DETEKSI DEEPHOAX TERHADAP SERANGAN ADVERSARIAL DENGAN MODEL EFFICIENTNET-B0 UNTUK KEAMANAN SIBER

SKRIPSI

ANDREW THOMAS AGUSTINUS
00000059999

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2026

DETEKSI DEEPHOAX TERHADAP SERANGAN ADVERSARIAL DENGAN MODEL EFFICIENTNET-B0 UNTUK KEAMANAN SIBER

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

ANDREW THOMAS AGUSTINUS
00000059999

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2026

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama	:	Andrew Thomas Agustinus
Nomor Induk Mahasiswa	:	00000059999
Program Studi	:	Informatika

Skripsi dengan judul:

Deteksi Deephoax terhadap Serangan Adversarial dengan model EfficientNet-B0 untuk Keamanan Siber

merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 8 Januari 2026

(Andrew Thomas Agustinus)

HALAMAN PENGESAHAN

Skripsi dengan judul

DETEKSI DEEPHOAX TERHADAP SERANGAN ADVERSARIAL DENGAN MODEL EFFICIENTNET-B0 UNTUK KEAMANAN SIBER

Oleh

Nama	:	Andrew Thomas Agustinus
NIM	:	00000059999
Program Studi	:	Informatika
Fakultas	:	Fakultas Teknik dan Informatika

Telah diujikan pada hari Kamis, 15 Januari 2026
Pukul 10.00 s.d 12.00 dan dinyatakan
LULUS
Dengan susunan penguji sebagai berikut

Ketua Sidang

(David Agustriawan, S.Kom., M.Sc., Ph.D.)
NIDN: 0525088601

Penguji

(Alethea Suryadibrata, S.Kom., M.Eng)
NIDN: 322099201

Pembimbing

(SY.Yuliani:,S.Kom.,MT.,Ph.D.,CEH)
NIDN: 0411037904

Ketua Program Studi Informatika,

(Arya Wicaksana, S.Kom., M.Eng.Sc., OCA)
NIDN: 0315109103

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini:

Nama	:	Andrew Thomas Agustinus
NIM	:	00000059999
Program Studi	:	Informatika
Jenjang	:	S1
Judul Karya Ilmiah	:	Deteksi Deephoax terhadap Serangan Adversarial dengan model EfficientNet-B0 untuk Keamanan Siber

Menyatakan dengan sesungguhnya bahwa saya bersedia (**pilih salah satu**):

☑ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.

☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.

☐ Lainnya, pilih salah satu:

- ☐ Hanya dapat diakses secara internal Universitas Multimedia Nusantara
- ☐ Embargo publikasi karya ilmiah dalam kurun waktu tiga tahun.

Tangerang, 8 Januari 2026

Yang menyatakan

(Andrew Thomas Agustinus)

*Centang salah satu tanpa menghapus opsi yang tidak dipilih

**Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

HALAMAN PERSEMBAHAN / MOTTO

”No temptation has overtaken you except what is common to mankind. And God is faithful; he will not let you be tempted beyond what you can bear. But when you are tempted, he will also provide a way out so that you can endure it.”

1 Corinthians 10:13 NIV

”God would never put us through all this suffering if He didn’t think we could bear it”

Konno Yuuki (Zekken)

KATA PENGANTAR

Puji Syukur atas selesainya penulisan Laporan Skripsi ini dengan judul: ”Deteksi Deephoax terhadap Serangan Adversarial dengan model EfficientNet-B0 untuk Keamanan Siber” dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan tugas akhir ini, sangatlah sulit bagi saya untuk menyelesaikan tugas akhir ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ir. Andrey Andoko, M.Sc., selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Bapak Arya Wicaksana, S.Kom., M.Eng.Sc., OCA, selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Ibu SY.Yuliani.,S.Kom.,MT.,Ph.D.,CEH, sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Orang Tua dan keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan laporan penelitian ini.

Semoga karya ilmiah ini dapat memberikan manfaat yang baik sebagai sumber informasi maupun inspirasi seluruh kalangan.

Tangerang, 8 Januari 2026

Andrew Thomas Agustinus

DETEKSI DEEPHOAX TERHADAP SERANGAN ADVERSARIAL DENGAN MODEL EFFICIENTNET-B0 UNTUK KEAMANAN SIBER

Andrew Thomas Agustinus

ABSTRAK

Perkembangan pesat teknologi *deephoax* memungkinkan manipulasi wajah manusia secara realistis, namun penggunaannya semakin meluas untuk disinformasi dan ancaman keamanan siber. Sistem deteksi *deephoax* berbasis *deep learning* masih rentan terhadap serangan *adversarial*, khususnya metode *Projected Gradient Descent* (PGD), yang dapat menurunkan akurasi model secara signifikan. Penelitian ini mengkaji penerapan *adversarial training* (AT) berbasis PGD pada arsitektur EfficientNet-B0 untuk meningkatkan ketahanan detektor terhadap serangan adversarial sekaligus mempertahankan efisiensi komputasi. Metodologi meliputi pembangunan model deteksi, pra-pemrosesan citra wajah, serta pelatihan menggunakan tiga skenario: Baseline, Adversarial Training, dan Mixed AT. Evaluasi dilakukan pada data bersih serta data yang dimodifikasi menggunakan PGD dengan iterasi 5, 10, dan 20. Hasil menunjukkan model Baseline memiliki *clean accuracy* 98,7% namun akurasinya 0% terhadap serangan PGD. Model AT mempertahankan akurasi adversarial sekitar 49,7%. Sementara itu, Mixed AT menunjukkan performa seimbang dengan *clean accuracy* 95,2% dan akurasi PGD 50,4 49,5% (PGD-5/10) serta 32,2% (PGD-20). Penelitian menegaskan bahwa *adversarial training* efektif meningkatkan robustness, dan Mixed AT menyediakan keseimbangan terbaik antara akurasi dan ketahanan terhadap serangan.

Kata kunci: *adversarial training*, *deephoax*, *EfficientNet-B0*, *Projected Gradient Descent* (PGD)

DEEPHOAX DETECTION AGAINST ADVERSARIAL ATTACKS USING EFFICIENTNET-B0 MODEL FOR CYBERSECURITY

Andrew Thomas Agustinus

ABSTRACT

The rapid development of deephoax technology enables realistic manipulation of human faces, yet its usage has increasingly been exploited for disinformation and cyber security threats. Deep learning–based deephoax detection systems remain vulnerable to adversarial attacks, particularly the Projected Gradient Descent (PGD) method, which can significantly reduce model accuracy. This study investigates the application of PGD-based adversarial training (AT) on the EfficientNet-B0 architecture to enhance detector robustness against adversarial attacks while maintaining computational efficiency. The methodology includes model development, facial image preprocessing, and training using three scenarios: Baseline, Adversarial Training, and Mixed AT. Evaluation was conducted on clean data as well as data modified with PGD at 5, 10, and 20 iterations. Results show that the Baseline model achieved 98.7% clean accuracy but 0% accuracy under PGD attacks. The AT model maintained adversarial accuracy around 49.7%. Meanwhile, the Mixed AT model demonstrated a balanced performance, with 95.2% clean accuracy and PGD accuracy of 50.4%,49,5% (PGD-5/10) and 32.2% (PGD-20). The study confirms that adversarial training effectively enhances robustness, and Mixed AT provides the best trade-off between accuracy and resilience against attacks.

***Keywords**: adversarial training, deephoax, EfficientNet-B0, Projected Gradient Descent (PGD)*

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN