

**ANALISIS KOMPARATIF VISION TRANSFORMER (VIT)
MENGUNAKAN HYBRID FINE-TUNING–XGBOOST
UNTUK KLASIFIKASI PENYAKIT KATARAK
BERBASIS CITRA MATA**

SKRIPSI

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



disusun oleh

NADIYATULHUSNA

22.11.5274

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS KOMPARATIF VISION TRANSFORMER (VIT)
MENGUNAKAN HYBRID FINE-TUNING–XGBOOST
UNTUK KLASIFIKASI PENYAKIT KATARAK
BERBASIS CITRA MATA**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

NADIYATULHUSNA

22.11.5274

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS KOMPARATIF VISION TRANSFORMER (VIT)
MENGUNAKAN HYBRID FINE-TUNING-XGBOOST UNTUK
KLASIFIKASI PENYAKIT KATARAK
BERBASIS CITRA MATA**

yang disusun dan diajukan oleh

Nadiyahulhusna

22.11.5274

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 25 Mei 2026

Dosen Pembimbing,



Ferian Fauzi Abdullah, S.Kom., M.Kom.
NIK. 190302276

HALAMAN PENGESAHAN
SKRIPSI
ANALISIS KOMPARATIF VISION TRANSFORMER (VIT)
MENGGUNAKAN HYBRID FINE-TUNING–XGBOOST UNTUK
KLASIFIKASI PENYAKIT KATARAK BERBASIS CITRA MATA

yang disusun dan diajukan oleh

Nadiyahulhusna

22.11.5274

Telah dipertahankan di depan Dewan Penguji
pada tanggal 25 Mei 2026

Susunan Dewan Penguji

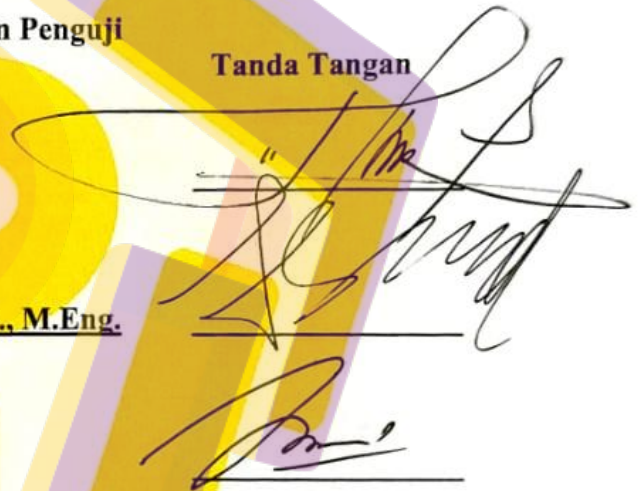
Nama Penguji

Haryoko, S.Kom., M.Cs.
NIK. 190302286

Surya Tri Atmaja Ramadhani, S.Kom., M.Eng.
NIK. 190302481

Bayu Setiaji, M.Kom.
NIK. 190302216

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 25 Mei 2026

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Nadiyahulhusna
NIM : 22.11.5274

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Komparatif Vision Transformer (VIT) Menggunakan Hybrid Fine-Tuning–XGBoost untuk Klasifikasi Penyakit Katarak Berbasis Citra Mata

Dosen Pembimbing : Ferian Fauzi Abdulloh, S.Kom., M.Kom.

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, 25 Mei 2026

Yang Menyatakan,



Nadiyahulhusna

HALAMAN PERSEMBAHAN

Sebagai bentuk syukur atas selesainya amanah pendidikan ini, skripsi ini penulis persembahkan untuk kedua orang tua dan keluarga, atas doa dan segala jalan kehidupan yang telah mengantarkan penulis hingga ke titik ini.

Terima kasih juga penulis haturkan kepada seluruh civitas akademika Universitas AMIKOM Yogyakarta. Terima kasih atas kesabaran, bimbingan, dan ruang diskusi yang mendukung penulis untuk terus berkembang dari awal hingga garis akhir.

Untuk sebuah perjalanan yang seringkali terasa berat, penulis sangat bersyukur dikelilingi oleh teman-teman, khususnya Hayya dan Niken. Terima kasih sudah bersedia menjadi pendengar yang baik, teman bertukar cerita, dan bahu untuk bersandar. Kehadiran kalian membuat beban dan keluh kesah di perjalanan ini terasa jauh lebih ringan.

Di sela-sela rasa lelah yang menghampiri, selalu ada ruang jeda yang menjadi penawar. Terima kasih untuk berbagai karya yang menjadi tempat pelarian ternyaman. Karakter seperti Kim Dokja yang menyadarkan bahwa sekadar menjadi manusia biasa yang bertahan membaca hingga akhir cerita pun sudah cukup, dan Sakata Gintoki yang mengingatkan untuk tetap menjalani hidup sesuai keinginan sendiri dan melindungi apa yang ada di depan mata. Tak lupa, lagu "*This Is Me*" dari *The Greatest Showman* yang senantiasa menjadi pengingat untuk berani menerima dan bangga pada diri sendiri.

Terakhir dan yang paling utama, persembahan ini adalah apresiasi untuk diri saya sendiri. Terima kasih sudah memilih untuk bertahan, berjuang, dan menolak kalah pada keadaan yang sulit. Halaman ini adalah wujud penghargaan paling jujur atas tanggung jawab, air mata, dan komitmen yang telah dijaga sedemikian rupa hingga perjalanan ini berhasil diselesaikan.

KATA PENGANTAR

Puji syukur ke hadirat Allah SWT atas segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul "Analisis Komparatif Vision Transformer (VIT) Menggunakan Hybrid Fine-Tuning–XGBoost untuk Klasifikasi Penyakit Katarak". Penulisan skripsi ini tidak lepas dari dukungan, bimbingan, serta bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Ferian Fauzi Abdulloh, S.Kom., M.Kom., selaku Dosen Pembimbing yang telah meluangkan waktu dan pikirannya untuk memberikan bimbingan kepada penulis. Gelar akademik dan nama beliau telah dituliskan secara lengkap sesuai dengan ketentuan penulisan.
2. Tim Dosen Penguji yang telah memberikan evaluasi dan masukan yang membangun demi kesempurnaan penyusunan skripsi ini.
3. Prof. Dr. Kusrini, M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
4. Seluruh dosen dan staf Program Studi Informatika yang telah memberikan ilmu dan pengalaman selama masa perkuliahan.
5. Kedua orang tua atas doa dan dukungan untuk penulis hingga dapat menyelesaikan pendidikan ini.
6. Sahabat dan rekan-rekan mahasiswa kelas IF 12 program studi Informatika Universitas Amikom Yogyakarta yang telah memberikan semangat.

Penulis menyadari bahwa skripsi ini masih memiliki kekurangan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan.

Yogyakarta, 25 Mei 2026

Penulis

DAFTAR ISI

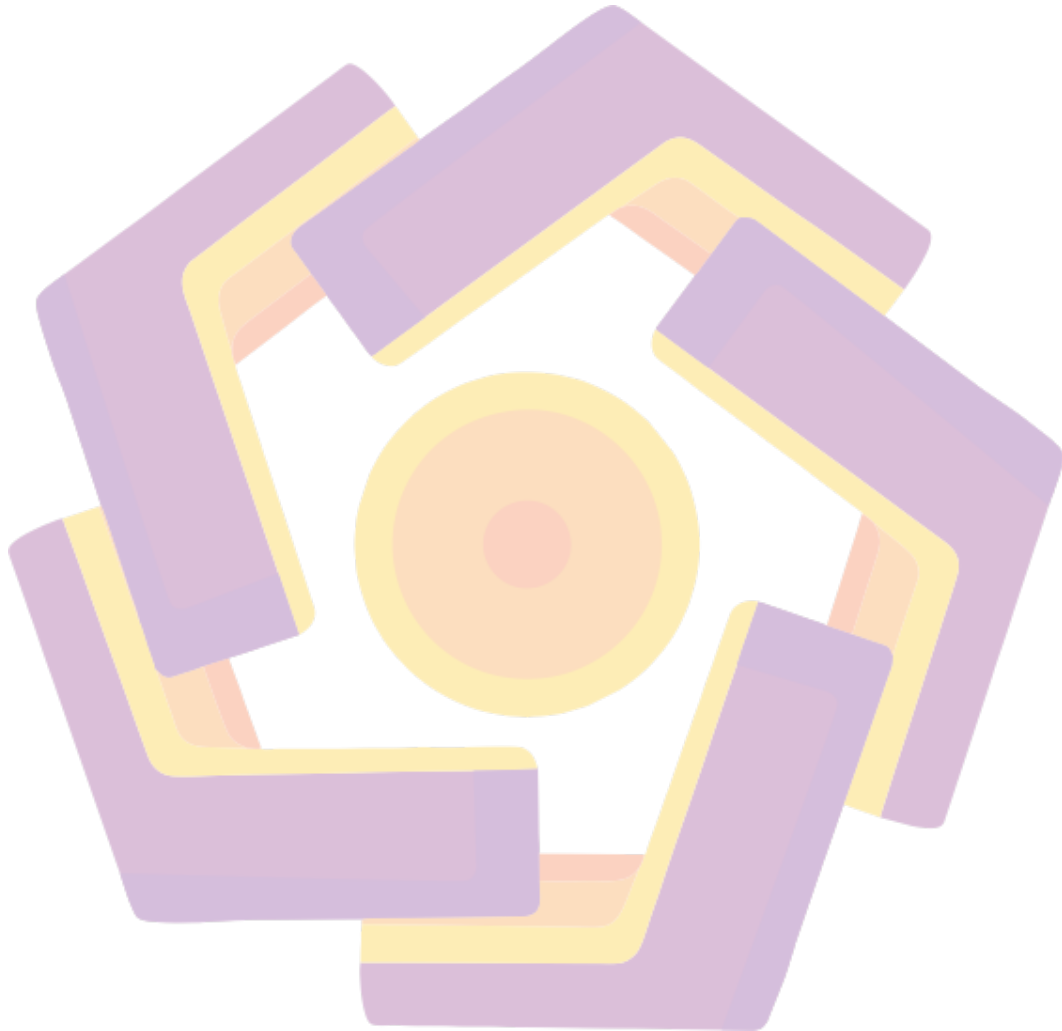
HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	i
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	Error! Bookmark not defined.
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xvi
INTISARI	xix
<i>ABSTRACT</i>	xx
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA	5

Studi Literatur	5
Dasar Teori	10
2.2.1 Katarak	10
2.2.2 Fotografi Fundus	11
2.2.3 Data Preprocessing	12
2.2.3.1 Digital Image Processing	12
2.2.3.2 Penyeimbangan Kelas Data (<i>Class Balancing</i>)	14
2.2.4 Deep Learning	14
2.2.5 Pergeseran Arsitektur Transformer ke Visi	14
2.2.6 Vision Transformer (ViT)	15
2.2.6.1 Patch Extraction	15
2.2.6.2 Proyeksi Linear (<i>Patch Embedding</i>)	16
2.2.6.3 <i>Positional Embedding</i>	16
2.2.6.4 Transformer Encoder	17
2.2.6.5 Classification Head	20
2.2.6.6 Transfer Learning dan Fine-tuning	21
2.2.7 Arsitektur <i>Hybrid ViT-XGBoost</i>	22
2.2.8 Validasi dan Evaluasi Model	22
2.2.8.1 Metode Holdout	22
2.2.8.2 Stratified K-Fold Cross Validation	23
2.2.8.3 Nested Cross-Validation	23
2.2.8.4 Metrik Evaluasi	23
2.2.8.5 Kurva ROC dan AUC	24
BAB III METODE PENELITIAN	26
3.1 Objek Penelitian	26

3.2	Alur Penelitian	26
3.3	Alat dan Bahan	29
3.3.1	Data Penelitian	29
3.3.2	Alat	29
BAB IV HASIL DAN PEMBAHASAN		31
4.1	Pengumpulan Data	31
4.2	Eksplorasi Data dan <i>Noise Analysis</i>	32
4.3	Data Preprocessing	34
4.4	<i>Holdout Split</i>	35
4.5	Pembagian <i>Stratified 5-Fold Cross Validation</i>	35
4.6	<i>Random Oversampling</i>	36
4.7	Implementasi dan Pelatihan Skenario Model	37
4.7.1	Pelatihan ViT-Base dengan Early Stopping	37
4.7.2	Ekstraksi Fitur Laten 768-Dimensi	39
4.7.3	Hyperparameter Tuning XGBoost dengan GridSearchCV	39
4.8	Uji Final Model Terbaik pada Unseen Test Data	41
4.8.1	Perbandingan Model	41
4.8.2	Visualisasi Confusion Matrix	42
4.8.3	Visualisasi Kurva ROC dan Skor AUC	44
BAB V PENUTUP		46
5.1	Kesimpulan	46
5.2	Saran	46
REFERENSI		48
LAMPIRAN		51

DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian	7
Tabel 4.1 Distribusi Kelas Hasil Pengecekan Eksplorasi Data	32
Tabel 4.2 Sebaran <i>Noise</i>	33
Tabel 4.3 Laporan Kinerja Hasil Klasifikasi ViT Murni	41
Tabel 4.4 Laporan Kinerja Hasil Klasifikasi Hybrid ViT-XGBoost	42



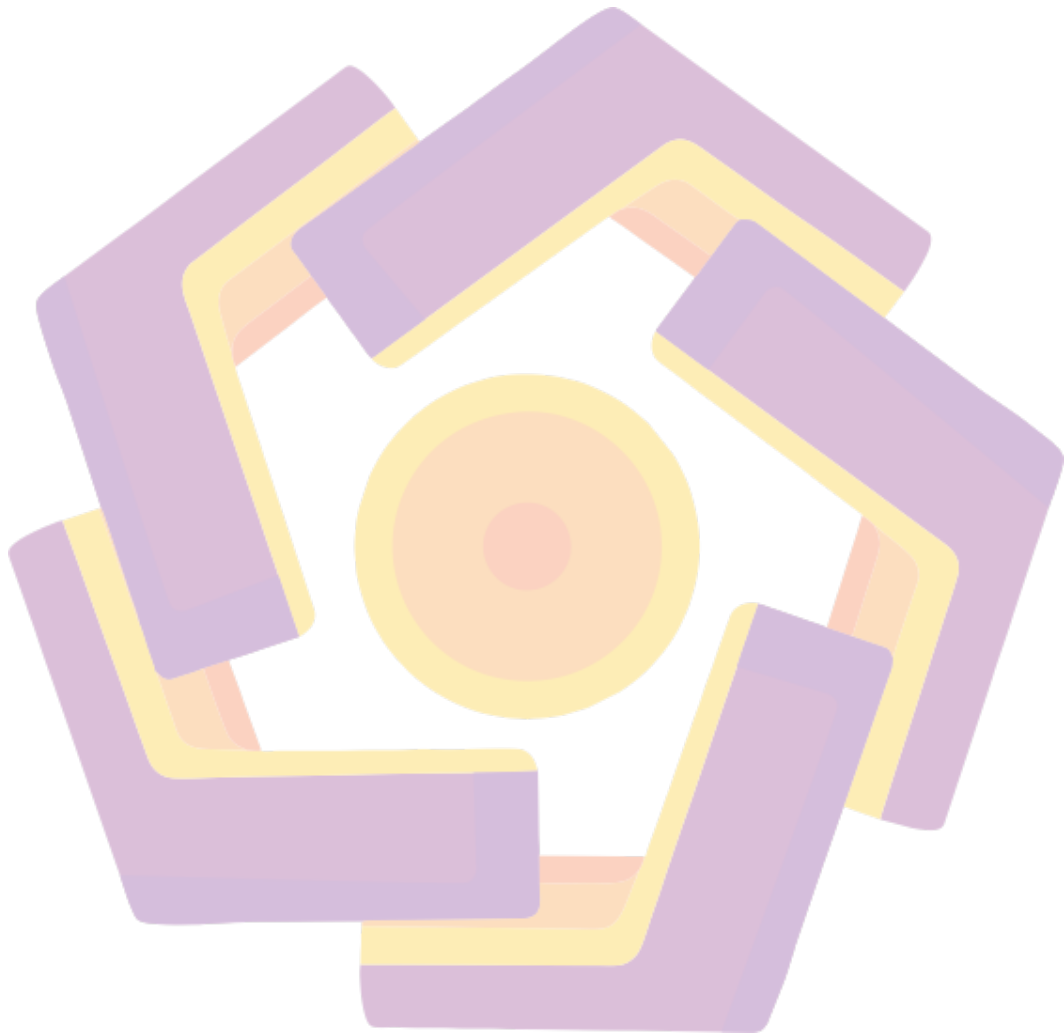
DAFTAR GAMBAR

Gambar 2.1 Arsitektur Umum Vision Transformer	15
Gambar 3.1 Alur Penelitian	26
Gambar 4.1 Konfigurasi Jalur Repositori Penyimpanan Data	31
Gambar 4.2 Mengaktifkan Fungsi rsync	31
Gambar 4. 3 Penyaringan Label Biner Target Klinis	32
Gambar 4.4 String Matching untuk Menghitung Noise Dataset	33
Gambar 4.5 Data <i>Preprocessing</i>	34
Gambar 4.6 Kode Program Pembagian Data Metode <i>Holdout</i>	35
Gambar 4.7 Inisialisasi Siklus Validasi Silang <i>Stratified</i>	36
Gambar 4.8 Random Over Sampling (ROS) pada <i>Internal Fold Train</i>	36
Gambar 4.9 Inisialisasi Model ViT	37
Gambar 4.10 Struktur Kelas <i>Early Stopping</i>	37
Gambar 4.11 Eksekusi <i>Early Stopping</i>	38
Gambar 4.12 Memuat Berkas Bobot Model Vit Terbaik	39
Gambar 4.13 Konfigurasi parameter pencarian kombinasi <i>decision tree</i>	40
Gambar 4.14 Inisialisasi GridSearchCV	40
Gambar 4.15 Pengecekan kinerja estimasi model hasil optimasi grid	41
Gambar 4.16 Visualisasi <i>Confusion Matrix</i> Kedua Model	42
Gambar 4.17 Hasil Ekstraksi Geometris Garis Tren Kurva ROC-AUC	44

DAFTAR LAMPIRAN

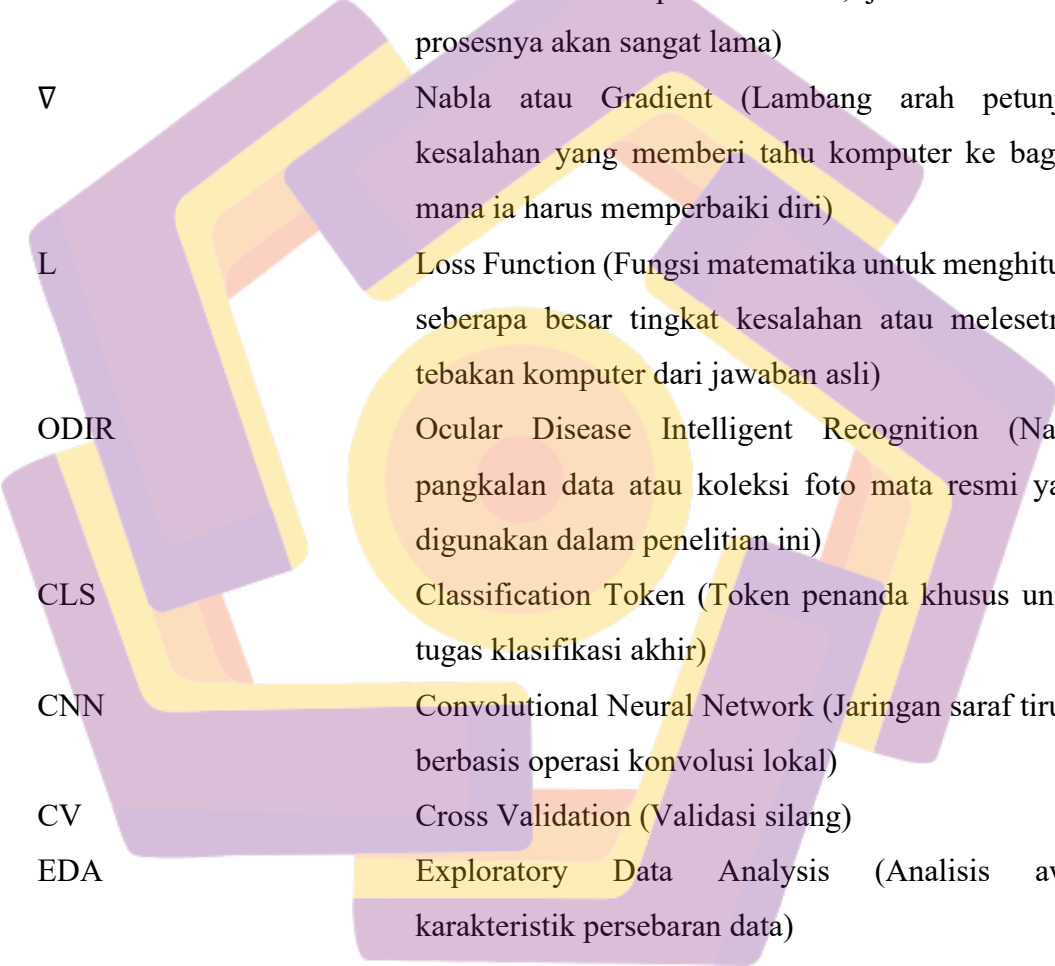
Lampiran 1. Dokumentasi Penelitian

51

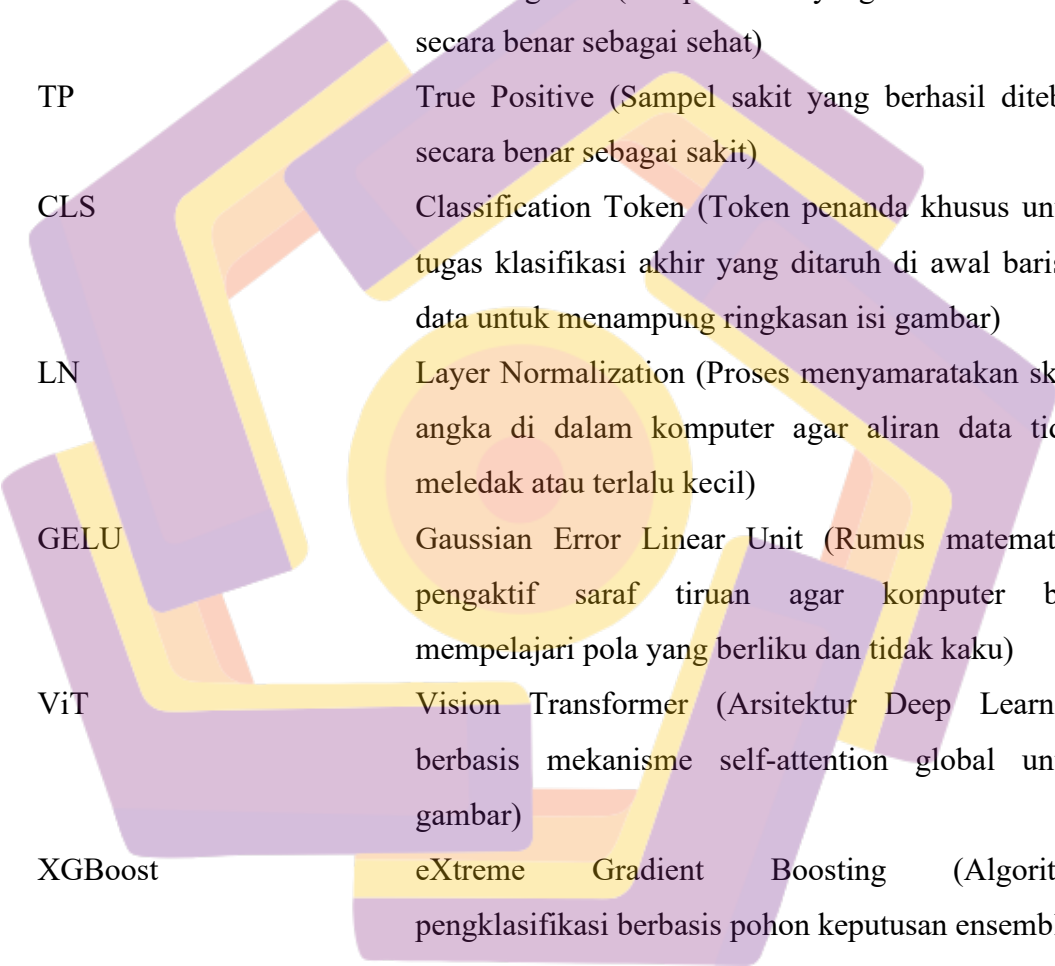


DAFTAR LAMBANG DAN SINGKATAN

AUC	Area Under the Curve (Luas area di bawah kurva evaluasi)
H	Height (Lambang untuk mengukur ukuran tinggi suatu gambar dalam satuan titik warna piksel)
W	Width (Lambang untuk mengukur ukuran lebar suatu gambar dalam satuan titik warna piksel)
C	Channel (Lambang jumlah warna dasar penyusun gambar, biasanya bernilai 3 untuk gambar berwarna Merah, Hijau, Biru)
P	Patch Size (Ukuran panjang dan lebar potongan kotak kecil gambar yang dipotong-potong oleh komputer)
N	Number of Patches (Jumlah total potongan kecil gambar yang berhasil dihasilkan dari satu foto utuh)
E	Embedding Matrix (Matriks pengali terpelajar yang digunakan untuk mengubah piksel gambar menjadi bahasa angka komputer)
z_0	Simbol untuk barisan data angka awal yang siap dimasukkan ke dalam mesin pemroses komputer
Q	Query (Vektor pencari data di komputer yang berfungsi layaknya kata kunci pencarian informasi)
K	Key (Vektor kunci penanda di komputer yang berfungsi layaknya judul artikel yang dicocokkan dengan kata kunci pencarian)
V	Value (Vektor nilai isi di komputer yang berisi informasi asli yang ingin diambil setelah kata kunci pencarian dinyatakan cocok)
dk	Ukuran panjang ruang dari vektor kunci, digunakan sebagai pembagi agar perhitungan matematika tidak macet atau error



σ	Sigma (Lambang fungsi matematika yang mengubah angka acak menjadi nilai persentase peluang antara 0% hingga 100%)
θ	Theta (Lambang parameter atau kumpulan ingatan bobot internal yang ada di dalam otak komputer)
η	Eta (Lambang kecepatan belajar komputer; jika terlalu besar komputer ceroboh, jika terlalu kecil prosesnya akan sangat lama)
∇	Nabla atau Gradient (Lambang arah petunjuk kesalahan yang memberi tahu komputer ke bagian mana ia harus memperbaiki diri)
L	Loss Function (Fungsi matematika untuk menghitung seberapa besar tingkat kesalahan atau melesetnya tebakan komputer dari jawaban asli)
ODIR	Ocular Disease Intelligent Recognition (Nama pangkalan data atau koleksi foto mata resmi yang digunakan dalam penelitian ini)
CLS	Classification Token (Token penanda khusus untuk tugas klasifikasi akhir)
CNN	Convolutional Neural Network (Jaringan saraf tiruan berbasis operasi konvolusi lokal)
CV	Cross Validation (Validasi silang)
EDA	Exploratory Data Analysis (Analisis awal karakteristik persebaran data)
FFN	Feed-Forward Network (Jaringan saraf tiruan satu arah maju)
FN	False Negative (Sampel sakit yang salah ditebak sebagai sehat oleh model)
FP	False Positive (Sampel sehat yang salah ditebak sebagai sakit oleh model)

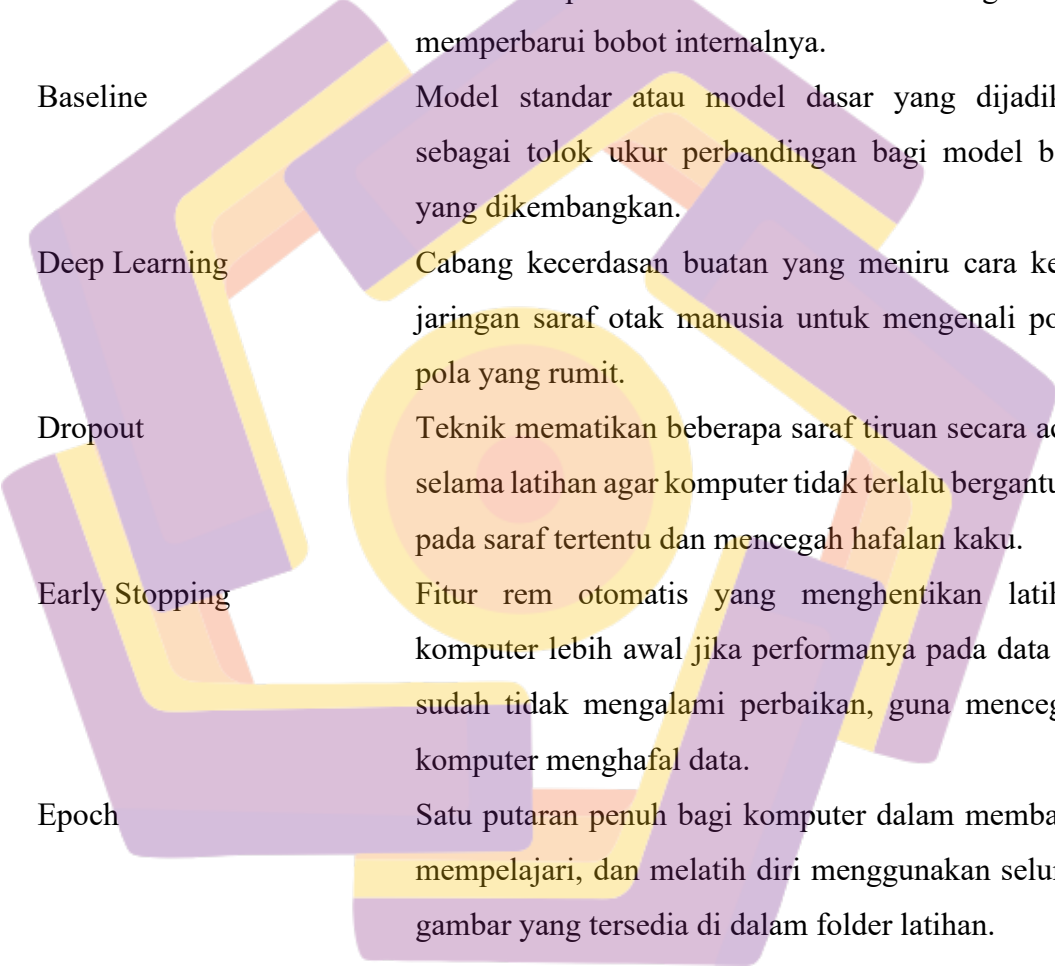


MLP	Multi-Layer Perceptron (Lapisan saraf tiruan multi-layer yang terhubung penuh)
ROC	Receiver Operating Characteristic (Kurva karakteristik performa grafis model)
ROS	Random Oversampling (Teknik duplikasi acak untuk menyeimbangkan kelas data)
TN	True Negative (Sampel sehat yang berhasil ditebak secara benar sebagai sehat)
TP	True Positive (Sampel sakit yang berhasil ditebak secara benar sebagai sakit)
CLS	Classification Token (Token penanda khusus untuk tugas klasifikasi akhir yang ditaruh di awal barisan data untuk menampung ringkasan isi gambar)
LN	Layer Normalization (Proses menyamaratakan skala angka di dalam komputer agar aliran data tidak meledak atau terlalu kecil)
GELU	Gaussian Error Linear Unit (Rumus matematika pengaktif saraf tiruan agar komputer bisa mempelajari pola yang berliku dan tidak kaku)
ViT	Vision Transformer (Arsitektur Deep Learning berbasis mekanisme self-attention global untuk gambar)
XGBoost	eXtreme Gradient Boosting (Algoritma pengklasifikasi berbasis pohon keputusan ensemble)

DAFTAR ISTILAH



<i>Embedding</i>	Representasi numerik dari teks dalam bentuk vektor yang menangkap makna semantik.
Akurasi	Nilai persentase yang menunjukkan seberapa sering model komputer berhasil menebak seluruh data dengan benar.
AUC	Area Under the Curve (Luas area di bawah kurva evaluasi)
CLS	Classification Token (Token penanda khusus untuk tugas klasifikasi akhir)
CNN	Convolutional Neural Network (Jaringan saraf tiruan berbasis operasi konvolusi lokal)
CV	Cross Validation (Validasi silang)
EDA	Exploratory Data Analysis (Analisis awal karakteristik persebaran data)
FFN	Feed-Forward Network (Jaringan saraf tiruan satu arah maju)
FN	False Negative (Sampel sakit yang salah ditebak sebagai sehat oleh model)
FP	False Positive (Sampel sehat yang salah ditebak sebagai sakit oleh model)
MLP	Multi-Layer Perceptron (Lapisan saraf tiruan multi-layer yang terhubung penuh)
ROC	Receiver Operating Characteristic (Kurva karakteristik performa grafis model)
ROS	Random Oversampling (Teknik duplikasi acak untuk menyeimbangkan kelas data)
TN	True Negative (Sampel sehat yang berhasil ditebak secara benar sebagai sehat)



TP	True Positive (Sampel sakit yang berhasil ditebak secara benar sebagai sakit)
ViT	Vision Transformer (Arsitektur Deep Learning berbasis mekanisme self-attention global untuk gambar)
Backpropagation	Proses hitung mundur yang dilakukan oleh komputer untuk memperbaiki kesalahan tebakan dengan cara memperbarui bobot internalnya.
Baseline	Model standar atau model dasar yang dijadikan sebagai tolok ukur perbandingan bagi model baru yang dikembangkan.
Deep Learning	Cabang kecerdasan buatan yang meniru cara kerja jaringan saraf otak manusia untuk mengenali pola-pola yang rumit.
Dropout	Teknik mematikan beberapa saraf tiruan secara acak selama latihan agar komputer tidak terlalu bergantung pada saraf tertentu dan mencegah hafalan kaku.
Early Stopping	Fitur rem otomatis yang menghentikan latihan komputer lebih awal jika performanya pada data uji sudah tidak mengalami perbaikan, guna mencegah komputer menghafal data.
Epoch	Satu putaran penuh bagi komputer dalam membaca, mempelajari, dan melatih diri menggunakan seluruh gambar yang tersedia di dalam folder latihan.
Fine-tuning	Melatih ulang model pintar yang sudah memiliki modal pengetahuan umum agar menjadi sangat ahli dalam mengerjakan satu tugas baru yang spesifik.
Ground Truth	Kondisi medis asli atau kebenaran mutlak di lapangan (misal, label asli dari dokter yang menyatakan mata tersebut benar-benar katarak atau normal).



Hyperparameter Tuning	Proses eksperimen mencari kombinasi setelan tombol parameter paling pas agar performa algoritma komputer mencapai titik maksimal.
Imbalanced Dataset	Kondisi tidak ideal di mana jumlah contoh gambar untuk satu kategori jauh lebih banyak (jomplang) dibanding kategori lainnya.
<i>Feature Extractor</i>	Komponen model yang digunakan untuk mengekstraksi representasi fitur dari data mentah.
<i>Hybrid Model</i>	Model yang menggabungkan dua atau lebih pendekatan berbeda dalam satu kerangka pemodelan.
<i>Gradient Boosting</i>	Metode <i>ensemble learning</i> yang membangun model secara bertahap untuk memperbaiki kesalahan sebelumnya.
<i>Overfitting</i>	Kondisi ketika model terlalu menyesuaikan diri terhadap data latih sehingga performa menurun pada data baru. Regresi Metode prediksi untuk memperkirakan nilai numerik kontinu.
<i>Data Leakage</i>	Kondisi ketika informasi dari data uji ikut digunakan dalam proses pelatihan model.
Interpretabilitas Model	Kemampuan untuk menjelaskan bagaimana model menghasilkan suatu prediksi.

INTISARI

Penyakit katarak merupakan penyebab utama kebutaan yang dapat dicegah jika didiagnosis secara dini. Namun, pemeriksaan manual citra fundus seringkali memakan waktu dan bergantung pada keahlian spesialis. Penelitian ini bertujuan mengembangkan dan menganalisis secara komparatif model klasifikasi citra katarak menggunakan arsitektur Vision Transformer (ViT) dengan pendekatan hibrida *fine-tuning* dan *eXtreme Gradient Boosting* (XGBoost). Eksperimen menggunakan data sekunder dari repositori *Ocular Disease Intelligent Recognition* (ODIR) yang dibagi dengan metode *Holdout* (80% data latih, 20% data uji murni). Untuk mengatasi data tidak seimbang dan mencegah *overfitting*, skema *Stratified 5-Fold Cross Validation* dikombinasikan dengan *Random Oversampling* (ROS) secara internal pada setiap lipatan. Skenario penelitian membandingkan model ViT murni *end-to-end* (baseline) dengan model Hybrid ViT-XGBoost, di mana ViT bertindak sebagai pengekstraksi fitur laten 768-dimensi dari token CLS dan XGBoost sebagai pengklasifikasi akhir berbasis *decision tree ensemble*. Hasil pengujian pada 684 sampel data uji murni menunjukkan model Hybrid ViT-XGBoost memberikan performa terbaik dengan Akurasi global 97%, Presisi kelas katarak 0,85, Recall 0,84, dan F1-Score 0,85. Sebaliknya, ViT murni menghasilkan Akurasi lebih rendah (96%) dan Presisi katarak jatuh di angka 0,75 akibat 18 kasus *False Positive*. Hambatan atensi global ViT terhadap *noise* visual (artifak pencahayaan dan debu lensa sebesar 8,93% dari data) berhasil diredam oleh batas keputusan non-linear XGBoost. Penelitian ini membuktikan model hibrida ini menghasilkan diagnosis yang lebih selektif dan andal untuk diimplementasikan pada sistem pendukung keputusan klinis masa depan bagi praktisi medis.

Kata kunci: Katarak, Vision Transformer, XGBoost, Model Hibrida, Citra Fundus.

ABSTRACT

Cataract disease is a primary cause of blindness that can be prevented if diagnosed early. However, manual examination of fundus images is often time-consuming and heavily reliant on specialist expertise. This study aims to develop and comparatively analyze a cataract image classification model utilizing the Vision Transformer (ViT) architecture combined with a hybrid fine-tuning and eXtreme Gradient Boosting (XGBoost) approach. Experiments were conducted using secondary data from the Ocular Disease Intelligent Recognition (ODIR) repository, partitioned via the Holdout method into 80% training data and 20% unseen test data. To address class imbalance and prevent overfitting, a Stratified 5-Fold Cross-Validation scheme was combined with internal Random Oversampling (ROS) within each fold. The research scenario compared an end-to-end pure ViT model (baseline) against the Hybrid ViT-XGBoost model, where ViT acts as a feature extractor yielding 768-dimensional latent features from the CLS token, and XGBoost serves as the final decision tree ensemble classifier. Evaluation results on 684 unseen test samples demonstrated that the Hybrid ViT-XGBoost model achieved the best performance, reaching a global accuracy of 97%, with a cataract class precision of 0.85, recall of 0.84, and an F1-Score of 0.85. Conversely, the pure ViT model produced a lower global accuracy of 96% and its cataract precision dropped to 0.75 due to 18 False Positive cases. The global attention limitation of pure ViT against visual noise (lighting artifacts and lens dust comprising 8.93% of the data) was successfully mitigated by the non-linear decision boundaries of XGBoost. This study proves that the hybrid model delivers a more selective and reliable diagnosis, making it suitable for future clinical decision support systems for medical practitioners.

Keywords: Cataract, Vision Transformer, XGBoost, Hybrid Model,