# MALWARE CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORK TO IMPROVE INDONESIAN GOVERNMENT CYBERSECURITY

A THESIS

A Partial Requirement to Fulfill for Master Degree in Computer Science


RACHMAD DWI KURNIAWAN
20210130039

Supervisor:
Prof. Teddy Mantoro, MSc., PhD., SMIEE
Harris Al qodri Maarif, S.T., M.Sc., Phd

**SCHOOL OF COMPUTER SCIENCE**
**NUSA PUTRA UNIVERSITY**

## STATEMENT OF AUTENTICITY

The undersigned below:

| | |
|---|---|
| Name | : Rachmad Dwi Kurniawan |
| ID of student | 20210130039 |
| Faculty | : Computer Science |
| The Tittle of Thesis | : Malware Classification Using Convolutional Neural Network to Improve Indonesian Government Cybersecurity |

Stating truthfully that this thesis has nothing in common with other thesis. Thus this statement is made without coercion from any party. If this statement is not true, it will be sanctioned by the faculty leader.



Sukabumi, Januari 2024

Writer,

*Materai 10000*

Rachmad Dwi Kurniawan
20210130039

**APPROVAL OF THESIS**

| | |
|---|---|
| The Tittle of Thesis | : Malware Classification Using Convolutional Neural Network to Improve Indonesian Government Cybersecurity |
| Name | : Rachmad Dwi Kurniawan |
| ID of student | 20210130039 |

This thesis has been reviewed and approved

Sukabumi, Januari 2024

| Head of Study Program, | Supervisor, |
|---|---|
| Prof. Ir.Teddy Mantoro, MSc., PhD., SMIEEE | Prof. Ir.Teddy Mantoro, MSc., PhD., SMIEEE |
| NIDN. 0323096491 | NIDN. 0323096491 |

**THESIS APPROVAL**

| | |
|---|---|
| The Tittle of Thesis | : Malware Classification Using Convolutional Neural Network to Improve Indonesian Government Cybersecurity |
| Name | : Rachmad Dwi Kurniawan |
| ID of student | 20210130039 |

This Thesis has been tested and defended in front of the Board of Examiners in Thesis session on Januari 14th, 2024. In our review, this Thesis adequate in terms of quality for the purpose of awarding the Master of Computer Degree.

| | |
|---|---|
| Supervisor 1 | Examiner 1 |
| Prof. Ir.Teddy Mantoro, MSc., PhD., SMIEEE<br>NIDN. 0323096491 | Deshinta Arrova Dewi, S.Kom., M.Si., PhD<br>NIDN. 0416127102 |
| Supervisor 2 | Examiner 2 |
| Harris Al qodri Maarif, S.T., M.Sc., Phd<br>NIDN. 0418068505 | Prof. Ir. Media Anugera Ayu, M.Sc., PhD<br>NIDN. 0315046903 |

## PUBLICATION APPROVAL

As a member of the academic community of Nusa Putra University Iundersigned:

| | |
|---|---|
| Name | : Rachmad Dwi Kurniawan |
| ID of student | 20210130039 |
| Study Program | : Computer Science |
| Type of Work | : Thesis |

For the sake of scientific development, agree to grant to the University of Nusa Putra the Non-Exclusive Royality-Free Right for my scientific work entitled: **Malware Classification Using Convolutional Neural Network to Improve Indonesian Government Cybersecurity**. With this non-exclusive royalty-free right, Nusa Putra University has the right to store, transfer media/formats, process in the form of a database, maintain and publish my thesis as long as I keep my name as the author/creator and as the copyright owner. This statement I made in truth.

Made in : Sukabumi

At the Date of : Januari 2024

That States

(Rachmad Dwi Kurniawan)

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURE

## FOREWORD

All praises to Allah SWT for His infinite mercy and guidance, enabling the author to complete this Thesis with the title "Malware Classification Using Convolutional Neural Network to Improve Indonesian Government Cybersecurity." This study serves as a requirement for completing the Masters Program at the SoCS, Nusa Putra University. The successful completion of this final project was made possible due to the guidance, assistance, prayers, and cooperation of various parties.

The author realizes that the Final Project research this is still far from perfection, so the author expects criticism and suggestions from readers. I am hopeful that this research will contribute significantly towards creating a safer cyberspace by helping to classify and mitigate the threat posed by malware. As you delve into this thesis, I hope you find it informative, enlightening, and a useful contribution to the field of cybersecurity.

## ABSTRACT

As the dependence on information technology and cyberspace intensifies across various critical sectors, it also presents a potential medium for cyber attacks. Particularly, the rapidly evolving nature of malware, with its myriad variants displaying distinct characteristics, often leaves victims grappling with effective mitigation. Current malware detection technologies, although numerous, are often deemed insufficient due to their inability to provide robust classification, an essential aspect that facilitates efficient cybersecurity analysis. This research addresses this pressing issue by harnessing Convolutional Neural Networks (CNNs) Inception-V3 for the classification of malware, with an aim to enhance the speed and effectiveness of malware mitigation efforts, especially in Indonesian Government. CNNs, subtypes of artificial neural networks, are mainly used for visual data examination. They employ a mathematical operation known as convolution in one or more of their layers, making them particularly adept at handling pixel data for tasks such as image recognition, processing, and classification. This research serves as a significant stride towards to improve Indonesian Government Cybersecurity on dealing with the cyber threat and reducing dependence on the use of signature-based malware detection which is considered to have many weaknesses.



Keywords: Malware Classification, Convolutional Neural Network

# CHAPTER I
# INTRODUCTION

## 1.1 Research Background

With the advancement of time, the use of information technology and cyberspace is increasing and widely used in critical sectors like government, military, industry, healthcare, manufacturing, education, and others. Besides being highly beneficial and making work easier, the use of information technology and cyberspace also has the potential to be a gap or source of cyber attacks.

It is not uncommon for victims of malware attacks to encounter difficulties in mitigating malware attacks, due to the continuous emergence and development of new variants of malware with different characteristics, each requiring different mitigation steps. Although there are currently many malware detection technologies, these devices are still considered unable to perform proper classification that facilitates cyber security analysis.

Despite the prevalence of many malware detection technologies today, these devices are often considered ineffective in conducting accurate classifications, largely due to their dependence on signature-based analysis. The drawback of dependence on signature-based analysis is that malware can only be identified or detected if the signature of the malware has been previously recorded. These signatures include file hashes, IP addresses of the attack source, file names, and so on.

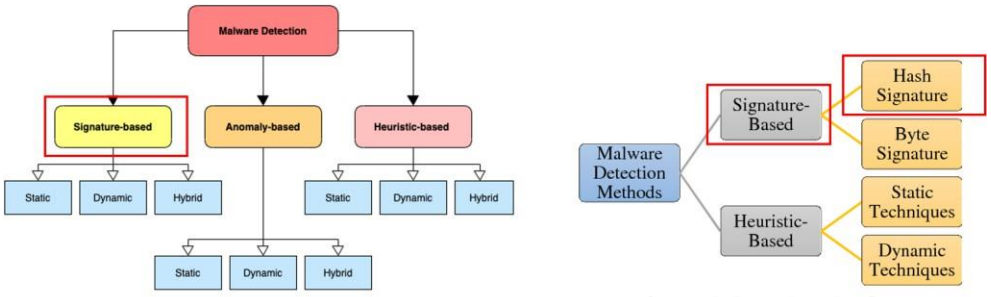


*Figure 1.1.1 Various Types of Malware Detection Methods*

Indonesian Government Network Infrastructure dealing with several malware stealer infection that targeting endpoint devices. Malware stealer is one of the type of malware that have capability to retrive account login credential data (username and password). Thus, attackers can carry out advanced cyber attacks, such as theft of important data, espionage and sabotage.

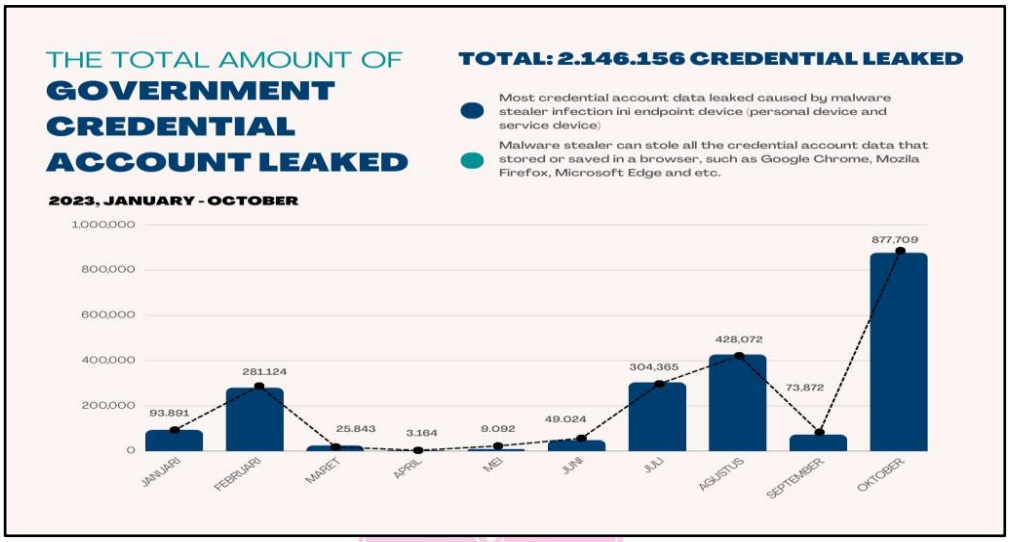


*Figure 1.1.2 Graph Showing the Total Amount of Government Credential Account Leaks Caused by Stealer Malware Infections.*

In the periode January to October 2023, there have been as many as 2,146,156 data breaches involving the credential information of login accounts for Indonesian government application websites, caused by infections/attacks from stealer malware.

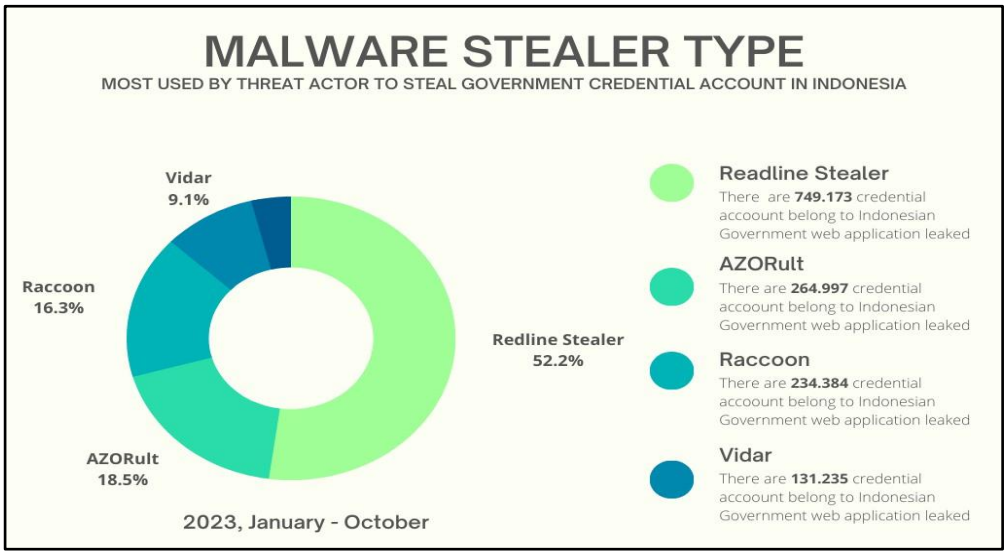


*Figure 1.1.3 Most Common Malware Stealer Infecting Devices Owned by the Indonesian Government*

It was noted that there are several stealer malwares most commonly used to steal credential data of login accounts for Indonesian government applications. One type of stealer malware that has most frequently infected devices owned by the Indonesian government is the RedlineStealer Malware. At least 749,173 government login account credentials have been successfully stolen by the RedlineStealer Malware.

The RedlineStealer malware infection is considered challenging to identify and detect because it has more than 750 different file hashes and varying Command and Control (CnC) IP addresses or sources of attack. Moreover, new RedlineStealer malware file hashes continue to emerge, making it difficult for signature-based identification/detection systems to keep up effectively.

Based on this, there is a need to develop a malware classification and identification system using Artificial Intelligence technology. Classification can be performed by converting malware files into binary and then transforming them into images.



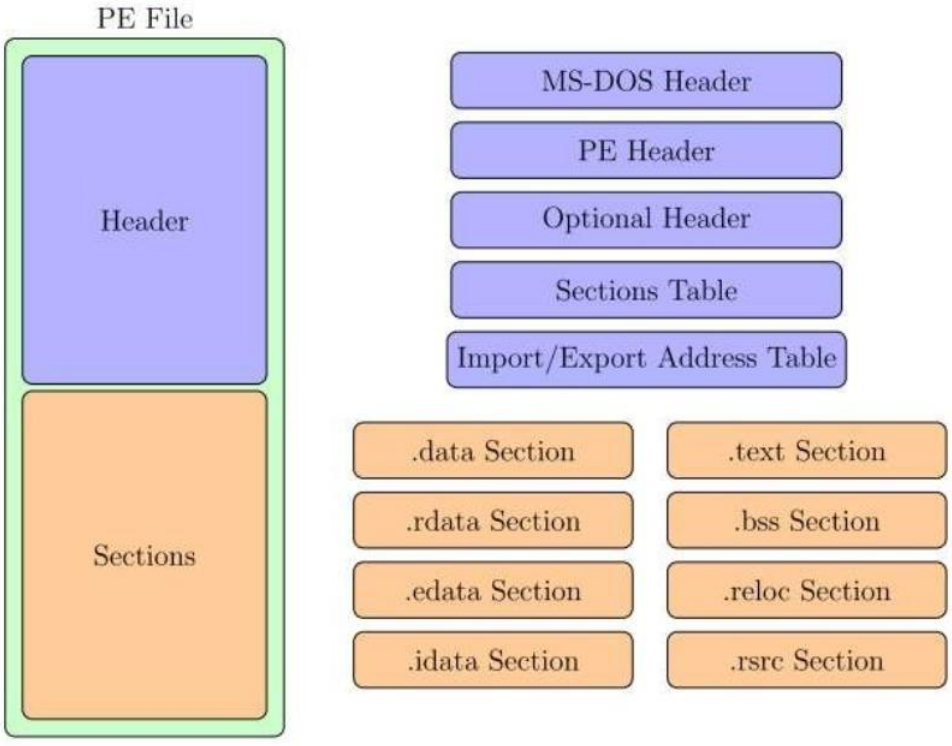


*Figure 1.1.4 PE (Portable Executable) File Structure (.exe).*

PE Files or .exe files contain several sections within them that store specific information and usually have distinctive characteristics from one file type to another. If a PE file is converted into an image, the image will still retain some characteristics that can be distinguished from others.

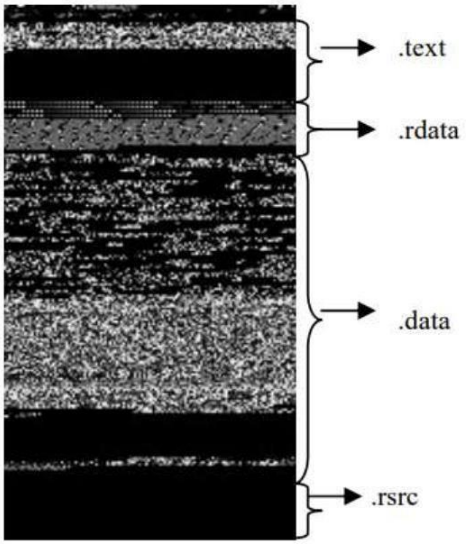


*Figure 1.1.5 Malware Image File Structure*

The following is an image of malware in .png format, which turns out to contain parts of the header including .text, .rdata, .data, and .rsrc sections. These images will then be classified to determine the type of malware family. Therefore, the author conducts research to classify malware using Convolutional Neural Network Models to assist cyber security analysts in mitigating malware attacks more quickly and effectively, Specifically regarding stealer malware, one of the types is the Redline Stealer Malware.

Thus, this research will focus on applying Convolutional Neural Network Models for malware classification. This work intends to bridge the gap in existing malware detection and classification methodologies, bolstering our defenses against the incessant barrage of cyber threats. By doing so, it is expected to contribute substantially to empowering cyber security analysts to navigate the complex terrain of malware attacks more adeptly, swiftly, and efficiently.

Ultimately, this research represents a forward-looking attempt to mitigate one of the most pressing cyber threats of our time, thereby making significant strides towards improve Indonesian Government Cyber Security.

## 1.2 Problem Statement

While the use of information technology and cyberspace is accelerating across various critical sectors, it simultaneously poses a potential conduit for cyber attacks. The constantly evolving landscape of malware, exhibiting new variants with unique characteristics, often leaves victims struggling with effective mitigation.

Despite numerous existing malware detection technologies, they are frequently deemed inadequate due to their inability to provide robust classification that facilitates cybersecurity analysis. This research aims to address this critical gap by employing Convolutional Neural Network Models for the classification of malware, thus enhancing the speed and effectiveness of malware mitigation efforts.

The problem, therefore, is to develop and validate a model that can accurately classify malware types, assisting cybersecurity analysts in combating these cyber threats more efficiently.



## 1.3 Research Objectives

a. Conduct training on malware image data using the Convolutional Neural Network (CNN) algorithm and assess the accuracy level of the resulting model.
b. Be able to classify malware files based on their types using CNN modeling. One of the malware types originates from the highest malware stealer detection results within the Indonesian Government's network infrastructure during the period of 2023.

## 1.4 Significance of Research

a. Improved Malware Identification and Classification: By utilizing Convolutional Neural Network Models, the study aims to develop and validate a more accurate and efficient system for identifying the type of

malware. This has the potential to improve the speed and effectiveness of malware mitigation efforts, thereby enhancing overall cybersecurity.

b. Empowering Cybersecurity Analysts: This research is vital in assisting cybersecurity analysts to better navigate the rapidly evolving landscape of cyber threats. By providing a more accurate classification model, it equips analysts with more precise information about malware attacks, enabling a more targeted, efficient, and swift response. This, in turn, can substantially improve the cybersecurity posture of critical sectors heavily reliant on information technology and cyberspace.

## 1.5 Limitation of Problems and Assumptions

The focus of this research is to implement the use of Convolutional Neural Network (CNN) in classifying malware, sourced from malware detection results within several Indonesian government network infrastructures. One type of malware that has most frequently infected devices owned by the Indonesian government is the RedlineStealer Malware. So, in this study, specific data on Redline Stealer malware will be added, which will then undergo training so that the Indonesian government network infrastructure cybersecurity can perform classification and detection of stealer malware by utilizing AI, rather than relying solely on signature-based detection, which is considered to have many shortcomings.

The outcome of this research is expected to assist Cyber Security analysts or government network infrastructure administrators in classification and identifying malware cyber attacks.

## 1.6 Thesis Structure

The rest of thesis is organized as follows:

a. Chapter I describes the background of problem that will be discussed in the thesis.
b. Chapter II describes the literature review of thesis.
c. Chapter III describes the methodology of thesis.
d. Chapter IV presents the expereiment result and discussion.
e. Chapter V presents the conclusion the thesis and future work.

Nusa Putra
UNIVERSITY
2002
AMOR DEUS, PARENTIUM, CONCERVIS

# CHAPTER V
# CONCLUSIONS AND RECOMMENDATIONS

## 5.1 Conclusions

a. In conclusion, this research presents a significant stride towards more efficient and accurate malware classification. The results obtained from our study suggest that the use of CNNs in classifying malware can potentially transform the way cybersecurity operations are conducted. However, the model's continuous improvement, evaluation, and adaptation to new malware types will remain a critical future task for this ongoing research.
b. A total of 25 types of malware used in this study can be accurately classified, including the Redline Stealer malware, which is one of the most frequently detected malware infecting endpoint devices owned by the Indonesian government, leading to the leakage of login credential data for government applications

## 5.2 Recommendations

a. Increasing the quantity of training datasets sourced especially from the latest type of Malware;
b. Enhancing and expanding the model with additional analytical features to improve accuracy;
c. Incorporating datasets from malware detected on the Indonesian government's network infrastructure, thus enabling the model to predict new types of malware, particularly those targeting government network infrastructure. This would also contribute to improving the Indonesian Government Cybersecurity

## REFERENCES


[1] Kalash, M., Rochan, M., Mohammed, N., Bruce, N., Wang, Y., and Iqbal, F. "A deep learning framework for malware classification." International Journal of Digital Crime and Forensics, vol. 12, no. 1, pp. 90–108, Jan. 2020. doi: 10.4018/IJDCF.2020010105.

[2] Abudawaood, N., Bensaoud, A., and Kalita, J. "Classifying Malware Images with Convolutional Neural Network Models." International Journal of Network Security, Oct. 2020. doi: 10.6633/IJNS.202011_22(6).17.

[3] Nazeer, A. K. "Malware Classification Using Deep Learning." [Online]. Available: https://ssrn.com/abstract=4296051

[4] Dilhara, B. A. S. "Classification of Malware using Machine learning and Deep learning Techniques." International Journal of Computer Applications, vol. 183, no. 32, pp. 12–17, Oct. 2021. doi: 10.5120/ijca2021921708.

[5] Saxe, J., and Berlin, K. "Deep Neural Network-Based Malware Detection Using Two Dimensional Binary Program Features." In Proceedings of the 10th International Conference on Malicious and Unwanted Software (MALWARE), pp. 20-22, Oct. 2015.

[6] Vinayakumar, R., Alazab, M., Soman, S., Poornachandran, P., and Sagayam, K. "Evaluating Deep Learning Approaches to Characterize and Classify the DGAs at Scale." Journal of Intelligent & Fuzzy Systems, vol. 34, no. 3, pp. 1265-1276, Mar. 2018.

[7] Yuxin, M., Zhuang, Y., and Zhang, X. "Deep Learning-Based Malware Analysis Techniques: A Comprehensive Survey." IEEE Access, 2021.

[8] Al-Dujaili, A., Huang, L., Zhao, L., and Sakurai, K. "Using Convolutional Neural Networks for Malware Classification and Analysis." Journal of Computer Virology and Hacking Techniques, vol. 15, pp. 15–28, Mar. 2019.

[9] Bensaoud, A. "Malware Detection using Deep Learning." Journal of Information Security and Applications, 2020.

[10] Z. Yuan, Y. Lu, and Y. Xue, "Automated Malware Classification using Deep Learning," Frontiers of Information Technology & Electronic Engineering, vol. 18, no. 9, pp. 1336-1347, 2017.

[11] Carlo Garcia, F. C., and Muga II, F. P. "Random Forest for Malware Classification." International Journal of Computer Applications, 2020.

[12] Gibert, D., Mateu, C., Planes, J., and Vicens, R. "Using convolutional neural networks for classification of malware represented as images." Journal of Computer Virology and Hacking Techniques, vol. 15, no. 1, pp. 15–28, Mar. 2019. doi: 10.1007/s11416-018-0323-0.

[13] Yeboah-Ofori, A. "Classification of Malware Attacks Using Machine Learning In Decision Tree." Journal of Cybersecurity, 2020.

[14] Nugraha, A., and Zeniarja, J. "Malware Detection Using Decision Tree Algorithm Based on Memory Features Engineering." 2022. [Online]. Available: Malware Memory Analysis | Datasets | Canadian Institute for Cybersecurity | UNB.

[15] Doshi, R., Yilmaz, E., and Kesidis, G. "A practical deep learning-based approach for malware detection," Journal of Network and Computer Applications, vol. 170, pp. 102827, May 2020. doi: 10.1016/j.jnca.2020.102827.

[16] Shijo, P., and Salim, A. "Malware detection using deep learning techniques: A review," Journal of Cybersecurity and Privacy, vol. 1, no. 1, pp. 72-103, 2021. doi: 10.3390/jcp1010005.
[17] Kolosnjaji, B., Demontis, A., Biggio, B., Maiorca, D., Giacinto, G., and Roli, F. "Adversarial Malware Binaries: Evading Deep Learning for Malware Detection in Executables," Proceedings of the 26th European Signal Processing Conference (EUSIPCO), pp. 533-537, 2018. doi: 10.23919/EUSIPCO.2018.8553418.
[18] Coull, S. E., and Gardner, K. "Activation analysis of a convolutional neural network-based malware classifier," Proceedings of the 32nd Annual Conference on Neural Information Processing Systems (NeurIPS), pp. 169-180, 2019.
[19] Firdausi, I., Erwin, A., Nugroho, A. S., and Jo, K. "Analysis of machine learning techniques used in behavior-based malware detection," Proceedings of the 2010 Second International Conference on Advances in Computing, Control, and Telecommunication Technologies (ACT), pp. 201-203, 2020. doi: 10.1109/ACT.2010.46.
[20] Cui, Y., Xie, G., Xu, L., Ma, Z., and Chen, Y. "Malware detection based on deep learning algorithm," Neural Computing and Applications, vol. 32, no. 4, pp. 2159-2171, April 2020. doi: 10.1007/s00521-019-04080-7.
[21] Yuan, Z., Lu, Y., Xue, Y., and Zhao, S. "Ensemble deep learning for automated malware detection," IEEE Transactions on Information Forensics and Security, vol. 13, no. 11, pp. 2870-2883, Nov. 2018. doi: 10.1109/TIFS.2018.2834900.
[22] Ahmadi, M., Ulyanov, D., Semenov, S., Trofimov, M., and Giacinto, G. "Novel feature extraction, selection and fusion for effective malware family classification," Proceedings of the 6th ACM Conference on Data and Application Security and Privacy (CODASPY), pp. 183-194, 2020. doi: 10.1145/2857705.2857732.
[23] Shafiq, M., Tian, Y., Bashir, A. K., Du, X., and Guizani, M. "IoT malicious traffic identification using wrapper-based feature selection mechanisms," Computers & Security, vol. 94, pp. 101863, May 2020. doi: 10.1016/j.cose.2020.101863.
[24] Su, J., Liu, D., and Sahin, G. "Lightweight classification of IoT malware based on image recognition," IEEE Access, vol. 8, pp. 77746-77756, April 2020. doi: 10.1109/ACCESS.2020.2989083.
[25] Nataraj, L., Karthikeyan, S., Jacob, G., and Manjunath, B. S. "Malware images: visualization and automatic classification," Proceedings of the 8th International Symposium on Visualization for Cyber Security (VizSec), pp. 1-7, 2011. doi: 10.1145/2016904.2016908.
[26] Zelinka, I., Szczypka, M., and Kuznetsov, N., "From malware samples to fractal images: A new paradigm for classification," Mathematics and Computers in Simulation, vol. 202, pp. 57-65, Apr. 2024.
[27] Vasan, D., Hammoudeh, M., and Alazab, M., "Broad learning: A GPU-free image-based malware classification," Applied Soft Computing, vol. 116, pp. 102325, Mar. 2024.

[28] Bensaoud, A., and Kalita, J., "CNN-LSTM and transfer learning models for malware classification based on opcodes and API calls," Knowledge-Based Systems, vol. 239, pp. 107724, Apr. 2024.
[29] Kumar, S., Shersingh, S., and Verma, K., "Malware Classification Using Machine Learning Models," Procedia Computer Science, vol. 202, pp. 45-54, 2024.

Nusa Putra
UNIVERSITY
2002
AMOR DEUS, PARENTIUM, CONCERVIS