**INTRODUCTION**

To this degree, it is guaranteed that mobile devices are an integral part of most people's daily lives. Furthermore, Android now controls the vast majority of mobile devices, with Android devices accounting for an average of 80% of the global market share over the past years. With the ongoing plan of Android to a growing range of smart phones and consumers around the world, malware targeting Android devices has increased as well. Since it is an open-source operating system, the level of danger it poses, with malware authors and programmers implementing unwanted permissions, features and application components in Android apps. The option to expand its capabilities with third-party software is also appealing, but this capability comes with the risk of malicious device attacks. When the number of smart phone apps increases, so does the security problem with unnecessary access to different personal resources. As a result, the applications are becoming more insecure, and they are stealing personal information, SMS frauds, ransom ware, etc.

In contrast to static analysis methods such as a manual assessment of AndroidManifest.xml, source files and Dalvik Byte Code and the complex analysis of a managed environment to study the way it treats a program, Machine Learning includes learning the fundamental rules and habits of the positive and malicious settings of apps and then data-venabling. The static attributes derived from an application are extensively used in machine learning methodologies and the tedious task of this can be relieved if the static features of reverse-engineered Android Applications are extracted and use machine learning SVM algorithm, logistic progression, ensemble learning and other algorithms to help train the model for prediction of these malware applications [1].

Machine learning employs a range of methodologies for data classification. SVM (Support Vector Machine) is a strong learner that plots each data item as a point in n-dimensional space (where n denotes the number of features you have), with the value of each feature becoming the vector value. Then it executes classification by locating the hyper-plane that best distinguishes the two groups, leading to an improvement identification property for any two parameters. Conversely, boosting or ensemble techniques like Adaboost are assigned higher weights to rectify the behavior of misclassified variables in conjunction with other machine algorithms. When combined alongside weak classifiers, our preliminary model benefits from deploying such models since they have a high degree of precision or classification. [2], [3], [4], supports classifiers in their system models to find the highest accuracy. Although using ensemble or strong classifiers can cause problems like multi collinearity, which in a regression model, occurs when two or more independent variables are strongly associated with one another. In multivariate regression, this indicates that one regression analysis may be forecasted from another independent variable. This scope of the study can be presented as a detection journal analysis itself and can present several experimentations and results based on machine learning models [5], [6].

When an app has access to a resource in the most recent versions of Android OS, it must ask the OS for approval, and the OS will ask the user if they wish to grant or refuse the request via a pop-up menu. Many reports have been performed on the success of this resource management approach. The studies showed consumers made decisions by giving all requested access to the applications to their privileges requests [7]. In contrast to this, over 70% of Android mobile applications seek extra access that is not needed. They also sought a permit that is not needed for the app to run. A chess game that asks for photographs or requests for SMS and phone call permits, or loads unwanted packages are an example of an extra requested authorization. So, trying to assess an app's vindictiveness and not understanding the app is a tough challenge. As a result, successful malicious app monitoring will provide extra information to customers to assist them and defend them from information disclosure [8]. Figure 1 elaborates the android risk framework through the Google Play platform, which is then manually configured by the android device developers.

Contrary to other smart phone formats, such as IOS, Android requires users to access apps from untrusted outlets like file- sharing sites or third-party app stores. The malware virus problem has become so severe that 97 % of all Smartphone malware now targets Android phones. In a year, approximately 3.25 million new malware Android applications are discovered as the growth of smartphones increases. This loosely amounts to a new malware android version being introduced every few seconds [9]. The primary aim of mobile malware is to gain entrance to user data saved on the computer and user information used in confidential financial activities, such as banking. Infected file extensions, files received via Bluetooth, links to infected code in SMS, and MMS application links are all ways that mobile malware can propagate [10]. There are some strategies for locating apps that need additional features. Hopefully, by using these techniques, it would be possible to determine whether the applications that were flagged as questionable and needed additional authorization are malicious.

Static analysis methodologies are the most fundamental of all approaches. Until operating programs, the permissions and source codes are examined [11]. For many machine learning tasks, such as enhancing predictive performance or simplifying complicated learning problems, ensemble learning is regarded as the most advanced method. It enhances a single model's prediction performance by training several models and combining their predictions. Boosting, bagging, and random forest are examples of common ensemble learning techniques [12]. In summary, the main contributions of our study are as follows:

1) We present a novel subset of features for static detection of Android malware, which consists of seven additional selected feature sets that are using around 56000 features from these categories. On a collection of more than 500k benign and malicious Android applications and the highest malware sample set than any state-of-the-art approach, we assess their stability. The results obtain a detection increase in accuracy to 96.24 % with 0.3% false-positives.

2) With the additional features, we have trained six classifier models or machine learning algorithms and also implemented a Boosting ensemble learning approach (AdaBoost) with a Decision Tree based on the binary classification to enhance our prediction rate.

3) Our model is trained on the latest and large time aware samples of malware collected within recent years including the latest Android API level than state-of-the-art approaches. This research paper incorporates binary vector mapping for classification by allocating 0 to malicious applications and 1 for non-harmful and for predictive analysis of each application fed to the model implemented in the study. The technique eases the process by reducing fault predictive errors. Figure 2 shows the procedure for a better understanding of the concept applied later in our study. The paper passes both the categories of applications through static analysis and then is further processed for feature extraction. We presented features in 0’s and 1’s after extraction. Matrix displays the extraction characteristics of each application used in the dataset. There are major issues to be addressed to incorporate our strategy. High measurements of the features will make it difficult to identify malware in many real-world Android applications. Certain features overlap with innocuous apps and malware [13]. In comparison, the vast number of features will cause high throughput computing. Therefore, we can learn from the features directly derived from Android apps, the most popular and significant features. The paper implements prediction models and various computer ensemble teaching strategies to boost and enhance accuracy to resolve this problem [14]. Feature selection is an essential step in all machine-based learning approaches. The optimum collection of features will not only help boost the outcomes of tests but will also help to reduce the compass of most machine-based learning algorithms [15]. Studies have extensively suggested three separate methods for identifying android malware: static, interactive meaning dynamically, and synthetic or hybrid. Static analysis techniques look at the code without ever running it, so they're a little sluggish if carried out manually and have to face a lot of false positives [16]. Data obfuscation and complex code loading are both significant pitfalls of the technique. That is why automated operation helps to achieve reliability, accuracy, and lesser time utilization [17]. Reverse engineer Android applications and extract features and do static analysis from them without having to execute them. This method entails examining the contents of two files: AndroidManifest.xml and classes.dex, and working on the file with the .apk extension. Feature selection techniques and classification algorithms are two crucial areas of feature- based types of fraudulent applications. Feature filtering methods are used to reduce the dimension size of a dataset. Any of the functions (attributes) that aren't helpful in the study are omitted from the data collection because of this. The remaining features are chosen by weighing the representational strength of all the dataset's features [18]. Parsing tools can help learn which permissions, packages or services an application offers by analyzing the AndroidManifest.xml file, such as permission android.permission.call phone, which allows an application to misuse calling abilities. The paper elaborates exactly what sort of sensitive API the authors could name by decoding the classes.dex file with the Jadx-gui disassembler [19]. In certain cases, including two permissions in a single app can signify the app's possible malicious attacks. For example, an application with RECEIVE SMS and WRITE SMS permissions can mask or interfere with receiving text messages [20] or applying sensitive API such as sendTextMessage() can also be harmful and lead to fraud and stealing. Until we started our main idea of the project. The fact explained that Android applications pose a lot of threats to its user because of the unnecessary programs compiled inside them and explained why it is necessary to automate the process of static analysis for the efficient detection of malware applications based on the extracted features. The rest of the paper is planned as follows. Related works are examined in Section II. Section III will present the design and method of our model. Section IV elaborates the assessment findings and future threats. The experiments and results will be dilated and performed in Sections V and VI. Section VII includes our research issues, recommendations, and conclusions for the future.