DETEKSI *MALWARE* DENGAN REPRESENTASI *IMAGE* MENGGUNAKAN *PRE-TRAINED* CNN

SKRIPSI

IHSAN ALAMAL AHMAD 2107411001

PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA DAN KOMPUTER
POLITEKNIK NEGERI JAKARTA
2025

DETEKSI *MALWARE* DENGAN REPRESENTASI *IMAGE* MENGGUNAKAN *PRE-TRAINED* CNN

SKRIPSI

Dibuat untuk Melengkapi Syarat-Syarat yang Diperlukan untuk Memperoleh Diploma Empat Politeknik

IHSAN ALAMAL AHMAD 2107411001

PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA DAN KOMPUTER
POLITEKNIK NEGERI JAKARTA
2025

SURAT PERNYATAAN BEBAS PLAGIARISME

Saya yang bertanda tangan di bawah ini:

Nama : Ihsan Alamal Ahmad

NIM : 2107411001

Jurusun/Prodi : Teknik Informatika dan Komputer / Teknik Informatika

Judul Skripsi : Deteksi *Malware* dengan representasi Image menggunakan *pre-trained* CNN

Menyatakan dengan sebenarnya bahwa skripsi ini benar-benar merupakan hasil dari karya dari saya sendiri, bebas dari peniruan terhadap karya dari orang lain. kutipan pendapat dan tulisan orang lain ditunjuk sesuai dengan cara-cara penulisan karya ilmiah yang berlaku.

Apabila di kemudian hari ini terbukti atau dapat dibuktikan bahwa dalam skripsi ini terkandung ciri-ciri plagiat dan bentuk-bentuk peniruan lain yang dianggap melanggar peraturan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Depok, 7 Juni 2025

Yang membuat pernyataan,

Ihsan Alamal Ahmad

NIM 2107411001

LEMBAR PENGESAHAN

Skripsi diajukan oleh:

Nama	:	Ihsan Alamal Ahmad
NIM	:	2107411001
Program Studi	:	Teknik Informatika
Judul Skripsi	:	Deteksi *malware* dengan representasi *image* menggunakan *Pre-trained* CNN

Telah diuji oleh tim penguji dalam Sidang Skripsi pada hari Kamis, tanggal 26, bulan Juni, tahun 2025, dan dinyatakan **LULUS.**

Disahkan oleh:

Pembimbing I	:	Iik Muhamad Malik Matin, M.T	( )
Penguji I	:	Dr. Dewi Yanti Liliana, S.Kom., M.Kom.	( )
Penguji II	:	Rizki Elisa Nalawati, S.T., M.T.	( )
Penguji III	:	Maria Agustin, S.Kom., M.Kom.	( )

Mengetahui:

Ketua Jurusan Teknik Informatika dan Komputer

Dr. Anita Hidayati, S.Kom., M.Kom.

NIP. 197908032003122003

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, karunia, serta petunjuk-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi ini yang berjudul "**Deteksi *Malware* dengan representasi *image* menggunakan *Pre-trained* CNN"** sebagai salah satu syarat untuk memperoleh gelar Sarjana Terapan pada Program Studi Teknik Informatika, Jurusan Teknik Informatika dan Komputer, Politeknik Negeri Jakarta.

Penyusunan skripsi ini tidak lepas dari bantuan, dukungan, bimbingan, dan doa dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Bapak Iik Muhammad Malik Matin, selaku dosen pembimbing yang telah membantu penulis dari awal hingga akhir pengerjaan skripsi ini;
2. Kedua Orang tua tercinta, yang telah memberikan dukungan selama masa perkuliahan penulis di Teknik Informatika Politeknik Negeri Jakarta.
3. Kedua Kakak kandung penulis, Rahma Kamila Ahmad dan Rahma Fitria Ahmad yang telah sabar memberikan dukungan dan bimbingan dalam berbagai aspek kehidupan.
4. Seluruh teman-teman dan kerabat yang telah banyak membantu penulis dalam pengerjaan skripsi ini, baik secara langsung maupun tidak langsung.

Penulis berharap semoga Tuhan Yang Maha Esa mengaruniakan rahmat dan hidayah-Nya kepada mereka semua. Semoga skripsi ini dapat bermanfaat bagi kita semua.

Depok, 7 Juni 2025

Yang Menyatakan,

Ihsan Alamal Ahmad

NIM 2107411001

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Politeknik Negeri Jakarta, saya yang bertanda tangan dibawah ini:

Nama : Ihsan Alamal Ahmad

NIM : 2107411001

Jurusan/Program Studi : T. Informatika dan Komputer / Teknik Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Politeknik Negeri Jakarta Hak Bebas Royalti Non-Eksklusif atas karya ilmiah saya yang berjudul:

DETEKSI *MALWARE* DENGAN REPRESENTASI IMAGE MENGGUNAKAN *PRE-TRAINED* CNN

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksklusif ini, Politeknik Negeri Jakarta berhak menyimpan, mengalihmediakan/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan skripsi saya tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta.

Demikian pernyataan ini saya buat dengan sebenar-benarnya.

Depok, 7 Juni 2025

Yang Menyatakan,

Ihsan Alamal Ahmad

NIM 2107411001

DETEKSI *MALWARE* DENGAN REPRESENTASI *IMAGE* MENGGUNAKAN *PRE-TRAINED CNN*

ABSTRAK

Evolusi cepat Malware menimbulkan tantangan signifikan bagi metode deteksi berbasis tanda tangan, khususnya pada perangkat dengan sumber daya terbatas. Penelitian ini mengatasi permasalahan tersebut melalui pengembangan sistem deteksi Malware berbasis gambar yang diimplementasikan pada aplikasi web, dengan tujuan membandingkan kinerja empat model pembelajaran mendalam, yaitu MobileNetV3, ResNet, EfficientNet, dan DenseNet. Metode penelitian mencakup konversi file biner (.exe, OLE, PE) menjadi citra grayscale berukuran 224x224 piksel, pelatihan model menggunakan dataset Dikedataset, serta evaluasi melalui tiga pendekatan: Black Box Testing, External Data Testing dengan sampel dari Malware Bazaar, dan User Acceptance Testing (UAT) (UAT). Hasil penelitian menunjukkan bahwa akurasi pelatihan model berkisar antara 95,20% (MobileNetV3) hingga 98,65% (EfficientNet), dengan tingkat keberhasilan 100% pada Black Box Testing dan waktu pemrosesan rata-rata 800 milidetik per file pada CPU. Namun, External Data Testing menunjukkan akurasi rata-rata sebesar 70%, menunjukkan adanya keterbatasan dalam generalisasi terhadap varian Malware baru. Sementara itu, UAT menghasilkan indeks penerimaan sebesar 80,63%, menunjukkan antarmuka yang diterima oleh pengguna. Penelitian ini menyimpulkan bahwa representasi visual efektif pada dataset terkontrol, dengan EfficientNet unggul dalam hal efisiensi, tetapi generalisasi terhadap data eksternal memerlukan peningkatan lebih lanjut.

Kata Kunci: *CNN , Malware , MobileNet , ResNet , EfficientNet*

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

BAB I
PENDAHULUAN

1.1 Latar Belakang Masalah

Masalah keamanan siber semakin mendesak seiring berkembangnya ancaman *Malware* di era digital. *Malware* adalah perangkat lunak berbahaya yang dirancang untuk merusak atau mencuri data dalam sistem komputer. Seiring meningkatnya penggunaan internet, volume serangan *Malware* global juga melonjak drastis. Penelitian menunjukkan bahwa kerusakan yang ditimbulkan *Malware* meningkat tajam karena internet yang makin populer dan kecanggihan kode jahat yang terus bertambah. Kondisi ini menuntut strategi deteksi *Malware* yang lebih canggih untuk menghadapi evolusi varian *Malware* baru yang semakin kompleks (Moser et al., 2020).

Tabel berikut merangkum tren ancaman *Malware* berdasarkan laporan Symantec:

Table 1.1 Symantec Internet Security Threat Report (ISTR) 2023-2025

Tahun	Jumlah *Malware* Baru Terdeteksi	Peningkatan	Sumber
2022	260 Juta	-	Symantec ISTR 2023
2023	320 Juta	+23%	Symantec ISTR 2024
2024	480 Juta	+50%	Symantec ISTR 2025

Meskipun demikian, penerapan *Machine Learning* dalam deteksi *Malware* menghadapi keterbatasan dimana model*Machine Learning*, termasuk *Deep Learning*, bergantung pada kualitas data pelatihan dan sering kali gagal mendeteksi varian baru jika pola tidak terwakili dalam *dataset* (Goodfellow et al., 2016). Selain itu, ketidakmampuan model untuk menangani obfuskasi canggih atau serangan adversarial dapat mengurangi akurasi, sementara overfit pada data spesifik menyebabkan performa buruk pada sampel eksternal (Chaganti et al., 2022). Keterbatasan ini menekankan perlunya pendekatan yang lebih adaptif dan robust.

Untuk mengatasi tantangan tersebut, pendekatan representasi citra (image-based representation) hadir sebagai solusi dimana teknik ini mengubah file biner *Malware* menjadi citra *grayscale*, sehingga memungkinkan pemanfaatan model *Deep Learning* berbasis *Convolutional Neural Network* (CNN). Citra *Malware* dari keluarga yang sama menunjukkan pola tekstur yang serupa, yang dapat dipelajari oleh CNN secara otomatis tanpa perlu rekayasa fitur manual. Selain mampu mendeteksi varian baru yang tidak ada dalam database *signature based*, pendekatan ini juga terbukti lebih tahan terhadap teknik seperti *padding* dan *redundant API injection,* yang sering digunakan dalam serangan adversarial. Penelitian menunjukkan bahwa konversi ke citra *grayscale* dan penggunaan CNN, seperti ResNet atau EfficientNet, dapat menghasilkan akurasi deteksi tinggi, sekaligus mengurangi ketergantungan pada proses ekstraksi fitur kompleks yang rentan kesalahan (Roseline et al., 2020; He & Kim, 2019).

Berbagai penelitian telah mengeksplorasi pendekatan representasi gambar untuk deteksi *Malware* . Nataraj et al. (2011) memperkenalkan metode konversi file biner menjadi citra *grayscale*, menunjukkan bahwa pola visual dapat digunakan untuk mengklasifikasikan keluarga *Malware* dengan akurasi hingga 97,18% menggunakan pendekatan k-*nearest neighbors* (k-NN). Roseline et al. (2020) melanjutkan dengan menerapkan CNN, seperti ResNet-50, pada *dataset Malware* yang dikonversi, mencapai akurasi 98,5% pada data pelatihan yang seimbang. He dan Kim (2019) mengembangkan pendekatan serupa dengan EfficientNet, menunjukan peningkatan performa hingga 99% pada *dataset* tertutup, meskipun keterbatasan generalisasi pada data eksternal tetap menjadi isu. Penelitian ini menyoroti potensi representasi gambar, namun sering kali terbatas pada *dataset* spesifik dan kurang mengevaluasi ketahanan terhadap serangan adversarial.

Berdasarkan pada keterbatasan penelitan sebelumnya, Penelitian ini difokuskan pada pengembangan metode deteksi malware berbasis representasi citra yang diperoleh melalui konversi file biner menjadi gambar grayscale. Model yang digunakan merupakan arsitektur *Convolutional Neural Network (CNN)* yang telah dilatih sebelumnya, yaitu MobileNetV3, ResNet, EfficientNet, dan DenseNet.

Keempat model ini dipilih untuk dibandingkan performanya dalam melakukan klasifikasi terhadap file yang telah direpresentasikan dalam bentuk citra.

Dataset yang digunakan dalam proses pelatihan berasal dari DikeDataset, yang menyediakan sampel file malware dan benign. Seluruh file biner dalam dataset dikonversi ke citra dengan pendekatan transformasi berbasis nilai byte ke dalam format dua dimensi. Setelah proses pelatihan selesai, model dievaluasi untuk menilai kemampuannya dalam membedakan malware dan non-malware. Selain itu, model dengan performa terbaik diterapkan dalam antarmuka web sederhana untuk menguji implementasi model dalam konteks penggunaan praktis.

1.2 Rumusan Masalah

Berdasarkan latar belakang, berikut adalah rumusan masalah dari penelitian ini:

1. Bagaimana merubah file biner menjadi representasi citra untuk kinerja deteksi *Malware* menggunakan *pre-trained* CNN?
2. Bagaimana mengembangkan empat arsitektur *CNN pre-trained*, yaitu MobileNetV3, ResNet, DenseNet, dan EfficientNet, dalam melakukan klasifikasi file *Malware* dan *Benign* berdasarkan citra *grayscale* hasil konversi biner?
3. Bagaimana performa *transfer learning* masing-masing arsitektur CNN dalam hal akurasi, waktu pelatihan, kecepatan inferensi, dan kebutuhan komputasi saat diterapkan pada Dikedataset yang memiliki karakteristik *Malware* dan *Benign* yang heterogen?
4. Bagaimana cara implementasi model yang telah dikembangkan dalam tampilan antar muka *web* dalam deteksi *Malware* ?

1.3 Batasan Masalah

Untuk menjaga fokus penelitian, batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Arsitektur CNN yang Digunakan

   Penelitian ini hanya menggunakan arsitektur Convolutional Neural Network yang telah dilatih sebelumnya, yaitu MobileNetV3, DenseNet, EfficientNet, dan ResNet. Model-model tersebut dipilih karena telah banyak digunakan dalam tugas klasifikasi citra dan memiliki karakteristik berbeda dalam hal

kompleksitas dan efisiensi. Arsitektur lain di luar keempat model ini tidak dijadikan objek evaluasi.

2. Pendekatan dan Metode Pelatihan

   Model dilatih menggunakan pendekatan *transfer learning*, di mana bobot awal dari model pre-trained disesuaikan kembali pada dataset representasi gambar *malware* dan *benign*. Penelitian ini tidak mencakup pelatihan model dari awal *(training from scratch)* atau pengembangan arsitektur CNN baru.

3. Parameter Evaluasi

   Penilaian kinerja model hanya akan difokuskan pada metrik utama seperti akurasi, presisi, *recall*, *F1-score*, waktu pelatihan, dan waktu inferensi.

4. Lingkup *Dataset*

   Dataset yang digunakan untuk pelatihan berasal dari DikeDataset, sedangkan pengujian eksternal dilakukan menggunakan sampel dari Malware Bazaar. Penelitian ini hanya mencakup deteksi berbasis citra dan tidak membahas metode berbasis signature, analisis perilaku, maupun rekayasa kode sumber secara langsung.

5. *Platform* Eksperimen

   Eksperimen dilakukan pada perangkat keras dengan spesifikasi:

Tabel 1. 2 Spesifikasi Perangkat Keras

Laptop	Huawei Matebook D-15
Prosesor	Intel Core i5-1135G7 8GB 512 SSD
Perangkat lunak	- *Jupyter Notebook* versi 7.2.2 - *Visual Studio Code* versi 1.100.

   Hasil penelitian ini dapat bergantung pada keterbatasan perangkat keras dan lingkungan perangkat lunak tersebut

6. Implementasi Sistem Prediksi

   Implementasi hasil penelitian ditampilkan dalam bentuk antarmuka web sederhana menggunakan Streamlit. Aplikasi ini berfungsi untuk menerima input file, menampilkan representasi gambar hasil konversi, dan memberikan hasil prediksi. Pengembangan sistem terbatas pada aspek fungsional dasar dan

tidak mencakup pengujian skalabilitas atau integrasi sistem keamanan tambahan.

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan

Penelitian ini bertujuan untuk:

1. Mengonversi file biner malware dan benign menjadi citra grayscale sebagai representasi visual.
2. Menerapkan dan membandingkan performa empat arsitektur *Convolutional Neural Network* (CNN) yang telah dilatih sebelumnya, yaitu MobileNetV3, ResNet, DenseNet, dan EfficientNet, dalam mendeteksi file *malware* berdasarkan representasi citra tersebut.
3. Mengevaluasi hasil pelatihan model menggunakan pendekatan *transfer learning* berdasarkan metrik evaluasi klasifikasi dan efisiensi komputasi.
4. Menerapkan model dengan performa terbaik ke dalam antarmuka *web* untuk mendemonstrasikan proses deteksi malware secara praktis.

1.4.2 Manfaat

Manfaat dari penelitian ini antara lain adalah:

1. Menyediakan pendekatan alternatif dalam representasi data biner.
   Konversi file biner menjadi citra grayscale memungkinkan pemanfaatan metode visual dalam deteksi malware. Representasi ini membuka kemungkinan untuk mengenali pola-pola visual yang tidak terlihat dalam format biner mentah, sehingga memperluas cakupan pendekatan klasifikasi berbasis data non-struktural dalam keamanan siber.
2. Memberikan perbandingan empiris antar arsitektur CNN pre-trained.
   Penelitian ini menghasilkan analisis terukur mengenai performa beberapa arsitektur CNN populer dalam mendeteksi malware berdasarkan data visual. Perbandingan ini memberikan informasi yang dapat dijadikan acuan dalam pemilihan model ketika dihadapkan pada keterbatasan sumber daya komputasi, kebutuhan efisiensi, atau karakteristik dataset tertentu.
3. Mengevaluasi efektivitas pendekatan transfer learning dalam domain keamanan siber.

Dengan menggunakan model pre-trained yang disesuaikan terhadap dataset khusus, penelitian ini menunjukkan bagaimana transfer learning dapat dimanfaatkan secara praktis dalam konteks deteksi malware. Evaluasi performa berdasarkan akurasi, presisi, recall, F1-score, dan efisiensi komputasi memberikan gambaran yang menyeluruh tentang potensi dan batasan pendekatan ini.

4. Mendemonstrasikan integrasi model klasifikasi ke dalam antarmuka aplikasi sederhana.

   Implementasi model ke dalam antarmuka web berbasis Python menunjukkan aplikasi langsung dari hasil pelatihan dalam konteks penggunaan nyata. Hal ini memberikan kontribusi terhadap pengembangan sistem pendeteksi malware yang dapat diakses secara interaktif, sekaligus membuktikan kelayakan integrasi model deep learning ke dalam sistem berbasis pengguna.

1.5 Sistematika Penulisan

Penelitian ini disusun dalam lima bab dengan sistematika penulisan sebagai berikut:

1. **BAB I: PENDAHULUAN** Bab ini berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.
2. **BAB II: TINJAUAN PUSTAKA DAN DASAR TEORI** Pada bab ini dijelaskan teori-teori dasar dan literatur yang relevan dengan penelitian, seperti konsep *Transfer Learning*, deteksi *Malware* berbasis representasi gambar dari kode biner, dan arsitektur model CNN *pre-trained* yang digunakan (MobileNetV3, ResNet, EfficientNet, dan DenseNet). Selain itu, bab ini juga menyajikan kajian pustaka dari penelitian-penelitian sebelumnya yang mendukung permasalahan yang diangkat.
3. **BAB III: PERENCANAAN DAN REALISASI** Bab ini menjelaskan metode yang digunakan dalam penelitian, meliputi langkah-langkah eksperimen, deskripsi *dataset* (Dikedataset), proses konversi kode biner ke representasi gambar, arsitektur *Transfer Learning* yang diuji (MobileNetV3, ResNet, EfficientNet, dan DenseNet), parameter evaluasi, serta alat dan perangkat lunak yang digunakan. Penjelasan rinci mengregarding proses pengolahan data dan analisis juga dipaparkan pada bagian ini.

4. **BAB IV: HASIL DAN PEMBAHASAN** Bagian ini menyajikan hasil penelitian berupa perbandingan kinerja empat arsitektur CNN *pre-trained* dalam mendeteksi *Malware* berbasis representasi gambar dari kode biner. Hasil analisis disertai dengan pembahasan mendalam mengenai efektivitas, efisiensi, dan generalisasi model yang diuji, termasuk metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score*. Grafik, tabel, dan visualisasi lain digunakan untuk memperjelas hasil penelitian.
5. **BAB V: KESIMPULAN DAN SARAN** Bab ini berisi kesimpulan dari penelitian yang telah dilakukan, kontribusi penelitian terhadap bidang teknologi informasi, serta saran untuk penelitian selanjutnya.

BAB V
PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian, berikut adalah kesimpulan yang didapat:

1. Transformasi *File Biner* ke Representasi Citra

Penelitian ini berhasil menerapkan metode transformasi *file biner* (.exe, OLE, PE) menjadi citra *grayscale* berukuran 224x224 piksel. Teknik ini memanfaatkan distribusi byte untuk membentuk pola visual yang mendukung deteksi *malware* menggunakan model *Convolutional Neural Network* (CNN) *pre-trained*. Proses transformasi ini efisien pada perangkat berbasis CPU dan menjawab rumusan masalah pertama dengan menyediakan representasi data yang optimal untuk analisis citra.

2. Pelatihan Empat Arsitektur CNN *Pre-trained*

Empat arsitektur CNN *pre-trained*, yaitu MobileNetV3, ResNet50, DenseNet121, dan EfficientNetB0, telah dilatih untuk mengklasifikasikan file *malware* dan *benign* berdasarkan citra *grayscale* yang dihasilkan. Pendekatan Transfer Learning digunakan untuk mengoptimalkan pelatihan pada dataset heterogen, memberikan jawaban atas rumusan masalah kedua dengan implementasi model yang sistematis dan terukur.

3. Evaluasi Performa Empat Model CNN

Analisis performa model dilakukan dengan memeriksa metrik evaluasi (akurasi, precision, *recall*, *F1-score*), waktu pelatihan, dan pola loss untuk memahami keunggulan dan kelemahan masing-masing model dalam mendeteksi *malware* :

- MobileNetV3: Mencapai akurasi 98,25%, precision 99,56%, *recall* 98,51%, dan *F1-score* 99,03%, dengan waktu pelatihan tercepat (24 menit). *Training loss* menurun dari 0,0265 menjadi 0,0054, dan validation loss dari 0,0383 menjadi 0,0063, menunjukkan efisiensi komputasi tinggi namun *recall* terendah.
- ResNet50: Mencatat akurasi tertinggi 98,78%, precision 99,71%, *recall* 98,94%, dan *F1-score* 99,32%, dengan waktu pelatihan terlama (90 menit). *Training loss* menurun dari 0,0106 menjadi 0,0034, dan validation loss dari 0,0070 menjadi 0,0041, menunjukkan stabilitas dan adaptasi unggul.

- DenseNet121: Menghasilkan akurasi 98,52%, precision 99,71%, *recall* 98,65%, dan *F1-score* 99,18%, dengan waktu pelatihan 84 menit. *Training loss* menurun dari 0,0254 menjadi 0,0043, dan validation loss dari 0,0074 menjadi 0,0048, menunjukkan performa solid namun kurang efisien.
- EfficientNetB0: Mencapai akurasi 98,65%, precision 99,47%, *recall* tertinggi 99,04%, dan *F1-score* 99,25%, dengan waktu pelatihan 33 menit. *Training loss* menurun dari 0,0295 menjadi 0,0079, dan validation loss dari 0,0081 menjadi 0,0054, mencerminkan keseimbangan antara akurasi dan efisiensi.

Evaluasi ini menjawab rumusan masalah ketiga, menegaskan efektivitas representasi citra *grayscale* dan mengidentifikasi EfficientNetB0 sebagai model dengan *recall* tertinggi dan efisiensi optimal.

4. Pemilihan Model Terbaik

Berdasarkan evaluasi performa dan pertimbangan efisiensi, EfficientNetB0 dipilih sebagai model terbaik untuk deteksi *malware* berbasis citra karena:

- Akurasi kompetitif (98,65%), hanya sedikit di bawah ResNet50 (98,78%).
- *Recall* tertinggi (99,04%), meminimalkan risiko *malware* yang tidak terdeteksi, yang krusial untuk keamanan siber.
- *F1-score* tinggi (99,25%), menunjukkan keseimbangan antara precision dan *recall*.
- Waktu pelatihan efisien (33 menit), lebih cepat dibandingkan ResNet50 (90 menit) dan DenseNet121 (84 menit), mendukung skalabilitas sistem. Meskipun ResNet50 unggul dalam akurasi dan *F1-score*, waktu pelatihannya yang lama kurang praktis untuk aplikasi berbasis *web*. MobileNetV3 cepat namun *recall*-nya rendah, sedangkan DenseNet121 kalah efisien. Pemilihan ini memperkuat jawaban atas rumusan masalah ketiga dan keempat.

5. Implementasi Model Terbaik pada Antarmuka *Web*

Model EfficientNetB0 diintegrasikan ke dalam aplikasi *web* berbasis Streamlit untuk deteksi *malware* . *User Acceptance Testing (UAT)* (UAT) menghasilkan indeks penerimaan 80,63% dari 8 responden, dengan antarmuka dinilai intuitif dan menarik, meskipun visualisasi citra memerlukan penjelasan tambahan.

Implementasi ini menjawab rumusan masalah keempat dengan solusi fungsional untuk deteksi awal *malware* yang dapat diakses secara praktis.

5.2 Saran

Berdasarkan temuan penelitian, beberapa saran dapat diajukan untuk pemanfaatan dan pengembangan lebih lanjut. Bagi akademisi, disarankan untuk memperluas *dataset* pelatihan dengan mengintegrasikan sampel dari MalwareBazaar dan sumber eksternal lainnya, guna meningkatkan generalisasi model dan mengatasi akurasi 70% pada data eksternal. Validasi *ground truth* melalui analisis sandbox atau *Platform* seperti VirusTotal dapat memperkuat akurasi prediksi jenis *Malware* , yang saat ini sering generik karena keterbatasan label. Praktisi keamanan siber disarankan memanfaatkan aplikasi ini sebagai alat bantu awal pada perangkat lokal, dengan mempertimbangkan efisiensinya (800-850 ms per file) pada CPU standar, serta menambahkan panduan teks di antarmuka misalnya, "Gambar ini menunjukkan distribusi *byte* untuk analisis visual" untuk meningkatkan kegunaan visualisasi.

Pembuat kebijakan di institusi teknologi dapat mendorong pengembangan versi terdistribusi dengan optimasi untuk file besar (>1 MB) dan dukungan format file tambahan (misalnya, APK, PDF), yang merupakan keterbatasan saat ini. Untuk penelitian masa depan, disarankan mengintegrasikan teknik *Fine-tuning* lanjutan atau model ensemble untuk mengatasi overlap fitur visual antar keluarga *Malware* , serta menguji aplikasi pada lingkungan cloud untuk skalabilitas. Saran ini bersumber dari pola penurunan akurasi eksternal dan umpan balik pengguna, menawarkan jalur konstruktif untuk meningkatkan kontribusi penelitian ini terhadap pengembangan solusi deteksi *Malware* yang lebih robust dan adaptif di era ancaman siber yang berkembang.

DAFTAR PUSTAKA

A. Ghaleb, F., Saeed, F., Al-Sarem, M., Ali Saleh Al-rimy, B., Boulila, W., Eljialy, A. E. M., ... & Alazab, M. (2020). Misbehavior-aware on-demand collaborative intrusion detection system using distributed ensemble learning for VANET. Electronics, 9(9), 1411.

Abhesa, R. A., & Ismail, S. J. I. (2021, November). Classification of *Malware* using *Machine Learning* based on image processing. In 2021 15th International Conference on Telecommunication Systems, Services, and *App*lications (TSSA) (pp. 1-4). IEEE.

Ahmed, A. K., Younus, S. Q., Ahmed, S. R., Algburi, S., & Fadhel, M. A. (2023, November). A *Machine Learning App*roach to Employee Performance Prediction within Administrative Information Systems. In 2023 7th International Symposium on Innovative *App*roaches in Smart Technologies (ISAS) (pp. 1-7). IEEE.

Altun, M., Gürüler, H., Özkaraca, O., Khan, F., Khan, J., & Lee, Y. (2023). Monkeypox detection using CNN with *Transfer Learning*. Sensors, 23(4), 1783.

Amos, Hua, An, Teo., Ching, Pang, Goh. (2024). 10. Oral Disease Image Detection System Using *Transfer Learning*. doi: 10.1109/icdxa61007.2024.1047051

Ashwaq, Katham, Mtashre., Dhakaa, Mohsin, Kareem., Zainab, AbdAlAbbas, Muhsen. (2024). Enhancing Object Detection Techniques Through *Transfer Learning* and *Pre-trained* Models. Mağallaẗ al-ʿulūm al-handasiyyaẗ wa-al-tiknūlūğiyā al-maʿlūmāt, 3(8):39-45. doi: 10.26389/ajsrp.k270724

Ding, Y., Zhang, X., Hu, J., & Xu, W. (2023). Android *Malware* detection method based on *byte*code image. Journal of Ambient Intelligence and Humanized Computing, 14(5), 6401-6410.

Gaber, M., Ahmed, M., & Janicke, H. (2023). *Malware* Detection with Artificial Intelligence: A Systematic Literature Review. ACM Computing Surveys, 56, 1 - 33.

Haile, M., Walle, Y. & Belay, A. (2024). Enhanced Image-Based *Malware* Multiclass Classification Method with the Ensemble Model and SVM. Open Information Science, 8(1), 20240003.

Helmi, Imaduddin., Alivia, Rahma, Sakina. (2024). 2. Eye disease detection using *Transfer Learning* based on retinal fundus image data. Indonesian Journal of Electrical Engineering and Computer Science, doi: 10.11591/ijeecs.v36.i1.pp509-516

Ismail, S.J., Gemilang, H.P., Rahardjo, B., & Hendrawan (2022). Self-Supervised Learning Implementation for *Malware* Detection. 2022 8th International Conference on Wireless and Telematics (ICWT), 1-6.

K. He and D. -S. Kim, "*Malware* Detection with *Malware* Images using *Deep Learning* Techniques," 2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), Rotorua, New Zealand, 2019, pp. 95-102, doi: 10.1109/TrustCom/BigDataSE.2019.00022.

Khaliki, M. Z., & Başarslan, M. S. (2024). Brain tumor detection from images and comparison with *Transfer Learning* methods and 3-layer CNN. Scientific Reports, 14(1), 2664.

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. ACM Computing Surveys, 55(9), 1-35.

Mao, W., Cai, Z., Towsley, D., Feng, Q., & Guan, X. (2017). Security importance assessment for system objects and *Malware* detection. Computers & Security, 68, 47-68.

Mercaldo, F., Martinelli, F., & Santone, A. (2024). Deep Convolutional Generative Adversarial Networks in Image-Based Android *Malware* Detection. Comput., 13, 154.

R. Frederick, J. Shapiro and R. A. Calix, "A Corpus of Encoded *Malware Byte* Information as Images for Efficient Classification," 2022 16th International Conference on Signal-Image Technology & Internet-Based

Systems (SITIS), Dijon, France, 2022, pp. 32-36, doi: 10.1109/SITIS57111.2022.00014.

Rajapaksha, R. W. V. P. C. (2020). Clickbait detection using multimodel fusion and *Transfer Learning* (Doctoral dissertation, Institut Polytechnique de Paris).

S. A. Roseline, S. Geetha, S. Kadry and Y. Nam, "Intelligent Vision-Based *Malware* Detection and Classification Using Deep Random Forest Paradigm," in IEEE Access, vol. 8, pp. 206303-206324, 2020, doi: 10.1109/ACCESS.2020.3036491.

Saridou, B., Moulas, I., Shiaeles, S., & Papadopoulos, B. (2023). Image-based *Malware* detection using α-cuts and *binary* visualisation. *Appl*ied Sciences, 13(7), 4624.

Tahir, R. (2018). A study on *Malware* and *Malware* detection techniques. International Journal of Education and Management Engineering, 8(2), 20

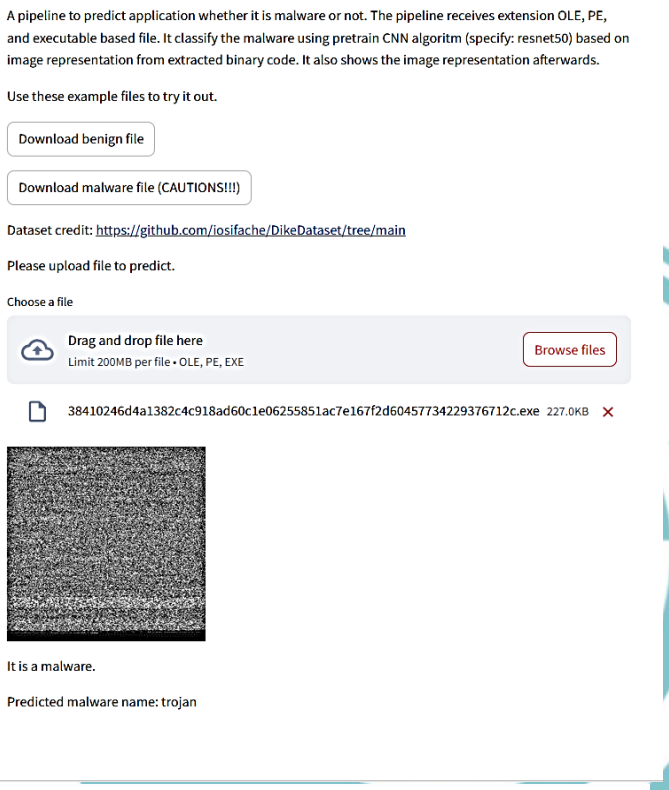
Venkatraman, S., Alazab, M., & Vinayakumar, R. (2019). A hybrid *Deep Learning* image-based analysis for effective *Malware* detection. Journal of Information Security and *App*lications, 47, 377-389.

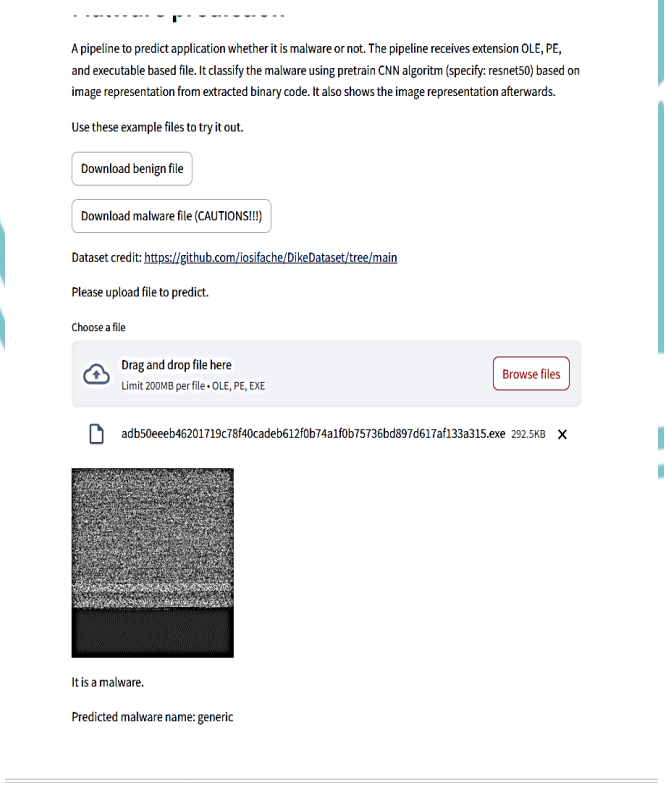
Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., & Venkatraman, S. (2019). Robust intelligent *Malware* detection using *Deep Learning*. IEEE access, 7, 46717-46738.

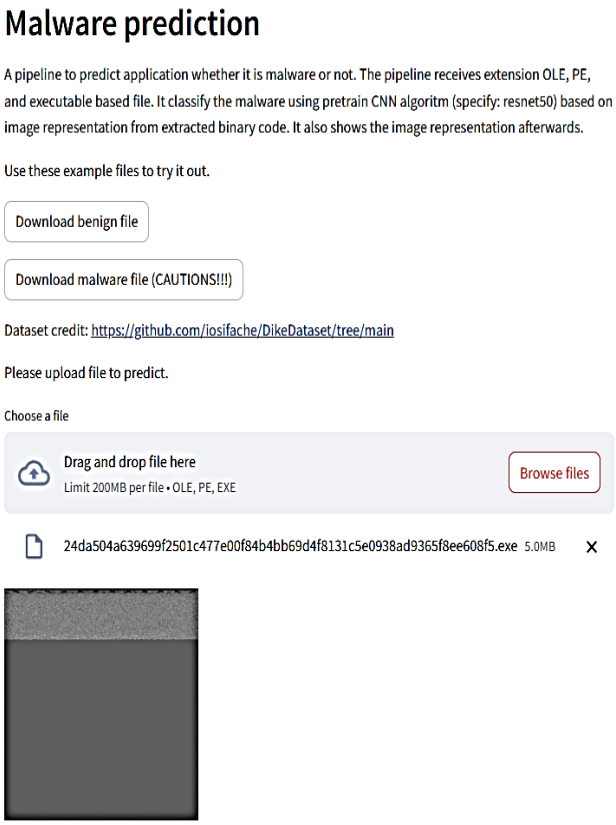
Lampiran 1. Riwayat Hidup Penulis

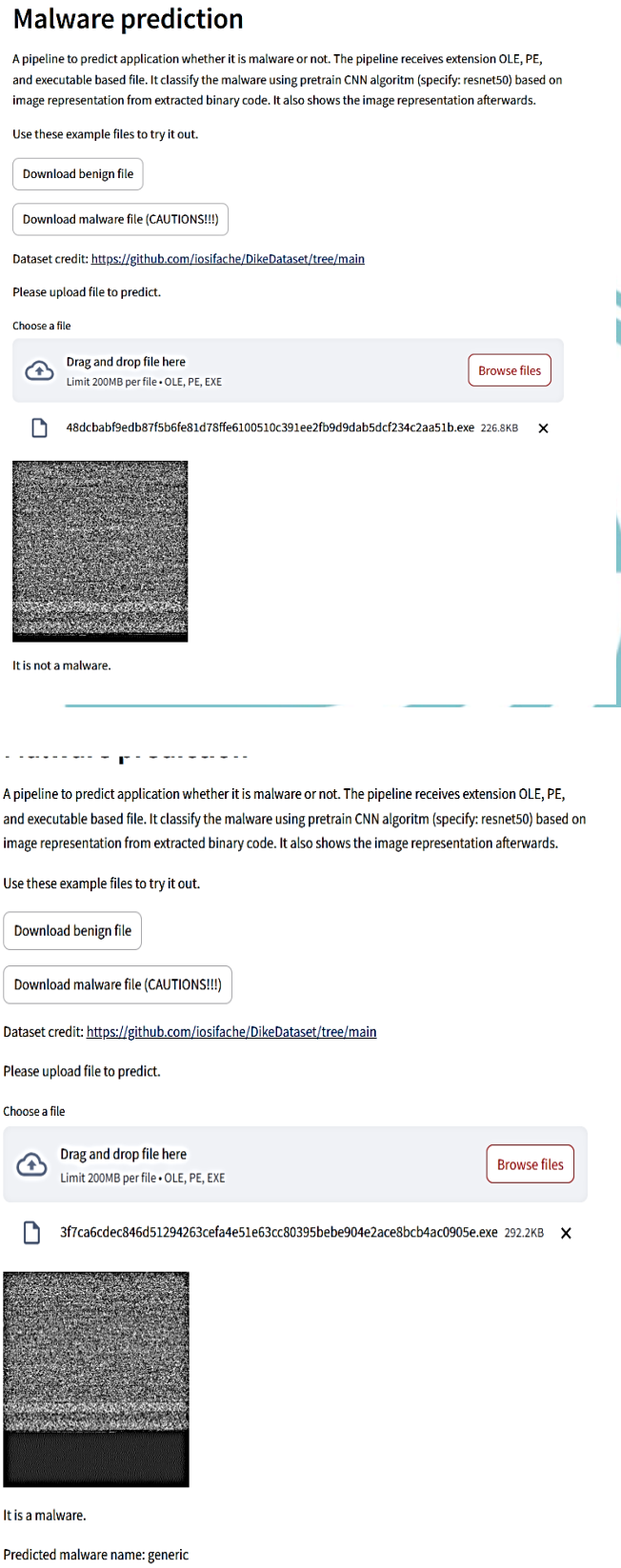
Ihsan Alamal Ahmad, lahir di Cilegon pada tanggal 13 Desember 2003, sebagai anak ketiga dari tiga bersaudara. Lulus dari SDIT Al Jannah pada tahun 2015, kemudian melanjutkan pendidikan di SMP Raudhatul Jannah dan lulus pada tahun 2018, serta menyelesaikan pendidikan di SMK IDN pada tahun 2021. Menjadi mahasiswa D4 Teknik Informatika, jurusan Teknik Informatika dan Komputer di Politeknik Negeri Jakarta pada tahun 2021.

Lampiran 2. Hasil Deteksi *Malware* Menggunakan data *External*









A pipeline to predict application whether it is malware or not. The pipeline receives extension OLE, PE, and executable based file. It classify the malware using pretrain CNN algoritm (specify: resnet50) based on image representation from extracted binary code. It also shows the image representation afterwards.

Use these example files to try it out.

Download benign file

Download malware file (CAUTIONS!!!)

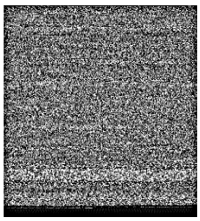
Dataset credit: https://github.com/iosifache/DikeDataset/tree/main

Please upload file to predict.

Choose a file

Drag and drop file here
Limit 200MB per file • OLE, PE, EXE
Browse files

e48deaeb2aae9b21193ac2222f7629f568dfe334c2f1ba5bfad8f5fa15e2de02.exe 226.9KB



It is a malware.

Predicted malware name: trojan

A pipeline to predict application whether it is malware or not. The pipeline receives extension OLE, PE, and executable based file. It classify the malware using pretrain CNN algoritm (specify: resnet50) based on image representation from extracted binary code. It also shows the image representation afterwards.

Use these example files to try it out.

Download benign file

Download malware file (CAUTIONS!!!)

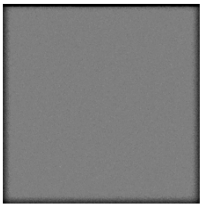
Dataset credit: https://github.com/iosifache/DikeDataset/tree/main

Please upload file to predict.

Choose a file

Drag and drop file here
Limit 200MB per file • OLE, PE, EXE
Browse files

0e04c54263dcd8deeab60c7f6f7d7223ccef8b72df707dc87557c6f5ce09b999.exe 54.8MB



It is a malware.

Predicted malware name: generic

Malware prediction

A pipeline to predict application whether it is malware or not. The pipeline receives extension OLE, PE, and executable based file. It classify the malware using pretrain CNN algoritm (specify: resnet50) based on image representation from extracted binary code. It also shows the image representation afterwards.

Use these example files to try it out.

Download benign file

Download malware file (CAUTIONS!!!)

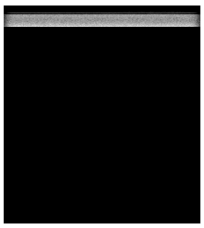
Dataset credit: https://github.com/iosifache/DikeDataset/tree/main

Please upload file to predict.

Choose a file

Drag and drop file here
Limit 200MB per file • OLE, PE, EXE
Browse files

4066909a1bbbf6cf45c130978171f682de238911c639aef2e9c4004d99016fce.exe 16.5MB



It is not a malware.

A pipeline to predict application whether it is malware or not. The pipeline receives extension OLE, PE, and executable based file. It classify the malware using pretrain CNN algoritm (specify: resnet50) based on image representation from extracted binary code. It also shows the image representation afterwards.

Use these example files to try it out.

Download benign file

Download malware file (CAUTIONS!!!)

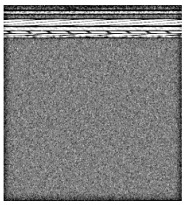
Dataset credit: https://github.com/iosifache/DikeDataset/tree/main

Please upload file to predict.

Choose a file

Drag and drop file here
Limit 200MB per file • OLE, PE, EXE
Browse files

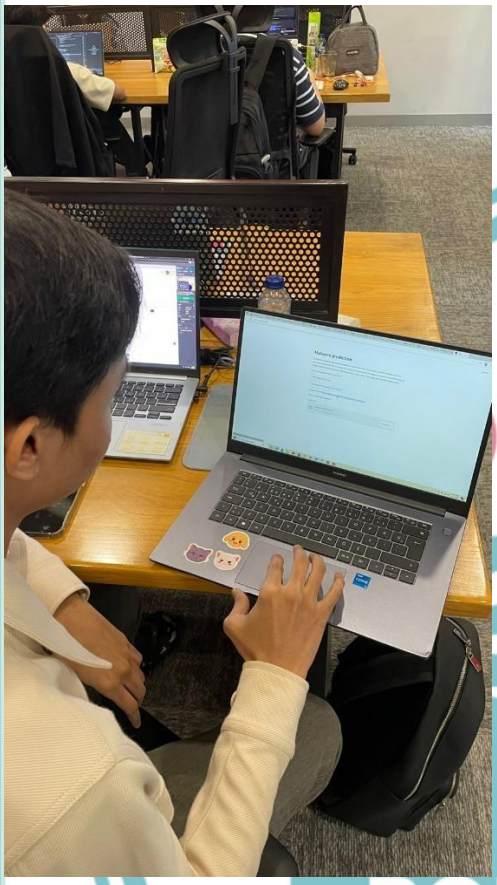
8991952056c10fe7ada82e9a944b33e372066e81101fcdf8e9de56e3ab1fbfde.exe 1.6MB

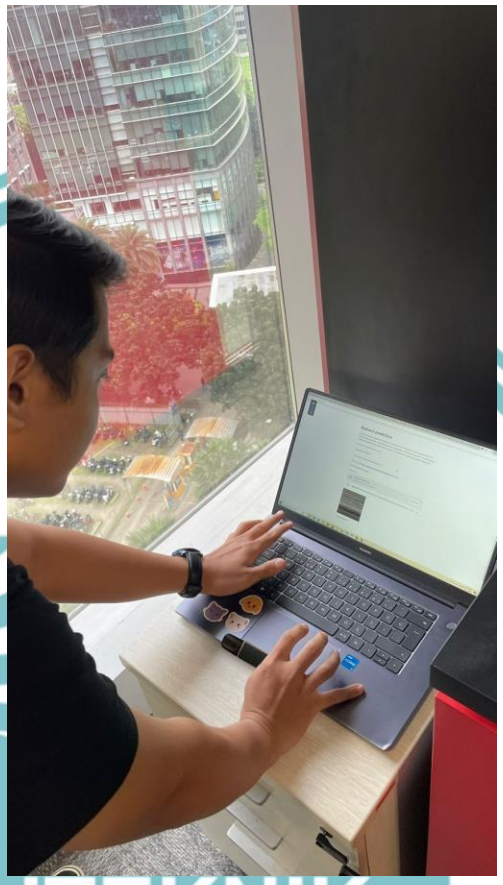


It is a malware.

Predicted malware name: generic

Lampiran 3. Foto responden mencoba simulasi deteksi *Malware* pada *web*





Lampiran 4. Transkrip Wawancara *User Acceptance Testing (UAT)* (UAT)

Tanggal Pelaksanaan: 11-31 May 2025

Waktu: 09:00 - 18:00 WIB

Lokasi: Virtual (via Google Meet), Kantor ITSEC

Peneliti: Ihsan Alamal Ahmad

Tujuan: Mengumpulkan umpan balik mengenai aplikasi *web* deteksi *malware* berbasis gambar yang dikembangkan menggunakan model *Convolutional Neural Network* (CNN) *pre-trained*, dengan representasi *grayscale* 224x224 piksel dari file biner (.exe, OLE, PE).

Responden 1: Yoga Effendy Nasution (QA Engineer, ITSEC ASIA)

Peneliti: Selamat pagi, Pak Yoga. Pertama, apakah tampilan antarmuka aplikasi ini menarik menurut Anda?

Yoga: Pagi. Iya, menurut saya cukup menarik, desainnya simpel tapi rapi. Skor saya 4 dari 5.

Peneliti: Terima kasih. Lalu, apakah kolom unggah file mudah ditemukan dan digunakan?

Yoga: Ya, sangat mudah. Tombolnya langsung terlihat di tengah, dan proses unggahnya cepat.

Peneliti: Bagus. Bagaimana dengan proses konversi file ke gambar, berjalan lancar?

Yoga: Lancar, cuma butuh beberapa detik. Tidak ada error saat saya coba upload file .exe.

Peneliti: Oke. Apakah hasil prediksi *Malware* atau *Benign* jelas dan mudah dipahami?

Yoga: Iya, hasilnya langsung muncul dengan label jelas, seperti '*Malware* ' atau '*Benign*', jadi mudah dimengerti.

Peneliti: Jika ada prediksi *Malware* , apakah informasi tambahan tentang jenisnya berguna?

Yoga: Ya, cukup membantu, terutama kalau ada detail seperti keluarga *malware* -nya, meskipun bisa lebih rinci lagi.

Peneliti: Terima kasih. Lalu, apakah visualisasi gambar *grayscale* membantu memahami struktur file?
Yoga: Sebenarnya agak membingungkan tanpa penjelasan, tapi kalau ada panduan singkat, mungkin lebih membantu.

Peneliti: Baik, saya catat. Apakah tombol unduh file contoh berfungsi dengan baik?
Yoga: Ya, berfungsi baik, file contohnya langsung terunduh tanpa masalah.

Peneliti: Terakhir, secara keseluruhan, apakah aplikasi ini memenuhi ekspektasi Anda sebagai alat deteksi *Malware* ?
Yoga: Secara umum iya, cukup memenuhi untuk deteksi awal, tapi perlu penyempurnaan di visualisasi dan detail prediksi. Skor 4 dari 5.

Responden 2: Ihza Novellino (Back-end Engineer, ITSEC ASIA)
Peneliti: Selamat pagi, Pak Ihza. Bagaimana pendapat Anda tentang tampilan antarmuka?
Ihza: Pagi. Tampilannya menarik, minimalis dan fungsional, saya beri skor 4.

Peneliti: Baik. Apakah kolom unggah file mudah ditemukan?
Ihza: Sangat mudah, posisinya strategis dan intuitif.

Peneliti: Proses konversi file ke gambar bagaimana?
Ihza: Berjalan lancar, tidak ada lag meskipun file saya cukup besar.

Peneliti: Hasil prediksi jelas dan mudah dipahami?
Ihza: Iya, labelnya langsung muncul dan informatif.

Peneliti: Informasi tambahan tentang jenis *Malware* berguna?

Ihza: Cukup berguna, tapi kalau ada analisis singkat, akan lebih baik.

Peneliti: Visualisasi *grayscale* membantu?
Ihza: Agak sulit dipahami tanpa konteks, mungkin perlu tooltip atau penjelasan.

Peneliti: Tombol unduh file contoh berfungsi?
Ihza: Ya, berfungsi dengan baik, file contohnya sesuai.

Peneliti: Keseluruhan, apakah aplikasi memenuhi ekspektasi?
Ihza: Ya, cukup memuaskan untuk prototipe, skor 4 dari 5, dengan ruang untuk perbaikan.

Responden 3: Dimas Maulana Rizky (IT Researcher, ITSEC ASIA)
Peneliti: Selamat pagi, Pak Dimas. Bagaimana menurut Anda tentang antarmuka?
Dimas: Pagi. Menarik, desainnya bersih, saya kasih 5.

Peneliti: Kolom unggah file mudah digunakan?
Dimas: Sangat mudah, langsung ketemu di halaman utama.

Peneliti: Proses konversi berjalan lancar?
Dimas: Lancar, cepat, dan tidak ada kendala.

Peneliti: Hasil prediksi jelas?
Dimas: Iya, sangat jelas, langsung tahu *Malware* atau *Benign*.

Peneliti: Informasi jenis *Malware* berguna?
Dimas: Ya, menambah nilai, terutama untuk analisis lebih lanjut.

Peneliti: Visualisasi *grayscale* membantu?
Dimas: Lumayan membantu, tapi butuh penjelasan untuk pemahaman lebih dalam.

Peneliti: Tombol unduh berfungsi?

Dimas: Berfungsi baik, file contohnya bagus untuk referensi.

Peneliti: Secara keseluruhan memenuhi ekspektasi?

Dimas: Iya, sangat memenuhi, skor 5 dari 5 untuk konsep awal.

Responden 4: Akbar Fajar Ramadhan (Fullstack Engineer, Universitas Gajah Mada)

Peneliti: Selamat pagi, Pak Akbar. Bagaimana tampilan antarmuka?

Akbar: Pagi. Cukup menarik, simpel tapi efektif, skor 4.

Peneliti: Kolom unggah file mudah ditemukan?

Akbar: Ya, mudah sekali, desainnya user-friendly.

Peneliti: Proses konversi lancar?

Akbar: Lancar, tapi agak lambat pada file besar.

Peneliti: Hasil prediksi jelas?

Akbar: Iya, cukup jelas dengan label yang tegas.

Peneliti: Informasi jenis *Malware* berguna?

Akbar: Ya, membantu untuk konteks tambahan.

Peneliti: Visualisasi *grayscale* membantu?

Akbar: Kurang membantu tanpa panduan, agak membingungkan.

Peneliti: Tombol unduh berfungsi?

Akbar: Ya, berfungsi dengan baik.

Peneliti: Keseluruhan memenuhi ekspektasi?

Akbar: Cukup memenuhi, skor 4, dengan perbaikan pada kecepatan.

Responden 5: Abidul Fikri (DevOps Engineer, ITSEC ASIA)

Peneliti: Selamat pagi, Pak Abidul. Apa pendapat Anda tentang antarmuka?

Abidul: Pagi. Menarik, desainnya modern, skor 4.

Peneliti: Kolom unggah file mudah?

Abidul: Sangat mudah, langsung terlihat.

Peneliti: Proses konversi lancar?

Abidul: Lancar, tidak ada masalah teknis.

Peneliti: Hasil prediksi jelas?

Abidul: Iya, mudah dipahami dengan output yang langsung.

Peneliti: Informasi jenis *Malware* berguna?

Abidul: Cukup berguna, bisa jadi acuan awal.

Peneliti: Visualisasi *grayscale* membantu?

Abidul: Kurang membantu, perlu penjelasan lebih.

Peneliti: Tombol unduh berfungsi?

Abidul: Ya, berfungsi dengan baik.

Peneliti: Keseluruhan memenuhi ekspektasi?

Abidul: Ya, memenuhi, skor 4 dari 5.

Responden 6: I Made Ocy (DevSecOps Engineer, ITSEC ASIA)

Peneliti: Selamat pagi, Pak Ocy. Bagaimana antarmuka menurut Anda?

Ocy: Pagi. Menarik, rapi, skor 5.

Peneliti: Kolom unggah file mudah?

Ocy: Sangat mudah, penggunaannya intuitif.

Peneliti: Proses konversi lancar?

Ocy: Lancar, prosesnya cepat.

Peneliti: Hasil prediksi jelas?

Ocy: Iya, sangat jelas dan langsung.

Peneliti: Informasi jenis *Malware* berguna?

Ocy: Ya, menambah wawasan, cukup membantu.

Peneliti: Visualisasi *grayscale* membantu?

Ocy: Agak sulit dipahami, butuh panduan.

Peneliti: Tombol unduh berfungsi?

Ocy: Berfungsi baik, tidak ada masalah.

Peneliti: Keseluruhan memenuhi ekspektasi?

Ocy: Iya, sangat memenuhi, skor 5.

Responden 7: Daffa Moreri (Front-end Engineer, Universitas Alazhar)

Peneliti: Selamat pagi, Pak Daffa. Apa pendapat Anda tentang antarmuka?

Daffa: Pagi. Cukup menarik, skor 4.

Peneliti: Kolom unggah file mudah ditemukan?

Daffa: Ya, mudah, desainnya bagus.

Peneliti: Proses konversi lancar?

Daffa: Lancar, tapi agak lambat pada file besar.

Peneliti: Hasil prediksi jelas?

Daffa: Iya, mudah dimengerti.

Peneliti: Informasi jenis *Malware* berguna?

Daffa: Cukup berguna, tapi bisa lebih detail.

Peneliti: Visualisasi *grayscale* membantu?

Daffa: Kurang membantu, agak membingungkan.

Peneliti: Tombol unduh berfungsi?

Daffa: Ya, berfungsi dengan baik.

Peneliti: Keseluruhan memenuhi ekspektasi?

Daffa: Cukup memenuhi, skor 4.

Responden 8: Haidar Rais (QA Engineer, Mekari)

Peneliti: Selamat pagi, Pak Haidar. Bagaimana tampilan antarmuka?

Haidar: Pagi. Menarik, desainnya profesional, skor 5.

Peneliti: Kolom unggah file mudah digunakan?

Haidar: Sangat mudah, langsung ketemu.

Peneliti: Proses konversi lancar?

Haidar: Lancar, tidak ada kendala.

Peneliti: Hasil prediksi jelas?

Haidar: Iya, sangat jelas dan informatif.

Peneliti: Informasi jenis *Malware* berguna?

Haidar: Ya, sangat membantu untuk analisis lanjut.

Peneliti: Visualisasi *grayscale* membantu?

Haidar: Kurang membantu tanpa penjelasan, agak membingungkan.

Peneliti: Tombol unduh berfungsi?

Haidar: Berfungsi baik, file contohnya sesuai.

Peneliti: Keseluruhan memenuhi ekspektasi?

Haidar: Iya, sangat memenuhi, skor 5.

Catatan Peneliti: Wawancara dilakukan secara online dan juga offline. antara pukul 09:00-18:00 WIB, di tanggal antara 11-31 may 2025. Responden memberikan skor rata-rata berdasarkan skala 1-5 untuk setiap pertanyaan, dengan indeks penerimaan keseluruhan 80,63% (dihitung dari rata-rata skor). Umpan balik menunjukkan antarmuka dan fungsionalitas diterima baik, namun visualisasi *grayscale* memerlukan penjelasan tambahan untuk meningkatkan pemahaman pengguna.