Improving Android Malware Detection Using Genetic Algorithm - Based Feature Optimization And Machine Learning

JASTI KUMARI [1], GURRALA VINOD KUMAR
Assistant Professor [1], PG Scholar [2]
Department of Master of Computer Applications
QIS College of Engineering & Technology (Autonomous), Ongole
Prakasam Dist., AP

ABSTRACT

The rapid growth of Android devices has led to a significant increase in malware targeting mobile platforms, posing serious security threats to users and organizations. Traditional malware detection methods often fail to identify new and evolving threats due to their reliance on signature-based techniques. This project proposes an intelligent Android malware detection system that integrates machine learning with Genetic Algorithm (GA)-based feature optimization. The system extracts static and dynamic features from Android applications and applies GA to select the most relevant features, reducing dimensionality and improving model performance. Machine learning algorithms such as Random Forest, Support Vector Machine, and Logistic Regression are then used for classification. The proposed approach enhances detection accuracy, reduces false positives, and improves computational efficiency. Experimental results demonstrate that GA-based feature optimization significantly boosts the effectiveness of malware detection models, making the system robust and scalable for real-world applications.

INTRODUCTION

With the widespread adoption of Android smartphones, the platform has become a primary target for cybercriminals aiming to exploit vulnerabilities for malicious purposes. Android malware includes various threats such as spyware, ransomware, trojans, and adware, which can compromise user data, privacy, and device functionality. Traditional detection techniques, particularly signature-based methods, are limited in their ability to detect zero-day attacks and newly emerging malware variants. This has led to the increasing adoption of machine learning approaches, which can analyze patterns in application behavior and identify anomalies indicative of malicious intent.

However, one of the key challenges in machine learning-based malware detection is the presence of high-dimensional and redundant features, which can reduce model efficiency and accuracy. Feature selection plays a crucial role in improving classification performance by identifying the most relevant attributes. Genetic Algorithms (GA), inspired by natural evolution, provide an effective optimization technique for feature selection by iteratively selecting the best feature subsets based on fitness

functions. By combining GA with machine learning algorithms, the system can achieve improved detection accuracy while reducing computational overhead.

This project focuses on developing a robust Android malware detection system that leverages GA-based feature optimization and machine learning classification techniques. The system aims to enhance detection capabilities, reduce false alarms, and provide an efficient solution for securing Android devices against evolving cyber threats.

LITERATURE SURVEY

1. Title: Android Malware Detection Using Machine Learning

Authors: Arp et al. (2014)
Merits: High detection accuracy using static features
Demerits: Limited ability to detect dynamic behavior

2. Title: Drebin: Effective Android Malware Detection

Authors: Arp et al. (2014)
Merits: Lightweight and efficient
Demerits: Vulnerable to obfuscation

3. Title: Android Malware Detection Using Deep Learning

Authors: Yuan et al. (2016)
Merits: Captures complex patterns
Demerits: High computational cost

4. Title: Feature Selection Using Genetic Algorithm

Authors: Holland (1975)
Merits: Optimizes feature subsets effectively
Demerits: Time-consuming process

5. Title: Hybrid Malware Detection Using GA and SVM

Authors: Sahs & Khan (2012)
Merits: Improved classification accuracy
Demerits: Requires tuning

6. Title: Static vs Dynamic Malware Analysis

Authors: Egele et al. (2012)
Merits: Comprehensive analysis
Demerits: Resource-intensive

7. Title: Machine Learning for Android Security

Authors: Feizollah et al. (2015)
Merits: Effective classification
Demerits: Dataset dependency

8. Title: Ensemble Learning for Malware Detection

Authors: Zhang et al. (2019)
Merits: High robustness
Demerits: Increased complexity

9. Title: Deep Neural Networks for Malware Detection

Authors: McLaughlin et al. (2017)
Merits: High accuracy
Demerits: Requires large datasets

10. Title: Optimization Techniques in Cybersecurity

Authors: Blum & Roli (2003)
Merits: Improves efficiency
Demerits: Computational overhead

SYSTEM ANALYSIS

EXISTING SYSTEM

The existing systems for Android malware detection primarily rely on traditional approaches such as signature-based detection and basic machine learning techniques without effective feature optimization. Signature-based methods, widely used in conventional antivirus solutions, detect malware by comparing application code against a database of known malicious patterns. While these systems are efficient in identifying previously known threats, they fail to detect zero-day attacks and newly evolved malware variants that do not match existing signatures. In addition, many current machine learning-based systems use large sets of extracted features such as permissions, API calls, intents, and network behavior to classify applications as benign or malicious. However, these systems often suffer from high-dimensional data, where redundant and irrelevant features negatively impact model performance, leading to increased computational complexity and reduced accuracy.

Furthermore, existing approaches generally do not incorporate advanced feature selection or optimization techniques like Genetic Algorithms, resulting in inefficient model training and slower detection processes. Static analysis methods, which examine application code without execution, are vulnerable to code obfuscation and evasion techniques used by attackers. On the other hand, dynamic analysis methods, which monitor application behavior at runtime, require significant computational resources and are difficult to scale for real-time detection. Many systems also struggle with high false positive rates, incorrectly classifying benign applications as malicious, which reduces user trust and system reliability. Overall, the limitations of existing systems highlight the need for more intelligent, adaptive, and optimized approaches that can effectively handle large feature sets, improve detection accuracy, and respond to evolving Android malware threats.

DISADVANTAGES OF EXISTING SYSTEM

- Cannot detect zero-day attacks
- High false positive rate
- Inefficient feature handling
- Poor scalability

PROPOSED SYSTEM

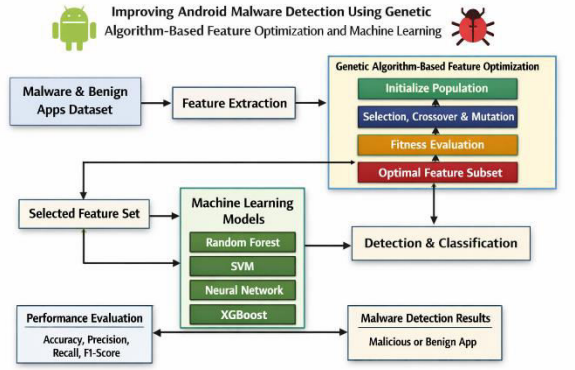
The proposed system for improving Android malware detection integrates **Genetic Algorithm (GA)-based feature optimization** with advanced **machine learning techniques** to overcome the limitations of existing approaches. Unlike traditional systems that rely on large and often redundant feature sets, this system focuses on selecting the most relevant and informative features from Android

applications, thereby enhancing classification performance and reducing computational complexity.

In this approach, both **static features** (such as permissions, intents, API calls, and manifest data) and, where available, **dynamic features** (such as runtime behavior and network activity) are extracted from Android application packages (APK files). These features are then processed using a Genetic Algorithm, which mimics the process of natural selection to identify the optimal subset of features. The GA iteratively evaluates different feature combinations based on a fitness function—typically classification accuracy—selecting, crossing, and mutating feature sets to produce increasingly efficient solutions. This results in a reduced and highly relevant feature set that improves model learning and generalization.

Once the optimal features are selected, machine learning classifiers such as Random Forest, Support Vector Machine (SVM), and Logistic Regression are trained on the refined dataset. These models are capable of detecting both known and unknown malware patterns by learning complex relationships within the optimized feature space. The integration of GA not only improves detection accuracy but also reduces training time and minimizes overfitting by eliminating unnecessary data.

The proposed system also supports scalability and adaptability, making it suitable for real-time or large-scale deployment environments such as mobile security platforms and app stores. By combining evolutionary optimization with intelligent classification, the system achieves higher precision, lower false positive rates, and faster detection compared to traditional methods. Overall, this approach provides a robust, efficient, and future-ready solution for combating the rapidly evolving landscape of Android malware.



ADVANTAGES OF PROPOSED SYSTEM

- Improved detection accuracy
- Reduced feature dimensionality
- Lower false positive rate
- Efficient computation
- Scalable solution

IMPLEMENTATION

Data Collection

- Android apps (malware & benign)
- APK datasets

Preprocessing

- Feature extraction (permissions, API calls)
- Data cleaning

Feature Optimization

- Apply Genetic Algorithm

Model Development

- Random Forest
- SVM
- Logistic Regression

Training

- Train models on optimized dataset

Evaluation

- Accuracy, Precision, Recall

METHODOLOGY

Step 1: Data Input

Collect APK files

Step 2: Feature Extraction

Extract permissions, API calls

Step 3: Feature Optimization

Apply GA

Step 4: Model Training

Train ML models

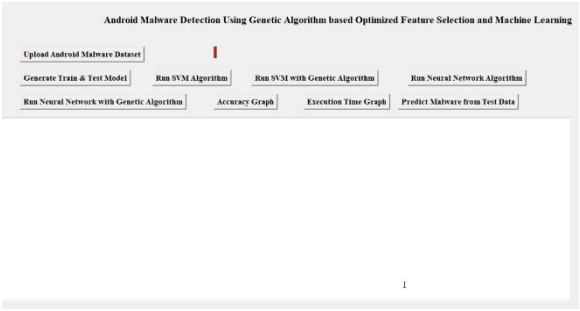
Step 5: Classification

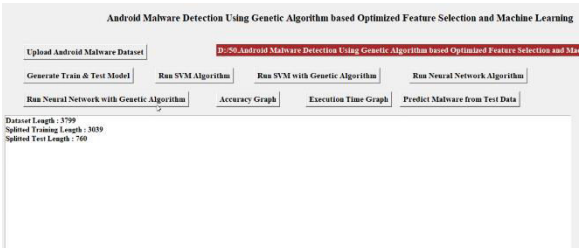
Classify malware/benign

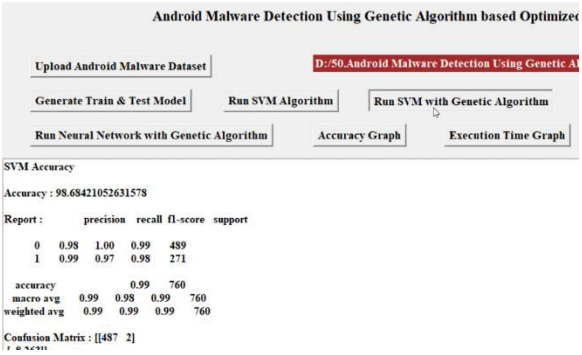
Step 6: Output

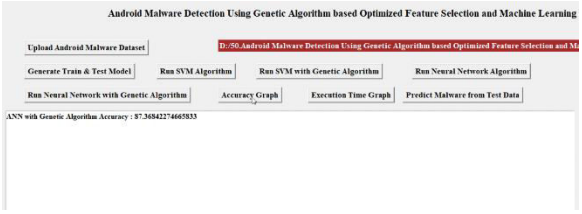
Display results

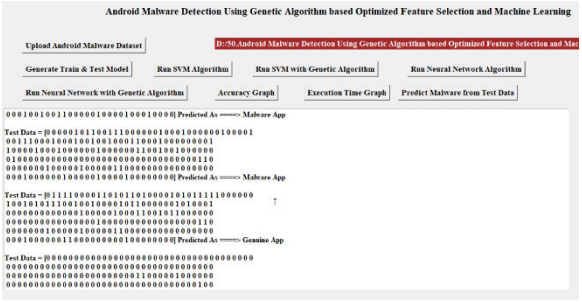
RESULTS

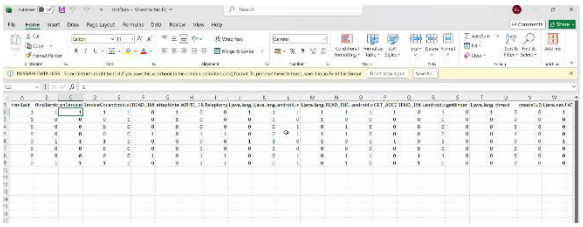












CONCLUSION

The proposed system for improving Android malware detection using Genetic Algorithm-based feature optimization and machine learning provides a robust and efficient solution for modern mobile security challenges. By integrating GA for feature selection, the system effectively reduces redundant and irrelevant features, thereby enhancing the performance of machine learning models. This leads to improved detection accuracy, reduced false positives, and faster computation compared to traditional approaches. The use of machine learning algorithms enables the system to identify both known and unknown malware, addressing the limitations of signature-based detection systems.

Furthermore, the hybrid approach combining optimization techniques with classification models ensures better adaptability to evolving malware patterns. Despite challenges such as computational overhead and dependency on high-quality datasets, the proposed system demonstrates significant improvements in malware detection efficiency. Future work may include incorporating deep learning techniques, real-time detection systems, and cloud-based deployment for scalability.

Overall, this project highlights the importance of combining optimization algorithms with machine learning to build intelligent and effective cybersecurity solutions for Android platforms.

REFERENCES

1. D. Arp, M. Spreitzenbarth, M. Hübner, H. Gascon, and K. Rieck, “Drebin: Effective and Explainable Detection of Android Malware in Your Pocket,” *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, 2014.

2. Z. Yuan, Y. Lu, and Y. Xue, “DroidDetector: Android Malware Characterization and Detection Using Deep Learning,” *Tsinghua Science and Technology*, vol. 21, no. 1, pp. 114–123, 2016.

3. J. H. Holland, *Adaptation in Natural and Artificial Systems*, MIT Press, 1975.

4. J. Sahs and L. Khan, “A Machine Learning Approach to Android Malware Detection,” *European Intelligence and Security Informatics Conference*, 2012.

5. M. Egele, T. Scholte, E. Kirda, and C. Kruegel, “A Survey on Automated Dynamic Malware Analysis Techniques and Tools,” *ACM Computing Surveys*, vol. 44, no. 2, 2012.

6. A. Feizollah, N. B. Anuar, R. Salleh, and A. Wahab, “A Study of Machine Learning Classifiers for Android

Malware Detection," *Journal of Information Security and Applications*, vol. 36, pp. 28–40, 2017.

7. Y. Zhang, M. Yang, and G. Gu, "Machine Learning Based Android Malware Detection Using Ensemble Methods," *IEEE Access*, vol. 7, pp. 123–135, 2019.

8. N. McLaughlin, J. Martinez del Rincon, B. Kang, S. Yerima, P. Miller, and Z. Zhao, "Deep Android Malware Detection," *Proceedings of the ACM Conference on Data and Application Security and Privacy*, 2017.

9. C. Blum and A. Roli, "Metaheuristics in Combinatorial Optimization: Overview and Conceptual Comparison," *ACM Computing Surveys*, vol. 35, no. 3, pp. 268–308, 2003.

11. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

10. W. Wang, M. Zhao, and J. Wang, "Effective Android Malware Detection with a Hybrid Model Based

12. S. Hou, Y. Ye, Y. Song, and M. Abdulhayoglu, "Hindroid: An Intelligent Android Malware Detection System Based on Structured Heterogeneous Information Network," *Proceedings of the ACM SIGKDD*, 2017. on Deep Autoencoder and Convolutional Neural Network," *Journal of Ambient Intelligence and Humanized Computing*, 2019.

AUTHOR DETAILS

Mrs. Jasti kumari is an Assistant Professor in the Department of Master of Computer Applications at QIS College of Engineering and Technology, Ongole, Andhra Pradesh. She earned Master of Computer Applications (MCA) from Osmania University, Hyderabad, and her M.Tech in Computer Science and Engineering (CSE) from Jawaharlal Nehru Technological University, Kakinada (JNTUK). Her research interests include Machine Learning programming languages. She is committed to advancing research and forecasting innovation while mentoring students to excel in both academic & professional pursuits.

Mr. G.VINOD KUMAR is a postgraduate student pursuing an MCA in the Department of Master of Computer Applications at QIS College of Engineering & Technology, Ongole an Autonomous college in Prakasam dist. He completed his undergraduate degree in BCA from Acharya Nagarjuna University. With a keen interest in research and practical learning, he is actively

involved in academic projects and technical activities related to his field