

**PENERAPAN EXPLAINABLE ARTIFICIAL INTELLIGENCE
(XAI) PADA MODEL DEEP LEARNING UNTUK DETEKSI
GAMBAR DEEPPAKE MENGGUNAKAN GRAD-CAM**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

NIZAR MOHAMMAD BINTANG TRI PRATAMA
22.11.4813

Kepada

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026

**PENERAPAN EXPLAINABLE ARTIFICIAL INTELLIGENCE
(XAI) PADA MODEL DEEP LEARNING UNTUK DETEKSI
GAMBAR DEEPPAKE MENGGUNAKAN GRAD-CAM**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

NIZAR MOHAMMAD BINTANG TRI PRATAMA
22.11.4813

Kepada

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026

HALAMAN PERSETUJUAN

SKRIPSI

**PENERAPAN EXPLAINABLE ARTIFICIAL INTELLIGENCE
(XAI) PADA MODEL DEEP LEARNING UNTUK DETEKSI
GAMBAR DEEPPAKE MENGGUNAKAN GRAD-CAM**

yang disusun dan diajukan oleh

Nizar Mohammad Bintang Tri Pratama

22.11.4813

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 21 Mei 2026

Dosen Pembimbing,



I Made Artha Agastva, S.T., M.Eng., Ph.D.

NIK. 190302352

HALAMAN PENGESAHAN

SKRIPSI

PENERAPAN EXPLAINABLE ARTIFICIAL INTELLIGENCE
(XAI) PADA MODEL DEEP LEARNING UNTUK DETEKSI
GAMBAR DEEPPAKE MENGGUNAKAN GRAD-CAM

yang disusun dan diajukan oleh

Nizar Mohammad Bintang Tri Pratama

22.11.4813

Telah dipertahankan di depan Dewan Penguji
pada tanggal 21 Mei 2026

Susunan Dewan Penguji

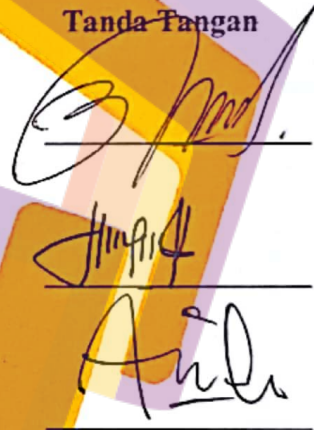
Nama Penguji

Ali Mustopa, S.Kom., M.Kom.
NIK. 190302192

Dr. Hartatik, S.T., M.Cs.
NIK. 190302232

I Made Artha Agastya, S.T., M.Eng., Ph.D.
NIK. 190302352

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 21 Mei 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Nizar Mohammad Bintang Tri Pratama
NIM : 22.11.4813

Menyatakan bahwa Skripsi dengan judul berikut:

**PENERAPAN EXPLAINABLE ARTIFICIAL INTELLIGENCE (XAI)
PADA MODEL DEEP LEARNING UNTUK DETEKSI GAMBAR
DEEPPAKE MENGGUNAKAN GRAD-CAM**

Dosen Pembimbing : I Made Artha Agastya, S.T., M.Eng., Ph.D.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 21 Mei 2026

Yang Menyatakan,



Nizar Mohammad Bintang Tri Pratama

HALAMAN PERSEMBAHAN

Alhamdulillahirabbil'alamin, segala puji dan syukur kehadiran Allah SWT, Tuhan Yang Maha Esa, atas limpahan berkat, rahmat, dan karunia-Nya yang tiada habisnya. Sholawat serta salam juga penulis curahkan pada Nabi Muhammad SAW, yang menjadi suri tauladan bagi seluruh umat manusia. Dengan penuh rasa syukur, penulis mempersembahkan skripsi ini kepada:

1. Allah SWT, Yang Maha Pengasih lagi Penyayang, atas kasih sayang, kesehatan, perlindungan, beserta petunjuk dan kenikmatan yang tak terhingga, sehingga skripsi ini dapat diselesaikan dengan tepat waktu.
2. Orang tua dan keluarga tercinta, atas segala doa, kasih sayang, dan dukungan yang tiada hentinya, baik secara moral maupun material di setiap langkah perjalanan penulis hingga sampai di titik ini.
3. Dosen Pembimbing, Bapak I Made Artha Agastya, S.T., M.Eng., Ph.D, yang telah membimbing, memberi arahan, masukan, serta kemudahan selama proses penelitian dan penulisan skripsi.
4. Seluruh pimpinan, dosen, dan staf Universitas AMIKOM Yogyakarta yang telah memberikan ilmu dan kesempatan yang sangat berharga selama menempuh jenjang studi di perguruan tinggi.
5. Rekan-rekan dan sahabat selama masa perkuliahan, yang telah berbagi ilmu, dukungan, kebersamaan, serta canda tawa yang akan selalu dikenang.
6. Diri penulis sendiri, Nizar Mohammad Bintang Tri Pratama, atas keteguhan, kesabaran, serta kerja keras dalam segala rintangan yang dihadapi. Terima kasih telah bertahan, pantang menyerah, dan berjuang hingga mampu menempuh langkah sejauh ini. Masih akan ada banyak lagi tantangan yang akan menghadang di masa depan, tetaplah semangat dan jangan putus asa. Sekali lagi, terima kasih.

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadirat Allah SWT atas limpahan rahmat, karunia, dan petunjuk-Nya sehingga penulis bisa menyelesaikan skripsi ini dengan judul “Penerapan Explainable Artificial Intelligence (XAI) Pada Model Deep Learning Untuk Deteksi Gambar Deepfake Menggunakan Grad-CAM”. Skripsi ini disusun sebagai salah satu syarat kelulusan dan untuk memperoleh gelar Sarjana program studi S1 Informatika, Fakultas Ilmu Komputer, di Universitas AMIKOM Yogyakarta. Penyusunan skripsi ini tidak terlepas dari dukungan, bimbingan, dan bantuan dari berbagai pihak. Oleh karena itu, penulis sampaikan hormat dan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Prof. Dr. Kusriani, M.Kom., selaku Dekan Fakultas Ilmu Komputer.
3. Ibu Eli Pujiastuti, M.Kom., selaku Ketua Program Studi Informatika.
4. Bapak I Made Artha Agastya, S.T, M.Eng., Ph.D., selaku Dosen Pembimbing, atas bimbingan, arahan, dan masukan yang membangun selama proses pembuatan skripsi.
5. Bapak/Ibu Dosen Penguji, atas kritik, saran, dan masukan yang diberikan untuk penyempurnaan skripsi ini.
6. Seluruh Dosen dan Staf Universitas AMIKOM Yogyakarta, yang telah memberikan ilmu dan pengalaman berharga selama masa perkuliahan.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan dan jauh dari kata sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun untuk perbaikan di masa depan. Semoga skripsi ini dapat memberikan manfaat pada perkembangan ilmu pengetahuan dan teknologi, khususnya deep learning dan sistem deteksi deepfake.

Yogyakarta, 21 Mei 2026

Penulis

DAFTAR ISI

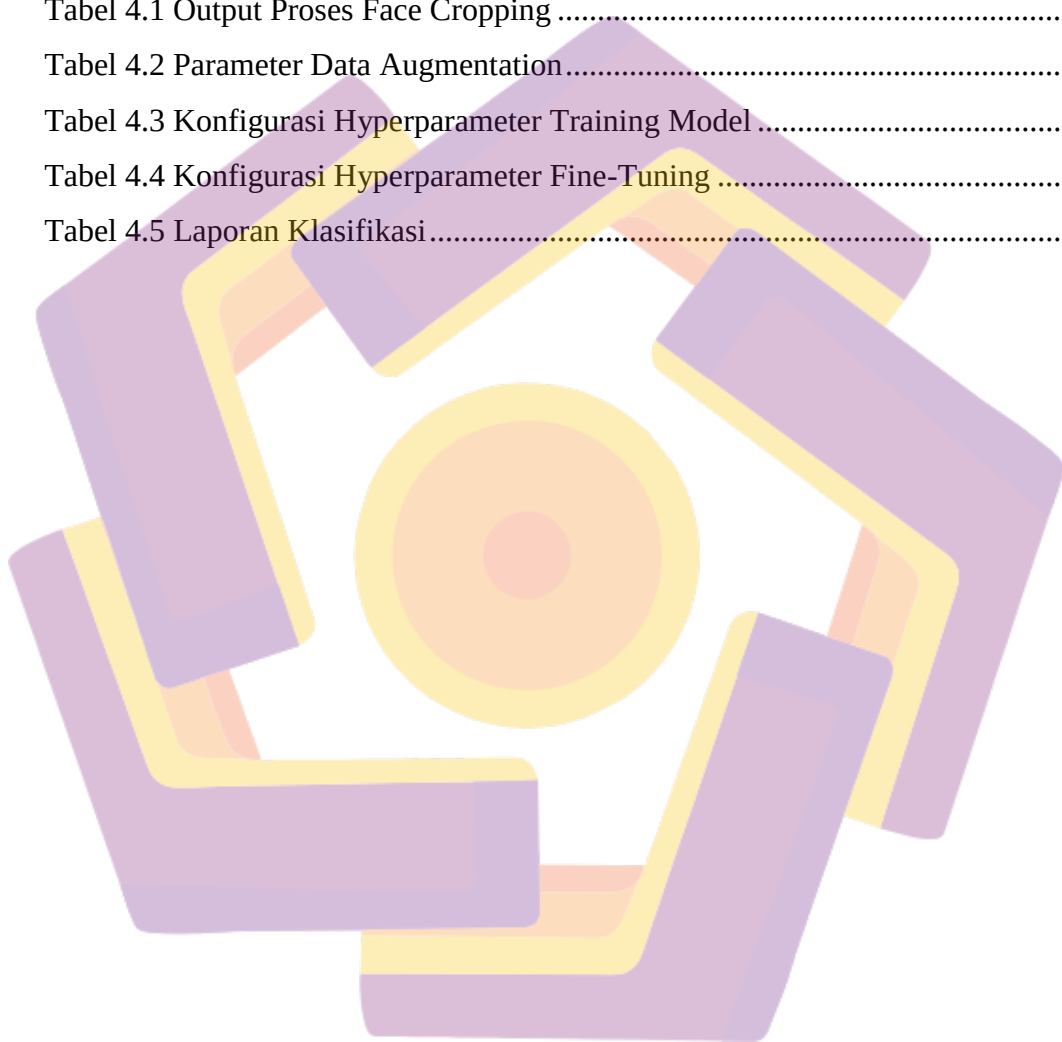
HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH	xiii
INTISARI	xvi
ABSTRACT.....	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	7
2.1 Studi Literatur	7
2.2 Dasar Teori	16
2.2.1 Deepfake	16
2.2.2 Deteksi Deepfake	16
2.2.3 Deep Learning.....	16
2.2.4 Convolutional Neural Network (CNN).....	17
2.2.5 Transfer Learning.....	18

2.2.6	Arsitektur Xception.....	19
2.2.7	Multi-Task Cascaded Convolutional Networks (MTCNN).....	21
2.2.8	Explainable Artificial Intelligence (XAI)	22
2.2.9	Gradient Class Activation Mapping (Grad-CAM)	22
2.2.10	Evaluasi Model	24
BAB III METODE PENELITIAN		26
3.1	Objek Penelitian.....	26
3.2	Alur Penelitian	26
3.2.1	Identifikasi Masalah.....	27
3.2.2	Studi Literatur	28
3.2.3	Pengumpulan Data	28
3.2.4	Data Preprocessing.....	29
3.2.5	Resampling Data.....	29
3.2.6	Face Cropping.....	29
3.2.7	Augmentasi Data.....	30
3.2.8	Modeling.....	30
3.2.9	Training Model	31
3.2.10	Fine Tuning.....	31
3.2.11	Evaluasi.....	32
3.2.12	Implementasi Grad-CAM	32
3.3	Alat dan Bahan.....	32
3.3.1	Data Penelitian	32
3.3.2	Alat/Instrumen Penelitian	33
BAB IV HASIL DAN PEMBAHASAN		34
4.1.	Pengumpulan Data.....	34
4.2.	Data Preprocessing	35

4.2.1	Resampling Data	35
4.2.2	Face Cropping	36
4.2.3	Data Augmentation	38
4.3.	Modeling	39
4.3.1	Training Model	40
4.3.2	Fine-Tuning	42
4.4.	Evaluasi	44
4.5.	Implementasi Grad-CAM	46
4.6.	Pembahasan	50
BAB V	PENUTUP	52
5.1	Kesimpulan	52
5.2	Saran	53
REFERENSI		54
LAMPIRAN		59

DAFTAR TABEL

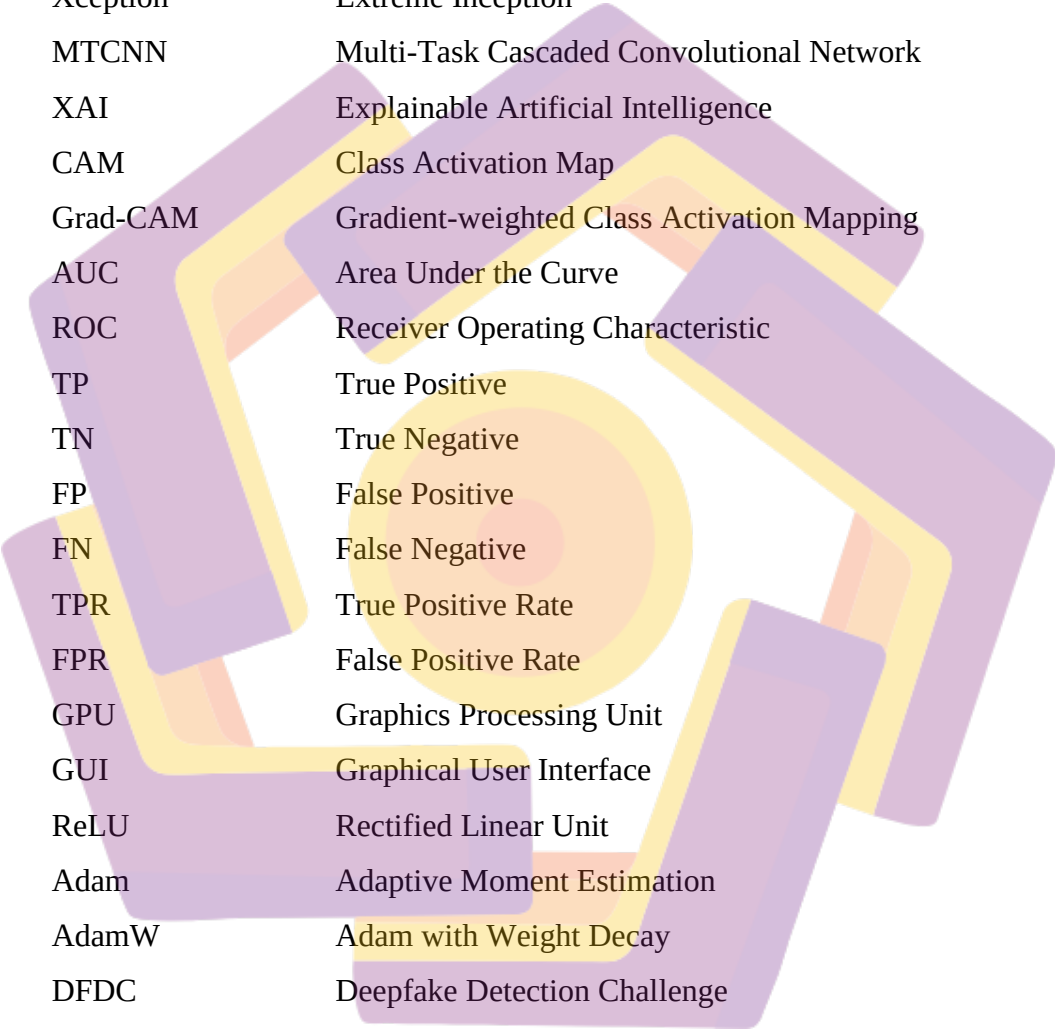
Tabel 2.1 Keaslian Penelitian	10
Tabel 2.2 Confusion Matrix	24
Tabel 3.1 Konfigurasi Model	31
Tabel 4.1 Output Proses Face Cropping	37
Tabel 4.2 Parameter Data Augmentation	39
Tabel 4.3 Konfigurasi Hyperparameter Training Model	41
Tabel 4.4 Konfigurasi Hyperparameter Fine-Tuning	43
Tabel 4.5 Laporan Klasifikasi	44



DAFTAR GAMBAR

Gambar 2.1 Arsitektur CNN	17
Gambar 2.2 Ilustrasi Konsep Transfer Learning.....	19
Gambar 2.3 Perbedaan <i>Depthwise Separable</i> dan <i>Standard Convolution</i>	20
Gambar 2.4 Arsitektur Model Xception	20
Gambar 2.5 Arsitektur MTCNN	21
Gambar 2.6 Ilustrasi Proses Grad-CAM pada Klasifikasi Deepfake.....	23
Gambar 3.1 Alur Penelitian	27
Gambar 4.1 Sampel Wajah Asli dan Wajah Palsu.....	34
Gambar 4.2 Grafik Distribusi Dataset Setelah Resampling.....	36
Gambar 4.3 Contoh Hasil Proses Face Cropping	38
Gambar 4.4 Contoh Gambar Sebelum dan Sesudah Augmentasi.....	39
Gambar 4.5 Grafik Hasil Training Model.....	42
Gambar 4.6 Grafik Hasil Fine-Tuning.....	43
Gambar 4.7 Hasil Confusion Matrix.....	45
Gambar 4.8 Hasil AUC dan ROC Curve	46
Gambar 4.9 Implementasi Grad-CAM pada Wajah Deepfake	47
Gambar 4.10 Implementasi Grad-CAM pada Wajah Asli.....	48
Gambar 4.11 Implementasi Grad-CAM pada Kasus Kesalahan Prediksi	49

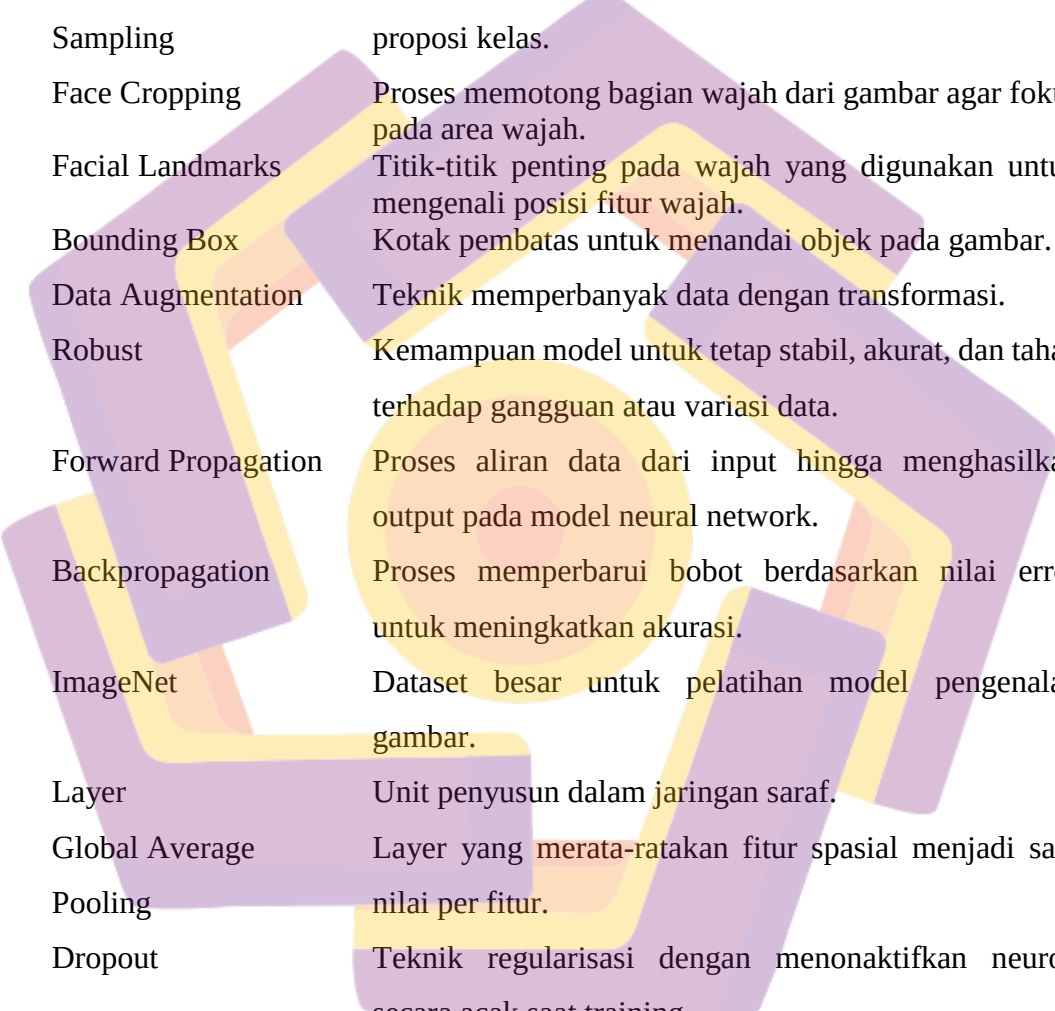
DAFTAR LAMBANG DAN SINGKATAN



AI	Artificial Intelligence
GAN	Generative Adversarial Network
CNN	Convolutional Neural Network
Xception	Extreme Inception
MTCNN	Multi-Task Cascaded Convolutional Network
XAI	Explainable Artificial Intelligence
CAM	Class Activation Map
Grad-CAM	Gradient-weighted Class Activation Mapping
AUC	Area Under the Curve
ROC	Receiver Operating Characteristic
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
TPR	True Positive Rate
FPR	False Positive Rate
GPU	Graphics Processing Unit
GUI	Graphical User Interface
ReLU	Rectified Linear Unit
Adam	Adaptive Moment Estimation
AdamW	Adam with Weight Decay
DFDC	Deepfake Detection Challenge
ViT	Vision Transformer
LIME	Local Interpretable Model-agnostic Explanation
SHAP	Shapely Additive Explanations

DAFTAR ISTILAH

Deepfake	Teknologi AI untuk membuat atau memanipulasi video/gambar agar tampak nyata.
Face Swap	Teknik mengganti wajah seseorang dengan wajah orang lain dalam gambar/video.
Artificial Intelligence	Cabang ilmu komputer yang membuat mesin dapat meniru kecerdasan manusia.
Machine Learning	Metode AI yang memungkinkan sistem belajar dari data untuk membuat prediksi atau keputusan.
Deep Learning	Metode AI berbasis jaringan saraf dalam untuk mempelajari pola kompleks.
Black-box	Model yang proses internalnya sulit dipahami oleh manusia.
Xception	Arsitektur CNN berbasis depthwise separable convolution untuk efisiensi dan performa tinggi.
Transfer Learning	Menggunakan model yang sudah dilatih sebelumnya untuk tugas baru.
Fine-Tuning	Penyesuaian lanjutan pada model pretrained agar sesuai dengan data baru.
Pretrained Model	Model yang telah dilatih sebelumnya pada dataset besar.
Library	Kumpulan fungsi atau modul siap pakai dalam pemrograman
TensorFlow	Library open-source untuk pengembangan model AI.
Keras	API tingkat tinggi untuk membangun model deep learning.
Matplotlib	Library untuk visualisasi data dalam bentuk grafik.
Seaborn	Library visualisasi berbasis Matplotlib.
Scikit-learn (Sklern)	Library Python untuk machine learning klasik.
OpenCV	Library untuk pengolahan citra dan computer vision.



Pipeline	Rangkaian proses pengolahan data dan pelatihan model.
Data Preprocessing	Tahap awal untuk mempersiapkan data sebelum digunakan dalam model.
Resample	Teknik mengambil ulang data untuk menyeimbangkan distribusi atau meningkatkan kualitas data.
Stratified Random Sampling	Pengambilan sampel acak dengan mempertahankan proporsi kelas.
Face Cropping	Proses memotong bagian wajah dari gambar agar fokus pada area wajah.
Facial Landmarks	Titik-titik penting pada wajah yang digunakan untuk mengenali posisi fitur wajah.
Bounding Box	Kotak pembatas untuk menandai objek pada gambar.
Data Augmentation	Teknik memperbanyak data dengan transformasi.
Robust	Kemampuan model untuk tetap stabil, akurat, dan tahan terhadap gangguan atau variasi data.
Forward Propagation	Proses aliran data dari input hingga menghasilkan output pada model neural network.
Backpropagation	Proses memperbarui bobot berdasarkan nilai error untuk meningkatkan akurasi.
ImageNet	Dataset besar untuk pelatihan model pengenalan gambar.
Layer	Unit penyusun dalam jaringan saraf.
Global Average Pooling	Layer yang merata-ratakan fitur spasial menjadi satu nilai per fitur.
Dropout	Teknik regularisasi dengan menonaktifkan neuron secara acak saat training.
Dense	Layer fully connected dalam jaringan saraf.
Sigmoid	Fungsi aktivasi yang menghasilkan output antara 0 dan 1.
Overfitting	Kondisi saat model terlalu menyesuaikan data training sehingga kurang baik pada data baru.
Hyperparameter	Parameter yang ditentukan sebelum training.



Batch	Sekelompok data yang diproses dalam satu iterasi.
Epoch	Satu siklus penuh pelatihan pada seluruh data training.
Optimizer	Algoritma untuk memperbarui bobot model agar meminimalkan loss.
Adam	Optimizer yang menggabungkan momentum dan adaptive learning rate.
AdamW	Varian optimizer Adam dengan weight decay terpisah untuk regularisasi yang lebih baik.
Learning Rate	Parameter yang mengatur besar langkah pembaruan bobot.
Loss Function	Fungsi untuk mengukur kesalahan antara prediksi model dan nilai sebenarnya.
Binary Cross-Entropy	Fungsi loss untuk klasifikasi biner.
Callbacks	Fungsi tambahan yang berjalan selama proses training.
ModelCheckpoint	Callback untuk menyimpan model terbaik selama training.
EarlyStopping	Callback untuk menghentikan training saat performa tidak meningkat.
ReduceLROnPlateau	Callback untuk menurunkan learning rate saat performa stagnan.
Confidence	Tingkat keyakinan prediksi model.
Overlay	Proses menampilkan objek atau informasi tambahan di atas gambar utama.
Heatmap	Visualisasi yang menunjukkan intensitas atau fokus model pada bagian tertentu.
Runtime	Fase saat program atau aplikasi sedang berjalan pada sistem komputer

INTISARI

Perkembangan teknologi kecerdasan buatan telah memungkinkan pembuatan konten manipulasi wajah yang dikenal sebagai deepfake, yang berpotensi menimbulkan dampak serius seperti penyebaran informasi palsu, penipuan digital, dan ancaman terhadap privasi individu. Meskipun model deep learning mampu mendeteksi deepfake dengan akurasi tinggi, sifat *black-box* dari model tersebut menyebabkan sulitnya memahami proses pengambilan keputusan yang dihasilkan. Penelitian ini bertujuan untuk mengembangkan sistem deteksi deepfake berbasis arsitektur Xception dengan pendekatan *transfer learning* dan *fine-tuning*, serta menerapkan metode *Explainable Artificial Intelligence* (XAI) menggunakan teknik *Gradient-weighted Class Activation Mapping* (Grad-CAM) untuk meningkatkan transparansi model. Dataset yang digunakan adalah CelebDF-V2 yang terdiri dari 20.000 citra wajah asli dan deepfake, dengan tahapan praproses meliputi *resampling*, *face cropping* menggunakan MTCNN, dan augmentasi data. Hasil evaluasi menunjukkan bahwa model berhasil mencapai akurasi sebesar 93,6%, nilai *recall* kelas Fake sebesar 0,98, dan nilai AUC sebesar 0,9888, yang mengindikasikan kemampuan diskriminasi model yang sangat baik. Visualisasi Grad-CAM membuktikan bahwa model memfokuskan perhatiannya pada area wajah yang relevan secara semantik dalam proses klasifikasi, sehingga meningkatkan interpretabilitas dan kepercayaan terhadap sistem. Hasil penelitian ini dapat dimanfaatkan oleh peneliti, praktisi keamanan digital, jurnalis, serta penegak hukum dalam upaya verifikasi keaslian konten visual di era kecerdasan buatan.

Kata kunci: Deepfake, Xception, Deep Learning, Grad-CAM, Explainable Artificial Intelligence

ABSTRACT

The advancement of artificial intelligence technology has enabled the creation of manipulated facial content known as deepfake, which poses serious threats such as the spread of misinformation, digital fraud, and violations of individual privacy. Although deep learning models are capable of detecting deepfakes with high accuracy, the black-box nature of these models makes it difficult to understand their decision-making process. This study aims to develop a deepfake detection system based on the Xception architecture using transfer learning and fine-tuning approaches, and to apply an Explainable Artificial Intelligence (XAI) method through Gradient-weighted Class Activation Mapping (Grad-CAM) to enhance model transparency. The dataset used is CelebDF-V2, consisting of 20,000 real and deepfake facial images, with preprocessing stages including resampling, face cropping using MTCNN, and data augmentation. Evaluation results show that the model achieved an accuracy of 93,6%, a recall of 0.98 for the Fake class, and an AUC value of 0.9888, indicating excellent discriminative capability. Grad-CAM visualizations confirm that the model focuses its attention on semantically relevant facial regions during classification, thereby improving interpretability and trustworthiness of the system. The findings of this study can be utilized by researchers, digital security practitioners, journalists, and law enforcement agencies in verifying the authenticity of visual content in the era of artificial intelligence.

Keyword: Deepfake, Xception, Deep Learning, Grad-CAM, Explainable Artificial Intelligence