

**KLASIFIKASI PENYAKIT MATA BERDASARKAN CITRA FUNDUS
MENGUNAKAN METODE *MULTI-LAYER PERCEPTRON***

SKRIPSI

Oleh :
LAUDZA ATSILA PRASETYO
NIM. 220605110144



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025**

**KLASIFIKASI PENYAKIT MATA BERDASARKAN CITRA FUNDUS
MENGUNAKAN METODE *MULTI-LAYER PERCEPTRON***

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh :
LAUDZA ATSILA PRASETYO
NIM. 220605110144

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025**

HALAMAN PERSETUJUAN

KLASIFIKASI PENYAKIT MATA BERDASARKAN CITRA FUNDUS MENGUNAKAN METODE *MULTI-LAYER PERCEPTRON*

SKRIPSI

Oleh :
LAUDZA ATSILA PRASETYO
NIM. 220605110144

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 9 Desember 2025

Pembimbing I,



Okta Qomaruddin Aziz, M.Kom
NIP. 19911019 201903 1 013

Pembimbing II,



Tri Mukti Lestari, M.Kom
NIP. 19911108 202012 2 005

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Shapriyono, M.Kom
NIP. 19841010 201903 1 012

HALAMAN PENGESAHAN

KLASIFIKASI PENYAKIT MATA BERDASARKAN CITRA FUNDUS MENGUNAKAN METODE *MULTI-LAYER PERCEPTRON*

SKRIPSI

Oleh :
LAUDZA ATSILA PRASETYO
NIM. 220605110144

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 24 Desember 2025

Susunan Dewan Penguji

Ketua Penguji	: <u>Dr. Cahyo Crysdian, M.Cs</u> NIP. 19740424 200901 1 008
Anggota Penguji I	: <u>Dr. Zainal Abidin, M.Kom</u> NIP. 19760613 200501 1 004
Anggota Penguji II	: <u>Okta Qomaruddin Aziz, M.Kom</u> NIP. 19911019 201903 1 013
Anggota Penguji III	: <u>Tri Mukti Lestari, M.Kom</u> NIP. 19911108 202012 2 005

()
()
()
()

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

()

Supriyono, M.Kom
NIP. 19841010 201903 1 012

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Laudza Atsila Prasetyo
NIM : 220605110144
Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika
Judul Skripsi : Klasifikasi Penyakit Mata Berdasarkan Citra Fundus Menggunakan Metode *Multi-Layer Perceptron*.

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 24 Desember 2025
Yang membuat pernyataan,



Laudza Atsila Prasetyo
NIM 220605110144

MOTTO

“Memento Mori”

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur ke hadirat Allah SWT atas segala rahmat, kemudahan, dan kekuatan-Nya, skripsi ini dapat terselesaikan dengan baik. Untuk itu, penulis persembahkan skripsi ini kepada;

1. Kedua orang tua saya, yakni Alm Nanang Prasetyo, S.H dan Metta Oktaviana, AMd. RO. yang selalu memberikan dukungan baik moril maupun materi, serta kasih sayang dan doa yang tidak pernah terputus dari awal penulis lahir hingga saat ini, terima kasih telah menjadi orang tua yang sempurna, menjadi teladan dalam hidupku. Adik penulis Syauqi Aludra Prasetyo, terimakasih atas perhatian, dukungan dan doa yang tak pernah terputus.
2. Dosen penulis, yakni Bapak Okta Qomaruddin Aziz, M.Kom., dan Ibu Roro Inda Melani, M.T, M.Sc yang telah banyak membimbing dan memberikan motivasi kepada penulis baik dalam perkuliahan serta dalam proses penulisan skripsi ini.
3. Teman-teman penulis, Infinity 2022 khususnya RAR NIM 044 terimakasih telah menjadi teman diskusi yang sangat sabar dan pendukung penulis selama masa perkuliahan. Teman se-bimbingan Husnul Khatimah dan Gavril Pandita, terimakasih atas dukungan, bantuan dan diskusi selama proses perkuliahan dan proses penulisan ini. Teman *internship* di Estungkara, Hilya Zahra Alifa yang telah menjadi teman *sharing* selama 2

bulan. Teman *sharing*, Fadhillah, Citra, Sivana dan Silvi terimakasih telah menghibur penulis saat masa perkuliahan dan masa penulisan skripsi.

4. Rekan-rekan "BPH LOPE", Kak Maulidya, Renata, Kak Aisha, Kak Reyhan dan Kak Gianda terimakasih atas kebersamaan, dukungan dan kerja sama yang sangat berkesan selama menjalankan tugas di HMPS Teknik Informatika "Encoder".
5. Kakak-kakak magister, Kak Alfina dan Kak Salma, yang secara tidak langsung selalu memberikan afirmasi positif, memberikan motivasi dan semangat kepada penulis dalam 1 semester terakhir ini.
6. Diriku sendiri, yang telah bersabar, bertahan, belajar, dan berjuang melewati setiap tahapan pembelajaran dari awal masa perkuliahan hingga akhirnya skripsi ini dapat terselesaikan.
7. Paul, Hikedan, Cup Of Joe, PMK, DYSWIS dan seluruh penyanyi dan *podcaster* dalam playlist penulis yang menemani hari-hari penulis saat menyusun skripsi.
8. Serta semua pihak yang telah membantu dan memberikan dukungan dalam bentuk apapun, yang tidak dapat disebutkan satu per satu.

KATA PENGANTAR

Assalamualaikum warahmatullahi wabarakatuh

Alhamdulillah rabbilalamin, segala puji dan syukur senantiasa penulis panjatkan ke hadirat Allah Subhanahu Wa Ta'ala atas segala nikmat dan kasih sayang-Nya yang telah memudahkan penulis untuk menyelesaikan skripsi yang berjudul “Klasifikasi Penyakit Mata Berdasarkan Citra Fundus Menggunakan Metode *Multi-Layer Perceptron*” dengan baik. Shalawat serta salam tetap tercurahkan kepada Nabi Muhammad SAW, yang telah membawa kita dari zaman kebodohan menuju zaman kebenaran yakni Islam, zaman yang penuh dengan ilmu pengetahuan sebagaimana yang di rasakan pada saat ini. Dan semoga kita semua mendapat syafaatnya di hari akhir kelak, Aamiin.

Penulis mengucapkan rasa syukur dan terima kasih yang tak terhingga kepada semua pihak-pihak yang selalu memberikan bantuan dan motivasi kepada penulis untuk dapat menyelesaikan skripsi ini. Ucapan ini penulis sampaikan kepada:

1. Okta Qomaruddin Aziz, M.Kom. selaku dosen pembimbing I dan Tri Mukti Lestari M.Kom. selaku dosen pembimbing II yang selalu memberikan banyak bantuan, bimbingan, *support*, motivasi dan arahan kepada penulis, sehingga bisa menyelesaikan skripsi ini.
2. Dr. Cahyo Crysdiyan, M.Cs. selaku dosen penguji I dan Dr. Zainal Abidin M.Kom. selaku dosen penguji II yang telah menguji serta memberikan saran dan masukan sehingga penulis dapat menyelesaikan skripsi dengan baik.

3. Seluruh dosen dan staf Fakultas Saintek Universitas Islam Negeri Malang, atas ilmu, bimbingan, dan bantuan administratif yang telah diberikan selama penulis menempuh studi.

Akhir kata, penulis mengakui bahwa penulisan pada skripsi ini masih banyak kekurangan. Penulis berharap semoga skripsi ini dapat menjadi bagian dari kontribusi dalam rangka memperkuat dan mengembangkan ilmu pengetahuan .Wassalamualaikum Warahmatullahi Wabarakatuh.

Malang, 24 Desember 2025

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	xi
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL	xiv
ABSTRAK	xvi
ABSTRACT	xvii
البحث مستخلص.....	xviii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Pernyataan Masalah	6
1.3 Batasan Masalah.....	6
1.4 Tujuan Penelitian	7
1.5 Manfaat Penelitian	7
BAB II STUDI PUSTAKA	8
2.1 Sistem Klasifikasi Penyakit Mata	8
2.2 Karakteristik Penyakit Mata.....	13
2.3 Multi-Layer Perceptron (MLP)	20
BAB III DESAIN DAN IMPLEMENTASI	24
3.1 Pengambilan Data	24
3.2 Desain Sistem.....	25
3.2.1 Pre-processing	25
3.2.2 Ekstraksi Fitur	28
3.2.3 Desain MLP.....	34
BAB IV UJI COBA DAN PEMBAHASAN.....	40
4.1 Persiapan Data.....	40
4.2 Skenario Uji Coba.....	41
4.3 Hasil Uji Coba.....	46
4.3.1 Skenario Uji Coba 1 (A-2H-32)	46
4.3.2 Skenario Uji Coba 2 (A-2H-64)	48
4.3.3 Skenario Uji Coba 3 (A-3H-64)	50
4.3.4 Skenario Uji Coba 4 (A-3H-128)	52
4.3.5 Skenario Uji Coba 5 (B-2H-32)	54
4.3.6 Skenario Uji Coba 6 (B-2H-64)	56
4.3.7 Skenario Uji Coba 7 (B-3H-64)	57
4.3.8 Skenario Uji Coba 8 (B-3H-128)	59
4.3.9 Skenario Uji Coba 9 (C-2H-32)	60
4.3.10 Skenario Uji Coba 10 (C-2H-64)	62

4.3.11 Skenario Uji Coba 11 (C-3H-64)	64
4.3.12 Skenario Uji Coba 12 (C-3H-128)	66
4.3.13 Skenario Uji Coba 13 (D-2H-32)	68
4.3.14 Skenario Uji Coba 14 (D-2H-64)	69
4.3.15 Skenario Uji Coba 15 (D-3H-64)	71
4.3.16 Skenario Uji Coba 16 (D-3H-128)	73
4.4 Pembahasan.....	74
4.4.1 Pengaruh Perlakuan Data	81
4.4.2 Pengaruh Arsitektur MLP	86
BAB V KESIMPULAN DAN SARAN	93
5.1 Kesimpulan	93
5.2 Saran.....	94
DAFTAR PUSTAKA	
LAMPIRAN	

DAFTAR GAMBAR

Gambar 2. 1 Citra Normal.....	15
Gambar 2. 2 Citra Diabetic	17
Gambar 2. 3 Citra Cataract.....	18
Gambar 2. 4 Citra Glaucoma	19
Gambar 2. 5 Arsitektur Multi-Layer Perceptron.....	21
Gambar 3. 1 Desain Sistem.....	25
Gambar 3. 2 Penerapan <i>Green Channel Extraction</i>	26
Gambar 3. 3 Penerapan <i>Illumination Correction</i>	27
Gambar 3. 4 Penerapan <i>Contrast Enhancement</i>	28
Gambar 3. 5 Contoh matriks GLCM probabilitas.....	29
Gambar 3. 6 Arsitektur Model MLP	34
Gambar 4. 1 Visualisasi Rata-rata <i>evaluation matrix</i> Skenario 1	47
Gambar 4. 2 Visualisasi Rata-rata <i>evaluation matrix</i> Skenario 2	49
Gambar 4. 3 Visualisasi Rata-rata <i>evaluation matrix</i> Skenario 3	51
Gambar 4. 4 Visualisasi Rata-rata <i>evaluation matrix</i> Skenario 4	53
Gambar 4. 5 Visualisasi Rata-rata <i>evaluation matrix</i> Skenario 5	55
Gambar 4. 6 Visualisasi Rata-rata <i>evaluation matrix</i> Skenario 6	56
Gambar 4. 7 Visualisasi Rata-rata <i>evaluation matrix</i> Skenario 7	58
Gambar 4. 8 Visualisasi Rata-rata <i>evaluation matrix</i> Skenario 8	60
Gambar 4. 9 Visualisasi Rata-rata <i>evaluation matrix</i> Skenario 9	61
Gambar 4. 10 Visualisasi Rata-rata <i>evaluation matrix</i> Skenario 10	63
Gambar 4. 11 Visualisasi Rata-rata <i>evaluation matrix</i> Skenario 11	65
Gambar 4. 12 Visualisasi Rata-rata <i>evaluation matrix</i> Skenario 12	67
Gambar 4. 13 Visualisasi Rata-rata <i>evaluation matrix</i> Skenario 13	69
Gambar 4. 14 Visualisasi Rata-rata <i>evaluation matrix</i> Skenario 14	70
Gambar 4. 15 Visualisasi Rata-rata <i>evaluation matrix</i> Skenario 15	72
Gambar 4. 16 Visualisasi Rata-rata <i>evaluation matrix</i> Skenario 16	73
Gambar 4. 17 Perbandingan Model A (Data <i>Unbalanced</i>).....	75
Gambar 4. 18 Perbandingan Model B (Augmentasi pada Data <i>Unbalanced</i>).....	76
Gambar 4. 19 Perbandingan Model C (Data <i>Undersampling</i>).....	77
Gambar 4. 20 Perbandingan Model D (Data <i>Undersampling</i> + Augmentasi).....	78
Gambar 4. 21 Perbandingan Perlakuan Data	86
Gambar 4. 22 Perbandingan Arsitektur MLP	88

DAFTAR TABEL

Tabel 2. 1 Penelitian Terkait Klasifikasi Penyakit Mata.....	11
Tabel 3. 1 Penjelasan Atribut.....	24
Tabel 4. 1 Jumlah Data Pada Proses <i>Data Cleaning</i>	40
Tabel 4. 2 Jumlah Data pada Proses Augmentasi	41
Tabel 4. 3 Jumlah Data Tiap Model.....	41
Tabel 4. 4 Variasi Arsitektur Jaringan	42
Tabel 4.5 Kombinasi skenario pengujian.....	43
Tabel 4. 6 <i>Confussion Matrix</i>	44
Tabel 4. 7 <i>Evaluation Matrix</i> K-Fold Skenario 1 (Model A)	46
Tabel 4. 8 Rata-rata <i>Evaluation Matrix</i> per kelas	48
Tabel 4. 9 <i>Evaluation Matrix</i> K-Fold Skenario 2 (Model A)	48
Tabel 4. 10 Rata-rata <i>Evaluation Matrix</i> per kelas	49
Tabel 4. 11 <i>Evaluation Matrix</i> K-Fold Skenario 3 (Model A)	50
Tabel 4. 12 Rata-rata <i>Evaluation Matrix</i> per kelas	51
Tabel 4. 13 <i>Evaluation Matrix</i> K-Fold Skenario 4 (Model A)	52
Tabel 4. 14 Rata-rata <i>Evaluation Matrix</i> per kelas	53
Tabel 4. 15 <i>Evaluation Matrix</i> K-Fold Skenario 5 (Model B).....	54
Tabel 4. 16 Rata-rata <i>Evaluation Matrix</i> per kelas	55
Tabel 4. 17 <i>Evaluation Matrix</i> K-Fold Skenario 6 (Model B).....	56
Tabel 4. 18 Rata-rata <i>Evaluation Matrix</i> per kelas	57
Tabel 4. 19 <i>Evaluation Matrix</i> K-Fold Skenario 7 (Model B).....	57
Tabel 4. 20 Rata-rata <i>Evaluation Matrix</i> per kelas	59
Tabel 4. 21 <i>Evaluation Matrix</i> K-Fold Skenario 8 (Model B).....	59
Tabel 4. 22 Rata-rata <i>Evaluation Matrix</i> per kelas	60
Tabel 4. 23 <i>Evaluation Matrix</i> K-Fold Skenario 9 (Model C).....	61
Tabel 4. 24 Rata-rata <i>Evaluation Matrix</i> per kelas	62
Tabel 4. 25 <i>Evaluation Matrix</i> K-Fold Skenario 10 (Model C).....	63
Tabel 4. 26 Rata-rata <i>Evaluation Matrix</i> per kelas	64
Tabel 4. 27 <i>Evaluation Matrix</i> K-Fold Skenario 11 (Model C).....	64
Tabel 4. 28 Rata-rata <i>Evaluation Matrix</i> per kelas	65
Tabel 4. 29 <i>Evaluation Matrix</i> K-Fold Skenario 10 (Model C).....	66
Tabel 4. 30 Rata-rata <i>Evaluation Matrix</i> per kelas	67
Tabel 4. 31 <i>Evaluation Matrix</i> K-Fold Skenario 13 (Model D)	68
Tabel 4. 32 Rata-rata <i>Evaluation Matrix</i> per kelas	69
Tabel 4. 33 <i>Evaluation Matrix</i> K-Fold Skenario 14 (Model D)	70
Tabel 4. 34 Rata-rata <i>Evaluation Matrix</i> per kelas	71
Tabel 4. 35 <i>Evaluation Matrix</i> K-Fold Skenario 15 (Model D)	71
Tabel 4. 36 sRata-rata <i>Evaluation Matrix</i> per kelas.....	72
Tabel 4. 37 <i>Evaluation Matrix</i> K-Fold Skenario 16 (Model D)	73
Tabel 4. 38 Rata-rata <i>Evaluation Matrix</i> per kelas	74
Tabel 4. 39 Hasil <i>K-Fold Cross Validation</i>	79
Tabel 4. 40 Hasil Uji Signifikansi.....	81
Tabel 4. 41 Rata-rata akurasi pada Model A (<i>Unbalanced</i>)	81

Tabel 4. 42 Rata-rata Akurasi pada Model B (Data Augmentasi)	82
Tabel 4. 43 Rata-rata Akuradi pada Model C (Data Undersampling)	83
Tabel 4. 44 Rata-rata Akurasi pada Model D (Data <i>Undersampling</i> + Augmentasi)	84
Tabel 4. 45 Akurasi pada arsitektur 2 <i>Hidden Layer</i>	86
Tabel 4. 46 Akurasi pada arsitektur 3 <i>Hidden Layer</i>	87

ABSTRAK

Prasetyo, Laudza Atsila. 2025. **Klasifikasi Penyakit Mata Berdasarkan Citra Fundus Menggunakan Metode *Multi-Layer Perceptron***. Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Okta Qomaruddin Aziz, M.Kom (II) Tri Mukti Lestari, M.Kom.

Kata kunci: Klasifikasi Penyakit Mata, MLP, Fundus, Ekstraksi Fitur.

Penelitian ini bertujuan untuk mengukur performa *Multi-Layer Perceptron (MLP)* dalam mengklasifikasikan penyakit mata berdasarkan citra fundus dari dataset ODIR yang terdiri atas empat kelas: *Normal*, *Diabetic*, *Glaucoma*, dan *Cataract*. Metode yang digunakan meliputi ekstraksi ciri menggunakan GLCM dan Gabor, pengolahan data melalui cleaning, augmentasi, dan undersampling, serta pengujian 16 skenario model dengan variasi jumlah *hidden layer* yaitu 1 dan 2 *hidden layer* serta variasi jumlah neuron. Hasil menunjukkan bahwa keseimbangan dan jumlah data merupakan faktor yang paling menentukan performa model, di mana perlakuan undersampling dan augmentasi menghasilkan kinerja terbaik. Arsitektur optimal diperoleh pada konfigurasi 64–32 neuron dengan rerata akurasi 73,06%. Secara keseluruhan, penelitian ini menyimpulkan bahwa kombinasi antara dataset seimbang dan arsitektur MLP yang proporsional mampu meningkatkan kemampuan model dalam klasifikasi penyakit mata berbasis citra fundus.

ABSTRACT

Prasetyo, Laudza Atsila. 2025. **Eye Disease Classification Based on Fundus Images Using Multi-Layer Perceptron Method**. Undergraduate Thesis. Department of Informatics Engineering, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Supervisors: (I) Okta Qomaruddin Aziz, M.Kom (II) Tri Mukti Lestari, M.Kom.

Keywords: Eye Disease Classification, MLP, Fundus, Feature Extraction.

This research aims to evaluate the performance of the Multi-Layer Perceptron (MLP) in classifying eye diseases based on fundus images from the ODIR dataset, which consists of four classes: Normal, Diabetic, Glaucoma, and Cataract. The methodology includes feature extraction using GLCM and Gabor, data preprocessing through cleaning, augmentation, and undersampling, as well as testing 16 model scenarios with variations in the number of hidden layers (1 and 2 layers) and neuron configurations. The results show that data balance and dataset size are the most influential factors affecting model performance, with the best results achieved through the combination of undersampling and augmentation. The optimal architecture was obtained using the 64–32 neuron configuration, yielding an mean accuracy of 73.06%. Overall, this study concludes that combining a balanced dataset with a proportional MLP architecture significantly improves the model's ability to classify eye diseases from fundus images.

البحث مستخلص

فراستيو، لوذا اثل. 2025. تصنيف أمراض العين اعتمادًا على صور قاع العين باستخدام طريقة الشبكة العصبية متعددة الطبقات (MLP). رسالة بكالوريوس. قسم هندسة المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانغ. المشرفان: (الأول) أوكتا قمر الدين عزيز، ماجستير في علوم الحاسوب، (الثانية) تري موكتي ليستاري، ماجستير في علوم الحاسوب.

الكلمات المفتاحية: تصنيف أمراض العين، الشبكة العصبية متعددة الطبقات، قاع العين، استخراج الخصائص.

يهدف هذا البحث إلى تقييم أداء الشبكة العصبية متعددة الطبقات (MLP) في تصنيف أمراض العين اعتمادًا على صور قاع العين المأخوذة من مجموعة بيانات ODIR، والتي تتكون من أربع فئات هي: العين السليمة، اعتلال الشبكية السكري، الجلوكوما، وإعتام عدسة العين. تتضمن المنهجية استخراج الخصائص باستخدام طريقي Gabor و GLCM، ومعالجة البيانات من خلال التنقية، وزيادة البيانات، وتقليل التحيز عبر التقليل العشوائي (undersampling)، بالإضافة إلى اختبار 16 سيناريو مختلف للنماذج مع تنوعات في عدد الطبقات المخفية (طبقة واحدة وطبقتان) وتكوينات مختلفة لعدد العصبونات. أظهرت النتائج أن توازن البيانات وحجم مجموعة البيانات يُعدّان من أكثر العوامل تأثيرًا على أداء النموذج، حيث تم تحقيق أفضل النتائج من خلال الجمع بين التقليل العشوائي وزيادة البيانات. وقد تم الحصول على البنية المثلى باستخدام تكوين العصبونات 32-64، محققًا متوسط دقة بلغ 73.06%. وبشكل عام، خلصت هذه الدراسة إلى أن الجمع بين مجموعة بيانات متوازنة وبنية متناسبة للشبكة العصبية متعددة الطبقات يسهم بشكل ملحوظ في تحسين قدرة النموذج على تصنيف أمراض العين اعتمادًا على صور قاع العين.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penyakit mata seperti *Diabetik retinopati*, *Glaucoma*, dan *Cataract* merupakan penyebab utama kebutaan di dunia (World Health Organization, 2019). Hal tersebut ditunjukkan oleh data *Blindness and Vision Impairment* yang diterbitkan oleh WHO pada 10 Agustus tahun 2023, di antara 1 miliar orang kondisi utama yang menyebabkan gangguan pengelihatian jarak jauh atau kebutaan adalah *Cataract* (94 juta), kelainan refraksi (88,4 juta), degenerasi makula yang berkaitan dengan usia (8 juta), *Glaucoma* (7,7 juta), *Diabetic Retinopathy* (3,9 juta). Deteksi dini dan klasifikasi penyakit mata sangat penting untuk mencegah komplikasi lebih lanjut dan meningkatkan kualitas hidup pasien, deteksi dini untuk mengklasifikasikan penyakit yang diderita dapat dilakukan melalui pengambilan citra fundus (Yadav & Mallick, 2024). Citra fundus memungkinkan visualisasi langsung struktur penting di bagian belakang mata seperti pembuluh darah, optic disc, dan area makula, yang merupakan lokasi utama berbagai penyakit mata seperti *Diabetik Retinopati*, *Glaucoma*, dan *Cataract* (Indraswari dkk., 2022). Namun, proses diagnosis manual dari citra fundus mata membutuhkan waktu, biaya, serta keahlian khusus, dan rentan terhadap subjektivitas antar dokter.

Seiring berkembangnya teknologi kecerdasan buatan, khususnya *Artificial Neural Network* (ANN) dengan arsitektur *Multi-Layer Perceptron* (MLP), beberapa

studi telah menunjukkan bahwa ANN mampu memproses fitur-fitur numerik hasil ekstraksi dari citra medis dan memberikan hasil klasifikasi yang akurat dengan akurasi rata-rata 93,87% untuk klasifikasi fitur numerik hasil ekstraksi citra medis, MLP unggul dalam fleksibilitas pemrosesan data tabular dan memiliki hasil yang konsisten (Molina & Parraga-Alava, 2024). MLP lebih efisien dan membutuhkan daya komputasi yang lebih ringan untuk diterapkan dalam mengolah fitur numerik hasil ekstraksi manual jika dibandingkan dengan CNN yang unggul untuk mengolah citra mentah namun membutuhkan daya komputasi yang tinggi (Zhang dkk., 2023). ANN (MLP) dapat dilatih untuk mengenali pola tekstur, morfologi, dan karakteristik statistik pada citra fundus, sehingga dapat membantu proses diagnosis otomatis dan mengurangi beban kerja tenaga medis.

Beberapa penelitian sebelumnya telah berhasil menerapkan ANN (MLP) untuk klasifikasi berbagai penyakit dari citra medis, seperti deteksi pendarahan pada klasifikasi retinopati diabetik dari citra fundus mencapai akurasi 97,9% dalam membedakan citra normal dan citra penyakit (Bhardwaj dkk., 2020), perbandingan CNN dan MLP untuk klasifikasi citra X-Ray paru menunjukkan bahwa MLP lebih efektif untuk fitur numerik hasil ekstraksi dengan akurasi 95,59% pada 15 epoch serta dapat menjadi alternatif pada perangkat yang memiliki keterbatasan komputasi (Wulandari & Fitriana, 2021). Dengan demikian, ANN (MLP) berpotensi besar untuk diadaptasi dalam klasifikasi *multi-label*, untuk mendukung deteksi otomatis berbagai penyakit mata berbasis citra fundus.

Namun, karena MLP tidak dirancang untuk mengenali pola spasial secara langsung, performanya sangat bergantung pada kualitas fitur masukan. Oleh karena

itu, penting untuk melakukan proses ekstraksi fitur manual terlebih dahulu sebagai bagian dari proses klasifikasi agar informasi visual dari citra yang diubah menjadi data numerik dapat terbaca oleh mesin (Costaner & Harun, 2024). Fitur-fitur yang biasanya diekstrak dari citra fundus meliputi fitur statistik (seperti *mean*, *standar deviasi*, *skewness*, *kurtosis*), fitur tekstur (seperti *Local Binary Pattern* dan *Gray-Level Co-occurrence Matrix*), serta fitur frekuensi spasial (seperti hasil transformasi wavelet dan filter Gabor) (Pathan dkk., 2023). Dengan ekstraksi fitur tekstur menggunakan *Gray-Level Co-occurrence Matrix* (GLCM), terutama pada area seperti optic disc, pembuluh darah, dan lesi patologis yang merupakan indikator penting dalam diagnosis penyakit mata. Perubahan tekstur sering kali menjadi tanda awal dari gangguan seperti *Glaucoma* atau *Diabetic Retinopathy*, yang tidak dapat ditangkap dengan baik oleh fitur statistik biasa, serta dilanjutkan dengan Filter Gabor yang mengekstraksi informasi frekuensi dan orientasi yang membantu dalam mendeteksi perubahan halus pada struktur retina. Kombinasi keduanya menghasilkan representasi fitur yang lebih informatif, sehingga diharapkan dapat meningkatkan efektivitas model MLP dalam mengklasifikasikan citra fundus *multi-label* (Yudono dkk., 2025).

Kemudian, untuk memastikan model tidak hanya belajar dari data pelatihan, tetapi juga mampu menggeneralisasi terhadap data baru serta menghindari overfitting atau underfitting perlu dilakukan evaluasi model. Evaluasi menggunakan kurva loss dan metrik performa seperti akurasi, sensitivitas, dan confusion matrix penting untuk menilai keandalan model dalam konteks nyata. Validasi eksternal dengan dataset berbeda juga diperlukan untuk mengukur

ketahanan model terhadap perubahan distribusi data. Tanpa evaluasi yang komprehensif, model dapat terlihat baik secara statistik, namun gagal saat diterapkan dalam praktik (Faghani dkk., 2022). Sehingga penting juga untuk melakukan evaluasi secara menyeluruh terhadap performa model ANN (MLP) ini guna memastikan keakuratan, konsistensi, dan kemampuan model dalam mendeteksi data baru, yang mana dapat dilakukan melalui hasil dari matrik evaluasi (Zai dkk., 2025). Sehubungan dengan tujuan dari penelitian ini, yaitu menganalisis performa *Artificial Neural Network* (ANN), khususnya *Multi-Layer Perceptron* (MLP), dalam mengklasifikasikan citra fundus mata setelah dilakukan preprocessing dan ekstraksi fitur, sehingga dapat diketahui nilai efektivitas dan keandalan model dalam mengklasifikasikan penyakit mata multi-label.

Upaya mendeteksi penyakit mata juga merupakan salah satu wujud dari mensyukuri anugerah yang telah diberikan oleh Allah. Dalam ajaran Islam, kisah Nabi Ayyub mengajarkan tentang kesabaran dan ikhtiar dalam menjalani ujian berupa penyakit, sebagaimana telah dijelaskan dalam Al-Anbiya ayat 83 yang berbunyi:

وَأَيُّوبَ إِذْ نَادَىٰ رَبَّهُ أَنِّي مَسَّنِيَ الضُّرُّ وَأَنْتَ أَرْحَمُ الرَّحِيمِينَ ﴿٨٣﴾

“(Ingatlah) Ayyub ketika dia berdoa kepada Tuhannya, “(Ya Tuhanku,) sesungguhnya aku telah ditimpa penyakit, padahal Engkau Tuhan Yang Maha Penyayang dari semua yang penyayang.” (QS Al-Anbiya: 83)

Menurut penjelasan Hidayat dkk., (2024), kisah Nabi Ayyub menggambarkan bagaimana seorang hamba yang saleh menghadapi ujian penyakit dengan penuh kesabaran dan keimanan kepada Allah. Dalam penafsiran Ibn ‘Ashur

terhadap QS. Al-Anbiya: 83–84, doa Nabi Ayyub menunjukkan adab seorang hamba ketika berdoa tidak mengeluh, tetapi menyampaikan penderitaan dengan penuh ketundukan dan harapan kepada rahmat Allah. Sementara itu, tafsir Al-Maraghi menambahkan bahwa perintah Allah kepada Nabi Ayyub untuk menghentakkan kakinya hingga keluar air pada QS Shad: 42 mengajarkan bahwa kesembuhan dapat diperoleh melalui sebab fisik yang Allah sediakan, bukan semata-mata mukjizat gaib. Dengan demikian, kisah Nabi Ayyub mengandung pelajaran tentang pentingnya kesabaran, doa, dan ikhtiar dalam menghadapi penyakit, serta keyakinan bahwa setiap penderitaan memiliki jalan kesembuhan melalui rahmat dan izin Allah.

Di sisi lain Rasulullah SAW juga menegaskan mengenai pentingnya menjaga kesehatan tubuh dan berikhtiar untuk mencari pengobatan, sebagaimana hadits yang diriwayatkan oleh Abu Daud:

إِنَّ اللَّهَ أَنْزَلَ الدَّاءَ وَالْأَدْوِيَّةَ، وَجَعَلَ لِكُلِّ دَاءٍ دَوَاءً، فَتَدَاوُوا وَلَا تَتَدَاوُوا بِحَرَامٍ

“Sesungguhnya Allah telah menurunkan penyakit dan penawarannya, dan Dia menjadikan untuk setiap penyakit penawar; maka berobatlah kalian, tetapi jangan menggunakan sesuatu yang haram” (HR. Abu Dawud, Mishkat al-Masabih No 4538).

Hadist ini menegaskan bahwa Islam mendorong umatnya untuk berikhtiar mencari kesembuhan melalui pengobatan yang halal dan sesuai dengan syariat. Rasulullah SAW mengajarkan bahwa setiap penyakit memiliki obatnya, sehingga manusia dituntut untuk berusaha menemukan dan memanfaatkannya. Prinsip ini sejalan dengan nilai-nilai yang diajarkan dalam kisah Nabi Ayyub, bahwa

kesembuhan diperoleh melalui perpaduan antara doa, kesabaran, dan usaha menggunakan sarana yang Allah sediakan.

Dalam konteks perkembangan teknologi modern, nilai ikhtiar ini juga mencakup penggunaan teknologi dan inovasi medis seperti penerapan kecerdasan buatan dalam diagnosis dan deteksi dini penyakit. Selama teknologi tersebut digunakan untuk kemaslahatan dan tidak bertentangan dengan prinsip syariah, maka hal ini merupakan bentuk nyata dari pengamalan ajaran Islam untuk menjaga kesehatan tubuh sebagai amanah sekaligus wujud rasa syukur atas nikmat penglihatan yang telah Allah karuniakan.

1.2 Pernyataan Masalah

Berapakah performa *Multi-Layer Perceptron* (MLP), dalam mengklasifikasikan penyakit mata berdasarkan dataset ODIR?

1.3 Batasan Masalah

Penelitian ini dibatasi pada penggunaan dataset *Ocular Disease Intelligent Recognition* (ODIR) yang diperoleh dari Kaggle. Data ini merupakan data sekunder yang berisi 5.000 data pasien dengan citra fundus mata kiri dan kanan. Data ini dikumpulkan oleh Shanggong Medical Technology Co., Ltd. dari berbagai rumah sakit/pusat medis di Tiongkok. Label penyakit yang digunakan adalah *Normal*, *Diabetic*, *Glaucoma*, *Cataract*.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk mengukur performa model *Multi-Layer Perceptron (MLP)*, dalam mengklasifikasikan penyakit mata berdasarkan dataset ODIR-5K.

1.5 Manfaat Penelitian

Penelitian diharapkan memberikan manfaat bagi tenaga medis melalui pengembangan model berbasis citra fundus untuk mendukung proses deteksi dini penyakit mata. Model ini diharapkan dapat mempercepat proses analisis, meningkatkan objektivitas penilaian pada citra fundus dan mendukung pengambilan keputusan klinis tanpa menggantikan peran utama dokter mata dalam proses diagnosis.

BAB II

STUDI PUSTAKA

2.1 Sistem Klasifikasi Penyakit Mata

Klasifikasi penyakit adalah proses pengelompokan kondisi kesehatan atau gangguan medis ke dalam kategori tertentu berdasarkan ciri, gejala, atau data medis yang relevan (Jutel, 2011). Tujuannya adalah untuk membantu diagnosis, pengobatan, pemantauan, penelitian epidemiologi, serta memungkinkan perbandingan antar kelompok pasien yang serupa (Wadmann, 2023). Dalam konteks komputasi, berbagai pendekatan modern seperti *machine learning* dan *deep learning* telah banyak digunakan untuk membantu mendeteksi serta mengklasifikasikan penyakit mata dengan memanfaatkan data klinis (misalnya usia, riwayat kesehatan, tekanan darah), data citra medis (seperti fundus retina, OCT), maupun data gejala pasien.

Penelitian oleh Aziz dkk., (2025), menggunakan metode kombinasi *multifractal analysis* untuk ekstraksi fitur dan *Multi-Layer Perceptron (MLP)* sebagai model klasifikasi. Data yang digunakan berupa citra *Optical Coherence Tomography (OCT)* retina, dengan preprocessing meliputi *Gaussian blur*, *CLAHE*, *thresholding*, dan *binarisasi*. Fitur yang diekstrak mencakup dimensi *multifractal* (Db , Di , Dc) serta parameter dari *singularity spectrum* ($amin$, $amax$, $acenter$, $f(\alpha)max$, *spectrum width*, dan *symmetry shift*). Variabel input berupa fitur-fitur numerik hasil ekstraksi citra OCT, sedangkan output berupa klasifikasi citra normal

dan penderita *diabetic retinopathy*. Hasil penelitian menunjukkan bahwa MLP memberikan performa terbaik dengan akurasi 98,02%, precision 98,24%, recall 97,80%, dan F1-score 98,01%, mengungguli metode lain seperti Random Forest (94,45%), Decision Tree (91,92%), SVM (91,52%), Gradient Boosting (94,97%), XGBoost (95,27%), dan LightGBM (95,58%). Temuan ini menunjukkan bahwa kombinasi *multifractal analysis* dengan *MLP* mampu mendeteksi *diabetic retinopathy* secara efektif dengan tingkat akurasi yang sangat tinggi.

Penelitian yang dilakukan oleh Yudono dkk., (2025) membandingkan performa beberapa metode *Artificial Neural Network* (ANN), yaitu *Backpropagation Neural Network* (BPNN), *Probabilistic Neural Network* (PNN), dan *Radial Basis Function Neural Network* (RBFNN), dalam mengklasifikasikan tingkat keparahan *Cataract* berdasarkan citra optic disc fundus. Data yang digunakan berasal dari citra fundus *smartphone* yang melalui tahap *preprocessing* (*resize, konversi HSV, CLAHE, dan smoothing morfologi*), segmentasi optic disc, serta ekstraksi fitur menggunakan GLCM (*ASM, contrast, correlation, IDM, entropy*) dan Filter Gabor (*mean, variance, entropy*). Hasil penelitian menunjukkan bahwa metode PNN dengan fitur GLCM memberikan performa terbaik dengan akurasi pelatihan mencapai 100% dan akurasi pengujian 89,29%, lebih unggul dibandingkan kombinasi metode ANN lainnya.

Penelitian oleh Chavan & Pete (2024), metode yang digunakan adalah *Multilevel Glowworm Swarm Convolutional Neural Network* (MGSCNN) dengan optimasi hyperparameter CNN menggunakan *Glowworm Swarm Optimizer* (GSO). Data yang digunakan adalah *RFMiD* (*Retinal Fundus Multi-Disease dataset*) yang

terdiri dari 3.200 citra fundus dengan 39 jenis penyakit retina. Variabel input berupa citra fundus hasil pre-processing (normalisasi, *smoothing*, *resizing*), sedangkan output berupa klasifikasi multi-penyakit retina, misalnya *Diabetic Retinopathy (DR)*, *Age-Related Macular Degeneration (ARMD)*, *Optic Disc Cupping (ODC)*, *Central Retinal Vein Occlusion (CRVO)*, dan *Branch Retinal Vein Occlusion (BRVO)*. Hasil penelitian menunjukkan bahwa MGSCNN mencapai akurasi 95,09%, sensitivitas 96,59%, spesifisitas 93,59%, presisi 93,53%, recall 96,67%, dan F1-score 95,07%, lebih baik dibanding CNN standar (90,35%), SVM (89,42%), maupun LSTM (69,62%). Temuan ini membuktikan bahwa optimasi hyperparameter menggunakan GSO dapat meningkatkan performa CNN dalam klasifikasi multi-penyakit retina.

Penelitian oleh Yadav & Mallick (2024) menggunakan metode *EfficientNet-B3* dengan pendekatan *transfer learning* untuk klasifikasi citra fundus ke dalam empat kategori, yaitu Normal, *Cataract*, *Diabetic retinopathy*, dan *Glaucoma*. Proses penelitian meliputi preprocessing citra berupa resize ke ukuran 300x300 piksel, normalisasi, serta augmentasi data, kemudian dilakukan *fine-tuning* dan pelatihan dengan *dropout*, *batch size* 40, dan 40 epoch. Variabel input berupa citra fundus hasil preprocessing, sedangkan variabel output berupa performa klasifikasi berdasarkan akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model mencapai akurasi keseluruhan 96%, dengan performa tertinggi pada klasifikasi *diabetic retinopathy* (precision 0,99; recall 1,00; F1-score 1,00), diikuti *Cataract* (F1-score 0,98), *Glaucoma* (F1-score 0,94), dan Normal (F1-score 0,93).

Temuan ini membuktikan bahwa *EfficientNet-B3* mampu menjadi model untuk deteksi dini berbagai penyakit mata berbasis citra fundus.

Penelitian oleh Mansour & Al-Marghilnai (2021), menggunakan metode yang mengkombinasikan *Convolutional Neural Network (CNN)* dengan arsitektur *Multi-Layer Perceptron (MLP)* berbasis perceptron. Proses penelitian diawali dengan segmentasi *optic disc (OD)* dan *optic cup (OC)* menggunakan CNN berbasis *entropy estimation*, kemudian dilakukan ekstraksi fitur menggunakan *Weighted Least Square (WLS)* untuk memperoleh fitur holistik, termasuk penghitungan *Cup-to-Disc Ratio (CDR)* yang menjadi indikator penting dalam deteksi *Glaucoma*. Variabel input berupa citra fundus dari dataset Drishti-GS dan RIM-ONE, sedangkan output berupa klasifikasi normal atau *Glaucoma*. Hasil penelitian menunjukkan performa sangat tinggi, dengan nilai Dice 0,99 dan Jaccard 0,97 pada segmentasi OC di dataset Drishti-GS, serta akurasi segmentasi OD sebesar 0,99. Pada klasifikasi, sensitivitas mencapai 100% dengan spesifisitas 97,8% dan AUC 0,971 pada Drishti-GS, serta AUC 0,991 (OD) dan 0,974 (OC) pada dataset RIM-ONE. Temuan ini membuktikan bahwa integrasi CNN dan MLP berbasis perceptron dengan ekstraksi fitur WLS mampu meningkatkan akurasi deteksi *Glaucoma* secara signifikan dibanding metode sebelumnya.

Tabel 2. 1 Penelitian Terkait Klasifikasi Penyakit Mata

Peneliti	Judul Penelitian	Metode	Hasil	Perbedaan Penelitian
(Aziz dkk., 2025)	<i>Early Diabetic Retinopathy Detection from OCT Images Using Multifractal Analysis and Multi-Layer</i>	<i>Multifractal Analysis</i> pada tahap ekstraksi fitur dan <i>Multi-Layer Perceptron</i> untuk tahap klasifikasi	Akurasi yang dihasilkan dari kombinasi dari <i>Multifractal Analysis</i> dan metode <i>MLP</i> yaitu 98,02%	Penelitian Aziz dkk. (2025) memakai citra OCT dan <i>multifractal analysis</i> untuk mendeteksi <i>diabetic retinopathy</i> (dua kelas), sedangkan penelitian ini memakai citra fundus dengan fitur GLCM dan

	<i>Perceptron Classification</i>			Gabor untuk empat kelas penyakit mata.
(Yudono dkk., 2025)	Klasifikasi Katarak Berdasarkan Optic Disc Citra Fundus Smartphone: Perbandingan Ekstraksi Ciri Tekstur dan Metode <i>Neural Network</i>	<i>Probabilistic Neural Network, Backpropagation Neural Network, Radial Basis Function Neural Network</i>	GLCM + PNN (100%), GLCM + BPNN (96%), GLCM + RFBNN (<96%) PNN lebih unggul dibanding metode lainnya.	Perbedaannya terletak pada cakupan penyakit, sumber data, dan arsitektur jaringan yang digunakan. Penelitian oleh Yudono et al. lebih aplikatif untuk mendeteksi <i>Cataract</i> berbasis <i>smartphone</i> serta membandingkan BPNN, PNN, dan RFBNN untuk mendapatkan hasil terbaik.
(Chavan & Pete, 2024)	<i>Automatic multi-disease classification on retinal images using multilevel glowworm swarm convolutional neural network</i>	<i>Multilevel Glowworm Swarm Optimization Convolutional Neural network (MGSCNN)</i>	Akurasi MGSCNN mencapai 95,09%, lebih baik dibanding CNN yaitu 90,35%	Penelitian oleh Chavan & Pete (2024) menggabungkan CNN dan <i>Glowworm Swarm Optimization</i> untuk ekstraksi dan optimasi fitur otomatis dengan akurasi 95.09%, sedangkan penelitian ini menggunakan GLCM dan Gabor secara manual dengan MLP.
(Yadav & Mallick, 2024)	<i>Early Detection Of Cataract, Diabetic Retinopathy, And Glaucoma Using Deep Learning</i>	<i>Deep Learning</i> dengan arsitektur <i>EfficientNet-B3</i>	Model <i>EfficientNet-B3</i> mencapai akurasi 96% pada keseluruhan kelas	Perbedaannya, penelitian oleh Yadav & Mallick (2024) menggunakan CNN <i>EfficientNet-B3</i> dengan akurasi 96%, sedangkan penelitian ini menggunakan ANN-MLP dengan fitur GLCM dan Gabor, yang lebih sederhana.
(Mansour & Al-Marghilnai, 2021)	<i>Glaucoma detection using novel perceptron based convolutional multi-layer neural network classification</i>	<i>Convolutional Neural Network (CNN)</i> dengan arsitektur <i>Multi-Layer Perceptron (MLP)</i> , dan menggunakan <i>Weighted Least Square (WLS)</i> untuk ekstraksi fitur	<i>CNN+MLP</i> berbasis perceptron dengan <i>WLS</i> unggul dalam segmentasi dan klasifikasi <i>glaukoma</i> dengan sensitivitas tinggi	Penelitian oleh Mansour & Al-Marghilnai (2021) menggabungkan CNN dan MLP (PC-MLNN) dengan fitur WLS dan segmentasi entropi, menghasilkan akurasi tinggi ($\approx 99.7\%$) khusus untuk <i>Glaukoma</i> , sedangkan penelitian ini menggunakan ANN-MLP dengan fitur GLCM dan Gabor untuk klasifikasi multi-penyakit.

Tabel 2.1 memaparkan ringkasan penelitian terkait penyakit mata yang menggunakan berbagai metode *machine learning* dan *deep learning* dalam klasifikasi penyakit mata. Tabel ini menampilkan peneliti, judul penelitian, metode yang digunakan, hasil dari masing-masing penelitian dan perbedaan dengan penelitian yang akan dilakukan. Dalam tabel ini terdapat beberapa penelitian yang mengaplikasikan berbagai metode dalam mendeteksi dan mengklasifikasikan berbagai penyakit mata berdasarkan data medis berupa citra fundus dan citra OCT. Informasi pada tabel ini memberikan gambaran mengenai efektivitas metode yang digunakan dalam konteks mengklasifikasikan penyakit mata, serta nilai akurasi yang dihasilkan oleh masing-masing penelitian.

Berdasarkan hasil perbandingan tersebut, sebagian besar penelitian tersebut menggunakan model *deep learning* seperti CNN dan EfficientNet yang membutuhkan sumber daya besar, sedangkan penelitian berbasis *Artificial Neural Network* (ANN) seperti BPNN, PNN, dan RBFNN umumnya hanya berfokus pada satu jenis penyakit dan belum mengintegrasikan fitur tekstur serta spasial. Selain itu, penelitian MLP sebelumnya menggunakan citra OCT, bukan citra fundus yang lebih mudah diakses. Oleh karena itu, penelitian ini mengisi celah tersebut dengan mengembangkan model *Multi-Layer Perceptron* (MLP) dengan fitur GLCM dan Gabor dari citra fundus untuk meng-klasifikasikan penyakit mata.

2.2 Karakteristik Penyakit Mata

Karakteristik penyakit mata dapat diamati melalui citra fundus karena setiap kelainan pada retina menimbulkan perubahan visual yang khas pada berbagai struktur anatomis. Struktur utama yang sering dianalisis adalah pembuluh darah

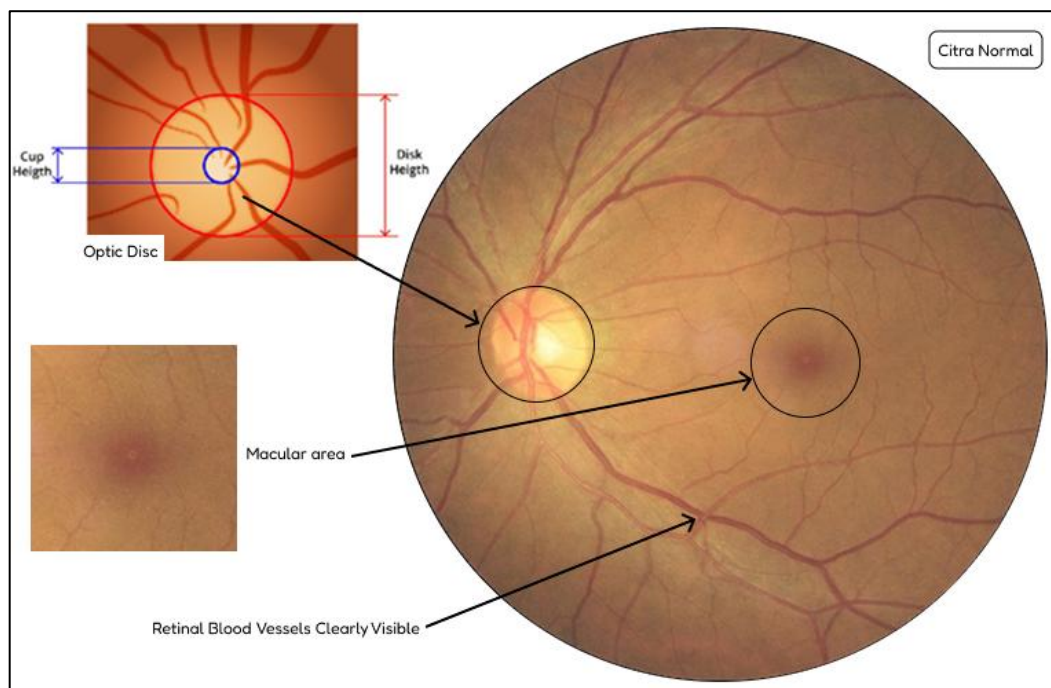
retina, yang berfungsi menyalurkan oksigen dan nutrisi ke jaringan retina. Perubahan pada pembuluh darah, seperti penyempitan, pelebaran, perdarahan, atau pola percabangan yang tidak normal, menjadi indikator penting gangguan vaskular retina, terutama pada diabetic retinopathy. Selain itu, makula, sebagai area pusat retina yang bertanggung jawab terhadap penglihatan sentral dan ketajaman visual, juga dapat mengalami perubahan intensitas dan tekstur akibat kondisi patologis.

Struktur penting lainnya adalah *optic disc*, yaitu area keluarnya serabut saraf retina menuju saraf optik yang tampak sebagai area terang pada citra fundus. Perubahan bentuk dan rasio *cup-to-disc* pada *optic disc* sering digunakan sebagai indikator utama dalam diagnosis glaukoma. Pada penyakit retina akibat diabetes, sering ditemukan *microaneurysm*, pelebaran kecil pada kapiler retina yang tampak sebagai titik merah halus dan menjadi tanda awal *diabetic retinopathy*. Seiring perkembangan penyakit, dapat muncul *exudates*, endapan lipid atau protein akibat kebocoran pembuluh darah, yang terlihat sebagai bercak terang berwarna kuning atau putih pada citra fundus (Costaner & Harun, 2024).

Pada tahap lanjut karakteristik penyakit pada retina, dapat terjadi *neovascularization*, pertumbuhan pembuluh darah baru yang abnormal sebagai respons terhadap iskemia retina (kondisi ketika retina kekurangan oksigen), yang tampak sebagai jaringan pembuluh darah halus dan tidak teratur. Selain itu, perubahan juga dapat terjadi pada *Foveal Avascular Zone (FAZ)*, yaitu area di pusat fovea yang secara normal tidak mengandung pembuluh darah, di mana perubahan ukuran atau bentuknya mencerminkan gangguan mikrosirkulasi retina. Struktur penting lainnya adalah *retinal nerve fiber layer (RNFL)*, yaitu lapisan serabut saraf

retina yang membawa sinyal visual ke otak, penipisan atau kerusakan RNFL menjadi indikator utama kerusakan saraf optik pada glaucoma (Riina dkk., 2024).

Dengan demikian, citra fundus memungkinkan proses identifikasi penyakit mata dilakukan secara non-invasif melalui analisis perubahan warna, tekstur, kontras, dan bentuk anatomi akibat kondisi patologis. Citra fundus memungkinkan proses identifikasi penyakit dilakukan secara *non-invasif* dengan memanfaatkan perubahan warna, tekstur, kontras, dan bentuk anatomi yang muncul akibat kondisi patologis tertentu (Horie & Ohno-Matsui, 2022). Melalui perbedaan visual ini, berbagai penyakit seperti *Diabetic Retinopathy*, *Glaucoma*, dan *Cataract* dapat dikenali berdasarkan lesi khas, perubahan struktural, maupun degradasi kualitas citra yang ditimbulkannya.

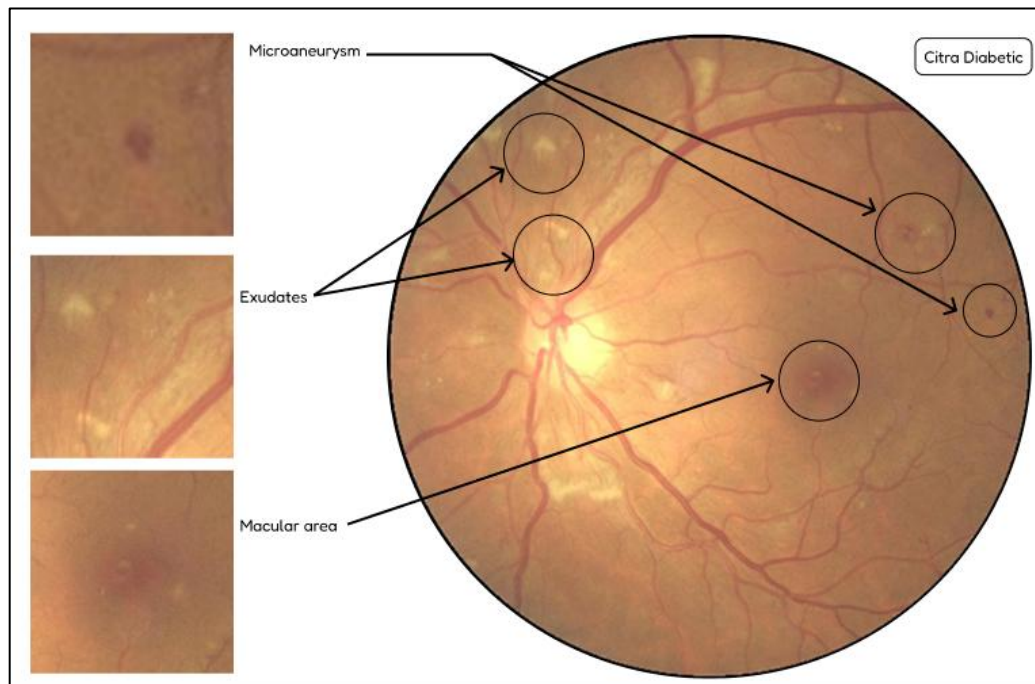


Gambar 2. 1 Citra Normal

Citra mata normal pada Gambar 2.1 memperlihatkan pembuluh darah yang tampak jelas, warna retina yang seragam, serta *optic disc* dengan batas tegas dengan

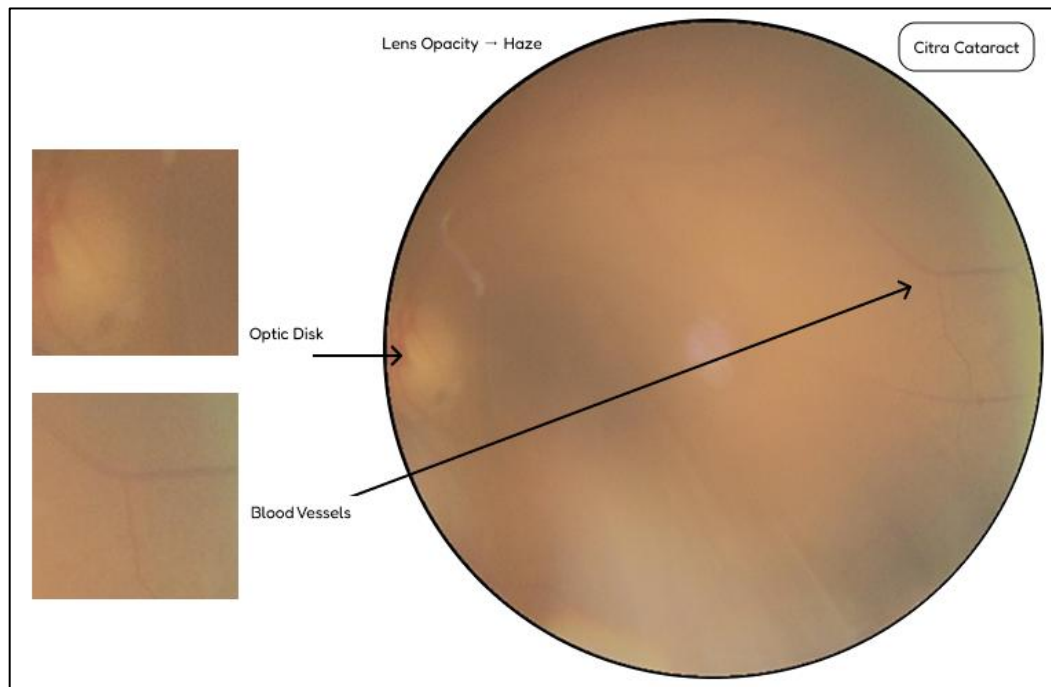
warna kuning cerah sebagai ciri khas kondisi retina sehat. Pada kondisi normal, permukaan retina terlihat jernih dengan pencahayaan yang baik, sehingga struktur anatomis seperti pembuluh darah, makula, dan fovea dapat diamati secara detail (World Health Organization, 2019). *Optic disc* tampak berwarna kuning pucat dengan batas yang tegas, *cup-to-disc* ratio berada pada rentang normal tanpa pelebaran atau pendalaman cup, menunjukkan tidak adanya kerusakan saraf optik. Pembuluh darah retina tampak jelas dan terdistribusi secara teratur dari *optic disc* menuju *perifer* retina dengan ketebalan yang konsisten. Makula terlihat lebih gelap dibanding area retina lainnya, sementara fovea tampak sebagai area kecil dengan pigmentasi lebih pekat yang menjadi pusat ketajaman visual. Secara keseluruhan, citra mata normal memperlihatkan pola tekstur yang halus, kontras yang stabil, serta tidak ditemukan microaneurysm, perdarahan, eksudat, kekeruhan lensa, ataupun defek serabut saraf—yang umumnya menjadi indikator patologis pada berbagai penyakit mata (Yadav & Mallick, 2024).

Kejelasan struktur pada citra fundus normal yang ditandai oleh warna retina homogen, pembuluh darah dengan pola distribusi yang teratur, serta *optic disc* dengan batas tegas dan cup yang proporsional menjadi acuan dalam evaluasi kesehatan mata. Tampilan ini memudahkan identifikasi dini terhadap perubahan patologis pada kondisi seperti *diabetic retinopathy*, *glaucoma*, maupun *cataract*.



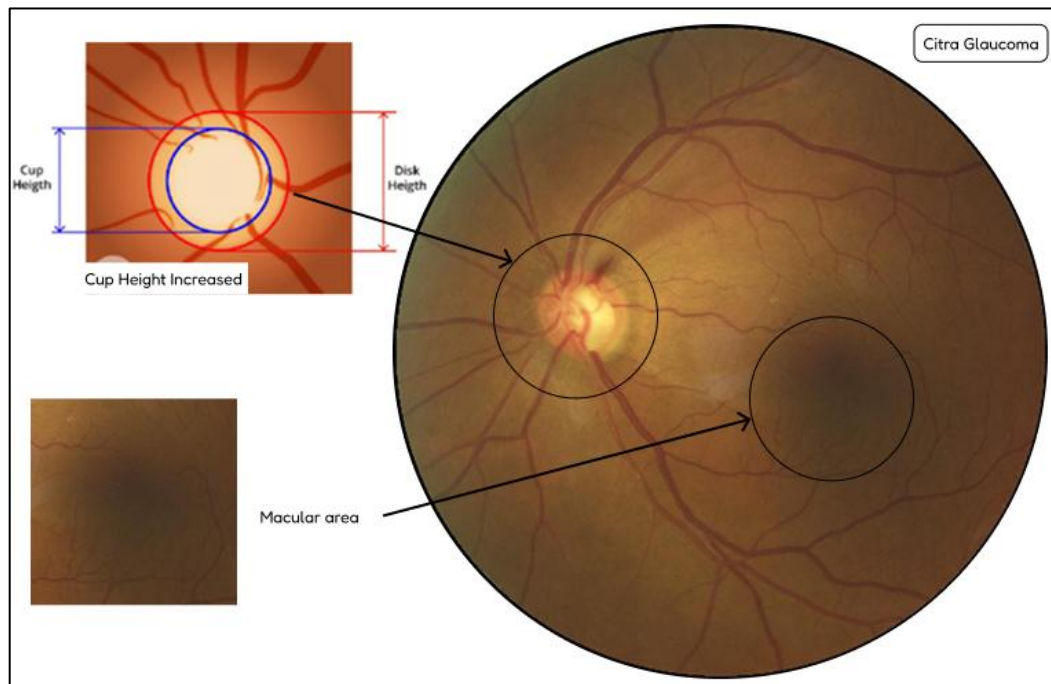
Gambar 2. 2 Citra Diabetic

Citra *Diabetic Retinopathy* (DR) pada Gambar 2.2 menunjukkan lesi khas akibat kerusakan pembuluh darah retina. Pada tahap awal tampak *microaneurysm* (MA) sebagai titik merah kecil, diikuti oleh *exudates*. Pada tahap lanjut dapat muncul *neovascularization* (NV). Perubahan pada *Foveal Avascular Zone* (FAZ), seperti pelebaran dan ketidakstabilan struktur, juga menjadi indikator awal DR (Ikram dkk., 2024). Secara keseluruhan, kombinasi *microaneurysm*, *exudates*, perubahan FAZ pada Gambar 2.2 menjadi karakteristik DR pada citra fundus. Lesi-lesi ini menimbulkan perubahan signifikan pada tekstur, kontras, dan distribusi intensitas, sehingga dapat dikenali melalui teknik ekstraksi fitur berbasis tekstur maupun frekuensi.



Gambar 2. 3 Citra Cataract

Citra *Cataract* pada Gambar 2.3 menampilkan kabut keputihan pada citra fundus karena kekeruhan lensa yang mengaburkan detail retina. Citra fundus penderita *cataract* tampak kabur, berkabut (*haze*), dan mengalami penurunan kontras yang signifikan. Kekeruhan lensa menyebabkan hamburan cahaya (*light scattering*) sehingga detail retina, termasuk pembuluh darah, makula, dan batas *optic disc*, menjadi tidak jelas (World Health Organization, 2019). Efek haze menyeluruh inilah yang menjadi ciri khas fundus *cataract*. Karena gangguan terjadi sebelum cahaya mencapai retina, citra fundus tidak menampilkan pola atau lesi tertentu, melainkan penurunan kualitas visual secara global. Karakteristik ini dapat terdeteksi melalui fitur tekstur yang menangkap tingkat kabur, rendahnya kontras, serta hilangnya detail struktural pada citra.



Gambar 2. 4 Citra Glaucoma

Citra *Glaucoma* pada Gambar 2.4 menunjukkan pelebaran dan pendalaman *optic cup* serta penipisan serat saraf retina di sekitar optic disc (World Health Organization, 2019). Citra glaukoma ditampilkan dengan “*cupping*” yang semakin menonjol dan disk optik yang tampak lebih cekung dibanding retina normal, perubahan ini terjadi akibat kerusakan progresif pada serabut saraf retina, di mana penipisan *retinal nerve fiber layer* (RNFL thinning) menyebabkan hilangnya jaringan di sekitar disk optik, sehingga cup tampak lebih besar dan batasnya semakin dalam. Selain perubahan tersebut, glaukoma juga memunculkan defek pada serabut saraf retina yang tampak sebagai area gelap atau pola seperti busur (*arc-shaped defect*) pada fundus. Defek ini terjadi akibat hilangnya berkas serabut saraf yang menuju ke optic disc, sehingga menyebabkan penurunan reflektivitas retina pada jalur RNFL (Pathan dkk., 2023).

Secara keseluruhan, karakteristik utama *glaucoma* pada citra fundus meliputi, peningkatan cup-to-disc ratio, pendalaman cup, penipisan RNFL, defek serabut saraf, serta perubahan arah pembuluh darah di sekitar disk optik. Perubahan-perubahan ini bersifat struktural, bukan berupa lesi, ciri ini menjadi dasar bagi metode deteksi glaucoma berbasis fitur tekstur dan bentuk pada citra fundus.

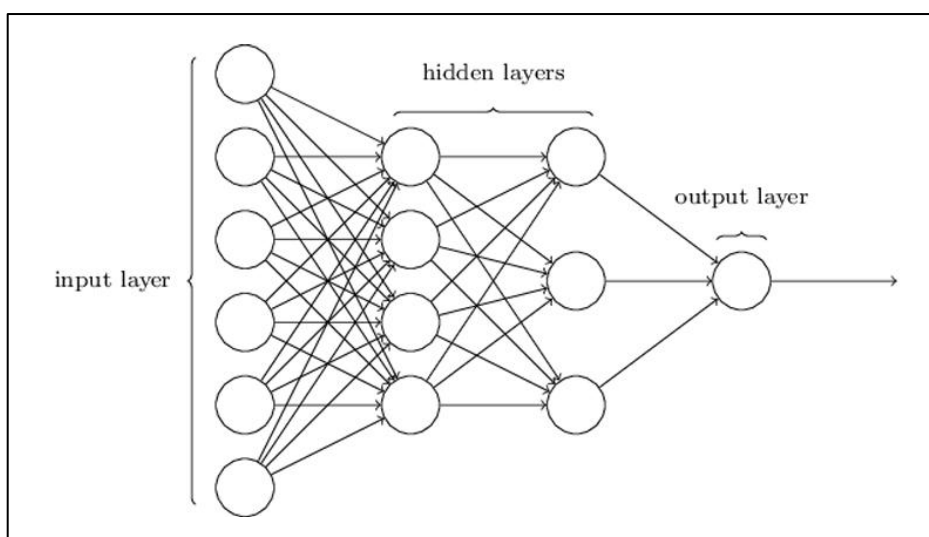
Beberapa penelitian menunjukkan bahwa model berbasis kecerdasan buatan mampu mencapai performa setara dengan dokter spesialis dalam mendiagnosis penyakit mata. Misalnya, Butt dkk., (2022) mengembangkan metode *hybrid deep learning features* untuk mendeteksi *diabetic retinopati* dari citra fundus yang menghasilkan akurasi mencapai 96,57%. Penelitian lain oleh Indraswari dkk., (2022) mengembangkan metode berbasis *Convolutional Neural Network (CNN)* dengan arsitektur MobileNetV2 sebagai base model untuk mendeteksi penyakit mata pada citra fundus, dengan akurasi mencapai 72%.

2.3 Multi-Layer Perceptron (MLP)

Multi-Layer Perceptron (MLP) merupakan salah satu arsitektur dari *Artificial Neural Network (ANN)* yang tersusun atas beberapa lapisan neuron, yaitu lapisan *input*, satu atau lebih lapisan tersembunyi (*hidden layer*), dan lapisan *output*. Setiap neuron pada MLP terhubung secara penuh dengan neuron pada lapisan berikutnya dan memiliki bobot yang akan diperbarui selama proses pelatihan. Pembaruan bobot ini dilakukan menggunakan algoritma backpropagation untuk meminimalkan kesalahan antara hasil prediksi dan target yang sebenarnya (Saputra dkk., 2017).

MLP memiliki kemampuan untuk mempelajari hubungan non-linear yang kompleks di antara variabel input dan output. Melalui lapisan tersembunyinya, MLP mampu mengekstraksi representasi fitur yang lebih dalam dan abstrak dari data yang diberikan, sehingga sangat efektif digunakan dalam berbagai kasus klasifikasi dan prediksi berbasis citra. Dalam konteks medis, kemampuan ini memungkinkan MLP untuk mengenali pola halus pada data citra, seperti tekstur atau struktur jaringan yang tidak mudah diidentifikasi oleh metode konvensional (Lee dkk., 2017).

Dalam bidang analisis citra medis, terutama pada citra fundus mata, MLP terbukti mampu mengklasifikasikan kondisi retina berdasarkan ciri-ciri tekstur dan struktur yang diekstraksi dari gambar. Misalnya, MLP dapat dilatih untuk mendeteksi adanya kelainan retina atau indikasi penyakit mata dengan menggunakan fitur yang merepresentasikan pola-pola tertentu pada citra fundus (Aziz dkk., 2025). Contoh arsitektur MLP dapat dilihat pada Gambar 2.5



Gambar 2. 5 Arsitektur Multi-Layer Perceptron
Sumber: (Heriz dkk., 2018)

1. *Input Layer*, Lapisan yang menerima fitur numerik hasil ekstraksi dari citra fundus, misalnya tekstur atau pola pembuluh darah. Data ini kemudian diproses lebih lanjut ke hidden layer.
2. *Hidden Layer*, Lapisan yang berisi neuron yang melakukan penjumlahan berbobot dengan bias yang dihitung menggunakan persamaan 2.1,

$$z_j^{(1)} = \sum_{i=1}^n w_{ij}^{(1)} x_i + b_j^{(1)} \quad (2.1)$$

kemudian hasilnya diproses melalui fungsi aktivasi non-linear seperti sigmoid atau ReLU dengan persamaan 2.2.

$$a_j = f(z_j) = \frac{1}{1+e^{-z_j}} \quad \text{atau} \quad a_j = \max(0, z_j) \quad (2.2)$$

Fungsi aktivasi ini memungkinkan jaringan mengenali pola kompleks dalam data.

3. *Output Layer*, Lapisan terakhir yang menghasilkan prediksi akhir berupa probabilitas kelas menggunakan fungsi softmax yang dihitung dengan persamaan 2.3,

$$y_k = \frac{e^{z_k}}{\sum_{j=1}^K e^{z_j}}, \quad k = 1, 2, \dots, K \quad (2.3)$$

sehingga model dapat memutuskan kategori penyakit tertentu berdasarkan citra fundus.

Proses *feedforward* merupakan tahapan saat data input dipropagasi maju melalui jaringan dari *input layer* menuju *output layer*. Setiap neuron menghitung penjumlahan berbobot dari input yang diterimanya, menambahkan bias, dan kemudian memproses hasil tersebut melalui fungsi aktivasi non-linear. Hasil dari

seluruh neuron pada layer sebelumnya menjadi input bagi layer berikutnya hingga akhirnya dihasilkan output prediksi.

Setelah *feedforward*, model dilatih menggunakan *backpropagation* untuk meminimalkan *error* antara prediksi dan label sebenarnya. Nilai *error* dihitung menggunakan fungsi *loss categorical cross-entropy*, dan gradiennya dihitung untuk setiap bobot jaringan kemudian digunakan untuk memperbarui bobot secara iteratif hingga fungsi loss mencapai nilai minimum dan performa model menjadi optimal. Kinerja MLP juga dipengaruhi oleh hyperparameter seperti jumlah hidden layer, jumlah neuron, learning rate, batch size, dan jumlah epoch. Pemilihan hyperparameter yang tepat sangat penting agar model mencapai akurasi optimal dan tidak mengalami overfitting (Goodfellow dkk., 2016).

Penelitian oleh Riina dkk., (2024) menunjukkan bahwa MLP mampu mendiagnosis *Glaucoma* dengan akurasi tinggi menggunakan biomarker struktural mata. Sementara, pada penelitian oleh Aziz dkk., (2025) menunjukkan bahwa MLP dapat mendeteksi retinopati diabetik dari citra OCT dengan akurasi 98,02%, lebih unggul dibandingkan metode lain. Temuan ini menunjukkan potensi MLP dalam klasifikasi citra fundus untuk deteksi dini berbagai penyakit mata.

BAB III

DESAIN DAN IMPLEMENTASI

3.1 Pengambilan Data

Data yang digunakan merupakan data sekunder *Ocular Disease Recognition/ODIR-5k* yang diperoleh melalui website resmi *Kaggle*. Dataset ini dikumpulkan oleh Shanggong Medical Technology Co, Ltd dari berbagai rumah sakit dan pusat medis di China menggunakan kamera Canon, Zeiss, dan Kowa, yang menghasilkan resolusi gambar berukuran 512 x 512 piksel. Anotasi citra dilakukan oleh pembaca ahli dengan manajemen kontrol kualitas, dataset ini mencakup 5.000 data pasien.

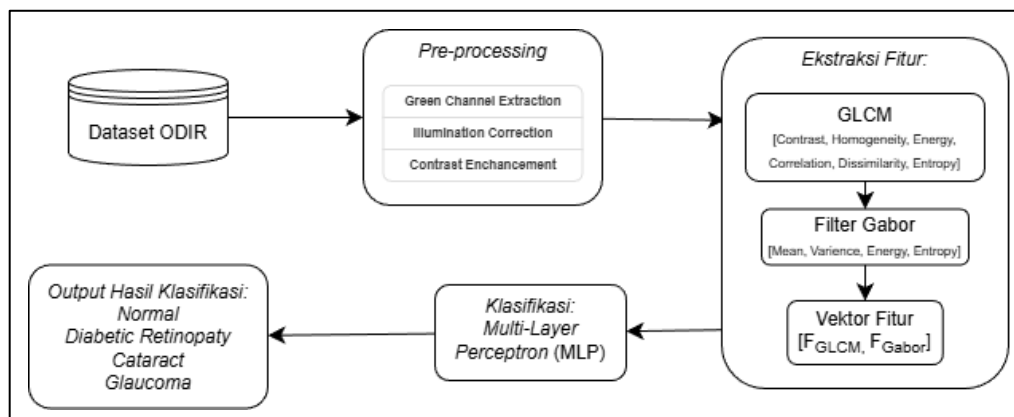
Atribut yang digunakan dalam penelitian ini ditunjukkan pada Tabel 3.1, yang terdiri dari *filename* sebagai nama file citra fundus dan label sebagai penanda kelas penyakit. Label kelas yang digunakan meliputi ['N'] untuk fundus mata normal, ['D'] untuk fundus mata penderita *diabetic retinopathy*, ['G'] untuk fundus mata penderita *glaucoma*, dan ['C'] untuk fundus mata penderita *cataract*. Setiap kelas penyakit tersebut memiliki karakteristik visual yang berbeda pada citra fundus, seperti yang telah dibahas pada Bab II, meliputi perbedaan tekstur, struktur anatomi, serta kualitas visual citra. Perbedaan karakteristik inilah yang selanjutnya dimanfaatkan dalam proses ekstraksi fitur tekstur dan fitur frekuensi spasial sebagai dasar klasifikasi.

Tabel 3. 1 Penjelasan Atribut

Atribut	Penjelasan
filename	Nama file dari citra fundus
labels	Label kelas penyakit dari citra fundus

3.2 Desain Sistem

Desain sistem yang digunakan dalam penelitian ini pada Gambar 3.1 terdiri dari 3 tahapan utama, yaitu tahap *pre-processing* dataset dilakukan dengan tujuan meningkatkan kualitas citra sebelum diproses pada tahap ekstraksi fitur. Setelah dilakukan *pre-processing* tahapan selanjutnya adalah ekstraksi fitur menggunakan metode GLCM untuk fitur tekstur dan filter gabor untuk fitur frekuensi spasial yang akan digunakan sebagai input ke model MLP.



Gambar 3. 1 Desain Sistem

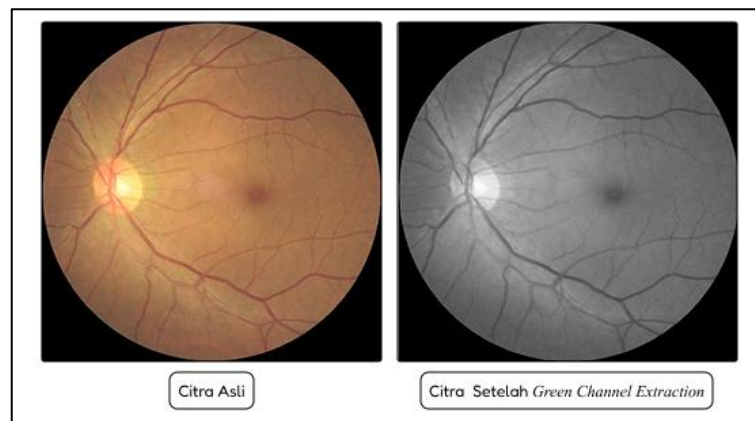
Tahapan klasifikasi menggunakan model MLP kemudian dirancang dengan menentukan jumlah layer, neuron, dan fungsi aktivasi, sebelum dilakukan proses training model menggunakan *feedforward* dan *backpropagation*. Setelah pelatihan selesai, performa model dievaluasi dengan metrik akurasi, presisi, recall, F1-score, dan *confusion matrix*.

3.2.1 Pre-processing

Tahap *pre-processing* dataset dilakukan dengan tujuan meningkatkan kualitas citra sebelum diproses pada tahap ekstraksi fitur. *Pre-processing* dilakukan dalam beberapa tahapan meliputi:

1. *Green Channel Extraction*

Green Channel Extraction dilakukan untuk menonjolkan struktur pembuluh darah yang ada pada retina, kanal hijau dipilih karena memiliki kontras paling tinggi terhadap struktur retina, dibandingkan kanal merah atau biru (Kaya dkk., 2020). Citra RGB diekstraksi dengan mengambil kanal hijau secara langsung dari matriks citra. Kanal hijau yang dihasilkan kemudian diproses melalui normalisasi intensitas menggunakan metode *min-max scaling* agar nilai piksel berada pada rentang 0–255 dan dikonversi ke format uint8. Proses normalisasi ini bertujuan untuk menyeragamkan skala intensitas citra sehingga lebih stabil dan konsisten untuk digunakan pada tahapan selanjutnya. Gambar 3.2 merupakan contoh citra sebelum dan sesudah dilakukan *Green Channel Extraction*.

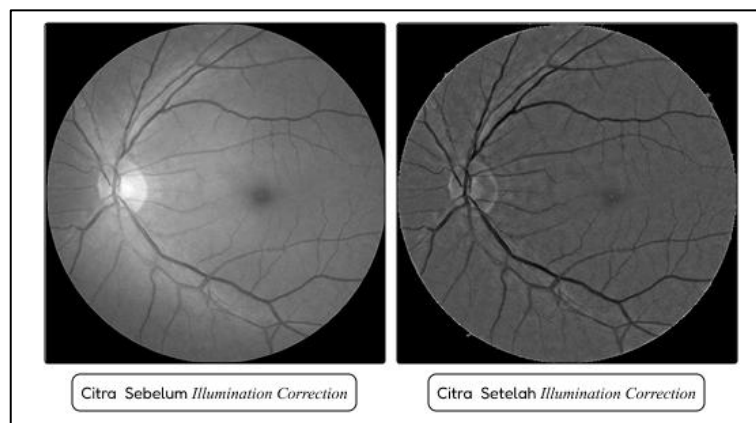


Gambar 3. 2 Penerapan *Green Channel Extraction*

2. *Illumination Correction*

Illumination Correction dilakukan untuk meratakan pencahayaan (*vignetting*) pada citra sehingga intensitas retina menjadi lebih seragam (Saha dkk., 2018). Proses *correction* hanya diterapkan pada area *Field of View (FOV)*

dengan memanfaatkan mask retina, sehingga area di luar retina tidak memengaruhi estimasi pencahayaan. sementara pencahayaan latar belakang diestimasi menggunakan *median filtering* berukuran besar. Selanjutnya, *illumination correction* dilakukan dengan metode *difference-based* dengan mengurangkan citra asli terhadap citra latar belakang, kemudian hasilnya dinormalisasi secara robust sebelum dikonversi ke format citra 8-bit. Gambar 3.3 merupakan contoh citra sebelum dan sesudah dilakukan *Illumination Correction*.

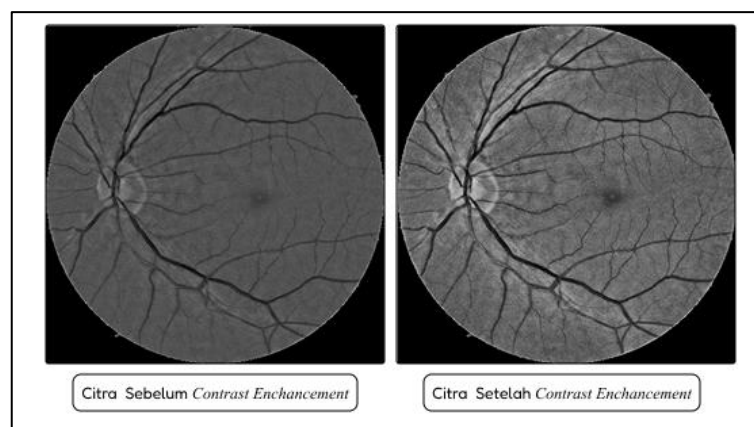


Gambar 3. 3 Penerapan *Illumination Correction*

3. *Contrast Enhancement*

Contrast Enhancement dilakukan dengan meningkatkan kontras citra untuk memperjelas detail halus pada jaringan retina tanpa memperbesar noise. Citra hasil *illumination correction* melalui proses peredaman noise ringan menggunakan *Gaussian filtering* untuk mengurangi fluktuasi intensitas skala kecil terlebih dahulu. Kemudian, peningkatan kontras dilakukan menggunakan metode *Contrast Limited Adaptive Histogram Equalization (CLAHE)* dengan membagi citra ke dalam beberapa tile berukuran tertentu dan membatasi

peningkatan kontras melalui parameter clip limit. Pendekatan ini memungkinkan peningkatan kontras yang adaptif dan terkendali pada setiap wilayah citra. Proses enhancement hanya diterapkan pada area retina dengan memanfaatkan *mask Field of View (FOV)*, sehingga area di luar retina tidak terpengaruh. Gambar 3.4 merupakan contoh citra sebelum dan sesudah dilakukan *Contrast Enhancement*.



Gambar 3. 4 Penerapan *Contrast Enhancement*

3.2.2 Ekstraksi Fitur

Tahap ekstraksi fitur dilakukan dengan tujuan mengubah citra fundus hasil *pre-processing* menjadi representasi numerik yang dapat dipahami oleh model *Multi-Layer Perceptron* (MLP). Proses ini diperlukan karena MLP tidak secara langsung mengenali pola spasial dari citra, sehingga dibutuhkan transformasi informasi visual ke dalam bentuk vektor fitur. Ekstraksi fitur dilakukan dalam beberapa tahapan meliputi:

a. GLCM

Tahap ekstraksi fitur tekstur menggunakan GLCM dilakukan dengan tujuan merepresentasikan variasi intensitas piksel dalam citra, seperti tingkat kekasaran

atau homogenitas pola retina. Matriks GLCM dibentuk dengan menghitung frekuensi kemunculan pasangan intensitas piksel (i, j) pada jarak dan arah tertentu. Misalnya, dari matriks 4 x 4 dengan jarak $d = 1$ dan arah $\theta = 0^\circ$, maka akan diperoleh matriks probabilitas sebagai berikut.

Matriks 4 x 4	Matriks Probabilitas
$I_{4 \times 4} = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$	$P = \begin{bmatrix} 0 & \frac{1}{12} \\ 0 & \frac{11}{12} \end{bmatrix}$

Gambar 3. 5 Contoh matriks GLCM probabilitas

Gambar 3.5 menunjukkan konversi dari matriks citra 4 x4 menjadi matriks probabilitas menggunakan *Gray Level Co-occurrence Matrix (GLCM)*. Pada contoh tersebut, citra dengan ukuran 4 x 4 memiliki dua tingkat keabuan, yaitu 0 (latar belakang) dan 1 (objek). Matriks probabilitas diperoleh dengan menghitung frekuensi kemunculan pasangan intensitas piksel (i, j) . Dari hasil perhitungan diperoleh bahwa pasangan (0,1) muncul sebanyak 1 kali, dan pasangan (1, 1) muncul sebanyak 11 kali, maka jumlah total pasangan piksel adalah 12. Sehingga, setelah membagi masing-masing frekuensi dengan total pasangan dan menghasilkan matriks probabilitas yang selanjutnya digunakan untuk menghitung fitur tekstur GLCM sebagai berikut:

1. Menghitung fitur *contrast* untuk mengukur variasi lokal intensitas dengan memberikan bobot kuadrat pada selisih level keabuan antara pasangan piksel menggunakan persamaan 3.1

$$Contrast = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i - j)^2 P(i, j) \quad (3.1)$$

Apabila nilai contrast relatif kecil, maka matriks tersebut memiliki sedikit variasi intensitas lokal pada arah yang diperiksa sehingga teksturnya relatif homogen pada skala arah tersebut.

2. Menghitung *homogeneity* untuk mengukur seberapa dekat massa probabilitas GLCM berkumpul di sekitar diagonal utama semakin besar nilainya, semakin seragam / halus tekstur pada arah yang diperiksa.

$$Homogeneity = \sum_{i=0}^1 \sum_{j=0}^1 \frac{P(i, j)}{1 + |i - j|} \quad (3.2)$$

Perhitungan *Homogeneity* dihitung menggunakan persamaan 3.2, apabila nilai yang dihasilkan mendekati 1 maka menandakan sedikit variasi antar-piksel pada arah tersebut.

3. Menghitung *energy (ASM)* untuk mengukur tingkat keteraturan atau keseragaman distribusi probabilitas dalam matriks GLCM. Semakin tinggi nilai ASM, semakin teratur (kurang acak) tekstur citra.

$$Energy = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (P(i, j))^2 \quad (3.3)$$

Perhitungan *energy* dihitung menggunakan persamaan 3.3, fitur *ASM/Energy* memberikan indikasi tekstur yang relatif homogen dan dominan pada nilai intensitas tertentu di arah yang diuji.

4. Menghitung *correlation* untuk mengukur hubungan linear antara intensitas piksel baris (i) dan kolom (j) dalam distribusi probabilitas. *Correlation* memberikan gambaran seberapa kuat nilai intensitas piksel dalam citra saling berkaitan pada arah dan jarak tertentu. Perhitungan *correlation* dihitung menggunakan persamaan 3.4,

$$Correlation = \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i-\mu)(j-\mu)P(i,j)}{\sigma^2} \quad (3.4)$$

μ_i = rata-rata baris
 μ_j = rata-rata kolom
 σ_i = standar deviasi baris
 σ_j = standar deviasi kolom

5. Menghitung *dissimilarity* untuk mengukur tingkat perbedaan rata-rata antara pasangan piksel. Nilai ini memberikan gambaran seberapa besar jarak antara nilai intensitas piksel baris(i) dan kolom (j) dalam matriks. Perhitungan *dissimilarity* dihitung menggunakan persamaan 3.5. Semakin tinggi nilai *dissimilarity*, semakin besar perbedaan intensitas antar pasangan piksel.

$$Dissimilarity = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |i - j| P(i, j) \quad (3.5)$$

6. Menghitung *entropy* untuk mengukur tingkat ketidakpastian atau kompleksitas tekstur dalam citra. Nilai *entropy* yang tinggi menunjukkan distribusi piksel yang lebih acak (kompleks), sedangkan nilai rendah menunjukkan tekstur yang lebih homogen atau teratur. Perhitungan *dissimilarity* dihitung menggunakan persamaan 3.6,

$$Entropy = - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i, j) \log P(i, j) \quad (3.6)$$

b. Gabor Filter

Tahap ekstraksi fitur spasial menggunakan filter Gabor dilakukan dengan tujuan untuk mengekstrak informasi tekstur citra dalam domain spasial–frekuensi, khususnya yang berkaitan dengan pola frekuensi dan orientasi. Filter Gabor efektif dalam merepresentasikan karakteristik tekstur karena mampu menangkap variasi

struktur lokal pada berbagai arah dan skala, sehingga banyak digunakan dalam analisis citra medis.

Secara umum, fungsi kernel Gabor dua dimensi didefinisikan sebagai kombinasi fungsi Gaussian dan gelombang sinusoidal, yang dapat dirumuskan pada persamaan 3.7,

$$g(x, y) = \exp \left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2} \right) \cdot \cos \left(2\pi \frac{x'}{\lambda} + \psi \right) \quad (3.7)$$

Dengan transformasi koordinat pada persamaan 3.8 dan 3.9:

$$x' = x \cos \theta + y \sin \theta \quad (3.8)$$

$$y' = -x \sin \theta + y \cos \theta \quad (3.9)$$

σ = lebar fungsi Gaussian

λ = panjang gelombang sinusoidal

θ = orientasi filter

ψ = fase

γ = rasio aspek

Kernel Gabor yang dibentuk kemudian dikonvolusikan dengan citra hasil pra-pemrosesan untuk memperoleh respons filter. Proses konvolusi ini bertujuan untuk menonjolkan pola tekstur tertentu yang sesuai dengan parameter frekuensi dan orientasi kernel yang digunakan. Secara matematis, proses konvolusi dapat dilihat pada persamaan 3.10,

$$S(x, y) = \sum_i \sum_j I(x - i, y - j) g(i, j) \quad (3.10)$$

$I(x, y)$ = citra masukan

$S(x, y)$ = respons hasil konvolusi

Matriks S menunjukkan hasil dari proses konvolusi antara citra input dengan kernel Gabor menghasilkan matriks respons. Setiap elemen pada matriks diperoleh dengan cara menggeser kernel Gabor ke setiap posisi yang memungkinkan pada

matriks input, kemudian menghitung hasil perkalian elemen-sejajar dan menjumlahkannya. Setelah diperoleh matriks respons gabor tersebut, kemudian dihitung *statistic* spasialnya, yaitu:

1. Menghitung *mean* untuk memberi nilai rata-rata kekuatan respons filter pada area retina menggunakan persamaan 3.11,

$$\mu_R = \frac{1}{M} \sum_{x,y} R(x,y) \quad (3.11)$$

2. Menghitung *variance* untuk mengukur penyebaran (variasi) respons di antara piksel; standard deviation adalah akar kuadratnya. *Variance* dihitung menggunakan persamaan 3.12,

$$\sigma^2 = \frac{1}{N} \sum_{k=1}^N (R_k - \mu)^2 \quad (3.12)$$

3. Menghitung *energy* untuk memberi nilai rata-rata kekuatan respons filter pada area retina menggunakan persamaan 3.13,

$$Energy_R = \sum_{x,y} R(x,y)^2 \quad (3.13)$$

4. Menghitung *entropy* menggunakan magnitudo (*absolute value*) dari respons dan merubahnya menjadi distribusi probabilitas dengan normalisasi terhadap jumlah magnitudo seluruh piksel. *Entropy* dihitung menggunakan persamaan 3.14,

$$H_R = - \sum_k p_k \log_2 p_k \quad (3.14)$$

c. Vektor Fitur

Tahap vektor fitur dilakukan setelah proses ekstraksi fitur menggunakan metode GLCM dan Gabor Filter dilakukan. Pada tahap GLCM, terdapat enam parameter yaitu *contrast*, *homogeneity*, *energy*, *correlation*, *dissimilarity*, dan *entropy*. Sedangkan pada tahap ekstraksi fitur dengan gabor filter, terdapat empat

statistik spasial yaitu, *mean*, *variance*, *energy*, dan *entropy*. Berikut penggabungannya menjadi vektor fitur menggunakan persamaan 3.15,

$$F = [F_{GLCM}, F_{Gabor}] \quad (3.15)$$

$$F_{GLCM} = [Contrast, Homogeneity, Energy, Correlation, Dissimilarity, Entropy]$$

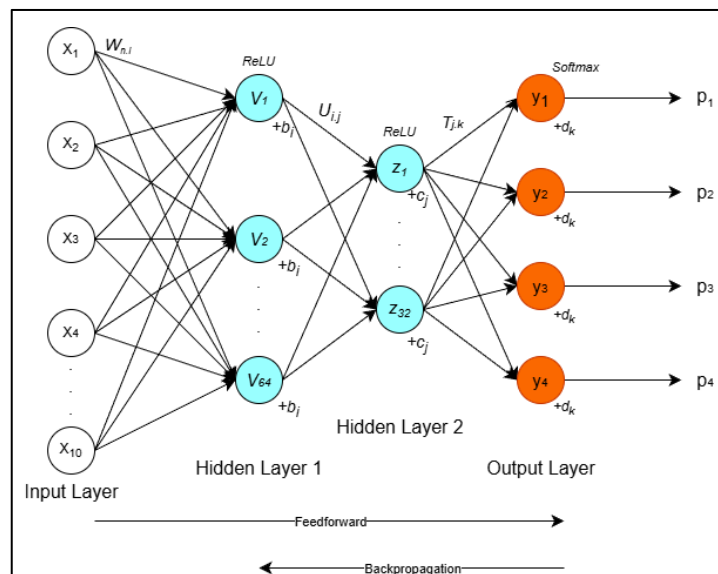
$$F_{Gabor} = [Mean, Variance, Energy, Entropy]$$

$$F = \begin{bmatrix} Contrast, Homogeneity, Energy, Correlation, Dissimilarity, \\ Entropy, Mean, Variance, Energy, Entropy \end{bmatrix}$$

Setelah fitur-fitur tersebut disatukan dalam bentuk numerik, vektor hasil ekstraksi digunakan sebagai masukan (*input layer*) pada model *Multi-Layer Perceptron (MLP)* untuk melakukan klasifikasi penyakit mata.

3.2.3 Desain MLP

Tahapan implementasi *Multi-Layer Perceptron (MLP)* untuk melakukan klasifikasi penyakit mata terdiri dari empat lapisan utama, yaitu *input layer*, dua *hidden layer*, dan *output layer*.



Gambar 3. 6 Arsitektur Model MLP

Contoh arsitektur jaringan *Multi-Layer Perceptron (MLP)* pada Gambar 3.6 dengan penjelasan tiap-tiap layer sebagai berikut:

1. *Input layer*, terdiri dari 10 neuron yang masing-masing merepresentasikan fitur numerik hasil ekstraksi citra fundus, terdiri atas 6 fitur dari GLCM (*contrast, homogeneity, energy, correlation, dissimilarity, entropy*) dan 4 fitur dari Gabor (*mean response, variance, energy, entropy*). Setiap neuron pada input layer merepresentasikan satu fitur, tanpa melakukan transformasi, dan kemudian meneruskan nilainya ke hidden layer.

$$X = \begin{bmatrix} \text{Contrast, Homogeneity, Energy, Correlation, Dissimilarity,} \\ \text{Entropy, Mean, Variance, Energy, Entropy} \end{bmatrix}$$

2. *Hidden layer*, menerima data dari input layer, kemudian memprosesnya melalui operasi linier (penjumlahan berbobot) yang ditambah dengan bias, lalu dilewatkan ke fungsi aktivasi non-linear agar jaringan mampu mengenali pola kompleks pada citra fundus.
 - a. Pada *hidden layer* pertama, setiap neuron menerima seluruh nilai masukan dari sepuluh fitur yang telah diekstraksi dari citra fundus. Setiap nilai input (X_n) dikalikan dengan bobot ($W_{n.i}$) yang menghubungkan neuron input ke neuron hidden, kemudian hasilnya dijumlahkan dan ditambahkan dengan nilai bias (b_i) untuk membentuk nilai net input (V_i). Nilai ini selanjutnya diproses menggunakan fungsi aktivasi *Rectified Linear Unit (ReLU)*, di mana hanya nilai positif yang diteruskan, sedangkan nilai negatif akan dipetakan menjadi nol. Dengan mekanisme ini, hanya neuron yang relevan terhadap pola tertentu pada citra fundus yang akan “aktif”, sementara neuron lainnya akan non-

aktif. Perhitungan neuron pada *hidden layer* pertama sebelum aktivasi dihitung menggunakan persamaan 3.16, kemudian hasilnya diproses melalui fungsi aktivasi non-linear (ReLU) dengan persamaan 3.17.

$$V_i = \sum_{n=1} W_{n,i} \cdot X_n + b_i \quad (3.16)$$

$$V_i = \max(0, V_i) \quad (3.17)$$

- b. *Hidden layer* kedua menerima masukan dari *hidden layer* pertama dalam bentuk representasi yang telah disaring oleh fungsi aktivasi. Pada tahap ini, neuron-neuron kembali melakukan operasi linier berupa penjumlahan berbobot $U_{i,j}$ dan keluaran *hidden layer* pertama V_i , dimana indeks i merupakan jumlah neuron pada *hidden layer* pertama, kemudian nilai tersebut ditambahkan dengan bias c_j , dan hasilnya disebut sebagai nilai net input Z_j , yang selanjutnya diproses menggunakan fungsi aktivasi ReLU. Perhitungan neuron pada *hidden layer* kedua sebelum aktivasi dihitung menggunakan persamaan 3.18, kemudian hasilnya diproses melalui fungsi aktivasi non-linear (ReLU) dengan persamaan 3.19.

$$Z_j = \sum_{i=1} U_{i,j} \cdot V_i + c_j \quad (3.18)$$

$$Z_j = \max(0, Z_j) \quad (3.19)$$

Dengan demikian, *hidden layer* kedua berfungsi untuk membentuk representasi yang lebih abstrak dari data, sehingga jaringan mampu mengenali pola-pola yang lebih kompleks.

3. *Output layer*, merupakan lapisan terakhir pada jaringan MLP yang bertugas menghasilkan prediksi akhir. Pada tahap ini, setiap neuron output menerima

masukan dari neuron-neuron di *hidden layer* kedua (Z_j), kemudian menghitung kombinasi linier bobot ($T_{j,k}$) dan bias (d_k). Jumlah neuron pada output layer ditentukan oleh banyaknya kelas target yang telah ditentukan, pada penelitian ini terdapat empat kategori penyakit mata yang diklasifikasikan, yaitu Normal, *Diabetic*, *Glaucoma*, dan *Cataract*. Sehingga output layer memiliki nilai $K = 4$ neuron. Persamaan 3.20 merupakan perhitungan untuk mendapatkan nilai neuron ke empat pada output layer,

$$y_k = \sum_{j=1} T_{j,k} \cdot Z_j + d_k \quad (3.20)$$

Nilai dari neuron pada lapisan output berada dalam rentang $[0,1]$ karena menggunakan fungsi aktivasi Softmax. Fungsi ini mengubah nilai keluaran dari setiap neuron (y_k) menjadi bentuk probabilitas ($\hat{y}_k$), kemudian fungsi Softmax digunakan untuk menormalisasi keseluruhan skor kelas (y_i) sehingga total seluruh probabilitas pada lapisan output berjumlah 1. Perhitungan softmax dilakukan dengan persamaan 3.21,

$$\hat{y}_k = \frac{e^{y_k}}{\sum_{j=1}^K e^{y_j}} \quad (3.21)$$

4. Setelah proses feedforward menghasilkan nilai aktivasi akhir pada output layer, dilakukan perhitungan *loss* dengan *categorical cross-entropy*, untuk menghitung kesalahan prediksi (*error*) dengan membandingkan hasil prediksi ($\hat{y}_k$) dengan label sebenarnya yang berbentuk *one-hot encoding* (t_i) yang dihitung menggunakan persamaan 3.22.

$$L = - \sum_{i=1}^K t_i \cdot \log(\hat{y}_k) \quad (3.22)$$

Nilai loss ini menunjukkan tingkat kesalahan prediksi model. Semakin kecil nilai loss, semakin baik performa model dalam menghasilkan prediksi yang sesuai dengan label sebenarnya.

5. *Backpropagation*, merupakan tahapan yang dilakukan setelah nilai *loss* diperoleh. Proses ini bertujuan untuk menyebarkan *error* dari *output layer* ke *hidden layer* sebelumnya dengan tujuan memperbaiki bobot bias pada setiap lapisan jaringan. *Backpropagation* menggunakan prinsip aturan rantai turunan untuk menghitung gradien error terhadap bobot. Error pada neuron output ($\delta_k^{(out)}$) dihitung berdasarkan selisih antara probabilitas hasil prediksi ($\hat{y}_k$) dan label sebenarnya (y_k), sebagaimana ditunjukkan pada persamaan 3.23,

$$\delta_k^{(out)} = \hat{y}_k - y_k \quad (3.23)$$

6. Pembaruan bobot dan bias, bertujuan untuk memperbarui bobot dan bias pada setiap lapisan jaringan, dilakukan berdasarkan nilai gradien *error* yang telah dihitung pada tahap *backpropagation*. Prinsip optimisasi adalah memperbarui bobot ke arah yang berlawanan dengan gradien, karena gradien menunjukkan arah peningkatan *error*. Pembaruan bobot dilakukan menggunakan persamaan 3.24 bobot ($W_{n,i}^{(t)}$) pada iterasi ke- t , dilakukan pembaruan yang menghasilkan bobot baru ($W_{n,i}^{(t+1)}$), dihitung menggunakan gradien fungsi loss terhadap bobot ($\frac{\partial L}{\partial W_{n,i}}$), sedangkan pembaruan bias dihitung menggunakan persamaan 3.25 bias ($b_i^{(t)}$) pada iterasi ke- t ,

dilakukan pembaruan yang menghasilkan bias baru $(b_i^{(t+1)})$ dihitung menggunakan gradien fungsi loss terhadap bias neuron ke- i ($\frac{\partial L}{\partial b_i}$).

$$W_{n,i}^{(t+1)} = W_{n,i}^{(t)} - \eta \cdot \frac{\partial L}{\partial W_{n,i}} \quad (3.24)$$

$$b_i^{(t+1)} = b_i^{(t)} - \eta \cdot \frac{\partial L}{\partial b_i} \quad (3.25)$$

Keterangan,
 η = *learning rate*

Proses pembaruan ini dilakukan secara iteratif untuk seluruh bobot dan bias pada setiap epoch, hingga model mencapai kondisi stabil dengan nilai *loss* yang minimum dan mampu menghasilkan klasifikasi citra fundus dengan akurasi yang optimal.

7. Dalam konteks klasifikasi multi-kelas, proses pengambilan keputusan dilakukan dengan memilih kelas yang memiliki nilai probabilitas tertinggi. Output layer menghasilkan vektor probabilitas kelas $P = [p_1, p_2, \dots, p_K]$, di mana p_k menyatakan probabilitas prediksi untuk kelas ke- k . Dengan demikian, hasil akhir klasifikasi ditentukan berdasarkan neuron dengan skor aktivasi terbesar, sebagaimana ditunjukkan pada persamaan 3.26,

$$P = \arg \max_k (\hat{y}_k) \quad (3.26)$$

Keterangan,
 P = Hasil kelas prediksi
 k = Indeks kelas
 $\hat{y}_k$ = Nilai probabilitas hasil softmax untuk kelas ke- k

BAB IV

UJI COBA DAN PEMBAHASAN

4.1 Persiapan Data

Pada penelitian ini, data yang digunakan berasal dari dataset ODIR-5k. Namun, data yang digunakan masih belum sepenuhnya valid/rusak. Sehingga perlu dilakukan *data cleaning* untuk menyaring kualitas citra dengan menghapus gambar yang bentuknya tidak sesuai dengan menyaring citra berdasarkan proporsionalitas bentuk fundus menggunakan pendekatan geometris, sehingga data yang digunakan pada proses pelatihan hanya berisi citra yang valid. Perubahan jumlah data sebelum dan sesudah proses *data cleaning* ditunjukkan pada Tabel 4.1.

Tabel 4. 1 Jumlah Data Pada Proses *Data Cleaning*

Kelas	Sebelum <i>Data Cleaning</i>	Setelah <i>Data Cleaning</i>
<i>Normal</i>	2.873	2.413
<i>Diabetic</i>	1.608	1.412
<i>Cataract</i>	293	225
<i>Glaucoma</i>	284	216
Total	5.058	4.266

Setelah *data cleaning* dilakukan, selanjutnya dilakukan augmentasi data untuk meningkatkan keragaman data latih dan memperkaya variasi pola visual pada citra fundus yang telah dinyatakan valid. Augmentasi yang digunakan yaitu rotasi (*rotating*) dan pembalikan citra (*flipping*). Penerapan augmentasi dilakukan dengan dua cara, yaitu augmentasi satu kali pada setiap citra untuk meningkatkan variasi visual, serta augmentasi tambahan pada kelas minoritas untuk menyamakan jumlah data antar kelas. Pendekatan bertujuan untuk memperkaya pola data latih sekaligus mengurangi ketidakseimbangan data. Hasil dari proses augmentasi ini ditunjukkan

melalui peningkatan jumlah data pada masing-masing kelas, sebagaimana ditunjukkan pada Tabel 4.2.

Tabel 4. 2 Jumlah Data pada Proses Augmentasi

Kelas	Jumlah Data		
	Setelah <i>Data Cleaning</i>	Augmentasi 1x	Augmentasi Tambahan
<i>Normal</i>	2.413	400	2.500
<i>Diabetic</i>	1.412	400	2.500
<i>Cataract</i>	225	400	2.500
<i>Glaucoma</i>	216	400	2.500
Total	4.266	1.600	10.000

4.2 Skenario Uji Coba

Pada penelitian ini, uji coba dilakukan pada dataset citra fundus menggunakan kelas *Normal*, *Cataract*, *Glaucoma* dan *Diabetic*. Skenario uji coba divariasikan menjadi beberapa model yaitu, variasi jumlah data dan variasi jumlah *neuron* pada *hidden layer*.

Variasi jumlah data pada skenario uji coba dilakukan dengan menggunakan beberapa konfigurasi dataset yang berbeda. Konfigurasi tersebut meliputi data hasil *data cleaning* tanpa penyeimbangan (*unbalanced*), data hasil *data cleaning* dengan augmentasi pada kelas yang memiliki jumlah data lebih sedikit, data hasil *data cleaning* yang kemudian dilakukan *undersampling* sejumlah kelas terkecil, serta data hasil *data cleaning* yang kemudian dilakukan *undersampling* dan augmentasi sebanyak satu kali pada tiap kelas.

Tabel 4. 3 Jumlah Data Tiap Model

Kelas	Jumlah Data			
	Model A (<i>Unbalance</i>)	Model B (Data Augmentasi)	Model C (<i>Undersampling</i>)	Model D (Data <i>Undersampling</i> + <i>Augmentasi</i>)
Normal	2.413	2.500	200	400
Diabetic	1.412	2.500	200	400
Cataract	225	2.500	200	400
Glaucoma	216	2.500	200	400
Total	4.266	10.000	800	1.600

Perbandingan jumlah data dari masing-masing kelas dapat dilihat pada Tabel 4.3. Variasi uji coba tersebut dirancang untuk mengevaluasi pengaruh distribusi dan jumlah data terhadap kinerja model MLP dalam melakukan klasifikasi citra fundus.

Selain variasi jumlah data, penelitian ini juga menerapkan variasi arsitektur jaringan untuk menguji performa model MLP dalam memproses fitur numerik dari citra fundus. Variasi tersebut mencakup perbedaan jumlah hidden layer dan jumlah neuron pada setiap layer, sehingga setiap skenario dapat merepresentasikan tingkat kompleksitas model yang berbeda. Kombinasi antara variasi jumlah data dan arsitektur jaringan tersebut menghasilkan sejumlah skenario uji coba yang dirancang untuk mengevaluasi pengaruh kedua faktor tersebut terhadap kemampuan klasifikasi. Untuk keseluruhan detail dari skenario uji coba dapat dilihat pada Tabel 4.4.

Tabel 4. 4 Variasi Arsitektur Jaringan

Skenario	Arsitektur MLP	Neuron
1	2 Hidden Layer	32-16
2		64-32
3	3 Hidden Layer	64-32-16
4		128-64-32

Berdasarkan variasi jumlah data dan arsitektur jaringan yang pada Tabel 4.4, dilakukan penggabungan kedua faktor tersebut untuk membentuk skenario pengujian yang lebih spesifik. Setiap skenario merepresentasikan kombinasi antara jenis dataset dan konfigurasi arsitektur MLP yang digunakan. Untuk memudahkan identifikasi dan pembahasan hasil pengujian, setiap skenario diberi kode penamaan tertentu yang mencerminkan konfigurasi yang digunakan. Sebagai contoh, kode A-

2H-32 menunjukkan bahwa skenario tersebut menggunakan dataset model A (*unbalanced*), dengan arsitektur MLP yang terdiri dari 2 hidden layer dan 32 neuron pada hidden layer pertama. Kombinasi skenario pengujian disajikan pada Tabel 4.5.

Tabel 4.5 Kombinasi skenario pengujian

No Skenario	Jumlah Data Skenario	Arsitektur MLP	Neuron	Skenario Pengujian
1	Model A (4.266 data)	2 Hidden Layer	32-16	A-2H-32
2			64-32	A-2H-64
3		3 Hidden Layer	64-32-16	A-3H-64
4			128-64-32	A-3H-128
5	Model B (10.000 data)	2 Hidden Layer	32-16	B-2H-32
6			64-32	B-2H-64
7		3 Hidden Layer	64-32-16	B-3H-64
8			128-64-32	B-3H-128
9	Model C (800 data)	2 Hidden Layer	32-16	C-2H-32
10			64-32	C-2H-64
11		3 Hidden Layer	64-32-16	C-3H-64
12			128-64-32	C-3H-128
13	Model D (1.600 data)	2 Hidden Layer	32-16	D-2H-32
14			64-32	D-2H-64
15		3 Hidden Layer	64-32-16	D-3H-64
16			128-64-32	D-3H-128

Pada setiap skenario uji coba, dataset dibagi menjadi tiga bagian, yaitu 80% untuk pelatihan, 10% untuk validasi, dan 10% untuk pengujian. Proses pelatihan pada model C dan D menggunakan jumlah *epoch* maksimum sebanyak 150 dan *epoch* maksimal 200 untuk model A dan B dengan menerapkan *early stopping* untuk mencegah *overfitting* apabila performa validasi tidak meningkat, dengan *learning rate* 0.0015. Parameter *batch size* ditetapkan berbeda sesuai kategori arsitektur, yaitu batch 16 pada model C dan D, serta batch 32 pada model A dan B.

Untuk mengukur performa *Multi-Layer Perceptron (MLP)* dalam mengklasifikasikan citra fundus dari dataset ODIR-5k pada empat kelas, yaitu Normal, *Diabetic*, *Glaucoma*, dan *Cataract*. Evaluasi dilakukan dengan menggunakan metrik performa berupa akurasi, presisi, recall, dan F1-score.

Evaluasi performa model klasifikasi melalui *confusion matrix*. *Confusion matrix* terdiri atas empat komponen utama: *True Positive* (TP), *False Positive* (FP), *False Negative* (FN), *True Negative* (TN).

1. *True Positive* (TP) mengacu pada jumlah data dari kelas positif yang diklasifikasikan dengan benar sebagai kelas positif.
2. *True Negative* (TN) adalah jumlah data dari kelas negatif yang diklasifikasikan dengan benar sebagai kelas negatif.
3. *False Positive* (FP) adalah data dari kelas negatif yang salah diklasifikasikan sebagai kelas positif.
4. *False Negative* (FN) merupakan data dari kelas positif yang salah diklasifikasikan sebagai kelas negatif yang bertujuan untuk menunjukkan perbandingan antara kelas sebenarnya dengan kelas hasil prediksi yang dilakukan oleh model.

Tabel 4. 6 *Confussion Matrix*

Kelas Sebenarnya	Kelas Prediksi			
	Normal	<i>Diabetic</i>	<i>Glaucoma</i>	<i>Cataract</i>
Normal	a_{11}	a_{12}	a_{13}	a_{14}
<i>Diabetic</i>	a_{21}	a_{22}	a_{23}	a_{24}
<i>Glaucoma</i>	a_{31}	a_{32}	a_{33}	a_{34}
<i>Cataract</i>	a_{41}	a_{42}	a_{43}	a_{44}

Confusion matrix pada Tabel 4.6 berukuran 4 x 4 karena terdapat empat kelas target, yaitu Normal, *Diabetic*, *Glaucoma*, dan *Cataract*. Setiap baris pada matriks merepresentasikan kelas sebenarnya, sedangkan setiap kolom merepresentasikan kelas hasil prediksi model. Nilai a_{ii} merepresentasikan jumlah data pada kelas i yang berhasil diprediksi dengan benar (TP untuk kelas tersebut), sementara nilai a_{ij} dengan $i \neq j$ menunjukkan jumlah data dari kelas i yang

diprediksi salah sebagai kelas j . Dimana diagonal utama merepresentasikan jumlah data yang berhasil diprediksi dengan benar, sedangkan nilai di luar diagonal menunjukkan kesalahan klasifikasi antar kelas.

Informasi pada *confussion matrix* menjadi dasar dalam perhitungan metrik evaluasi performa model seperti akurasi, presisi, recall, dan F1-score. Berikut persamaan dari masing-masing performa:

1. Presisi, digunakan untuk mengukur seberapa akurat model dalam memprediksi kelas positif. Presisi dihitung menggunakan persamaan 4.1.

$$Presisi (\%) = \frac{TP}{TP + FP} \times 100\% \quad (4.1)$$

2. *Recall*, digunakan untuk mengukur seberapa baik model dalam menemukan semua contoh positif yang sebenarnya. *Recall* dihitung menggunakan persamaan 4.2.

$$Recall (\%) = \frac{TP}{TP + FN} \times 100\% \quad (4.2)$$

3. *F1-score*, digunakan untuk menggabungkan presisi dan *recall* menjadi satu nilai, memberikan gambaran seimbang antara keduanya. *F1-score* dihitung menggunakan persamaan 4.3.

$$F1-Score = \frac{2 \times Presisi \times Recall}{Presisi + Recall} \quad (4.3)$$

4. Akurasi, digunakan untuk mengukur seberapa sering model membuat prediksi yang benar secara keseluruhan. Akurasi dihitung menggunakan persamaan 4.4.

$$Akurasi(\%) = \frac{TP_i}{N} \times 100\% \quad (4.4)$$

Penelitian ini juga menerapkan metode *k-fold cross validation* untuk mengevaluasi kinerja model neural network pada beragam variasi dan subset data.

Pada penerapannya, dataset dibagi menjadi k bagian yang disebut *fold*. Model dilatih sebanyak k kali, dengan setiap *fold* secara bergiliran dijadikan sebagai data pengujian, sementara *fold* lainnya digunakan sebagai data pelatihan. Kinerja model dari seluruh iterasi kemudian dihitung dan dirata-ratakan untuk memperoleh gambaran performa secara menyeluruh. Pada penelitian ini, nilai k yang digunakan adalah 5.

4.3 Hasil Uji Coba

Dataset yang telah melalui proses ekstraksi fitur selanjutnya diklasifikasikan menggunakan model *Multilayer Perceptron (MLP)* untuk membedakan empat kelas, yaitu Normal, *Diabetic*, *Glaucoma*, *Cataract*. Hasil dari beberapa skenario uji coba adalah sebagai berikut:

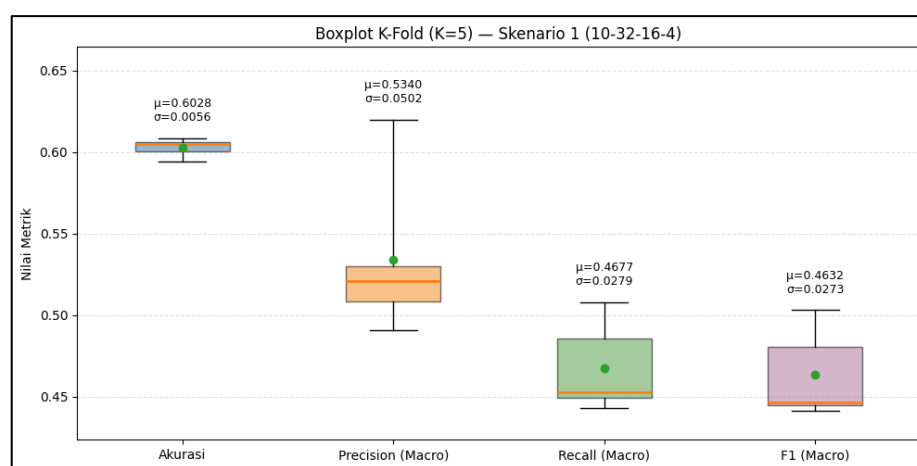
4.3.1 Skenario Uji Coba 1 (A-2H-32)

Pada skenario pertama, data yang digunakan terdiri dari 4.266 citra fundus yaitu 2.873 citra Normal, 1.608 citra *Diabetic*, 293 citra *Cataract*, dan 284 citra *Glaucoma*. Data tersebut dibagi ke dalam data *training*, data *validation* dan data *testing* dengan rasio 80:10:10. *Hidden layer* berjumlah 32 dan 16 *neuron* dengan 200 *epoch* dan *batch* sebesar 32.

Tabel 4. 7 *Evaluation Matrix K-Fold Skenario 1 (Model A)*

FOLD	Akurasi	Precision	Recall	F1-Score
1	60.02%	50.83%	45.25%	44.14%
2	59.44%	52.10%	44.92%	44.63%
3	60.84%	61.99%	50.81%	50.31%
4	60.49%	52.98%	48.54%	48.05%
5	60.61%	49.08%	44.32%	44.48%
Rata-rata	60.28%	53.40%	46.77%	46.32%
Standard Deviasi	0,0056	0.0502	0.0279	0,0273

Hasil evaluasi performa model pada Tabel 4.7 menggunakan *metode k-fold cross validation* dengan nilai $k = 5$. Evaluasi dilakukan dengan menghitung nilai akurasi, precision, recall dan f1-score pada setiap fold, kemudian dirata-ratakan untuk memperoleh gambaran kinerja model secara keseluruhan. Berdasarkan tabel tersebut, diperoleh rata-rata akurasi sebesar 60.28% dengan standar deviasi sebesar 0,0056.



Gambar 4. 1 Visualisasi Rata-rata *evaluation matrix* Skenario 1

Boxplot k-fold pada Skenario 1 pada Gambar 4.1 menunjukkan bahwa akurasi memiliki sebaran paling sempit dan standar deviasi kecil, menandakan performa model yang stabil di setiap fold. Sebaliknya, precision, recall, dan F1-score (macro) memiliki variasi yang lebih besar dan nilai rata-rata lebih rendah. Hal ini mengindikasikan bahwa meskipun akurasi relatif konsisten, kemampuan model dalam mengenali seluruh kelas secara merata masih terbatas akibat ketidakseimbangan data.

Rata-rata *evaluation matrix* per kelas pada Tabel 4.8 memperlihatkan bahwa model memiliki performa yang stabil pada kelas *Cataract*, sementara kelas

Glaucoma masih menunjukkan nilai recall dan F1-score yang sangat rendah. Perbedaan ini mengindikasikan bahwa terdapat tantangan dalam mengenali kelas tertentu secara optimal.

Tabel 4. 8 Rata-rata *Evaluation Matrix* per kelas

Kelas	Precision	Recall	F1-Score
Normal	61.57%	85.04%	71.39%
<i>Diabetic</i>	51%	24.51%	32.89%
<i>Glaucoma</i>	25.67%	04.19%	06.93%
<i>Cataract</i>	75.34%	73.33%	74.07%

4.3.2 Skenario Uji Coba 2 (A-2H-64)

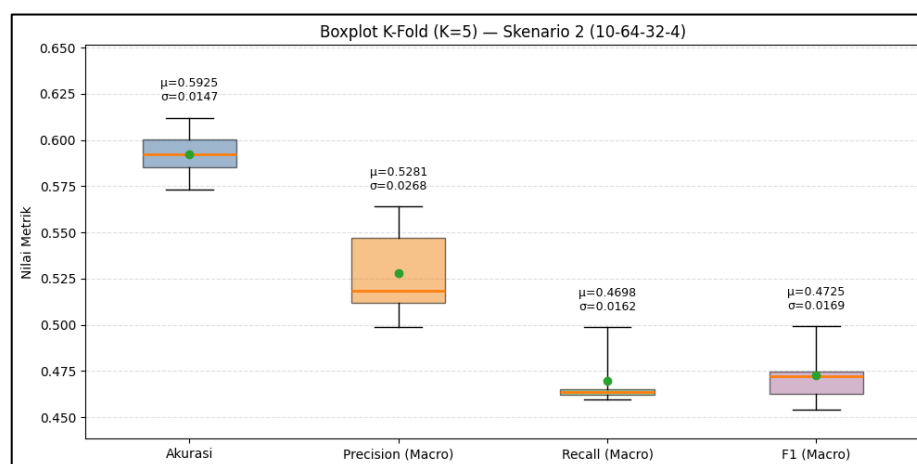
Pada skenario kedua, *Hidden layer* yang digunakan berjumlah 64 dan 32 *neuron* dengan 200 *epoch* dan *batch* sebesar 32. Tabel 4.9 menunjukkan hasil evaluasi performa model menggunakan *metode k-fold cross validation* dengan nilai $k = 5$. Evaluasi dilakukan dengan menghitung nilai akurasi, precision, recall dan f1-score pada setiap fold, kemudian dirata-ratakan untuk memperoleh gambaran kinerja model secara keseluruhan. Berdasarkan tabel tersebut, diperoleh rata-rata akurasi sebesar 59.25% dengan standar deviasi sebesar 0,0147.

Tabel 4. 9 *Evaluation Matrix K-Fold* Skenario 2 (Model A)

FOLD	Akurasi	Precision	Recall	F1-Score
1	59.20%	54.73%	46.51%	46.24%
2	57.33%	49.90%	45.97%	47.45%
3	60.02%	56.40%	49.86%	49.91%
4	58.50%	51.17%	46.22%	47.23%
5	61.20%	51.84%	46.36%	45.41%
Rata-rata	59.25%	52.81%	46.98%	47.25%
Standard Deviasi	0,0147	0.0268	0.0162	0,0169

Boxplot k-fold pada Skenario 2 pada Gambar 4.2 menunjukkan bahwa akurasi memiliki sebaran yang relatif sempit dengan standar deviasi kecil, menandakan performa model yang cukup stabil antar fold. Nilai precision

menunjukkan variasi yang lebih besar dibanding metrik lain, sementara recall dan F1-score (macro) berada pada kisaran rendah dengan sebaran yang relatif sempit. Menunjukkan bahwa meskipun model stabil secara akurasi, kemampuan mengenali seluruh kelas secara merata masih terbatas dan masih dipengaruhi oleh ketidakseimbangan data antar kelas.



Gambar 4. 2 Visualisasi Rata-rata evaluation matrix Skenario 2

Rata-rata *evaluation matrix* per kelas pada Tabel 4.10 memperlihatkan bahwa model memiliki performa yang stabil pada kelas *Cataract*, sementara kelas *Glaucoma* masih menunjukkan nilai recall dan F1-score yang sangat rendah. Perbedaan ini mengindikasikan bahwa terdapat tantangan dalam mengenali kelas tertentu secara optimal.

Tabel 4. 10 Rata-rata *Evaluation Matrix* per kelas

Kelas	Precision	Recall	F1-Score
Normal	62.07%	79.64%	69.70%
Diabetic	49.68%	30.59%	37.58%
Glaucoma	29.29%	08.36%	12.06%
Cataract	70.19%	69.33%	69.65%

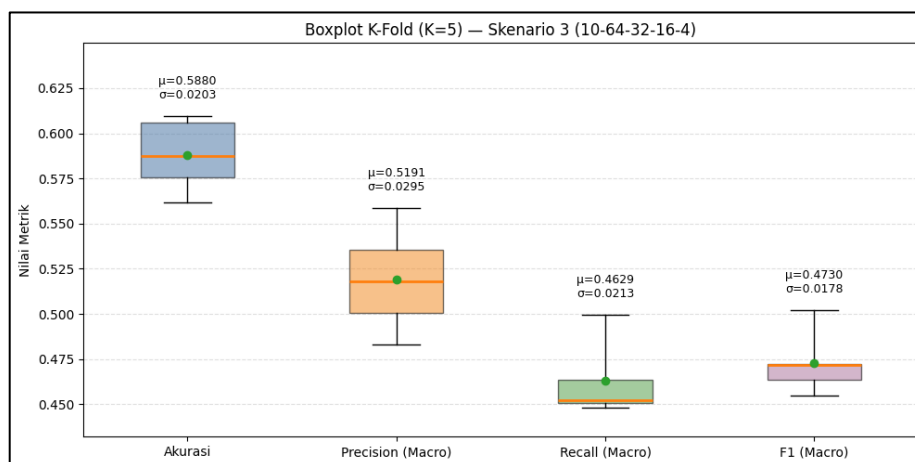
4.3.3 Skenario Uji Coba 3 (A-3H-64)

Pada skenario ketiga, *Hidden layer* yang digunakan berjumlah 64, 32 dan 16 *neuron* dengan 200 *epoch* dan *batch* sebesar 32. Tabel 4.11 menunjukkan hasil evaluasi performa model menggunakan *metode k-fold cross validation* dengan nilai $k = 5$. Evaluasi dilakukan dengan menghitung nilai akurasi, precision, recall dan f1-score pada setiap fold, kemudian dirata-ratakan untuk memperoleh gambaran kinerja model secara keseluruhan. Berdasarkan tabel tersebut, diperoleh rata-rata akurasi sebesar 58.80% dengan standar deviasi sebesar 0,0203.

Tabel 4. 11 *Evaluation Matrix K-Fold Skenario 3 (Model A)*

FOLD	Akurasi	Precision	Recall	F1-Score
1	56.15%	51.79%	44.82%	46.36%
2	57.56%	48.31%	45.21%	45.51%
3	60.96%	55.87%	49.95%	50.22%
4	58.73%	50.05%	46.36%	47.16%
5	60.61%	53.55%	45.09%	47.23%
Rata-rata	58.80%	51.91%	46.29%	47.30%
Standard Deviasi	0,0203	0.0295	0.0213	0,0178

Boxplot k-fold pada Skenario 3 pada Gambar 4.3 menunjukkan bahwa akurasi memiliki sebaran sedang, namun mengalami penurunan rata-rata dibanding skenario sebelumnya. Precision, recall, dan F1-score (macro) berada pada nilai yang lebih rendah dengan variasi yang masih terlihat antar fold. Kondisi ini menunjukkan bahwa penambahan *hidden layer* belum mampu meningkatkan kemampuan generalisasi model, dan ketidakseimbangan data masih menjadi faktor utama yang membatasi performa klasifikasi antar kelas.



Gambar 4. 3 Visualisasi Rata-rata *evaluation matrix* Skenario 3

Rata-rata *evaluation matrix* per kelas pada Tabel 4.12 memperlihatkan bahwa model memiliki performa terbaik pada kelas Normal dan Cataract, ditandai dengan nilai precision, recall, dan F1-score yang relatif tinggi. Kelas Normal memiliki recall paling tinggi, menandakan sebagian besar data Normal berhasil dikenali dengan baik. Sebaliknya, performa pada kelas Diabetic dan terutama *Glaucoma* masih rendah, khususnya pada nilai recall dan F1-score, yang mengindikasikan banyak kasus penyakit tersebut tidak terdeteksi oleh model. Ketimpangan ini menunjukkan bahwa model masih cenderung bias terhadap kelas mayoritas dan kesulitan membedakan kelas penyakit dengan karakteristik visual yang lebih kompleks atau jumlah data yang lebih sedikit.

Tabel 4. 12 Rata-rata *Evaluation Matrix* per kelas

Kelas	Precision	Recall	F1-Score
Normal	61.78%	80.22%	69.77%
Diabetic	48.64%	28.47%	35.57%
Glaucoma	24.08%	11.13%	15.00%
Cataract	73.16%	65.33%	68.84%

4.3.4 Skenario Uji Coba 4 (A-3H-128)

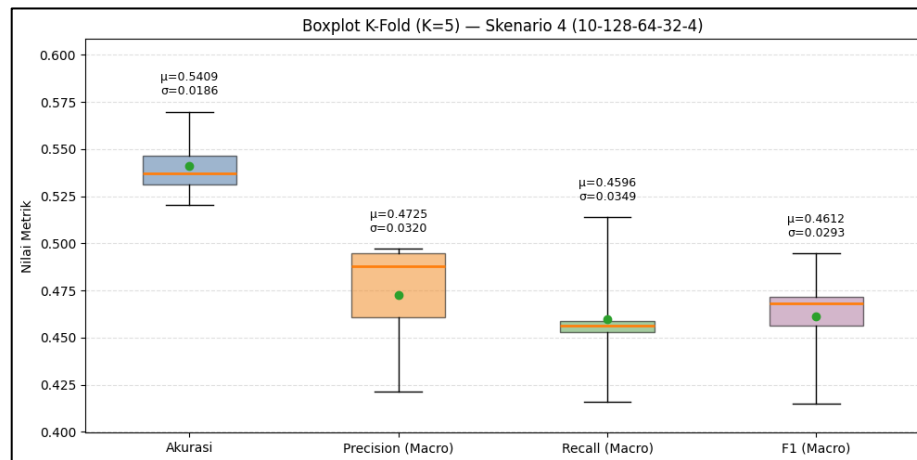
Pada skenario keempat, *Hidden layer* yang digunakan berjumlah 128, 64 dan 32 *neuron* dengan 200 *epoch* dan *batch* sebesar 32.

Tabel 4. 13 *Evaluation Matrix* K-Fold Skenario 4 (Model A)

FOLD	Akurasi	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1	53.69%	49.49%	45.31%	46.83%
2	53.11%	46.09%	45.63%	45.65%
3	56.98%	48.81%	51.38%	49.46%
4	52.05%	42.15%	41.61%	41.51%
5	54.63%	49.70%	45.87%	47.17%
Rata-rata	54.63%	47.25%	45.96%	46.12%
Standard Deviasi	0,0186	0.032	0.0349	0,0293

Hasil evaluasi performa model pada Tabel 4.13 menggunakan *metode k-fold cross validation* dengan nilai $k = 5$. Berdasarkan tabel tersebut, diperoleh rata-rata akurasi sebesar 54.63% dengan standar deviasi sebesar 0,0186.

Boxplot k-fold pada Skenario 4 pada Gambar 4.4 menunjukkan bahwa model relatif stabil antar fold dengan nilai standar deviasi yang tergolong kecil. Rata-rata akurasi berada di kisaran menengah, namun *precision*, *recall*, dan *F1-score* masih rendah dan memiliki sebaran yang lebih lebar dibanding akurasi. Hal ini menunjukkan bahwa model cukup konsisten dalam memprediksi kelas mayoritas, kemampuan dalam mendeteksi seluruh kelas secara seimbang terutama kelas minoritas masih terbatas. Peningkatan kompleksitas arsitektur pada skenario ini belum memberikan perbaikan signifikan terhadap kualitas prediksi per kelas.



Gambar 4. 4 Visualisasi Rata-rata *evaluation matrix* Skenario 4

Rata-rata *evaluation matrix* per kelas pada Tabel 4.12 memperlihatkan bahwa kelas *Cataract* dan *Normal* memiliki performa paling baik, dengan nilai precision, recall, dan F1-score yang sedikit tinggi, menandakan kedua kelas tersebut dapat dikenali secara konsisten oleh model. Sebaliknya, kelas *Diabetic* masih menunjukkan performa sedang dengan recall yang rendah, mengindikasikan banyak yang belum terdeteksi. Kelas *Glaucoma* memiliki performa terendah pada seluruh metrik, yang mencerminkan kesulitan model dalam mengenali pola *Glaucoma* akibat keterbatasan representasi data dan kemiripan fitur dengan kelas lain. Hal ini menunjukkan bahwa ketidakseimbangan data masih berdampak signifikan terhadap kemampuan klasifikasi per kelas.

Tabel 4. 14 Rata-rata *Evaluation Matrix* per kelas

Kelas	Precision	Recall	F1-Score
Normal	60.31%	69.65%	64.64%
Diabetic	41.19%	31.30%	35.49%
Glaucoma	19.69%	16.67%	17.69%
Cataract	67.80%	66.22%	66.67%

4.3.5 Skenario Uji Coba 5 (B-2H-32)

Pada skenario kelima, data yang digunakan terdiri dari 10.000 citra fundus yaitu 2.500 citra Normal, 2.500 citra *Diabetic*, 2.500 citra *Cataract*, dan 2.500 citra *Glaucoma*. Data tersebut dibagi ke dalam data *training*, data *validation* dan data *testing* dengan rasio 80:10:10. *Hidden layer* berjumlah 32 dan 16 *neuron* dengan 200 *epoch* dan *batch* sebesar 32.

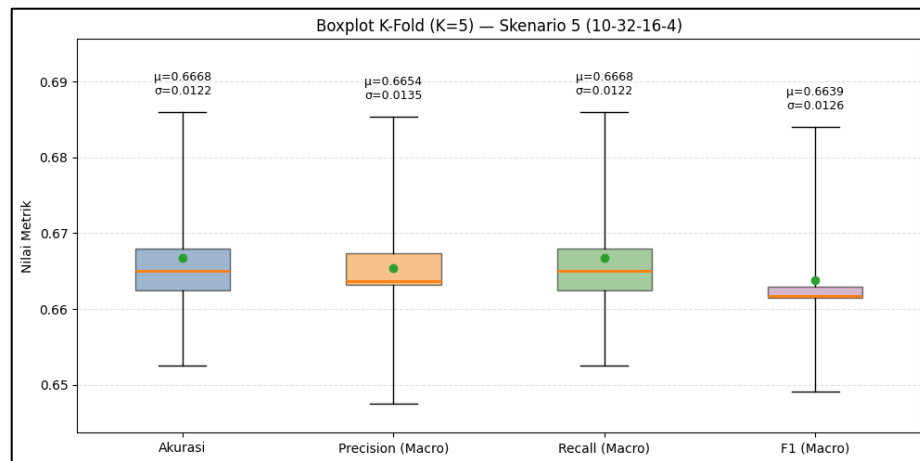
Tabel 4. 15 *Evaluation Matrix* K-Fold Skenario 5 (Model B)

FOLD	Akurasi	Precision	Recall	F1-Score
1	66.25%	66.73%	66.25%	66.17%
2	65.25%	64.75%	65.25%	64.91%
3	68.60%	68.54%	68.60%	68.41%
4	66.50%	66.32%	66.50%	66.30%
5	66.80%	66.37%	66.80%	66.14%
Rata-rata	66.68%	66.54%	66.68%	66.39%
Standard Deviasi	0,0122	0.0135	0.0122	0,0126

Hasil evaluasi performa model pada Tabel 4.15 menggunakan *metode k-fold cross validation* dengan nilai $k = 5$. Evaluasi dilakukan dengan menghitung nilai akurasi, precision, recall dan f1-score pada setiap fold, kemudian dirata-ratakan untuk memperoleh gambaran kinerja model secara keseluruhan. Berdasarkan tabel tersebut, diperoleh rata-rata akurasi sebesar 66.68% dengan standar deviasi sebesar 0,0122.

Boxplot k-fold pada Skenario 5 pada Gambar 4.5 menunjukkan bahwa seluruh metrik (akurasi, precision, recall, dan F1-score) memiliki rata-rata yang hampir seimbang di kisaran 66% dengan standar deviasi yang kecil, menandakan performa model yang cukup stabil dan konsisten pada setiap fold. Tidak terlihat perbedaan ekstrem antar metrik, yang mengindikasikan bahwa setelah penerapan

augmentasi, model mulai mampu mengenali seluruh kelas secara lebih merata dibandingkan skenario sebelumnya yang masih tidak seimbang.



Gambar 4. 5 Visualisasi Rata-rata *evaluation matrix* Skenario 5

Rata-rata *evaluation matrix* per kelas pada Tabel 4.16 menunjukkan bahwa kelas *Cataract* memiliki performa paling tinggi, menandakan model sangat baik dalam mengenali kelas ini. Kelas *Glaucoma* juga menunjukkan performa cukup baik dan seimbang. Sebaliknya, kelas *Diabetic* dan *Normal* masih memiliki nilai precision dan recall yang lebih rendah yang menunjukkan bahwa meskipun augmentasi meningkatkan performa secara umum, model masih mengalami kesulitan membedakan pola pada kelas-kelas tersebut.

Tabel 4. 16 Rata-rata *Evaluation Matrix* per kelas

Kelas	Precision	Recall	F1-Score
Normal	52.46%	47.44%	49.61%
Diabetic	52.79%	55.32%	52.76%
Glaucoma	72.31%	73.72%	72.80%
Cataract	88.61%	90.24%	89.37%

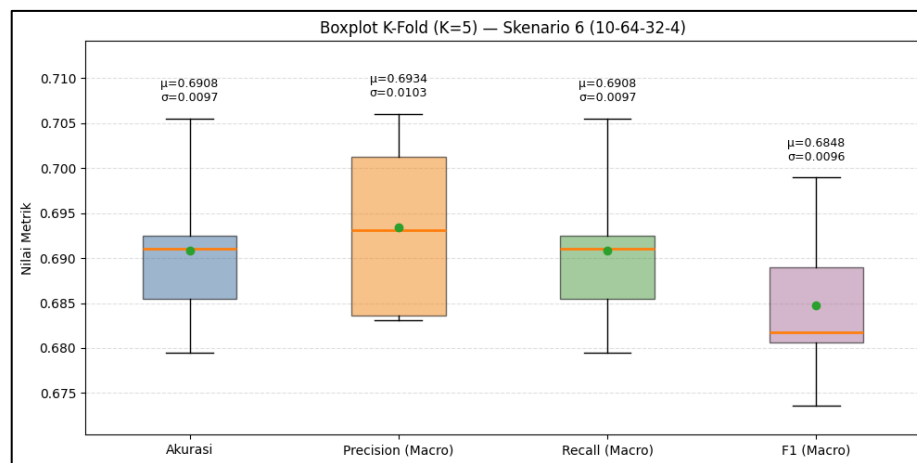
4.3.6 Skenario Uji Coba 6 (B-2H-64)

Pada skenario keenam, *Hidden layer* yang digunakan berjumlah 64 dan 32 *neuron* dengan 200 *epoch* dan *batch* sebesar 32.

Tabel 4. 17 *Evaluation Matrix* K-Fold Skenario 6 (Model B)

FOLD	Akurasi	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1	68.55%	68.31%	68.31%	67.36%
2	69.25%	70.13%	69.25%	68.18%
3	69.10%	69.31%	69.10%	68.90%
4	67.95%	68.36%	67.95%	68.06%
5	70.55%	70.60%	70.55%	69.89%
Rata-rata	69.08%	69.34%	69.08%	68.48%
Standard Deviasi	0,0097	0.0103	0.0097	0,0096

Hasil evaluasi performa model pada Tabel 4.17 menggunakan *metode k-fold cross validation* dengan nilai $k = 5$. Berdasarkan tabel tersebut, diperoleh rata-rata akurasi sebesar 69.08% dengan standar deviasi sebesar 0,0097.



Gambar 4. 6 Visualisasi Rata-rata *evaluation matrix* Skenario 6

Boxplot k-fold pada Skenario 6 pada Gambar 4.6 menunjukkan performa stabil pada seluruh metrik dengan standar deviasi kecil, yang menandakan konsistensi performa antar fold. Nilai akurasi, precision, dan recall berada pada kisaran 68-70%, sementara F1-score sedikit lebih rendah, mengindikasikan

keseimbangan klasifikasi sudah baik namun masih terdapat kesalahan ringan pada beberapa kelas.

Rata-rata *evaluation matrix* per kelas pada Tabel 4.18 menunjukkan bahwa kelas *Cataract* memiliki performa sangat baik dengan precision, recall, dan F1-score tinggi, menandakan model mampu mengenali kelas tersebut secara konsisten. Kelas *Diabetic* dan *Normal* masih menunjukkan performa rendah, khususnya pada nilai recall dan F1-score, yang mengindikasikan sebagian data masih salah diklasifikasikan. Secara umum, skenario ini memperlihatkan peningkatan kemampuan model pada kelas penyakit tertentu, meskipun tantangan klasifikasi masih terdapat pada kelas dengan karakteristik visual yang saling tumpang tindih.

Tabel 4. 18 Rata-rata *Evaluation Matrix* per kelas

Kelas	Precision	Recall	F1-Score
Normal	52.07%	57.40%	53.56%
<i>Diabetic</i>	56.35%	46.64%	49.86%
<i>Glaucoma</i>	76.26%	80.00%	78.02%
<i>Cataract</i>	92.69%	92.28%	92.47%

4.3.7 Skenario Uji Coba 7 (B-3H-64)

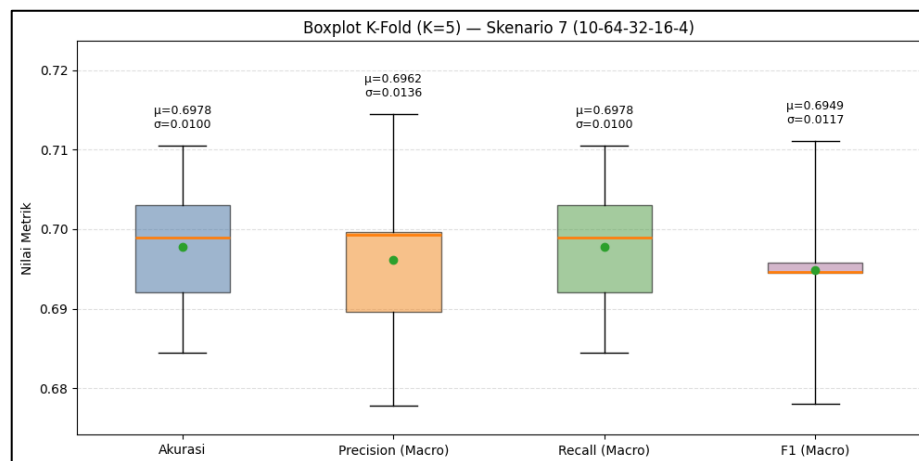
Pada skenario ketujuh, *Hidden layer* yang digunakan berjumlah 64, 32 dan 16 *neuron* dengan 200 *epoch* dan *batch* sebesar 32.

Tabel 4. 19 *Evaluation Matrix* K-Fold Skenario 7 (Model B)

FOLD	Akurasi	Precision	Recall	F1-Score
1	69.90%	69.92%	69.90%	69.58%
2	68.45%	67.78%	68.45%	67.81%
3	71.05%	71.45%	71.05%	71.12%
4	69.20%	69.97%	69.20%	69.47%
5	70.30%	68.96%	70.30%	69.45%
Rata-rata	69.78%	69.62%	69.78%	69.49%
Standard Deviasi	0,01	0.0136	0.01	0,0117

Hasil evaluasi performa model pada Tabel 4.19 menggunakan *metode k-fold cross validation* dengan nilai $k = 5$. Berdasarkan tabel tersebut, diperoleh rata-rata akurasi sebesar 69.78% dengan standar deviasi sebesar 0,01.

Boxplot k-fold pada Skenario 7 pada Gambar 4.7 menunjukkan keempat metrik evaluasi (akurasi, precision, recall, dan F1-score) menunjukkan sebaran yang relatif sempit dengan nilai standar deviasi yang kecil, yang menandakan konsistensi performa antar fold.



Gambar 4. 7 Visualisasi Rata-rata *evaluation matrix* Skenario 7

Rata-rata *evaluation matrix* per kelas pada Tabel 4.20 menunjukkan bahwa kelas *Cataract* memiliki performa sangat baik dengan precision, recall, dan F1-score tinggi, menandakan model mampu mengenali kelas tersebut secara konsisten. Kelas *Glaucoma* juga memiliki performa yang cukup tinggi menandakan kemampuan klasifikasi yang stabil. Kelas *Diabetic* dan *Normal* masih menunjukkan performa rendah, khususnya pada nilai recall dan F1-score, yang mengindikasikan sebagian data masih salah diklasifikasikan. Secara umum, skenario ini memperlihatkan peningkatan kemampuan model pada kelas penyakit

tertentu, meskipun tantangan klasifikasi masih terdapat pada kelas dengan karakteristik visual yang saling tumpang tindih.

Tabel 4. 20 Rata-rata *Evaluation Matrix* per kelas

Kelas	Precision	Recall	F1-Score
Normal	52.39%	55.44%	53.75%
Diabetic	55.60%	50.72%	52.75%
Glaucoma	79.71%	77.36%	78.33%
Cataract	90.76%	95.60%	93.11%

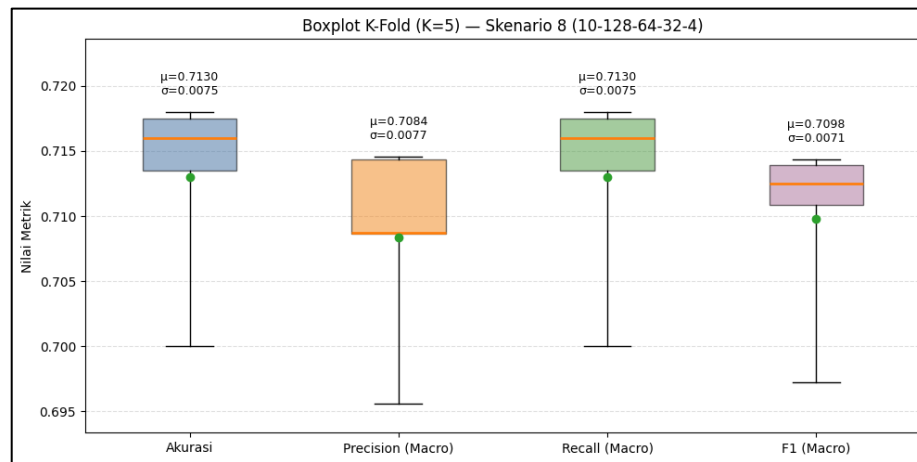
4.3.8 Skenario Uji Coba 8 (B-3H-128)

Pada skenario kedelapan, *Hidden layer* yang digunakan berjumlah 128, 64 dan 32 *neuron* dengan 200 *epoch* dan *batch* sebesar 32. Hasil evaluasi performa model pada Tabel 4.21 menggunakan *metode k-fold cross validation* dengan nilai $k = 5$. Berdasarkan tabel tersebut, diperoleh rata-rata akurasi sebesar 71.30% dengan standar deviasi sebesar 0,0075.

Tabel 4. 21 *Evaluation Matrix K-Fold* Skenario 8 (Model B)

FOLD	Akurasi	Precision	Recall	F1-Score
1	71.35%	71.44%	71.35%	71.39%
2	71.80%	70.88%	71.80%	71.25%
3	71.60%	70.86%	71.60%	71.08%
4	70.00%	69.56%	70.00%	69.73%
5	71.75%	71.46%	71.75%	71.44%
Rata-rata	71.30%	70.84%	71.30%	70.98%
Standard Deviasi	0,0075	0.0077	0.0075	0,0071

Boxplot k-fold pada Skenario 8 pada Gambar 4.8 menunjukkan sebaran nilai metrik yang konsisten di seluruh fold. Median antar metrik relatif sejajar dan tidak tampak outlier ekstrem, menandakan proses pelatihan yang stabil serta kemampuan generalisasi model yang baik. Konsistensi ini mengindikasikan bahwa arsitektur dan perlakuan data pada skenario 8 mampu menghasilkan performa yang tidak sensitif terhadap variasi pembagian data.



Gambar 4. 8 Visualisasi Rata-rata *evaluation matrix* Skenario 8

Rata-rata *evaluation matrix* per kelas pada Tabel 4.22 menunjukkan bahwa model efektif dalam mengenali kelas *Cataract* dan *Glaucoma* dengan ketepatan dan kelengkapan prediksi yang tinggi. Sebaliknya, performa pada kelas *Diabetic* dan *Normal* masih lebih rendah, menunjukkan adanya kemiripan pola antar kedua kelas tersebut yang menyulitkan pemisahan. Secara keseluruhan, model menunjukkan keseimbangan performa yang baik dengan keunggulan pada kelas yang memiliki ciri visual lebih kontras.

Tabel 4. 22 Rata-rata *Evaluation Matrix* per kelas

Kelas	Precision	Recall	F1-Score
Normal	52.32%	49.48%	50.75%
Diabetic	55.17%	54.28%	54.62%
Glaucoma	81.91%	86.56%	84.13%
Cataract	93.95%	94.88%	94.41%

4.3.9 Skenario Uji Coba 9 (C-2H-32)

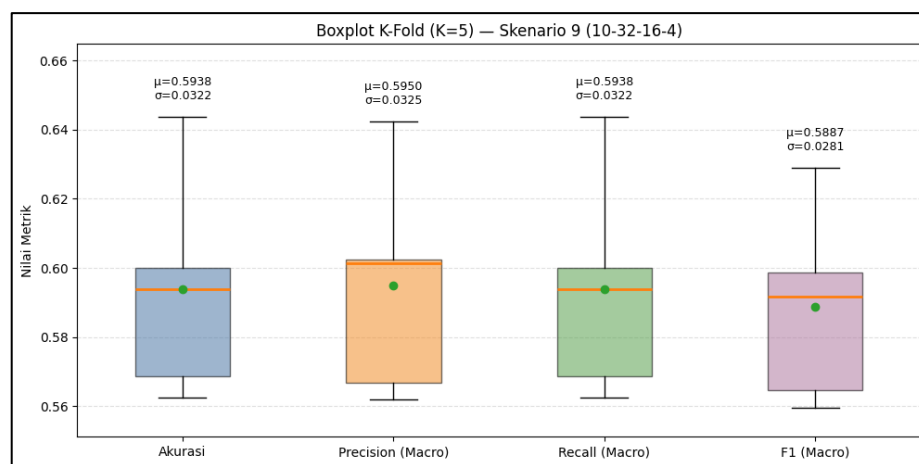
Pada skenario kesembilan, data yang digunakan adalah 800 citra fundus yang terdiri dari 200 data pada masing-masing kelasnya. Data tersebut dibagi ke

dalam data *training*, data *validation* dan data *testing* dengan rasio 80:10:10. *Hidden layer* berjumlah 32 dan 16 *neuron* dengan 150 *epoch* dan *batch* sebesar 16.

Tabel 4. 23 *Evaluation Matrix K-Fold Skenario 9 (Model C)*

FOLD	Akurasi	Precision	Recall	F1-Score
1	59.38%	60.14%	59.38%	59.18%
2	60.00%	60.23%	60.00%	59.87%
3	56.25%	56.20%	56.25%	55.96%
4	64.38%	64.24%	64.37%	62.90%
5	56.87%	56.68%	56.88%	56.47%
Rata-rata	59.38%	59.50%	59.38%	58.87%
Standard Deviasi	0,0322	0.0325	0.0322	0,0281

Hasil evaluasi performa model pada Tabel 4.23 menggunakan *metode k-fold cross validation* dengan nilai $k = 5$. Evaluasi dilakukan dengan menghitung nilai akurasi, precision, recall dan f1-score pada setiap fold, kemudian dirata-ratakan untuk memperoleh gambaran kinerja model secara keseluruhan. Berdasarkan tabel tersebut, diperoleh rata-rata akurasi sebesar 59.38% dengan standar deviasi sebesar 0,0322.



Gambar 4. 9 Visualisasi Rata-rata *evaluation matrix* Skenario 9

Boxplot k-fold pada Skenario 9 pada Gambar 4.9 menunjukkan sebaran metrik yang relatif lebih lebar dibanding skenario dengan data lebih besar. Variasi

antar fold masih terlihat pada seluruh metrik, khususnya pada precision dan F1-score, yang mengindikasikan bahwa performa model cukup sensitif terhadap pembagian data meskipun dataset sudah seimbang per kelas. Namun demikian, tidak tampak fluktuasi ekstrem, sehingga model masih memiliki kemampuan generalisasi yang dapat diterima.

Tabel 4. 24 Rata-rata *Evaluation Matrix* per kelas

Kelas	Precision	Recall	F1-Score
Normal	47.41%	36.50%	40.76%
<i>Diabetic</i>	48.97%	57.50%	52.81%
<i>Glaucoma</i>	59.55%	61.50%	59.96%
<i>Cataract</i>	82.06%	82.00%	81.97%

Rata-rata *evaluation matrix* per kelas pada Tabel 4.24 menunjukkan bahwa model paling baik dalam mengenali kelas *Cataract*, yang menunjukkan pola fitur lebih konsisten dan mudah dipisahkan. Kelas *Glaucoma* berada pada tingkat menengah dengan performa cukup stabil. Sebaliknya, kelas *Diabetic* dan terutama *Normal* masih menunjukkan kinerja yang rendah, menandakan adanya tumpang tindih karakteristik antar kedua kelas tersebut. Hal ini menunjukkan bahwa meskipun dataset telah diseimbangkan melalui *undersampling*, jumlah data masih membatasi kemampuan model dalam mempelajari variasi fitur pada kelas-kelas tertentu.

4.3.10 Skenario Uji Coba 10 (C-2H-64)

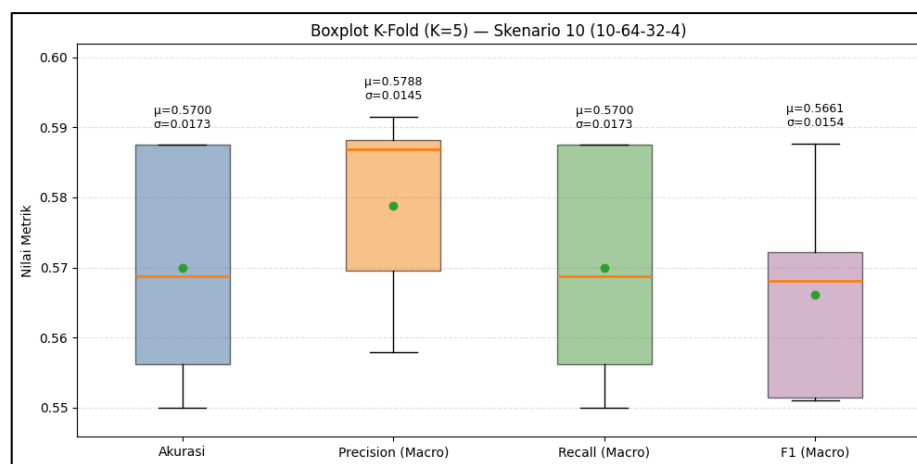
Pada skenario kesepuluh, *Hidden layer* yang digunakan berjumlah 64 dan 32 *neuron* dengan 150 *epoch* dan *batch* sebesar 16. Hasil evaluasi performa model pada Tabel 4.25 menggunakan *metode k-fold cross validation* dengan nilai $k = 5$.

Berdasarkan tabel tersebut, diperoleh rata-rata akurasi sebesar 57.88% dengan standar deviasi sebesar 0.173

Tabel 4. 25 *Evaluation Matrix* K-Fold Skenario 10 (Model C)

FOLD	Akurasi	Precision	Recall	F1-Score
1	58.75%	58.82%	58.75%	57.21%
2	58.75%	59.15%	58.75%	58.77%
3	56.87%	56.96%	56.87%	56.81%
4	55.00%	55.79%	55.00%	55.11%
5	55.63%	58.69%	55.63%	55.14%
Rata-rata	57.88%	57.88%	57.00%	56.61%
Standard Deviasi	0,0173	0.0145	0.0173	0,0154

Boxplot k-fold pada Skenario 10 pada Gambar 4.10 menunjukkan sebaran metrik yang relatif lebih sempit dibanding skenario 9, menandakan stabilitas model yang sedikit lebih baik. Variasi antar fold masih terlihat, namun fluktuasinya lebih terkendali terutama pada precision dan F1-score. Hal ini mengindikasikan bahwa konfigurasi arsitektur yang digunakan mampu memberikan pembelajaran yang lebih konsisten pada dataset seimbang, meskipun jumlah data masih terbatas.



Gambar 4. 10 Visualisasi Rata-rata *evaluation matrix* Skenario 10

Rata-rata *evaluation matrix* per kelas pada Tabel 4.26 menunjukkan performa terbaik tetap ditunjukkan pada kelas *Cataract*, yang paling mudah

dikenali oleh model. Kelas *Glaucoma* menunjukkan kinerja menengah dengan prediksi yang cukup stabil. Sementara itu, kelas *Diabetic* dan *Normal* masih menjadi tantangan utama, dengan kecenderungan kesalahan klasifikasi yang lebih tinggi akibat kemiripan pola fitur antar kelas. Hal ini menunjukkan bahwa peningkatan arsitektur membantu stabilitas model, namun keterbatasan jumlah data tetap membatasi kemampuan model dalam membedakan kelas-kelas yang memiliki karakteristik visual serupa.

Tabel 4. 26 Rata-rata *Evaluation Matrix* per kelas

Kelas	Precision	Recall	F1-Score
Normal	42.83%	34.50%	37.83%
<i>Diabetic</i>	45.31%	55.50%	49.80%
<i>Glaucoma</i>	62.32%	58.50%	59.16%
<i>Cataract</i>	81.06%	79.50%	79.64%

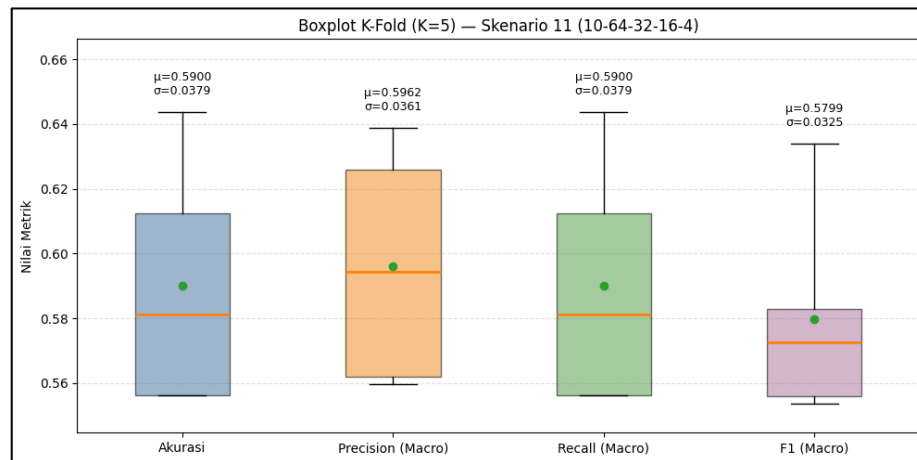
4.3.11 Skenario Uji Coba 11 (C-3H-64)

Pada skenario kesebelas, *Hidden layer* yang digunakan berjumlah 64, 32 dan 16 *neuron* dengan 150 *epoch* dan *batch* sebesar 16.

Tabel 4. 27 *Evaluation Matrix* K-Fold Skenario 11 (Model C)

FOLD	Akurasi	Precision	Recall	F1-Score
1	64.38%	63.90%	64.38%	63.39%
2	58.13%	59.44%	58.13%	57.27%
3	55.63%	55.98%	55.63%	55.61%
4	61.25%	62.61%	61.25%	58.30%
5	55.63%	56.21%	55.62%	55.37%
Rata-rata	59.00%	59.62%	59.00%	57.99%
Standard Deviasi	0.0379	0.0361	0.0379	0,0325

Hasil evaluasi performa model pada Tabel 4.27 menggunakan *metode k-fold cross validation* dengan nilai $k = 5$. Berdasarkan tabel tersebut, diperoleh rata-rata akurasi sebesar 59% dengan standar deviasi sebesar 0.0379.



Gambar 4. 11 Visualisasi Rata-rata *evaluation matrix* Skenario 11

Boxplot k-fold pada Skenario 11 pada Gambar 4.11 menunjukkan bahwa performa model relatif stabil di seluruh metrik dengan nilai standar deviasi yang kecil. Nilai rata-rata akurasi dan recall macro berada di sekitar 0,59, sementara precision macro sedikit lebih tinggi ($\approx 0,60$), menandakan prediksi benar cenderung lebih presisi daripada kemampuan menangkap seluruh sampel kelas. F1-score macro menjadi yang terendah ($\approx 0,58$), mencerminkan masih adanya ketidakseimbangan antara precision dan recall antar kelas. Secara keseluruhan, model cukup konsisten antar fold, namun kemampuan generalisasi multi-kelas pada skenario ini masih terbatas.

Tabel 4. 28 Rata-rata *Evaluation Matrix* per kelas

Kelas	Precision	Recall	F1-Score
Normal	50.68%	30.00%	36.98%
Diabetic	49.00%	63.00%	54.69%
Glaucoma	57.51%	61.50%	59.02%
Cataract	81.30%	81.50%	81.26%

Rata-rata *evaluation matrix* per kelas pada Tabel 4.28 menunjukkan performa model masih belum merata antar kelas. Kelas *Cataract* memiliki kinerja

terbaik menandakan pola fitur pada kelas ini dapat dipelajari dengan baik oleh model. Sebaliknya, kelas *Normal* menunjukkan performa terendah yang menunjukkan banyak data *Normal* salah diklasifikasikan ke kelas lain. Kelas *Diabetic* dan *Glaucoma* berada pada tingkat performa menengah. Hal ini menunjukkan bahwa meskipun model mampu mengenali kelas tertentu dengan baik, masih terdapat ketidakseimbangan kemampuan klasifikasi antar kelas pada skenario ini.

4.3.12 Skenario Uji Coba 12 (C-3H-128)

Pada skenario ke-dua belas, *Hidden layer* yang digunakan berjumlah 128, 64 dan 32 *neuron* dengan 150 *epoch* dan *batch* sebesar 16.

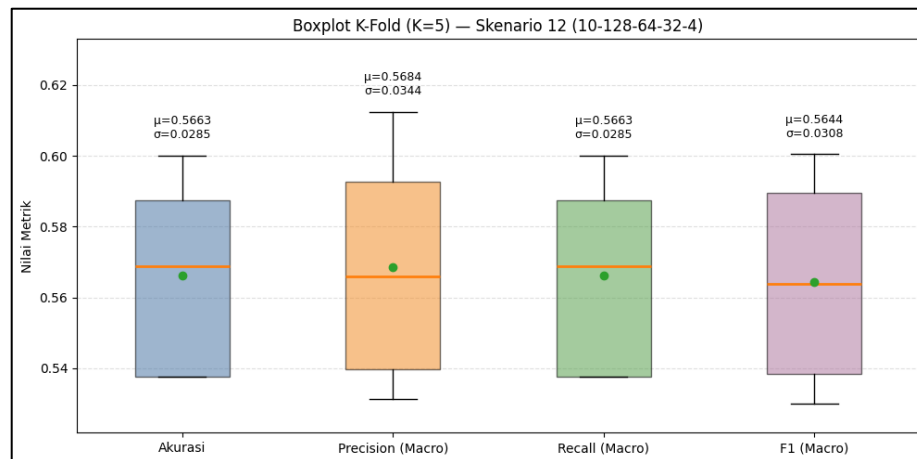
Tabel 4. 29 *Evaluation Matrix* K-Fold Skenario 10 (Model C)

FOLD	Akurasi	Precision	Recall	F1-Score
1	60.00%	61.24%	60.00%	60.05%
2	58.75%	59.27%	58.75%	58.94%
3	53.75%	53.13%	53.75%	53.00%
4	56.87%	56,60%	56.87%	56.37%
5	53.75%	53.97%	53.75%	53.85%
Rata-rata	56.63%	56.84%	56.63%	56.44%
Standard Deviasi	0,0285	0.0344	0.0285	0,0308

Hasil evaluasi performa model pada Tabel 4.29 menggunakan *metode k-fold cross validation* dengan nilai $k = 5$. Berdasarkan tabel tersebut, diperoleh rata-rata akurasi sebesar 56.63% dengan standar deviasi sebesar 0.0285.

Boxplot k-fold pada Skenario 12 pada Gambar 4.12 menunjukkan performa model yang cukup stabil dan standar deviasi relatif kecil. Hal ini menandakan bahwa model menghasilkan performa yang konsisten pada berbagai pembagian data. Precision macro sedikit lebih tinggi dibanding akurasi dan recall,

mengindikasikan bahwa prediksi yang benar relatif presisi, meskipun tidak selalu mampu menangkap seluruh sampel kelas secara merata.



Gambar 4. 12 Visualisasi Rata-rata *evaluation matrix* Skenario 12

Rata-rata *evaluation matrix* per kelas pada Tabel 4.30 model menunjukkan kinerja terbaik pada kelas *Cataract*, diikuti oleh *Glaucoma*. Sebaliknya, kelas *Diabetic* dan *Normal* masih memiliki F1-score yang rendah, menandakan model kesulitan dalam membedakan kedua kelas tersebut. Secara keseluruhan, Skenario 12 memiliki stabilitas yang baik antar fold, namun performa antar kelas masih belum seimbang, sehingga peningkatan representasi fitur atau perlakuan data tambahan masih diperlukan.

Tabel 4. 30 Rata-rata *Evaluation Matrix* per kelas

Kelas	Precision	Recall	F1-Score
Normal	42.75%	45.50%	44.03%
Diabetic	46.26%	43.00%	44.01%
Glaucoma	59.21%	58.00%	58.21%
Cataract	79.15%	80.00%	79.52%

4.3.13 Skenario Uji Coba 13 (D-2H-32)

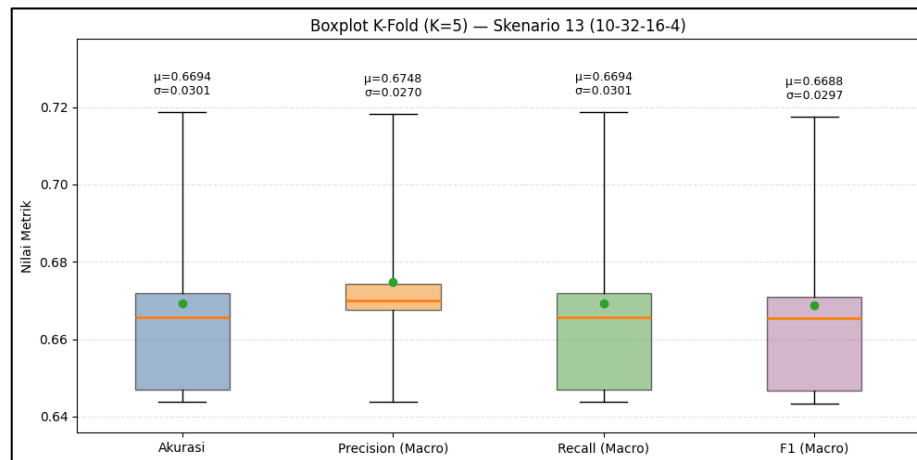
Pada skenario ke-tiga belas, data yang digunakan terdiri dari 1.600 citra fundus yaitu 400 citra Normal, 400 citra *Diabetic*, 400 citra *Cataract*, dan 400 citra *Glaucoma*. Data tersebut dibagi ke dalam data *training*, data *validation* dan data *testing* dengan rasio 80:10:10. *Hidden layer* berjumlah 32 dan 16 *neuron* dengan 150 *epoch* dan *batch* sebesar 16.

Tabel 4. 31 *Evaluation Matrix* K-Fold Skenario 13 (Model D)

FOLD	Akurasi	Precision	Recall	F1-Score
1	67.19%	67.43%	67.19%	67.10%
2	64.69%	64.39%	64.69%	64.34%
3	66.56%	66.77%	66.56%	66.54%
4	64.38%	66.99%	64.37%	64.68%
5	71.88%	71.82%	71.88%	71.76%
Rata-rata	66.94%	67.48%	66.94%	66.88%
Standard Deviasi	0,0301	0.0270	0.0301	0,0297

Hasil evaluasi performa model pada Tabel 4.31 menggunakan *metode k-fold cross validation* dengan nilai $k = 5$. Evaluasi dilakukan dengan menghitung nilai akurasi, precision, recall dan f1-score pada setiap fold, kemudian dirata-ratakan untuk memperoleh gambaran kinerja model secara keseluruhan. Berdasarkan tabel tersebut, diperoleh rata-rata akurasi sebesar 66.94% dengan standar deviasi sebesar 0,0301.

Boxplot k-fold pada Skenario 13 pada Gambar 4.13 menunjukkan performa model yang cukup stabil dan konsisten, dengan nilai rata-rata akurasi, precision, recall, dan F1-score berada di kisaran 66–67% serta standar deviasi yang relatif kecil. Hal ini menunjukkan bahwa model memiliki stabilitas generalisasi yang baik pada berbagai pembagian data fold.



Gambar 4. 13 Visualisasi Rata-rata *evaluation matrix* Skenario 13

Rata-rata *evaluation matrix* per kelas pada Tabel 4.32 menunjukkan bahwa model mampu mengenali kelas Cataract dengan sangat baik ($F1 \approx 79\%$), diikuti oleh Glaucoma ($F1 \approx 69\%$). Kelas Diabetic dan Normal menunjukkan performa yang cukup, meskipun masih lebih rendah dibanding dua kelas lainnya. Secara keseluruhan, Skenario 13 memperlihatkan peningkatan keseimbangan performa antar kelas dibanding skenario sebelumnya, yang menunjukkan bahwa penggunaan data yang lebih seimbang mulai memberikan dampak positif terhadap kemampuan klasifikasi model.

Tabel 4. 32 Rata-rata *Evaluation Matrix* per kelas

Kelas	Precision	Recall	F1-Score
Normal	61.00%	56.25%	58.51%
Diabetic	58.82%	63.00%	60.42%
Glaucoma	68.01%	71.00%	69.24%
Cataract	82.09%	77.50%	79.36%

4.3.14 Skenario Uji Coba 14 (D-2H-64)

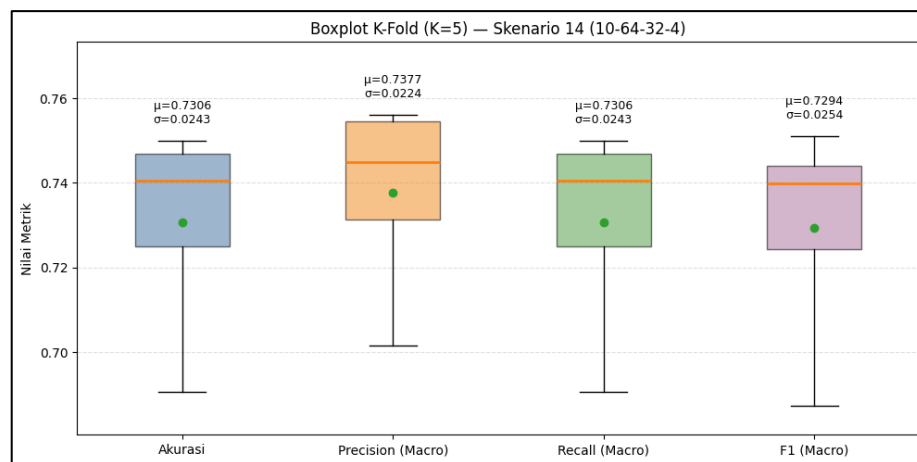
Pada skenario ke-empat belas, *Hidden layer* yang digunakan berjumlah 64 dan 32 *neuron* dengan 150 *epoch* dan *batch* sebesar 16. Hasil evaluasi performa

model pada Tabel 4.33 menggunakan *metode k-fold cross validation* dengan nilai $k = 5$. Berdasarkan tabel tersebut, diperoleh rata-rata akurasi sebesar X% dengan standar deviasi sebesar X.

Tabel 4. 33 *Evaluation Matrix K-Fold Skenario 14 (Model D)*

FOLD	Akurasi	Precision	Recall	F1-Score
1	72.50%	73.13%	72.50%	72.44%
2	69.06%	70.16%	69.06%	68.73%
3	75.00%	75.61%	75.00%	75.11%
4	74.06%	75.44%	74.06%	73.99%
5	74.69%	74.48%	74.69%	74.40%
Rata-rata	73.06%	73.77%	73.06%	72.94%
Standard Deviasi	0,0243	0.0224	0.0243	0,0254

Boxplot k-fold pada Skenario 14 pada Gambar 4.14 menunjukkan performa lebih baik dan stabil dibandingkan skenario sebelumnya, dengan nilai rata-rata akurasi, precision, recall, dan F1-score berada pada kisaran 73–74% serta standar deviasi yang relatif kecil. Hal ini menunjukkan bahwa model tidak hanya mencapai performa tinggi, tetapi juga konsisten pada berbagai pembagian data fold.



Gambar 4. 14 Visualisasi Rata-rata *evaluation matrix* Skenario 14

Rata-rata *evaluation matrix* per kelas pada Tabel 4.34 menunjukkan kelas *Cataract* menunjukkan performa tinggi ($F1 \approx 83.50\%$), diikuti oleh *Glaucoma* ($F1$

$\approx 73.41\%$). Kelas *Diabetic* dan *Normal* juga mengalami peningkatan yang cukup signifikan dengan F1-score di atas 66%, menandakan bahwa model mampu mengurangi bias antar kelas. Secara keseluruhan, Skenario 14 merupakan konfigurasi yang paling optimal karena mampu mengombinasikan akurasi tinggi, kestabilan model, dan keseimbangan performa antar kelas, sehingga dapat dijadikan model terbaik dalam penelitian ini.

Tabel 4. 34 Rata-rata *Evaluation Matrix* per kelas

Kelas	Precision	Recall	F1-Score
Normal	67.75%	69.25%	68.08%
<i>Diabetic</i>	67.21%	67.50%	66.76%
<i>Glaucoma</i>	77.96%	70.25%	73.41%
<i>Cataract</i>	82.14%	85.25%	83.50%

4.3.15 Skenario Uji Coba 15 (D-3H-64)

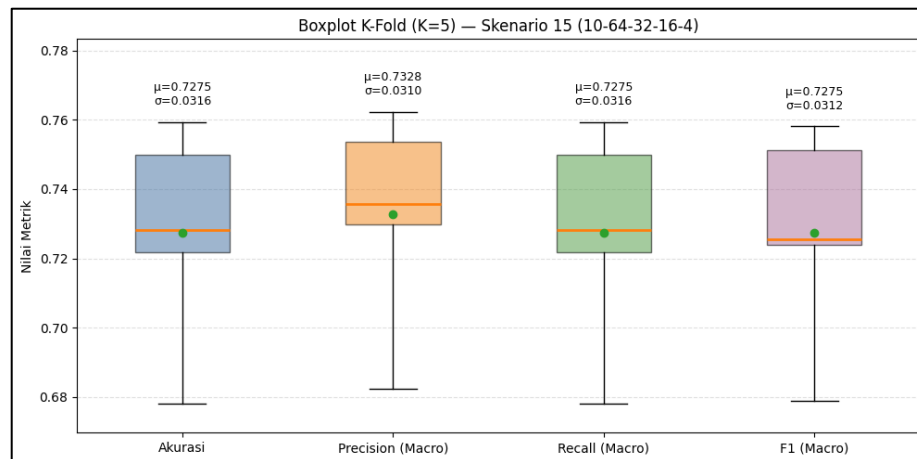
Pada skenario ke-lima belas, *Hidden layer* yang digunakan berjumlah 64, 32 dan 16 *neuron* dengan 150 *epoch* dan *batch* sebesar 16. Hasil evaluasi performa model pada Tabel 4.35 menggunakan *metode k-fold cross validation* dengan nilai $k = 5$. Berdasarkan tabel tersebut, diperoleh rata-rata akurasi sebesar X% dengan standar deviasi sebesar X.

Tabel 4. 35 *Evaluation Matrix K-Fold* Skenario 15 (Model D)

FOLD	Akurasi	Precision	Recall	F1-Score
1	75.94%	76.22%	75.94%	75.81%
2	67.81%	68.25%	67.81%	67.88%
3	75.00%	75.35%	75.00%	75.11%
4	72.19%	73.57%	72.19%	72.55%
5	72.81%	72.99%	72.81%	72.38%
Rata-rata	72.75%	73.28%	72.75%	72.75%
Standard Deviasi	0,0316	0.031	0.0316	0,0312

Boxplot k-fold pada Skenario 15 pada Gambar 4.15 menunjukkan performa yang cukup tinggi dan relatif stabil. Hal ini menandakan bahwa model memiliki

konsistensi yang baik antar fold, meskipun variasinya sedikit lebih besar dibanding skenario 14..



Gambar 4. 15 Visualisasi Rata-rata *evaluation matrix* Skenario 15

Rata-rata *evaluation matrix* per kelas pada Tabel 4.36 kelas *Cataract* kembali menunjukkan performa terbaik ($F1 = 83.97\%$), diikuti oleh *Glaucoma* ($F1 = 73.94\%$). Sementara itu, kelas *Diabetic* dan *Normal* memiliki F1-score di kisaran 66%, yang mengindikasikan bahwa peningkatan kompleksitas arsitektur (3 *hidden layer*) belum memberikan peningkatan signifikan pada kelas yang memiliki pola fitur saling tumpang tindih. Secara keseluruhan, Skenario 15 menghasilkan performa yang baik, namun tidak melampaui kestabilan dan keseimbangan performa yang dicapai pada Skenario 14.

Tabel 4. 36 Rata-rata *Evaluation Matrix* per kelas

Kelas	Precision	Recall	F1-Score
Normal	66.62%	66.25%	66.19%
Diabetic	67.18%	67.00%	66.89%
Glaucoma	74.10%	74.25%	73.94%
Cataract	85.21%	83.50%	83.97%

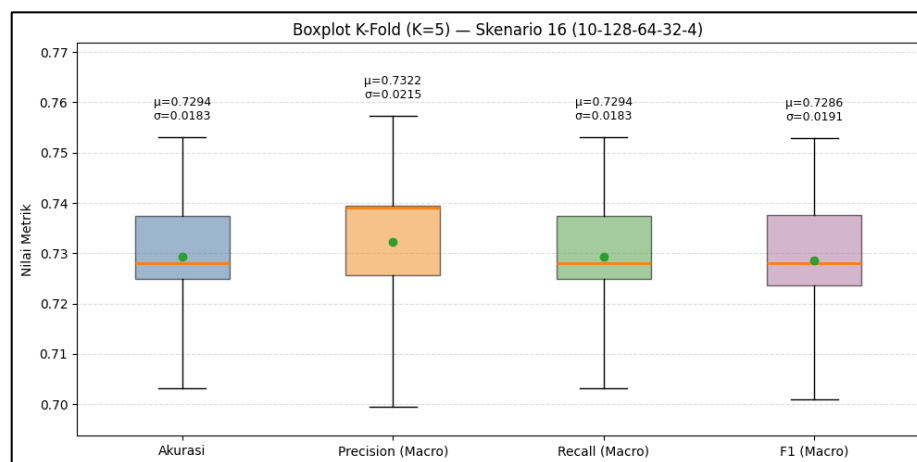
4.3.16 Skenario Uji Coba 16 (D-3H-128)

Pada skenario ke-enam belas, *Hidden layer* yang digunakan berjumlah 128, 64 dan 32 *neuron* dengan 150 *epoch* dan *batch* sebesar 16. Hasil evaluasi performa model pada Tabel 4.37 menggunakan *metode k-fold cross validation* dengan nilai $k = 5$. Berdasarkan tabel tersebut, diperoleh rata-rata akurasi sebesar x% dengan standar deviasi sebesar x.

Tabel 4. 37 *Evaluation Matrix K-Fold Skenario 16 (Model D)*

FOLD	Akurasi	Precision	Recall	F1-Score
1	73.75%	73.90%	73.75%	73.76%
2	70.31%	69.95%	70.31%	70.09%
3	75.31%	75.73%	75.31%	75.29%
4	72.81%	73.95%	72.81%	72.80%
5	72.50%	72.57%	72.50%	72.36%
Rata-rata	72.94%	73.22%	72.94%	72.86%
Standard Deviasi	0,0183	0.0215	0.0183	0,0191

Boxplot k-fold pada Skenario 16 pada Gambar 4.16 menunjukkan performa model yang stabil dan konsisten pada seluruh metrik, dengan sebaran antar fold yang relatif sempit. Hal ini mengindikasikan kemampuan generalisasi yang baik meskipun arsitektur yang digunakan memiliki kompleksitas lebih tinggi



Gambar 4. 16 Visualisasi Rata-rata *evaluation matrix* Skenario 16

Rata-rata *evaluation matrix* per kelas pada Tabel 4.38 model mampu mengklasifikasikan kelas *Cataract* dan *Glaucoma* dengan sangat baik, sementara kelas *Diabetic* menunjukkan performa yang cukup stabil. Kelas *Normal* masih menjadi kelas yang paling rendah karena kemiripan pola fitur dengan kelas lain. Secara keseluruhan, Skenario 16 memperlihatkan keseimbangan performa antar kelas dan kestabilan model yang baik, meskipun peningkatannya tidak melampaui skenario terbaik sebelumnya.

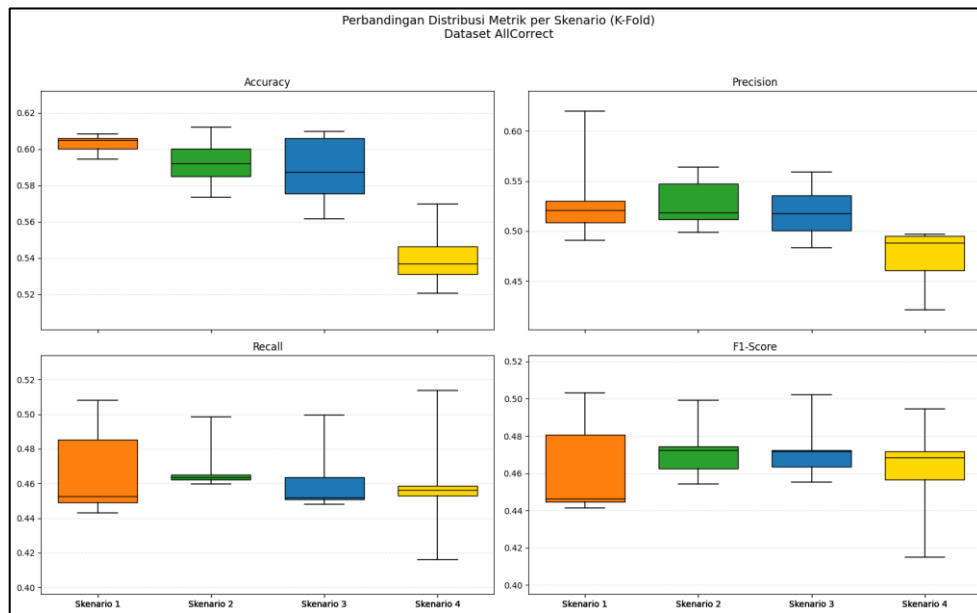
Tabel 4. 38 Rata-rata *Evaluation Matrix* per kelas

Kelas	Precision	Recall	F1-Score
Normal	65.57%	62.50%	63.91%
<i>Diabetic</i>	67.08%	70.00%	68.19%
<i>Glaucoma</i>	75.04%	74.00%	74.17%
<i>Cataract</i>	85.20%	85.25%	85.17%

4.4 Pembahasan

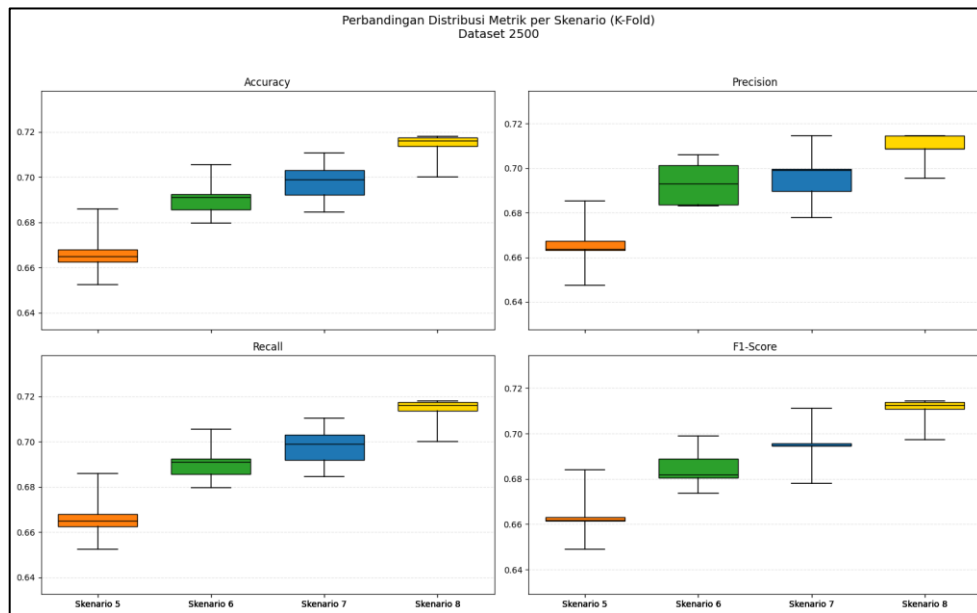
Setiap model dianalisis untuk melihat bagaimana perubahan pada kualitas, jumlah, serta distribusi data, maupun konfigurasi MLP, memengaruhi kemampuan sistem dalam melakukan proses klasifikasi.

Perbandingan pada skenario 1 hingga 4 pada Gambar 4.17 menunjukkan bahwa penggunaan data yang masih tidak seimbang antar kelas menghasilkan distribusi performa yang relatif seragam dan cenderung stagnan di seluruh skenario. Hal ini terlihat dari boxplot akurasi yang memiliki median dan rentang nilai yang saling berdekatan, meskipun kompleksitas arsitektur ditingkatkan dari konfigurasi 32–16 hingga 128–64–32 neuron.



Gambar 4. 17 Perbandingan Model A (Data *Unbalanced*)

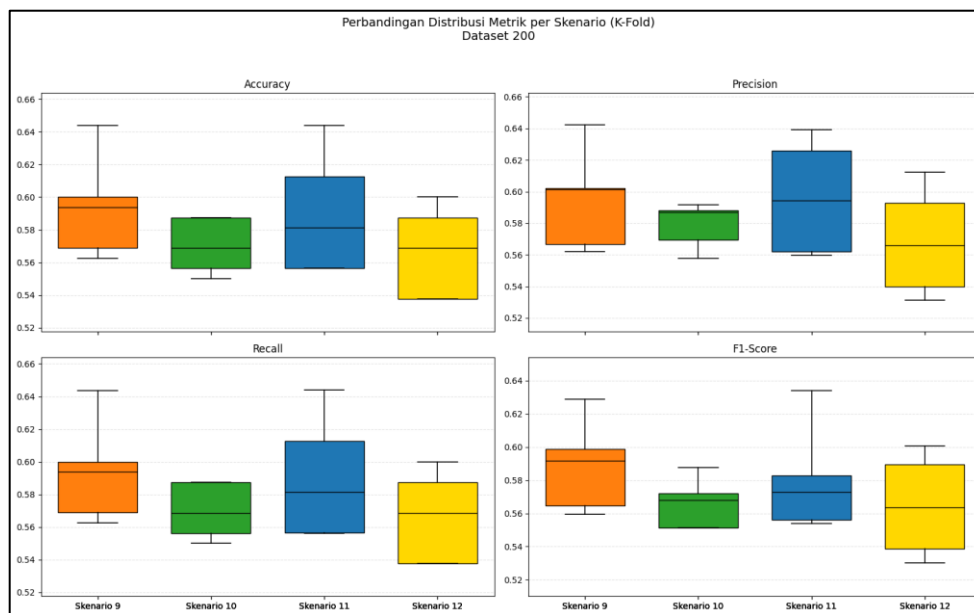
Peningkatan arsitektur tersebut tidak mampu mendorong kenaikan performa yang signifikan, karena akurasi tetap berada pada kisaran yang sempit. Pola serupa juga tampak pada metrik precision, recall, dan F1-score, yang secara konsisten berada di bawah nilai akurasi dan menunjukkan sebaran yang tidak jauh berbeda antar skenario. Distribusi recall dan F1-score yang relatif rendah mengindikasikan bahwa model belum mampu mengenali kelas minoritas secara seimbang. Kondisi ini mengonfirmasi bahwa ketidakseimbangan jumlah data dengan dominasi kelas *Normal* dan *Diabetic* menyebabkan model lebih berfokus pada kelas mayoritas, sehingga representasi fitur untuk kelas *Cataract* dan *Glaucoma* tetap lemah. Akibatnya, peningkatan kompleksitas jaringan tidak memberikan dampak terhadap kenaikan performa selama permasalahan ketidakseimbangan data belum ditangani.



Gambar 4. 18 Perbandingan Model B (Augmentasi pada Data Unbalanced)

Perbandingan pada skenario 5–8 pada Gambar 4.18 menunjukkan adanya peningkatan performa yang lebih konsisten dibandingkan skenario sebelumnya yang masih menggunakan data tidak seimbang. Hal ini terlihat dari pergeseran median dan penyempitan rentang boxplot pada metrik akurasi, precision, recall, dan F1-score yang cenderung meningkat secara bertahap dari skenario 5 hingga skenario 8. Distribusi nilai pada setiap metrik juga tampak lebih stabil dengan variasi yang relatif kecil antar fold, menandakan bahwa model mampu melakukan generalisasi dengan lebih baik. Peningkatan ini menandakan bahwa penggunaan dataset berukuran besar dan seimbang melalui augmentasi (2500 data per kelas) memberikan dampak positif dalam memperkaya representasi fitur yang dipelajari model. Selain itu, kenaikan performa yang konsisten pada berbagai konfigurasi arsitektur menunjukkan bahwa augmentasi data berperan penting dalam mengurangi bias terhadap kelas tertentu dan meningkatkan keseimbangan performa

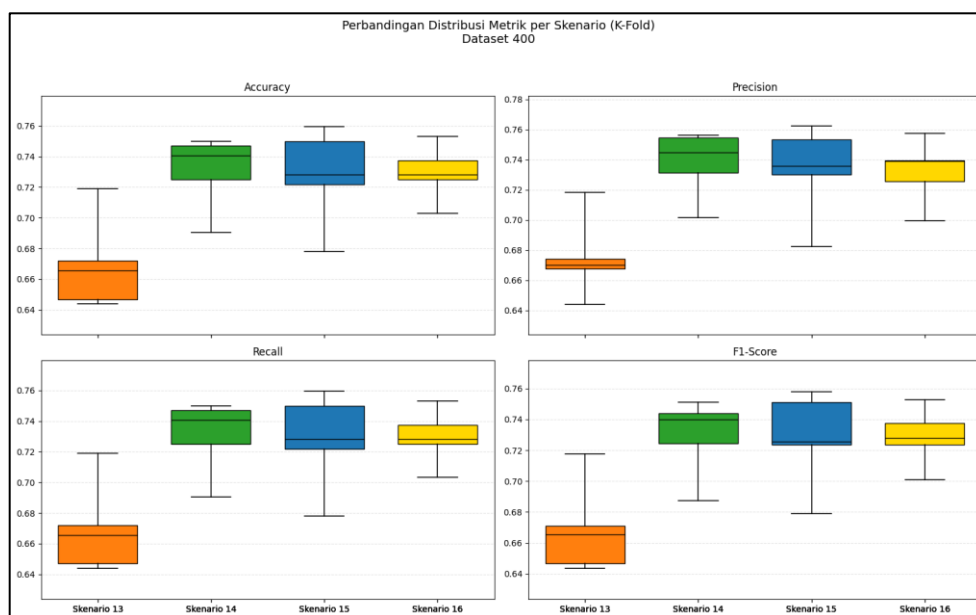
antar kelas, sehingga model tidak lagi terlalu bergantung pada kelas mayoritas seperti pada skenario sebelumnya.



Gambar 4. 19 Perbandingan Model C (Data Undersampling)

Performa model skenario 9–12 pada Gambar 4.19, menggunakan dataset hasil *undersampling* menjadi 200 data per kelas, cenderung berada pada tingkat menengah dengan variasi yang relatif lebih besar dibandingkan skenario dengan jumlah data lebih banyak. Distribusi akurasi, precision, recall, dan F1-score tampak saling berdekatan antar skenario, menandakan bahwa perubahan konfigurasi arsitektur belum memberikan peningkatan performa yang signifikan. Pada beberapa metrik, khususnya akurasi dan recall, terlihat rentang nilai yang cukup lebar pada skenario 11 dan 12, yang menandakan ketidakstabilan performa antar fold akibat keterbatasan jumlah data. Meskipun *undersampling* berhasil menghilangkan bias kelas, pengurangan jumlah sampel menyebabkan model kesulitan menangkap kompleksitas pola citra fundus secara konsisten. Akibatnya, peningkatan

kompleksitas arsitektur tidak selalu diikuti oleh perbaikan performa, dan secara keseluruhan kinerja model pada dataset 200 data per kelas masih dibatasi oleh minimnya variasi data yang tersedia.



Gambar 4. 20 Perbandingan Model D (Data *Undersampling* + Augmentasi)

Performa model skenario 13-16 pada Gambar 4.20, menunjukkan bahwa penggunaan dataset 400 data per kelas menghasilkan peningkatan performa yang lebih konsisten dibandingkan skenario dengan jumlah data lebih sedikit. Pada seluruh metrik akurasi, precision, recall, dan F1-score terlihat bahwa skenario 14, 15, dan 16 memiliki median nilai yang lebih tinggi dibandingkan skenario 13, menandakan bahwa penambahan kapasitas arsitektur yang didukung oleh jumlah data yang cukup mampu meningkatkan kemampuan generalisasi model. Distribusi nilai pada skenario 14 relatif lebih sempit, yang mengindikasikan performa yang stabil antar fold, sedangkan skenario 15 menunjukkan rentang yang sedikit lebih lebar akibat peningkatan kompleksitas arsitektur. Skenario 16 tetap

mempertahankan performa yang tinggi dengan variasi yang terkendali, menunjukkan keseimbangan antara kapasitas model dan jumlah data. Secara keseluruhan, pola boxplot ini menegaskan bahwa kombinasi dataset seimbang dengan ukuran menengah dan arsitektur MLP yang proporsional mampu menghasilkan performa klasifikasi yang lebih optimal dan stabil pada berbagai metrik evaluasi.

Berikut merupakan rangkuman hasil evaluasi k-fold cross validation yang mencakup nilai akurasi terendah, tertinggi, rata-rata, serta standar deviasi pada setiap skenario pengujian. Penyajian ini bertujuan untuk memberikan gambaran performa dan kestabilan model terhadap variasi data pada tiap fold. Evaluasi tidak hanya berfokus pada capaian akurasi terbaik, tetapi juga pada konsistensi kinerja model yang ditunjukkan melalui nilai standar deviasi, di mana nilai yang lebih kecil menandakan performa yang lebih stabil.

Tabel 4. 39 Hasil *K-Fold Cross Validation*

No Skenario	Kode Skenario	Akurasi K-Fold Terendah	Akurasi K-Fold Tertinggi	Rata-rata Akurasi	Standar Deviasi
1	A-2H-32	59.44%	60.84%	60.28%	0,0056
2	A-2H-64	57.33%	61.20%	59.25%	0,0147
3	A-3H-64	56.15%	60.96%	58.80%	0,0203
4	A-3H-128	52.05%	56.98%	54.09%	0,0186
5	B-2H-32	65.25%	68.60%	66.68%	0,0122
6	B-2H-64	67.95%	70.55%	69.08%	0,0097
7	B-3H-64	68.45%	71.05%	69.78%	0,0100
8	B-3H-128	70%	71.80%	71.30%	0,0075
9	C-2H-32	56.25%	64.38%	59.38%	0,0322
10	C-2H-64	55%	58.75%	57.00%	0,0173
11	C-3H-64	55.63%	64.38%	59.00%	0,0379
12	C-3H-128	53.75%	60%	56.63%	0,0285
13	D-2H-32	64.38%	71.88%	66.94%	0,0301
14	D-2H-64	69.06%	75%	73.06%	0,0243
15	D-3H-64	67.81%	75.94%	72.75%	0,0316
16	D-3H-128	70.31%	75.31%	72.94%	0,0183

Berdasarkan Tabel 4.39 terlihat bahwa performa model meningkat secara bertahap seiring dengan perbaikan perlakuan data dan penyesuaian arsitektur jaringan. Model A yang masih menggunakan data tidak seimbang menunjukkan rata-rata akurasi terendah dan standar deviasi yang relatif kecil, menandakan model stabil tetapi terbatas oleh kualitas data. Peningkatan mulai tampak pada Model B yang memanfaatkan augmentasi data, di mana rata-rata akurasi meningkat secara konsisten dengan standar deviasi yang lebih rendah, menunjukkan kestabilan sekaligus peningkatan kemampuan generalisasi. Pada Model C, meskipun dataset telah diseimbangkan melalui *undersampling*, rata-rata akurasi cenderung menurun dan standar deviasi membesar, mengindikasikan bahwa pengurangan jumlah data berdampak pada menurunnya kemampuan model dalam menangkap pola yang kompleks. Performa terbaik diperoleh pada Model D, khususnya Skenario 14 (D-2H-64) dan Skenario 16 (D-3H-128), yang mencatat rata-rata akurasi tertinggi dengan standar deviasi relatif kecil, menunjukkan kombinasi optimal antara keseimbangan data, jumlah sampel yang memadai, dan arsitektur jaringan yang proporsional.

Meskipun kedua skenario tersebut menunjukkan performa yang unggul dan stabil, diperlukan analisis lebih lanjut untuk memastikan apakah perbedaan kinerja di antara Skenario 14 dan Skenario 16 bersifat signifikan secara statistik. Uji signifikansi menggunakan paired t-test terhadap hasil akurasi per fold dari kedua skenario terbaik tersebut dilakukan untuk menilai apakah perbedaan performa yang

dihasilkan benar-benar bermakna atau hanya terjadi secara kebetulan. Pada uji signifikansi antara 2 skenario ini nilai alpha yang digunakan adalah 0.05.

Tabel 4. 40 Hasil Uji Signifikansi

Metrik	Mean A	Mean B	p-value	Keputusan
Akurasi	0.7306	0.7294	0.8638	Tidak Signifikan
F1-Score	-	-	0.9183	Tidak Signifikan

Hasil uji paired t-test pada Tabel 4.40 menunjukkan bahwa perbedaan performa antara Skenario 14 dan Skenario 16 tidak signifikan secara statistik, dengan nilai p-value sebesar 0,8638 untuk akurasi dan 0,9183 untuk F1-score macro ($p > 0.05$). Hal ini menunjukkan bahwa peningkatan kompleksitas arsitektur pada Skenario 16 tidak memberikan peningkatan performa yang bermakna dibandingkan Skenario 14. Dengan demikian, pemilihan model yang lebih sederhana menjadi lebih rasional karena memberikan performa yang sebanding dengan kompleksitas komputasi yang lebih rendah.

4.4.1 Pengaruh Perlakuan Data

Perbedaan perlakuan data yang menunjukkan bahwa variasi pada proses *data cleaning*, *balancing data* menggunakan *undersampling*, dan augmentasi data menghasilkan perubahan yang signifikan terhadap performa model.

Tabel 4. 41 Rata-rata akurasi pada Model A (*Unbalanced*)

No	Model A (<i>Unbalance</i>)		
	Skenario	Akurasi	F1-score
1	Skenario 1	60.28%	46.32%
2	Skenario 2	59.25%	47.25%
3	Skenario 3	58.80%	47.30%
4	Skenario 4	54.09%	46.12%
	Rata-rata	58.11%	46.75%

Performa Model A pada Tabel 4.41 menggunakan data hasil cleaning tanpa penyeimbangan jumlah sampel antar kelas, sehingga *Normal* dan *Diabetic* tetap mendominasi dibandingkan *Glaucoma* dan *Cataract*. Terlihat bahwa performa model masih relatif rendah dan tidak stabil, dengan rata-rata akurasi sebesar 58,11% dan F1-score 46,75%. Kondisi ini menunjukkan dampak negatif dari penggunaan data yang belum seimbang antar kelas, di mana model cenderung lebih mudah mengenali kelas mayoritas dan kesulitan mempelajari pola pada kelas minoritas. Hal tersebut tercermin dari nilai F1-score yang konsisten lebih rendah dibanding akurasi, menandakan ketimpangan performa antar kelas. Penurunan akurasi yang cukup signifikan pada skenario 4 juga mengindikasikan bahwa peningkatan kompleksitas arsitektur tanpa perbaikan distribusi data tidak mampu meningkatkan kinerja model. Dengan demikian, perlakuan data berupa data cleaning saja belum cukup untuk menghasilkan model klasifikasi yang andal, sehingga diperlukan strategi lanjutan seperti penyeimbangan data atau augmentasi untuk meningkatkan kemampuan generalisasi model.

Tabel 4. 42 Rata-rata Akurasi pada Model B (Data Augmentasi)

No	Model B (Data Augmentasi)		
	Skenario	Akurasi	F1-score
1	Skenario 5	66.68%	66.39%
2	Skenario 6	69.08%	68.48%
3	Skenario 7	69.78%	69.49%
4	Skenario 8	71.30%	70.90%
	Rata-rata	69.21%	68.83%

Penerapan augmentasi pada Model B ditunjukkan pada Tabel 4.42 menggunakan data hasil cleaning yang kemudian diperluas melalui augmentasi hingga mencapai 2.500 sampel per kelas, sehingga distribusi data menjadi lebih

seimbang dan bervariasi. Penambahan jumlah data ini memberi dampak pada performa model terlihat adanya peningkatan performa dibandingkan Model A yang hanya menggunakan data hasil cleaning. Rata-rata akurasi meningkat menjadi 69,21% dengan F1-score 68,83%, menunjukkan bahwa augmentasi mampu memperkaya variasi data dan membantu model mempelajari pola citra fundus secara lebih representatif. Kenaikan akurasi dan F1-score yang konsisten dari skenario 5 hingga 8 mengindikasikan bahwa penambahan data berhasil mengurangi bias terhadap kelas mayoritas dan memperbaiki kemampuan model dalam mengenali kelas minoritas. Selain itu, selisih antara akurasi dan F1-score yang semakin kecil menandakan distribusi performa antar kelas menjadi lebih seimbang. Dengan demikian, perlakuan data berupa augmentasi pada data yang belum seimbang terbukti efektif dalam meningkatkan kemampuan generalisasi model, meskipun ketidakseimbangan kelas secara struktural masih belum sepenuhnya diatasi.

Tabel 4. 43 Rata-rata Akurasi pada Model C (Data Undersampling)

No	Model C (Data Undersampling)		
	Skenario	Akurasi	F1-Score
1	Skenario 9	59.38%	58.87%
2	Skenario 10	57.00%	56.60%
3	Skenario 11	59.00%	57.98%
4	Skenario 12	56.63%	56.44%
	Rata-rata	58.00%	57.47%

Penerapan *undersampling* pada Model C ditunjukkan pada Tabel 4.43 menggunakan data hasil cleaning yang kemudian di-undersampling menjadi 200 data per kelas, sehingga distribusi data menjadi seimbang. Perlakuan ini bertujuan untuk menghilangkan bias kelas mayoritas yang sering muncul pada model

sebelumnya, sehingga setiap kelas memiliki peluang yang sama untuk dipelajari oleh model. Namun, konsekuensinya adalah jumlah data menjadi jauh lebih sedikit, sehingga ragam informasi visual yang tersedia untuk proses pembelajaran ikut berkurang.

Hal ini terlihat pada performa model cenderung menurun dibandingkan Model B dan tidak menunjukkan peningkatan berarti dibandingkan Model A. Rata-rata akurasi sebesar 58,00% dan F1-score 57,47% mengindikasikan bahwa meskipun *undersampling* berhasil menciptakan keseimbangan jumlah data antar kelas, pengurangan jumlah sampel secara signifikan justru membatasi kemampuan model dalam mempelajari karakteristik fitur citra fundus secara mendalam. Hal ini menyebabkan representasi pola penyakit menjadi kurang kaya, sehingga model kesulitan membedakan kelas dengan baik. Dengan demikian, perlakuan data berupa *undersampling* saja belum cukup efektif, karena keseimbangan data perlu diimbangi dengan jumlah sampel yang memadai agar model mampu mencapai performa klasifikasi yang optimal.

Model C membuktikan bahwa keseimbangan data saja belum cukup untuk meningkatkan performa secara signifikan apabila jumlah data terlalu kecil. *Undersampling* memang mengurangi bias, tetapi juga mengurangi keberagaman fitur yang diperlukan model untuk membedakan pola antar kelas secara optimal.

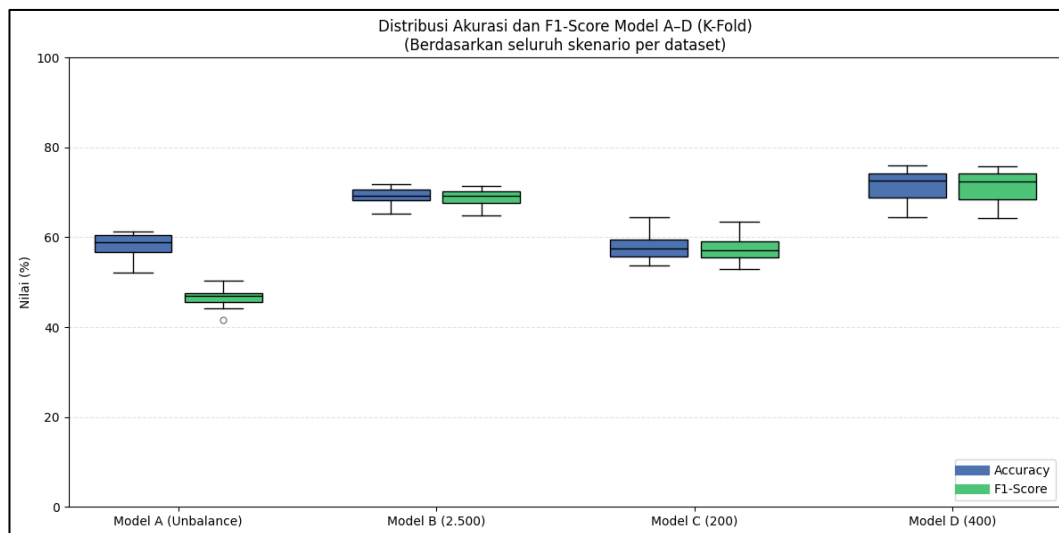
Tabel 4. 44 Rata-rata Akurasi pada Model D (Data *Undersampling* + Augmentasi)

No	Model D (Data <i>Undersampling</i> + Augmentasi)		
	Skenario	Akurasi	F1-score
1	Skenario 13	66.94%	66.88%
2	Skenario 14	73.06%	72.93%
3	Skenario 15	72.75%	72.75%
4	Skenario 16	72.93%	72.86%
	Rata-rata	71.42%	71.35%

Performa dari penerapan kombinasi *undersampling* dan augmentasi pada Model D ditunjukkan pada Tabel 4.44 menunjukkan peningkatan performa yang paling tinggi dibandingkan model lainnya. Setelah data diseimbangkan melalui *undersampling* dan ditambahkan dengan augmentasi menjadi 400 data untuk tiap kelas, model memperoleh rata-rata akurasi 71,42% dengan F1-score 71,35%, menunjukkan bahwa model tidak hanya lebih akurat tetapi juga memiliki performa yang lebih seimbang antar kelas. Kombinasi keseimbangan data melalui *undersampling* dan peningkatan jumlah serta variasi data melalui augmentasi memungkinkan model mempelajari pola fitur penyakit mata secara lebih representatif tanpa kehilangan informasi penting. Hasil ini menegaskan bahwa keseimbangan kelas saja tidak cukup, tetapi harus disertai dengan jumlah data yang memadai dan keragaman citra, sehingga Model D menjadi perlakuan data yang paling efektif dalam meningkatkan kemampuan generalisasi MLP pada klasifikasi citra fundus.

Berdasarkan perbandingan rata-rata akurasi pada Model A hingga Model D yang ditunjukkan pada Gambar 4.21, perlakuan data memiliki pengaruh yang lebih dominan dibandingkan variasi arsitektur dalam menentukan performa model. Model dengan data tidak seimbang menunjukkan keterbatasan dalam menjaga konsistensi performa antar kelas, sedangkan augmentasi data mampu meningkatkan akurasi sekaligus F1-score secara signifikan. Namun, *undersampling* dengan jumlah data yang terlalu sedikit justru menurunkan kemampuan generalisasi model. Kombinasi data seimbang dan augmentasi pada Model D menghasilkan performa

paling optimal dan stabil, yang ditunjukkan oleh tingginya nilai akurasi dan F1-score serta sebaran yang relatif rapat. Hal ini menegaskan bahwa keseimbangan dan kecukupan data merupakan kunci utama dalam meningkatkan keberhasilan klasifikasi citra fundus.



Gambar 4. 21 Perbandingan Perlakuan Data

4.4.2 Pengaruh Arsitektur MLP

Pengaruh arsitektur MLP terhadap performa model dianalisis melalui variasi jumlah *hidden layer* dan konfigurasi jumlah neuron untuk mengetahui sejauh mana kapasitas jaringan memengaruhi kemampuan model dalam mempelajari pola pada citra fundus.

Tabel 4. 45 Akurasi pada arsitektur 2 *Hidden Layer*

No	32-16			64-32		
	Skenario	Akurasi	F1-score	Skenario	Akurasi	F1-score
1	Skenario 1	60.28%	46.32%	Skenario 2	59.25%	47.25%
2	Skenario 5	66.68%	66.39%	Skenario 6	69.08%	68.48%
3	Skenario 9	59.38%	58.87%	Skenario 10	57%	56.61%
4	Skenario 13	66.94%	66.88%	Skenario 14	73.06%	72.93%
	Rata-rata	63.32%	59.62%	Rata-rata	64.60%	61.32%

Konfigurasi 2 hidden layer yang ditunjukkan pada Tabel 4.45, menunjukkan bahwa arsitektur 32–16 menghasilkan rata-rata akurasi sebesar 63.32% dan F1-score 59.62%, sedangkan arsitektur 64–32 memberikan performa lebih tinggi dengan rata-rata akurasi 64.60% dan F1-score 61.32%. Peningkatan ini menunjukkan bahwa penambahan jumlah neuron memberikan kapasitas representasi fitur yang lebih baik, sehingga model mampu mengenali pola citra fundus dengan lebih akurat dan konsisten. Meski begitu, selisih peningkatan tidak terlalu besar, menandakan bahwa manfaat dari meningkatkan jumlah neuron hanya efektif ketika didukung oleh ketersediaan data yang memadai. Secara keseluruhan, arsitektur 64–32 sedikit lebih stabil dan unggul baik dari sisi akurasi maupun F1-score, menandakan kemampuan model dalam menjaga keseimbangan antara prediksi benar pada berbagai kelas.

Tabel 4. 46 Akurasi pada arsitektur 3 *Hidden Layer*

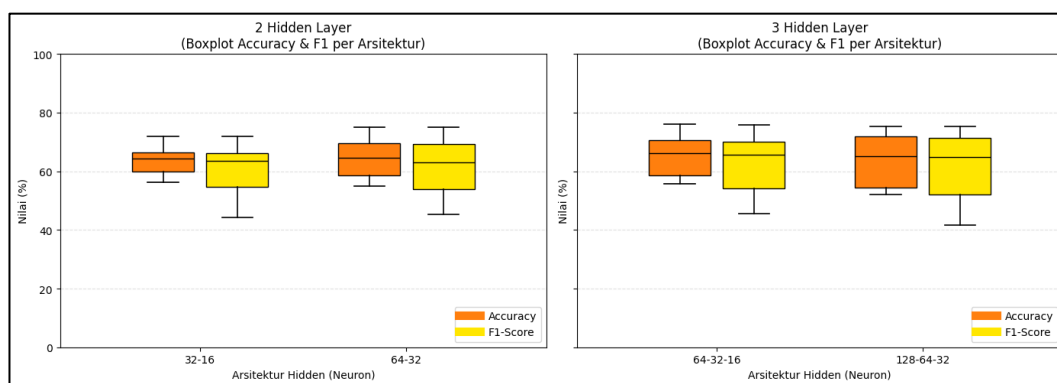
No	64-32-16			128-64-32		
	Skenario	Akurasi	F1-score	Skenario	Akurasi	F1-score
1	Skenario 3	58.80%	47.30%	Skenario 4	54.09%	46.12%
2	Skenario 7	69.78%	69.49%	Skenario 8	71.30%	70.98%
3	Skenario 11	59%	57.99%	Skenario 12	56.63%	56.44%
4	Skenario 15	72.75%	72.75%	Skenario 16	72.94%	72.86%
	Rata-rata	65.08%	61.88%	Rata-rata	63.73%	61.60%

Arsitektur 3 hidden layer yang ditunjukkan pada Tabel 4.46, menunjukkan bahwa konfigurasi 64–32–16 menghasilkan rata-rata akurasi sebesar 65.08% dan F1-score 61.88%. Sementara itu, arsitektur yang lebih besar, yaitu 128–64–32, menunjukkan rata-rata akurasi yang sedikit lebih rendah sebesar 63,73% dan F1-score 61,60%. Meskipun arsitektur 128–64–32 mampu mencapai performa tinggi

pada beberapa skenario tertentu, variasi nilai yang lebih besar menyebabkan rata-rata performanya berada di bawah konfigurasi 64–32–16.

Hal menunjukkan bahwa peningkatan jumlah neuron pada arsitektur 3 hidden layer tidak secara otomatis meningkatkan performa model. Arsitektur yang lebih kompleks cenderung memerlukan dukungan data yang lebih besar dan seimbang agar dapat bekerja secara optimal. Sebaliknya, konfigurasi yang lebih sederhana seperti 64–32–16 mampu memberikan performa yang lebih stabil ketika data telah diproses dan diperlakukan dengan baik.

Perbandingan rata-rata akurasi pada berbagai arsitektur yang ditunjukkan pada Gambar 4.22, terlihat bahwa peningkatan jumlah hidden layer tidak secara konsisten menghasilkan peningkatan performa model. Pada arsitektur dengan 2 hidden layer, konfigurasi 64–32 menunjukkan median akurasi yang sedikit lebih tinggi dibandingkan arsitektur 32–16, dengan sebaran nilai antar fold yang relatif serupa. Hal ini mengindikasikan adanya peningkatan performa, namun tidak signifikan secara stabilitas.



Gambar 4. 22 Perbandingan Arsitektur MLP

Pada arsitektur 3 hidden layer, pola peningkatan performa tidak bersifat linear. Arsitektur 64–32–16 menunjukkan median akurasi yang lebih rendah dan

sebaran nilai yang lebih lebar dibandingkan arsitektur 128–64–32, yang memiliki median akurasi lebih tinggi serta distribusi yang lebih stabil. Temuan ini menegaskan bahwa penambahan jumlah layer tidak otomatis meningkatkan akurasi, dan arsitektur yang lebih kompleks baru memberikan manfaat ketika didukung oleh kapasitas representasi yang memadai serta kondisi data yang cukup mendukung.

Pengembangan model klasifikasi citra fundus pada penelitian ini tidak hanya berfokus pada penyajian hasil evaluasi atau grafik performa, tetapi diawali dengan rangkaian proses perancangan yang terukur. Setiap skenario dirancang dengan penyesuaian parameter seperti jumlah hidden layer, konfigurasi neuron, jumlah epoch, serta strategi perlakuan data, sehingga memungkinkan analisis yang komprehensif mengenai pengaruh masing-masing faktor terhadap kemampuan model dalam mengenali pola penyakit mata pada dataset ODIR. Pendekatan ini menegaskan bahwa performa model bukanlah hasil dari proses coba-coba, melainkan dari serangkaian keputusan yang berpijak pada prinsip keteraturan dan keseimbangan. Hal ini sejalan dengan nilai-nilai Islam yang mengajarkan bahwa segala sesuatu di alam semesta memiliki ukuran dan ketentuan yang telah ditetapkan, sebagaimana tercantum dalam QS. Al-Qamar ayat 49 yang berbunyi:

إِنَّا كُلَّ شَيْءٍ خَلَقْنَاهُ بِقَدَرٍ ﴿٤٩﴾

“Sesungguhnya Kami menciptakan segala sesuatu sesuai dengan ukuran.” (QS Al-Qamar: 49)”

Tafsir para ulama terhadap QS. Al-Qamar ayat 49 menegaskan bahwa seluruh ciptaan Allah berjalan berdasarkan ukuran, ketetapan, dan ketelitian yang

sangat presisi. Ibn Katsir dan Al-Tabari menjelaskan bahwa “*bi qadar*” berarti segala sesuatu diatur sesuai takaran yang paling tepat, tanpa kelebihan maupun kekurangan. Pesan ini memiliki relevansi filosofis dalam penelitian ini, terutama dalam proses mencari konfigurasi terbaik pada model klasifikasi. Sebagaimana ciptaan Allah berjalan atas dasar ukuran yang optimal, performa model juga sangat bergantung pada parameter yang terukur dimulai dari jumlah neuron, hidden layer, epoch, hingga keseimbangan data yang semuanya berpengaruh dalam menentukan akurasi dan stabilitas prediksi. Dengan demikian, optimasi model melalui berbagai skenario bukan hanya proses teknis, tetapi juga bentuk ikhtiar untuk menemukan hasil akhir sebagaimana prinsip ketelitian dan keteraturan yang diajarkan dalam ayat tersebut.

Ketelitian dan keteraturan sebagaimana ditegaskan dalam QS. Al-Qamar ayat 49 tidak hanya bersifat konseptual, tetapi juga tercermin dalam cara Allah menciptakan keberagaman yang memiliki karakteristik masing-masing dan tidak diseragamkan. Prinsip ini sejalan dengan QS. Al-Hujurāt ayat 13 yang berbunyi:

يَا أَيُّهَا النَّاسُ إِنَّا خَلَقْنَاكُمْ مِنْ ذَكَرٍ وَأُنْثَىٰ وَجَعَلْنَاكُمْ شُعُوبًا وَقَبَائِلَ لِتَعَارَفُوا إِنَّ أَكْرَمَكُمْ عِنْدَ اللَّهِ أَتْقَاهُ إِنَّ اللَّهَ عَلِيمٌ خَبِيرٌ ﴿١٣﴾

“Wahai manusia, sesungguhnya Kami telah menciptakan kamu dari seorang laki-laki dan perempuan. Kemudian, Kami menjadikan kamu berbangsa-bangsa dan bersuku-suku agar kamu saling mengenal. Sesungguhnya yang paling mulia di antara kamu di sisi Allah adalah orang yang paling bertakwa. Sesungguhnya Allah Maha Mengetahui lagi Mahateliti.” (QS Hujurat: 13)

Menurut Tafsir Ibn Katsir, Allah SWT menjelaskan bahwa seluruh manusia berasal dari satu asal yang sama, yaitu dari Nabi Adam dan Hawa. Hal ini menegaskan bahwa tidak ada perbedaan derajat kemanusiaan berdasarkan asal-

usul, suku, bangsa, atau ras. Perbedaan yang diciptakan Allah berupa bangsa dan suku bertujuan agar manusia dapat saling mengenal (*li-ta 'ārafū*), sehingga masing-masing individu atau kelompok dapat dibedakan identitasnya tanpa adanya unsur penghinaan atau kesombongan. Ibn Katsir menekankan bahwa keberagaman tersebut bukan untuk membanggakan nasab atau kedudukan, melainkan sebagai sarana pengenalan dan keteraturan sosial. Kemuliaan seseorang di sisi Allah tidak ditentukan oleh kategori lahiriah, tetapi oleh ketakwaan, sementara Allah Maha Mengetahui seluruh kondisi manusia, baik yang tampak maupun yang tersembunyi.

Makna tafsir tersebut relevan dengan konsep klasifikasi dalam penelitian ini. Sebagaimana keberagaman manusia diciptakan agar dapat dikenali dan dibedakan secara tepat tanpa menyamaratakan atau mendiskreditkan, proses klasifikasi penyakit mata juga bertujuan untuk mengenali perbedaan karakteristik visual pada citra fundus secara objektif. Setiap kondisi mata, seperti *Normal*, *Diabetic*, *Glaucoma*, dan *Cataract*, memiliki ciri khas yang perlu dipisahkan ke dalam kelas yang sesuai berdasarkan pola warna, tekstur, dan struktur pembuluh darah. Klasifikasi ini tidak dimaksudkan untuk menyamakan seluruh citra, melainkan untuk mengidentifikasi perbedaan yang ada secara terukur dan adil, sehingga mendukung proses diagnosis yang lebih akurat dan bertanggung jawab.

Keberagaman karakteristik visual tersebut memerlukan kejelasan batas agar dapat dikenali dan dibedakan secara tepat. Prinsip ini sejalan dengan penjelasan Rasulullah ﷺ dalam menafsirkan sebagian QS. Al-Baqarah ayat 187 yang berbunyi:

...وَكُلُوا وَاشْرَبُوا حَتَّى يَتَبَيَّنَ لَكُمُ الْخَيْطُ الْأَبْيَضُ مِنَ الْخَيْطِ الْأَسْوَدِ مِنَ الْفَجْرِ...

“...Makan dan minumlah hingga jelas bagimu (perbedaan) antara benang putih dan benang hitam, yaitu fajar. Kemudian, sempurnakanlah puasa sampai (datang) malam. ...” (QS Al-Baqarah: 187)

Ayat tersebut kemudian dijelaskan oleh Rasulullah ﷺ dalam sebuah hadis yang diriwayatkan oleh Imam al-Bukhari dan Imam Muslim dari sahabat ‘Adi bin Hatim r.a. Rasulullah ﷺ menegaskan bahwa yang dimaksud dengan “benang putih” dan “benang hitam” bukanlah benda fisik, melainkan batas visual antara cahaya fajar dan gelapnya malam. Penafsiran ini menunjukkan bahwa penentuan suatu kondisi harus didasarkan pada kejelasan perbedaan visual yang nyata dan dapat dibedakan secara objektif, bukan semata-mata pada persepsi subjektif.

Makna ini berkaitan dengan karakteristik citra fundus, variasi warna retina dan pola pembuluh darah menjadi indikator penting dalam membedakan kondisi mata normal dan abnormal. Perubahan warna pada area tertentu, intensitas kemerahan, maupun kepekatan struktur pembuluh darah mencerminkan kondisi biologis yang berbeda dan memerlukan pemisahan intensitas yang jelas agar dapat dianalisis secara akurat. Dengan demikian, pemanfaatan fitur warna dan tekstur dalam citra fundus dapat dipahami sebagai upaya ilmiah untuk menganalisis tanda-tanda visual yang telah diciptakan Allah dengan ukuran dan ketelitian tertentu, selaras dengan prinsip *bi qadar* yang menegaskan bahwa setiap detail ciptaan memiliki makna dan fungsi.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan seluruh rangkaian penelitian yang meliputi proses pre-processing citra fundus, ekstraksi fitur GLCM dan Gabor, pengembangan arsitektur Multi-Layer Perceptron (MLP), serta evaluasi performa melalui 16 skenario uji coba, dapat disimpulkan bahwa performa MLP dalam mengklasifikasikan penyakit mata pada dataset ODIR sangat bergantung pada kualitas dan keseimbangan data yang digunakan. Pada skenario awal dengan data tidak seimbang, model menunjukkan bias kuat terhadap kelas mayoritas sehingga akurasinya terbatas pada kisaran 54–60%. Setelah dilakukan perlakuan data berupa augmentasi dan penyeimbangan melalui undersampling, performa model meningkat sedikit. Perlakuan data terbaik terdapat pada Model D, yang mengombinasikan data seimbang melalui *undersampling* dan augmentasi sehingga mencapai rata-rata akurasi 71.42% dengan kestabilan F1-score pada seluruh kelas. Dari sisi arsitektur, konfigurasi MLP dengan dua hidden layer, khususnya arsitektur 64–32 neuron pada Skenario 14, menghasilkan performa tertinggi dengan rata-rata akurasi 73.06% dan mampu mengklasifikasikan keempat kelas penyakit mata secara merata. Hal ini menunjukkan bahwa MLP dapat menjadi metode klasifikasi yang efektif untuk citra fundus apabila didukung oleh data yang seimbang, bervariasi, dan diolah dengan baik, serta arsitektur jaringan yang dirancang secara proporsional terhadap kompleksitas pola pada data.

5.2 Saran

Penelitian ini telah menghasilkan model klasifikasi penyakit mata berdasarkan citra fundus menggunakan metode MLP. Namun, masih terdapat beberapa aspek yang dapat dikembangkan untuk meningkatkan kualitas dan keakuratan hasil penelitian di masa mendatang. Berikut saran yang dapat digunakan untuk pengembangan penelitian selanjutnya:

1. Menambah jumlah dataset dan memperluas keberagaman citra, sehingga model akan memperoleh representasi fitur yang lebih kuat jika dilatih menggunakan lebih banyak variasi citra fundus.
2. Mempertimbangkan kombinasi metode ekstraksi fitur lainnya, seperti *Local Binary Pattern (LBP)*, kombinasi fitur Haralick, atau *wavelet transform* dapat memberikan representasi tekstur yang lebih kuat bagi model.

Dengan menerapkan saran-saran tersebut, penelitian berikutnya diharapkan dapat menghasilkan model yang lebih akurat dan stabil untuk klasifikasi penyakit mata berdasarkan citra fundus.

DAFTAR PUSTAKA

- Aziz, A., Tezel, N. S., Kaçmaz, S., & Attallah, Y. (2025). Early Diabetic Retinopathy Detection from OCT Images Using Multifractal Analysis and Multi-Layer Perceptron Classification. *Diagnostics*, 15(13), 1616. <https://doi.org/10.3390/diagnostics15131616>
- Bhardwaj, C., Jain, S., & Sood, M. (2020). Diabetic Retinopathy Lesion Discriminative Diagnostic System for Retinal Fundus Images. *Advanced Biomedical Engineering*, 9(0), 71–82. <https://doi.org/10.14326/abe.9.71>
- Butt, M. M., Iskandar, D. N. F. A., Abdelhamid, S. E., Latif, G., & Alghazo, R. (2022). Diabetic Retinopathy Detection from Fundus Images of the Eye Using Hybrid Deep Learning Features. *Diagnostics*, 12(7), 1607. <https://doi.org/10.3390/diagnostics12071607>
- Chavan, R., & Pete, D. (2024). Automatic multi-disease classification on retinal images using multilevel glowworm swarm convolutional neural network. *Journal of Engineering and Applied Science*, 71(1), 26. <https://doi.org/10.1186/s44147-023-00335-0>
- Costaner, L., & Harun, N. (2024a). Fundus Image Feature Extraction in Diabetic Retinopathy: A Bibliometric Analysis Approach. *Proceedings of the 2nd International Conference on Environmental, Energy, and Earth Science, ICEEES 2023, 30 October 2023, Pekanbaru, Indonesia*. Proceedings of the 2nd International Conference on Environmental, Energy, and Earth Science, ICEEES 2023, 30 October 2023, Pekanbaru, Indonesia, Pekanbaru, Indonesia. <https://doi.org/10.4108/eai.30-10-2023.2343101>
- Costaner, L., & Harun, N. (2024b). Fundus Image Feature Extraction in Diabetic Retinopathy: A Bibliometric Analysis Approach. *Proceedings of the 2nd International Conference on Environmental, Energy, and Earth Science, ICEEES 2023, 30 October 2023, Pekanbaru, Indonesia*. Proceedings of the 2nd International Conference on Environmental, Energy, and Earth Science, ICEEES 2023, 30 October 2023, Pekanbaru, Indonesia, Pekanbaru, Indonesia. <https://doi.org/10.4108/eai.30-10-2023.2343101>
- Faghani, S., Khosravi, B., Zhang, K., Moassefi, M., Jagtap, J. M., Nugen, F., Vahdati, S., Kuanar, S. P., Rassoulinejad-Mousavi, S. M., Singh, Y., Vera Garcia, D. V., Rouzrokh, P., & Erickson, B. J. (2022). Mitigating Bias in Radiology Machine Learning: 3. Performance Metrics. *Radiology: Artificial Intelligence*, 4(5). <https://doi.org/10.1148/ryai.220061>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. The MIT Press.

- Heriz, H. H., Salah, H. M., Abdu, S. B. A., Sbihi, M. M. E., & Abu-Naser, S. (2018). *English Alphabet Prediction Using Artificial Neural Networks*. 2(11).
- Hidayat, F. T., Rahman, S., Turgut, E., & Azira, V. (2024). *Approaching the Maqasid al-Qur'an through the Story of Ayub and Its Impact on Ibn Ashur's Interpretation*. 1(2).
- Horie, S., & Ohno-Matsui, K. (2022). Progress of Imaging in Diabetic Retinopathy—From the Past to the Present. *Diagnostics*, 12(7), 1684. <https://doi.org/10.3390/diagnostics12071684>
- Ikram, A., Imran, A., Li, J., Alzubaidi, A., Fahim, S., Yasin, A., & Fathi, H. (2024). A Systematic Review on Fundus Image-Based Diabetic Retinopathy Detection and Grading: Current Status and Future Directions. *IEEE Access*, 12, 96273–96303. <https://doi.org/10.1109/ACCESS.2024.3427394>
- Indraswari, R., Herulambang, W., & Rokhana, R. (2022a). Deteksi Penyakit Mata Pada Citra Fundus Menggunakan Convolutional Neural Network (CNN). *Techno.Com*, 21(2), 378–389. <https://doi.org/10.33633/tc.v21i2.6162>
- Indraswari, R., Herulambang, W., & Rokhana, R. (2022b). Deteksi Penyakit Mata Pada Citra Fundus Menggunakan Convolutional Neural Network (CNN). *Techno.Com*, 21(2), 378–389. <https://doi.org/10.33633/tc.v21i2.6162>
- Jutel, A. (2011). Classification, Disease, and Diagnosis. *Perspectives in Biology and Medicine*, 54(2), 189–205. <https://doi.org/10.1353/pbm.2011.0015>
- Kaya, E., Sarıtaş, İ., & Ozkan, I. A. (2020). *Supervised Segmentation of Retinal Vessel Structures Using ANN* (Versi 1). arXiv. <https://doi.org/10.48550/ARXIV.2001.05549>
- Lee, J.-G., Jun, S., Cho, Y.-W., Lee, H., Kim, G. B., Seo, J. B., & Kim, N. (2017). Deep Learning in Medical Imaging: General Overview. *Korean Journal of Radiology*, 18(4), 570. <https://doi.org/10.3348/kjr.2017.18.4.570>
- Mansour, R. F., & Al-Marghilnai, A. (2021). Glaucoma detection using novel perceptron based convolutional multi-layer neural network classification. *Multidimensional Systems and Signal Processing*, 32(4), 1217–1235. <https://doi.org/10.1007/s11045-021-00781-0>
- Molina, E., & Parraga-Alava, J. (2024). Artificial Neural Networks for Classification Tasks: A Systematic Literature Review. *Enfoque UTE*, 15(4), 1–10. <https://doi.org/10.29019/enfoqueute.1058>
- Pathan, S., Kumar, P., Pai, R. M., & Bhandary, S. V. (2023). An automated classification framework for glaucoma detection in fundus images using ensemble of dynamic selection methods. *Progress in Artificial Intelligence*, 12(3), 287–301. <https://doi.org/10.1007/s13748-023-00304-x>

- Riina, N., Harris, A., Siesky, B. A., Ritzer, L., Pasquale, L. R., Tsai, J. C., Keller, J., Wirostko, B., Arciero, J., Fry, B., Eckert, G., Verticchio Vercellin, A., Antman, G., Sidoti, P. A., & Guidoboni, G. (2024). Using Multi-Layer Perceptron Driven Diagnosis to Compare Biomarkers for Primary Open Angle Glaucoma. *Investigative Ophthalmology & Visual Science*, 65(11), 16. <https://doi.org/10.1167/iovs.65.11.16>
- Saha, S. K., Xiao, D., & Kanagasingam, Y. (2018). A Novel Method for Correcting Non-uniform/Poor Illumination of Color Fundus Photographs. *Journal of Digital Imaging*, 31(4), 553–561. <https://doi.org/10.1007/s10278-017-0040-0>
- Saputra, W., Tulus, Zarlis, M., Sembiring, R. W., & Hartama, D. (2017). Analysis Resilient Algorithm on Artificial Neural Network Backpropagation. *Journal of Physics: Conference Series*, 930, 012035. <https://doi.org/10.1088/1742-6596/930/1/012035>
- Wadmann, S. (2023). Disease classification: A framework for analysis of contemporary developments in precision medicine. *SSM - Qualitative Research in Health*, 3, 100217. <https://doi.org/10.1016/j.ssmqr.2023.100217>
- World Health Organization. (2019). *World report on vision*. World Health Organization. <https://iris.who.int/handle/10665/328717>
- Wulandari, N. P., & Fitriana, D. (2021). *Analisa Perbandingan Algoritma CNN Dan MLP Dalam Mendeteksi Penyakit COVID-19 Pada Citra X-Ray Paru*.
- Yadav, H., & Mallick, S. (2024a). *Early Detection Of Cataract, Diabetic Retinopathy, And Glaucoma Using Deep Learning*. 12(12).
- Yadav, H., & Mallick, S. (2024b). *Early Detection Of Cataract, Diabetic Retinopathy, And Glaucoma Using Deep Learning*. 12(12).
- Yudono, M. A. S., Ridha, F. F., Mardiyana, D., Al-Ghozi, F., & Maulana, A. (2025). Klasifikasi Katarak Berdasarkan Optic Disc Citra Fundus Smartphone: Perbandingan Ekstraksi Ciri Tekstur Dan Metode Neural Network. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 12(1), 203–212. <https://doi.org/10.25126/jtiik.20251219254>
- Zai, A. R., Rambe, L. H., Putra, R. A., Rosnelly, R., Sagala, T. S., & Jaya, I. K. (2025). *OPTIMASI ALGORITMA GENETIKA PADA PERBANDINGAN ANN DAN KNN UNTUK KLASIFIKASI PENYAKIT JANTUNG*. <https://ejurnal.methodist.ac.id/index.php/methoda/article/view/3959/2269>

Zhang, R., Wang, L., Cheng, S., & Song, S. (2023). MLP-based classification of COVID-19 and skin diseases. *Expert Systems with Applications*, 228, 120389. <https://doi.org/10.1016/j.eswa.2023.120389>

LAMPIRAN

Lampiran I Implementasi Perhitungan Ekstraksi Fitur

Implementasi perhitungan GLCM

Contoh perhitungan *Contrast* dengan persamaan 3.1 berdasarkan contoh matriks probabilitas menghasilkan:

$$\begin{aligned}\text{Contrast} &= (0 - 0)^2 \cdot P(0,0) + (0 - 1)^2 \cdot P(0,1) + (1 - 0)^2 \cdot P(1,0) + (1 - 1)^2 \\ &\quad \cdot P(1,1) \\ &= 0 \cdot 0 + 1 \cdot \frac{1}{12} + 1 \cdot 0 + 0 \cdot \frac{11}{12} \\ &= \frac{1}{12} \approx 0.0833\end{aligned}$$

Nilai *contrast* yang relatif kecil 0.0833 menunjukkan bahwa pada matriks 4×4 terdapat sedikit variasi intensitas lokal pada arah yang diperiksa (lebih banyak pasangan berintensitas sama (1,1) daripada pasangan berbeda), sehingga teksturnya relatif homogen pada skala arah horizontal.

Contoh perhitungan *Homogeneity* dengan persamaan 3.2 berdasarkan contoh matriks probabilitas menghasilkan:

$$\begin{aligned}\text{Homogeneity} &= \frac{P(0,0)}{1 + |0 - 0|} + \frac{P(0,1)}{1 + |0 - 1|} + \frac{P(1,0)}{1 + |1 - 0|} + \frac{P(1,1)}{1 + |1 - 1|} \\ &= 0 + \frac{1}{24} + 0 + \frac{11}{12} \\ &= \frac{1}{24} + \frac{22}{24} \\ &= \frac{23}{24} \approx 0.95833\end{aligned}$$

Hasil homogeneity bernilai 0.9583 menunjukkan bahwa pada matriks 4 x 4 dengan arah horizontal yang diuji, teksturnya relatif homogen (banyak pasangan berintensitas sama (1,1)). Nilai Homogeneity mendekati 1 menandakan sedikit variasi antar-piksel pada arah tersebut.

Contoh perhitungan *Energy* dengan persamaan 3.3 berdasarkan contoh matriks probabilitas menghasilkan:

$$\begin{aligned}
 ASM &= P(0,0)^2 + P(0,1)^2 + P(1,0)^2 + P(1,1)^2 \\
 &= 0^2 + \left(\frac{1}{12}\right)^2 + 0^2 + \left(\frac{11}{12}\right)^2 \\
 &= \frac{1}{144} + \frac{121}{144} \\
 &= \frac{122}{144} = \frac{61}{72} \approx 0.84722
 \end{aligned}$$

$$Energy = \sqrt{ASM} = \sqrt{\frac{61}{72}} \approx 0.9204$$

Nilai $ASM \approx 0.8472$ (dan $Energy \approx 0.9204$) menunjukkan bahwa distribusi pasangan intensitas ini cukup teratur mayoritas pasangan adalah (1,1). Oleh karena itu fitur $ASM/Energy$ memberikan indikasi tekstur yang relatif homogen dan dominan pada nilai intensitas tertentu di arah yang diuji.

Contoh perhitungan *Correlation* dengan persamaan 3.4 berdasarkan contoh matriks probabilitas menghasilkan:

$$\mu_i = \sum_i \sum_j i \cdot P(i, j) = \frac{11}{12} \approx 0.9167$$

$$\mu_j = \sum_i \sum_j j \cdot P(i, j) = 1$$

$$\sigma_i = \sqrt{\sum_i \sum_j (i - \mu_i)^2 P(i, j)} \approx 0.276$$

$$\sigma_j = \sqrt{\sum_i \sum_j (j - \mu_j)^2 P(i, j)} = 0$$

$\Rightarrow$ Correlation tidak terdefinisi (pembagian dengan nol).

Correlation tidak dapat dihitung (tak terdefinisi) karena terjadi pembagian dengan nol, hal ini disebabkan karena probabilitas terkonsentrasi di $j = 1$, maka $\sigma_j = 0$ yang mana artinya tidak ada variasi pada arah kolom.

Contoh perhitungan *Dissimilarity* dengan persamaan 3.5 berdasarkan contoh matriks probabilitas menghasilkan:

$$\begin{aligned}
 &= (0 \cdot 0) + \left(1 \cdot \frac{1}{12}\right) + (1 \cdot 0) + \left(0 \cdot \frac{11}{12}\right) \\
 &= \frac{1}{12} \approx 0.0833
 \end{aligned}$$

Contoh perhitungan *Entropy* dengan persamaan 3.6 berdasarkan contoh matriks probabilitas menghasilkan:

$$\begin{aligned}
 &= - \left[\frac{1}{12} \cdot \log_2 \left(\frac{1}{12} \right) + \frac{11}{12} \cdot \log_2 \left(\frac{11}{12} \right) \right] \\
 &= - \left[\frac{1}{12} \cdot (-3.585) + \frac{11}{12} \cdot (-0.129) \right] \\
 &= -[-0.299 - 0.118] \\
 &= 0.417
 \end{aligned}$$

Implementasi perhitungan Filter Gabor

Berikut merupakan contoh implementasinya pada kernel gabor 3 x 3, dengan orientasi 0 derajat. Contoh parameter yang digunakan adalah $\sigma = 1$, $\lambda = 2$, $\theta = 0^\circ$, $\psi = 0$, $\gamma = 1$ maka persamaan menjadi,

$$g(x, y) = \exp \left(-\frac{x^2 + y^2}{2} \right) \cdot \cos(\pi x)$$

Koordinat diambil pada $(x, y \in \{-1, 0, 1\})$

Maka perhitungan dengan persamaan 3.7 untuk tiap elemen kernel adalah,

1. Untuk $((x, y) = (-1, -1))$:

$$r^2 = (-1)^2 + (-1)^2 = 2$$

$$\exp\left(-\frac{2}{2}\right) = e^{-1} = 0.367879$$

$$\cos(\pi(-1)) = \cos(-\pi) = -1$$

$$g(-1, -1) = 0.367879 \times (-1) = -0.367879$$

2. Untuk $((x, y) = (0, -1))$:

$$r^2 = 0^2 + (-1)^2 = 1$$

$$\exp\left(-\frac{1}{2}\right) = e^{-0.5} = 0.606531$$

$$\cos(\pi \cdot 0) = 1$$

$$g(0, -1) = 0.606531 \times 1 = 0.606531$$

3. Untuk $((x, y) = (1, -1))$:

$$r^2 = 1^2 + (-1)^2 = 2, \quad \exp(-1) = 0.367879$$

$$\cos(\pi \cdot 1) = \cos(\pi) = -1$$

$$g(1, -1) = 0.367879 \times (-1) = -0.367879$$

4. Untuk $((x, y) = (-1, 0))$:

$$r^2 = 1, \quad \exp(-0.5) = 0.60653$$

$$\cos(\pi(-1)) = -1$$

$$g(-1, 0) = 0.606531 \times (-1) = -0.606531$$

5. Untuk $((x, y) = (0, 0))$:

$$r^2 = 0, \quad \exp(0) = 1$$

$$\cos(0) = 1$$

$$g(0, 0) = 1 \times 1 = 1.000000$$

6. Untuk $((x, y) = (1, 0))$:

$$r^2 = 1, \quad \exp(-0.5) = 0.606531$$

$$\cos(\pi) = -1$$

$$g(1, 0) = -0.606531$$

7. Untuk $((x, y) = (-1, 1))$:

$$r^2 = 2, \quad \exp(-1) = 0.367879$$

$$\cos(-\pi) = -1$$

$$g(-1, 1) = -0.367879$$

8. Untuk $((x, y) = (0,1))$:

$$r^2 = 1, \quad \exp(-0.5) = 0.606531$$

$$\cos(0) = 1$$

$$g(0,1) = 0.606531$$

9. Untuk $((x, y) = (1,1))$:

$$r^2 = 2, \quad \exp(-1) = 0.367879$$

$$\cos(\pi) = -1$$

$$g(1,1) = -0.367879$$

Maka hasil akhir kernel gabor (3 x 3) adalah

$$G = \begin{bmatrix} -0.367879 & 0.606531 & -0.367879 \\ -0.606531 & 1 & -0.606531 \\ -0.367879 & 0.606531 & -0.367879 \end{bmatrix}$$

Matriks G menunjukkan hasil perhitungan kernel gabor yang kemudian akan dikonvolusikan dengan citra 4 x 4 yang akan menghasilkan matriks respons akhir sebagai berikut,

$$S = \begin{bmatrix} S(0,0) & S(0,1) \\ S(1,0) & S(1,1) \end{bmatrix} \approx \begin{bmatrix} -0.103637 & -0.471516 \\ -0.471516 & -0.471516 \end{bmatrix}$$

Maka implementasi pada persamaan 3.8 berdasarkan contoh matriks respons akhir menghasilkan:

$$\begin{aligned} \mu_R &= \frac{-0.103637 + (-0.471516) + (-0.471516) + (-0.471516)}{4} \\ &= \frac{-1.518185}{4} = -0.37954625 \end{aligned}$$

Maka implementasi pada persamaan 3.9 berdasarkan contoh matriks respons akhir menghasilkan:

$$\begin{aligned} \sigma^2 &= \frac{1}{4} [(-0.1036 - (-0.3795))^2 + (-0.4715 - (-0.3795))^2 + (-0.4715 - (-0.3795))^2 \\ &\quad + (-0.4715 - (-0.3795))^2] \\ \sigma^2 &= \frac{1}{4} [(0.2759)^2 + (-0.0920)^2 + (-0.0920)^2 + (-0.0920)^2] \end{aligned}$$

$$\sigma^2 = \frac{1}{4} [0.0761 + 0.0085 + 0.0085 + 0.0085]$$

$$\sigma^2 = \frac{0.1016}{4} = 0.0254$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{0.0254} \approx 0.1593$$

Maka implementasi pada persamaan 3.10 berdasarkan contoh matriks respons akhir menghasilkan:

$$\begin{aligned} E_R &= (-0.103637)^2 + 3 \times (-0.471516)^2 \\ &\approx 0.010744 + 3 \times 0.222327 \\ &= 0.010744 + 0.666981 \\ &= 0.677725 \end{aligned}$$

$$E_R^{(norm)} = \frac{E_R}{M} = \frac{0.677725}{4} \approx 0.169431$$

Maka implementasi pada persamaan 3.11 berdasarkan contoh matriks respons akhir menghasilkan:

$$\begin{aligned} H_R &\approx -(0.0683 \log_2 0.0683 + 3 \times 0.3106 \log_2 0.3106) \\ H_R &\approx 0.2647 + 1.5720 = 1.8367 \text{ bits} \end{aligned}$$

Kemudian hasil dari ekstraksi fitur GLCM dan Filter Gabor digabungkan menggunakan persamaan 3.12

$$F_{GLCM} = [0.0833, 0.9583, 0.8472, 0, 0.0833, 0.417]$$

$$F_{Gabor} = [-0.3795, 0.1593, 0.169431, 1.8367]$$

$$F = [0.0833, 0.9583, 0.8472, 0, 0.0833, 0.417, -0.3795, 0.1593, 0.169431, 1.8367]$$

Lampiran II Implementasi Perhitungan MLP

Sebagai contoh sederhana menggunakan 3 neuron, 10 fitur masukan, dengan nilai bobot dan bias:

$$X = [0.0833, 0.9583, 0.8472, 0, 0.0833, 0.417, -0.3795, 0.1593, 0.169431, 1.8367]$$

$$W_n = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0.928 & 0 & 0 \end{bmatrix} \text{ dan } b_i^{(1)} = \begin{bmatrix} 0 \\ -0.167 \\ -1.048 \end{bmatrix}$$

maka perhitungan untuk neuron pertama hidden layer pertama adalah,

$$\begin{aligned} V_1 &= (w_{1.1} \cdot 0.0833) + (w_{2.2} \cdot 0.9583) + (w_{3.3} \cdot 0.8472) + (w_{4.4} \cdot 0) \\ &\quad + (w_{5.5} \cdot 0.0833) + (w_{6.6} \cdot 0.417) + (w_{7.7} \cdot -0.3795) \\ &\quad + (w_{8.8} \cdot 0.1593) + (w_{9.9} \cdot 0.1694) + (w_{10.10} \cdot 1.8367) + b_1^{(1)} \end{aligned}$$

$$a_1 = \max(0, V_1) = 1.7053$$

Untuk neuron kedua diperoleh,

$$V_2 \approx -0.1673 \Rightarrow a_2 = 0$$

Untuk neuron ketiga diperoleh,

$$V_3 \approx -1.0487 \Rightarrow a_3 = 0$$

Sehingga keluaran dari hidden layer pertama adalah,

$$\alpha^{(1)} = [1.7053, 0, 0]$$

Dengan demikian, hidden layer kedua berfungsi untuk membentuk representasi yang lebih abstrak dari data, sehingga jaringan mampu mengenali pola-pola yang lebih kompleks. Sebagai contoh sederhana menggunakan 2 neuron, dengan keluaran hidden layer pertama yang sudah didapatkan serta bobot dan bias senilai,

$$U_i = \begin{bmatrix} 0 & 0.473 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \text{ dan } b_i^{(2)} = \begin{bmatrix} -1.168 \\ 0 \end{bmatrix}$$

maka perhitungan untuk neuron pertama hidden layer kedua adalah,

$$Z_1 = (U_{1.1} \cdot 1.7053) + (U_{2.2} \cdot 0) + (U_{3.3} \cdot 0) + b_1^{(2)}$$

$$\alpha^{(1)} = \max(0, Z_1) = 0$$

Sedangkan untuk neuron kedua adalah,

$$Z_2 = (U_{1.1} \cdot 1.7053) + (U_{2.2} \cdot 0) + (U_{3.3} \cdot 0) + b_2^{(2)}$$

$$\alpha^{(2)} = \max(0, z_2) = 0.8075$$

Sehingga keluaran hidden layer kedua adalah,

$$\alpha^{(2)} = [0, 0.8075]$$

Sehingga output layer memiliki $K = 4$ neuron. Persamaan 3.20 merupakan perhitungan untuk mendapatkan nilai neuron ke empat pada output layer,

Dengan bobot dan bias senilai,

$$T_j = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0.069 & -0.339 & 0.485 & -0.148 \end{bmatrix} \text{ dan } b_k^{(out)} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

maka,

$$y_k = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0.069 & -0.339 & 0.485 & -0.148 \end{bmatrix} \cdot [0 \ 0.8075] + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.056 \\ -0.274 \\ 0.392 \\ -0.120 \end{bmatrix}$$

Kemudian nilai probabilitas kelas dihitung dengan fungsi softmax menggunakan nilai dari y_k dengan persamaan 3.21,

$$Y_1 = \frac{e^{0.056}}{e^{0.056} + e^{-0.274} + e^{0.392} + e^{-0.120}} \approx 0.245$$

$$Y_2 = \frac{e^{-0.274}}{e^{0.056} + e^{-0.274} + e^{0.392} + e^{-0.120}} \approx 0.174$$

$$Y_3 = \frac{e^{0.392}}{e^{0.056} + e^{-0.274} + e^{0.392} + e^{-0.120}} \approx 0.354$$

$$Y_4 = \frac{e^{-0.120}}{e^{0.056} + e^{-0.274} + e^{0.392} + e^{-0.120}} \approx 0.227$$

Sehingga nilai dari output layer adalah,

$$Y = [0.245, 0.174, 0.354, 0.227]$$

Apabila semisal label sebenarnya adalah $Y = [1, 0, 0, 0]$ maka dapat dihitung nilai

$loss$,

$$L = -(1 \cdot \log(0.245) + 0 \cdot \log(0.174) + 0 \cdot \log(0.354) + 0 \cdot \log(0.227))$$

$$L = -\log(0.245) \approx 1.407$$

Error pada neuron output dapat dihitung dengan persamaan 3.23,

$$\delta^{(out)} = [-0.755, 0.174, 0.354, 0.227]$$

Untuk menghitung gradien bobot pada output layer yaitu,

$$\frac{\partial L}{\partial T_j} = \delta^{(out)} \cdot \alpha^{(2)} = [-0.755, 0.174, 0.354, 0.227] \cdot [0, 0.8075]$$

$$\frac{\partial L}{\partial T_j} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ -0.609 & 0.140 & 0.286 & 0.183 \end{bmatrix}$$

Perhitungan untuk pembaruan bobot dapat dihitung dengan persamaan 3.24 dan untuk pembaruan bias dapat dihitung dengan persamaan 3.25

Sebagai contoh, jika *learning rate* bernilai $\eta = 0.1$ dan hasil perhitungan gradien pada bobot antara neuron ke-2 pada hidden layer kedua dan neuron output pertama adalah $\frac{\partial L}{\partial T_{j,k}} \approx -0.609$, serta nilai bobot awal $T_{j,k} = 0.0609$, maka pembaruan

bobot dapat dihitung sebagai berikut,

$$T_{j,k}^{(baru)} = 0.0609 - 0.1 \cdot (-0.609) = 0.13$$

Kemudian untuk pembaruan bias dapat dihitung sebagai berikut,

$$b_k^{(baru)} = 0 - 0.1 \cdot (-0.755) = 0.755$$

Lampiran III Hasil Uji Coba Kejenuhan Arsitektur Neural Network

Menggunakan model D dengan 256 neuron pada *hidden layer* 1, dengan 3 *hidden layer*

Kelas Sebenarnya	Kelas Prediksi			
	Normal	<i>Diabetic</i>	<i>Glaucoma</i>	<i>Cataract</i>
Normal	25	11	3	1
<i>Diabetic</i>	5	31	2	2
<i>Glaucoma</i>	1	3	26	10
<i>Cataract</i>	1	1	4	34

Matrix	Nilai
Accuracy	72,50%
Precision	73,04%
Recall	72,50%
F1-score	72,26%

Menggunakan model D dengan 64 neuron pada *hidden layer* 1, dengan 4 *hidden layer*

Kelas Sebenarnya	Kelas Prediksi			
	Normal	<i>Diabetic</i>	<i>Glaucoma</i>	<i>Cataract</i>
Normal	3	33	3	1
<i>Diabetic</i>	1	34	4	1
<i>Glaucoma</i>	1	6	25	8
<i>Cataract</i>	1	1	5	33

Matrix	Nilai
Accuracy	59.38%
Precision	60.06%
Recall	59.38%
F1-score	54.29%

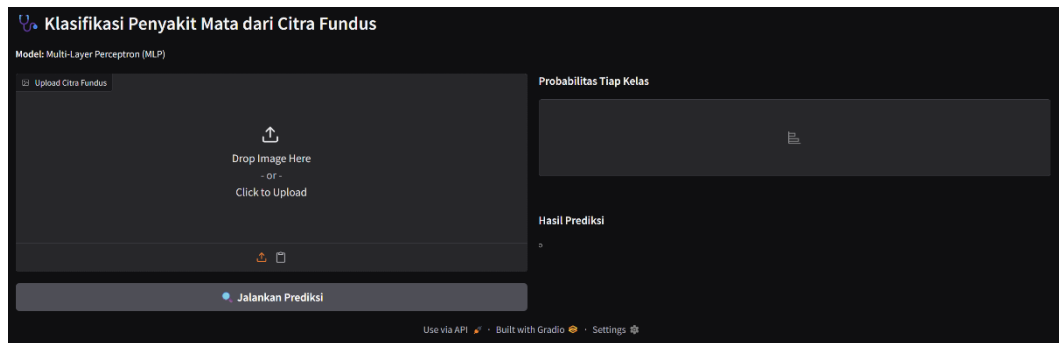
Menggunakan model D dengan 128 neuron pada *hidden layer* 1, dengan 4 *hidden layer*

Kelas Sebenarnya	Kelas Prediksi			
	Normal	<i>Diabetic</i>	<i>Glaucoma</i>	<i>Cataract</i>
Normal	29	7	3	1
<i>Diabetic</i>	16	20	3	1
<i>Glaucoma</i>	0	5	29	6
<i>Cataract</i>	0	1	4	35

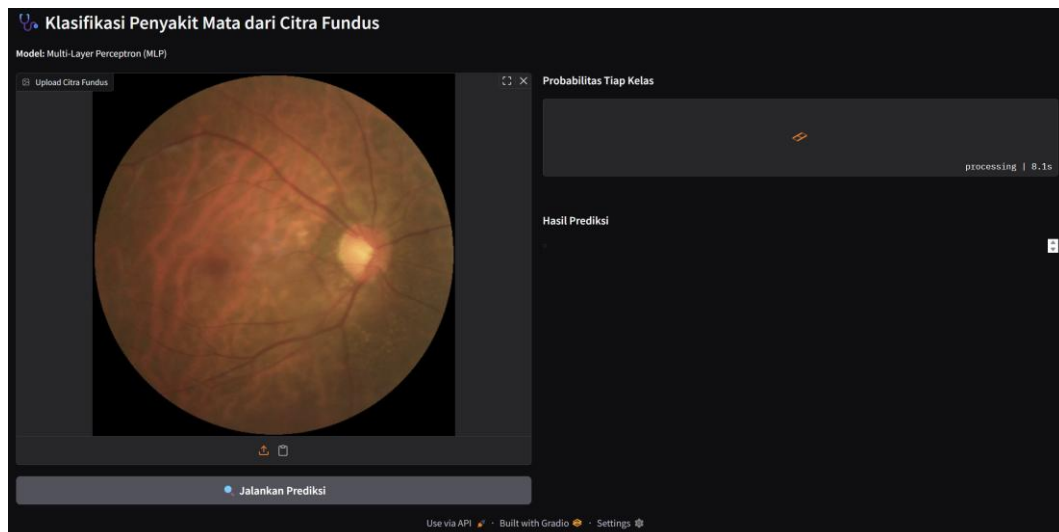
Matrix	Nilai
Accuracy	70.63%
Precision	70.20%
Recall	70.63%
F1-score	70.20%

Lampiran IV Interface Test

Tampilan awal



Tampilan setelah user memasukkan *input* data



Tampilan ketika *output* berupa hasil prediksi penyakit mata dari citra fundus

