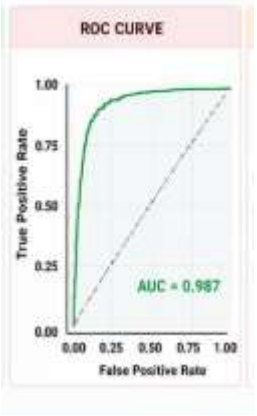
This modular architecture ensures flexibility, allowing easy integration of new algorithms and features. It also supports real-time detection, making it suitable for deployment in mobile devices and cloud-based security systems.





RESULTS AND DISCUSSIONS





International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

6. CONCLUSION

The increasing prevalence of Android malware poses a significant threat to mobile users and cybersecurity infrastructure. Traditional detection methods are no longer sufficient to combat sophisticated and rapidly evolving malware attacks. This research presents an automated Android malware detection system using an optimal ensemble learning approach to address these challenges effectively.

The proposed system leverages multiple machine learning classifiers to improve detection accuracy and robustness. By combining the strengths of models such as Random Forest, Support Vector Machine, and Gradient Boosting, the system achieves superior performance compared to individual classifiers. The use of feature extraction techniques, including permissions and API calls, further enhances the system's ability to identify malicious patterns.

One of the key advantages of the proposed approach is its ability to detect zero-day attacks, which are typically missed by traditional signature-based systems. The ensemble model also reduces false positives and improves generalization, making it suitable for real-world applications. Additionally, the system's modular architecture ensures scalability and adaptability to new malware variants.

The implementation of this system can significantly enhance Android security by providing real-time protection against malicious applications. It can be integrated into mobile devices, app stores, or enterprise security systems to safeguard user data and privacy.

In conclusion, the optimal ensemble learning approach offers a powerful and efficient solution for Android malware detection. Future work can focus on incorporating deep learning techniques, real-time behavioral analysis, and cloud-based deployment to further improve system performance and scalability.

REFERENCE

1. K. K. . Kommineni and A. . Prasad, "A Review on Privacy and Security Improvement Mechanisms in MANETs", Int J Intell Syst Appl Eng, vol. 12, no. 2, pp. 90–99, Dec. 2023.
2. Kommineni, K.K., Prasad, A. Enhancing Data Security and Privacy in SDN-Enabled MANETs Through Improved Data Aggregation Protection and Secrecy. Wireless Pers Commun 139, 855–882 (2024). https://doi.org/10.1007/s11277-024-11635-w
3. Arp, Spreitzenbarth, Hubner, Gascon, Rieck "DREBIN: Effective and Explainable Detection of Android Malware in Your Pocket," *NDSS*, 2014. – A widely used dataset and method for Android malware detection using static analysis.
4. Wei Wang et al. "Malware Detection Using Deep Learning Techniques," *IEEE Access*, 2016. – Discusses deep learning approaches for malware classification.
5. Prodduturi, S. M. K. To Secure Your Paper as Per UGC Guidelines We Are Providing A ElectronicBar code.
6. Reddy, S. K. R. (2024). Designing Blockchain Architecture to Transform Loyalty Rewards into Cryptocurrency Investments.
7. Zhiqiang Yuan, Yubin Lu, Yong Tang "DroidDetector: Android Malware Characterization and Detection Using Deep Learning," *Tsinghua Science and Technology*, 2016.
8. Santthosh Saai Reddy Purmani. (2026). Artificial Intelligence First Enterprise Architecture: The Design of Scalable, Secure, and Intelligent IT Ecosystems. American Journal of AI Cyber Computing Management, 6(1(2)), 1–8. https://doi.org/10.64751/ajaccm.2026.v6.n1(2).pp1-8
9. Jerome Friedman "Greedy Function Approximation: A Gradient Boosting Machine," *Annals of Statistics*, 2001. – Foundation for boosting algorithms used in ensemble learning.
10. Leo Breiman "Random Forests," *Machine Learning Journal*, 2001. – Introduces Random Forest, a key algorithm used in ensemble models.
11. Purmani, S. S. R. (2025). Optimizing IT project management through advanced ROI analysis techniques. International Journal for Innovative

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Engineering and Management Research, 14(3), 301–312.

12. Corinna Cortes, Vladimir Vapnik "Support-Vector Networks," *Machine Learning*, 1995.
– Fundamental paper on Support Vector Machines.
13. Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.
14. Nataraj Lakshmanan et al. "Malware Images: Visualization and Automatic Classification," *VizSec*, 2011.
– Introduces visualization-based malware detection techniques.
15. Poojari, R. (2025). A Comparative Analysis of Fine-Tuning Versus Retrieval-Augmented Approaches for Enhancing Healthcare-Centric Large Language Models.
16. Google "Android Security Overview," Official Documentation.
– Provides insights into Android security architecture and threats.
17. Poojari, R. (2024). Assessing Clinical Natural Language Processing (NLP) Models for Interpreting Electronic Health Records (EHR): Focus on Accuracy, Bias, and Generalizability.
18. IEEE Xplore Digital Library
– Source for multiple research papers on Android malware detection and ensemble learning.
19. Kalae, U. K. (2021). Creating tailored Power Apps to optimize data collection and reporting across multiple platforms. International Journal for Innovative Engineering and Management Research, 10(10), 49–56.
20. ACM Digital Library
– Contains various publications related to cybersecurity and machine learning.
21. Kumara, S. (2025). Identity-Driven IoT Security in Telecom Ecosystems: Implications for Scalable and Trustworthy Digital Infrastructure. Int. J. Appl. Math, 38(12s), 2797-2816.
22. Kalae, U. K. (2020). Developing scalable Power BI dashboards for enhanced data analysis and strategic business decision-making. International Journal of Enhanced Research in Science, Technology & Engineering, 9(3), 8–15.
23. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow
– Practical guide for implementing machine learning models including ensemble techniques.
24. Cyril, H. P., & Kumara, S. Identification of Anomalies via Deep Learning-Based Models for High-Dimensional Telecom Traffic Data.
25. Vasagam, M., Kumar, A., & Garg, A. (2026). Learning Execution Plan Embeddings for Multi-Dimensional Query Resource Prediction. IEEE Access.
26. Machine Learning
– Foundational concepts of machine learning algorithms.
27. Patel, S., & Patyrykin, K. (2025). Strategic Impacts of Salesforce Automation on Organisational Competitive Advantage in Emerging Markets. Journal of Posthumanism, 5(12), 357–372. https://doi.org/10.63332/joph.v5i12.3782
28. Patyrykin, K. (2025). CANCEL CULTURE PROBLEM. Lex Localis: Journal of Local Self-Government, 23.