

COMPARATIVE STUDY OF CNN-BASED ARCHITECTURES ON EYE DISEASES CLASSIFICATION USING FUNDUS IMAGES TO AID OPHTHALMOLOGIST

Evelyn Callista Yaurentius^{*1}, Theresia Ratih Dewi Saputri², Evan Tanuwijaya³, Richard Evan Sutanto⁴

^{1,2,3,4}Informatics, School Of Information Technology, Universitas Ciputra Surabaya, Indonesia
Email: ¹ecallista01@student.ciputra.ac.id, ²theresia.ratih@ciputra.ac.id, ³evan.tanuwijaya@ciputra.ac.id,
⁴richard.evan@ciputra.ac.id

(Article received: September 19, 2024; Revision: October 30, 2024; published: February 20, 2025)

Abstract

Eye health has a significant impact on quality of life, with more than 2.2 billion people experiencing vision problems. Many of these cases can be prevented or treated. The use of AI for eye disease classification helps healthcare professionals provide optimal care. However, the complexity of fundus images challenges classification performance. This study examines various Convolutional Neural Network (CNN) architectures using Transfer Learning and Adam optimization. Fundus images are processed using CLAHE (clip limit and grid size) and the Wiener filter (size) to enhance contrast and reduce noise. Afterward, ResNet-152, EfficientNet, MobileNetV1, and DenseNet-121 are tested to identify the most effective model. The study aims to determine the optimal CNN architecture for eye disease classification, assisting ophthalmologists in diagnosing eye diseases through fundus images. The best CNN model, ResNet-152, achieved an accuracy of 94.82%, outperforming other models by 3.95 - 8.29%.

Keywords: Classification, Convolutional Neural Network, Deep Learning, Fundus Image, Image Processing.

STUDI KOMPARATIF ARSITEKTUR BERBASIS CNN TERHADAP KLASIFIKASI PENYAKIT MATA MENGGUNAKAN CITRA FUNDUS UNTUK MEMBANTU DOKTER OFTALMOLOGI

Abstrak

Kesehatan mata berdampak signifikan terhadap kualitas hidup, dengan lebih dari 2,2 miliar orang mengalami gangguan penglihatan. Banyak di antaranya dapat dicegah atau diobati. Penggunaan AI untuk klasifikasi penyakit mata membantu profesional kesehatan memberikan perawatan optimal. Namun, kompleksitas gambar fundus menantang kinerja klasifikasi. Penelitian ini mengkaji berbagai arsitektur CNN menggunakan Transfer Learning dan optimasi Adam. Gambar fundus diproses dengan CLAHE (clip limit dan grid size) dan filter Wiener (ukuran) untuk meningkatkan kontras dan mengurangi *noise*. Setelah itu, ResNet-152, EfficientNet, MobileNetV1, dan DenseNet-121 diuji untuk mengidentifikasi model yang paling efektif. Penelitian bertujuan menentukan arsitektur CNN yang optimal untuk klasifikasi penyakit mata, membantu dokter oftalmologis dalam melakukan klasifikasi penyakit mata melalui gambar fundus. Hasil model CNN terbaik diperoleh dari ResNet-152 dengan nilai akurasi mencapai 94.82% mengungguli model lain sebesar 3.95 - 8.29%.

Kata kunci: Classification Neural Network, Deep Learning, Gambar Fundus, Klasifikasi, Pemrosesan Gambar.

1. PENDAHULUAN

Memastikan perawatan mata yang optimal sangatlah penting, mengingat potensi ancaman berbagai penyakit mata. Kesehatan mata kita secara signifikan mempengaruhi kualitas hidup secara keseluruhan, mencakup aspek-aspek seperti aktivitas fisik, kesejahteraan mental, dan interaksi sosial. Secara global, setidaknya 2,2 miliar orang mengalami gangguan penglihatan dengan berbagai tingkat keparahan, di mana sekitar satu miliar kasus sebenarnya dapat dicegah atau tidak mendapatkan

pengobatan yang memadai, dan penyebab utamanya adalah kesalahan refraksi, katarak, dan glaukoma [1]. Selain itu, degenerasi makula terkait usia dan retinopati diabetik juga berkontribusi terhadap hal ini [2].

Diagnosa penyakit mata sejak dini sangat penting untuk mencegah gangguan penglihatan. Meskipun demikian, kompleksitas gambar fundus dapat menghadirkan tantangan bagi tenaga kedokteran. Hal ini mengakitbatkan terjadinya peningkatan kebutuhan akan sistem *Eye Diseases*

Classification (EDC) dengan memanfaatkan *Artificial Intelligence* (AI) yang mampu mengidentifikasi penyakit mata pada tahap awal [3]. Adapun kesulitan untuk mendeteksi perubahan morfologis yang meliputi mikroaneurisma, eksudat keras, eksudat lunak (*cotton wool spot*), perdarahan, makula, diskus optik, kepala saraf optik, dan peningkatan pembuluh darah dalam gambar fundus masih merupakan tugas yang kompleks. Perubahan morfologis ini dapat dideteksi baik melalui inspeksi manual yang memakan waktu atau dengan diagnosis berbantuan komputer yang dapat membantu dokter mata mengidentifikasi masalah. Hipertensi Retinopati (HR) dan Diabetik Retinopati (DR) merupakan dua contoh masalah kesehatan mata paling umum yang biasanya dideteksi melalui gambar fundus [4].

Metode tradisional seperti analisis citra fundus, retinal, dan *optical coherence tomography* (OCT) oleh dokter spesialis mata merupakan proses yang membutuhkan tenaga dan waktu yang signifikan. Selain itu, jumlah citra medis yang terus meningkat menimbulkan kekurangan tenaga ahli yang dapat menganalisisnya. Hal ini menyebabkan keterlambatan dalam diagnosis dan pengobatan. Konsistensi hasil diagnosis juga menjadi tantangan, karena penilaian dari dokter yang berbeda seringkali tidak seragam, dan keterlibatan banyak dokter dalam satu diagnosis dapat semakin memperlambat proses penanganan pasien [5]. Untuk mengatasi tantangan ini, otomatisasi dapat diimplementasikan melalui penerapan *machine learning*. *Convolutional Neural Network* (CNN) sangat tepat untuk menangani tugas-tugas tersebut karena mereka memiliki kemampuan untuk mengekstrak informasi hierarki dari gambar [6]. *Convolutional Neural Network* (CNN) merupakan jenis jaringan saraf tiruan yang berkembang dari *Multilayer Perceptron* (MLP). Arsitektur ini dapat dilatih dan terdiri dari beberapa tahap. Berbeda dengan MLP yang terutama memproses array satu dimensi, CNN menangani array input mulai dari dua dimensi. CNN umumnya digunakan dalam menganalisis gambar visual seperti *object recognition* dan *detection* [7].

Berbagai model arsitektur CNN yang telah diterapkan untuk klasifikasi gambar fundus penyakit mata, seperti katarak dan glaukoma. Metode yang umum digunakan termasuk EfficientNet-B0 [8], AlexNet, DenseNet, Inception-V3, InceptionRes, NetV2, Lenet-5, MobileNetV2, ResNet50, VGG16, dan VGG19 [9]. Efektivitas metode ini bervariasi, dengan akurasi berkisar antara 67% hingga 79,22%. Pemilihan arsitektur CNN dalam penelitian ini EfficientNetB0, ResNet-152, MobileNetV1, dan DenseNet-121 mencerminkan kinerja yang terbukti dalam tugas *image recognition*. EfficientNetB0 dikenal karena efisiensinya, dibandingkan dengan model CNN lainnya, EfficientNet dengan jumlah parameter dan *floating-point operations per second*

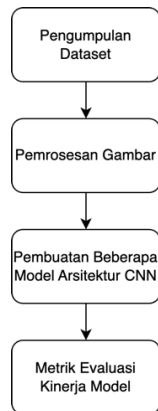
(FLOPS) yang jauh lebih sedikit, menjadikannya sangat efisien dalam penggunaan sumber daya komputasi [10]. Arsitektur ResNet-152 memperkenalkan konsep *residual blocks*, input asli dari suatu lapisan dilewati langsung lapisan berikutnya melalui koneksi *shortcut* sehingga memungkinkan jaringan untuk mempertahankan informasi penting dan memastikan bahwa gradien tetap cukup besar untuk mendukung pembelajaran yang efektif sehingga dapat mengatasi masalah degradasi kinerja dan *vanishing-gradient* [4]. DenseNet dapat mengatasi masalah *vanishing-gradient*, memperkuat propagasi fitur, mendorong penggunaan kembali fitur, dan membuat *training* lebih efisien [11]. MobileNetV1 memperkenalkan *depthwise separable convolutions*. Pendekatan ini memisahkan *spatial filtering* dari pembuatan fitur, dengan membagi konvolusi tradisional menjadi dua lapisan terpisah sehingga secara signifikan mengurangi parameter model dan kebutuhan komputasi jika dibandingkan dengan model lainnya, dengan hanya sedikit penurunan akurasi [12].

Peneliti akan menerapkan metode preprocessing *Contrast Limited Adaptive Histogram Equalization* (CLAHE) untuk meningkatkan kualitas gambar melalui peningkatan kontras adaptif. Metode ini umumnya digunakan pada bidang *biomedical* seperti gambar mamografi, ultrasonografi, segmentasi gambar sel, gambar patah tulang, dan deteksi pembuluh darah retina. Berdasarkan studi terdahulu didapatkan metode CLAHE berhasil meningkatkan akurasi model CNN sebesar 13% [13][14]. *Transfer Learning* digunakan untuk meningkatkan pengumpulan data dalam *machine learning* menggunakan data yang telah disusun sebelumnya sehingga menghasilkan hasil yang lebih logis, cepat dan baik [15]. Beberapa penelitian terdahulu yang akan digunakan sebagai acuan penelitian diantara lain penelitian penerapan MobileNet, VGG-19, ResNet50V2, DenseNet-121, dan NASNetMobile pada dataset wayang, *Drone-Bird*, dan *Grapevine Leaves* [16], penerapan EfficientNet untuk klasifikasi 4 kelas penyakit mata [17], klasifikasi penyakit mata retinopati diabetik menggunakan ResNet-152 [18], serta penelitian deteksi otomatis penyakit retinopati diabetik dan hipertensi menggunakan ResNet-101 [4]. Dalam penelitian ini, peneliti akan mencoba melakukan analisis komparasi lapisan arsitektur model Resnet-152, EfficientNetB0, MobileNetV1, dan DenseNet-121 dengan memanfaatkan model *transfer learning*.

2. METODE PENELITIAN

Penelitian analisis pengaruh arsitektur model CNN terhadap klasifikasi penyakit mata menggunakan citra fundus ini dilakukan dengan menggunakan beberapa tahap yang terdiri dari: pengumpulan *dataset*, pemrosesan gambar, pembuatan model arsitektur CNN, EfficientNetB0t, MobileNetV1, ResNet-152, dan DenseNet-121, dan

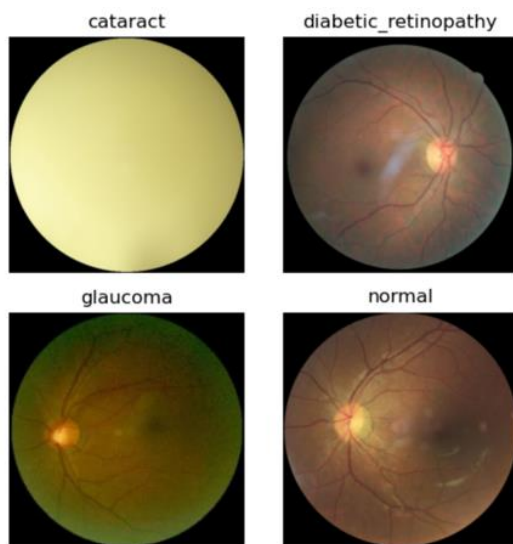
analisis perbandingan hasil arsitektur yang digunakan. Skema proses penelitian sistem ditunjukkan pada Gambar 1.



Gambar 1. Skema Proses Penelitian

2.1. Pengumpulan Dataset

Penelitian ini menggunakan *dataset* sekunder publik sesuai dengan Gambar 1 yang menunjukkan skema proses penelitian *dataset* yang digunakan berupa gambar fundus RGB beberapa penyakit mata, yang terdiri dari kelas *cataract*, *diabetic retinopathy*, *glaucoma*, dan normal [19]. Jumlah gambar terdiri dari 1038 gambar *cataract*, 1098 gambar *diabetic retinopathy*, 1007 gambar *glaucoma*, dan 1074 gambar normal. Dimensi ukuran gambar *cataract* dan *glaucoma* adalah 256x256 sedangkan dimensi gambar *diabetic retinopathy* dan normal adalah 512x512. Contoh tampilan citra fundus dari tiap kelas dapat dilihat pada Gambar 2.



Gambar 2. Contoh Citra Fundus yang Digunakan

2.2. Pemrosesan Gambar

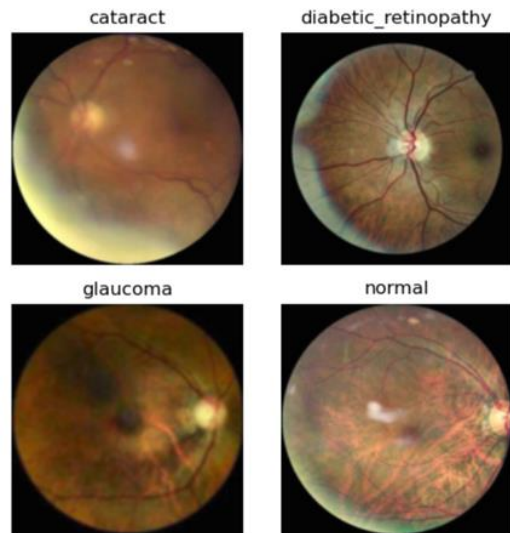
Tahap selanjutnya dari skema penelitian sesuai dengan Gambar 1 yaitu pemrosesan gambar. *Dataset* akan di *resize* menjadi format 224x224 kemudian dibagi menjadi 80% *training*, 10% *validation*, dan 10% test. *Dataset* akan ditingkatkan kontrasnya

menggunakan CLAHE dengan tujuan mempermudah pengolahan gambar. CLAHE akan diakses melalui *library cv2*. *Dataset* kemudian akan di filter menggunakan *wiener filter* untuk menghilangkan *noise* pada gambar akibat proses CLAHE. *Hyperparameter* yang akan diterapkan pada metode CLAHE antara lain *clip limit* dan *tileGridSize*. Metode *wiener filter* akan diakses menggunakan *library scipy.signal*. Parameter utama yang akan digunakan adalah *mysize*. Nilai dari *hyperparameter* CLAHE dan Wiener Filter dapat dilihat pada Tabel 1.

Tabel 1. Nilai Hyperparameter CLAHE dan Wiener Filter

Metode	Parameter	Nilai
CLAHE	clip limit	0.5
	tileGridSize	(8,8)
Wiener Filter	mysize	5

Ukuran ini mengontrol seberapa banyak informasi di sekitar setiap piksel yang akan digunakan dalam perhitungan filter. Diharapkan dengan menerapkan *image processing* ini akan meningkatkan kontras gambar serta menghilangkan *noise*. Contoh tampilan citra fundus dari tiap kelas setelah penerapan *image processing* CLAHE dapat dilihat pada Gambar 3.



Gambar 3. Contoh Citra Fundus Setelah Penerapan CLAHE

2.3. Pembuatan Model Arsitektur CNN

Penelitian akan menerapkan *hyperparameter finetuning* terhadap beberapa arsitektur CNN antara lain Resnet-152, EfficientNet, MobileNetV1, dan DenseNet-121. Model diinisialisasi dengan bobot ImageNet. *input_shape* model akan menggunakan (224, 224, 3). Penelitian ini akan menerapkan perbandingan beberapa nilai *parameter* yang dapat dilihat pada Tabel 2.

Tabel 2. Nilai Perbandingan Hyperparameter Model

Parameter	Nilai
Epoch	25 & 50
Neuron	128, 256, 512, & 1024
Activation	ReLU, Elu, Swish, Gelu, & Leaky Relu
Regularization	0.01 & 0.001

Optimasi metode deep learning akan membandingkan beberapa *optimizer* antara lain, RMS, Adam, dan SGD. Peneliti juga akan menerapkan *early stopping* dengan memonitor *val_loss* serta *patience* 5. Penerapan *transfer learning* yang digunakan akan diambil dari Keras dengan asal *dataset* imagenet *Library* yang akan digunakan tensorflow, matplotlib, cv2, os, numpy, dan scipy.signal *Framework* yang digunakan Jupyter Notebook untuk pelatihan model *deep learning* dan *image processing*.

2.4. Metrik Evaluasi Kinerja Model

Untuk mengevaluasi kinerja model peneliti akan menerapkan metrik confusion matrix adalah metode untuk mengukur kinerja model klasifikasi dalam *machine learning* yang menghasilkan output berupa dua kelas atau lebih. Matriks ini terdiri dari empat komponen utama yang menggambarkan kinerja model: *true positive*, *true negative*, *false positive*, dan *false negative*. Berdasarkan *confusion matrix* ini, *classification report* dapat dihasilkan untuk memberikan metrik seperti *accuracy*, *precision*, *recall*, dan *f1-score* yang dapat dihitung menggunakan persamaan berikut.

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

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

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$f1 - score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

Pada persamaan (1), *accuracy* mengukur seberapa sering model membuat prediksi yang benar secara keseluruhan, baik positif maupun negatif. Pada persamaan (2), *precision* mengukur seberapa tepat prediksi model untuk suatu kelas. Pada persamaan (3), *recall* mengukur seberapa baik model mengidentifikasi semua contoh dari suatu kelas. Persamaan (4), *f1-score* mengukur rata-rata harmonis dari *precision* dan *recall*. Selain itu, *support* menunjukkan jumlah total sampel sebenarnya untuk kelas tersebut dalam data uji. *Classification* report juga mencakup *macro avg*, yaitu rata-rata sederhana dari *precision*, dan *recall* untuk semua kelas, serta *weighted avg*, yaitu rata-rata tertimbang dari *precision*, dan *recall* untuk semua kelas dengan bobot berdasarkan jumlah sampel tiap kelas.

3. HASIL DAN PEMBAHASAN

Pada hasil dan pembahasan penulis akan membahas pengujian beberapa model CNN dengan arsitektur seperti ResNet-152, EfficientNetB0, MobileNetV1, dan DenseNet-121. Pengujian

dilakukan dengan menggunakan berbagai kombinasi hyperparameter, termasuk optimasi, jumlah neuron, dan fungsi aktivasi, untuk menentukan konfigurasi yang memberikan hasil terbaik. Analisis terhadap hasil akurasi dan loss model akan dijelaskan secara mendetail, di mana model ResNet-152 digunakan sebagai patokan utama. Selain itu, bab ini juga akan mengkaji dampak perubahan pada jumlah layer Dense dan jenis aktivasi terhadap kinerja model, serta mengevaluasi kombinasi yang menghasilkan akurasi tertinggi.

3.1. Pengujian Model CNN

Hasil training beberapa model CNN yaitu ResNet-152, EfficientNetB0, MobileNetV1, dan DenseNet-121 dengan menggunakan beberapa kombinasi *hyperparameter* yang telah ditetapkan mendapat hasil yang bervariasi. Peneliti menggunakan ResNet-152 sebagai patokan akurasi model serta optimasi Adam serta *dropout layer* 0.15. Untuk data yang lebih lengkap dapat dilihat pada Tabel 3.

Tabel 3. Hasil Akurasi Pengujian Model CNN

Model	Accuracy (%)	Loss
ResNet-152	94.82	0.2565
DenseNet121	86.53	0.4237
EfficientNetB0	90.87	0.3307
MobileNetV1	88.31	0.4027

Tabel 3 menunjukkan hasil akurasi model tertinggi diraih oleh arsitektur ResNet-152. Semua kombinasi menunjukkan angka diatas 85%. Peneliti kemudian mencoba melakukan percobaan terhadap beberapa kombinasi jumlah neuron. peneliti mencoba menambahkan *layer* Dense yang dapat dilihat pada Tabel 2.

Tabel 4. Hasil Akurasi Pengujian Jumlah Neuron

Layer	ResNet-152(%)	DenseNet 121(%)	EfficientNetB0(%)	MobileNetV1(%)
1024 & 1024	93.33	87.57	92.94	86.66
512 & 512	94.10	87.01	92.84	88.20
256 & 256	93.33	84.11	93.83	78.38
128 & 128	92.19	87.96	87.77	83.34
1024 & 512	90.58	88.79	94.02	86.35
1024 & 256	91.30	89.99	93.16	89.53
1024 & 128	92.65	84.65	93.08	81.13
512 & 256	92.24	92.07	91.06	78.70
512 & 128	94.67	89.56	93.13	74.05
256 & 128	92.03	91.85	86.36	74.21

Tabel 4 menunjukkan hasil akurasi dari berbagai kombinasi jumlah neuron pada dua lapisan model CNN. Hasil terbaik dicapai dengan kombinasi 512 & 128 neuron pada ResNet-152, yang memberikan akurasi 94.67%. Beberapa pola menarik terlihat, seperti penggunaan neuron yang sama besar (misalnya 1024 & 1024) menghasilkan akurasi lebih rendah (93.33%) dibanding kombinasi neuron yang lebih kecil di lapisan kedua, seperti 512 & 128 (94.67%). Ini menunjukkan bahwa kombinasi lapisan dengan jumlah neuron yang berbeda membantu

mengurangi *overfitting* dan meningkatkan generalisasi model, seperti terlihat pada perbandingan 1024 & 1024 dengan 512 & 512. Selain itu, penggunaan jumlah neuron yang lebih besar di kedua lapisan, seperti 1024 & 1024 menghasilkan akurasi lebih rendah dibanding kombinasi yang lebih seimbang. Misalnya, kombinasi 1024 & 1024 memberikan akurasi lebih rendah daripada 512 & 512 pada ResNet-152. Hal ini mungkin disebabkan oleh *overfitting* karena terlalu banyak *parameter*, sehingga model sulit untuk generalisasi pada data uji. Penurunan dibandingkan hasil sebelumnya yaitu penerapan 1 *layer* Dense. Peneliti kemudian mencoba menerapkan perbandingan jenis aktivasi yang digunakan antara lain ReLu, Elu, Swish, Gelu dan Leaky (Leaky Relu) yang dapat dilihat pada Tabel 3.

Tabel 5. Hasil Akurasi Pengujian Jenis Aktivasi

Aktivasi	ResNet-152 (%)	Dense Net 121 (%)	Efficient NetB0 (%)	Mobile NetV1 (%)
2x ReLu	94.10	87.01	92.84	88.20
2x Elu	91.34	91.55	91.60	88.28
2x Swish	93.76	84.75	93.11	75.20
2x Gelu	91.61	84.29	92.71	80.28
2x Leaky	90.95	84.55	93.83	85.21
ReLu & Elu	92.94	90.59	93.48	81.89
ReLu & Swish	92.75	86.77	83.15	72.58
ReLu & Gelu	94.36	87.25	91.96	79.59
ReLu & Leaky	93.72	89.97	94.39	81.71
Elu & Swish	87.18	83.03	89.11	82.96
Elu & Gelu	88.40	89.98	91.36	84.06
Elu & Leaky	93.25	82.96	93.14	80.07
Swish & Gelu	92.43	85.20	92.92	86.39
Swish & Leaky	92.77	88.41	92.01	78.19
Gelu & Leaky	90.89	79.35	88.05	80.94

Tabel 5 menunjukkan hasil perbandingan penerapan *activation* ReLu, Elu, Swish, Gelu, dan Leaky (Leaky Relu). Akurasi tertinggi diperoleh dari penerapan 2 *layer* Dense dengan kombinasi aktivasi ReLu dan Gelu pada ResNet-152 sebesar 94.36%, menunjukkan bahwa kombinasi linearitas ReLU dan Gelu meningkatkan performa. Namun, hasil ini masih lebih rendah dibandingkan dengan penggunaan 1 *layer* Dense aktivasi ReLu pada Tabel 3 sebelumnya. Aktivasi ReLU menunjukkan performa baik di beberapa model, seperti DenseNet 121 dan EfficientNetB0, sementara kombinasi seperti Elu & Swish memberikan kinerja yang kurang stabil. Selain itu, 2x Elu pada DenseNet 121 stabil dengan akurasi 91.55%, dan kombinasi Swish & Gelu memberikan hasil baik pada EfficientNetB0 dengan 92.92%. Kombinasi aktivasi campuran seperti ReLU & Leaky pada MobileNetV1 juga menunjukkan peningkatan performa hingga 81.71%, dibandingkan dengan ReLU & Swish yang hanya mencapai 72.58%. Secara keseluruhan, kombinasi aktivasi yang berbeda dapat membantu menyeimbangkan performa dengan memanfaatkan keunggulan masing-masing aktivasi dalam menangani *non-linearity* dan *gradient flow*.

Dari hasil uji coba pelatihan model didapatkan bahwa penggunaan urutan *fully connected layer* seperti pada Tabel 6.

Tabel 6. Fully Connected Layer

Layer	Neuron (Layer)	Aktivasi
Flatten()		
Dense	512	ReLu
Dropout	0.15	
Dense	4	Softmax

Serta penerapan Adam *optimizer*, 25 *epoch* dan *early stopping* yang memonitor *val_loss* serta *patience* 5. Nilai akurasi untuk model CNN yang digunakan dapat dilihat pada Tabel 3.

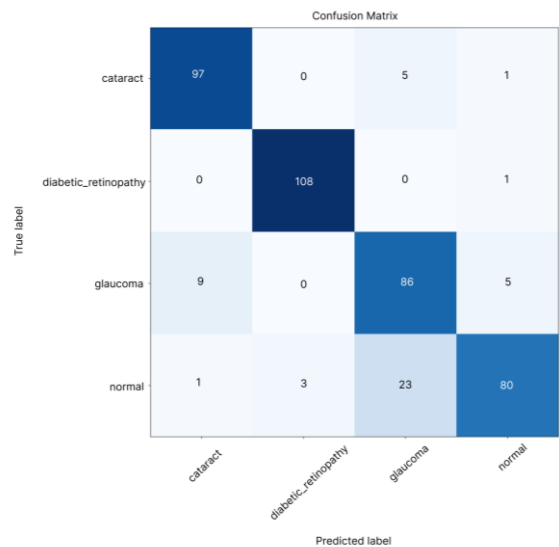
Peneliti melakukan komparasi *f1-score* untuk model yang digunakan yang dapat dilihat pada Tabel 7.

Tabel 7. Perbandingan F1-Score Model

Model	F1-Score
ResNet-152	0.89
DenseNet121	0.84
EfficientNetB0	0.91
MobileNetV1	0.84

Berdasarkan Tabel 7, EfficientNetB0 dengan nilai *f1-score* 0.91 menunjukkan kinerja terbaik dalam keseimbangan antara *precision* dan *recall*. Kemudian diikuti dengan ResNet-152 dengan nilai *f1-score* 0.89. Sementara itu, DenseNet121 dan MobileNetV1 keduanya memiliki skor yang sama, yaitu 0.84, yang menandakan performa lebih rendah dibandingkan dua model sebelumnya. Model-model ini menunjukkan variasi kinerja dalam mengklasifikasikan penyakit mata berdasarkan gambar fundus.

Peneliti melakukan komparasi *fine tuning hyperparameter* untuk mendapatkan performa terbaik dan dapat dilihat melalui *confusion matrix*. Hasil *Confusion matrix* untuk model ResNet-152 dapat dilihat pada Gambar 3.

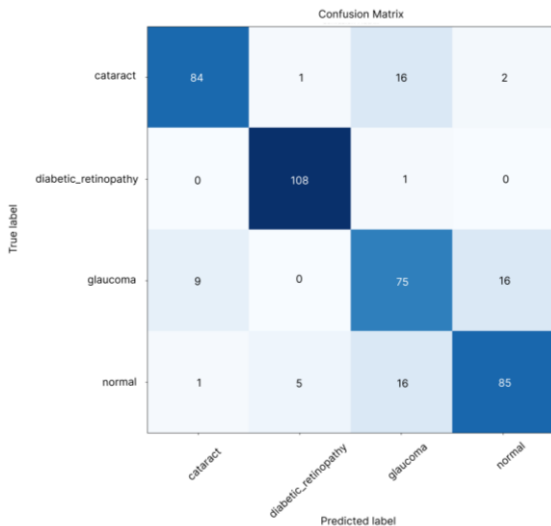


Gambar 3. Confusion Matrix ResNet-152

Berdasarkan Gambar 3, hasil evaluasi dengan *confusion matrix* menunjukkan bahwa model ResNet-152 memberikan performa terbaik pada kelas *diabetic_retinopathy*, dengan 108 *True Positive* dan

hanya 1 kesalahan prediksi sebagai kelas normal. Namun, model mengalami kesulitan dalam mengklasifikasikan kelas normal, dengan hanya 80 *True Positive* dan 29 kesalahan prediksi, di mana sebagian besar dari kesalahan tersebut diklasifikasikan sebagai *glaucoma* (23 kasus). Model juga menunjukkan tantangan dalam membedakan antara kelas *glaucoma* dan *cataract*, yang tercermin dari jumlah *False Positives* pada kedua kelas tersebut.

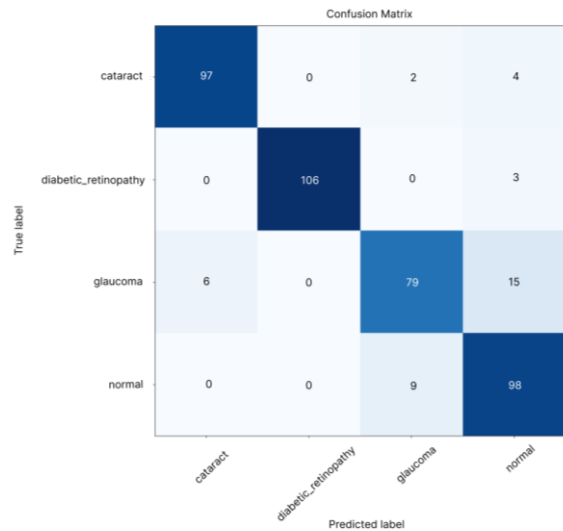
Hasil *Confusion matrix* untuk model DenseNet121 dapat dilihat pada Gambar 4.



Gambar 4. Confusion Matrix DenseNet121

Berdasarkan Gambar 4, hasil evaluasi dengan *Confusion Matrix* menunjukkan bahwa model DenseNet121 memberikan performa terbaik pada kelas *diabetic_retinopathy*, dengan 108 *True Positive* dan hanya 1 kesalahan prediksi sebagai kelas *glaucoma*. Namun, model menunjukkan tantangan dalam mengklasifikasikan kelas *glaucoma* dan normal. Untuk kelas *glaucoma*, model hanya menghasilkan 75 *True Positive* dan membuat 25 kesalahan prediksi, dengan sebagian besar kesalahan diklasifikasikan sebagai normal (16 kasus). Selain itu, untuk kelas *cataract*, model menghasilkan 84 *True Positive*, tetapi membuat 19 kesalahan prediksi, yang sebagian besar diklasifikasikan sebagai *glaucoma* (16 kasus). Meskipun demikian, model berhasil mengklasifikasikan kelas normal dengan cukup baik, dengan 85 *True Positive* dan 22 kesalahan prediksi, di mana 16 dari kesalahan ini terjadi pada kelas *glaucoma*.

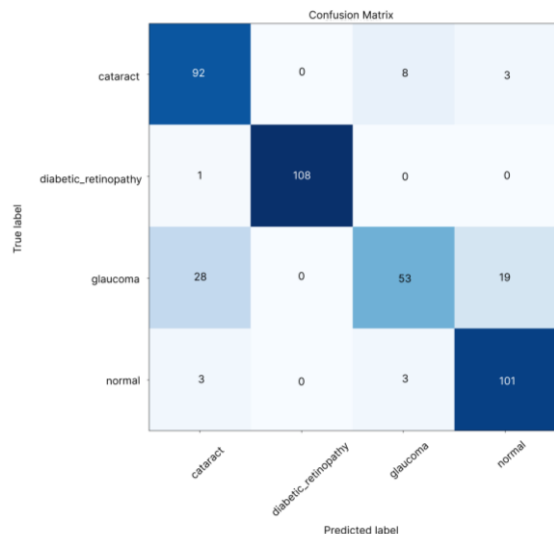
Hasil *Confusion matrix* untuk model EfficientNetB0 dapat dilihat pada Gambar 5.



Gambar 5. Confusion Matrix EfficientNetB0

Berdasarkan Gambar 5, hasil evaluasi dengan *Confusion Matrix* menunjukkan bahwa model EfficientNetB0 memberikan performa terbaik pada kelas normal, dengan 98 *True Positive* dan hanya 9 kesalahan prediksi yang semuanya diklasifikasikan sebagai *glaucoma*. Model juga menunjukkan kinerja yang kuat pada kelas *diabetic_retinopathy*, dengan 106 *True Positive* dan 3 kesalahan prediksi yang semuanya diklasifikasikan sebagai normal. Namun, untuk kelas *glaucoma*, model menunjukkan kesulitan dengan hanya 79 *True Positive* dan membuat 21 kesalahan prediksi, di mana 15 dari kesalahan tersebut diklasifikasikan sebagai normal dan 6 sebagai *cataract*. Sementara itu, untuk kelas *cataract*, model berhasil mengklasifikasikan dengan baik, menghasilkan 97 *True Positive*, tetapi membuat 6 kesalahan prediksi yang sebagian besar diklasifikasikan sebagai *glaucoma* (2 kasus) dan normal (4 kasus).

Hasil *Confusion matrix* untuk model MobileNetV1 dapat dilihat pada Gambar 6.



Gambar 6. Confusion Matrix MobileNetV1

Berdasarkan Gambar 6, hasil evaluasi dengan *Confusion Matrix* menunjukkan bahwa model MobileNetV1 memberikan performa terbaik pada kelas *diabetic_retinopathy* dan normal, dengan masing-masing 108 dan 101 *True Positive* serta sangat sedikit kesalahan prediksi (1 dan 6 kesalahan, masing-masing). Namun, model menunjukkan tantangan yang signifikan dalam mengklasifikasikan kelas *glaucoma*, dengan hanya 53 *True Positive* dan 47 kesalahan prediksi. Sebagian besar kesalahan pada kelas *glaucoma* diklasifikasikan sebagai *cataract* (28 kasus) dan normal (19 kasus). Untuk kelas *cataract*,

model menghasilkan 92 *True Positive*, tetapi membuat 11 kesalahan prediksi, sebagian besar diklasifikasikan sebagai *glaucoma* (8 kasus) dan normal (3 kasus).

Peneliti melakukan uji statistik untuk membandingkan perbedaan kinerja antar model dengan mengambil masing-masing 10 gambar dari tiap kelas (*cataract*, *diabetic_retinopathy*, *glaucoma*, dan normal) dan mencoba melakukan prediksi menggunakan keempat model untuk gambar-gambar tersebut. Beberapa sampel hasil data yang digunakan dapat dilihat pada Tabel 9.

Tabel 9. Hasil Sampel Uji Statistik Model

Sampel	True Label	ResNet-152	DenseNet 121	EfficientNetB0	MobileNetV1
1	cataract	cataract	glaucoma	cataract	cataract
2	cataract	cataract	cataract	cataract	cataract
3	diabetic	diabetic	diabetic	diabetic	diabetic
4	diabetic	diabetic	diabetic	normal	diabetic
5	glaucoma	glaucoma	glaucoma	cataract	glaucoma
6	glaucoma	glaucoma	glaucoma	glaucoma	glaucoma
7	glaucoma	glaucoma	cataract	cataract	cataract
8	normal	normal	normal	normal	normal
9	normal	normal	diabetic	normal	normal
10	normal	normal	cataract	normal	glaucoma

Tabel 9 menunjukkan hasil nilai F sebesar 2.935 dan p-value sebesar 0.059. Meskipun hasil ini menunjukkan adanya perbedaan kinerja di antara model-model yang diuji, nilai p yang sedikit di atas ambang signifikansi 0.05 ($p > 0.05$) mengindikasikan bahwa perbedaan tersebut tidak cukup kuat untuk dianggap signifikan secara statistik pada tingkat kepercayaan 95%. Dengan demikian, berdasarkan analisis ini, tidak terdapat bukti statistik yang cukup untuk menyimpulkan adanya perbedaan yang signifikan dalam kinerja prediksi model-model yang dibandingkan.

4. DISKUSI

Hasil akurasi yang didapat pada penelitian ini antara beberapa arsitektur model CNN berada pada rentang 89.54% hingga 94.82%. Penelitian ini menjawab saran dari penelitian referensi ke-1 untuk mencoba menerapkan arsitektur selain yang telah digunakan di penelitian referensi ke-1 yaitu ResNet-152 dan EfficientNetB0 [13].

Perbedaan dengan hasil referensi ke-1 [13], terdapat pada jenis penggunaan *dataset* yang digunakan dimana pada penelitian ini menggunakan dataset dari 4 jenis penyakit mata sedangkan pada referensi ke-1 menggunakan 3 *dataset* berbeda yaitu *drone-bird*, *wayang*, dan *grapevine*. Pada referensi ke-1 arsitektur DenseNet-121 mendapatkan akurasi tertinggi dibandingkan dengan model lain yang digunakan yaitu 80%. Sedangkan pada penelitian ini berhasil mendapatkan akurasi 87% untuk penerapan arsitektur Dense-Net-121.

Perbedaan hasil penelitian ini dengan referensi ke-2 [14] terletak pada penggunaan arsitektur CNN. Penelitian sebelumnya menggunakan arsitektur EfficientNet dan berhasil mencapai akurasi sebesar 94%. Dalam penelitian ini, arsitektur EfficientNet

juga diterapkan, namun menghasilkan akurasi tertinggi sebesar 90,87%. Perbedaan ini kemungkinan disebabkan oleh perbedaan jenis EfficientNet yang digunakan, di mana penelitian sebelumnya tidak secara spesifik menyebutkan varian EfficientNet yang digunakan, serta perbedaan dalam penggunaan *hyperparameter*. Meskipun demikian, penelitian ini menggunakan ResNet-152 sebagai nilai *baseline* dan berhasil mencapai akurasi sebesar 94,82%. ResNet-152 menunjukkan performa yang lebih unggul pada kelas *cataract* dan *diabetic_retinopathy*, dengan jumlah *True Positive* yang lebih tinggi dan kesalahan prediksi yang lebih rendah dibandingkan EfficientNet, model ini mengalami lebih banyak kesalahan pada kelas normal. Di sisi lain, penerapan EfficientNet pada referensi menunjukkan performa yang lebih baik pada kelas normal, meskipun kurang akurat pada kelas *cataract* dan *diabetic_retinopathy*. Hasil ini menunjukkan bahwa pemilihan arsitektur dan pengaturan *hyperparameter* yang tepat sangat penting untuk mendapatkan hasil yang optimal pada tugas klasifikasi citra fundus.

Perbedaan dengan hasil referensi ke-3 [15], terdapat pada jumlah kelas yang digunakan. Pada referensi dataset yang digunakan berfokus pada tingkat keparahan *diabetic_retinopathy*. Sedangkan pada penelitian ini berfokus pada klasifikasi jenis penyakit mata yang berbeda yaitu *diabetic_retinopathy*, *glaucoma*, *cataract*, dan normal.

Perbedaan dengan hasil referensi ke-4 [4], terdapat pada jenis kelas *dataset* yang digunakan. Kelas *dataset* yang digunakan pada referensi antara lain adalah normal, *mild diabetic*, *moderate diabetic*, *severe diabetic*, dan *hypertensive*. 3 dari kelas tersebut berfokus pada *tingkat* keparahan *diabetic*.

Sehingga penyakit yang digunakan terbatas pada mata *diabetic* dan *hypertensive*.

Penelitian ini menunjukkan bahwa setiap arsitektur CNN memiliki kekuatan dan kelemahan dalam klasifikasi mata. ResNet-152 lebih sering mengalami kesalahan dalam membedakan kelas *glaucoma* dan normal, sementara DenseNet121 kesulitan dalam membedakan antara kelas *glaucoma* dan *cataract*. EfficientNetB0 secara konsisten membuat kesalahan pada kelas *glaucoma*, meskipun memiliki performa yang sangat baik pada kelas normal dan *diabetic retinopathy*. Kesalahan pada MobileNetV1 sebagian besar terkait dengan prediksi kelas *glaucoma* dan *cataract*, yang menunjukkan bahwa karakteristik visual penyakit-penyakit ini memerlukan perhatian lebih dalam pemodelan.

Ukuran dataset dan distribusi kelas mungkin menjadi salah satu keterbatasan penelitian ini, mengingat beberapa kelas memiliki kesalahan prediksi yang cukup besar, terutama pada kelas *glaucoma*. *Dataset* yang lebih besar dan lebih seimbang dapat membantu model untuk lebih baik dalam mendeteksi semua kelas. Penelitian di masa mendatang sebaiknya mencakup metode untuk mengatasi bias ini, misalnya dengan augmentasi data atau penggunaan arsitektur model yang lebih kompleks. Penelitian di masa mendatang juga dapat menjajaki penggunaan ensembel model atau teknik *hybrid* untuk mengurangi kesalahan prediksi pada kelas-kelas yang sulit dibedakan, seperti *glaucoma* dan *cataract*.

5. KESIMPULAN

Penelitian ini menganalisis pengaruh arsitektur model CNN terhadap klasifikasi penyakit mata menggunakan citra fundus dengan membandingkan empat model utama: ResNet-152, DenseNet-121, EfficientNetB0, dan MobileNetV1. Berbagai *hyperparameter* diuji, termasuk jumlah *epoch*, regularisasi, dan penggunaan *early stop*. Hasil penelitian menunjukkan bahwa ResNet-152 mencapai akurasi tertinggi sebesar 94.82% dengan kombinasi 25 *epoch*, regularisasi 0.001, *early stopping*, Adam *optimizer*, dan 0.15 *dropout dense layer*. DenseNet-121, EfficientNetB0, dan MobileNetV1 juga menunjukkan kinerja yang baik, namun tidak melebihi ResNet-152. Semua performa dari model CNN menunjukkan hasil di atas 85%.

Kontribusi utama penelitian ini adalah penerapan *preprocessing* gambar menggunakan CLAHE dan *wiener filter*, yang meningkatkan akurasi klasifikasi sehingga sangat penting bagi dokter oftalmologi dalam diagnosis penyakit mata yang lebih cepat menggunakan model *deep learning* yang lebih akurat. Dengan akurasi tinggi, model ini dapat mendukung deteksi dini penyakit mata sehingga memungkinkan perawatan lebih cepat bagi pasien.

Penerapan *image preprocessing* menggunakan CLAHE dan *wiener filter* terbukti efektif dalam

meningkatkan kualitas gambar untuk klasifikasi. Penelitian ini menekankan pentingnya pemilihan arsitektur model dan pengaturan *hyperparameter* yang tepat untuk mencapai performa terbaik dalam klasifikasi penyakit mata menggunakan citra fundus. Pada penelitian ini, beberapa hal yang dapat ditingkatkan adalah jumlah data dengan menggunakan *image augmentation*, peningkatan akurasi model, dan pengolahan gambar lebih lanjut untuk menghilangkan *noise*.

UCAPAN TERIMA KASIH

Penelitian ini didanai dengan melalui hibah dana internal penelitian (DIP) dengan nomor kontrak 012/UC-LPPM/DIP/SP3H/IX/2024.

DAFTAR PUSTAKA

- [1] WHO, "Blindness and vision impairment," 2023. <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment> (accessed June. 15, 2024).
- [2] M. J. Burton, J. Ramke, A. P. Marques, R. R. A. Bourne, N. Congdon, I. Jones, B. A. M. Ah Tong, S. Arunga, D. Bachani, C. Bascaran, A. Bastawrous, K. Blanchet, T. Braithwaite, J. C. Buchan, J. Cairns, A. Cama, M. Chagunda, C. Chuluunkhuu, A. Cooper, ... H. B. Faal, "The Lancet Global Health Commission on Global Eye Health: vision beyond 2020," *The Lancet Global Health*, vol. 9, no. 4, 2021. [https://doi.org/10.1016/S2214-109X\(20\)30488-5](https://doi.org/10.1016/S2214-109X(20)30488-5).
- [3] A. R. W. Sait, "Artificial Intelligence-Driven Eye Diseases Classification Model," *Multidisciplinary Digital Publishing Institute (MDPI)*, vol. 13, no. 20, 2023. <https://doi.org/10.3390/app132011437>.
- [4] D. Nagpal, N. Alsubaie, B. O. Soufiene, M. S. Alqahtani, M. Abbas, and H. M. Almohiy, "Automatic Detection of Diabetic Hypertensive Retinopathy in Fundus Images Using Transfer Learning," *Applied Sciences (Switzerland)*, vol. 13, no. 8, 2023. <https://doi.org/10.3390/app13084695>.
- [5] C. Guo, M. Yu, and J. Li, "Prediction of different eye diseases based on fundus photography via deep transfer learning," *J. Clin. Med.*, vol. 10, no. 23, 2021. <https://doi.org/10.3390/jcm10235481>.
- [6] M. Juneja, S. Singh, N. Agarwal, S. Bali, S. Gupta, N. Thakur, and P. Jindal, "Automated detection of Glaucoma using deep learning convolution network (G-net)," *Multimedia Tools and Applications*, vol. 79, no. 21-22, 2020. <https://doi.org/10.1007/s11042-019-7460-4>.
- [7] Y. N. Fuadah, I. D. Ubaidullah, N. Ibrahim,

- F. F. Taliningsing, N. K. Sy, and M. A. Pramuditho, "Optimasi Convolutional Neural Network dan K-Fold Cross Validation pada Sistem Klasifikasi Glaukoma," *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, vol. 10, no. 3, 2022. <https://doi.org/10.26760/elkomika.v10i3.728>
- [8] Z. Arif, R. Y. Nur Fu'adah, S. Rizal, and D. Ilhamdi, "Classification of eye diseases in fundus images using Convolutional Neural Network (CNN) method with EfficientNet architecture," *JRTI (Jurnal Riset Tindakan Indonesia)*, vol. 8, no. 1, 2023. <https://doi.org/10.29210/30032835000>.
- [9] K. Saputra, et al., "Penerapan Algoritma Convolution Neural Network untuk Menentukan Retinopati Hipertensi Melalui Citra Retina Fundus," *Jurnal Penelitian Teknik Informatika Universitas Prima Indonesia (UNPRI) Medan*, vol. 6, no. 2, pp. 143-149, Oct. 2023. <https://doi.org/10.34012/jutikomp.v6i2.4307>
- [10] J. Wang, L. Yang, Z. Huo, W. He, and J. Luo, "Multi-Label Classification of Fundus Images with EfficientNet," *IEEE Access*, vol. 8, pp. 212499-212508, 2020. doi: 10.1109/ACCESS.2020.3040275
- [11] H. Li, Z. Yue, J. Liu, Y. Wang, H. Cai, K. Cui, and X. Chen, "Sccdnnet: A pixel-level crack segmentation network," *Appl. Sci.*, vol. 11, no. 11, 2021. doi: 10.3390/app11115074.
- [12] B. Jiang, N. Li, X. Cui, et al., "Research on facial expression recognition algorithm based on improved MobileNetV3," *J. Image Video Process.*, vol. 2024, no. 22, 2024. doi: 10.1186/s13640-024-00638-z.
- [13] A. Wanto, Y. Yuhandri, and O. Okfalisa, "Optimization Accuracy of CNN Model by Utilizing CLAHE Parameters in Image Classification Problems," in *Proceedings - 2023 International Conference on Networking, Electrical Engineering, Computer Science, and Technology, IConNECT 2023*, 2023. <https://doi.org/10.1109/IConNECT56593.2023.10327100>.
- [14] K. Aurangzeb, S. Aslam, M. Alhussein, R. A. Naqvi, M. Arsalan, and S. I. Haider, "Contrast Enhancement of Fundus Images by Employing Modified PSO for Improving the Performance of Deep Learning Models," *IEEE Access*, vol. 9, pp. 55253-55264, 2021. doi: 10.1109/ACCESS.2021.3068477.
- [15] A. Hosna, E. Merry, J. Gyalmo, Z. Alom, Z. Aung, and M. A. Azim, "Transfer Learning: A Friendly Introduction," *Journal of Big Data*, vol. 9, no. 102, 2022. doi: 10.1186/s40537-022-00652-w.
- [16] M. D. A. Putra, T. S. Winanto, R. Hendrowati, A. Primajaya, and F. D. Adhinata, "A Comparative Analysis of Transfer Learning Architecture Performance on Convolutional Neural Network Models with Diverse Datasets," *Komputika: Jurnal Sistem Komputer*, vol. 12, no. 1, 2023. doi: 10.34010/komputika.v12i1.8626.
- [17] T. Babaqi, M. Jaradat, A. E. Yildirim, S. Al-Nimer, and D. Won, "Eye Disease Classification Using Deep Learning Techniques," in *IISE Annual Conference and Expo 2023*, 2023. doi: 10.21872/2023IISE_1944.
- [18] U. Bhimavarapu and G. Battineni, "Deep Learning for the Detection and Classification of Diabetic Retinopathy with an Improved Activation Function," *Healthcare (Switzerland)*, vol. 11, no. 1, 2023. doi: 10.3390/healthcare11010097.
- [19] G. V. Doddi, "Eye Diseases Classification Dataset," 2022. <https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification> (accessed June. 10, 2024)..