PENERAPAN METODE CONVOLUTIONAL NEURAL NETWORK TERHADAP CITRA *GRAYSCALE* DAN *PSEUDOCOLOR* RGB UNTUK KLASIFIKASI *MALWARE*

TUGAS AKHIR

Chris Christian
1118041

INSTITUT
TEKNOLOGI
HARAPAN
BANGSA
Veritas vos liberabit

PROGRAM STUDI INFORMATIKA
INSTITUT TEKNOLOGI HARAPAN BANGSA
BANDUNG
2022

PENERAPAN METODE CONVOLUTIONAL NEURAL NETWORK TERHADAP CITRA *GRAYSCALE* DAN *PSEUDOCOLOR* RGB UNTUK KLASIFIKASI *MALWARE*

TUGAS AKHIR

Diajukan sebagai salah satu syarat untuk memperoleh
gelar sarjana dalam bidang Informatika

Chris Christian
1118041

INSTITUT
TEKNOLOGI
HARAPAN
BANGSA
Veritas vos liberabit

PROGRAM STUDI INFORMATIKA
INSTITUT TEKNOLOGI HARAPAN BANGSA
BANDUNG
2022

ABSTRAK

Nama	:	Chris Christian
Program Studi	:	Informatika
Judul	:	Penerapan Metode Convolutional Neural Network terhadap Citra *Grayscale* dan *Pseudocolor* RGB untuk Klasifikasi *Malware*

Malware atau *malicious software* merupakan perangkat lunak yang sengaja dibuat untuk merusak fungsionalitas sistem komputer. Pada era internet saat ini, terdapat peningkatan yang signifikan pada jumlah malware dalam tahun-tahun belakangan. Oleh karena itu, diperlukan sistem yang dapat mengidentifikasi dan mengklasifikasikan sampel *malware*.

Penelitian ini melakukan eksperimen untuk melakukan klasifikasi *malware* menggunakan visualisasi citra *grayscale* dan citra RGB dengan metode Convolutional Neural Network (CNN). *Dataset* yang digunakan berupa visualisasi citra *grayscale* yang dikonversi dari *file executable*, dan terdiri dari 9339 citra dan 25 *family*. Sedangkan, citra RGB diperoleh dari konversi citra *grayscale* menjadi citra berwarna dengan teknik *pseudocolor*. Sebelum citra malware diklasifikasi, Principal Component Analysis (PCA) digunakan untuk menangani tingginya dimensi data dan meningkatkan akurasi klasifikasi dengan mengekstrak fitur-fitur penting.

Hasil penelitian menunjukkan model Convolutional Neural Network mencapai akurasi tertinggi 97.752% dan *recall* tertinggi 95.026% dengan citra *grayscale malware*. Sementara itu, untuk citra RGB *malware*, model Convolutional Neural Network mendapatkan akurasi terbaik 97.591% dan *recall* terbaik 94.824%. Berdasarkan hasil penelitian, klasifikasi *malware* pada kedua jenis citra memperoleh hasil yang cukup baik.

Kata kunci: *Cybersecurity*, Klasifikasi Gambar, *Malware*, Convolutional Neural Network (CNN).

ABSTRACT

Name	:	Chris Christian
Department	:	*Informatics*
Title	:	*Application of Convolutional Neural Network for Grayscale and Pseudocolor RGB Image to Classify Malware*

Malware or malicious software is any software intentionally made to harm the function of a computer system. In this internet era, there has been a significant increase in the volume of malware in recent years. Therefore, it is essential to develop a system in order to identify and classify malware samples.

This research conducts experiments to classify malware using grayscale and RGB image visualization with Convolutional Neural Network (CNN) method. The dataset is visualized grayscale images which are converted from executable files and consists of 9339 images and 25 families. Whereas, the RGB images are obtained by converting grayscale images into color images using the pseudocolor technique. Before malware images are classified, Principal Component Analysis (PCA) is utilized to handle high dimensional data and improve the classification accuracy by extracting important features.

Experimental results show that the Convolutional Neural Network model achieves the highest accuracy 97.591% and the highest recall of 95.026%. As well, for the RGB images, the Convolutional Neural Network model obtains the best accuracy of 97.591% and the best recall of 94.824%. Overall, malware classification with both types of images delivers quite good results.

Keywords: *Cybersecurity, Image Classification, Malware, Convolutional Neural Network (CNN).*

KATA PENGANTAR

Terima kasih kepada Tuhan yang Maha Esa karena dengan bimbingan-Nya dan karunia-Nya penulis dapat melaksanakan Tugas Akhir yang berjudul ”PENERAPAN METODE CONVOLUTIONAL NEURAL NETWORK TERHADAP CITRA *GRAYSCALE* DAN *PSEUDOCOLOR* RGB UNTUK KLASIFIKASI *MALWARE*”. Laporan ini disusun sebagai salah satu syarat kelulusan di Institut Teknologi Harapan Bangsa. Pada kesempatan ini penulis menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Tuhan Yang Maha Esa, karena oleh bimbingan-Nya penulis selalu mendapat pengharapan untuk menyelesaikan tugas akhir ini.
2. Bapak Ventje Jeremias Lewi Engel S.T., M.T., CEH selaku Pembimbing I Tugas Akhir yang senantiasa memberi dukungan, semangat, ilmu-ilmu, dan saran kepada penulis selama pengerjaan laporan Tugas Akhir ini.
3. Bapak Sriwisnu Noloadi, S.Sc., M.Kom., selaku Penguji I Tugas Akhir. Terima kasih atas dukungan, semangat, ilmu-ilmu, dan masukan yang telah diberikan kepada penulis dalam menyelesaikan laporan Tugas Akhir ini.
4. Bapak Yoyok Yusman Gamaliel, S.T., M.Eng., selaku Penguji II Tugas Akhir. Terima kasih atas dukungan, semangat, ilmu-ilmu, dan masukan yang telah diberikan kepada penulis dalam menyelesaikan laporan Tugas Akhir ini.
5. Seluruh dosen dan staff Departemen Informatika ITHB yang telah membantu dalam menyelesaikan Laporan Tugas Akhir ini.
6. Segenap jajaran staf dan karyawan ITHB yang turut membantu kelancaran dalam menyelesaikan Laporan Tugas Akhir ini.
7. Teman seangkatan dan kakak tingkat, baik yang berkuliah di satu tempat maupun yang berbeda tempat yang telah menyediakan waktu untuk memberikan dukungan, semangat dan masukan kepada penulis dalam menyelesaikan laporan Tugas Akhir ini.
8. Kedua orang tua tercinta yang selalu menyediakan waktu untuk memberikan doa, semangat, dan dukungan yang tak habis-habisnya kepada penulis dalam menyelesaikan Laporan Tugas Akhir ini. Terima kasih untuk nasihat, masukan, perhatian, teguran, dan kasih sayang yang diberikan hingga saat ini.

Penulis menyadari bahwa laporan ini masih jauh dari sempurna karena keterbatasan waktu dan pengetahuan yang dimiliki oleh penulis. Oleh karena itu, kritik dan saran untuk membangun kesempurnaan tugas akhir ini sangat diharapkan. Semoga tugas akhir ini dapat membantu pihak-pihak yang membutuhkannya.

Bandung, 5 Juli 2022

Hormat penulis,

Chris Christian

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

DAFTAR REFERENSI

[1] M. Kalash, M. Rochan, N. Mohammed, N. Bruce, Y. Wang, and F. Iqbal, ”A Deep Learning Framework for Malware Classification,” *International Journal of Digital Crime and Forensics (IJDCF)*, vol. 12, no.1, pp. 90-108, March 2020. [Online], Available: http://doi.org/10.4018/IJDCF.2020010105.

[2] B. Yadav and S. Tokekar, “Recent Innovations and Comparison of Deep Learning Techniques in Malware Classification: A Review,” *International Journal of Information Security Science*, vol. 9, no. 4, pp. 230-247, 2020. [Online], Available: http://www.ijiss.org/ijiss/index.php/ijiss/article/view/852.

[3] A. Makandar and A. Patrot, ”Trojan Malware Image Pattern Classification,” *Proceedings of International Conference on Cognition and Recognition*, Springer Singapore, 2018. [Online], Available: https://doi.org/10.1007/978-981-10-5146-3_24.

[4] J. Fu, J. Xue, Y. Wang, Z. Liu, and C. Shan, ”Malware Visualization for Fine-Grained Classification,” *IEEE Access*, vol. 6, pp. 14510-14523, 2018. [Online], Available: https://doi.org/10.1109/ACCESS.2018.2805301.

[5] M. Jain, W. Andreopoulos, and M. Stamp, “Convolutional neural networks and extreme learning machines for malware classification,” *Journal of Computer Virology and Hacking Techniques*, vol. 16, pp. 229-244, 2020. [Online], Available: https://doi.org/10.1007/s11416-020-00354-y.

[6] A. Darwaish and F. Naït-Abdesselam, ”RGB-based Android Malware Detection and Classification Using Convolutional Neural Network,” *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, pp. 1-6, 2020. [Online], Available: https://doi.org/10.1109/GLOBECOM42002.2020.9348206.

[7] J. Rasheed, A.A. Hameed, C. Djeddi, A. Jamil, and F. Al-Turjman, ”A machine learning-based framework for diagnosis of COVID-19 from chest X-ray images,” *Interdiscip Sci Comput Life Sci*, vol. 13, pp. 103-117, 2021. [Online], Available: https://doi.org/10.1007/s12539-020-00403-6.

[8] M. Stamp, M. Alazab, and A. Shalaginov, *Malware Analysis Using Artificial Intelligence and Deep Learning*, Berlin/Heidelberg, Germany: Springer, 2021.

[9] O. Aslan and A.A.Yilmaz, "A New Malware Classification Framework Based on Deep Learning Algorithms," *IEEE Access*, vol. 9, pp. 87936-87951, 2021. [Online], Available: https://doi.org/10.1109/ACCESS.2021.3089586.

[10] M. Alazab, S. Venkatraman, P. Watters, M. Alazab, and A. Alazab, "Cybercrime: The Case of Obfuscated Malware," *Global security, safety and sustainability & e-Democracy*, pp. 204-211, 2011. [Online], Available: https://doi.org/10.1007/978-3-642-33448-1_28.

[11] J. Singh and J. Singh, "Challenges of Malware Analysis: Obfuscation Techniques," *International Journal of Information Security Science*, vol. 7, no. 3, pp. 100-110, 2018. [Online], Available: http://50.87.218.19/ijiss/index.php/ijiss/article/view/327.

[12] L. Nataraj, S. Karthikeyan, G. Jacob, and B.S. Manjunath, "Malware images: visualization and automatic classification," *Proceedings of the 8th International Symposium on Visualization for Cyber Security*, article 4, pp. 1-7, 2011. [Online], Available: https://doi.org/10.1145/2016904.2016908.

[13] H. Abdi and L.J. Williams, "Principal component analysis," *WIREs Comp Stat*, vol. 2, no.4m, pp. 433-459, 2010. [Online]. Available: https://doi.org/10.1002/wics.101.

[14] R.C. Gonzalez and R.E.Woods, *Digital Image Processing*, 3^{rd} Edition, New Jersey:Pearson Prentice Hall, 2008.

[15] Western Sydney University, "Colour modes explained". [Online], Available: https://www.westernsydney.edu.au/tld/home/how_to/how-to_resources/images_and_graphics/colour_modes. [Accessed: 16-Feb-2022].

[16] R. Yamashita, M. Nishio1, R.K.G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights Imaging*, vol. 9, no. 4, pp. 611-629, 2018. [Online], Available: https://doi.org/10.1007/s13244-018-0639-9.

[17] S. Indolia, A.K. Goswami, S.P. Mishra, and P. Asopa, "Conceptual Understanding of Convolutional Neural Network - A Deep Learning

Approach," *Procedia Computer Science*, vol. 132, pp. 679-688, 2018. [Online], Available: https://doi.org/10.1016/j.procs.2018.05.069.

[18] A. Ghosh, A. Sufian, F. Sultana, A. Chakrabarti, and D. De, "Fundamental Concepts of Convolutional Neural Network," *Recent Trends and Advances in Artificial Intelligence and Internet of Things*, vol. 172, Springer, pp. 519-567, 2019. [Online], Available: https://doi.org/10.1007/978-3-030-32644-9_36.

[19] P. Ligade, "Why cautiously initializing deep neural networks matters?", Towards Data Science, 18 April 2019. [Online], Available: https://towardsdatascience.com/what-is-weight-initialization-in-neural-nets-and-why-it-matters-ec45398f99fa. [Accessed: 21-Feb-2022].

[20] A. G´eron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow 2nd Edition*, O'Reilly Media, Inc, 2019.

[21] H.H. Aghdam and E.J. Heravi, *Guide to Convolutional Neural Networks*, 1st ed. Switzerland: Springer International Publishing AG, 2017.

[22] L. Panneerselvam, "Activation Functions and their Derivatives-A Quick & Complete Guide", Analytics Vidhya, 14 April 2021. [Online], Available: https://www.analyticsvidhya.com/blog/2021/04/activation-functions-and-their-derivatives-a-quick-complete-guide/. [Accessed: 18-Feb-2022]

[23] W. Di, A. Bhardwaj, and J. Wei, *Deep Learning Essentials: Your hands-on guide to the fundamentals of deep learning and neural network modeling*, Packt Publishing, 2018.

[24] H. Kinsley and D. Kukieła, *Neural Networks from Scratch in Python*, 1st ed. Harrison Kinsley, 2020.

[25] OpenCV, "ColorMaps in OpenCV". [Online], Available: https://docs.opencv.org/4.x/d3/d50/group__imgproc__colormap.html. [Accessed: 19-Feb-2022]

[26] Microsoft Security Intelligence, "Malware Encyclopedia Description". [Online], Available: https://www.microsoft.com/en-us/wdsi/threats/malware-encyclopedia-description. [Accessed: 5-Mar-2022]

[27] Trend Micro, "Threat Encyclopedia". [Online], Available: https://www.trendmicro.com/vinfo/us/threat-encyclopedia/. [Accessed: 5-Mar-2022]