MALWARE IMAGE CLASSIFICATION USING DEEP LEARNING MODELS

Prof. Jayasudha M[1], Aaryan Mehta[2], Ladi Jeevan Sai[3], Sarvesh Chandak[4], Parasu Sai Nikhil[5]

1.Associate Professor, 2,3,4,5 Undergraduate Students, Department of Computer Science, Vellore Institute of Technology,
Chennai – 600127, TamilNadu, India.

1. jayasudha.m@vit.ac.in 2. aryanmehta376@gmail.com, 3 ladijeevansai2002@gmail.com, 4.sarveshchandak18@gmail.com
5. nikhil.parasu517@gmail.com

***Abstract*—** Malware is a type of harmful software created with the intent to harm or to take full control of your computer systems. Here in this study, we have used Malimg dataset to examine the performance of different deep learning models in classifying malware images. Malware detection and classification are essential for ensuring computer system security.
Wen have employed various transfer learning algorithms for this task namely, VGG16, ResNet50, and InceptionV3.
Malware image classification is a critical cyber security task that involves finding and classifying malicious images in order to prevent malware attacks. This study demonstrates the potential of deep learning models in cyber security for malware identification and prevention

Keywords—Malware, CNN, Transfer learning, Deep Learning, Cyber Security

I. INTRODUCTION

Malware attacks are escalating every passing day. Using image classification techniques to classify malware images is one method for malware detection. This is one of the most effective approach because malware often has distinct visual features that can be detected through image analysis.
In this project, we will classify malware images using the Malimg dataset and deep learning models. The Malimg dataset, includes more than 9000 malware images from 25 different families. It is widely used dataset for malware image classification.
Our objective is to create models that can correctly classify malware images and identify the malware family to which they belong. By early detection of malware attacks, the security of computer systems and networks can be increased.

II. OBJECTIVE

- To convert a malware into an image and analyze it.
- To develop a deep learning model that can accurately classify malware and non-malware images.
- To compare the performance of the CNN-basedapproach with other malware detection methodsand evaluate its advantages and limitations.
- To improve the detection rate of unknown malware by using visual analysis instead of traditional methods.
- To reduce the number of false positives and false negatives in malware detection.
- To improve the overall efficiency.

III. Related Work

Over the past few years, there has been a growing interest in using deep learning models for malware image classification. The Malimg dataset has been widely used in this area and has served as a benchmark dataset for evaluating the performance of different deep learning models.

In [1] The authors proposed a novel deep learning and Markov image classification method for malware. The suggested methodology first takes the byte-level representation of malware and extracts n-gram features from it. Then Markov images are created from their characteristics. In order to classify the malware samples, a deep convolutional neural network is trained on the Markov pictures. They achieved a accuracy of 97.36% In [2] the authors proposed a comprehensive study of malware detection using image analysis techniques. The authors evaluated various CNN architectures on the Malimg dataset and achieved an accuracy of 98% In [3] The authors used CNN architecture, consisting of several convolutional layers followed by pooling layers and fully connected layers. They also used techniques like dropout and batch normalization to prevent overfitting.
They achieved a accuracy of 96%

In [4] The authors proposed a novel approach which involved multiclass categorization. First they converted raw malware binaries into colour images which is utilized by the refined CNN architecture for detecting and identifying malware families. They also utilised various data augmentation trchniques. They used 2 datasets for their study, Malimg malware dataset and the IoT-Android mobile dataset. They achieved a accuracy of 98.82% in the Malimg malware dataset and 97.35% in the IoT-android mobile dataset. In [5] the authors proposed a method for classifying malware programmes based on transfer learning techniques. Windows portable executable files are converted to grayscale images. The specialised deep CNN architecture receives input in the form of grayscale images. Features are flattened and put into a fully connected dense layer after being recovered from the convolutional layers of the deep CNN model. Early Stopping is also used to prevent overfitting. They have used 2 datasets in this study, the MalImg dataset and the Microsoft BIG dataset. They achieved a accuracy of 93.19% for Microsoft datasets and 98.92% for MalImg datasets. The author proposed a deep learning-based malware detection technique based on static methods for classifying different malware families in [6]. The proposed technique uses grayscale images of malware samples for feature engineering and thus classifying malware families.
The proposed technique uses a pre-trained VGG16 model and fine-tunes it on the grayscale images of malware samples. They achieved a accuracy of 99.5%

In [7] authors proposed a novel method that uses deep learning to improve the detection of malware variants. They converted the malicious code into grayscale images and classified them using a convolutional neural network (CNN) that could extract the features of the malware images automatically. To address the data imbalance among the different malware families they utilized a bat algorithm. They achieved a accuracy of 94.5%.

For malware classification, the authors in [8] presented a deep learning approach. The authors developed a convolutional neural network (CNN) for classification after converting malware binaries into grayscale images. For the Microsoft malware dataset, the suggested technique produced classification results with an accuracy of 99.97%. In [9] the authors proposd a malware classification approach that integrates deep learning techniques with the static malware genes. The malware gene sequences with both material and informational attributes are extracted using the model. They achieved a accuracy of 98.5%.on malimg dataset. In [10] the author suggested a hybrid model-based deep learning architecture for classifying malware variants. Tested on the Malimg, Microsoft BIG 2015, and Malevis datasets, they achieved an accuracy of 97.78%. For the Malimg dataset, they employed a hybrid model that combined two deep neural networks and achieved a accuracy of 98.7%.

Overall, these studies demonstrate the effectiveness of deep learning models for malware image classification using the Malimg dataset. However, there is still room for improvement, and further research is needed to develop more robust and accurate models for malware detection and classification.

IV. Research Gap

Malware image classification using deep learning models is an active research area, and there are several studies that have been focused on this topic. However, there are still some researchgaps that need to be addressed.

One of the main gaps is the lack of research on the performance of deep learning models for malware image classification using the malimg dataset. The malimg dataset is a publicly available dataset that contains over 9,000 malware images, which are collected from different sources. However, there is limited research onthe use of this dataset for malware image classification usingdeep learning models. Therefore, there is a need to investigate the performance of deep learning models using the malimg dataset and compare it with other existing datasets.

Another research gap is the lack of studies on the use of transfer learning models for malware image classification. Transfer learning is a technique that has been widely used in other domains to improve the performance of deep learning models by leveraging pretrained models..

Lastly, there is a need to investigate the effectiveness of different deep learning architectures for malware image classification. While there are several deep learning architectures that have been proposed for image classification tasks, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), there islimited research on the effectiveness of these architectures for malware image classification. Therefore, there is a need to compare the performance of different deep learning architectures.

In summary, there are several research gaps that need to be addressed in the area of malware image classification using deep learning models, including the performance of deep learning models using the malimg dataset, the transferability of deep learning models, and the effectiveness of different deep learning architectures. Addressing these gaps can help to improve the accuracy and effectiveness of malware detection systems.

V. Proposed Architecture

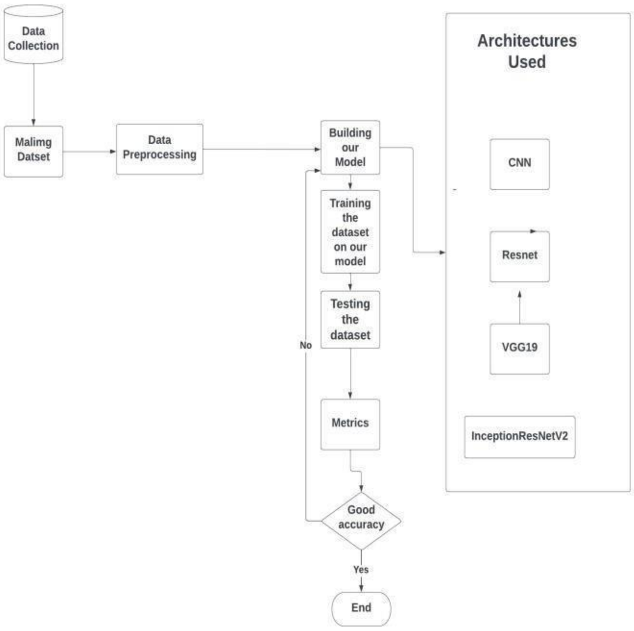

Fig 1. Architecture of our proposed method.

VI. Proposed Methodology

Deep learning models for malware image classification are a common method for identifying and categorising harmful malware. We have used malimg dataset for classification. Below is our proposed methodology for classifying malware images using deep learning models:

- **Data Collection and Pre-processing:**

This module involves collection of dataset of labeled malware and non-malware images, and preprocessing the dataset which we will use for training and testing. First, the Malimg dataset needs to be preprocessed to make it suitable for deep learning models. This includes resizing the images to a fixed size, normalizing the pixel values, and splitting the dataset into training and validation sets.

- **Deep learning Model Design:**

This module involves collection of dataset of malware and non-malware images and preprocessing it so that it can be used for training and testing. This module involves some pre-processing techniques like scaling the photos to a fixed size, normalizing the pixel values, and splitting the dataset into training and testing.

- **Training and Validation:**

The preprocessed dataset is used to train the chosen model. In order to reduce the classification error, the model is fed the training data and the corresponding weights are adjusted. To improve the performance of the model, the model's hyperparameters—such as learning rate and batch size—can be further changed. This can be achieved by splitting the dataset into training and testing sets, and training the model on the training set.

- **Testing and Evaluation:**

The training set must be used to test the model once it has been trained in order to determine how well it performs. The performance of the model can be measured using various metrics like accuracy, precision, recall, and F1-score. Based on the accuracy we will choose the best model.

- **Model Optimization:**

Several methods can be employed to optimise the model performance incase of bad accuracy. The can be done by changing the model architecture, hyperparameters, by adding more data and using some data augmentation techniques.

- **Model Deployment:**

The model can be used to classify malware once it has been optimised. The model can predict any malware from an input image. A malware detection system that uses the approach can automatically categorise malicious photos.

In summary, data preparation, model selection, model training, model evaluation, model optimization, and model deployment together can be used for classifying malware images using deep learning models.

VII. Architectures used

CNN

Convolution neural networks(CNN) are frequently employed for image classification. They use convolutional filters on the input data to automatically learn and and extract features from photos.
CNN can be used to categorise malware images into various groups based on their visual properties. The malimg dataset, consists of various malware sources from multiple sources. CNN has achieved high accuracy for classification of malware images.

InceptionResNetV2

InceptionResNetV2 is a modern deep learning architecture which is used for image classification tasks. It is a combination of two popular deep learning models, Inception and ResNet architectures. The InceptionResNetV2 architecture has shown outstanding performance in various computer vision applications, including segmentation, object detection, and image classification.
InceptionResNetV2 can be used to classify various malware image by utilizing the Malimg dataset. The Malimg dataset, , has been extensively used to compare the performance of different deep learning models.

VIII. Results and Discussions

- Using CNN we obtained a accuracy of 96.13%
- Using InceptionResNetV2 we obtained accuracy of 94.84%
- Using VGG19 we obtained a accuracy of 33.90%

Model Name	Accuracy	F1 Score
InceptionResNetV2	94.84	94.14
VGG19	33.09	17.16

Overall, it is clearly visible that deep learning models can be used to classify malware images using the Malimg dataset which shows the potential of deep learning models for enhancing security experts' capacity to identify and mitigate malware attack.

Further study can be done to investigate the use of these models for other malware datasets and applications. We can clearly see that CNN outperforms all other models.

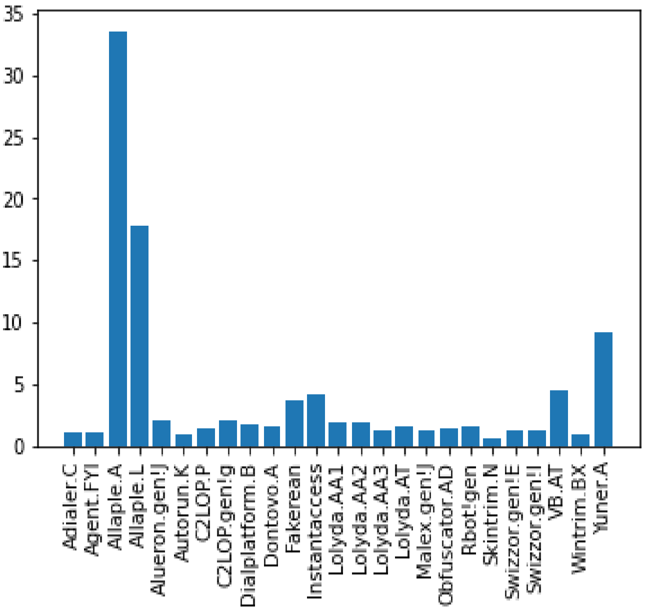

Fig 2. Graphical representation of number of images in each malware family.

IX. Conclusion and Future Scope

In conclusion, deep learning models can be used for classification of malware images using malimg dataset. Malimg dataset, consists of 9,335 malware images from 25 different families.
The malimg dataset has been used to classify malware images using a number of deep learning models, including Convolutional Neural Networks (CNNs) and Transfer Learning. The malware families found in the photos have been successfully identified by these algorithms with commandable accuracy rates.
VGGNet, ResNet, and InceptionNet are a few of the well-known deep learning models that we have used for malware image classification.
One of the challenge in this task is small size of dataset which can lead to overfitting. But methods like data augmentation and transfer learning can be employed to overcome this problem.
It is safe to say that deep learning models have a great potential and can be used for detecting and mitigating any cyber security or malware threats. Further study can be done to implement different models and use combination of dataset to enhance its performance and using it for various other cyber security application.

ACKNOWLEDGMENT

We express our profound gratitude to Guide Prof. Jayasudha M, Department of Computer Science, Vellore Institute of Technology, Chennai for her expert guidance, encouragement and inspiration during this research work.

REFERENCES

[1]. Yuan, B., Wang, J., Liu, D., Guo, W., Wu, P. and Bao,X., 2020. Bytelevel malware classification based on markov images and deep learning. Computers & Security, 92, p.101740.

[2]. Kabanga, E.K. and Kim, C.H., 2017. Malware images classification using convolutional neural network. Journal of Computer and Communications, 6(1), pp.153-158.

[3]. Choi, S., Jang, S., Kim, Y. and Kim, J., 2017, October. Malware detection using malware image and deep learning. In 2017 International Conference on Information and Communication Technology Convergence (ICTC) (pp. 1193-1195). IEEE.

[4]. Vasan, D., Alazab, M., Wassan, S., Naeem, H., Safaei, B. and Zheng, Q., 2020. IMCFN: Image-based malware classification using fine-tuned convolutional neural network architecture. Computer Networks, 171, p.107138.

[5]. Kumar, S. and Janet, B., 2022. DTMIC: Deep transfer learning for malware image classification. Journal of Information Security and Applications, 64, p.103063.

[6]. Pant, D. and Bista, R., 2021, November. Image-based Malware Classification using Deep Convolutional Neural Network and Transfer Learning. In 2021 3rd International Conference on Advanced Information Science and System (AISS 2021) (pp. 1-6).

[7]. Cui, Z., Xue, F., Cai, X., Cao, Y., Wang, G.G. and Chen, J., 2018. Detection of malicious code variants based on deep learning. IEEE Transactions on Industrial Informatics, 14(7), pp.3187-3196.

[8]. Lad, S.S. and Adamuthe, A.C., 2020. Malware classification with improved convolutional neural network model. Int. J. Comput. Netw. Inf. Secur, 12, pp.30-43.

[9]. Meng, X., Shan, Z., Liu, F., Zhao, B., Han, J., Wang, H. and Wang, J., 2017, October. MCSMGS: malware classification model based on deep learning. In 2017 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC) (pp. 272-275). IEEE.

[10]. Aslan, Ö. and Yilmaz, A.A., 2021. A new malware classification framework based on deep learning algorithms. Ieee Access, 9, pp.87936-87951.