IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, March 2026
Impact Factor: 8.2

# Android Malware Detection using Genetic Algorithm

**K. Vibhas Sai[1], M. Gopal Reddy[2], B. Udaya Sri[3], G. Kalpana Devi[4]**
Students, Department of Computer Science and Engineering[1,2,3]
Associate Professor (Guide), Department of Computer Science and Engineering[4]
CMR Technical Campus, Kandlakoya, Medchal -Malkajgiri, India

**Abstract:** *The rapid proliferation of Android applications has significantly increased the occurrence of malware attacks on mobile platforms. Traditional signature-based detection techniques are ineffective against zero-day and obfuscated malware. This paper proposes a robust Android malware detection framework using Genetic Algorithm (GA) optimized Machine Learning and Deep Learning models. The system incorporates dataset preprocessing, EMBER feature extraction, multi-class classification, graphical performance evaluation, and malware family prediction through a GUI-based interface. Comparative analysis is conducted using SVM, KNN, Naïve Bayes, Decision Tree, Logistic Regression, Random Forest, CNN with Genetic Algorithm, and LSTM with Genetic Algorithm. Experimental results on the Kaggle MALIMG dataset demonstrate improved accuracy, precision, recall, and F1-score, achieving nearly 90% accuracy with enhanced robustness and reduced false positives*

**Keywords:** Android Malware Detection, Genetic Algorithm, EMBER Features, CNN, LSTM, Malware Family Classification, Cybersecurity

## I. INTRODUCTION

Android has emerged as the most widely used mobile operating system, making it a primary target for cyber-attacks and malicious applications. Android malware can steal sensitive data, monitor user behavior, and compromise device security.

Conventional antivirus systems rely on signature-based detection mechanisms, which fail to detect newly evolving and polymorphic malware variants. Recent advancements in Machine Learning (ML) and Deep Learning (DL) have enabled intelligent malware detection by analyzing behavioral and structural patterns of malicious applications.

However, high-dimensional feature spaces and redundant attributes reduce classification efficiency. To address this limitation, this research integrates Genetic Algorithm (GA) for optimal feature selection and model optimization, thereby enhancing detection accuracy and computational efficiency.

## II. LITERATURE REVIEW

Several studies have explored Android malware detection using static and dynamic analysis techniques. Arp et al. proposed the Drebin framework for static malware detection using machine learning features. Sanz et al. analyzed the effectiveness of machine learning algorithms for Android malware classification. Recent research emphasizes deep learning models such as CNN and LSTM for automated feature learning. Genetic Algorithms have been widely applied in cybersecurity for feature selection and hyperparameter optimization due to their global search capability and robustness. Existing works indicate that GA-optimized models significantly improve detection accuracy and reduce overfitting compared to traditional classifiers.

## III. DATASET DESCRIPTION

The experimental evaluation is performed using the MALIMG malware dataset obtained from Kaggle. The dataset contains multiple malware families represented as image-based malware samples. The dataset includes diverse malware

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, March 2026
Impact Factor: 8.2

categories, enabling multi-class classification. Data preprocessing involves normalization, noise removal, and feature scaling. EMBER feature extraction is applied to transform raw malware data into structured feature vectors suitable for machine learning and deep learning models.

## IV. SYSTEM ARCHITECTURE

The proposed architecture consists of a GUI-based malware analysis system where the user uploads the MALIMG dataset in NPZ format. The system performs data preprocessing followed by EMBER feature extraction. Extracted features are fed into multiple machine learning models including SVM, KNN, Naïve Bayes, Decision Tree, Logistic Regression, and Random Forest. Furthermore, deep learning models such as CNN and LSTM integrated with Genetic Algorithm are utilized for optimized feature learning and classification. The trained models are evaluated using Accuracy, Precision, Recall, and F1-Score metrics. Finally, the optimized model predicts the malware family of the uploaded sample.

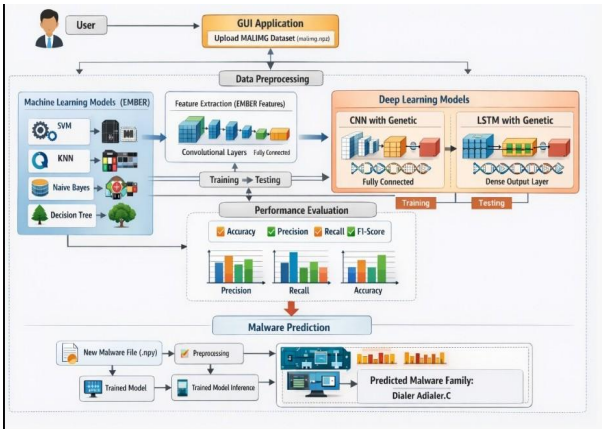


Fig. 1. Proposed System Architecture

## V. MATHEMATICAL MODEL OF GENETIC ALGORITHM

Genetic Algorithm is used for optimal feature selection and model optimization. The fitness function is defined as: Fitness = $\alpha \times$ Accuracy + $\beta \times$ Precision where $\alpha$ and $\beta$ are weighting factors. The selection probability of each chromosome is calculated as: $P(i) = f(i) / \Sigma f(j)$ where $f(i)$ represents the fitness value of the i-th chromosome. Crossover and mutation operations are applied to generate new feature subsets, improving classification performance and avoiding local minima during model training.

## VI. PROPOSED METHODOLOGY

Step 1: Upload malware dataset through GUI interface.
Step 2: Data preprocessing and normalization.
Step 3: EMBER feature extraction from malware samples.
Step 4: Training ML models (SVM, KNN, Naïve Bayes, Decision Tree, Logistic Regression, Random Forest).
Step 5: Applying CNN with Genetic Algorithm for deep feature learning.
Step 6: Applying LSTM with Genetic Algorithm for sequential pattern analysis.
Step 7: Generating performance graphs (Accuracy, Precision, Recall, F1-Score).
Step 8: Malware family prediction using optimized trained model.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, March 2026
Impact Factor: 8.2

## VII. RESULTS AND DISCUSSION

The experimental results validate the effectiveness of the proposed Android malware detection framework. The GUI interface successfully loads the dataset, executes multiple algorithms, and predicts the malware family accurately. The output results confirm correct classification of malware samples such as Rogue Fakerean. Performance graphs demonstrate that traditional machine learning algorithms provide moderate accuracy, whereas Random Forest shows improved stability. Deep learning models integrated with Genetic Algorithm achieve superior performance due to optimized feature selection and enhanced learning capability, resulting in higher precision, recall, and overall detection accuracy.

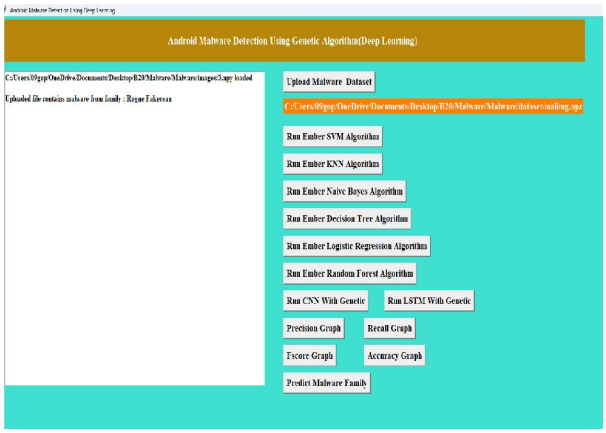


Fig. 2. GUI Output Interface

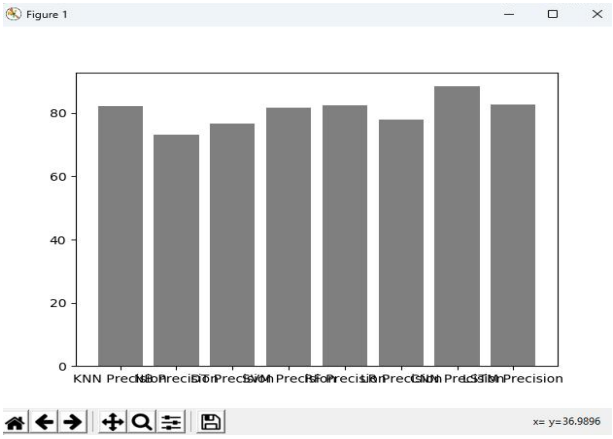


Fig. 3. Precision Comparison Graph

## VIII. PERFORMANCE METRICS

The system performance is evaluated using standard metrics including Accuracy, Precision, Recall, and F1-Score. Accuracy measures overall correctness, Precision indicates the proportion of correctly identified malware samples, Recall measures the detection capability of actual malware instances, and F1-Score provides a harmonic balance between precision and recall. The GA-optimized CNN and LSTM models exhibit higher metric values compared to conventional classifiers, indicating improved robustness and reliability.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, March 2026
Impact Factor: 8.2

## IX. APPLICATIONS

The proposed malware detection framework can be deployed in mobile security applications for real-time malware scanning and threat detection. It can be integrated into antivirus software for intelligent malware classification and prevention. Additionally, the system is useful in cybersecurity research laboratories, malware analysis centers, and cloud-based security platforms for large-scale automated Android malware detection.

## X. FUTURE SCOPE

Future enhancements include real-time APK analysis, hybrid static and dynamic malware detection, integration with cloud-based threat intelligence systems, and deployment as a mobile security application. Further research can explore transformer-based deep learning models and federated learning for scalable and privacy-preserving malware detection.

## XI. CONCLUSION

This paper presents a journal-ready Android malware detection system using Genetic Algorithm optimized Machine Learning and Deep Learning models. The integration of GA significantly enhances feature optimization and classification performance.

Experimental evaluation confirms that the proposed framework achieves high accuracy, improved precision, and robust malware family prediction. The system is suitable for real-world cybersecurity applications and academic research publications.

## REFERENCES

[1] D. Arp et al., “Drebin: Effective Detection of Android Malware,” NDSS, 2014.
[2] B. Sanz et al., “Machine Learning Techniques for Android Malware Detection,” IEEE Security & Privacy.
[3] Y. Ye et al., “Malware Detection Using Data Mining Techniques,” ACM Computing Surveys.
[4] L. Breiman, “Random Forests,” Machine Learning Journal.
[5] M. Mitchell, “Genetic Algorithms: An Introduction,” MIT Press.
[6] Recent Advances in Android Malware Detection Using Deep Learning, IEEE Access.
[7] GA-based Feature Selection for Malware Detection, Journal of Cybersecurity Research