ELG7186 AI for Cyber Security
“Malware classification using Malimg dataset”

Project description and state-of-the-art review

Instructor: Dr. Paula Branco

Group: 1

Contents

Problem Definition:

Malwares can be harmful to any computer and they need to be classified to the known types after detection to prevent system breaches and capturing of sensitive data. Building a system from scratch can be a hassle for such classification systems. Image-based malwares are should be classified automatically instead of manual classification by experts which can varies in accuracy due to different levels of knowledge and intellectual skills between them.

State-of-the-art Review:

MCFT-CNN: Malware classification with fine-tune convolution neural networks using traditional and transfer learning in Internet of Things (Sudhakar, Sushil Kumar)[2]

Summary:

Security experts are endeavoring to make a method that precisely perceives all malware. In this article, the author has proposed a one-of-a-kind convolution mind network-based malware-gathering method. Without featuring the planning, sorting out, or even the advanced dubious strategies used to make the disease, the MCFT-CNN model sees the dark malware test.

Considering its qualities and capacities, the malware might be arranged into many kinds, like worms, diversions, spyware, ransomware, infections, secondary passages, and adware. A new malware assortment presents extra snags for the online protection specialist to distinguish precisely.

Text and picture information is altogether different from the information gathered and made by malware tests. Nataraj et al put out an original technique for malware ID and grouping using visual qualities. They changed the parallel executable's construction into two-layered, greyscale pictures. Afterwards, the AI calculation for malware discovery was prepared to utilize these qualities. Without including designing or past information on twofold code investigation or picking apart, the MCFT-CNN model might foresee unidentified malware types.

By using polymorphic change and morphological jumbling, it can unequivocally identify malware variations intended to get past antivirus programs. Also, the model is more exact and quicker in foreseeing known malware.

A MCFT-CNN model was recommended in this review to sort malware tests into malware families. With the Adam enhancing compiler, it has a 99.05% exactness rate, while prepared utilizing the traditional learning strategy has a 99.22% precision rate. The model likewise accomplished a critical accomplishment with its 5.14 ms gauge time for obscure malware tests.

Research goal:

This paper has introduced an MCFT-CNN model for classifying malware tests into malware families. The advantage of utilizing a profound learning-based CNN model is that no component designing is required. Besides precision, proposing a model that characterizes new malware low time. Their methodology is to outperform past state of art concentrates on malware picture characterization involving profound learning as far as exactness and expectation time.

Methodology:

Deep learning is a subfield of AI that falls under the umbrella of computerized reasoning. In profound learning, input is taken care of into the model, which then, at that point, registers through layers to anticipate or arrange obscure information. The layers of the profound learning model depend altogether on counterfeit brain frameworks for information preparation.

A transfer learning procedure was utilized to order malware parallels. The malware doubles are displayed in greyscale. Pictures of malware family varieties show underlying likenesses. Accordingly, the order might be concluded that the infection pictures are utilized as picture classification. For this situation, the entire data of malware tests are used.

Experiments and results:

The model is prepared to utilize a standard learning technique utilizing benchmark datasets (MalImg and Microsoft datasets). A few examples are moreover muddled and incorporate a powerful motor inside the code. MCFT-CNN outflanks ResNet50 with regards to ordering varieties of a certain malware family, like CLOP, Swizzor.gen! E, as well as Wintrim.BX.

(Preciseness:.96 to 1) The classification report portrays Examination I and Trial III, each utilizing Adam and NAdam optimizers. Our model anticipated obscure examples in 5.14 ms, which is altogether quicker than the present status of-the-craftsmanship research on the Huge 2015 dataset.

Strengthen and weakness:

The model's strengths include 99.05% accuracy with the Adam optimizer and 99.22% accurateness with the NAdam optimizer when constructed using the traditional learning technique. The model also achieved a crucial success in the anticipation time of cryptic malware testing, which is 5.14 ms. Furthermore, the model obtained 99.18% accuracy with the Adam optimizer and 99.10% preciseness with the NAdam optimizer in the interchange learning technique.

Regardless, the limitation is, that there is a limit of employing the model benefits from uniform image size. The author should have used the spatial pyramid pooling layer, which may handle any size information image, however he has stated that he will consider this in the future.

Malware Classification Using Automated Transmutation and CNN (Ritu Agarwal, Saurabh Patel, Sparsh Katiyar , Sharad Nailwal)[1]

Summary:

Malware is a piece of software that damage or cause harmful for the computers, this software mainly used to steel sensitive information.

Due to the increases of malicious softwares we need to analyze them to develop a software that can detect the malicious attack before it happens, also it`s essential that detect the malware type, so now days the machine learning and deep learning can be used to detect the pattern of malware and block any malicious attack.

In this conference paper the authors proposed a solution to study the pattern of the malware it`s a binary data and show it as image to classify malware images using a deep learning CNN architecture pretrained model to extract the different malware signatures

The authors here use a pretrained model to build his solution VGG16 and use the weight of this model to build his approach.

Research Goal:

This conference paper aims to classify the malware types using malware images using deep neural networks pretrained model VGG16, set the suitable parameters for model training, remove some layers and evaluate the model performance, the authors use accuracy as evaluation matrix.

The authors started their work by a literature review for the past related works for classical supervised machine learning techniques such as K nearest neighbor and support vector machine model (SVM) which achieve on 14 types of malwares with average accuracy 88%.

And compare the previous result to his approach that achieve accuracy 98.97%.

Methodology:

The authors proposed a methodology that meets the research goal which is to find a suitable method to determine the malware type starting with visualization as known the malware is present in executable format so looking at binary file and divided it for each 8 bits that help to visualize the malware type as image and detect the malware signature here the authors use gray scale image the size is 224*224,

The proposed model uses a deep neural network pretrained model VGG16 and use transfer learning to apply his approach the loss function used is categorical cross entropy this loss function gives probability that can distinguish between two classes of each other, the activation function used in the convolutional layer is Relu and the activation function of the output layer is softmax to give the probability for each class.

Experiments and results:

In this paper, the authors provided an applied result, for the experiments they used a malware images dataset which contains 9339 images from different 25 malware families, split the data into 90% training (8405) images and 10% testing (934).

For experimental purpose, the authors used Google Colab with GPU enabled. Configuration of the virtual machine was as follows: Dual Core (TM) Intel Xenon CPU (2.30 GHz) with 12 GB RAM and Nvidia P40 GPU 12 GB RAM. Operating System used was Ubuntu 20.04 64bit they trained the model on number of epochs = 20 and the Batch size= 10000

The authors apply many different models such as GIST with SVM and achieve 93.23% also M-CNN and achieve 98.52% and the VGG16 his chosen model achieve 98.97%.

Strengthen and weakness:

This paper achieved high accuracy in classifying the malware types, one of the strengthen points is that the authors used a very simple approach transfer learning to achieve his result and didn`t use any complex feature engineering to classify the malware images but they didn`t mention why they use transfer learning also, they used accuracy as evaluation matrix and they didn`t mention that the data is imbalance or not because if it`s not imbalance the accuracy metrics didn`t fit well.

IMCLNet: A lightweight deep neural network for Image-based Malware Classification[3]

Summary:

Malware is a software program or code that can be harmful for your computing device, it specifically designed to damage, distribute, or gain an unauthorized access to your device. By the time, the number of malwares increased, so, detecting malware type is a very important challenge. To solve this challenge, it's essential to detect the malware type by the machine learning techniques.

The authors proposed a machine learning model to classify the malware images, which is the lightweight malware classification model (IMCLNET) that doesn't need complex feature engineering, large domain knowledge, pre-training parameters and data enhancement. By studying the impact of the images size, the authors said that the images size can affect the performance of the classification model by decreasing or increasing the classification model performance, but the accuracy doesn't change as the performance. The prediction time of the IMCLNET model also affected by the images size. So, it's a great challenge to reduce the prediction time and the impact of images size without affecting the classification model accuracy.

Research goal:

This paper aims to detect the malware types using malware images using the lightweight malware classification model (IMLCNET) using deep neural networks, set the suitable parameters for model training, evaluate the model performance, and increase the classification model accuracy without affecting the model performance and predictions.

The authors started their work by a literature review for the past related works for classical supervised machine learning techniques such as support vector machine model (SVM) which doesn't achieve a high performance for classifying the malware types, image-based methods, which are about dealing with the malware images on gray scale to classify them with the DenseNet model, this method are not sufficient and the detecting efficiency is very low.

The lightweight malware classification model has an efficiency on detecting the malware type with high model performance, limited preprocessing and feature engineering and without data augmentation.

Methodology:

The authors proposed a methodology that meets the research goal which is to find a suitable method that can classify the malware images. They performed malware images visualization as a grayscale image because it doesn't require complex feature engineering, complex feature extraction and expert domain knowledge. After applying the grayscale to the images, resize the images to a uniform size which is 224 * 224. While normalizing the data, most of malware images may lose the most important features, but most of them save their texture features and layouts. The similarity of layouts between the two images indicates that they are from the same family, then we can use these images in malware classification task.

The IMCLNET which consists of full convolutional layer, coordinate attention layer, Depth wise separable convolutional layer, and pooling layer can classify malware images and calculate the number of parameters and weight accuracy.

The attention mechanism helps in malware classification by capturing the positional information and finds the relationships between the channels to generate the attention maps. The Depth wise separable convolutional layer applies only one filter for each input channel, and makes the model performance efficient by reducing the complexity of the classification model.

Finally, the global context embedding designed to work at different stages to capture different feature embeddings of malware images while extracting features from the malware images.

Experiments and results:

In this paper, the authors provided an applied result, for the experiments they used a malware images dataset which contains 9339 images from different 25 malware families, and a dataset contains bytecode files from Kaggle which belongs to a competition published in 2015. The bytecode files used to generate the malware images for classification task.

The authors used 75% of the data for training and 25% for testing, using python 3.9 environment with pytorch and Nvidia 12GP GPU. They used 50 epochs and learning rate = 0.0038 for training the model. They used FLOPs metric to evaluate the model and calculate the model complexity. Then they used accuracy score, precision, recall and F1-score for model evaluation.

They implemented several models to compare with IMCLNET model such as SVM, Adaboost, and several classification techniques, and the IMCLNET model with cross validation was the best one with 99.272% accuracy score for the image dataset, and 98.666 % accuracy score for the BIG2015 dataset.

Strengthen and weakness:

This paper achieved high accuracy in classifying the malware types, one of the strengthen points is that the authors used cross validation, filtering mechanism and classifying the images without performing complex feature engineering on the images. But they didn’t mention the disadvantages of IMCLNET model and the effect of hyperparameter tuning of the model.

Solution:

By using Malimg dataset which contains 9339 malware images that belong to 25 families/classes, a pre-trained model will be used with our own architecture, probably VGG-19 which is a 19 layers depth convolutional neural network that can classify images into 1000 object categories. The data should be deployed and trained with a fine-tuning to be done. Cross-validation will also be used and the data will be split into 80% train, 10% validation and 10% test. The classification will be evaluated using common metrics such as accuracy, recall, precision and F1-score. There will be a deployment and data representation to be made on streamlit.

Dataset:

malimg_dataset.zip. (n.d.). Dropbox.

https://www.dropbox.com/s/ep8qjakfwh1rzk4/malimg_dataset.zip?dl=0

References:

[1] Agarwal, R., Patel, S., Katiyar, S., & Nailwal, S. (2021). Malware Classification Using Automated Transmutation and CNN. *Advanced Computing and Intelligent Technologies*, 73–81. https://doi.org/10.1007/978-981-16-2164-2_6

[2] Sudhakar, & Kumar, S. (2021). MCFT-CNN: Malware classification with fine-tune convolution neural networks using traditional and transfer learning in Internet of Things. *Future Generation Computer Systems*, *125*, 334–351. https://doi.org/10.1016/j.future.2021.06.029

[3] Zou, B., Cao, C., Tao, F., & Wang, L. (2022). IMCLNet: A lightweight deep neural network for Image-based Malware Classification. *Journal of Information Security and Applications*, *70*, 103313. https://doi.org/10.1016/j.jisa.2022.103313