Image Classification Pada Malware Menggunakan Pendekatan Metode Deep Learning VGG-16

Laporan Tugas Akhir

Diajukan Untuk Memenuhi
Persyaratan Guna Meraih Gelar Sarjana
Informatika Universitas Muhammadiyah Malang

Faiq Azmi Nurfaizi
201810370311047

Bidang Minat
Data Science

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH MALANG
2023

LEMBAR PERSETUJUAN

Image Classification Pada Malware Menggunakan Pendekatan Metode Deep Learning VGG-16

TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1
Informatika Universitas Muhammadiyah Malang

Menyetujui,
Malang, *29 September 2023*

Dosen Pembimbing 1	Dosen Pembimbing 2
Didih Rizki Chandranegara S.kom., M.Kom	**Zamah Sari ST., MT.**
NIP. 180302101992PNS.	**NIP. 10814100555PNS.**

LEMBAR PENGESAHAN

Image Classification Pada Malware Menggunakan Pendekatan Metode Deep Learning VGG-16

TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1
InformatikaUniversitas Muhammadiyah Malang

Disusun Oleh :

Faiq Azmi Nurfaizi

201810370311047

Tugas Akhir ini telah diuji dan dinyatakan lulus melalui sidang majelis penguji pada tanggal 29 September 2023

Menyetujui,

Dosen Penguji 1	Dosen Penguji 2
Wildan Suharso S.Kom., M.Kom	**Hardianto Wibowo S.Kom, MT.**
NIP. 10817030596PNS.	**NIP. 10816120592PNS.**

Mengetahui,
Ketua Jurusan Informatika

Ir. Galih Wasis Wicaksono S.kom. M.Cs.
NIP. 10814100541PNS.

LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini :

NAMA : **Faiq Azmi Nurfaizi**

NIM : **201810370311047**

FAK./JUR. : **Informatika**

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul **"Image Classification Pada Malware Menggunakan Pendekatan Metode Deep Learning VGG-16"** beserta seluruh isinya adalah karya saya sendiri dan bukan merupakan karya tulis orang lain, baik sebagian maupun seluruhnya, kecuali dalam bentuk kutipan yang telah disebutkan sumbernya.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini, atau ada klaim dari pihak lain terhadap keaslian karya saya ini maka saya siap menanggung segala bentuk resiko/sanksi yang berlaku.

Mengetahui,
Dosen Pembimbing

Didih Rizki Chandranegara S.kom., M.Kom

Malang, 29 September 2023
Yang Membuat Pernyataan

Faiq Azmi Nurfaizi

ABSTRAK

Pengimplementasian *Malware* dirancang untuk merusak perangkat keras seperti komputer, server, klien atau perangkat keras yang memiliki hubungan dengan jaringan. Secara umum, *malware* diinterpretasikan sebagai program yang dirancang untuk merusak komputer atau server dengan tujuan jahat seperti mendapatkan akses tidak sah ke sistem tertentu dengan memanfaatkan celah dalam suatu sistem atau jaringan. Sebagian besar *malware* saat ini dirancang dengan memiliki banyak variasi dengan dampak kerusakan yang berbeda dan hal ini menjadikan kemampuan mengklasifikasikan suatu karakteristik varian *malware* yang serupa dalam keluarga malware merupakan strategi yang baik dan juga kompleks dalam menghentikan *malware*. Riset pada penelitian ini mencoba mengklasifikasikan *malware* menggunakan dataset gambar *malware malimg* yang memiliki komposisi berbentuk *grayscale bytemap* dengan total 9.029 gambar dari 25 jenis *malware* yang berbeda. Dengan mengimplementasikan arsitektur VGG-16 dan model pembelajaran pembanding yaitu InceptionResNet-V2 pada 2 skenario yang berbeda dengan skenario 1 menggunakan dataset asli dan skenario 2 menggunakan dataset asli hasil proses *undersampled*. Setiap skenario yang dikembangkan menghasilkan nilai *metrics* evaluasi berupa akurasi, presisi, recall, dan f1-score dengan hasil akhir yang menunjukkan perolehan skor tertinggi pada skenario 2 pada arsitektur VGG-16 dengan skor akurasi 94,8% dan terendah pada skenario 2 pada arsitektur pembanding InceptionResNet-V2 dengan skor 85,1%.

Kata Kunci: Convolutional Neural Network, Klasifikasi Gambar, Malware, Pembelajaran Mesin, VGG-16 .

ABSTRACT

Implementation of Malware is designed to damage hardware devices such as computers, servers, clients or hardware that has a connection with the network. In general, malware is interpreted as a program designed to damage a computer or server with malicious purposes such as gaining unauthorized access to certain systems by exploiting loopholes in a system or network. Most of today's malware is designed with many variations with different damage effects and this makes the ability to classify a similar variant of malware characteristics in a malware family is a good and also complex strategy in stopping malware. Research in this study attempts to classify malware using a dataset of malimg malware images which have a composition in the form of a grayscale bytemap with a total of 9,029 images of 25 different types of malware. By implementing the VGG-16 architecture and the comparative learning model, namely InceptionResNet-V2 in 2 different scenarios with scenario 1 using the original dataset and scenario 2 using the original dataset resulting from the undersampled process. Each scenario developed produces evaluation metrics values in the form of accuracy, precision, recall, and f1-score with the final result showing the highest score obtained in scenario 2 on the VGG-16 architecture with an accuracy score of 94.8% and the lowest in scenario 2 on the comparison architecture InceptionResNet-V2 with a score of 85.1%.

Keywords: Convolutional Neural Networks, Image Classification, Malware, Machine Learning, VGG-16

LEMBAR PERSEMBAHAN

Puji syukur kepada Allah SWT atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini. Penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada :

1. Bapak Didih Rizki Chandranegara, S.kom., M.Kom, selaku dosen pembimbing satu dan Bapak Zamah Sari, ST., MT, selaku dosen pembimbing dua yang senantiasa membimbing selama proses penelitian jurnal berlangsung.
2. Kedua orang tua peneliti, Bapak Far'i Hidayat Hadi dan Ibu Ika Susi Layarati yang selalu memberikan dukungan dan doa untuk melaksanakan penelitian hingga selesai.
3. Bapak Ir. Galih Wasis Wicaksono S.kom. M.Cs, selaku Ketua Jurusan Prodi Informatika Universitas Muhammadiyah Malang.
4. Bapak/Ibu Dosen Informatika Universitas Muhammadiyah Malang, yang telah memberikan ilmu bermanfaat selama proses perkuliahan.
5. Ayu Wulandari sebagai orang terdekat yang selalu menjadi *support system* dan selalu memberikan semangat dalam penulisan dan penelitian jurnal.
6. Rekan-rekan eFishery, terutama tim Business Performance yang selalu memotivasi dan memberikan kesempatan kepada peneliti untuk dapat menyelesaikan penelitian jurnal dengan baik dan sempurna.
7. Rekan-rekan mahasiswa informatika angkatan 2018 Universitas Muhammadiyah Malang yang senantiasa membantu saya dalam masa perkuliahan.

Malang, 29 September 2023

Faiq Azmi Nurfaizi

KATA PENGANTAR

Dengan memanjatkan puji syukur kehadirat Allah SWT. Atas limpahan rahmat dan hidayah-NYA sehingga peneliti dapat menyelesaikan tugas akhir yang berjudul:

"**Image Classification Pada Malware Menggunakan Pendekatan Metode Deep Learning VGG-16**".

Di dalam tulisan ini disajikan pokok-pokok bahasan yang meliputi penggunakan dataset *Malimg* sebagai dataset penelitian dan rancangan arsitektur model *VGG-16* serta evaluasi skor Akurasi, presisi, recall, dan f1-score sebagai performa model yang dibangun.

Peneliti menyadari sepenuhnya bahwa dalam penulisan tugas akhir ini masih banyak kekurangan dan keterbatasan. Oleh karena itu peneliti mengharapkan saran yang membangun agar tulisan ini bermanfaat bagi perkembangan ilmu pengetahuan.

Malang, 29 September 2023

Faiq Azmi Nurfaizi

DAFTAR ISI

LEMBAR PERSETUJUAN ..... **i**
LEMBAR PENGESAHAN ..... **ii**
LEMBAR PERNYATAAN ..... **iii**
ABSTRAK ..... **iv**
ABSTRACT ..... **v**
LEMBAR PERSEMBAHAN ..... **vi**
KATA PENGANTAR ..... **vii**
DAFTAR ISI ..... **viii**
DAFTAR GAMBAR ..... **x**
DAFTAR TABEL ..... **xi**
BAB I PENDAHULUAN **1**
1.1 Latar Belakang 1
1.2 Rumusan Masalah 4
1.3 Tujuan Penelitian 4
1.4 Batasan Masalah 4
BAB II TINJAUAN PUSTAKA **5**
2.1 Penelitian Terdahulu 5
2.2 Image Classification 7
2.3 Deep Learning 8
2.4 VGG-16 8
BAB III METODE PENELITIAN **11**
3.1 Desain Penelitian 11
3.2 Dataset 13
3.3 Data Preprocessing 14
3.3.1 Data Balancing 14
3.3.2 Data Grayscaling 14
3.3.3 Data Resizing 15
3.3.4 Data Rescaling 16
3.4 Pembuatan Model 17
3.4.1 VGG-16 17
3.4.2 InceptionResNet-V2 18
3.5 Evaluasi Model 20

BAB IV HASIL DAN PEMBAHASAN 21
4.1 Persiapan dan Pengolahan Data 21
4.1.1 Proses Data Balancing 21
4.1.2 Proses Normalisasi Data 23
4.2 Inisialisasi Model 24
4.3 Evaluasi Performa Model 25
4.3.1 Skenario 1 25
4.3.2 Skenario 2 28
4.3.3 Rangkuman Hasil Kinerja Model.....30
BAB V PENUTUP.....33
5.1 Kesimpulan 33
5.2 Saran.....33
DAFTAR PUSTAKA.....35

DAFTAR GAMBAR

Gambar 2. 1 Model arsitektur deep learning VGG-16 9
Gambar 3. 1 Alur desain penelitian 11
Gambar 3. 2 Contoh gambar dataset *malware* 15
Gambar 3. 3 Contoh hasil proses *grayscaling* pada dataset 15
Gambar 3. 4 Contoh gilustrasi penerapan *resizing* pada gambar 15
Gambar 3. 5 Contoh gilustrasi penerapan *rescaling* pada gambar 15
Gambar 3. 6 Arsitektur model VGG-16 15
Gambar 3. 7 Arsitektur model InceptionResNet-V2 15
Gambar 4. 1 Potongan kode proses *data balancing* 26
Gambar 4. 2 Potongan kode proses normalisasi dataset..... 23
Gambar 4. 3 Potongan kode inisialisasi model VGG-16 26
Gambar 4. 4 Potongan kode inisialisasi model InceptionRestNet-V2 26
Gambar 4. 5 Grafik *Accuracy* dan *Loss* pada model VGG-16 di skenario 1 26
Gambar 4. 6 Grafik *Accuracy* dan *Loss* pada model pembanding InceptionResNet-V2 di skenario 1 26
Gambar 4. 7 Grafik *Accuracy* dan *Loss* pada model VGG-16 di skenario 2 28
Gambar 4. 8 Grafik *Accuracy* dan *Loss* pada model pembanding InceptionResNet-V2 di skenario 2 26

DAFTAR TABEL

Tabel 2. 1 Mapping studi literatur 6
Tabel 4. 1 Perbandingan dataset 22
Tabel 4. 2 Perbandingan hasil akurasi model di skenario 1 26
Tabel 4. 3 Hasil evaluasi model di skenario 1 27
Tabel 4. 4 Perbandingan hasil akurasi model di skenario 2 29
Tabel 4. 5 Hasil evaluasi model di skenario 2 30
Tabel 4. 6 Perbandingan hasil akurasi model di setiap skenario 30
Tabel 4. 7 Perbandingan hasil akurasi model dengan penelitian sebelumnya 31

DAFTAR PUSTAKA

[1] M. N. Alenezi, H. Alabdulrazzaq, A. A. Alshaher, and M. M. Alkharang, "Evolution of Malware Threats and Techniques: A Review," *Int. J. Commun. Networks Inf. Secur.*, vol. 12, no. 3, pp. 326–337, 2020, doi: doi.org/10.17762/ijcnis.v12i3.4723.

[2] Alena Yuryna Connolly and H. Borrion, "Reducing Ransomware Crime: Analysis of Victims' Payment Decisions," 2022, doi: doi.org/10.1016/j.cose.2022.102760.

[3] M. Naseer *et al.*, "Malware Detection: Issues and Challenges," *J. Phys. Conf. Ser.*, vol. 1807, no. 1, pp. 1–6, 2021, doi: 10.1088/1742-6596/1807/1/012011.

[4] M. A. Hama Saeed, "Malware in Computer Systems: Problems and Solutions," *IJID (International J. Informatics Dev.*, vol. 9, no. 1, p. 1, 2020, doi: 10.14421/ijid.2020.09101.

[5] F. A. Aboaoja, A. Zainal, F. A. Ghaleb, B. A. S. Al-rimy, T. A. E. Eisa, and A. A. H. Elnour, "Malware Detection Issues, Challenges, and Future Directions: A Survey," *Appl. Sci.*, vol. 12, no. 17, 2022, doi: 10.3390/app12178482.

[6] Z. Zhao, D. Zhao, S. Yang, and L. Xu, "Image-Based Malware Classification Method with the AlexNet Convolutional Neural Network Model," *Secur. Commun. Networks*, vol. 2023, 2023, doi: 10.1155/2023/6390023.

[7] M. S. Akhtar and T. Feng, "Malware Analysis and Detection Using Machine Learning Algorithms," *Symmetry (Basel).*, vol. 14, no. 11, p. 2304, 2022, doi: 10.3390/sym14112304.

[8] S. Choi, J. Bae, C. Lee, Y. Kim, and J. Kim, "Attention-based automated feature extraction for malware analysis," *Sensors (Switzerland)*, vol. 20, no. 10, pp. 1–17, 2020, doi: 10.3390/s20102893.

[9] H. Naeem, B. Guo, M. R. Naeem, F. Ullah, H. Aldabbas, and M. S. Javed, "Identification of malicious code variants based on image visualization,"

Comput. Electr. Eng., vol. 76, pp. 225–237, 2019, doi: 10.1016/j.compeleceng.2019.03.015.

[10] M. R. Naeem, R. Amin, S. S. Alshamrani, and A. Alshehri, "Digital Forensics for Malware Classification: An Approach for Binary Code to Pixel Vector Transition," *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/6294058.

[11] U.-H. Tayyab, F. B. Khan, M. H. Durad, A. Khan, and Y. S. Lee, "A Survey of the Recent Trends in Deep Learning Based Malware Detection," *J. Cybersecurity Priv.*, vol. 2, no. 4, pp. 800–829, 2022, doi: 10.3390/jcp2040041.

[12] A. Thomas, P. M. Harikrishnan, P. Palanisamy, and V. P. Gopi, "Moving Vehicle Candidate Recognition and Classification Using Inception-ResNet-v2," *Proc. - 2020 IEEE 44th Annu. Comput. Software, Appl. Conf. COMPSAC 2020*, pp. 467–472, 2020, doi: 10.1109/COMPSAC48688.2020.0-207.

[13] Q. Guan *et al.*, "Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: A pilot study," *J. Cancer*, vol. 10, no. 20, pp. 4876–4882, 2019, doi: 10.7150/jca.28769.

[14] D. Pant and R. Bista, "Image-based Malware Classification using Deep Convolutional Neural Network and Transfer Learning," *ACM Int. Conf. Proceeding Ser.*, 2021, doi: 10.1145/3503047.3503081.

[15] A. Bensaoud, N. Abudawaood, and J. Kalita, "Classifying Malware Images with Convolutional Neural Network Models," 2020, doi: 10.48550/arXiv.2010.16108.

[16] H. U. Sharif, N. Jiwani, K. Gupta, M. A. Mohammed, and M. F. Ansari, "a Deep Learning Based Technique for the Classification of Malware Images," *J. Theor. Appl. Inf. Technol.*, vol. 101, no. 1, pp. 135–160, 2023, [Online]. Available: http://www.jatit.org/volumes/Vol101No1/12Vol101No1.pdf

[17] K. R. Goyal Manish, "AVMCT: API Calls Visualization based Malware Classification using Transfer Learning," vol. 13, no. 1, pp. 31–41, 2022,

doi: doi.org/10.52783/jas.v13i1.59.

[18] Y. Kim, C. Schmid, J. M. Kim, Z. Akata, J. Jeong, and J. Lee, "Partially Annotated Multi-label Classification", doi: https://doi.org/10.48550/arXiv.2304.01804.

[19] K. H. Janiesch Christian, Patrick Zschech, "Machine Learning and Deep Learning," *Adv. Inf. Secur.*, vol. 103, pp. 347–384, 2023, doi: 10.1007/978-3-031-26845-8_8.

[20] M. A. Rafidison *et al.*, "Image Classification Based on Light Convolutional Neural Network Using Pulse Couple Neural Network," *Comput. Intell. Neurosci.*, vol. 2023, pp. 1–17, 2023, doi: 10.1155/2023/7371907.

[21] D. Albashish, R. Al-Sayyed, A. Abdullah, M. H. Ryalat, and N. Ahmad Almansour, "Deep CNN Model based on VGG16 for Breast Cancer Classification," *2021 Int. Conf. Inf. Technol. ICIT 2021 - Proc.*, pp. 805–810, 2021, doi: 10.1109/ICIT52682.2021.9491631.

[22] L. Alzubaidi *et al.*, *Review of deep learning: concepts, CNN architectures, challenges, applications, future directions*, vol. 8, no. 1. Springer International Publishing, 2021. doi: 10.1186/s40537-021-00444-8.

[23] C. L. Fan and Y. J. Chung, "Design and Optimization of CNN Architecture to Identify the Types of Damage Imagery," *Mathematics*, vol. 10, no. 19, 2022, doi: 10.3390/math10193483.

[24] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A Survey of the Recent Architectures of Deep Convolutional Neural Networks," *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5455–5516, 2020, doi: 10.1007/s10462-020-09825-6.

[25] M. R. Alwanda, R. Putra, K. Ramadhan, D. Alamsyah, P. Studi, and T. Informatika, "Implementasi Metode Convolutional Neural Network Menggunakan Arsitektur LeNet-5 untuk Pengenalan Doodle," *J. Algoritm.*, vol. 1, no. 1, 2020, doi: doi.org/10.35957/algoritme.v1i1.434.

[26] R. Nirthika and S. Manivannan, "Pooling in convolutional neural networks for medical image analysis : a survey and an empirical study," *Neural Comput. Appl.*, vol. 34, no. 7, pp. 5321–5347, 2022, doi: 10.1007/s00521-022-06953-8.

[27] M. A. Saleem, N. Senan, F. Wahid, M. Aamir, A. Samad, and M. Khan, “Comparative Analysis of Recent Architecture of Convolutional Neural Network,” vol. 2022, 2022, doi: doi.org/10.1155/2022/7313612.

FORM CEK PLAGIARISME LAPORAN TUGAS AKHIR

Nama Mahasiswa : **Faiq Azmi Nurfaizi**

NIM : **201810370311047**

Judul TA : **Image Classification Pada Malware Menggunakan Pendekatan Metode Deep Learning VGG-16**

Hasil Cek Plagiarisme dengan Turnitin

No.	Komponen Pengecekan	Nilai Maksimal Plagiarisme (%)	Hasil Cek Plagiarisme (%) *
1.	Bab 1 – Pendahuluan	10 %	6%
2.	Bab 2 – Daftar Pustaka	25 %	8%
3.	Bab 3 – Analisis dan Perancangan	25 %	5%
4.	Bab 4 – Implementasi dan Pengujian	15 %	9%
5.	Bab 5 – Kesimpulan dan Saran	5 %	3%
6.	Makalah Tugas Akhir	20%	17%

) Hasil cek plagiarism diisi oleh pemeriksa (staf TU)

***) Maksimal 5 kali (4 Kali sebelum ujian, 1 kali sesudah ujian)**

Mengetahui,

Pemeriksa (Staff TU)

(.....deny.....)