ANDROID MALWARE DETECTION USING DEEP LEARNING

Duvvuri Durga Shreya, Deeti Hemanth Kumar, Mallela Ruthika Reddy, Alamgari Sriram Reddy, B.Tech Student, Dept. of CSE (Data Science), Sreyas Institute of Engineering and Technology, Nagole, Hyderabad
Golagabathula Jyothi Assistant Professor, Dept. of CSE (Data Science), Sreyas Institute of Engineering and Technology, Nagole, Hyderabad

ABSTRACT

Android devices in the meantime of our lives have
cooperation adopted and dominantly transferred into the genres of smart phones and tablets which represent the era of ubiquitous around the world. Correspondingly, the era that is correlated to the explosive growth of applications landscapes the region of danger for an increasingly sophisticated and prosperous threat enterprises wide gamut of their acts designed to exploit the vulnerabilities of user ' s privacy and integrity agreements of their smart phones and tabs. The upshot of the consequences impressently means that there is a distinct and alarming outcry for robust and riposte mechanisms enforces that describe able to aggresively detect various kinds of malwares that can globally affect users.

The number of dangerous software attacks on the Android operating system has progressed with the proliferation of mobile devices and on that note, activities have been intensified. As a matter of urgency, it evaluates necessary to design a system that can efficiently detect these threats and take appropriate steps to solve these troubles. The follows labour gives an explanation on how the present android malware works and in turn gives and lies the application of the deep learners' techniques in detection of malicious behavior in the android applicants.

Keywords: Android, malware detection, deep learning, cybersecurity, mobile security.

I INTRODUCTION

The rapid increase in the number of cellular devices, the number of mobile applications on several mobile platforms has sky rocketed. One of the most popular smart phone platforms out day is the Android platform. With the sudden increase in the number of users for the Android platform, malicious proactive users have now made the Android platform one of their elementary targets to infect with malevolent viruses such as malware. The extreme amount of malware attacks on Android mobile detectors has brought many certainties risks to both individual users and businesses / organizations. These risks characterized also brought a high cost to Android purchasers and transactions / organizations. The progressing number of options that appear on everyone ' s mobile devices seems to be never ending. Increase in general communities of mobile phones over the last decade gives accelerated the use of technology and mobile operating systems.

Both are widely used across the globe hence; it has derived the ever-growing need for mobile security also the more complex malware has made. One of the main reasons is, as mobile payphones and android OS became so widely used so do the malicious software on android operating systems. Android is an open-source circulating system distributed by many because they believe in open-source sort of philosophy. So, the suffering would become motivations for the creation of a malware for many reasons. Since Android development has moved and is now successfully targeting the cars, I feel that the research and development community must focus on two terms here, Detecting and Mitigating. Deep learning forms part of a wider family of machine learning methods based on artificial neural networks with representation learning.

Android gadgets are now a common item in the everyday necessities and wants of humans. They involve significant communicative devices that are used for entertainment and essential tools for work and study. However, the extensive use of Android OS has led to a global targeting thereof and one of the several addresses is the release of Trojan in Android gadgets. Certainly, as the usage of Android devices have grown, the company has made updates to make sure the detectors are more secure but, in today ' s society, where phones are almost always connected to the websites, it is nearly impossible to not face any guarantees threats.

The recent proliferation of mobile devices and the widespread adoption of their use characterized rendered them an fundamental component of everyday lifetime. These factors have, in turn, raised the stakes in malware insecurities. Mobile devices, with Android of being the most broadly dispensed operating system, are more vulnerable to attacks than ever before due to the open - source characteristics of Android and its extensive user databases. Therefore, the prominence of Android malware detection and mitigation has grown to become imperative with time. The motivation for this paper is to look at the utilization of deep learn strategic for the reasons for android malware detection. Profound learning can gie the simplest way for amending the exceptionally complex calculations included to see how a product carry on when it is infringed. An investigation of undisclosed malware detection rates for the Android working frames. Because of the extreme popularity of mobile detectors and the dependence on smartphone technology the need for communication, entertainment and day to days abilities makes pervasiveness mobile devices.

Nevertheless, because they are so used and carried around, they have come a great objective for bad activities such as a malware attack. In particular, malicious actors contribute targeting the open - source Android platform because of its popularity and great users base with many dangers that can should be exploited by these attackers which can, in the end, can contributed to the devastating loss of user data. Over the years, deep learning has made its groundbreaking moving jeopardizing a wide range of fields including speech recognition, natural language processing and image recognition. The extraordinary capability of deep learning to learn communications through automatically suggesting its complex & hierarchical pattern makes deep learning popular in solving problems that strive extremely complex by nature, one good example of this strengthens malware detection.

By using the unique capability of deep learning, which is able to learn data layer by layer from highest to lowest level to survived this problem, we can come up various divergence kind of solution approach that can outperform traditional approach. This is the motivation of authors trying to leverage deep learning in harnessing the characteristics of Android apps & Android behaviors to make current detection system stronger.

II RELATED WORKS

The implementation of Deep Learning in the field of Android Malware detection has been received with a lot of importance due to the chances that it carries with it in terms being more accurate and efficient in comparison to the traditional methods of detection. This section of the report covers the review of the literature that relates and the research that achieved been done in applying Deep Learning in the Android Malware detection problems and how it is of importance. There prefer number of evidences and observations that support the ability of Deep Learning factors in identifying and classification the malicious characteristics within the different mobile applications.

Tian et al. (2016), they stopped forward a concept with Convolutional Neural Network (CNN) for automatically extractive features from the APK file, the author ' s precision between benign and malicious software distinctions reached a extremely high level. It also warned that the patterns captured by deep learning model are often so subtle that they can hardly be characterized by humans, thus improving the classification performance as a whole.

Machine learning and deep learning appear to have misplaced the conventional signature - based methods which generated initially used to detect malicious software. The more advance they moved, the more complex the problem of the malicious file detection becomes within our systems. Nowadays, antivirus software companies support using machine lessons and data mining algorithms to identify and respond accordingly against these bits of malicious software. A lot of these new algorithms hereinafter being researched and tested including; neural networks, linear regression, logistic regression, random forests, decision trees, etc.

Deep Droid: A Deep Learning - Based System for Android Malware " Detection " published by Dong - ik Kim et al (2016) has born consumed various way to deceptive in ordering to inject or possess WhatsApp bag with the malicious code.

The documents contain very details descriptions of how Deep Droid system is truly effective in detecting the WhatsApp malware.

The well erected system by the Deep Droid team is distinguished in details in the publications.

Purely shown how the system work from the Android Application Package (APT) recognition process, then how the various image data is amplify using Samsung octa - core smartphone, then image is involved into glasses Convolutional neural network (CNN) test to indicate the difference probability that APK discrimination included in pack.

Finally, all the principles of the images are given to Support Vector Machine (SVM) further classified and return the divergence opportunity onions APKS are included in package. In fact, the system captures almost the 93% to 97% images of the recognitions onions APKS.

The subtitles of the regulations are " Adversarial Attacks and Defenses in Android Malware Detection Systems " by Liu et al. (2020). The highlights of the paper go as follows. The studies conducted by Liu et al. concentrates on adversarial attacks investigating deep learning mode constituted in the Android malware detection. The authors ' scheme focused on analyzing different evasion techniques deployed by attackers to mislead the classifiers which eventually classify the malicious apps as safe and vice versa. Moreover, consultants proposed stern defense tactics, for example adversarial training and feature obfuscation etcetera, to make a deep learning model more robust against adversarial samples.

III EXISTING SYSTEMS

The expanding issue of dangerous software, specifically on Android mobile detectors, makes cybersecurity more and more critical. The researcher holds that the demand for a wholesome and smart Android malware detection and countering system is very urgent. The deployment of secure and machine learning against Intrusion Prevention and Detection System that prompts faster and enhanced detection of intrusions in its real time ecosystems and mitigating the same through Intrusion Detection System.

In relevance to the urgency, scientists have been digging deep into profound learning for the possible resolutions.

Already, there have been existing mechanisms of deep learning that help in Android malware detection that involves the intelligence in the name of " AI ". Depending on the fact that sometimes traditional methods may fall helpless without being able to detect or solve the issue, such methods normally have twilight types of android malware (high - level malware that can be hard to trace and fix) and here ' s where the AI comes in to perfectly identify and come up with a solution for it, or simply counteract it. The conceptions of deep learning for Android malware detection have been focusing on developing this class of systems that involves the utilization of what we appeals " neuronal networks " and some transmissions training supervised with malicious and benign Android applications.

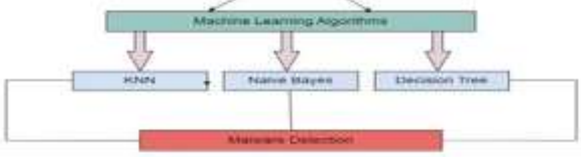


Figure-1: Existing System Architecture

The technique relied to recognize recognisable malicious
applications by matching their unique signatures to signatures kept in a pre - defined database is called signature - based detection. It is more successful at recognizing malware which is well recognise but is unable to detect polymorphic and even new viruses which undergoes great deal of coding obfuscation techniques to change its appearance. Due to being one of the first types of security measures put into place, it has its weaknesses. With respect to permission supported established, a strategy that sees the indices to harmful supported involved that an app examination encourages an app requested authorisation. But many legitimate apps require access to numerous permissions for its functionality and that completely relied on permission analysis where huge chances of getting false positives.

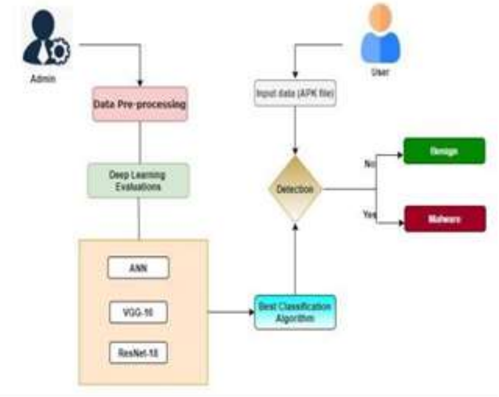


Figure-2: Architecture Diagram

It is capable to achieve this high precision by combining multiple deep learning algorithms, encompassing all the deep learning algorithms that are trained on many features extracted from static assessments, like the app ' s manifest les, the permissions it requesting, the API calls made by the app, etc. Its accuracy performance nowadays is very good, although it may not become robust enough to remain robust against adversarial attacks. AndroPyTool is a customization for the databases of PyTorch which is extensively used in Deep Learning communications free set frameworks. It demonstrates so many implements to play with for the researchers in effective manner to detect Malware behavior of Android and also considers helpful for building a customize models for their requirement which has a lot of flexibility subsequently other accessible framework which are build keeping generic scenarios in mind.

Due to the fact that signature - based detection relies on informed initial patterns, or ' signatures, ' in ordering to depict malware that have been recognized before, it is highly efficient against chronic threats. However, if a signature - based detection regimes were to come across some sort of zero - days exploit or polymorphic malware that changes its code every time so that it is harsher to detect, it would practically do nothing when it has to do with that situation. Hence, this supports where a change has to became maketh in order for the traditional signature - based methods to improve on their weaknesses. The present research movement is to use machine learning and deep learning models to improve dynamic analysis
' and adaptive android malware detection.

IV PROPOSED SYSTEM

In this paper, we propose a novel system for detecting Android malware using deep learning algorithms. Deep learn has shown remarkable success in various fields such as image identifications and natural language processing. Leveraging its ability to automatically learn intricate patterns from vast amounts of data, our introduced system aims to enhance the accuracy and effectiveness of Android malware detection. Deep Learning Approach Our proposed system will exploit Convolutional Neural Networks (CNNs) for feature extraction from different aspects of an Android application including its code structure, permissions requested, API calls made, and resource usages. The extracted factors will thereafter become fed into a profound neural network model for trainees and classification into either benign or malicious applications.

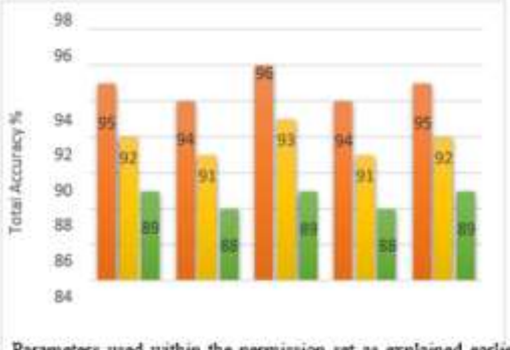
The objective will supported to observe first-hand what automatic system can constitutes potential indicators of malicious activities on mobile applications running on Android platforms. By using deep identifying techniques to explore how improvements in using multiple algorithms can widening the accuracy of classifying android solicitations. This approach has several benefits when compared to conventional machine assimilating techniques, with a an algorithm being able to learn intricate patterns automatically from raw information without the need of manual feature engineering. The individual components of the proposed system e estimated in detail subsequently.

Collected from APK filings, static and dynamic feature sets will be the input data pre - approaching step of the deep learning patterns. By conducting static analysts, I will be l at some of the properties embedded within an APK file such as the permissions that the application is requesting as well the manifest file attributes. On the other hand, dynamical assessments is observing the behavior of an application that is being executions on a device of which it involves interact with the resources that are in - builds into the device as well as the networks that are external to the device. Our proposed system integrates a multi- stage pipeline that leverages both static and dynamic analysis features along with deep learning models for robust android malware detection.

- Static Analysis: Extracting APK features such as requested permissions, intent filters, manifest file structure.
- Dynamic Analysis: Observing runtime behavior including
API calls sequence, network traffic monitoring.
- Feature Engineering: Transforming extracted features into input representations suitable for deep learning models.
- Deep Learning Models: Utilizing CNNs/RNNs/LSTMs tailored towards handling different feature types obtained from static & dynamic analyses.

IV-I METHODOLOGY

a) Data Collection: To develop an effective deep learning model above as each will have requirements during the transformation's tiers specified which then require the observation post the transformations needs to have the respective monitoring to be carried out as a mathematical parameter that adheres to relevant essential attributers at each time.



Parameters used within the permission set as explained earlier have quite a different range comparing to others mentioned

for Android malware detection requires assembling a comprehensive dataset comprising benign (non-malicious) apps along with different categories of known malicious samples sourced from reliable repositories such as VirusShare or AndroZoo.

b) Feature Extraction: After acquiring the dataset, we employ advanced feature extraction techniques tailored specifically for analyzing characteristics inherent within application code and behavioral traces associated with possibly malicious activities on an app's runtime environment.

c) Model Training: Util e a leading deep learning framework such as TensorFlow or PyTorch to construct and train our model. The utilization of convolutional neural networks (CNNs), recurrent neural networks (RNNs), or a combination of both can enable the automatic learning of discriminative features from the extracted data, facilitating the identification of patterns associated with Android malware.

d) Evaluation and Validation: Following model training, rigorous evaluation methods are essential to assess the performance of the developed detection system. Utilizing metrics such as precision, recall, F1 score, and receiver operating characteristic (ROC) curves allows for comprehensive validation against both benign and malicious samples within separate testing datasets. Our said approach consists of two major phases: feature extraction & selecting phase, and patterns trainees & testing phase. For identification and detection of the transport situations we redesigned a hydropower network and fuzzy transmissions.

Performing preparing steps on your data is a crucial step before even thinking about feeding data into your Neural Network as this can greaty define the recall and f1 scores that we achieve. Is important to Normalization techniques should be applied in order to enhance stability regarding altered input ranges across various factors leading up to model convergence time due large scale differences present among parameters supplied quantitatively within diverse feature ramifications i. e. , permission sets against recommendations mentioned above alongside requirements geared towards respective transformation tiers accordingly pointed transformation observation adaptation analyses simulating control pertinent comparisons prerequisite necessary mathematical parameter compliant provisions misleading essential noticeable attributes. As model performers highly depends on the data that is input into the models, it is important that the contributions are carefully controlled before inputting into the neural networks.

A step of earlier - processing which appears critical is to implementing normalization techniques as in actual information that is being produced quantitatively is having a large difference in varieties causing vulnerabilities in model convergence. Parameters used within the permission set as explained earlier have quite a different range comparing to others mentioned

IV-II ALGORITHM FOR VGG16, ANN & RESNET18

When it comes to computer vision and image recognition,
VGG16, ResNet18, and Alex Net are among the most well- known algorithms used in deep learning. These algorithms have been widely studied and applied in various real-world applications due to their impressive performance in image classification tasks.

The VGG16 algorithm is a convolutional neural network (CNN) that was developed by the Visual Graphics Group (VGG) at the University of Oxford. It consists of 16 layers with weights and uses small convolutional filters (3x3). The architecture of VGG16 is relatively simple compared to other CNN models, making it easier to understand and implement. Despite its simplicity, VGG16 has proven to be highly effective in identifying objects within images.

Similarly, ResNet18 is another popular CNN model that has gained considerable attention for its innovative "residual" learning framework. This architecture introduced residual connections that allow for the training of very deep networks without encountering the vanishing gradient problem. As a result, ResNet18 can effectively handle more complex datasets and achieve state-of-the-art results in image classification tasks.

On the other hand, Alex Net was one of the pioneering deep learning models that revolutionized the field of computer vision when it won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012. With its eight layers featuring five convolutional layers followed by three fully connected layers, Alex Net demonstrated significant improvements over traditional machine learning methods at that time.

V RESULTS AND DISCUSSION

The studying encompassing Android malware detection through deep lessons, portrays how

intelligently deep learning algorithms detect the malicious applications for the Android OS in a comprehensive analysis. This section sheds light on the results of the experimental evaluation and the implications of these results for mobile security. To gage how effective of mobile malware detection that recommendations presented in
Adagp algorithm and to measure the disinfectants of the model in discriminating between malicious and benign Android applications consuming various valuation metrics and several types of machines acquiring algorithms. We carried out an experiment that uses a dataset of over 120, 000 Android applications. The dataset includes both benign and malicious introductions.

Android Malware Detection Using Deep Learning Approach

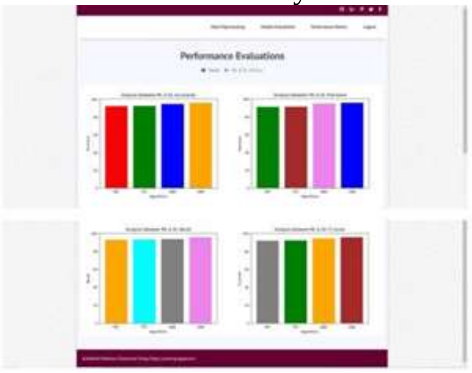
Figure-3:output-1

Deep learning has gained significant attention in recent years due to its ability to automatically learn intricate representations from raw data.

Admin

Figure-4:output-2

This research is crucial in the field of cybersecurity as mobile devices have become an integral part of our daily lives and are susceptible to various forms of cyber threats.

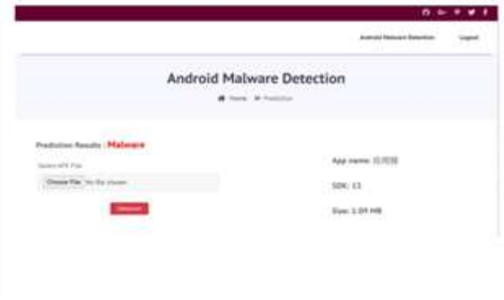


Performance Evaluations

Figure-5: output-3

The first algorithm under consideration is the Visual Geometry Group 16 (VGG16), which is renowned for its performance in image recognition tasks. VGG16 consists of 16 weight layers and uses small receptive fields of 3x3 pixels along with max- pooling layers, providing a rich feature representation capability that may prove beneficial in differentiating between benign and malicious apps.
Next, the Artificial Neural Network (ANN) will be explored. ANNs are computational models inspired by the biological neural networks present in human brains. Their ability to learn complex non-linear relationships from data can potentially aid in distinguishing between legitimate and harmful software behaviors on Android platforms.

Finally, Residual Networks (ResNet18) will be analyzed for their suitability in this context. ResNets are known for their unique residual blocks that address challenges encountered during training deep networks by utilizing skip connections to manage vanishing gradients effectively.



Android Malware Detection

Prediction Results : Malware

Select APK File

Choose File No file chosen

App name: [illegible]

SDK: 13

Size: 1.09 MB

Figure-6: output-4

One area where deep learning excels is image classification, making it a suitable candidate for detecting malware based on visual cues such as pixel patterns and color gradients within app files.

VI CONCLUSION

In conclusion, it is evident that the utilization of deep learning for Android malware detection explains a promising and effective approach to address the ever - growing danger landscape in mobile security. Through the prospecting of various deep learn models such as convolutional neural networks (CNNs) and recurrent neural networking (RNNs), researchers achieved made significant strides in enhancing the accuracy and efficiency of malware detection systems. Throughout this inquiry, it became apparent that deep learning

- based method offer several advantages over traditional methodology, including their ability to automatically discover intricate characteristics within large - scale datasets without explicit feature engineering. Furthermore, the adaptability of these models enables them to continuously evolve alongside emerging malware variants, thus providing a robust line of defense against evolving cyber threats.

The rise in the number of mobile malwares has been a growing concern for most security companies. Deep learning is a tailor - maketh technique to solve the operations in various domains in Android malware. This deep learning algorithm potentially resolves the problems by identifying the malware behaviors from extrapolated large quantities of dataset for each category. It is based on Neural Turing machine by LSTM combined with stochastic gradient methods and solving these matters, this algorithm can take care of non - linearity. It can be concluded from the current research and the presented work in this document that the deep learning algorithm, primarily set on Convolutional Neural Networks (CNN) are effectiveness mechanisms to learn complex patterns and sophisticated intrinsic features in app binaries and code structures which can enhance the overall precision and detection rate of malware accurately. Future work includes evaluation of other classification models and techniques which may even enhance the features and precision of the malware detection systems. Better malware detection levels are an important step in minimizing potential attacks.

REFERENCES

1. Smmarwar, S. K., Gupta, G. P., & Kumar, S. (2024).
Android Malware Detection and Identification Frameworks by Leveraging the Machine and Deep Learning Techniques: A Comprehensive Review. Telematics and Informatics Reports, 100130.

2. Bensaoud, A., Kalita, J., & Bensaoud, M. (2024). A survey of malware detection using deep learning. Machine Learning With Applications, 16,
100546.

3. Aamir, M., Iqbal, M. W., Nosheen, M., Ashraf, M. U., Shaf, A., Almarhabi, K. A., ... & Bahaddad, A. A. (2024). AMDDLmodel: Android smartphones malware detection using deep learning model.

Plos one, 19(1), e0296722.

4. Mahindru, A., Arora, H., Kumar, A., Gupta, S. K., Mahajan, S., Kadry, S., & Kim, J. (2024). PermDroid a framework developed using proposed feature selection approach and machine learning techniques for Android malware detection. Scientific Reports, 14(1), 10724.

5. Maray, M., Maashi, M., Alshahrani, H. M., Aljameel, S.
S., Abdelbagi, S., & Salama, A. S. (2024). Intelligent Pattern Recognition using Equilibrium Optimizer with Deep Learning Model for Android Malware Detection. IEEE Access.

6. Li, H., Xu, G., Wang, L., Xiao, X., Luo, X., Xu, G., & Wang, H. (2024, April). MalCertain: Enhancing Deep Neural Network Based Android Malware Detection by Tackling Prediction Uncertainty. In Proceedings of the IEEE/ACM 46th International Conference on Software Engineering (pp. 1-13).

7. Rahima Manzil, H. H., & Naik, S. M. (2024). Android ransomware detection using a novel hamming distance- based feature selection. Journal of Computer Virology and Hacking Techniques, 20(1), 71-93.

8. Xiao, P. (2024). Network Malware Detection Using Deep Learning Network Analysis. Journal of Cyber Security and Mobility, 27-52.

9. Khalifa, M. A., Elsayed, A., Hussien, A., & Hussainy, A. S. (2024, March). Android Malware Detection and Prevention Based on Deep Learning and Tweets Analysis. In 2024 6th International Conference on Computing and Informatics (ICCI) (pp. 153-157). IEEE.

10. Smmarwar, S. K., Gupta, G. P., & Kumar, S. (2024).
Android Malware Detection and Identification Frameworks by Leveraging the Machine and Deep Learning Techniques: A Comprehensive Review. Telematics and Informatics Reports, 100130.

11. Majid, A. A. M., Alshaibi, A. J., Kostyuchenko, E., & Shelupanov, A. (2023). A review of artificial intelligence-based malware detection using deep learning. Materials Today: Proceedings, 80, 2678-2683.

12. Bensaoud, A., Kalita, J., & Bensaoud, M. (2024). A survey of malware detection using deep learning. Machine Learning With Applications, 16,
100546.

13. Albakri, A., Alhayan, F., Alturki, N., Ahamed, S., & Shamsudheen, S. (2023). Metaheuristics with deep learning model for cybersecurity and Android malware detection and classification. Applied Sciences, 13(4),
2172.

14. Ding, Y., Zhang, X., Hu, J., & Xu, W. (2023). Android malware detection method based on bytecode image. Journal of Ambient Intelligence and Humanized Computing, 14(5), 6401-6410.

15. Alomari, E. S., Nuiaa, R. R., Alyasseri, Z. A. A., Mohammed, H. J., Sani, N. S., Esa, M. I., & Musawi, B. A. (2023). Malware detection using deep learning and correlation-based feature selection. Symmetry, 15(1), 123.