ANDROID MALWARE DETECTION

A.DURGA DEVI[1], KALIDINDI JYOTHI DURGA[2]

[1] **Assistant Professor MCA, DEPT,** Dantuluri Narayana Raju College, **Bhimavaram, Andhra Pradesh**
Email id:- adurgadevi760@gmail.com

[2]**PG Student of MSc Computer Science,** Dantuluri Narayana Raju College, **Bhimavaram, Andhra Pradesh**
Email id:- jyothidurgakalidindi@gmail.com

ABSTRACT

,

Android malware growth has been increasing dramatically along with increasing the diversity and complicity of their developing techniques. Machine learning techniques are the current methods to model patterns of static features and dynamic behaviorsof Android malware. Whereas the accuracy rates of the deep learning classifiers increase with increasing the quality of the features, we relate between the apps' featuresand the features that are needed to deliver its category's functionality. Differently, our classification approach defines legitimate static features for benign apps under a specific category as opposite to identifying malicious patterns. We utilize the features of thetop-rated apps in a specific category to train a malware detection classifier for thatgiven category. Android apps stores organize apps into different categories, for instance,26 categories on Google Play Store. Each category has its distinct functionalities which means the apps under a specific category are similar in their static and dynamic features.In general, benign apps under a certain category tend to share a common set of features.On the contrary, malicious apps tend to request abnormal features, less or more than what is common for the category that they belong to. This study proposes category-based machine learning classifiers to enhance the performance of classification models at detecting malicious apps under a certain category. The intensive deep learning experiments proved that category-based classifiers report a remarkable higher average performance compared to non-category based.

Keyword :- MALWARE DETECTION, THETOP-RATED APPS

1 INTRODUCTION

According to International Data Corporation (IDC), Android OS is the most popular smartphone platform with 82.2% of the market share of smartphones, while 13.9% for iOS apple in the second quarter of 2015 [3]. Statistically speaking, it is also the first targeted platform by malware authors seeking to take the control over millions of Android smartphones over the world. Due to the popularity of Android's smartphones, its apps' security is a serious issue concerning 80% of smartphones users.

Android is an open-source development environment that offers a rich SDK that enables developers to deploy their own apps and distribute them through Android apps centers. Android's popularity is a result of being an open source, third-party distribution centers, a rich SDK, and the popularity of Java as a programming language. Importantly, due to this open environment, malware authors can develop malicious apps that abuse the features that the platform offers or pack a legitimate app with a piece of malicious code; besides, exploiting vulnerabilities in the platform, hardware, or other installed apps to lunch malicious behaviors. Mainly, malware authors seek access confidential data of a device's user, monetary benefits via premium SMS, or joining the device to a botnet. Even legitimate apps introduce the risk of

privacy-invading; McAfee reported in Q1 2014 that 82% of Android apps track user's and 80% gather location data.

Research studies in the Android malware detection field work in three approaches static, dynamic or hybrid. In static analysis, malware is disassembled into a source code from where specific features are extracted. In dynamic analysis, malware is monitored at run-time in a virtual environment. In the both approaches, machine learning algorithms have been used to build classification models by training classifiers with datasets of malware features that collected from static or dynamic analysis. The learned classification models are then used to detect malicious apps and classify them into their families.

In this study, we approach the problem differently by utilizing the features of benign apps for malware detection. We relate between the features that the app requests and the common features for its category.

Android apps stores organize apps into different categories; for example, Google play store organizes apps in 26 categories such as:" Health & Fitness"," News & Magazine"," Books & References"," Music & Audio", etc. Each category has its distinct functionalities which means the apps under a certain category share similar features. One group of these features are the permissions; per- missions are the privileges that enable apps to access the system's resources to perform their functions. Each built-in permission is responsible for providing the capabilities to execute a particular process. Apps belong to a specific category deliver the same functionality as a result they require a common combination of permissions. For instance, apps under" Communication" category commonly request READ CONTACTS but it is un- common if it is requested by apps under" News & Magazines". In general, benign apps under a certain category tend to have a common set of features: permissions, intents filters, hardware components, broadcast receivers, APIs, etc. On the contrary, malicious apps tend to request abnormal features, less or more than what is common for the category that they belong to. Repeatedly from that point of view, this study proposes category-based machine learning classifiers to enhance the performance of classification models at detecting malicious apps under a certain category.

2. LITERATURE SURVEY AND RELATED WORK

The initial studies on smart phone malware were chiefly targeted on understanding the threats behaviors of rising malware. There has been vital work on the matter of police work malware on mobile devices. Many approaches monitor the facility usage of applications and report abnormal consumption. Others monitor system calls and arrange to discover uncommon system call patterns. Different approaches additional ancient comparison with acknowledged malware or different heuristics. Signatures primarily based ways, introduced within the mid-90s area unit ordinarily employed in malware detection. The main weakness of this kind of approach is its weakness in police work metamorphic and unseen malware. Rather than victimization predefined signatures for malware detection, data processing and machine learning techniques give a good thanks to dynamically extract malware patterns. For smart phone-based mobile computing platforms, recent years have witnessed an increasing range of additional sophisticated malware attacks like repackaging. Recent analysis

consistently characterizes existing mechanical man malware from varied aspects, together with their installation ways, activation mechanism moreover because the nature of carried malicious payloads. supported the analysis with four representative mobile security software packages over 1200 collected malware, their experiments show the weakness of current malware detection solutions and need the necessity to develop next-generation automobile-malware solutions. One existing work has used data processing and options generated from windows workable API calls. They achieved sensible leads to a really giant scale dataset with concerning 35,000 transportable workable files. Another activity foot printing methodology additionally provides a dynamic approach to discover self-propagating malware. All these existing ways have basically advanced the mechanical man malware detection; however, the misuse detection isn't reconciling to the novel mechanical man malware and continually needs frequent change of the signatures. Here lies the analysis gap.

In comparison, our work is motivated by a number of the higher than techniques and approaches, however with a spotlight on developing straightforward and effective malware detection approaches, while not looking forward to advanced dynamic runtime analysis and any static predefined malware signatures.

3 EXISTING SYSTEM

One existing work has used data processing and options generated from windows workable API calls. They achieved sensible leads to a really giant scale dataset with concerning 35,000 transportable workable files. Another activity foot printing methodology additionally provides a dynamic approach to discover self-propagating malware. All these existing ways have basically advanced the mechanical man malware detection; however, the misuse detection isn't reconciling to the novel mechanical man malware and continually needs frequent change of the signatures. Here lies the analysis gap. In exiting system, they implemented the classifiers like naive bayes and decision tree which gives the poor accuracy

4 PROPOSED WORK

In the proposed system we implement a better feature extraction technique and then we apply CNN Algorithm for feature extraction and classify android malware detection which gives the better accuracy ratio when compare to existing system

Advantages

Easy to identify and block malware

Accuracy is more

Dynamic feature extraction using genetic algorithm

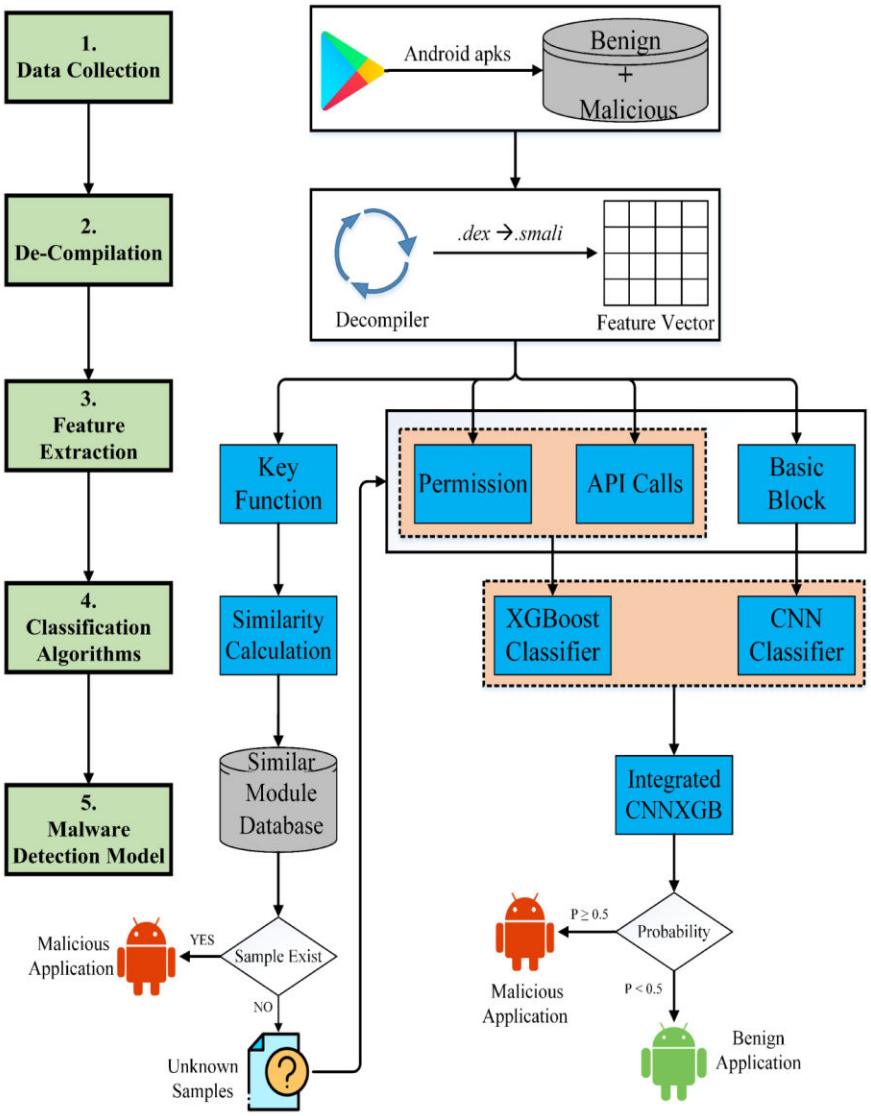

Fig 1: System architecture

5 METHODOLOGIES

6 MODULES

(DATA COLLECTION AND FEATURE FILTERING)

1. Collect all applications in separate folders which contain benign as well as suspicious applications respectively.

2. Using "Glob" framework in python create an array of files is for further processing.

   Analyze each application in the array using "pyaxmlparser" and "androguard" framework.

3. Extract the following things in the analysis phase:

   a. Permissions

   b. Activities

c. Intents

d. API calls

4. Taking these four attributes into consideration a program maps all attributes to a CSV file and mentions a class for each application.

5.Once CSV files are generated, analyze themfor any redundancy present, and if found, eliminate the entire row.

6. Another program extracts the total permissions from these APK files. These permissions will work as attributes in the Dataset CSV File (Here if permission is present, it is marked as 1else it is marked as 0).

An N-bit Vector extracts search line in theCSV file, these vectors work as input to the machine learning algorithm.

7 RESULTS AND DISCUSSION SCREEN SHOTS

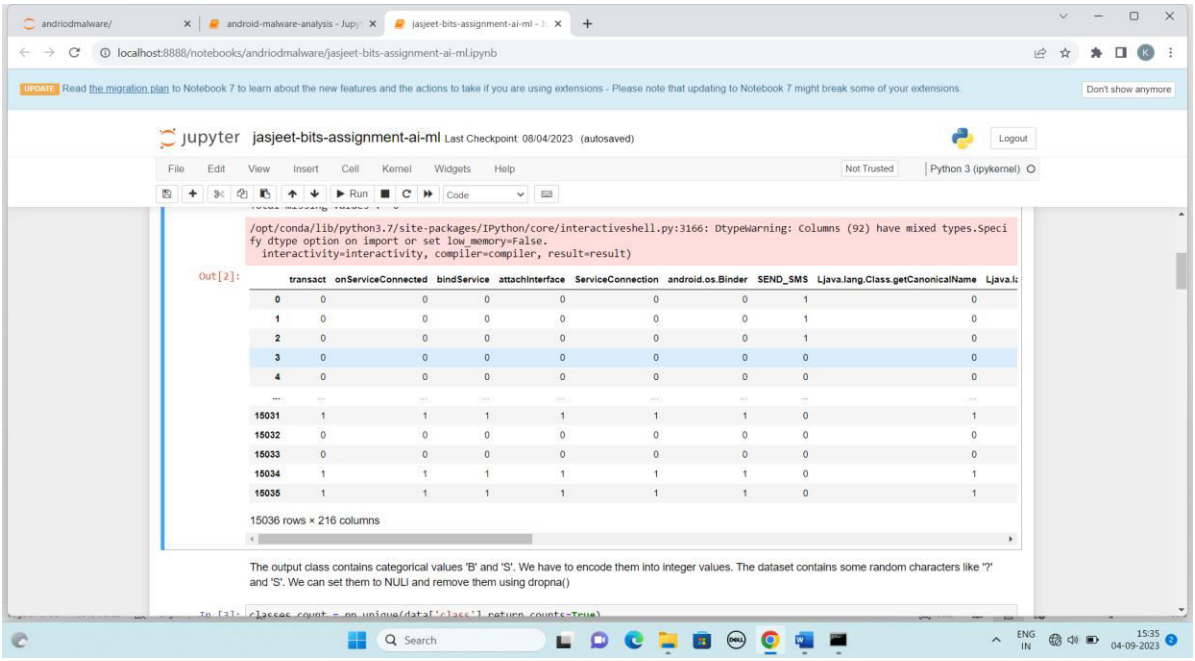

Fig 2: - sample dataset

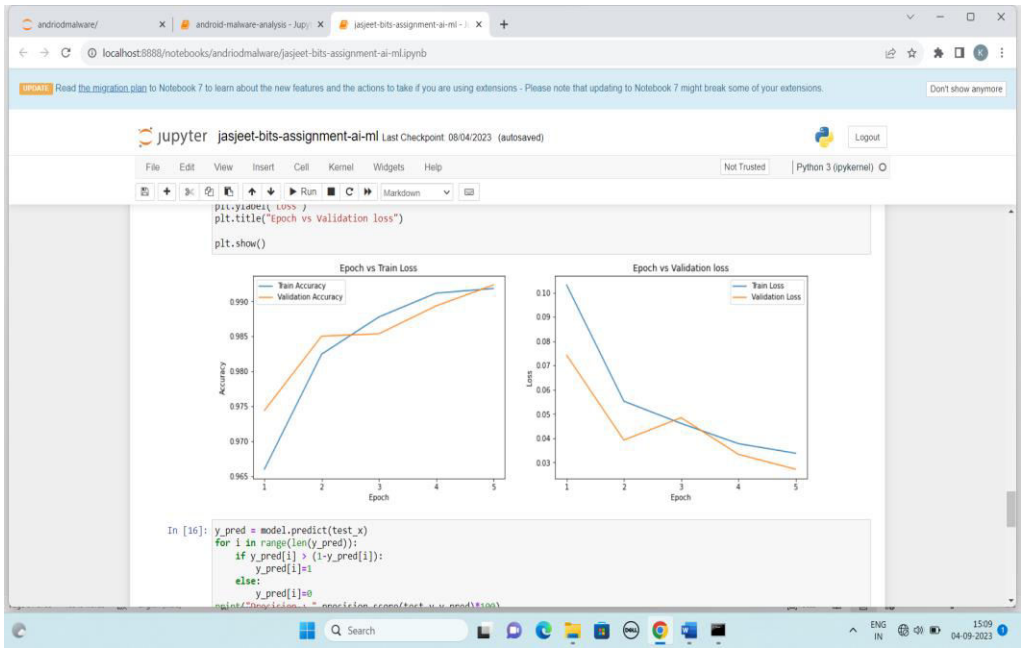

Fig 3: - training and Testing Graph

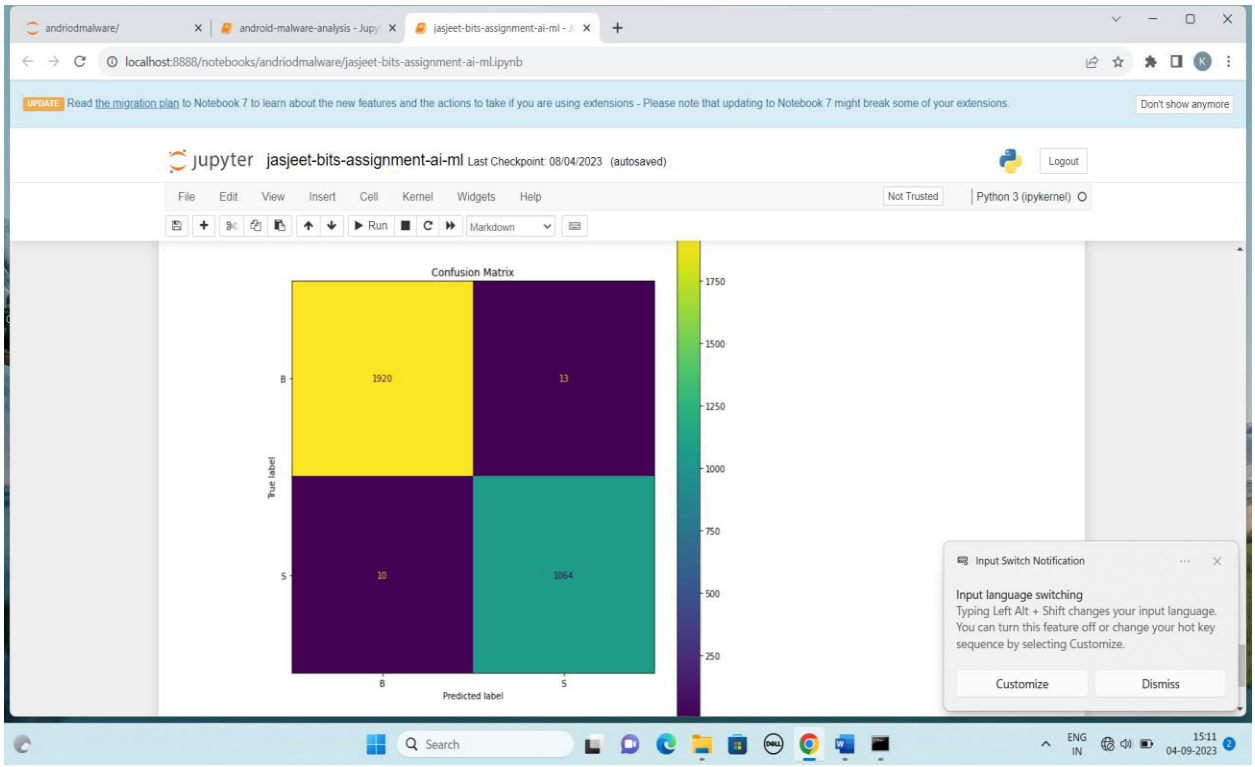

Fig 4: - Confusion matrix

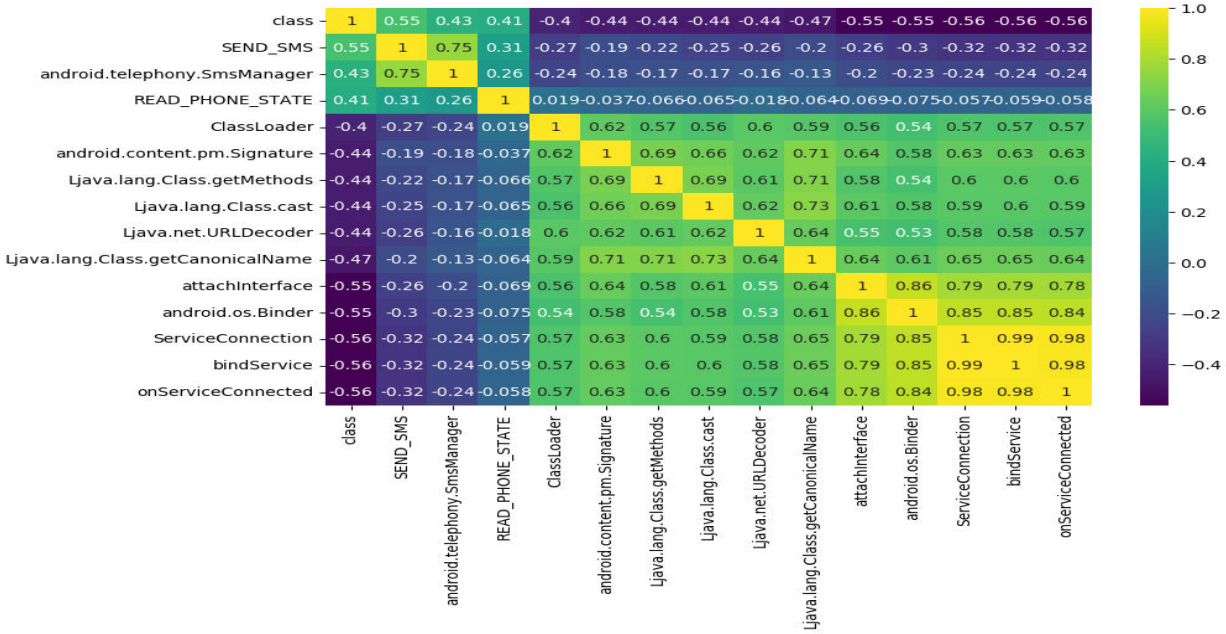

Fig 5 -CORRELATION MATRIX OF FETURES

6.CONCLUSION AND FUTURE SCOPE

In our study, we propose category-based deep learning classifiers to improve the performance of the classification models. In static analysis of Android malware, machine learning algorithms have been used to train classifiers with features of malicious apps to build models that capable of detecting malicious patterns. Differently, our classification approach defines legitimate static features for benign apps as opposite to identifying malicious patterns. We utilize the features of the top-rated apps in a specific categoryto define a profile of the common sets of features for that category. In other words, to detect whether or not the app possesses the characteristics of benign, we relate between the app's features and the features that are needed to deliver the category's functionality that the app belongs to. Android stores organize apps into different categories; 26 categories on the Google Play Store, for example. In each category, the apps

delivera similar functionality as a result the they tend to request a common set of featureslike same permissions, APIs, hardware components, broadcast receivers, intents filters, etc. On the contrary, malicious apps tend to have abnormal features, less or more than what is common for the category that they belong to. Malicious apps can be identified by comparing between the features they request to the features that are requested by benign apps in the same category. For example, malicious apps, compared to the benign apps in the same category, tend to request over-privileged permissions, listen to specificevents that broadcast by the Android system, or using unneeded APIs for the app's category functionality that can be used to lunch malicious behaviors.

Our future work will consider three aspects. First, including other static features such as: functions call in building the classification models to get a better understanding of the processes that apps may lunch in a way to increase the detection accuracy of the classifiers. Second, implementing the proposed solution on a large-scale level by building profile models for other categories and sub categories. Third, testing the feasibility of integrating our solution with dynamic detection techniques by profiling dynamic features for each category; dynamic features like system calls, network connections, resources' usage, and etc.

7 REFERENCES

1. Androguard-usagehttps://code.google.com/p/androguard/wiki/Usage.Ac-cussed April 24, 2015.
2. Android-statistics&facts—Statista. http://www.statista.com/topics/876/ android/. Accessed April 19, 2015.
3. Android and iOS continue to dominate the worldwide smartphone market with an- droid shipmentsjustshyof800millionin2013,accordingtodice.http://www.idc. com/getdoc.jsp?containerId=prUS24676414.Accessed April 19, 2015.
4. Applicationfundamentals—androiddevelopers. http://developer.android.com/guide/components/fundamentals.html. Accessed April 19, 2015.
5 Are—download/installationhttps://redmine.honeynet.org/projects/are/wiki. Accessed April 28, 2015.
6 Dynamicanalysistoolsforandroidfailtodetectmalwarewithheuristicevasiontechniques.http://thehackernews.com/2014/05/dynamic-analysis-tools-for-android-fail.html. Accessed April 19, 2015.
7 " Global smartphone sales exceed 1.2b units in 2014." gfk - we see the big picture. http://www.gfk.com/news-and-events/press-room/press-releases/pages/global-smartphone-sales-exceed-1-2b-units-in-2014.aspx. Accessed April 19, 2015.
8 Google:Wehavebillionmonthlyactiveandroidusershttp://www.businessinsider.com/google-we-have-1-billion-monthly-active-android-users-2014-6. Accessed April 19, 2015.
9 Report: 97- Forbes.http://www.forbes.com/sites/gordonkelly/2014/03/24/

report-97-of-mobile-malware-is-on-android-this-is-the-easy-way-you-stay-safe/. Accessed April 19, 2015.

10 Smartphoneosmarketshare,q42014http://www.idc.com/prodserv/smartphone-os-market-share's. Accessed April 19, 2015.

11 Arp, D., Spreitzenbarth, M., Hu¨bner, M., Gascon, H., Rieck, K., and Siemens, C. (2014). Drebin: Effective and explainable detection of android malware in your pocket. In *Proc. Of* NDSS

12 Aung, Z. and Zaw, W. (2013). Permission-based android malware detection. *Inter- national Journal of Scientific and Technology Research*, 2(3):228–234.