ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 02 (10), October 2023, pp. 36-48

Article Title: Advanced Malware detection with Inception V3 for Enhanced Computer Security

Advanced Malware detection with Inception V3 for Enhanced Computer Security

R. Arun Kumar, J. B. Shriram

[1]Department of Electrical and Electronics Engineering, V.S.B. Engineering College, Karur, Tamilnadu, India.

[2]Department of Information Technology, University College of Engineering, BIT Campus, Anna University, Tiruchirappalli, Tamilnadu, India.

ABSTRACT

In the domain of computer security, identifying and categorizing malware is of utmost importance. Malicious software frequently employs sophisticated methods to avoid detection, underscoring the urgency of early identification to protect computer networks and the Internet from extensive harm. This study tackles the obstacles related to malware detection and presents an innovative solution leveraging Convolutional Neural Network (CNN) classification through Inception V3.The input dataset is pre-processed to make the procedure easier. Initial stage of pre-processing is performed on dataset followed by data visualization. These data visualizations help identify the abnormalities and their existence or absence, as well as the structure of data. The categorization process is performed by Inception V3, a CNN-based classification network. The Inception V3 is compared with the most accurate models to assess the effectiveness of proposed system. The implementation of the proposed approach is carried out using the Python programming language. The outcomes of this research showcase the proficiency of the developed system, highlighting its potential in enhancing the accuracy and efficiency of malware detection.

Keywords: Convolutional neural networks, Malimg dataset, machine learning, malware detection, InceptionV3.

1 Introduction

Malware is software that is specifically created to inflict harm on a network and cause it to malfunction in various ways, as data volumes rise, so do the amount of harmful attacks that seriously endanger the security of both consumers and commercial entities [1]. Malicious software producers typically make minor changes to the malware application's original source code to create new malicious software variations in order to avoid detection by antivirus software [2]. The primary reason for the rise in malware sample counts is the widespread adoption of obfuscation techniques by malevolent software authors to create new versions of their dangerous software a method for classifying malware is needed to manage malware variations that belong to the same family in order to combat the proliferation of malware, to tackle it by developing deep learning and machine learning architectures for malware categorization [3]. Traditional malware often consists of a single process and doesn't employ intricate hiding mechanisms. However, modern malware makes use of several distinct

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 02 (10), October 2023, pp. 36-48

Article Title: Advanced Malware detection with Inception V3 for Enhanced Computer Security

concurrently with new or current procedures, and makes use of some hidden methods to blend in and become enduring in the framework [4]. More devastating attacks, like persistent and targeted ones that have never been seen before, can be launched by new generation malware, and these attacks combine multiple virus types [5]. The traditional machine learning approach has a number of drawbacks, including poor integration and transfer learning capabilities as well as a high feature engineering time and resource requirement. Such algorithms as Random false positive rate for Forest and Naïve Bayes is typically high on the training data that is inappropriate for classifying malware [6]. While SVM and decision trees perform admirably in known malware assault scenarios, they are unable to identify critical patterns for malware identification. In recent years, deep convolutional neural networks have emerged as the most effective method for classification problems [7]. It performs better on tasks like the Image Net categorization Challenge, which explains deeper models yield greater performance [8]. Technologies like LR, ANN, CNN, transfer learning on CNN, and LSTM are used in machine learning and deep learning research. Deep learning has demonstrated strong performance recently in a variety of fields, including speech recognition and computer vision [9]. The field of cyber security can also benefit from deep learning. As CNNs have demonstrated their superiority over time in solving image-related issues, initially the in order for CNNs to perform picture categorization, the task simply deal with it [10]. To use Inception V3 to classify and identify malware families, enabling a performance matrix to improve accuracy and overall performance, by using pre-processing techniques to improve quality. The list below demonstrates how the paper is organised: Section I gives a description of the introduction. The recent works are described in Section II. The proposed model and description are explained in Part III. The results are presented in Section IV. In Section V, the conclusion is presented.

2 Recent Works

S. Abijah Roseline *et al* [2020] [1] have recommended system performing deep learning techniques by using a layered ensemble approach that mimics the essential elements of those method in which the system operates with less model complexity and doesn't require hyperparameter adjustment or backpropagation. For the Malimg, BIG 2015, and MaleVis malware datasets, respectively, the model surpassed existing cutting-edge methods with a detection rate of 98.65%, 97.2%, and 97.43%. The outcomes show that, because to its variety of properties, the approach was successful in recognizing novel and sophisticated malware.

Xichen Zhang *et al* [2020] [2] have created a simple detection system that solely uses lexical-based criteria to identify advertisements through uniform resource locators (URLs). With a false negative rate as low as 1.31%, this method produce good results once the deep neural network design has been optimized. Also create a brand-new unsupervised data clustering technique. Using AutoEncoder for preprocessing features and t-distributed stochastic neighbor embedding for clustering and visualization, our model generates distinct clustering for various URL families, outperforming previous dimensionality reduction algorithms.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 02 (10), October 2023, pp. 36-48

Article Title: Advanced Malware detection with Inception V3 for Enhanced Computer Security

Ömer Aslan *et al* [2020] [3] have provided a thorough analysis of malware detection techniques and more modern techniques that make use of these techniques. Before it infects a large number of computers, the malware must be found in order to safeguard computer systems and the Internet from it. Malware identification, however, is still a challenge. Both heuristic- and signature-based detection techniques are quick and effective in finding known malware, but signature-based techniques in particular have not been able to find unidentified malware.

Ruitao Feng *et al* [2020] [4] have defined the MobiTive malware detection solution for Android, which uses specialized deep neural networks to offer a dynamic and real-time detection environment for mobile devices. Original deep learning models cannot be directly installed and operated on mobile devices because to several performance limits, including processing power, memory size, and energy. However, a deep learning-based solution for malware detection can be efficiently maintained on the server side, effectively and instantly defends mobile devices against malware attacks by utilizing binary features and specialized deep neural networks. Absence of robustness analysis and quality assurance.

Wei Yuan *et al* [2020] [5] have focused On-Device Lightweight Malware Detection Method for Android. In actuality, the need for offline upgrades makes on-device training especially crucial. However, on-device training is challenging to accomplish due to the restricted resources of mobile devices, particularly for those high-complexity malware detectors. For model training, our detector primarily use one-shot computing. It can therefore be trained immediately on mobile devices, either completely or gradually.

3 Proposed Work Explanation

Millions of sensors routinely record millions of dangerous threat events every second, according to a recent research. Malware has become more common as mobile devices and the Internet of Things have grown in popularity. Antimalware solutions from the past sometimes don't have enough power to handle the kind and volume of malware that exists nowadays. The field of malware analysis is expanding quickly, and with the growth of technology in social networks, mobile environments, cloud computing, the Internet of Things (IoT), and the Industrial Internet of Things (IIoT), it demands significant attention. Conventional malware is the term used to describe this kind of malware. These days, malware that is more destructive and challenging to identify than ordinary malware—as well as having kernel mode functionality—may be categorized as next-generation malware.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 02 (10), October 2023, pp. 36-48

Article Title: Advanced Malware detection with Inception V3 for Enhanced Computer Security

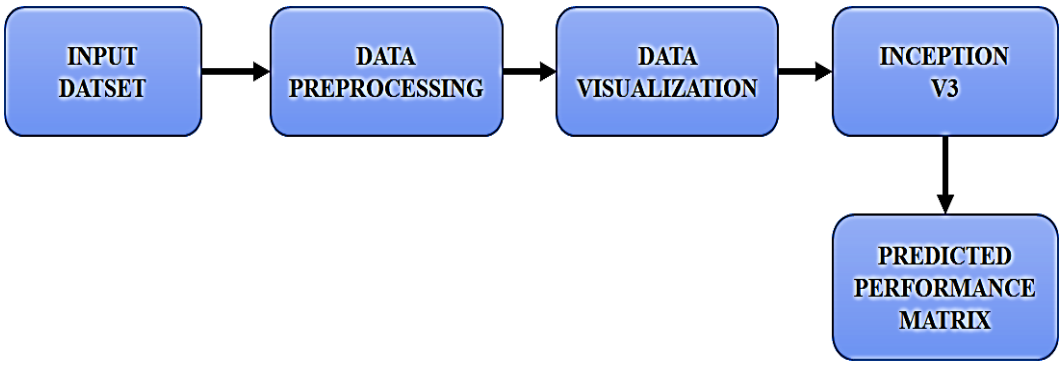

Figure 1: *Block diagram for proposed system*

The proposed methodology is outlined in Figure 1, illustrating the step-by-step process for malware detection. The approach incorporates CNN classification based on Inception V3. Initially, the input dataset undergoes a preprocessing procedure, where raw data is prepared for further analysis, a crucial step known as data preparation. Subsequently, the pre-processed dataset is fed into the data visualization block, aiming to identify trends, patterns, and outliers within large datasets. In parallel, Inception V3, a sophisticated CNN model, is employed to classify the pre-processed data. This model is chosen for its accuracy in predicting outcomes and its ability to speed up training while reducing the number of factors involved, ensuring precision in its predictions. The datasets generated are then subjected to a performance matrix, which assesses various aspects such as performance, precision, and specification. Finally, the evaluated data produces prediction results, culminating the process of malware detection.

3.1 Input Dataset

Malimg dataset, is utilized as input containing malware images. The binary content of each file is represented in hexadecimal in the raw data. The intention is to turn those files into PNG pictures so that CNN can use them as input.

3.2 Data Preprocessing

Any kind of processing done on raw data to get it ready for another data processing step is referred to as data preprocessing, which is a part of data preparation. It has always been a crucial first stage in the data mining procedure. Data preparation is necessary because high-quality data is crucial for models to work well, and it is indisputable that good data is more important than good models. There are two main categories of data preparation techniques:

Data Preparation: Any kind of processing done on raw data in order to get it ready for another data processing step. Historically, it has been a crucial first stage in the data mining process.

Data Augmentation: Used to increase the amount and quality of a given dataset in order to prevent overfitting and boost your network's capacity for generalization. A minimum mac normalization is applied to each image, ensuring that the mapping is limited to the interval [0, 1].

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 02 (10), October 2023, pp. 36-48

Article Title: Advanced Malware detection with Inception V3 for Enhanced Computer Security

3.3 Data Visualization

The process of converting information into a visual environment, like a map or graph, is known as data visualization, and it is used to make data easier for the human mind to comprehend and draw conclusions from. Facilitating the identification of patterns, trends, and outliers in huge data sets is the primary objective of data visualization. The field of data analysis that deals with the visual display of data is called data visualization.

3.4 Classification

Data categorization is the process of classifying information so that it may be easily retrieved, sorted, and stored for later use. Finding and retrieving crucial data is made simple with a well-thought-out data classification system. In general, data categorization refers to the practice of grouping pertinent data into categories for better protection and usage.

3.5 Inception V3

The idea of classification provides a way to display components, classes, data types, and interfaces. A classifier is a set of cases with similar structural and behavioural characteristics. A CNN-based network for classification is called Inception Net V3. It uses inception modules, which are 48 layers deep and comprise a concatenated layer with 1 × 1 3 × 3 and 5 × 5 convolutions. We can reduce the number of parameters and speed up training by doing this.

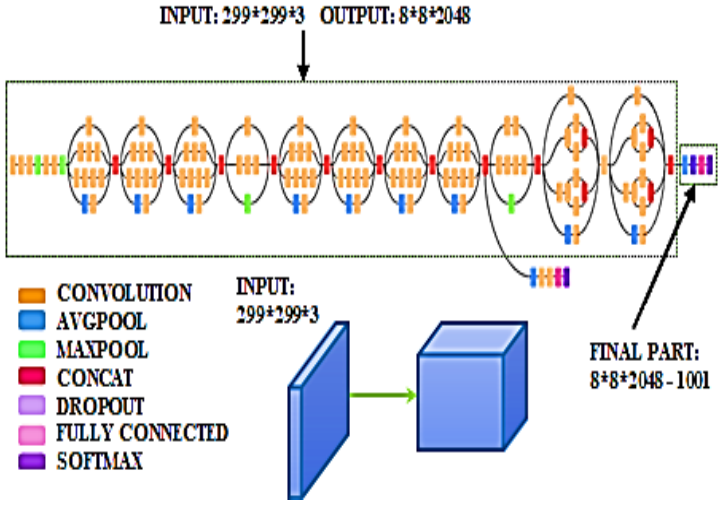

Figure 2: *Architecture of Inception V3*

Pictures of 1000 different object categories, including a keyboard, mouse, pencil, and numerous animals, may be classified by this pretrained network. Consequently, a vast array of image rich feature representations have been trained by the network. A picture input size of 299 by 299 is supported by the network. In the first phase, the model extracts generic features from the input images, and in the second half, it classifies the images based on those features.

The popular image recognition model Inception v3 has demonstrated around 78.1% accuracy on the Image Net dataset and approximately 93.9% accuracy in the top 5 results. The model is the result of numerous concepts that have been developed over time by various researchers.

3.5.1 Inception V3 Algorithm

A 2-node softmax classifier takes the place of Inception v3's last fully linked layer. This replaced layer's parameters were initialized at random:

$$\text{Softmax}\,(x_i) = \frac{\exp(x_i)}{\Sigma_j \exp(x_j)} \tag{1}$$

Where,
$\exp(x_i)$ → the current output logit's exponent.
$\Sigma_j \exp(x_j)$ → The total of each output logit's exponent.
The classifier's logits are transformed into a probability distribution using the softmax function. Note that there is an issue with class imbalance in the dataset, employing a weighted cross-entropy loss function to solve this issue. The formula provides this function.

$$Loss(x, class) = weight[class]\left(-x[class] + \log\left(\sum_j exp(x[j])\right)\right) \tag{2}$$

Where,
$weight[class]$- is used to describe the weight that is given to each class.

$$\text{loss} = \frac{\Sigma_i^N loss(i, class[i])}{\Sigma_i^N weight[class[i]]} \tag{3}$$

Several of the augmentation strategies discussed in Data Preprocessing were employed during this training phase. By using the picture augmentation, the model's performance and generalization are enhanced.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 02 (10), October 2023, pp. 36-48

4 Results and Discussion

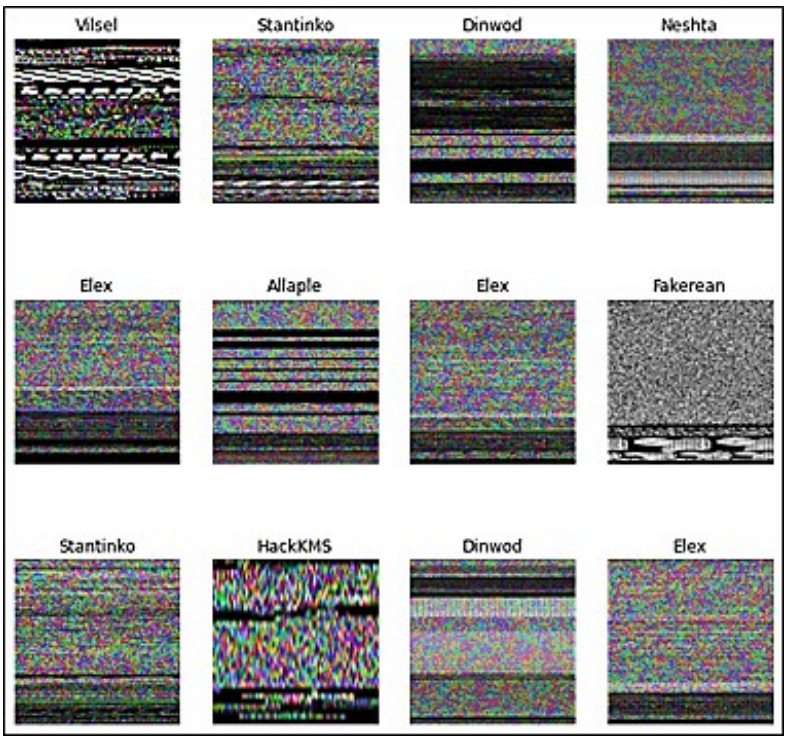

Figure 3: *Input Dataset*

The results of examining many malware samples are shown in this part in order to determine whether the byte sequences recovered using the suggested technique offer valuable data for manual investigation. There are 9339 malware pictures in the Malimg Dataset, organized into 25 families and classes. The validation of proposed methodology is done using python software. The input dataset is shown in figure 3.

Any kind of processing done on raw data to get it ready for another data processing step is referred to as data preprocessing, which is a part of data preparation. It has always been a crucial first stage in the data mining procedure. Figure 4 shows the dataset for image normalization.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 02 (10), October 2023, pp. 36-48

Article Title: Advanced Malware detection with Inception V3 for Enhanced Computer Security

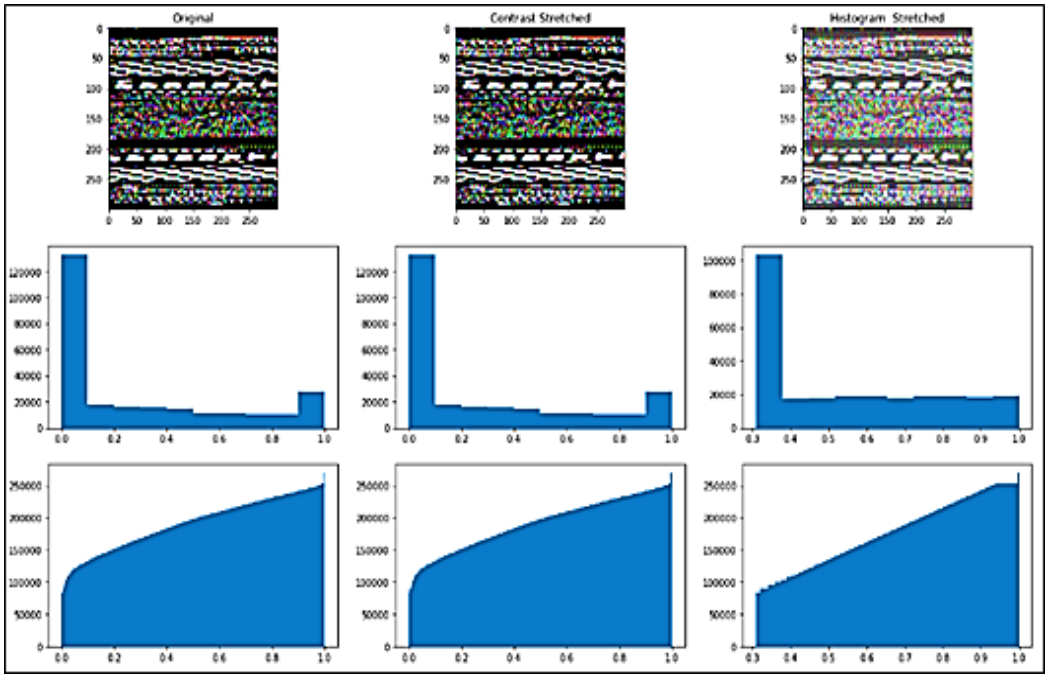

Figure 4: *Image Normalization*

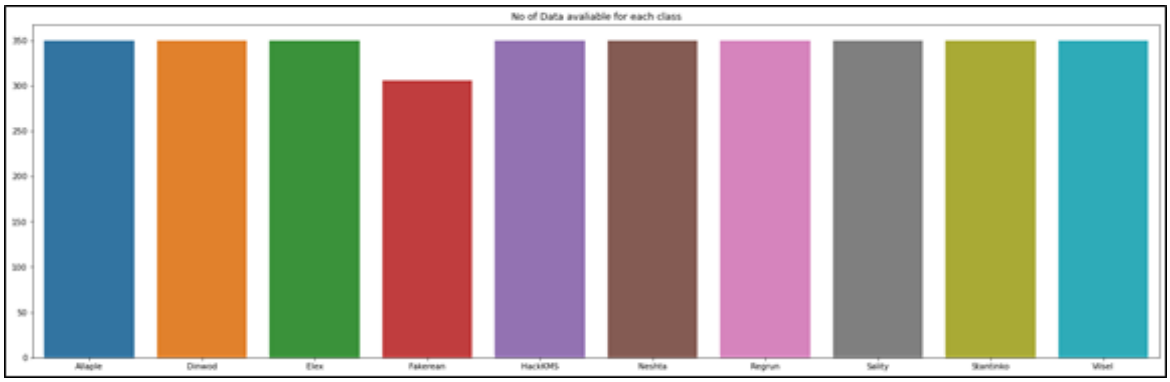

Figure 5: *Data Visualization*

The dataset's data visualization for malware categorization is shown in Figure 5. The graphical depiction of data that includes the specifics of the data is called dataset visualization. This is used to analyze data that can be visualized; it is specifically designed to analyze large amounts of data and attempt to calculate a result from it.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 02 (10), October 2023, pp. 36-48

Article Title: Advanced Malware detection with Inception V3 for Enhanced Computer Security

```
Epoch 1/5
108/108 [==============================] - 353s 3s/step - loss: 0.6448 - accuracy: 0.8058 - val_loss: 0.4087 - val_accuracy: 0.
8708
Epoch 2/5
108/108 [==============================] - 337s 3s/step - loss: 0.2065 - accuracy: 0.9245 - val_loss: 0.2261 - val_accuracy: 0.
9267
Epoch 3/5
108/108 [==============================] - 341s 3s/step - loss: 0.1297 - accuracy: 0.9505 - val_loss: 0.2089 - val_accuracy: 0.
9354
Epoch 4/5
108/108 [==============================] - 360s 3s/step - loss: 0.0810 - accuracy: 0.9722 - val_loss: 0.2446 - val_accuracy: 0.
9231
Epoch 5/5
108/108 [==============================] - 353s 3s/step - loss: 0.0522 - accuracy: 0.9797 - val_loss: 0.2341 - val_accuracy: 0.
9340
```

Figure 6: *Accuracy Report for the Inception V3*

The Accuracy Report for the Inception V3 Classification is shown in Figure 6. Overall, the optional method's classification performance is confirmed. These findings support the notion that the attention maps produced by the suggested technique accurately depict the target malware family.

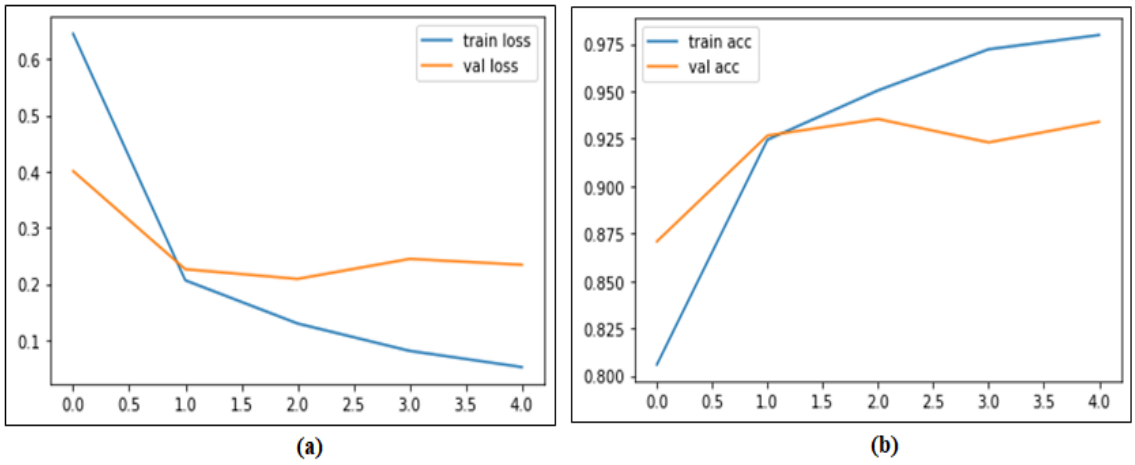

Figure 7: *Training and Validation of (a) Loss and (b) Accuracy of Inception V3*

The validation and training outcomes of proposed classifier in terms of loss and accuracy is illustrated in Figure 7(a) and (b). It is observed that the proposed system results with minimized loss and improved accuracy.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 02 (10), October 2023, pp. 36-48

Article Title: **Advanced Malware detection with Inception V3 for Enhanced Computer Security**

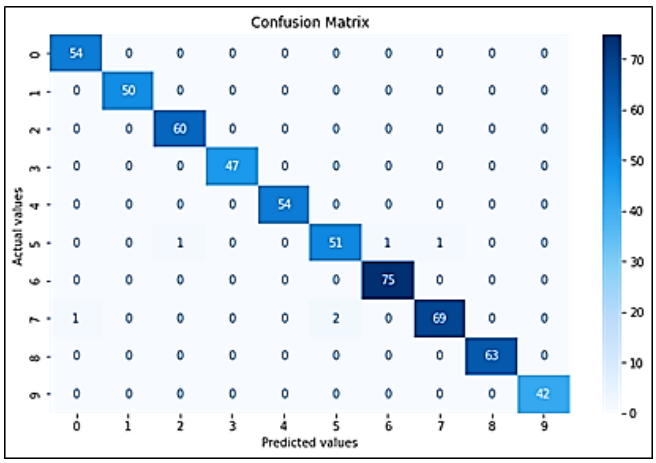

Figure 8: *Confusion matrix for Inception V3*

The confusion matrix for Inception V3 is shown in Figure 8. There is a description of a confusion matrix with real and expected values for Inception V3.

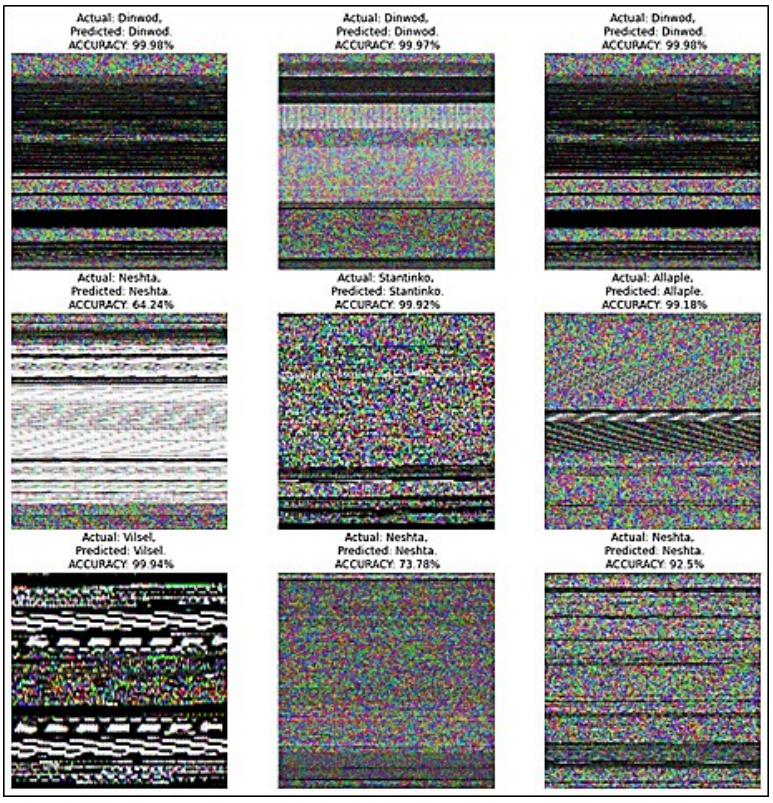

Figure 9: *Predicted Output*

The proposed method provides a precise classification of malware kinds. Figure 9 shows the individual output photos of the malware classification.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 02 (10), October 2023, pp. 36-48

Article Title: Advanced Malware detection with Inception V3 for Enhanced Computer Security

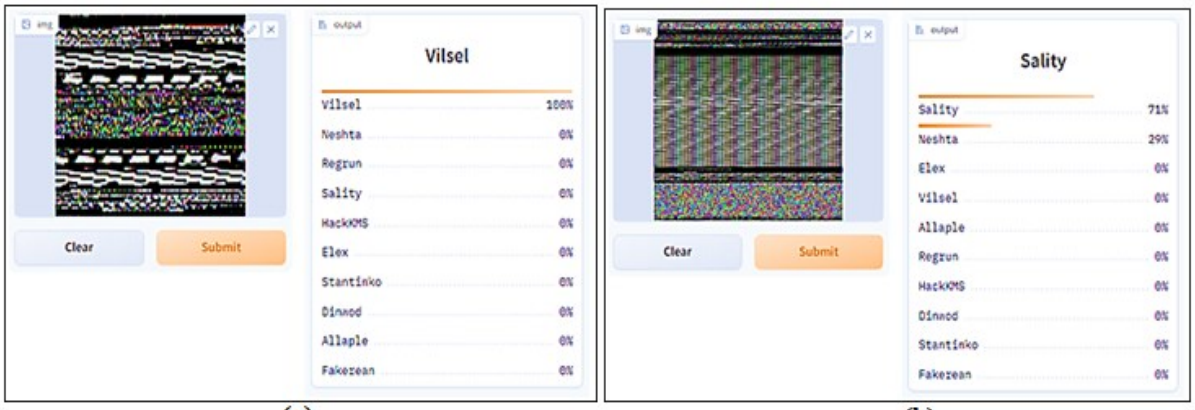

Figure 10: *(a) VILSEL and (b) SALITY Malware Families*

Anticipated is the ultimate result of classifying the families of malware. The families of malware VILSEL and SALITY are represented in Figures 10 (a) and (b) respectively. In order to detect VILSEL attacks, Figure 10 (a) lists malware families based on ten different kinds of datasets. Every single VILSEL attack is successful. Similarly, Figure 10 (b) identifies two distinct malware attack types by establishing malware families using ten distinct datasets. Successful attacks are expected to result in NESHTA malware for 29% and SALITY malware for 71%.

Table 1: *comparison of Malware detection between proposed and existing methods.*

Model	**Accuracy in (%)**
ANN	90.08
CNN	93.2
INCEPTION V3	97.5

The table 1 explains the comparison between the existing and proposed method used before in malware detection, it takes 97.5% for training in which the proposed method takes lower time than other existing method.

5 Conclusion

The proposed research underscores the critical significance of identifying and categorizing malware in the realm of computer security. The proposed methodology, centered on the utilization of CNN, specifically Inception V3, represents a significant step forward in the field of malware detection. The proposed classifier supports in categorizing and detecting more hazardous codes quickly and accurately, enabling response to the current rise in harmful code activity. By utilizing the power of CNN technique, this approach demonstrates the potential to substantially enhance the accuracy and efficiency of malware identification processes. In

conclusion, it has successfully classified and detected the different malware families using Inception V3.

Reference

1. Mumtaz Ahmed; Neda Afreen; Muneeb Ahmed; Mustafa Sameer; Jameel Ahamed, Year: 2023, "An inception V3 approach for malware classification using machine learning and transfer learning", International Journal of Intelligent Networks, Vol: 4, pp. 11 – 18.

2. S. Abijah Roseline; S. Geetha; Seifedine Kadry; Yunyoung Nam, Year: 2020, "Intelligent Vision-Based Malware Detection and Classification Using Deep Random Forest Paradigm", IEEE Access, Vol: 8, pp. 206303 – 206324.

3. Iman Almomani; Aala Alkhayer; Walid El-Shafai, Year: 2022, "An automated vision-based deep learning model for efficient detection of android malware attacks", IEEE Access, Vol: 10, pp. 2700 – 2720.

4. Ömer Aslan Aslan; Refik Samet, Year: 2020, "A comprehensive review on malware detection approaches", IEEE access, Vol: 8, pp. 6249 – 6271.

5. Nada Lachtar; Duha Ibdah; Anys Bacha, Year: 2020, "Toward mobile malware detection through convolutional neural networks", IEEE Embedded Systems Letters, Vol: 13, no: 3, pp. 134 – 137.

6. Jin Ho Go; Tony Jan; Manoranjan Mohanty; Om Prakash Patel; Deepak Puthal; Mukesh Prasad, Year: 2020, "Visualization approach for malware classification with ResNeXt", In 2020 IEEE Congress on Evolutionary Computation (CEC), pp. 1 – 7. IEEE.

7. Iman Almomani; Aala Alkhayer; Walid El-Shafai, Year: 2022, "An automated vision-based deep learning model for efficient detection of android malware attacks", IEEE Access, Vol: 10, pp. 2700 – 2720.

8. Xiaofei Xing; Xiang Jin; Haroon Elahi; Hai Jiang; Guojun Wang, Year: 2022, "A malware detection approach using autoencoder in deep learning", IEEE Access, Vol: 10, pp. 25696 – 25706.

9. Barriga, Jhonattan J; Sang Guun Yoo., Year: 2017, "Malware Detection and Evasion with Machine Learning Techniques: A Survey", ISSN, Vol: 12, no: 8, pp. 7207 – 7214.

10. Yuchen Liang; Shady Side Academy; Xiaodan Yan, Year: 2019, "Using Deep Learning to Detect Malicious URL", ICEL, Vol: 1, pp. 487 – 492.

11. S. Abijah Roseline; S. Geetha; Seifedine Kadry; Yunyoung Nam, Year: 2020, "Intelligent Vision-Based Malware Detection and Classification Using Deep Random Forest Paradigm", IEEE Access, Vol: 8, pp. 206303 – 206324.

12. Xichen Zhang; Arash Habibi Lashkari; Ali A. Ghorbani, Year: 2020, "Classifying and clustering malicious advertisement uniform resource locators using deep learning", Vol: 37, no: 1, pp. 511 – 537.

13. Ömer Aslan Aslan; Refik Samet, Year: 2020, "A Comprehensive Review on Malware Detection Approaches", in IEEE Access, Vol: 8, pp. 6249 – 6271.

14. Ruitao Feng; Sen Chen; Xiaofei Xie; Guozhu Meng; Shang-Wei Lin; Yang Liu, Year: 2020, "A Performance-Sensitive Malware Detection System Using Deep Learning on Mobile Devices", IEEE Transactions on Information Forensics and Security, Vol: 16, pp. 1563 – 1578.

15. Wei Yuan; Yuan Jiang; Heng Li; Minghui Cai, Year: 2021, "A Lightweight On-Device Detection Method for Android Malware", IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol: 51, no: 9, pp. 5600 – 5611.