

# THE USAGE OF GENERATIVE ADVERSARIAL NETWORK TO RESOLVE MALWARE DATA IMBALANCE FOR CLASSIFICATION

Nazri A. Zamani<sup>1,2</sup>, Siti S. Yuhaniz<sup>2\*</sup>, Aswami Ariffin<sup>1</sup>

<sup>1</sup>CyberSecurity Malaysia, Cyberjaya, Selangor, Malaysia

<sup>2</sup>Razak Faculty of Technology & Informatics, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia

## ABSTRACT

Malware detection and classification with image visualization of malware binary files and Machine Learning or Deep Learning is getting more attention by cybersecurity researchers. As image classification is already a proven viable solution to many imaging applications, the same is expected for malware binary files. The implementation of Deep Convolutional Neural Networks (DCNN) for malware image classification has shown convincing results. One issue still stands where the nature of data imbalance in malware dataset has caused overfitting in the malware image Deep Learning training, in which hugely affect the overall performance of the classification. This paper presents the use of Deep Convolutional Generative Adversarial Networks (DCGAN) to solve low samples number issues in certain malware family classes by generating identical adversarial images that unique to the specific classes. The implementation of DCGAN shows significant improvement on the DCNN classification accuracy and log-loss from 70.22% and 2.5388 to 91.78% and 0.2379.

## METHODOLOGY

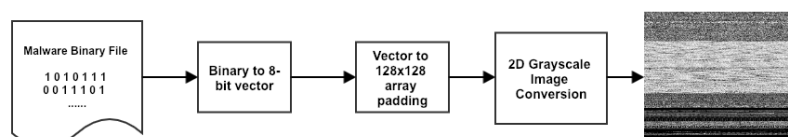
### Datasets

For the experiments, the malware dataset from Microsoft Malware Classification Challenge (BIG 2015) is used, in which the data is distributed via a competition hosted on Kaggle [7]. The datasets is 500 gigabytes worth of malware data, in which consist of 10868 and 10873 malwares in train and test set, respectively. The malware data are categorized into nine classes as shown in **Table 1**. With each of the labelled malware sample is provided with a hexadecimal byte and an assembly file which is produced by IDA Pro disassembler software.

Table 1. Malware classes and the associated number of samples in each class

Malware class	Number of Samples
Ramnit	1553
Lollipop	2478
Kelihos_ver3	2942
Vundo	475
Simda	42
Tracur	751
Kelihos_ver1	398
Obfuscator.ACY	1228
Gatak	1013

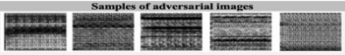
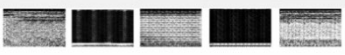



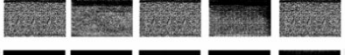



### Pre-processing



### Malware classification

The system is designed as a two-tiers Deep Learning in generating malware adversarial images for each malware class as a strategy to generate more data for malware classes that lost in term of having lower number of samples. The first tier is the *learn-generate* where Deep Learning learns malware images and subsequently generates adversarial malware images. These adversarial malware images are then added back to their respective original classes. These new enriched malware image classes are then subjected to the next tier of Deep Learning. At the *learn-classification* tier, these malware data will be cognitively learned and then tested for classification performance.

## RESULTS

Class	# of samples	# of generated	Samples of adversarial Images
Ramnit	1553	1663	
Lollipop	2478	2578	
Kelihos_ver3	2942	3042	
Vundo	475	575	
Simda	42	184	
Tracur	751	851	
Kelihos_ver1	398	498	
Obfuscator.ACY	1228	1328	
Gatak	1013	1113	

## CONCLUSIONS

In this paper, the proposed two-tiers Deep Learning has managed to increase the accuracy of the DCNN classification of malware images from 70.22% and 2.5388 (in the original imbalance dataset) to 91.78% and 0.2379 by the introduction of adversarial malware images generated by DCGAN. The appeal of this method is through the engineering of Deep Learning processes, the research manages to solve the data imbalance without involving any oversampling methods.