DOI: 10.5281/zenodo.17805319

International Journal of Research and Applications
Oct - Dec © 2025 Transactions 12(48): 3801-3806

eISSN : 2349 – 0020
pISSN : 2394 – 4544
www.ijraonline.com

COMPUTERS RESEARCH ARTICLE

Android Malware Detection Using Genetic Algorithm

B. Maheshwari [1] and T. Archana [2]

[1] M. Tech. (Pursuing), [2]Assistant Professor
[1, 2] Department of Computer Science and Engineering
[1, 2] Kakatiya University Campus, Warangal, Telangana, India.

maheshwarireddybiradar@gmail.com [1], archanapraneeth@gmail.com [2]

ABSTRACT

Android is an open source operating system which is free and Google assists developers to place the Android applications on its Play Store. Anyone can create an android game and place it at the play store at no cost. Hackers are also attracted by this attribute of Android and are creating malicious applications to be installed on the play store. When one installs such malware, it will steal information on the phone and forward it to hackers or provide the scammers with complete control of the phone. The way we do it is through a ML approach to detect malware in mobile applications so that the user does not get exposed to such apps. To detect malware within an app, we must reverse engineer it to retrieve all the code in it and then examine whether it is carrying out any malevolent actions, such sending SMS messages or stealing access to contact details. We will recognize that the application is malicious if such behavior is exposed in the code. More than 100 permissions, including transact, on Service Connected, bind Service, API call signature, Service Connection, and API call signature, can be granted to a single application, and so on. We have to drag these permissions out of the code and create a features dataset. In case the app is generally authorized to do that, then we will record value 1 into the features dataset, and vice versa. These characteristics will be used to identify the dataset app as malware or good software.

Keywords: Android Malware Detection, Genetic Algorithm, Machine Learning, Feature Selection, Static and Dynamic Analysis, Evolutionary Computing, Mobile Security, Optimization Techniques.

1. INTRODUCTION

Malware is a huge issue to the safety of computer users, and it could cost an organization a lot of money. The "Internet of Things (IoT)" became more convenient in use, so the attention of criminals turned to it. Depending on the nature of operation, malware is referred to by various terms, such as adware, root kit, backdoor, ransom ware, Trojans, worms, spyware, etc. This has made the researchers more difficult in detecting these malwares.

The two predominant methods of analysis and detection of malware are the static analysis and dynamic analysis. When one does not run a program, and simply examines and pulls out information (data) contained in an executable file, it is known as static analysis. Dynamic analysis this is the implementation of the malware and observing its behavior on the machine. When a new malware releases, the professionals tend to go through its sample manually or create a program that can be used to compare it with other malware of the same category. Image classification has improved significantly in the recent past with the advent of DL techniques. CNN yielded higher results. In this example, feature engineering, feature learning and feature representation is all carried out in a mechanical manner.

2. LITERATURE REVIEW

Atieh, A., et al. (2022):
In the article MDPI, the authors have outlined the enhancement of the ensemble learning models which are based on the Genetic Algorithms (GA). They used the GA to optimize the model parameters and select the most helpful features to the Android application. Their proposed model was superior in locating things and lowering the number of false positives compared to the traditional ensemble methods. However, by virtue of the optimization aspect, it was more expensive to train with a computer.

Xie, N., Qin, Z., & Di, X. (2023):
The study was referred to as "GA-StackingMD: Android Malware Detection Method Based on Genetic Algorithm Optimized Stacking" and was developed using stacking ensembles of GA-optimized feature subsets and base-learner settings (Applied Sciences, MDPI). The technique was more precise and recalled higher numbers on Android malware benchmark datasets. Although the outcome was favorable, stacking process increased the complexity of the model and trained it longer.

Beştaş, M. Ş., & Dinler, Ö. (2023):
The article authored by the International Journal of Pure and Applied Sciences examined the ways various metaheuristic techniques, including GA, could be utilized to locate malicious Android applications. The study examined the variants of classical metaheuristics such as "Genetic Algorithms (GA) and Particle Swarm Optimization (PSO)". It discovered that GA produced less fluctuating outcomes. Nevertheless, the effectiveness of the feature engineering required a significant amount of success.

Anđelić, N., & Baressi Šegota, S. (2024):
In the paper, "Achieving High Accuracy in Android Malware Detection through Genetic Programming Symbolic Classifier", Genetic Programming (symbolic classifiers) was adopted as an extension of GA (Computers, MDPI). This approach generated models that could be comprehended yet very precise. Although the GP models were effective, they required a high level of computing in order to evolve and be tested.

Polatidis, N., et al. (2024):
In a study named: "FSSDroid: Feature Subset Selection in Android Malware Detection (World Wide Web, SpringerLink)", it was proposed to use the feature selection technique which is based on GA to select the best sets of rights and API calls. Laboratory experiments indicated that the approach improved the performance of the classifier and reduced a lot of irrelevant features. However, it required a series of GA steps to converge.

Padmalatha, E., et al. (2023):
A paper named "Detection of Android Malware using Feature Selection with a Hybrid Genetic Algorithm and Simulated Annealing (IJRITCC)" describes a new feature selection scheme that will employ a combination of GA and simulated annealing. The analysis of the selected features was performed with the help of SVM and DBN classifiers. The combination method proved to be more precise, however, it was slower to train since it involved two optimization processes.

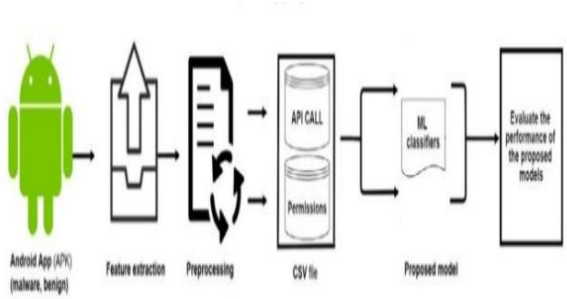
Swarajya Lakshmi, B., et al. (2023):
It was authored on the processes of locating malware on Android by the application of a superior genetic algorithm that employs feature selection and ML (IJRPR). Another better genetic algorithm discussed in this paper is that which has adaptable mutation rates and crossing rates. This technique served a splendid purpose of discovering significant fixed and dynamic characteristics, which resulted in better desegregation. The issue was that parameter values relied on the already set ones in order to determine GA stability.

Nayva Sree, V., et al. (2022):
A study was published by the Advanced Engineering Science and was referred to as Android Malware Detection using Genetic Algorithm and ML. It employed the GA-based feature selection and classical feature classifiers such as DT and RF. The model performed better in terms of performance measures than baseline methods. However, the research largely examined permissions-based features, which may not reveal how malware evolves the way it behaves with time.

Gupta, R., Sharma, K., & Garg, R.K. (2024):
The article "Innovative Approach to Android Malware Detection: Prioritization of Critical Features by Rough Set Theory (Electronics, MDPI)" also focused on feature prioritization, but prioritization via a similar structure to GA optimization, although it was not necessarily GA. It concentrated on feature dependence and reduction and the outcomes were as good as those of evolutionary techniques. Its disadvantage was that it based on static analysis which may fail to detect malicious activities occurring in the middle of execution.

Manzil, H.H.R., & Manohar Naik, S. (2023):
The authors of "Android Malware category Detection Using a Novel Feature Vector-Based Machine Learning Model (Cybersecurity, Springer)" recommended a new method of utilizing feature vectors in order to classify malware as families. The study did not apply GA per se, but it compared its approach to the GA-based features selection, which indicated that GA is an excellent option to disregard unnecessary features and to make the classification more precise.

3. SYSTEM ARCHITECTURE



The Malware and good ware are two types of Android apps (APKs) that are reverse-engineered to retrieve data, including permissions and the quantity of app components, including Activity, Services, Content Providers, and so on. They are displayed as feature vectors in the CSV format, with the class names for malware and good ware being 0 and 1, respectively. The CSV file is inputted into the Genetic Algorithm that picks the most desirable set of features by diminishing the number of dimensions. The SVM and the Neural Network are both ML classifiers taught using the best set of the features that was discovered. AndroidManifest.xml provides the name of the things that must not change in the suggested approach. This is an file with all the necessary information on apps that any android platform requires. The Androguard tool has been used to decompose APKs and retrieve its fundamental features.

Advantages of proposed system: Proposed a new and workable feature selection algorithm to improve the overall detection accuracy. The detection of new forms of Android malware that are threats of the so-called zero-day can be implemented through a method based on ML and the analysis of the environment in both its static and dynamic forms.

4. DATA ANALSISES

The main component of ML is feature selection which helps to diminish dimensions in data and a lot of research has been conducted to identify a dependable feature selection procedure. The filter was used to choose the features. method and the wrapping method. Under the filter method, features are selected according to their performance in a series of statistical tests that test the significance of the features examining their ability to correlate well with the dependent or result variable. A subset of features is found by the wrapper method through the dependence variable to estimate the usefulness of a subset of features. Consequently, filter methods do not depend on the ML algorithm which was employed to train the model, but on the best feature subset that the wrapper method chooses. A subset evaluator in the wrapper approach is the one that looks at all the possible subsets and then he employs a classification method to demonstrate that the features contained in each subset are sufficiently good to appeal to the classifiers.

The group of features that the classification process is best suited to is what the classifier considers. Various methods in which the evaluator can locate the group are there. They are random search, depth first search, breadth first search and hybrid search. The filter method is employed to rank all features

in the collection with the help of a ranker and an attribute evaluator. The process, in this instance, involves doing without one such feature at a time to determine how the classification algorithm can do a good guess on the next thing to happen. What the ranker algorithms are doing with weights or ranks is not the same thing that the classification algorithms are doing with them. Wrapper method is good in testing ML, whereas filter method is good in testing data mining, which contains numerous features (tens of millions).

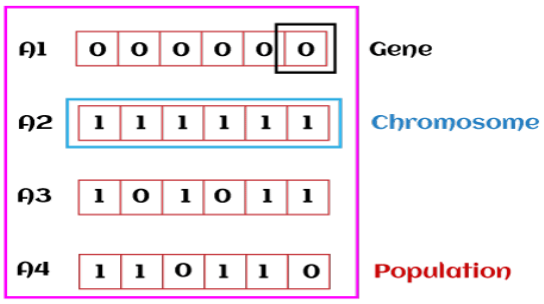
Genetic Algorithm:

The genetic algorithm is an algorithm that generates good answers through the evolutionary generational cycle. These algorithms transform the population in different ways that will either improve or transform it to get a better fit.

The following set of challenging optimization problems can be solved in just five steps:

1. *Initialization*

The initial phase in genetic algorithm is to create a population or a set of people. Every individual in this instance becomes the solution to the problem. Individuals consist or are characterized by a combination of aspects called Genes. It is possible to solve the problem by joining the genes together in a string to create chromosomes. One of the most typical methods to start up a program is by use of random binary strings.



2. **Fitness assignment**

The fitness tool is applied to calculate the fitness of an individual. In other words, the art of how to fight with other human beings. In each round, people are evaluated in terms of their role in fitness. Each person is assigned a fitness number via the fitness function. This figure informs us further concerning the possibilities of being selected to be reproduced. The health score is the higher, the more chances the animal is to be selected as an animal to reproduce.

3. **Selection**

During the selecting process, people are selected to bear children. The selected individuals are then grouped in pairs in order to enhance procreation. Subsequently, these people pass on their DNA to the next generation.

Selection methods are of three types, which include:

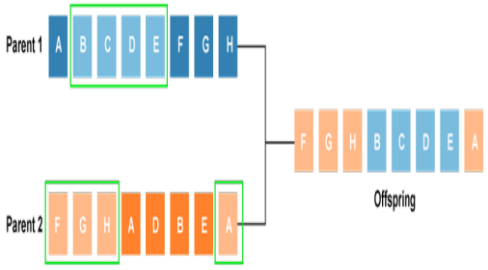
- Roulette wheel selection
- Tournament selection
- Rank-based selection

4. *Reproduction*

Once the step of choosing is done, the reproduction step follows in which a child is created. In this stage of the genetic method there are The parent population is subjected to two variation operators.

The two employees who participate in the reproduction phase are as follows:

Crossover: One of the most important phases in the genetic algorithms reproduction section is the crossover. This is accomplished by randomly choosing a gene crossing location. In order to create a new person who is the child, the crossover operator then modifies the genetic code of two parents who are members of the current generation.



Up to the time of the crossover point, the genes of the parents exchange between themselves. These are new babies who are introduced to the population.

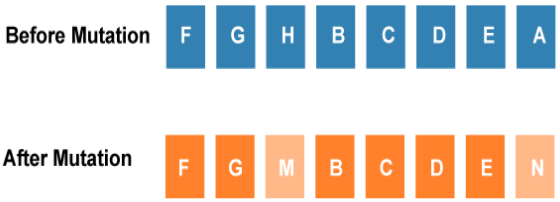
This process or crossing may also be called as such. Various types of crossing styles:

- One point crossover
- Two-point crossover
- Livery crossover
- Inheritable Algorithms crossover

Mutation: The mutation operator introduces random genes into the progeny (new child) to preserve population variety. It can be achieved by rearranging certain chromosomes.

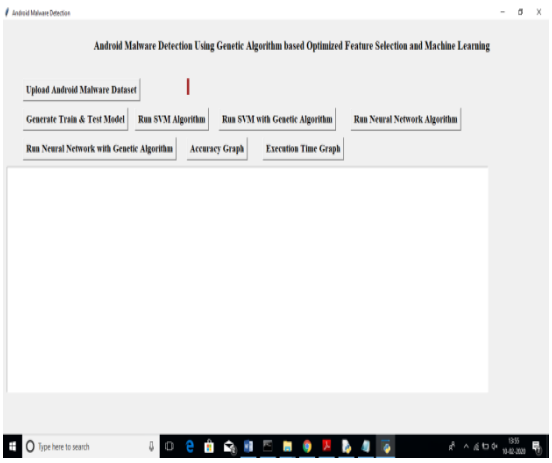
In order to improve diversity and address the problem of early merging, mutation is helpful. The following image illustrates the process of change: types of accessible mutation styles,

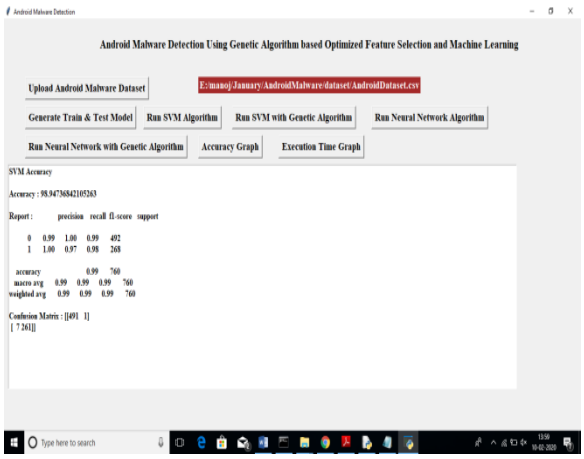
- Flip bit mutation
- Gaussian mutation
- Exchange/Swap mutation



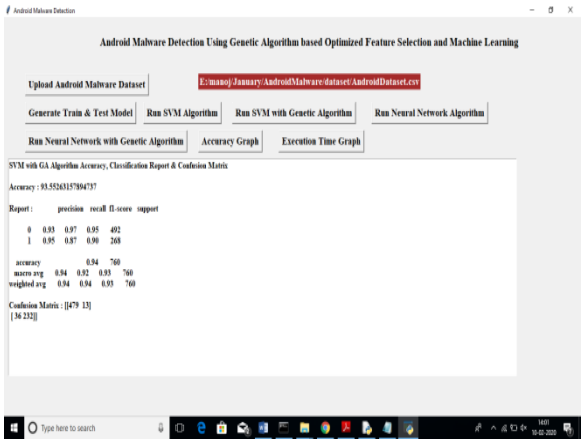
*5. **Termination:*** After the reproduction stage is done, the process is terminated by a stopping condition. Once the threshold fitness solution has been attained, the algorithm terminates. It will discover that the last solution will be the best solution of all those.

5. IMPLEMETATION RESULTS

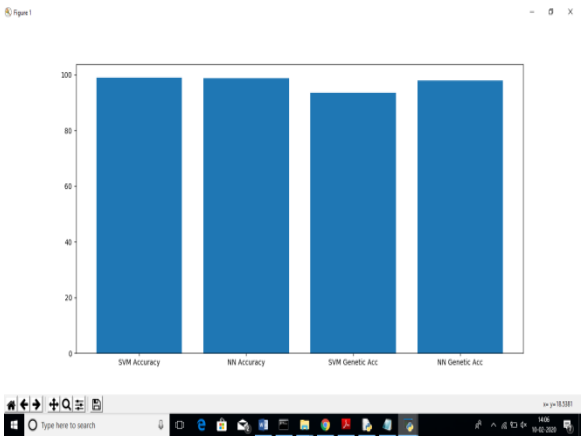




We achieved SVM of 98 percent accuracy on the above screen. It is time to select the best features and this is done by clicking on Run SVM with Genetic Algorithm and then SVM is run on those features in order to achieve accuracy.



As shown in the picture above, SVM using the Genetic algorithm was successful 93 percent of the times. Although Genetic with SVM is not as accurate, it will be faster, which is revealed by the graph of comparison.



The x-axis of the above graph displays the name of the algorithm whereas the y-axis displays the accuracy. All in all, SVM achieved a large accuracy. Go ahead and press the button which is titled "Execution Time Graph" to view the time every method took to execute.

6. CONCLUSION & FUTURE ENHANCEMENT

Android platforms are being faced with threats daily primarily by way of harmful apps or malware. Due to this reason, development of a system capable of detecting these types of malware at an accurate rate is highly important. The methods based on ML are implemented in cases when signature-based techniques are unable to locate new forms of malware that are considered a threat on a zero-day basis. The proposed algorithm attempts to utilize the evolutionary genetic algorithm to identify the most effective set of characteristics which may be employed in the most optimal manner to train ML algorithms.

The proposed approach to identifying malware could be improved with a better feature selection technique, dynamic malware analysis technique, adversarial ML, privacy-preserving technique, expanding to IoT devices, DL architecture, and real-time malware detection and response. A comprehensive test over a wide range of sets of data with various types of malware may provide valuable data concerning the efficiency of the system and its applicability to other cases.

REFERENCES

1) Atieh, A., & …, "Android Malware Classification Using Optimized Ensemble Learning Based on Genetic Algorithms", *Sustainability*, 2022, 14(21), 14406. MDPI+1
2) Xie, N., Qin, Z., Di, X., "GA-StackingMD: Android Malware Detection Method Based on Genetic Algorithm Optimized Stacking", *Applied Sciences*, 2023, 13(4), 2629. MDPI+1
3) Beştaş, M. Ş., & Dinler, Ö., "Detection of Android Based Applications with Traditional Metaheuristic Algorithms", *International Journal of Pure and Applied Sciences*, 2023, 9(2), 381-392. DergiPark
4) Anđelić, N. & Baressi Šegota, S., "Achieving High Accuracy in Android Malware Detection through Genetic Programming Symbolic Classifier", *Computers*, 2024, 13(8), 197. MDPI
5) Polatidis, N., Kapetanakis, S., Trovati, M., Korkontzelos, I., Manolopoulos, Y., "FSSDroid: Feature subset selection for Android malware detection", *World Wide Web*, 2024, article 50. SpringerLink+1
6) Padmalatha, E., Venkata Krishna Reddy, M., Suvarna Kumari, T., Kabeeruddin, "Detection of Android Malware using Feature Selection with a Hybrid Genetic Algorithm and Simulated Annealing (SVM and DBN)", *International Journal on Recent and Innovation Trends in Computing and Communication*, 2023, 11(10), 1481-1487. ijritcc.org+1
7) Swarajya Lakshmi, B., Pranavi, S., Jayalakshmi, C., Gayatri, K., Sireesha, M., Akhila, A., "Detecting Android Malware with an Enhanced Genetic Algorithm for Feature Selection and Machine Learning", *International Journal of Research Publication and Reviews*, 2023, 4(4), 894-901. IJRPR
8) Nayva Sree, V. et al., "Android Malware Detection using Genetic Algorithm and Machine Learning", *Advanced Engineering Science*, 2022, Vol 54, Issue 02. advancedengineeringscience.com
9) Gupta, R., Sharma, K., Garg, R.K., "Innovative Approach to Android Malware Detection: Prioritizing Critical Features Using Rough Set Theory", *Electronics*, 2024, 13(3), 482. MDPI
10) (While not strictly GA-only) Manzil, H.H.R., & Manohar Naik, S., "Android malware category detection using a novel feature vector-based machine learning model", *Cybersecurity*, 2023, 6, 6