International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Neural Network Collation: A Comparative Study on Novel Image-based Malware Classification through Neural Network

[1]P. M. Kavitha, [2]Dr. B. Muruganantham

Submitted: 26/05/2023 **Revised**: 06/07/2023 **Accepted**: 25/07/2023

Abstract: The term malware is defined as any malicious software that affects the system or software. A malware is a piece of program that sticks to the system and affects the same. Most of the times, it is found stealthy and infects the user without his knowledge. But this malware can be benign and the classification of the malware from benign needs to identified. Several algorithms come in hand in detecting the malware like KNN, SVM, and decision tree. This comparative study up brings the various malware classification methods and identifies the one with the best accuracy. The work portrays two classification algorithms Such as Mal_CNN and Mal_CapsNet by the author along with the standard CNN and Capsule Neural Network. The work delves by augmenting the Malimg dataset of 9339 malware with 25 malware families for training the model. With this result, segmentation is worked upon to produce 27890 images. With the resultant image, the work flows upon the Mal_CNN and Mal_CapsNet to produce a greater accuracy. After several experiments on the pretrained model, it is found that Mal_CapsNet achieves a significant accuracy of 97.6%. The study focuses a comparison on the four models like CNN, Capsule Neural Network, Mal_CNN and Mal_CapsNet, to identify the best model for malware classification.

Keywords: *Deep Learning, Neural Network, Convolution Neural Network, Capsule Neural Network, Feature engineering.*

1. Introduction

There is a drastic development in the cyber offence. Day by day the vulnerability focuses on the data stealing. Data stealing can be done in many ways. Generating malware has been a usual aspect for hackers and cyber criminals [11]. Malware can be represented in many formats. It can be represented in hashing function, image binary files etc. In traditional [17] approach static and dynamic analysis were used. This study works with the malware of image type. Malware can be classified through AI algorithms. Few classifiers [9,19,20] like SVM, KNN, Random Forest, CNN etc. excel better in classification. Many studies states a [3] hybrid approach in classification which combines the features extracted [18] through segmentation.

In this research, an IoT-based approach is suggested for enhancing water management in smart cities. The suggested remedy is creating a fundamental architecture for the water.

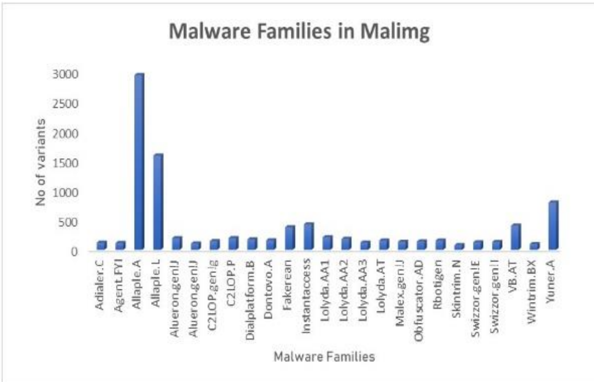

Fig.1 Representation of Malware families with number of malware image samples

Deep learning algorithms and Machine learning algorithms like [14,15,16] CNN, Capsule neural network, ANN, Deep CNN best outperforms in detection of malwares. Few of the study proven to be a better performs when there is a hybrid network.[10] These frameworks extracts features which leads to classification. [8] Few studies have proved; DL is better in malware detection.

This article is related to a comparative study performed through neural network frameworks. The two novel neural network classifiers are compared. In order to showcase the better accuracy of the models existing classifiers were also used in this study to differentiate. The dataset implemented in this study is Malimg with 9339 image-based malwares. The paper is organized in such a way that section 2 is a crips study on the related works. Section 3 discuss the feature engineering concepts implemented over the malware sample

[1]*Department of Computing Technologies, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Chennai., pmkavimca@gmail.com*
2Department of Computing Technologies, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Chennai, bmuruganantham@gmail.com

images available in the dataset. A detailed note on the model selection is discussed in the section 4 followed by the result identified through the two novel approaches discussion in section 5. Finally, section 6 is a detailed note on the comparison between the models based on the performance metrics.

2. Related Work

Image classification on the detection of Malware which uses Capsule network is the work by Xiaoliang Zhang et.al. [6] The malware detection encompasses with the MalCaps, by overcoming the limitations of CNN. With the grayscale malware images, a dynamic routing based capsule network is proposed, for malware detection produce a result of 99.34% accuracy with Microsoft challenge dataset. This consists of 9 families of malware. MalCaps, a revised malware classification for visualization and Capsnet is proposed. The architecture with two convolution layers increases the accuracy of the model. Further, a Random search runs along the variety of neurons in the initial convolution layer. [4] Vasan et.al proposed an enhancement of the malware detection on various malware families is proposed, a IMCFN method based on CNN framework. This method classifies the images with a fine-tuning CNN architecture, a multiclass classification. The primary data gets transferred into RGB images, which fine tunes for the CNN. Two datasets such as Malimg and IoT-android mobile data set with subsequent samples are used up in this work. Among the other deep learning methods, IMCFN produces the most empirical results, achieving an accuracy rate of 98.82%, 97.35% in Malimg and IoT-android mobile data set respectively. This study further gets enhanced by performing better on colored images.[2] Ding.et.al proposed in his paper an efficient malware classification method using feature extraction based on neural network.

Bensaoud et.al [5] works on about six models under deep learning that classifiers the malware images using the classic CNN models. Among them, three are from the ImageNet large scale visual recognition, and rest are the other enhanced framework. The Inception V3 model showed a greater accuracy of all the models. [6] The focus of the paper by Manoharan J.S, is the usage of Capsule Net for optimization of text classification. A classic Capsule network for the classification of hierarchical multi-label text is worked out. Further, the work goes by comparing SVM, LSTM, ANN, CNN and several other neural and non-neural networks. This enhances the performance of the model for the datasets. As a result, the algorithm encodes with the latent input and gives the varied categories.

The focus of the paper by Manoharan J.S,et.al [7] is the usage of Capsule Net for optimization of text classification. A classic Capsule network for the classification of hierarchical multi-label text is worked out. Further, the work goes by comparing SVM, LSTM, ANN, CNN and other such neural and non-neural based networks. This enhances the performance of the model for the Blurb Genre collection and Web of Science datasets. As a result, the algorithm encodes with the latent input and gives the varied categories.

3. Dataset

Malimg dataset [12] which is used in this study consist [1] of 9339 malware image samples. It comprises of 25 malware families. Each family has irregular image sample. The malware families are Fig. 1 Adialer.C, Agent.FYI, Allaple.A, Allaple.L, Alueron.gen! J, Alueron.gen! J, C2LOP.gen! g, C2LOP.P, Dialplatform.B, Dontovo.A, Instantaccess, Swizzor.gen!E, VB.AT, Wintrim.BX, Yuner.A, etc. Few researchers converted the malware image samples to RGB. We apply gray scale pattern of images. The dataset comprises of grey scale image, each with variable size.

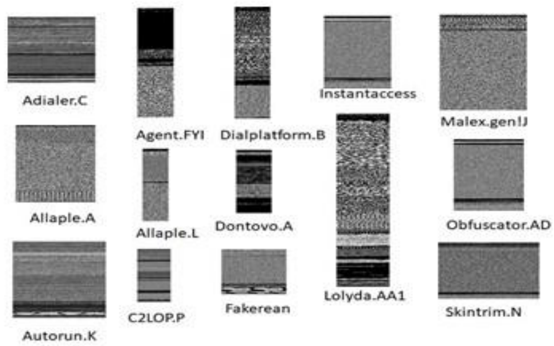

Fig. 2. (i) Representation of malware family.

In direction of the enhancement of the model performance benign samples were used. Fig 2. (i) is the representation of sample group of malware. malware families shown in Fig 2. (ii) is the representation of benign malware samples.In Table.1 The dataset used by various researchers are listed for a comparison. The classifiers and the performance were also discussed.

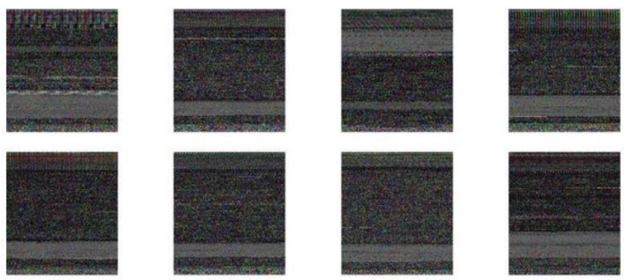

Fig.2. (ii) Representation of Benign Images

Dataset Used	Samples count	Technique Followed
Malimg dataset[1]	Totally 25 malware families with	Visualization and automatic classification

	9,458 samples images[1].	
Microsoft malware data set[2]	Nine classes of malware with 10,868 samples.	Feature extraction based on deep neural network
Malimg malware image dataset[3]	9,435 samples	Segmentation-based fractal texture analysis and deep convolution neural network features
1.Malimg malware dataset 2. IoT- android mobile dataset	1.9,435 samples 2. 14,733 malware and 2,486 benign samples	Fine-tuned convolutional neural network architecture
Malimg dataset[4]	9,435 samples of 25 malware families	Convolutional neural network models
Microsoft Malware Classification Challenge (MMCC) dataset[5]	21,741 samples	Capsule network-based model
Kaggle dataset.[6]		Capsule network algorithm for text classification
Malware image data from Vision Research Lab. [7]		Deep learning-based detection of malicious code variants
CICAndMal 2017 [8]	10.854 samples	Artificial intelligence-based malware detection using deep learning
1. MalImg dataset 2. Microsoft Malware Classification Challenge dataset.[9]	1. 9,435 samples 2. 21,741 samples	Convolutional neural networks for classification of malware represented as images
Malimg dataset[10]	9,435 samples	Image Visualization based Multiclass Malware Classification using Transfer Learning
1. MalImg dataset 2. Microsoft BIG dataset[11]	1. 9339 malware samples of 25 families 2. 10868 malware samples of 9 families	Deep transfer learning for malware image classification
1.MNIST dataset [12]	Not mentioned	Comparative study of capsule neural network in various applications
Malimg dataset[13]	9,435 samples	Deep Convolutional Neural Networks for malware classification

Table 1: Representation of various datasets used in different articles

4. Data Preprocessing

Image based malware samples available is not suitable for classification without preprocessing. The raw data is processed accordingly, and made suitable for the classification process. The image is resized to 200x200x1. Few researchers used RGB images for classification. They converted the original grey scale image to RGB mode. In this study we apply grey scale image with 200x200x1image pixel. Thus, processed image is been augmented to increase the dataset size. Hence forth the dataset is increased. The final size of malware image sample is 27890. Fig. 3 is the representation of few augmentation technique applied in the malware sample. These samples are preprocessed with few features engineering task like edge detection, noise removal etc.

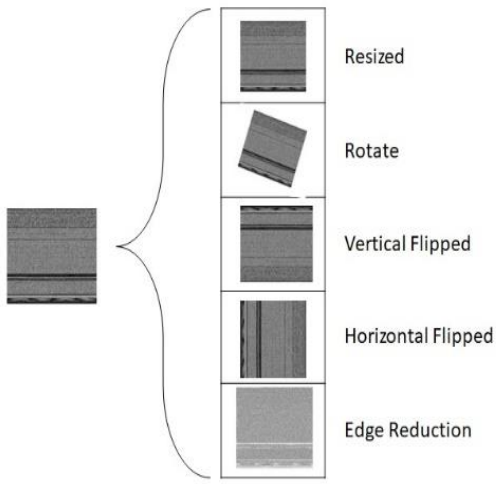

Fig. 3 Representation of Augmentation of malware image.

5. Model Selection

In this study we completely focus on comparison between the performance of two novel approaches Mal_CNN and Mal_CapsNet. This study also includes a detailed comparison with existing classifiers like CNN and Capsule neural network. Fig. 4 is the representation of malicious image classification through neural network. the images in dataset undergoes augmentation, segmentation and feature engineering task. Thus, calibrated images are applied in neural network classifiers and classification is done. The trained model classifies the image belongs to a malicious family or the benign family. Fig 6 and Fig 7 represents the Novel CNN and Capsule neural network framework.

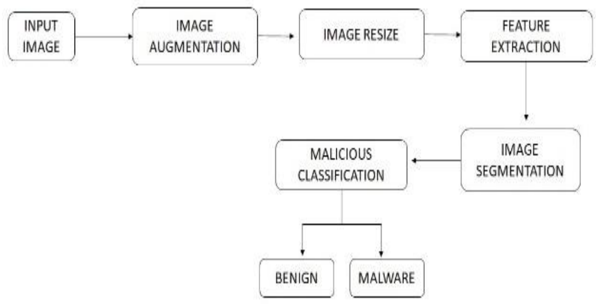

Fig 4 Representation of Malicious classification over feature extraction and segmentation

5.1 MAL_CNN Framework

This Mal CNN is the framework developed based on the convolution neural network framework. This classifier works better compared to the other machine learning and deep learning classifiers. In this Mal_CNN framework a Convolution layer of 2 dimension and an activation function relu in added at the end of each convolution layer was implemented. 2 dimensional Maxpooling layer was also upskilled. Three layers of dense network was applied. This novelty in the framework enhances the classification of malicious image and benign images. The Early stopping techniques was also incorporated to avoid overfitting. Both the segmentation and feature engineering were applied.

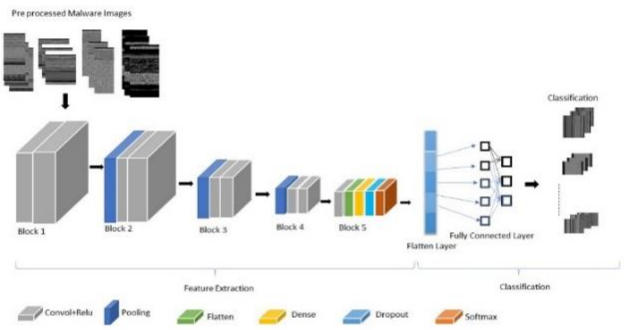

Fig. 5 Representation of MAL_CNN framework

5.2 MAL_CAPSNET Framework

Mal Capsnet is the novel approach applied in the malicious image classification. This framework helps in classification of malicious image with benign images. two level of addon convolutional layer along with the existing convolution layer. This enhances the feature extraction. The enhanced layer acts with the Relu squash. This outperforms by collecting the low-level features from the input image and sends to the next convolution layer. The striding is fixed to be 2 without padding. The next convolution layer is liked to primary capsule layer.

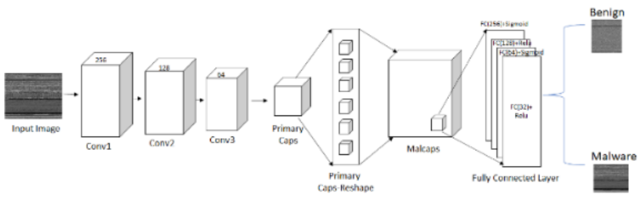

Fig. 6 Representation of Mal_CapsNet framework.

6. Result and Discussion

The Malimg dataset after enhancement amount to 27890 malicious image samples. Those images were used as input for the four neural network frameworks. In this study we proposed a comparison study of the four frameworks. The Frameworks CNN, Capsule neural network, Mal_CNN and Mal_CapsNet were pre-owned frameworks. Once the input is given to the framework, it turns out with classification. The frame works classifies the given image into malicious or a benign type. Benign samples were used for better comparison. Outcome of the framework is analyzed

Graph is been used for representing the training loss and validation accuracy through Mal_CNN and Mal_CapsNet architecture when applied with Malimg dataset.

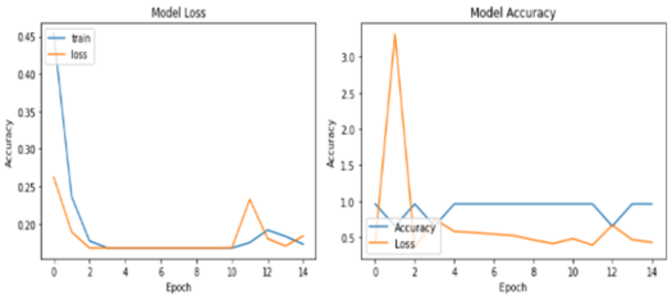

Fig. 7 Representation of MAL_CNN

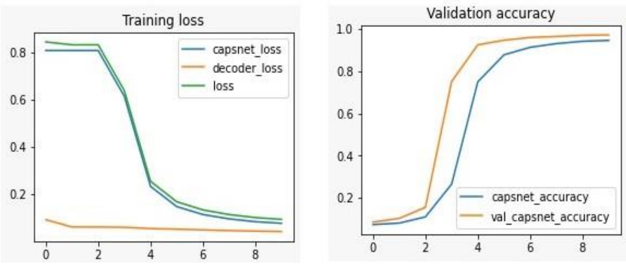

Fig.8 Representation of MAL_CAPSNET

Models manifest in this article are measure using three metrics Recall, precision and accuracy. Positive notation is for malware image and negative represents benign image. TP is the correctly classified malware samples from the dataset; TN mention number benign samples classified; While, FP is the malware samples that are not rightly classified. FN is the count of benign samples that are not classified correctly.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Recall = \frac{TP}{TP+FP} \quad (2)$$

$$Precision = \frac{TP}{TP+FN} \quad (3)$$

Table 2. represents the comparative analysis exist between the results acquired in the various four neural network concept. The classifiers of neural network used were CNN, Mal_CNN, Capsule Neural Network .and Mal_ CapsNet. All the four framework was trained and used to classy the malware with the same enhanced dataset from Malimg.

	Recall	Precision	Accuracy
CNN	0.86	0.86	92%
Mal_CNN	0.86	0.87	92%
CapsNet	0.95	0.96	96%
Mal_CapsNet	0.96	0.97	97.6%

Table 2: Representation of Comparative analysis between various classifiers.

7. Conclusion

This study deals with the image-based malware classification. The neural algorithms CNN, Capsule neural network and novel approaches Mal_CNN and Mal_CapsNet are used to classify the image as malware or benign image. The enhanced image data from Malimg dataset is fed as input to the models. The comparative analysis of the framework is discussed. The model accuracy is estimated through the performance metrics like recall, precision and accuracy. Mal_Capsule network performs better in this above study. This study may benefit the researchers to further classify the image based malicious dataset. This study may be extended by using various other malicious image-based datasets for classification.

Reference

[1] Nataraj, L., Karthikeyan, S., Jacob, G., & Manjunath, B. S. (2011, July). Malware images: visualization and automatic classification. In Proceedings of the 8th international symposium on visualization for cyber security (pp. 1-7).

[2] Ding, Y., Zou, D., Wang, S., & Zhang, Y. (2020). An efficient malware classification method using feature extraction based on deep neural network. Journal of Ambient Intelligence and Humanized Computing, 11(3), 1245-1257.

[3] Nisa, M., Shah, J. H., Kanwal, S., Raza, M., Khan, M. A., Damaševičius, R., & Blažauskas, T. (2020). Hybrid malware classification method using segmentation-based fractal texture analysis and deep convolution neural network features. Applied Sciences, 10(14), 4966.

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

[5] Bensaoud, A., Abudawaood, N., & Kalita, J. (2020). Classifying malware images with convolutional neural network models. International Journal of Network Security, 22(6), 1022-1031.

[6] Zhang, X., Wu, K., Chen, Z., & Zhang, C. (2021). MalCaps: A capsule network based model for the malware classification. Processes, 9(6), 929.

[7] Manoharan, J. S. (2021). Capsule network algorithm for performance optimization of text classification. Journal of Soft Computing Paradigm (JSCP), 3(01), 1–9..

[8] Z. Cui, F. Xue, X. Cai, Y. Cao, G. -g. Wang and J. Chen, "Detection of Malicious Code Variants Based on Deep Learning," in IEEE Transactions on Industrial Informatics, vol. 14, no. 7, pp. 3187-3196, July 2018, doi: 10.1109/TII.2018.2822680.

[9] Majid, A. A. M., Alshaibi, A. J., Kostyuchenko, E., & Shelupanov, A. (2021). A review of artificial intelligence based malware detection using deep learning. Materials Today: Proceedings

[10] Gibert, D., Mateu, C., Planes, J., & Vicens, R. (2019). Using convolutional neural networks for classification of malware represented as images. Journal of Computer Virology and Hacking Techniques, 15(1), 15-28

[11] Goyal, M., & Kumar, R. (2022). IVMCT: Image Visualization based Multiclass Malware Classification using Transfer Learning. Mathematical Statistician and Engineering Applications, 71(2), 42-50.

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

[13] Vijayakumar, T. (2019). Comparative study of capsule neural network in various applications. Journal of Artificial Intelligence, 1(01), 19-27.

[14] M. Kalash, M. Rochan, N. Mohammed, N. D. B. Bruce, Y. Wang and F. Iqbal, "Malware Classification with Deep Convolutional Neural Networks," 2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS), Paris, France, 2018, pp. 1-5, doi: 10.1109/NTMS.2018.8328749.

[15] Chou, P. C., Chang, C. C., & Su, K. W. (2019). A Study on Malware Classification using Capsule Neural

Network. In 2019 IEEE International Conference on Information Networking (ICOIN) (pp. 45-50). IEEE.

[16] Ijaz, M., Durad, M. H., & Ismail, M. (2019, January). Static and dynamic malware analysis using machine learning. In 2019 16th International bhurban conference on applied sciences and technology (IBCAST) (pp. 687-691). IEEE

[17] Jang, J., Kim, J., Kang, H., & Kim, H. (2020). Malware classification using image-based features extracted from dynamic malware analysis. Journal of Information Security and Applications, 53, 102453.

[18] Mahesh, B. (2020). Machine learning algorithms-a review. International Journal of Science and Research (IJSR).[Internet], 9, 381-386.

[19] Sihwail, R., Omar, K., & Ariffin, K. Z. (2018). A survey on malware analysis techniques: Static, dynamic, hybrid and memory analysis. Int. J. Adv. Sci. Eng. Inf. Technol, 8(4-2), 1662-1671.

[20] Dr. Naveen Jain. (2020). Artificial Neural Network Models for Material Classification by Photon Scattering Analysis. International Journal of New Practices in Management and Engineering, 9(03), 01 - 04. https://doi.org/10.17762/ijnpme.v9i03.88

[21] Dr. Anasica S, Mrs. Sweta Batra. (2020). Analysing the Factors Involved In Risk Management in a Business. International Journal of New Practices in Management and Engineering, 9(03), 05 - 10. https://doi.org/10.17762/ijnpme.v9i03.89