

Penerapan Transfer Learning VGG-16 untuk Mendeteksi Penyakit Mata Manusia Berbasis Citra Fundus

Willy¹, Ary Prabowo²

^{1,2}Teknik Informatika, Universitas Esa Unggul, Indonesia

¹willyueu0211@gmail.com, ²ary.prabowo@esaunggul.ac.id

Info Artikel

Riwayat Artikel:

Received 2025-07-14

Revised 2026-01-21

Accepted 2026-01-23

Abstract – Eye disorders represent a serious global health issue that can lead to a decline in quality of life and even permanent blindness. Early diagnostic for eye diseases such as glaucoma, diabetic retinopathy, age-related macular degeneration, cataract, myopia, and hypertension is crucial to prevent more severe complications. The objective of this study is to develop an image classification model for fundus images using a transfer learning approach with the VGG-16 architecture. The dataset used is ODIR-5K, which includes eight classes of eye diseases. The research stages involve image preprocessing, data augmentation, class balancing using SMOTE, and CNN for training the model. The model training process was conducted over 80 epochs with a combination of freezing layers, fine-tuning, and hyperparameter tuning. Model evaluation was carried out using metrics such as accuracy, precision, recall, F1-score, confusion matrix, and ROC AUC curve. The results show that the developed model achieved an accuracy of 89% compared to the previous study which only reached 45%, with a macro average F1-score of 0.89. The model demonstrated excellent performance in classes such as Hypertension, Glaucoma, and Myopia, although challenges remain in distinguishing the Diabetes and Normal classes. Therefore, the VGG-16-based approach has proven effective for multi-class classification of fundus images, and the results of this study may serve as a foundation for developing deep learning-based diagnostic support systems in the field of ophthalmology.

Keywords: CNN, Fundus Image, Image Classification, Transfer Learning, VGG-16

Corresponding Author:

Willy

Email: willyueu0211@gmail.com



This is an open access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.

Abstrak – Gangguan pada mata merupakan salah satu isu serius dalam bidang kesehatan secara global yang dapat mengakibatkan penurunan kualitas hidup hingga kebutaan permanen. Deteksi dini terhadap penyakit mata seperti glaukoma, retinopati diabetik, degenerasi makula terkait usia, katarak, miopia, dan hipertensi sangat penting untuk mencegah terjadinya komplikasi yang lebih serius. Penelitian ini bertujuan untuk merancang model klasifikasi untuk citra fundus mata menggunakan pendekatan transfer learning dengan arsitektur VGG-16. Dataset yang digunakan adalah ODIR-5K, yang mencakup delapan kelas penyakit mata. Tahapan penelitian meliputi preprocessing gambar, augmentasi data, penyeimbangan kelas menggunakan SMOTE, dan pelatihan model CNN. Proses pelatihan dilakukan sebanyak 80 epoch dengan kombinasi freezing layer, fine-tuning, dan hyperparameter tuning. Evaluasi model dilakukan dengan penggunaan dari metrik akurasi, precision, recall, F1-score, confusion matrix, dan kurva ROC AUC. Hasil menunjukkan bahwa model yang dikembangkan memiliki akurasi sebesar 89% dibandingkan dengan penelitian sebelumnya yang hanya 45%, dengan nilai macro average F1-score sebesar 0.89. Model menunjukkan performa yang sangat baik pada kelas seperti Hypertension, Glaucoma, dan Myopia, namun masih terdapat tantangan pada kelas Diabetes dan Normal. Dengan demikian, pendekatan VGG-16 terbukti efektif untuk klasifikasi multi-kelas citra fundus, dan hasil dari penelitian yang dilakukan dapat menjadi landasan awal untuk pengembangan sistem bantu diagnosis menggunakan deep learning untuk bidang oftalmologi.

Kata Kunci: Citra Fundus, CNN, Klasifikasi Gambar, Transfer Learning, VGG-16

I. PENDAHULUAN

Mata merupakan salah satu dari enam indra yang memiliki peran krusial pada kehidupan manusia, karena memungkinkan individu untuk beraktivitas secara optimal, baik dalam lingkungan sosial, pekerjaan, maupun aktivitas sehari-hari lainnya. Penyakit mata merupakan sebuah masalah kesehatan tingkat global yang sangat berpengaruh kepada kualitas hidup manusia[1]. Berdasarkan laporan kesehatan, setidaknya ada 2,2 miliar orang pada muka bumi mengalami gangguan penglihatan/penyakit mata, dengan setidaknya ada sekitar 1 miliar kasus penyakit mata yang dapat dicegah atau ditangani secara dini. Penyakit mata *glaucoma*, *diabetic retinopathy*, *age related macular degeneration*, *cataract*, *myopia*, dan *hypertension*. sering menyerang berbagai kalangan usia dan berpotensi menyebabkan kebutaan jika tidak segera didiagnosis[1],[2].

Seiring dengan perkembangan teknologi, metode kecerdasan buatan khususnya *deep learning* telah menunjukkan kontribusi yang signifikan dalam bidang citra medis. Khususnya *Convolutional Neural Network* (CNN) yang terbukti efektif dalam menganalisis citra fundus mata[3]. CNN memiliki kemampuan untuk mengekstraksi fitur-fitur yang kompleks secara otomatis tanpa memerlukan proses ekstraksi manual[4], sehingga metode ini sangatlah

cocok untuk dan efektif untuk klasifikasi visual dari citra medis[5]. Untuk meningkatkan efisiensi serta akurasi dari model, pendekatan *transfer learning* sering digunakan, seperti penggunaan arsitektur VGG-16 yang dimanfaatkan untuk pelatihan pada dataset besar agar terjadi pengulangan untuk tugas baru yang serupa[6].

Berdasarkan kajian dari penelitian terdahulu, menunjukkan bahwa VGG-16 memiliki performa yang baik dalam klasifikasi citra medis, seperti dari penelitian yang dilakukan oleh Shagun Sharma dan Kalpna Guleria (2022) yang menggunakan VGG-16 untuk mendeteksi penyakit *pneumonia* dengan kinerja model *accuracy*, *recall*, *precision*, dan *f1-score* sebesar 95.4%, 0.954, 0.954, dan 0.954[7]. Serta penelitian yang dilakukan oleh T. Gayathri dan K. Sundeep Kumar (2024) yang menggunakan VGG-16 untuk segmentasi dan klasifikasi tumor otak berdasarkan gambar MRI mendapatkan *accuracy* sebesar 99.6%, 95.35%, dan 94% untuk masing-masing bagian tumor TC, ET, dan WT[8]. Di sisi lain, penerapan arsitektur VGG-16 masih menghadapi sejumlah tantangan, khususnya ketika dihadapkan pada permasalahan ketidakseimbangan data, variasi kualitas dataset, serta penerapan teknik augmentasi yang belum maksimal[9]. Kondisi tersebut berpotensi mempengaruhi kinerja dan kemampuan generalisasi model. Oleh karena itu, penelitian ini mengusulkan pengembangan model klasifikasi penyakit mata dengan *transfer learning* VGG-16 dengan strategi perbaikan berupa penerapan augmentasi data yang lebih beragam, penyeimbangan kelas menggunakan teknik *Synthetic Minority Over-sampling Technique* (SMOTE), serta penyetelan *hyperparameter* untuk peningkatan performa model.

Penelitian terdahulu yang dilakukan oleh Muhammad Sirojul Qulub dan Soffiana Agustin (2024) untuk mendeteksi penyakit mata berbasis citra fundus menggunakan VGG-16 mendapatkan *accuracy* 45% dan terdapat bias pada kelas *Normal* berdasarkan *precision*, *recall*, dan *f1-score* yang lebih tinggi dibandingkan kelas lain, kondisi ini disebabkan ketidakseimbangan data antar kelas pada dataset yang digunakan[10]. Sebagai perbandingan, penelitian untuk mendeteksi penyakit mata berbasis citra fundus menggunakan arsitektur VGG-19 oleh Dwi Marcella dkk (2022) mendapatkan hasil *accuracy*, *precision*, *recall* dan *f1-score* sebesar 65.29%, 66.53%, 65.29%, dan 65.40%, yang menunjukkan bahwa kinerja model masih memiliki ruang untuk peningkatan[11]. Sementara itu, Erdiansyah dkk (2022) menerapkan metode *multilayer percept* untuk mengklasifikasikan penyakit mata *diabetic retinopathy* dan memperoleh nilai *accuracy* sebesar 71.80%, *precision* 72.50%, dan *recall* 71.80%[12]. Di samping itu, penelitian yang dilakukan oleh Ayu Evilia Suwanda dan Dwi Juniati (2022) menggunakan dimensi fraktal *box counting* dan *fuzzy k-means* untuk penyakit mata berbasis citra fundus mendapatkan *accuracy* sebesar 76%[13]. Selanjutnya, penelitian yang dilakukan oleh Syafiq Hilmi Abdullah dkk (2022) menggunakan *EfficientNet-B0* untuk klasifikasi penyakit *diabetic retinopathy* berbasis citra fundus dan *deep learning* menunjukkan *accuracy* sebesar 82.096%, *precision* 67.6%, *recall* 63.4%, dan *f1-score* 64.6%[14]. Berdasarkan kajian terhadap penelitian terdahulu, dapat disimpulkan bahwa meskipun berbagai pendekatan telah diterapkan untuk klasifikasi citra fundus mata, performa dari model yang dihasilkan masih memiliki ruangan untuk peningkatan.

Oleh karena itu, tujuan dari penelitian ini adalah mengembangkan dan mengevaluasi model klasifikasi penyakit mata berbasis CNN dengan arsitektur *transfer learning* VGG-16 yang memiliki kinerja lebih stabil dan seimbang antar penyakit serta menguji apakah CNN dengan arsitektur *transfer learning* VGG-16 dapat mengklasifikasikan penyakit mata pada manusia dengan hasil yang lebih baik. Kebaruan pada penelitian ini adalah strategi peningkatan kinerja secara komprehensif, yaitu penggunaan augmentasi data yang lebih beragam, penanganan ketidakseimbangan kelas menggunakan SMOTE, serta penyetelan *hyperparameter* untuk mengoptimalkan proses ketika pelatihan model. Adapun kontribusi yang dapat diharapkan agar penelitian ini dapat menghasilkan model klasifikasi penyakit mata dengan performa yang lebih stabil, sekaligus memberikan landasan bagi pengembangan sistem pendukung diagnosis berbasis *deep learning* pada bidang oftamologi.

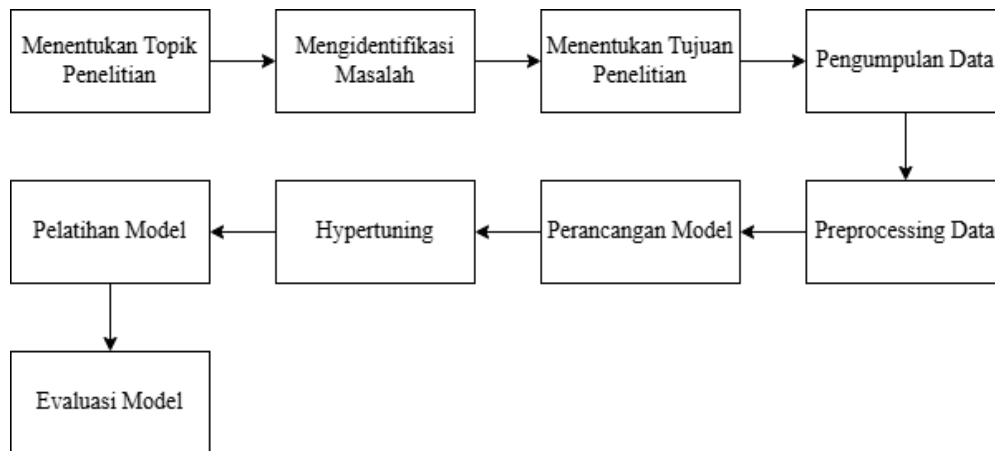
II. METODE

A. Metode Penelitian

Penelitian ini menggunakan pendekatan eksperimen berbasis *transfer learning* dengan memanfaatkan arsitektur dari CNN bernama VGG-16. Arsitektur VGG-16 telah banyak digunakan dalam klasifikasi citra medis karena memiliki struktur jaringan yang dalam dan teratur, sehingga mampu mengekstraksi fitur visual secara efektif dari citra medis yang umumnya memiliki pola dan tekstur yang kompleks. Kemudian VGG-16 ini berguna untuk ekstraksi fitur yang efektif dengan konvolusi bertumpuk kecil (3x3) dalam VGG-16, yang memungkinkan model untuk mendapatkan fitur citra yang lebih halus secara bertahap[15].

Penelitian ini dilakukan dengan pendekatan eksperimental yang berfokus pada pengembangan model klasifikasi citra fundus mata menggunakan *transfer learning* VGG-16. Tujuan penelitian ini adalah mengembangkan model *Convolutional Neural Network* (CNN) berbasis arsitektur VGG-16 untuk mengklasifikasikan penyakit mata menggunakan dataset *Ocular Disease Recognition* (ODR), yang mencakup delapan kelas, terdiri atas satu kelas normal, enam kelas penyakit, dan satu kelas penyakit lainnya. Data kemudian melalui tahap *preprocessing* dan augmentasi untuk meningkatkan kualitas dan keberagaman dataset sebelum digunakan dalam proses pelatihan model. Model VGG-16 dilatih menggunakan data latih dan dievaluasi menggunakan data uji. Kinerja model dievaluasi

berdasarkan *confusion matrix* yang mencakup metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Tahap akhir penelitian meliputi analisis hasil evaluasi, penarikan kesimpulan, serta perumusan rekomendasi untuk pengembangan penelitian selanjutnya.



Gambar 1. Alur Penelitian dalam Mengembangkan Model

1. Dataset

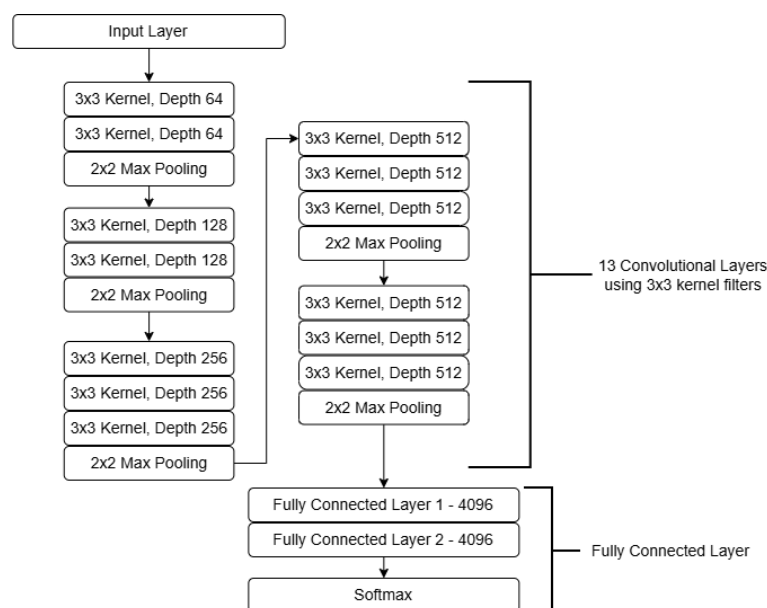
Data citra fundus mata pada penelitian bersumber dari dataset *Ocular Disease Recognition* yang tersedia pada platform Kaggle. Dataset ini mencakup informasi mengenai fundus mata yang memiliki penyakit dan normal dari 5000 pasien dengan citra fundus kiri dan kanan. Dataset ini disusun melalui kerja sama antara Shanggong Medical Technology Co., Ltd. dan Peking University, sehingga memiliki reliabilitas yang baik karena melibatkan institusi medis.

2. Preprocessing

Pada penelitian ini, langkah preprocessing data adalah tahapan penting untuk memastikan kualitas dataset citra fundus sebelum digunakan dalam pelatihan model klasifikasi. Proses preprocessing dimulai dengan menggabungkan gambar fundus kiri dan fundus kanan, serta menyederhanakan label multi-penyakit menjadi satu label utama berdasarkan urutan, kemudian mengubah ukuran setiap gambar ke resolusi 224x224 piksel agar sesuai dengan kebutuhan input layer dari model VGG-16. Kemudian metode *Synthetic Minority Over-sampling Technique* (SMOTE) diterapkan untuk menangani permasalahan ketidakseimbangan data pada kelas minoritas. SMOTE dilakukan pada ruang fitur hasil ekstraksi model *Convolutional Neural Network*, sehingga sampel sintetis yang dihasilkan merepresentasikan karakteristik fitur yang relevan tanpa mengubah struktur spasial citra fundus. Strategi ini bertujuan untuk mengurangi bias model terhadap kelas mayoritas serta meningkatkan kemampuan generalisasi model pada seluruh kelas penyakit mata. Selanjutnya, gambar diterapkan normalisasi dengan merubah intensitas piksel kedalam rentang 0 dan 1 untuk memastikan stabilitas dan efisiensi pelatihan model. Selain itu, teknik augmentasi data diterapkan untuk meningkatkan variasi dataset serta memperbaiki kemampuan generalisasi model. Teknik augmentasi yang terapkan mencakup rotasi, flipping horizontal, dan zooming. Langkah-langkah *preprocessing* ini bertujuan untuk meningkatkan kualitas data input serta mengurangi risiko overfitting selama proses pelatihan.

3. Perancangan Model

Pada tahap ini, model transfer learning VGG-16 dibuat untuk mengklasifikasikan penyakit mata berdasarkan citra fundus ke dalam beberapa kategori penyakit mata. Pemilihan model VGG-16 ini dilakukan karena kemampuan model tersebut dalam mengekstraksi fitur spasial dari citra medis, hal ini sesuai dengan sifat citra fundus yang memiliki struktur visual kompleks. Efektivitas dari model ini teruji efektif dalam sejumlah studi terdahulu untuk tugas klasifikasi citra medis, sehingga meningkatkan kepercayaan terhadap performanya dalam konteks penelitian ini. Pada penelitian ini, strategi *transfer learning* diterapkan dengan memanfaatkan bobot awal VGG-16 yang telah dilatih pada dataset ImageNet. Lapisan konvolusi awal VGG-16 dibekukan (*freezing*) untuk mempertahankan kemampuan ekstraksi fitur dasar seperti tepi, tekstur, dan pola visual umum. Selanjutnya, beberapa lapisan konvolusi pada bagian akhir model dibuka (*fine-tuning*) dan dilatih kembali menggunakan dataset citra fundus mata. Pendekatan ini bertujuan untuk menyesuaikan fitur tingkat tinggi agar lebih relevan dengan karakteristik spesifik penyakit mata, sekaligus mengurangi risiko overfitting dan kebutuhan komputasi yang berlebihan.



Gambar 2. Arsitektur VGG-16

Gambar 1 merupakan arsitektur model VGG-16, model ini dirancang dengan langkah pertama *input layer* dimana data berupa citra fundus mata dengan resolusi 224x224 piksel dimasukkan ke dalam model setelah melalui tahap preprocessing, termasuk normalisasi dan augmentasi[16]. Kemudian *convolutional layer* dimana model VGG-16 memiliki struktur 13 lapisan konvolusi yang dirancang untuk mengekstraksi fitur spasial dari citra fundus. Setiap lapisan konvolusi diikuti oleh operasi ReLU activation untuk menambahkan non-linearitas dan *max pooling* agar dapat mengurangi dimensi spasial tanpa adanya kehilangan fitur penting. Masuk pada *fully connected layer* lapisan ini bertugas mengintegrasikan fitur-fitur yang diperoleh melalui proses yang terjadi pada lapisan konvolusi. Pada tahap ini, neuron pada *fully connected layers* dilatih untuk mengenali karakteristik khas yang sesuai dengan kategori penyakit mata. Terakhir adalah *output layer* dimana lapisan ini menghasilkan prediksi multi-label berupa probabilitas untuk setiap kategori penyakit mata, menggunakan fungsi aktivasi *Softmax*.

4. Hyperparameter Tuning dan Training Model

Tahapan ini dilakukan untuk meningkatkan performa model yang dimana akan dilakukan penyesuaian *hyperparameter* melalui proses tuning. *Learning rate* digunakan untuk mengontrol besarnya pembaruan bobot model pada setiap iterasi pelatihan. Nilai *learning rate* diuji dengan rentang 0.001 hingga 0.0001 untuk menemukan nilai optimal yang mempercepat konvergensi tanpa melampaui global minimum. *Batch size* menentukan jumlah citra yang diproses sekaligus dalam satu iterasi pelatihan. Nilai yang diuji meliputi 32, 64, dan 128, dengan mempertimbangkan efisiensi komputasi dan stabilitas pelatihan. *Epochs* berfungsi untuk menentukan jumlah iterasi penuh melalui seluruh dataset selama proses pelatihan model. Model diuji dengan iterasi 50 hingga 100 untuk memastikan model belajar secara optimal tanpa *overfitting*. Untuk menghindari terjadinya *overfitting*, *dropout rate* diterapkan dengan cara menghilangkan neuron secara acak saat model dilatih. Nilai dropout rate diuji antara 0.3 hingga 0.6 untuk lapisan *fully connected*. Sebelum proses pelatihan model dilakukan, dataset dibagi menjadi tiga bagian, yaitu data pelatihan (*training set*), data validasi (*validation set*), dan data pengujian (*testing set*). Pembagian data dilakukan dengan proporsi 80% untuk data pelatihan, 10% untuk data validasi, dan 10% untuk data pengujian. Skema ini bertujuan untuk memastikan proses pelatihan berlangsung optimal, menghindari *data leakage*, serta memungkinkan evaluasi kinerja model secara objektif pada data yang tidak pernah dilihat sebelumnya. Kemudian proses pelatihan dilakukan menggunakan jumlah epoch yang telah ditentukan berdasarkan hasil tuning.

5. Evaluasi Model

Model VGG-16 yang dibuat ini akan dievaluasi menggunakan *confusion matrix* dengan *accuracy* untuk menunjukkan ringkasan performa dari model secara menyeluruh, *precision* digunakan untuk mengukur tingkat ketepatan model dalam menghasilkan prediksi positif, *recall* merepresentasikan kemampuan model dalam mendeteksi seluruh kasus positif, dan *F1-score* guna memberikan matrix yang seimbang untuk mengukur performa model. Pada *confusion matrix* ini terdapat TP (*True Positive*) menunjukkan jumlah kasus ketika model berhasil mengklasifikasikan data positif dengan benar, TN (*True Negative*) ini merujuk pada jumlah prediksi yang tepat ketika model membaca model negatif, FP (*False Positive*) adalah banyaknya kesalahan ketika model

mengidentifikasi data negatif menjadi positif, dan FN (*False Negative*) mengacu pada banyaknya kesalahan ketika model gagal membaca data positif dan mengklasifikasikannya sebagai negatif.

Accuracy mengukur proporsi prediksi model yang benar dengan membandingkan dari total prediksi yang dibuat. Matriks ini menggambarkan ringkasan performa model secara menyeluruh. Rumus untuk menghitung *accuracy*:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (1)$$

Precision mengukur seberapa tepat model saat membuat prediksi positif. *Precision* ini penting dikarenakan kasus prediksi positif yang keliru dapat menimbulkan masalah. Rumus untuk menghitung *precision*:

$$Precision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Positives\ (FP)} \quad (2)$$

Recall digunakan untuk mengukur kemampuan model ketika mengenali semua kasus positif. *Recall* ini penting ketika model mendeteksi kasus positif (*false negative*) yang harus diminimalkan. Rumus untuk menghitung *recall*:

$$Recall = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Negatives\ (FN)} \quad (3)$$

F1-score merupakan bentuk dari nilai rata-rata harmonis dari *Precision* dan *Recall*, guna memberikan penilaian yang seimbang. *F1-Score* berguna ketika distribusi perkelas dataset tidak merata atau ketika *trade-off* antara *Precision* dan *Recall* perlu dipertimbangkan. Rumus untuk menghitung *F1-score*:

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

III. HASIL DAN PEMBAHASAN

A. Hasil Pembuatan Dataset

Setelah proses pengunduhan dataset, dilakukan penyusunan ulang struktur data untuk memastikan konsistensi dan kemudahan pemrosesan pada tahap selanjutnya. Penyederhanaan label multi-penyakit menjadi satu label utama per citra dilakukan melalui proses *mapping*, sehingga setiap citra fundus direpresentasikan oleh satu kelas yang dominan. Selain itu, citra fundus mata kiri dan kanan digabungkan ke dalam satu *dataframe* terintegrasi, yang memungkinkan proses pelabelan dan pengolahan data dilakukan secara lebih sistematis dan efisien. Pendekatan ini bertujuan untuk mengurangi kompleksitas label serta meminimalkan potensi redundansi data, sehingga dataset yang dihasilkan lebih sesuai untuk kebutuhan pelatihan model klasifikasi.

B. Hasil Preprocessing Dataset

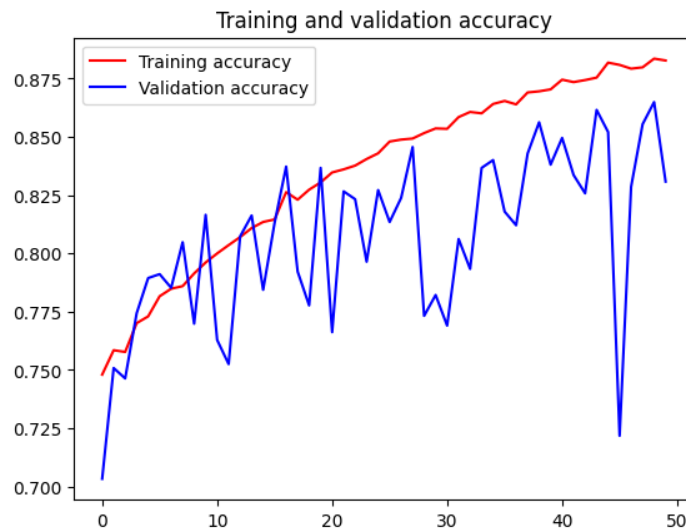
Distribusi data pada dataset awal menunjukkan ketidakseimbangan jumlah sampel antar kelas, yang berpotensi menyebabkan bias model terhadap kelas mayoritas. Untuk mengatasi permasalahan tersebut, diterapkan metode *Synthetic Minority Over-sampling Technique* (SMOTE) pada ruang fitur guna menyeimbangkan distribusi data pada setiap kelas. Setelah distribusi data menjadi lebih seimbang, dataset dialokasikan ke dalam data pelatihan, validasi, dan pengujian dengan proporsi masing-masing sebesar 80%, 10%, dan 10% secara proporsional pada setiap kelas. Selain itu, teknik *image augmentation* menggunakan *ImageDataGenerator* diterapkan untuk meningkatkan keragaman data latih, sehingga model memperoleh variasi pola visual yang lebih luas dan mampu meningkatkan kemampuan generalisasi selama proses pelatihan.

C. Hasil Pelatihan Model dan Transfer Learning

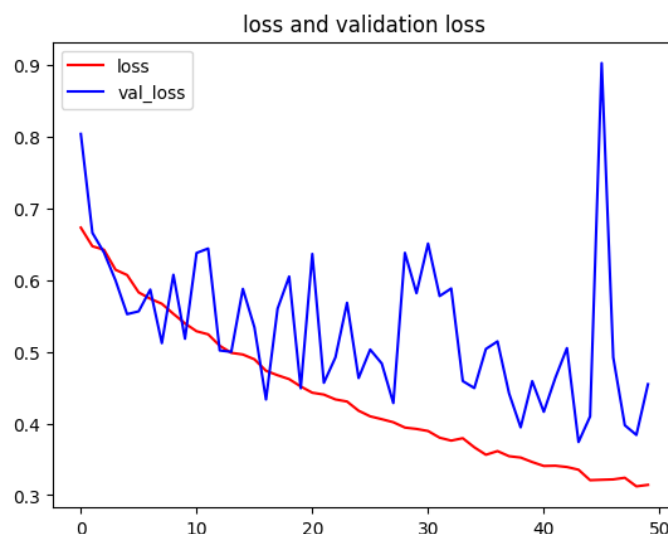
Penelitian ini menerapkan pendekatan *transfer learning* berbasis arsitektur VGG-16 untuk membangun model klasifikasi penyakit mata. Model awal dirancang menggunakan struktur *Convolutional Neural Network* yang terdiri atas lima blok konvolusi dengan jumlah filter yang meningkat secara bertahap (32, 64, 128, 256, dan 512), yang memungkinkan model mengekstraksi fitur visual mulai dari pola sederhana hingga fitur tingkat tinggi. Setiap blok konvolusi diikuti oleh *Batch Normalization* dan *MaxPooling* untuk menjaga stabilitas pelatihan serta mengurangi dimensi spasial tanpa kehilangan informasi penting. Untuk meminimalkan risiko *overfitting*, lapisan *Dropout* diterapkan pada bagian *fully connected*, dengan susunan akhir model mencakup *input layer*, blok konvolusi bertingkat,

Global Average Pooling, dan *dense layer* sebagai pengklasifikasi. Model dilatih menggunakan fungsi kerugian *categorical cross-entropy* dan *optimizer Adam* dengan *learning rate* sebesar 1×10^{-4} , serta metrik evaluasi *accuracy*.

Selanjutnya, strategi *transfer learning* diterapkan dengan memanfaatkan bobot awal VGG-16 yang telah dilatih pada dataset ImageNet. Lapisan konvolusi awal dibekukan (*freezing*) untuk mempertahankan kemampuan ekstraksi fitur dasar, sementara *custom classifier head* digunakan untuk menyesuaikan model dengan karakteristik citra fundus mata. Proses pelatihan tahap ini dilakukan selama 50 epoch. Untuk meningkatkan kemampuan generalisasi model, tahap *fine-tuning* selanjutnya dilakukan dengan membuka sebagian lapisan konvolusi akhir dan melatihnya kembali selama 10 epoch, sehingga model mampu mempelajari representasi fitur yang lebih spesifik terhadap klasifikasi penyakit mata.



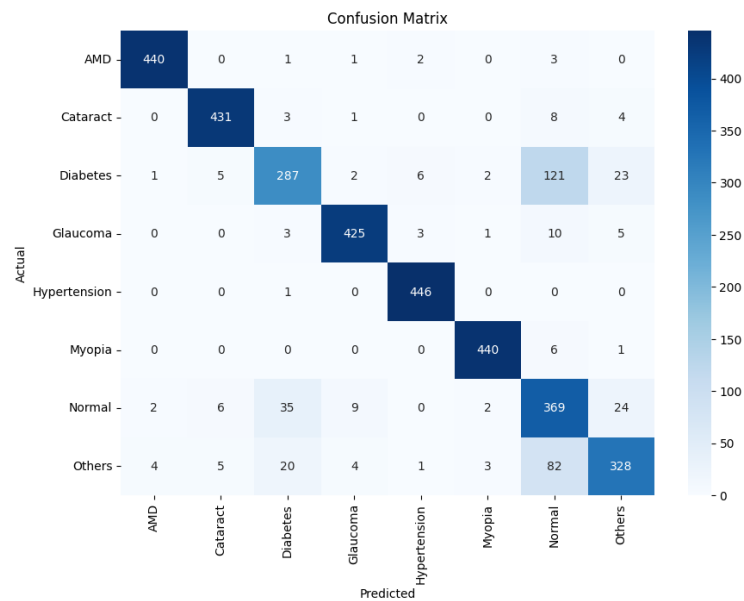
Gambar 3. Visualisasi Training dan Validation Accuracy



Gambar 4. Visualisasi Training dan Validation Loss

Gambar 3 dan Gambar 4 memperlihatkan kinerja model VGG-16 selama 50 epoch proses transfer learning berdasarkan metrik akurasi dan loss. Akurasi pelatihan menunjukkan peningkatan yang konsisten, sedangkan akurasi validasi cenderung meningkat meskipun mengalami fluktuasi pada beberapa epoch. Hal ini menunjukkan bahwa model mampu mempelajari pola fitur citra fundus sekaligus mempertahankan kemampuan generalisasi. Nilai loss pelatihan menurun secara stabil sepanjang proses pelatihan, menandakan proses optimasi yang efektif. Sementara itu, loss validasi menunjukkan fluktuasi dengan lonjakan pada sekitar epoch ke-46. Namun, lonjakan tersebut bersifat sementara dan tidak diikuti oleh peningkatan loss yang berkelanjutan, sehingga tidak mengindikasikan terjadinya overfitting. Secara keseluruhan, jarak antara kurva pelatihan dan validasi relatif terkendali, menunjukkan bahwa model memiliki performa yang stabil dan kemampuan generalisasi yang baik.

D. Hasil Evaluasi



Gambar 5. *Confusion Matrix*

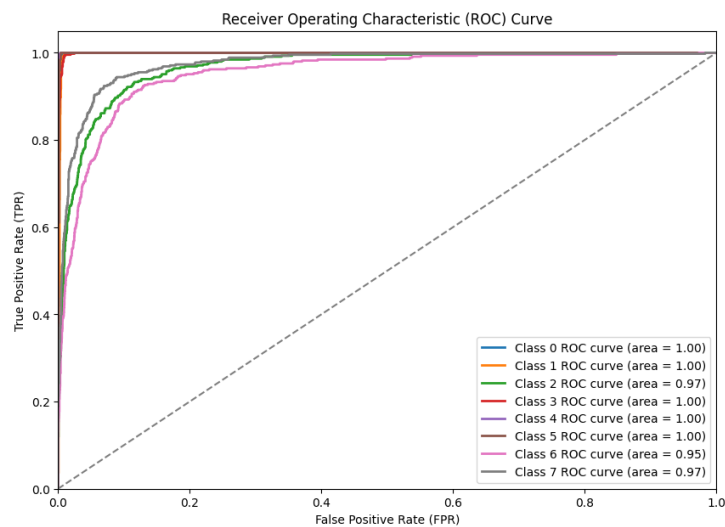
Gambar 5 merupakan *confusion matrix* pada model yang sudah dilatih, menunjukkan hasil evaluasi prediksi model terhadap data uji untuk delapan kelas penyakit mata yang terdapat dalam dataset ODIR-5K. Setiap baris pada matriks mewakili label aktual (*ground truth*), sedangkan setiap kolom menunjukkan label prediksi dari model. Secara umum, kinerja model cukup optimal dalam mendeteksi sebagian besar kelas dengan tingkat kesalahan klasifikasi yang rendah.

TABEL 1
CLASSIFICATION REPORT

Class	Precision	Recall	F1-Score	Support
AMD	0.98	0.98	0.98	447
Cataract	0.96	0.96	0.96	447
Diabetes	0.82	0.64	0.72	447
Glaucoma	0.96	0.95	0.96	447
Hypertension	0.97	1.00	0.99	447
Myopia	0.98	0.98	0.98	447
Normal	0.62	0.83	0.71	447
Others	0.85	0.73	0.79	447
Accuracy			0.89	3576
Macro Avg	0.89	0.89	0.89	3576
Weighted Avg	0.89	0.89	0.89	3576

Tabel 1 mengindikasikan bahwa model menunjukkan performa yang sangat baik dalam mengklasifikasikan sebagian besar kelas penyakit mata. Kelas seperti *AMD*, *Cataract*, *Glaucoma*, *Hypertension*, dan *Myopia* memperoleh nilai *precision*, *recall*, dan *F1-score* yang tinggi, menandakan bahwa model dapat mengklasifikasi citra dari kelas-kelas tersebut dengan akurasi tinggi. Namun, performa model pada kelas *Diabetes* dan *Normal* relatif lebih rendah. Hal ini terlihat dari nilai *recall*, *precision*, dan *f1-score* yang lebih kecil, mengindikasikan bahwa model masih cukup sering melakukan kesalahan dalam membedakan antara dua kelas tersebut, terutama karena kemiripan visual antar citra fundus. Akurasi keseluruhan model mencapai 89%, dengan nilai *Macro Avg* dan *Weighted Avg* yang juga berada di angka 0.89, menunjukkan bahwa model memiliki kinerja yang seimbang dan stabil secara umum dalam menangani klasifikasi multi-kelas pada dataset ODIR-5K.

Berdasarkan *classification report*, model *transfer learning* VGG-16 pada penelitian ini mencapai *accuracy* sebesar 89% dengan nilai *precision*, *recall*, dan *F1-score* yang relatif seimbang pada seluruh kelas. Hasil tersebut menunjukkan peningkatan yang signifikan dibandingkan penelitian Qulub dan Agustin (2024) yang hanya memperoleh *accuracy* sebesar 45% dan mengalami *class bias*. Selain itu, kinerja model yang diusulkan juga melampaui penelitian Dwi Marcella dkk. (2022) yang menggunakan VGG-19 dengan *accuracy* 65.29%, serta metode *multilayer perceptron* oleh Erdiansyah dkk. (2022) dengan *accuracy* 71.80%. Jika dibandingkan dengan metode non-*deep learning* seperti *fractal dimension* dan *fuzzy k-means* oleh Ayu Evilia Suwanda dan Dwi Juniati (2022) dengan *accuracy* 76%, maupun arsitektur *EfficientNet-B0* oleh Syafiq Hilmi Abdullah dkk (2022) dengan *accuracy* 82.096%, model yang diusulkan tetap menunjukkan performa yang lebih tinggi dan lebih stabil. Peningkatan kinerja ini mengindikasikan bahwa kombinasi *transfer learning* VGG-16, *data augmentation*, serta penanganan ketidakseimbangan kelas efektif dalam meningkatkan *accuracy* dan mengurangi bias antar kelas. Namun demikian, kelas *Diabetic* masih menunjukkan nilai *recall* yang relatif lebih rendah dibandingkan kelas lainnya dan jika dibandingkan dengan penelitian yang dilakukan oleh Erdiansyah dkk (2022) dengan *recall* 71.80% ini mendapatkan penurunan dibandingkan penelitian sebelumnya. Hal ini mengindikasikan bahwa klasifikasi *diabetic retinopathy* dalam skenario *multi-class classification* memiliki tingkat kompleksitas yang lebih tinggi dibandingkan pendekatan *single-class classification* pada penelitian terdahulu.



Gambar 6. Receiver Operating Characteristic (ROC)

Gambar 6 menunjukkan kurva ROC (*Receiver Operating Characteristic*) untuk masing-masing kelas, yang menggambarkan hubungan antara *True Positive Rate* (TPR) dan *False Positive Rate* (FPR) dalam klasifikasi multi-kelas. Setiap kurva mewakili performa model dalam membedakan satu kelas terhadap kelas lainnya (*One-vs-Rest*). Berdasarkan gambar 8, menunjukkan bahwa performa masing-masing kelas sangat baik dikarenakan mendekati satu dan sama dengan satu.

TABEL 2
AREA UNDER CURVE

Class	Area
AUC For Class 0 (AMD)	0.9995
AUC For Class 1 (Cataract)	0.9979
AUC For Class 2 (Diabetic)	0.9660
AUC For Class 3 (Glaucoma)	0.9989
AUC For Class 4 (Hypertension)	0.9999
AUC For Class 5 (Myopia)	0.9997
AUC For Class 6 (Normal)	0.9516
AUC For Class 7 (Others)	0.9742

Tabel 2 menyajikan nilai *Area Under Curve* (AUC) yang digunakan untuk mengevaluasi kemampuan model dalam melakukan diskriminasi antar kelas. Nilai AUC yang diperoleh menunjukkan bahwa model memiliki kemampuan klasifikasi yang sangat baik, ditandai dengan nilai AUC yang tinggi pada seluruh kelas dengan rata-rata

di atas 0.95. Capaian ini mengindikasikan bahwa model mampu membedakan secara efektif antara kelas penyakit mata yang berbeda, termasuk kelas normal dan kelas dengan kondisi patologis. Nilai AUC yang konsisten tinggi juga mencerminkan stabilitas performa model serta kemampuan generalisasi yang baik dalam menghadapi variasi data citra fundus.

Hasil pelatihan dan evaluasi model menunjukkan bahwa model transfer learning VGG-16 yang diusulkan memiliki kinerja yang baik dalam mengklasifikasikan citra fundus mata ke dalam masing-masing kelas penyakit. Berdasarkan *confusion matrix*, model mampu memberikan prediksi yang akurat pada kelas-kelas dengan karakteristik visual yang kuat dan relatif konsisten, seperti *Hypertension*, *Myopia*, dan *Glaucoma*. Tingginya tingkat prediksi benar pada kelas-kelas tersebut mengindikasikan bahwa model efektif dalam mengekstraksi dan mempelajari pola visual penting dari citra fundus. Namun demikian, masih ditemukan kesalahan klasifikasi pada kelas dengan karakteristik visual yang saling tumpang tindih, khususnya pada kelas *Diabetes* yang beberapa kali terklasifikasi sebagai *Normal*, serta pada kelas *Others* yang merupakan gabungan dari beberapa kondisi penyakit mata di luar kelas utama. Kondisi ini menunjukkan bahwa kompleksitas variasi visual pada kelas-kelas tersebut masih menjadi tantangan bagi model dalam melakukan pemisahan kelas secara optimal. Kinerja model secara keseluruhan menunjukkan peningkatan yang signifikan dibandingkan sejumlah penelitian terdahulu, terutama dalam hal stabilitas prediksi dan pengurangan *class bias*. Kendati demikian, penelitian ini masih memiliki beberapa keterbatasan. Dataset *Ocular Disease Recognition* (ODR) yang digunakan memiliki ketidakseimbangan jumlah data antar kelas, yang meskipun telah ditangani melalui penerapan *Synthetic Minority Over-sampling Technique* (SMOTE) dan *data augmentation*, tetap berpotensi membatasi representasi variasi klinis yang sesungguhnya. Selain itu, proses pelatihan dan evaluasi model hanya dilakukan pada satu sumber dataset, sehingga kemampuan generalisasi model terhadap data dari institusi atau kondisi klinis yang berbeda belum dapat dipastikan sepenuhnya. Potensi bias juga dapat muncul akibat dominasi kelas tertentu serta variasi kualitas citra fundus dan karakteristik pasien yang tidak terdistribusi secara merata.

Terlepas dari keterbatasan tersebut, hasil penelitian ini memberikan kontribusi yang relevan baik secara praktis maupun teoretis. Secara praktis, model yang diusulkan berpotensi dikembangkan sebagai sistem pendukung keputusan (*clinical decision support system*) untuk membantu tenaga medis dalam melakukan skrining awal penyakit mata berbasis citra fundus secara cepat dan objektif. Penerapan sistem semacam ini diharapkan dapat meningkatkan efisiensi layanan kesehatan mata serta mendukung deteksi dini penyakit mata, khususnya di wilayah dengan keterbatasan tenaga spesialis. Secara teoretis, penelitian ini memperkuat bukti bahwa kombinasi *transfer learning* VGG-16, penanganan ketidakseimbangan data, dan strategi pelatihan yang tepat dapat meningkatkan kinerja klasifikasi citra medis multi-kelas, sekaligus menjadi dasar bagi pengembangan arsitektur dan pendekatan yang lebih adaptif pada penelitian selanjutnya.

IV. SIMPULAN

Penelitian ini bertujuan mengembangkan model klasifikasi citra fundus mata menggunakan algoritma Transfer Learning VGG-16 guna mengklasifikasikan berbagai kategori penyakit mata, seperti diabetes, glaukoma, katarak, dan lainnya. Berdasarkan proses yang telah dilakukan, beberapa poin penting dapat disimpulkan adalah dataset *Ocular Disease Recognition* (ODIR-5K) yang digunakan telah diolah melalui langkah-langkah preprocessing seperti SMOTE, augmentasi data, dan perubahan ukuran gambar menjadi 224x224 piksel, sesuai dengan arsitektur VGG-16. Model VGG-16 dipilih karena kemampuannya mengekstraksi fitur spasial dari citra medis, sehingga cocok untuk menganalisis visual kompleks pada gambar fundus mata. *Hyperparameter tuning* dilakukan untuk meningkatkan performa model, termasuk penyesuaian learning rate, batch size, dan jumlah epochs. Evaluasi model dilakukan menggunakan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score* untuk mengukur kemampuan model dalam klasifikasi penyakit mata. Setelah melewati langkah-langkah tersebut, model VGG-16 dapat menghasilkan model yang mengalami peningkatan yang signifikan dengan *accuracy* 89% dari penelitian sebelumnya. *F1-score* masing-masing kelas yaitu AMD 0.98, *cataract* 0.96, *diabetic* 0.72, *glaucoma* 0.96, *hypertension* 0.99, *myopia* 0.98, *normal* 0.71, *others* 0.79. Selanjutnya, nilai *precision* masing-masing kelas yaitu AMD 0.98, *cataract* 0.96, *diabetic* 0.82, *glaucoma* 0.96, *hypertension* 0.97, *myopia* 0.98, *normal* 0.62, *others* 0.85. Adapun nilai *recall* masing-masing kelas yaitu AMD 0.98, *cataract* 0.96, *diabetic* 0.64, *glaucoma* 0.95, *hypertension* 1.00, *myopia* 0.98, *normal* 0.83, *others* 0.73. Meskipun model yang diusulkan telah menunjukkan peningkatan dalam kinerja yang cukup signifikan, model ini terdapat keterbatasan, terutama pada dataset antara kelas yang awalnya tidak seimbang sehingga diperlukannya penggunaan SMOTE sehingga dapat terjadi *bias* dalam model, serta ada kekurangan pada kelas *normal* dengan *recall* 0.71, *precision* 0.62, serta pada kelas *diabetic* dengan *f1-score* 0.72, dan *recall* 0.64, ini mengindikasikan bahwa data pada kedua kelas ini dapat ditingkatkan kembali agar hasil yang didapatkan bisa lebih memuaskan. Penelitian selanjutnya dapat difokuskan untuk pengumpulan dataset lapangan dari multi-sumber dan penerapan arsitektur deep learning yang lebih kompleks.

UCAPAN TERIMAKASIH

Ucapakan terima kasih yang sebesar-besarnya disampaikan kepada seluruh pihak yang telah memberikan dukungan serta kontribusi selama pelaksanaan penelitian ini. Secara khusus, ucapan terima kasih yang mendalam kepada Bapak Ary Prabowo, S.Kom, M.Kom, selaku dosen pembimbing tugas akhir dan penulisan jurnal atas bimbingan, ilmu, serta dukungan yang telah beliau berikan. Masukan serta arahan yang beliau berikan telah sangat membantu dalam memahami konsep-konsep mendasar dan implementasi dari penelitian ini.

DAFTAR PUSTAKA

- [1] C. R. Mulyasari, A. Id Hadiana, A. Komarudin, U. Jenderal, and A. Yani, "DETEKSI PENYAKIT DIABETES, KATARAK DAN GLAUKOMA PADA CITRA FUNDUS RETINA MATA MANUSIA MENGGUNAKAN CNN DENGAN ARSITEKTUR ALEXNET," vol. 8, no. 1, pp. 53–68, 2024.
- [2] M. Elkholy and M. A. Marzouk, "Deep learning-based classification of eye diseases using Convolutional Neural Network for OCT images," *Front. Comput. Sci.*, vol. 5, 2023, doi: 10.3389/fcomp.2023.1252295.
- [3] A. Hizriadi, S. Purnamawati, F. Angreni, and B. Gtg, "KLIK: Kajian Ilmiah Informatika dan Komputer Implementation of Extreme Learning Machine for Classification of Retina Ablasio Results on Retina Fundus Images," *Media Online*, vol. 3, no. 4, pp. 371–376, 2023, [Online]. Available: <https://djournals.com/klik>
- [4] H. Hardiman, "Pengenalan Pola Aksara Ulu Banyuasin dengan Metode Convolutional Neural Network dan Support Vector Machine," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 10, no. 4, pp. 1069–1079, Sep. 2025, doi: 10.30591/jpit.v10i4.8793.
- [5] K. Alifian Hernanda Wibowo, "Indonesian Journal of Mathematics and Natural Sciences Klasifikasi Diabetic Retinopathy Menggunakan CNN dengan Arsitektur yang Dimodifikasi," 2023. [Online]. Available: <http://journal.unnes.ac.id/nju/index.php/JM>
- [6] S. Godbole, A. Kattukaran, S. Savla, V. Pradhan, P. Kanani, and D. Patil, "Enhancing Paediatric Pneumonia Detection and Classification Using Customized CNNs and Transfer Learning Based Ensemble Models," *International Research Journal of Multidisciplinary Technovation*, vol. 6, no. 6, pp. 38–53, Nov. 2024, doi: 10.54392/irjmt2463.
- [7] S. Sharma and K. Guleria, "A Deep Learning based model for the Detection of Pneumonia from Chest X-Ray Images using VGG-16 and Neural Networks," in *Procedia Computer Science*, Elsevier B.V., 2022, pp. 357–366. doi: 10.1016/j.procs.2023.01.018.
- [8] T. Gayathri and K. S. Kumar, "BRAIN TUMOR SEGMENTATION AND CLASSIFICATION USING CNN PRE-TRAINED VGG-16 MODEL IN MRI IMAGES," *IIUM Engineering Journal*, vol. 25, no. 2, pp. 196–211, 2024, doi: 10.31436/iiumej.v25i2.2963.
- [9] D. Marcella and S. Devella, "Klasifikasi Penyakit Mata Menggunakan Convolutional Neural Network Dengan Arsitektur VGG-19," vol. 3, no. 1, pp. 60–70, 2022.
- [10] M. S. Qulub and S. Agustin, "IDENTIFIKASI PENYAKIT MATA DENGAN KLASIFIKASI CITRA FOTO FUNDUS MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK (CNN)," 2024.
- [11] D. Marcella and S. Devella, "Klasifikasi Penyakit Mata Menggunakan Convolutional Neural Network Dengan Arsitektur VGG-19," vol. 3, no. 1, pp. 60–70, 2022.
- [12] U. Erdiansyah, A. I. Lubis, and G. Syahputra, "JAISE : Journal of Artificial Intelligence and Software Engineering Klasifikasi Penyakit Diabetic Retinopathy Menggunakan Multilayer Perceptron," 2022.
- [13] J. Penelitian, M. Dan, P. Matematika, A. E. Suwanda, D. Juniati, and U. N. Surabaya, "Klasifikasi Penyakit Mata Berdasarkan Citra Fundus Retina Menggunakan Dimensi Fraktal Box Counting Dan Fuzzy K-Means," 2022.
- [14] S. H. Abdullah, R. Magdalena, and R. Y. N. Fu'adah, "KLASIFIKASI DIABETIC RETINOPATHY BERBASIS PENGOLAHAN CITRA FUNDUS DAN DEEP LEARNING," *JOURNAL OF ELECTRICAL AND SYSTEM CONTROL ENGINEERING*, vol. 5, no. 2, pp. 84–90, Feb. 2022, doi: 10.31289/jesce.v5i2.5659.
- [15] T. Berliani, E. Rahardja, and L. Septiana, "Perbandingan Kemampuan Klasifikasi Citra X-ray Paru-paru menggunakan Transfer Learning ResNet-50 dan VGG-16," *Journal of Medicine and Health*, vol. 5, no. 2, pp. 123–135, Aug. 2023, doi: 10.28932/jmh.v5i2.6116.
- [16] K. Kualitas Buah Salak dengan Transfer Learning Arsitektur VGG and A. Luthfiarta, "VGG16 Transfer Learning Architecture for Salak Fruit Quality Classification," *Jurnal Informatika dan Teknologi Informasi*, vol. 18, no. 1, pp. 37–48, 2021, doi: 10.31515/telematika.v18i1.4025.