

Klasifikasi Penyakit Mata Berdasarkan Citra Fundus Menggunakan Metode Multi-Layer Perceptron

Laudza Atsila Prasetyo ^{(1)*}, Okta Qomaruddin Aziz ⁽²⁾, Tri Mukti Lestari ⁽³⁾
Departemen Teknik Informatika, UIN Maulana Malik Ibrahim, Malang, Indonesia
e-mail : laudza.a.p@gmail.com, {okta.qomaruddin,trimuktilestari}@uin-malang.ac.id.

* Penulis korespondensi.

Artikel ini diajukan 13 Januari 2026, direvisi 7 April 2026, diterima 8 April 2026, dan dipublikasikan 25 Mei 2026.

Abstract

This research aims to evaluate the performance of the Multi-Layer Perceptron (MLP) for classifying eye diseases from fundus images in the ODIR dataset, which comprises four classes: Normal, Diabetic, Glaucoma, and Cataract. The methodology includes feature extraction using GLCM and Gabor, data pre-processing through cleaning, augmentation, and undersampling, and testing 16 model scenarios with variations in the number of hidden layers (2 and 3) 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 with the 64–32-neuron configuration, yielding a 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.

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

Abstrak

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 2 dan 3 *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 akurasi rata-rata sebesar 73,06%. Secara keseluruhan, penelitian ini menyimpulkan bahwa kombinasi antara dataset seimbang dan arsitektur MLP yang proporsional mampu meningkatkan kemampuan model dalam mengklasifikasikan penyakit mata berbasis citra fundus.

Kata Kunci: Klasifikasi Penyakit Mata, MLP, Fundus, Ekstraksi Fitur, Klasifikasi

1. PENDAHULUAN

Penyakit mata seperti *Diabetic Retinopathy, Glaucoma, dan Cataract* merupakan penyebab utama kebutaan di dunia (World Health Organization, 2019). Hal tersebut ditunjukkan oleh data *Blindness and Vision Impairment* (World Health Organization, 2023) yang diterbitkan pada 10 Agustus 2023. Di antara 1 miliar orang, kondisi utama yang menyebabkan gangguan pengelihatannya 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 serta 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



Diabetic Retinopathy, *Glaucoma*, dan *Cataract* (Indraswari dkk., 2022). Namun, proses diagnosis citra fundus secara manual bersifat subjektif, memerlukan keahlian klinis, dan berpotensi menghasilkan variasi interpretasi antarpemeriksa. Oleh karena itu, diperlukan pendekatan komputasi yang mampu membantu proses klasifikasi citra fundus secara lebih objektif dan konsisten.

Pendekatan *Artificial Neural Network (ANN)*, khususnya arsitektur *Multi-Layer Perceptron (MLP)*, telah banyak digunakan pada data berbasis fitur numerik hasil ekstraksi citra medis dan menunjukkan performa klasifikasi yang kompetitif (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). Selain itu, MLP dapat dilatih untuk mengenali pola tekstur, morfologi, dan karakteristik statistik pada citra fundus, sehingga berpotensi mendukung pengembangan sistem bantu diagnosis berbasis citra.

Beberapa penelitian sebelumnya telah berhasil menerapkan ANN (MLP) untuk klasifikasi berbagai penyakit dari citra medis, seperti deteksi pendarahan pada klasifikasi *Diabetic Retinopathy* 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 diadaptasi untuk 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 glaukoma atau retinopati *diabetik*, 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).

Sejumlah penelitian sebelumnya telah menerapkan MLP maupun metode *feature-based learning* pada citra medis dan menunjukkan bahwa pendekatan berbasis fitur numerik dapat menghasilkan performa klasifikasi yang baik. Meskipun demikian, sebagian besar studi masih berfokus pada satu jenis penyakit, atau hanya menguji model pada satu konfigurasi arsitektur tanpa mengevaluasi pengaruh variasi perlakuan data terhadap stabilitas dan akurasi model. Selain itu, penelitian yang memanfaatkan kombinasi fitur tekstur *GLCM* dan *Gabor* pada skenario klasifikasi *multi-class* citra fundus masih terbatas, terutama dalam konteks evaluasi sistematis dengan berbagai konfigurasi MLP.

Berdasarkan kondisi tersebut, penelitian ini difokuskan pada evaluasi performa MLP dalam klasifikasi penyakit mata berbasis citra fundus menggunakan kombinasi fitur *GLCM* dan *Gabor*, dengan membandingkan beberapa skenario perlakuan data serta variasi konfigurasi arsitektur jaringan. Evaluasi ini diharapkan dapat memberikan pemahaman yang lebih komprehensif

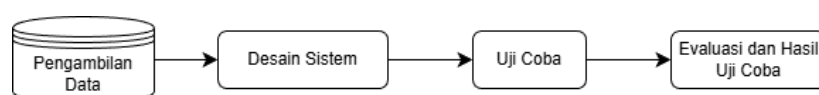


mengenai pengaruh representasi fitur tekstur, keseimbangan data, dan konfigurasi model terhadap efektivitas MLP dalam kasus klasifikasi *multi-class* penyakit mata.

2. METODE PENELITIAN

Penelitian ini menggunakan metode *Multi-Layer Perceptron* (MLP), yang merupakan salah satu arsitektur *Artificial Neural Network* (ANN) yang tersusun atas beberapa lapisan neuron, yaitu lapisan *input*, satu atau lebih lapisan tersembunyi (*hidden layer*), dan lapisan *output* (Saputra dkk., 2017). Setiap neuron pada MLP terhubung sepenuhnya 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 (Wulandari & Fitriyah, 2021). 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 kelainan retina atau indikasi penyakit mata menggunakan fitur yang merepresentasikan pola tertentu pada citra fundus (Aziz dkk., 2025).

Dalam penelitian ini, terdapat tahapan yang menggambarkan proses penelitian ini, bertujuan untuk membantu peneliti melaksanakan penelitian secara sistematis dan terstruktur. Tahapan ini mencakup pengambilan data, desain sistem yang mencakup implementasi metode *Multi-Layer Perceptron* (MLP), uji coba, serta evaluasi dan analisis hasil uji coba. Penelitian ini menggunakan bahasa pemrograman Python untuk membangun model. Keseluruhan tahapan penelitian ditampilkan pada Gambar 1.



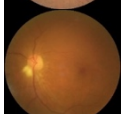


Gambar 1 Desain Tahapan Penelitian

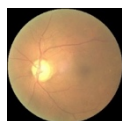
2.1 Pengambilan Data

Dalam penelitian ini, untuk melakukan klasifikasi penyakit mata berdasarkan citra fundus, dibutuhkan *dataset* citra fundus dari kelas yang akan diklasifikasikan, yaitu *Normal*, *Diabetic Retinopathy*, *Cataract* dan *Glaucoma*. *Dataset* yang digunakan adalah “*Ocular Disease Recognition/ODIR-5k*” yang diperoleh dari situs 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 gambar berukuran 512 x 512 piksel. Anotasi citra dilakukan oleh pembaca ahli dengan manajemen kontrol kualitas. *Dataset* ini mencakup 5.000 data pasien. Tabel 1 menunjukkan contoh data dan jumlah data yang ada pada *dataset* serta digunakan dalam penelitian.

Tabel 1 Contoh Data

No.	Citra Fundus	Labels	Jumlah Citra
1		Normal	2.873
2		Diabetic	1.608
3		Cataract	293

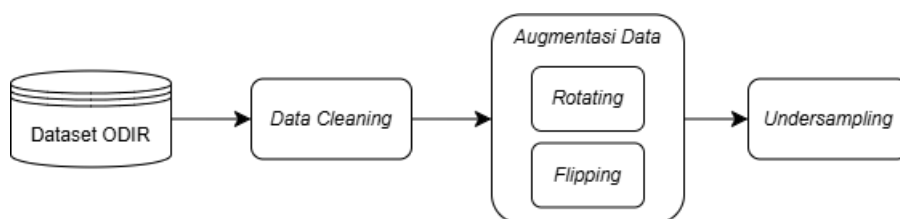




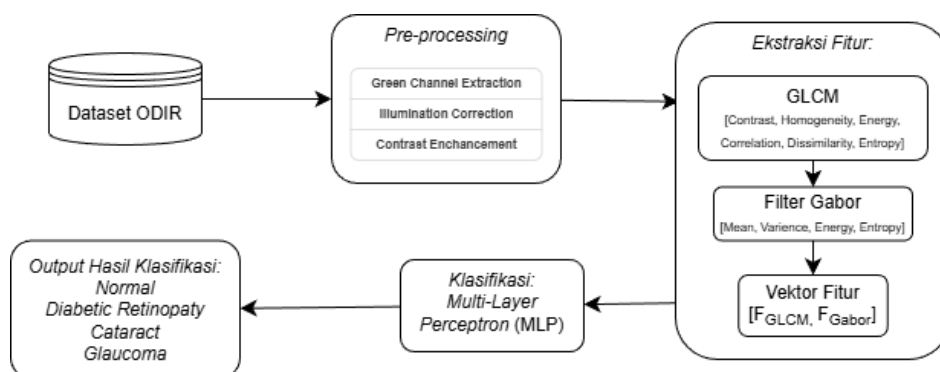
2.2 Desain Sistem

Proses klasifikasi penyakit mata ditunjukkan pada Gambar 2 dan 3. Proses ini dimulai dengan input data yang berasal dari dataset ODIR setelah melalui proses *data cleaning* (menyaring kualitas citra), augmentasi data dengan rotasi citra (*rotating*) dan membalikan citra (*flipping*), serta *undersampling*. Kemudian, data tersebut diproses pada tahapan *pre-processing*, yaitu *Green Channel Extraction* untuk menonjolkan struktur pembuluh darah pada retina. Kanal hijau dipilih karena memiliki kontras paling tinggi terhadap struktur retina dibandingkan dengan kanal merah atau biru (Kaya dkk., 2020). Dilanjutkan dengan *illumination correction* untuk meratakan pencahayaan (*vignetting*) pada citra sehingga intensitas retina menjadi lebih seragam (Saha dkk., 2018). Kemudian, dilanjutkan dengan *contrast enhancement* untuk meningkatkan kontras citra dan memperjelas detail halus pada jaringan retina tanpa memperbesar noise. Keseluruhan proses *pre-processing* bertujuan untuk meningkatkan kualitas citra sebelum diproses pada tahap ekstraksi fitur. Setelah dilakukan *pre-processing*, tahapan selanjutnya adalah ekstraksi fitur menggunakan metode GLCM untuk fitur tekstur (Siswanto dkk., 2020). Serta filter Gabor untuk fitur frekuensi spasial yang akan digunakan sebagai input pada model MLP.

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 menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* berdasarkan *confusion matrix* (Faghani dkk., 2022). Penelitian ini menghasilkan 4 kelas dari proses klasifikasi, yaitu *Normal*, *Diabetic Retinopathy*, *Cataract*, dan *Glaucoma*. Hasil klasifikasi tersebut kemudian ditentukan dengan argumen max untuk ditampilkan sebagai *output* yang memberikan informasi tentang penyakit pada citra fundus (Jason Brownlee, 2020).



Gambar 2 Alur Proses Persiapan Data



Gambar 3 Desain Sistem



2.3 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 (Costaner & Harun, 2024). Ekstraksi fitur dilakukan dalam beberapa tahapan meliputi:

2.3.1 GLCM

Ekstraksi fitur tekstur pada citra fundus dilakukan menggunakan *Gray Level Co-occurrence Matrix (GLCM)* untuk merepresentasikan hubungan spasial antarpasangan piksel pada jarak dan arah tertentu. Matriks GLCM dibentuk dengan menghitung frekuensi kemunculan pasangan intensitas piksel (i, j) yang kemudian dinormalisasi menjadi matriks probabilitas $P(i, j)$. Dari matriks tersebut dihitung enam fitur tekstur utama, yaitu *contrast*, *homogeneity*, *energy (ASM)*, *correlation*, *dissimilarity*, dan *entropy*, sebagaimana dirumuskan pada Pers. (1) sampai (6). Perhitungan GLCM dilakukan pada jarak $d = 1$ dengan arah $0^\circ, 45^\circ, 90^\circ,$ dan 135° , kemudian nilai fitur dirata-ratakan untuk setiap arah.

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

$$Homogeneity = \sum_{i=0}^1 \sum_{j=0}^1 \frac{P(i, j)}{1 + |i - j|} \quad (2)$$

$$Energy = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (P(i, j))^2 \quad (3)$$

$$Correlation = \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i - \mu)(j - \mu)P(i, j)}{\sigma^2} \quad (4)$$

$$Dissimilarity = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |i - j| P(i, j) \quad (5)$$

$$Entropy = - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i, j) \log P(i, j) \quad (6)$$

2.3.2 Filter Gabor

Selain fitur berbasis GLCM, ekstraksi tekstur juga dilakukan menggunakan *filter Gabor* untuk merepresentasikan informasi frekuensi dan orientasi pada citra fundus (Dakshayani dkk., 2022). *Filter Gabor* dua dimensi didefinisikan pada Pers. (7) dengan transformasi koordinat Pers. (8) dan (9) di mana λ adalah panjang gelombang, θ adalah orientasi, σ adalah deviasi standar *Gaussian envelope*, γ adalah *aspect ratio*, dan ϕ adalah fase. Kemudian, citra difilter menggunakan *bank Gabor* pada beberapa orientasi dan frekuensi. Dari respons magnitudo hasil konvolusi $R(x, y)$, dihitung empat deskriptor statistik utama, yaitu *mean*, *variance*, *energy*, dan *entropy*, dengan formulasi pada Pers. (10) sampai (13), dengan p_k merupakan probabilitas intensitas respons pada histogram magnitudo Gabor.

$$g(x, y) = \exp \left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2} \right) \cdot \cos \left(2\pi \frac{x'}{\lambda} + \psi \right) \quad (7)$$



$$x' = x \cos \theta + y \sin \theta \quad (8)$$

$$y' = -x \sin \theta + y \cos \theta \quad (9)$$

$$\mu_R = \frac{1}{M} \sum_{x,y} R(x,y) \quad (10)$$

$$\sigma^2 = \frac{1}{N} \sum_{k=1}^N (R_k - \mu)^2 \quad (11)$$

$$Energy_R = \sum_{x,y} R(x,y)^2 \quad (12)$$

$$H_R = - \sum_k p_k \log_2 p_k \quad (13)$$

2.3.3 Vektor Fitur

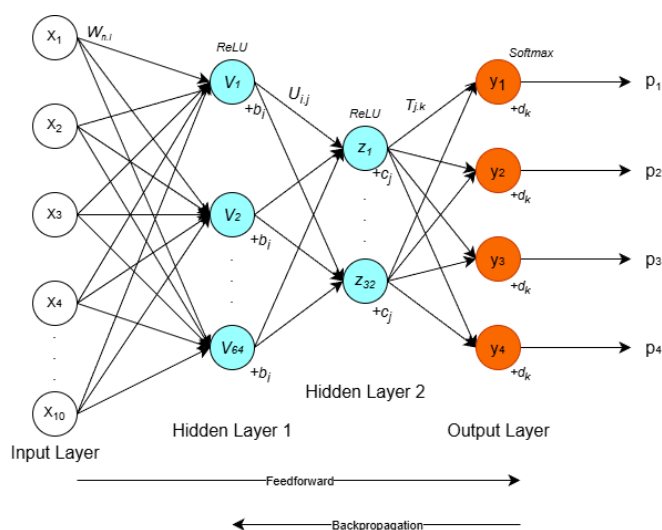
Tahap vektor fitur dilakukan setelah proses ekstraksi fitur menggunakan metode GLCM dan *filter Gabor* selesai. Pada tahap GLCM, terdapat enam parameter, yaitu *contrast*, *homogeneity*, *energy*, *correlation*, *dissimilarity*, dan *entropy*. Sedangkan pada tahap ekstraksi fitur dengan *Filter Gabor*, terdapat empat statistik spasial, yaitu *mean*, *variance*, *energy*, dan *entropy*. Berikut penggabungannya menjadi vektor fitur menggunakan Pers. (14). 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.

$$\begin{aligned} F &= [F_{GLCM}, F_{Gabor}] \\ F_{GLCM} &= [Contrast, Homogeneity, Energy, Correlation, Dissimilarity, Entropy] \\ F_{Gabor} &= [Mean, Variance, Energy, Entropy] \\ F &= \left[\begin{array}{l} Contrast, Homogeneity, Energy, Correlation, Dissimilarity, \\ Entropy, Mean, Variance, Energy, Entropy \end{array} \right] \end{aligned} \quad (14)$$

2.4 Metode Multi-Layer Perceptron

Dalam penelitian ini, model *Multi-Layer Perceptron (MLP)* diaplikasikan untuk mengklasifikasikan penyakit mata berdasarkan citra fundus. Arsitektur *Multi-Layer Perceptron* terdiri dari *input layer*, *hidden layer*, dan *output layer* (Heriz dkk., 2018). Arsitektur yang digunakan pada penelitian ini ditunjukkan oleh Gambar 3 dengan menerapkan beberapa modifikasi di antaranya ialah penerapan 2 *hidden layer* dan 3 *hidden layer*. Pemilihan *hyperparameter* yang tepat sangat penting agar model mencapai akurasi optimal dan tidak mengalami *overfitting* (Goodfellow dkk., 2016). Berikut penjelasan aplikatif mengenai proses klasifikasi data hasil *pre-processing* dan ekstraksi fitur yang diproses oleh MLP.





Gambar 4 Arsitektur MLP

Pada tahap awal, data dimasukkan ke dalam model dalam bentuk vektor fitur numerik hasil ekstraksi citra fundus. Setiap citra direpresentasikan sebagai vektor berdimensi 10 yang terdiri atas enam fitur tekstur GLCM, yaitu contrast, homogeneity, energy, correlation, dissimilarity, dan entropy, serta empat fitur berbasis Gabor, yaitu mean, variance, energy, dan entropy. Setiap elemen pada vektor fitur dipetakan ke satu neuron pada input layer, sehingga input layer terdiri dari 10 neuron yang masing-masing merepresentasikan satu fitur numerik sebelum diteruskan ke hidden layer untuk proses pemetaan nonlinier.

Vektor fitur pada input layer selanjutnya dipetakan ke beberapa hidden layer untuk melakukan proses pemodelan non-linier terhadap hubungan antar fitur. Pada penelitian ini, konfigurasi arsitektur MLP terdiri dari 2 dan 3 hidden layer dengan jumlah neuron masing-masing 32, 64, dan 128 sesuai skenario pengujian. Setiap neuron pada hidden layer menghitung kombinasi linier dari masukan yang diterima, kemudian dilewatkan melalui fungsi aktivasi ReLU sehingga menghasilkan representasi fitur laten yang lebih diskriminatif. Pers. (15) digunakan untuk menghitung nilai neuron pada hidden layer, di mana \$X_n\$ merupakan nilai fitur pada layer sebelumnya (input layer), sedangkan \$W_{\{n,i\}}\$ dan \$b_i\$ masing-masing menyatakan bobot dan bias neuron ke-\$i\$. Keluaran setiap neuron pada hidden layer menghasilkan representasi fitur yang kemudian diteruskan ke output layer untuk proses klasifikasi.

$$V_i = ReLU \left(\sum_{n=1} W_{n,i} \cdot X_n + b_i \right) \quad (15)$$

Keluaran dari hidden layer diteruskan ke output layer yang terdiri dari \$K\$ neuron sesuai jumlah kelas. Setiap neuron ke-\$k\$ menghitung kombinasi linier dari aktivasi pada layer sebelumnya \$Z_j\$ menggunakan bobot \$T_{\{j,k\}}\$ dan bias \$d_k\$, sebagaimana ditunjukkan pada Pers. (16), di mana nilai \$y_k\$ merepresentasikan skor logit sebelum dilakukan normalisasi probabilitas. Nilai logit tersebut kemudian diproses menggunakan fungsi softmax untuk memperoleh probabilitas prediksi pada setiap kelas. Pada Pers. (17), nilai \$\hat{y}_k\$ menyatakan probabilitas model bahwa sampel termasuk ke dalam kelas ke-\$k\$.

$$y_k = \sum_{j=1} T_{j,k} \cdot Z_j + d_k \quad (16)$$

$$\hat{y}_k = \frac{e^{y_k}}{\sum_{j=1}^K e^{y_j}} \quad (17)$$



Pelatihan model dilakukan dengan meminimalkan categorical cross-entropy, yang mengukur selisih antara distribusi target t_k dan probabilitas prediksi \hat{y}_k . Pada kasus one-hot encoding, nilai t_k bernilai 1 untuk kelas benar dan 0 untuk kelas lainnya. Pers. (18) menunjukkan fungsi loss yang digunakan selama proses pelatihan model untuk mengoptimalkan parameter jaringan. Selanjutnya, proses backpropagation dilakukan untuk menghitung gradien error pada neuron output. Gradien error pada neuron output ke- k dihitung sebagaimana ditunjukkan pada Pers. (19), di mana nilai gradien ini digunakan dalam proses pembaruan bobot pada tahap propagasi balik. Proses ini memungkinkan model untuk menyesuaikan parameter secara iteratif agar menghasilkan prediksi yang semakin akurat.

$$L = - \sum_{i=1}^K t_i \cdot \log(\hat{y}_k) \tag{18}$$

$$\delta_k^{(out)} = \hat{y}_k - y_k \tag{19}$$

Tahap akhir model adalah menentukan keputusan kelas berdasarkan probabilitas terbesar pada vektor keluaran softmax. Kelas akhir dipilih menggunakan fungsi arg-max sebagaimana ditunjukkan pada Pers. (20), sehingga sampel diklasifikasikan ke kelas dengan nilai probabilitas tertinggi. Dengan mekanisme ini, model dapat menghasilkan keputusan klasifikasi akhir berdasarkan probabilitas prediksi yang diperoleh dari seluruh neuron output.

$$P = \arg \max_k (\hat{y}_k) \tag{20}$$

3. HASIL DAN PEMBAHASAN

Penelitian ini menerapkan beberapa skenario pengujian yang mengombinasikan perlakuan data dan variasi arsitektur MLP untuk mengevaluasi performa klasifikasi citra fundus berbasis fitur GLCM dan Gabor. Variasi perlakuan data terdiri dari Model A (*unbalanced*), Model B (*augmentasi*), Model C (*undersampling*), dan Model D (*undersampling + augmentasi*), sebagaimana ditunjukkan pada Tabel 2. Variasi tersebut dirancang untuk menganalisis pengaruh jumlah dan distribusi data terhadap kemampuan model dalam mempelajari pola pada citra fundus. Model yang menggunakan *augmentasi*, dilakukan *transformasi* citra berupa rotasi dengan sudut kecil (-15° hingga $+15^\circ$) dan *horizontal flipping*, dengan tetap mempertahankan area *field of view (FOV)* retina melalui masking, sehingga variasi data meningkat tanpa mengubah karakteristik anatomis citra secara signifikan.

Tabel 2 Variasi Jumlah Data

Kelas	Jumlah Data				
	Sebelum Data Cleaning	Setelah Data Cleaning (Model A)	Augmentasi (Model B)	Undersampling (Model C)	Undersampling + Augmentasi (Model D)
Normal	2.873	2.413	2.500	200	400
Diabetic	1.608	1.412	2.500	200	400
Cataract	293	225	2.500	200	400
Glaucoma	284	216	2.500	200	400
Total	5.058	4.266	10.000	800	1.600

Tabel 3 Skenario Pengujian

No Skenario	Jumlah Data Skenario	Arsitektur MLP	Neuron	Kode Skenario Pengujian
1	Model A (4.266 data)	2 Hidden	32-16	A-2H-32
2		Layer	64-32	A-2H-64
3			64-32-16	A-3H-64



4		3 <i>Hidden Layer</i>	128-64-32	A-3H-128
5	Model B (10.000 data)	2 <i>Hidden Layer</i>	32-16	B-2H-32
6		3 <i>Hidden Layer</i>	64-32	B-2H-64
7		3 <i>Hidden Layer</i>	64-32-16	B-3H-64
8		3 <i>Hidden Layer</i>	128-64-32	B-3H-128
9	Model C (800 data)	2 <i>Hidden Layer</i>	32-16	C-2H-32
10		3 <i>Hidden Layer</i>	64-32	C-2H-64
11		3 <i>Hidden Layer</i>	64-32-16	C-3H-64
12		3 <i>Hidden Layer</i>	128-64-32	C-3H-128
13	Model D (1.600 data)	2 <i>Hidden Layer</i>	32-16	D-2H-32
14		3 <i>Hidden Layer</i>	64-32	D-2H-64
15		3 <i>Hidden Layer</i>	64-32-16	D-3H-64
16		3 <i>Hidden Layer</i>	128-64-32	D-3H-128

Model C dan D dilatih hingga maksimum 150 epoch, sedangkan model A dan B dilatih hingga maksimum 200 epoch dengan menerapkan *early stopping* pada seluruh model guna 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.

Selain variasi jumlah data, pengujian juga dilakukan pada beberapa konfigurasi arsitektur MLP yang berbeda sebagaimana ditunjukkan pada Tabel 3, mencakup 2 *hidden layer* (32–16, 64–32) serta 3 *hidden layer* (64–32–16, 128–64–32). Kombinasi antara variasi data dan arsitektur tersebut menghasilkan sejumlah skenario uji yang digunakan untuk mengevaluasi pengaruh kedua faktor terhadap kinerja klasifikasi.

Dari 16 kombinasi skenario pengujian, analisis difokuskan pada skenario terbaik dari masing-masing model perlakuan data. Pemilihan dilakukan berdasarkan nilai rata-rata akurasi tertinggi pada hasil *k-fold cross-validation*. Ringkasan performa skenario terbaik dari masing-masing model ditunjukkan pada Tabel 4.

Tabel 4 Ringkasan Performa Terbaik

No Skenario	Kode Pengujian	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
8	B-3H-128		70%	71.80%	71.30%	0,0075
9	C-2H-32		56.25%	64.38%	59.38%	0,0322
14	D-2H-64		69.06%	75%	73.06%	0,0243

Pada Model A, skenario terbaik diperoleh pada arsitektur 2 *hidden layer* (32–16 neuron) dengan rata-rata akurasi 60.28%. Meskipun arsitektur ini relatif sederhana, performa model terbatas karena dataset tidak seimbang, sehingga model lebih banyak belajar dari kelas mayoritas. Nilai standar deviasi yang rendah menunjukkan bahwa model stabil, namun stabilitas ini tidak disertai kemampuan generalisasi yang baik terhadap kelas *Glaucoma*, sebagaimana terlihat pada Tabel 5. Kondisi ini menegaskan bahwa pada dataset yang tidak seimbang, peningkatan kapasitas jaringan tidak memberikan kontribusi signifikan karena bias distribusi data masih mendominasi proses pembelajaran.

Tabel 5 Rata-rata *Evaluation Matrix* Skenario A-2H-32

Kelas	Precision	Recall	F1-Score
-------	-----------	--------	----------



Normal	61.57%	85.04%	71.39%
Diabetic	51%	24.51%	32.89%
Glaucoma	25.67%	04.19%	06.93%
Cataract	75.34%	73.33%	74.07%

Model B menghasilkan skenario terbaik pada arsitektur 3 *hidden layer* (128–64–32) dengan rata-rata akurasi 71.30% dan standar deviasi rendah (0.0075). *Augmentasi* data meningkatkan variasi pola tekstur citra, sehingga model dengan kapasitas lebih besar mampu memanfaatkan informasi tambahan tersebut secara efektif. Pada kasus ini, arsitektur yang lebih kompleks justru memberikan keuntungan karena tersedia cukup data untuk mendukung proses pembelajaran. Tabel 6 menunjukkan bahwa skenario B-3H-128 efektif mengenali kelas *Cataract* dan *Glaucoma*, namun pada kelas *Diabetic* dan *Normal* masih lebih rendah, menunjukkan adanya kemiripan pola antar kedua kelas tersebut yang menyulitkan pemisahan. Hal ini menunjukkan bahwa *augmentasi* tidak hanya menambah jumlah data, tetapi juga memperluas ruang fitur, sehingga model mampu membentuk representasi yang lebih diskriminatif.

Pada Model C, skenario terbaik justru kembali diperoleh pada arsitektur 2 *hidden layer* (32–16 neuron) dengan rata-rata akurasi 59.38% dan standar deviasi terbesar (0.0322). Meskipun *undersampling* menghasilkan distribusi kelas yang lebih seimbang metode ini secara signifikan mengurangi jumlah data pelatihan, sehingga model kehilangan variasi pola tekstur. Akibatnya arsitektur yang lebih kompleks tidak memberikan peningkatan performa karena kapasitas model tidak sebanding dengan jumlah data, yang berpotensi menyebabkan *underfitting* serta meningkatkan ketidakstabilan performa antar-*fold*. Tabel 7 menunjukkan bahwa skenario C-2H-32 dapat mengenali kelas *Cataract* dengan baik, namun kinerja dalam mengenali kelas *Glaucoma*, *Diabetic* dan *Normal* masih 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.

Tabel 6 Rata-rata *Evaluation Matrix* Skenario B-3H-128

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%

Tabel 7 Rata-rata *Evaluation Matrix* Skenario C-2H-32

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

Model D menghasilkan performa terbaik di antara seluruh skenario, yaitu pada arsitektur 2 *hidden layer* (64–32 neuron) dengan rata-rata akurasi 73.06%. Kombinasi keseimbangan data melalui *undersampling* dan penambahan variasi melalui *augmentasi* membuat model memperoleh distribusi kelas yang proporsional sekaligus tetap memiliki keragaman fitur yang memadai. Menariknya, arsitektur dengan kapasitas sedang justru lebih optimal dibanding arsitektur yang lebih besar, yang mengindikasikan bahwa model dengan kompleksitas moderat lebih sesuai dengan karakteristik data hasil kombinasi tersebut. Tabel 8 menunjukkan performa paling seimbang antar kelas menandakan bahwa skenario D-2H-64 mampu mengurangi bias antar kelas. Secara keseluruhan, Skenario ini 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 8 Rata-rata *Evaluation Matrix* Skenario D-2H-64

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

Meskipun demikian, nilai akurasi yang diperoleh dalam penelitian ini relatif lebih rendah dibandingkan beberapa penelitian terdahulu. Perbedaan tersebut disebabkan oleh perbedaan kompleksitas tugas klasifikasi yang dilakukan. Penelitian sebelumnya umumnya menerapkan skenario klasifikasi *biner* (misalnya normal dan penyakit), yang lebih sederhana karena model hanya perlu membedakan dua kelas utama. Sebaliknya, penelitian ini menggunakan pendekatan klasifikasi *multi-class*, sehingga model harus membedakan beberapa jenis penyakit mata dengan karakteristik tekstur yang seringkali saling tumpang tindih. Kondisi ini meningkatkan tingkat kesulitan klasifikasi secara signifikan, sehingga berdampak pada penurunan nilai akurasi. Selain itu, penggunaan fitur berbasis ekstraksi manual (GLCM dan *Filter Gabor*) juga memiliki keterbatasan dalam menangkap kompleksitas pola spasial dibandingkan pendekatan berbasis *deep learning end-to-end*.

4. KESIMPULAN

Penelitian ini mengevaluasi kinerja model MLP pada klasifikasi citra fundus berbasis fitur GLCM dan *Filter Gabor* melalui beberapa skenario perlakuan data dan variasi arsitektur jaringan. Berdasarkan hasil pengujian, dapat disimpulkan bahwa jumlah dan distribusi data memiliki pengaruh lebih dominan dibandingkan kompleksitas arsitektur MLP dalam menentukan kemampuan model ketika mempelajari pola citra. Model pada data tidak seimbang menghasilkan akurasi terendah karena distribusi kelas yang tidak proporsional menyebabkan bias terhadap kelas mayoritas. Sedangkan peningkatan performa dicapai pada skenario dengan augmentasi data serta mencapai hasil terbaik pada kombinasi undersampling dan augmentasi, karena distribusi kelas menjadi lebih seimbang namun tetap mempertahankan keragaman pola tekstur.

Selain itu, hasil menunjukkan bahwa penambahan jumlah hidden layer dan neuron tidak selalu meningkatkan akurasi. Arsitektur dengan kapasitas moderat memberikan performa yang lebih stabil pada kondisi data terbatas, sementara arsitektur yang lebih kompleks baru efektif ketika variasi data yang mencukupi. Dengan demikian, pemilihan arsitektur MLP perlu diselaraskan dengan karakteristik dan kualitas data, bukan hanya berorientasi pada peningkatan ukuran model. Temuan ini menegaskan bahwa pendekatan klasifikasi berbasis fitur tekstur (GLCM–*Filter Gabor*) dan MLP masih relevan untuk skenario data terbatas, selama didukung oleh strategi perlakuan data yang tepat. Secara khusus, penelitian ini memberikan pemahaman bahwa keseimbangan distribusi kelas dan kecukupan jumlah data berperan penting dalam meningkatkan kemampuan model dalam merepresentasikan pola citra secara lebih akurat.

Untuk meningkatkan performa klasifikasi, terutama pada kelas dengan karakteristik tekstur yang sulit dibedakan, penelitian selanjutnya dapat mempertimbangkan beberapa strategi, seperti menambah jumlah dataset dan memperluas keberagaman citra sehingga model memperoleh representasi fitur yang lebih kuat jika dilatih menggunakan lebih banyak variasi citra fundus. Selain itu, kombinasi metode ekstraksi fitur lainnya, seperti Local Binary Pattern (LBP), kombinasi fitur Haralick, atau wavelet transform dapat dipertimbangkan untuk memberikan representasi tekstur yang lebih kuat bagi model. Dengan penyempurnaan lebih lanjut, pendekatan klasifikasi berbasis fitur tekstur dan MLP diharapkan dapat menghasilkan model yang lebih akurat dan stabil untuk menjadi solusi dalam mendukung deteksi dini penyakit mata berbasis 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>
- Costaner, L., & Harun, N. (2024). 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>
- Dakshayani, V., Locharla, G. R., Pławiak, P., Datti, V., & Karri, C. (2022). Design of a Gabor Filter-Based Image Denoising Hardware Model. *Electronics*, 11(7), 1063. <https://doi.org/10.3390/electronics11071063>
- 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).
- Indraswari, R., Herulambang, W., & Rokhana, R. (2022). 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>
- Jason Brownlee. (2020, Agustus 19). What Is Argmax in Machine Learning? What Is Argmax in Machine Learning? <https://machinelearningmastery.com/argmax-in-machine-learning/>
- 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>
- 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>
- 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>
- Siswanto, A., Fadlil, A., & Yudhana, A. (2020). Ekstraksi Ciri Metode Gray Level Co-Occurrence Matrix Untuk Identifikasi Sel Darah Putih. *JOINTECS (Journal of Information Technology and Computer Science)*, 5(2), 71. <https://doi.org/10.31328/jointecs.v5i2.1334>
- World Health Organization. (2019). World report on vision. World Health Organization. <https://iris.who.int/handle/10665/328717>
- World Health Organization. (2023, Agustus). Blindness and Vision Impairment. <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment>
- Wulandari, N. P., & Fitrihanah, D. (2021). Analisa Perbandingan Algoritma CNN Dan MLP Dalam Mendeteksi Penyakit COVID-19 Pada Citra X-Ray Paru.
- Yadav, H., & Mallick, S. (2024). 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>
- 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>

