Optimasi Model CNN untuk Klasifikasi *Diabetic Retinopathy* pada Citra Fundus

Fathia Darin Aqliya[1*], Teguh Iman Hermanto[2], Syariful Alam[3]
[1,2,3] Program Studi Teknik Informatika, STT Wastukancana Purwakarta
[1]fathiadarin45@wastukancana.ac.id, [2]teguhiman@wastukancana.ac.id, [3]syarifulalam@wastukancana.ac.id
*Penulis Korespondensi
(Fathia Darin Aqliya, STT Wastukancana Purwakarta, [1*]fathiadarin45@wastukancana.ac.id)

ABSTRAK

Kebutaan akibat diabetes melitus sering kali dipicu oleh komplikasi *Diabetic Retinopathy* (DR). Pemeriksaan mata secara konvensional kerap terkendala oleh keterbatasan jumlah tenaga spesialis dan waktu proses yang panjang. Oleh karena itu, penelitian ini dilakukan untuk mengoptimalkan kinerja *Convolutional Neural Network* (CNN) berarsitektur DenseNet-121 dalam mengklasifikasikan lima level keparahan DR. Metodologi penelitian mengacu pada kerangka kerja CRISP-DM, memanfaatkan 2.930 citra fundus retina dari dataset APTOS 2019. Pembagian data diproses menggunakan teknik *stratified split* dengan rasio 80% pelatihan, 10% validasi, dan 10% pengujian. Tahap optimasi mencakup modifikasi arsitektur akhir melalui implementasi *Sequential Deep Classifier* dan penambahan *Dropout* 30% guna mencegah *overfitting*. Selain itu, tantangan ketidakseimbangan data (*class imbalance*) ditangani secara efektif dengan mengaplikasikan metode *Focal Loss*serta *Class Weights*. Evaluasi sistem melibatkan delapan skenario pengujian dengan ragam variasi *batch size* (16, 32) dan *epoch* (50, 100) untuk membandingkan versi *Baseline* dengan model optimasi, sekaligus memanfaatkan teknik Grad-CAM sebagai visualisasi area keputusan. Hasil eksperimen mencatatkan bahwa model yang dioptimasi mengalami peningkatan generalisasi yang signifikan, di mana tingkat akurasi tertinggi mencapai 87% pada parameter *batch size* 32 dan 100 *epoch*. Penerapan *Focal Loss*juga terbukti ampuh mendongkrak metrik *recall* pada kategori kelas minoritas. Model terbaik kemudian diintegrasikan ke dalam aplikasi web interaktif berbasis Shiny. Kesimpulannya, pengembangan model klasifikasi ini berpotensi besar menjadi alat bantu diagnostik yang efektif dalam menunjang efisiensi skrining klinis.

Kata kunci: *Diabetic Retinopathy*, *CNN, DenseNet-121, Focal Loss, Class Imbalance*

ABSTRACT

Blindness caused by diabetes mellitus is often triggered by Diabetic Retinopathy (DR) complications. Conventional eye examinations are frequently constrained by the limited number of specialists and prolonged processing times. Therefore, this study was conducted to optimize the performance of a Convolutional Neural Network (CNN) using the DenseNet-121 architecture to classify five severity levels of DR. The research methodology follows the CRISP-DM framework, utilizing 2,930 retinal fundus images from the APTOS 2019 dataset. Data division was processed using a stratified split technique with a ratio of 80% training, 10% validation, and 10% testing. The optimization phase included modifying the final architecture through the implementation of a Sequential Deep Classifier and the addition of a 30% Dropout to prevent overfitting. Furthermore, the challenge of class imbalance was effectively handled by applying Focal Lossand Class Weights methods. System evaluation involved eight testing scenarios with variations in batch size (16, 32) and epochs (50, 100) to compare the Baseline version with the optimized model, while simultaneously utilizing the Grad-CAM technique to visualize decision areas. Experimental results demonstrated that the optimized model experienced a significant improvement in generalization, where the highest accuracy reached 87% at a batch size of 32 and 100 epochs. The application of Focal Lossalso proved effective in boosting the recall metric for minority class categories. The best model was then integrated into a Shiny-based interactive web application. In conclusion, the development of this classification model has great potential to become an effective diagnostic tool in supporting the efficiency of clinical screening.

Keywords*: Diabetic Retinopathy, CNN, DenseNet-121, Focal Loss, Class Imbalance*

1. PENDAHULUAN

Diabetes melitus merupakan salah satu tantangan kesehatan global abad ke-21 yang terus mengalami tren peningkatan secara drastis. Berdasarkan proyeksi *International Diabetes Federation* (IDF), jumlah orang dewasa penderita diabetes diperkirakan melonjak dari 589 juta pada tahun 2024 menjadi 853 juta pada tahun 2050 [1]. Di Indonesia, prevalensi penyakit ini juga meningkat signifikan, dengan jumlah penderita diperkirakan telah melampaui 16 juta jiwa berdasarkan data Riset Kesehatan Dasar (Riskesdas) [2]. Peningkatan kasus ini berbanding lurus dengan tingginya risiko komplikasi kronis, salah satunya adalah *Diabetic Retinopathy* (DR). Kondisi yang

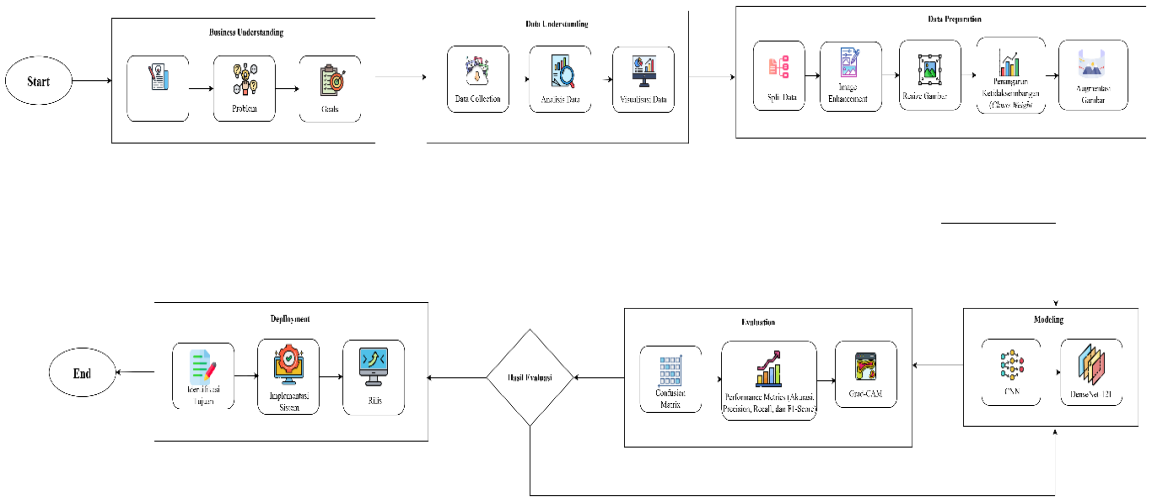
dipicu oleh kerusakan pembuluh darah kapiler di retina ini dialami oleh sekitar 30% penderita diabetes dan menjadi penyebab utama kebutaan apabila tidak terdeteksi serta ditangani sejak dini [1], [3].

Deteksi DR umumnya memerlukan pemeriksaan fundus retina secara langsung oleh dokter spesialis mata. Namun, identifikasi secara visual dan manual memiliki berbagai kendala, mulai dari proses yang memakan waktu lama, tingginya unsur subjektivitas, hingga terbatasnya jumlah tenaga spesialis ahli di berbagai daerah. Oleh karena itu, pengembangan sistem diagnosis berbantuan komputer menggunakan pendekatan *Deep Learning* menjadi solusi krusial. *Convolutional Neural Network* (CNN) telah terbukti andal dan bersifat *data-driven* dalam mengekstraksi fitur citra medis secara otomatis untuk kebutuhan klasifikasi, termasuk pada citra fundus [4], [5]

Sejumlah studi pustaka terdahulu telah mengeksplorasi efektivitas CNN dalam mendeteksi DR. Penggunaan model pre-trained seperti NASNet-Large, ResNet-50, hingga EfficientNetB0 dengan teknik transfer learning terbukti mampu memberikan performa tinggi dan efisiensi komputasi dalam mengenali pola lesi retina [6],[7]. Arsitektur DenseNet-121 juga secara spesifik diakui keandalannya dan mampu mencapai akurasi hingga 98,36% dalam mengklasifikasikan stadium DR pada dataset publik [8], serta efektif digunakan sebagai *backbone* pada klasifikasi dua tahap [[9]. Untuk mencapai tingkat akurasi tersebut, berbagai inovasi arsitektur terus dikembangkan, mulai dari model *Dual Branch* [10], *Region-based* CNN (R-CNN) untuk lokalisasi lesi [11], model *classifier* terpisah untuk deteksi dini [12], hingga pendekatan hibrida HOG-CNN yang menggabungkan fitur tekstur lokal dengan fitur semantik [13]. Kendati demikian, tinjauan literatur sistematis menyoroti bahwa tantangan utama dalam riset DR adalah kualitas citra yang bervariasi serta fenomena ketidakseimbangan kelas (*class imbalance*), yang dapat menyebabkan model bias terhadap kelas mayoritas [50]. Pendekatan seperti fusi fitur dan perbaikan kualitas citra multi-kanal terbukti dapat meningkatkan *visibilitas* lesi untuk mengatasi kendala ini [14].

Di samping masalah ketidakseimbangan data, penerapan model CNN yang memiliki jutaan parameter sering kali berhadapan dengan masalah *overfitting*, di mana model gagal menggeneralisasi data baru dengan baik [15]. Selain itu, sifat *black-box* dari *deep learning* menjadi hambatan dalam implementasi medis karena kurangnya transparansi keputusan, sehingga memunculkan kebutuhan akan metode seperti *Gradient-weighted Class Activation Mapping* (Grad-CAM) untuk memvisualisasikan area fokus model [9].

Berdasarkan identifikasi persoalan tersebut, kebaruan (*novelty*) dari penelitian ini terletak pada pendekatan optimasi komprehensif terhadap model dasar DenseNet-121 yang terintegrasi dengan metode interpretabilitas keputusan. Penelitian ini secara khusus menspesifikasi modifikasi arsitektur akhir menggunakan *Sequential Deep Classifier* dan *Dropout* 30% untuk menekan laju *overfitting*. Secara bersamaan, kelemahan terkait *class imbalance* pada dataset APTOS 2019 diatasi secara langsung pada tahap pelatihan menggunakan kombinasi fungsi kerugian *Focal Loss*dan *Class Weights*. Tujuan dari penelitian ini adalah mengimplementasikan serta mengoptimasi model CNN berbasis DenseNet-121 untuk klasifikasi lima tingkat keparahan DR, sekaligus meningkatkan interpretabilitas model melalui metode Grad-CAM. Hasil akhir dari optimasi ini kemudian diterapkan ke dalam aplikasi berbasis web yang interaktif, sehingga diharapkan dapat memberikan dukungan *skrining* klinis yang efisien, transparan, dan dapat dipercaya oleh tenaga medis.



Gambar 1. Diagram Alur Penelitian Berbasis Kerangka Kerja CRISP-DM

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan kerangka kerja *Cross Industry Standard Process for Data Mining* (CRISP-DM). Kerangka ini digunakan untuk menggambarkan tahapan penelitian secara sistematis, mulai dari pemahaman permasalahan hingga tahap implementasi sistem, seperti yang ditunjukkan pada Gambar 1.

1. *Business Understanding*

Tahap ini guna memahami konteks diagnosis *Diabetic Retinopathy* (DR). Masalah utama yang diidentifikasi meliputi rendahnya kualitas citra retina, ketidakseimbangan data antar stadium penyakit, dan kurangnya interpretasi visual pada model *Deep Learning* (*black-box*). Oleh karena itu, penelitian ini bertujuan mengoptimasi kualitas citra dan membangun model klasifikasi lima stadium DR yang transparan menggunakan visualisasi Grad-CAM.

2. *Data Understanding*

Dataset yang digunakan adalah subset dari dataset "APTOS-2019" dari Kaggle (*Aravind Eye Hospital*, India) yang terdiri dari 2.930 citra fundus retina berlabel stadium DR berdasarkan standar medis ICDRSS. Analisis awal menunjukkan adanya variasi pencahayaan dan ketidakseimbangan kelas (*class imbalance*) yang sangat ekstrem, di mana sampel citra mata normal jauh lebih mendominasi dibandingkan stadium berat.

3. *Data Preparation*

Untuk mengatasi permasalahan tersebut, dilakukan tahap *Data Preparation*. Data dibagi menjadi data latih, data validasi, dan data uji menggunakan teknik *stratified split* untuk menjaga proporsi kelas. Prapemrosesan citra (*image enhancement*) dilakukan menggunakan kombinasi metode *Green Channel Extraction* (GCE) dan *Contrast Limited Adaptive Histogram Equalization* (CLAHE) guna mempertegas fitur klinis seperti lesi dan pembuluh darah. Selanjutnya, citra diubah ukurannya (*resize*) menjadi 224x224 piksel. Untuk menangani ketidakseimbangan data, diterapkan teknik pembobotan kelas (*Class Weight*) (W_j) secara otomatis:

$$W_j = \frac{N}{K \times N_j}$$

di mana $N = 2930$ total sampel, $K = 5$ jumlah kelas, dan N_j jumlah sampel pada kelas tertentu), serta augmentasi citra sintetik (rotasi, *flipping*, penyesuaian kecerahan, dan zoom) guna meningkatkan kemampuan generalisasi model.

4. *Modeling*

Tahap inti membangun dan melatih algoritma CNN dengan arsitektur DenseNet-121. Untuk memperkuat kekokohan (*robustness*) model secara ilmiah, lapisan pengklasifikasi bawaan dimodifikasi menjadi *Sequential Deep Classifier*. Modifikasi ini bekerja secara matematis melalui:

- Transformasi Dimensi Vektor **(**$1024 \rightarrow 512 \rightarrow 5$**),** mengompresi 1024 fitur spasial DenseNet menjadi 512 fitur perantara sebelum diproyeksikan ke 5 kelas stadium DR.
- Aktivasi ReLU **(**$\mathrm{ReLU}(x) = \max(0, x)$**)**, memungkinkan model mempelajari pola lesi medis yang kompleks.
- Regularisasi Dropout **(**$p = 0{,}4$**):** Menonaktifkan 40% neuron secara acak untuk mencegah model menghafal data (*overfitting*).

Pelatihan menggunakan fungsi Focal Loss dengan parameter gamma sebesar 2,0, parameter alpha yang berasal dari pembobotan kelas (*class weights*), *Optimizer Adam* (*learning rate* 0.0001), dan fungsi aktivasi ReLU. Parameter gamma pada *Focal Loss* berfungsi mengatur tingkat penekanan terhadap sampel yang mudah diklasifikasikan, sehingga model memberikan perhatian relatif lebih besar pada sampel yang sulit diklasifikasikan. Penggunaan fungsi ini bertujuan membantu proses pembelajaran model pada dataset dengan distribusi kelas yang tidak seimbang.

Untuk menguji performa secara komprehensif, dirancang mekanisme pelatihan otomatis yang mengeksekusi delapan skenario eksperimen dengan memadukan parameter ukuran tumpukan (*batch size* 16 dan 32) serta jumlah siklus (*epochs* 50 dan 100) pada jalur arsitektur *Baseline* dan optimasi. Eksperimen ini bertujuan menemukan kombinasi *hyperparameter* yang paling optimal sekaligus membandingkan efektivitas peningkatan performa model yang telah dioptimasi terhadap model *Baseline*.

5. *Evaluation*

Setelah model DenseNet-121 berhasil dilatih, dilakukan tahap evaluasi untuk mengukur sejauh mana kinerja dan kemampuan model. Proses ini diawali dengan analisis *Learning Curve* melalui plot grafik akurasi dan *loss* data validasi guna memantau stabilitas konvergensi model serta mendeteksi ada tidaknya gejala *overfitting*. Selanjutnya, performa model diuji secara kuantitatif menggunakan *Confusion Matrix* untuk memetakan distribusi tingkat ketepatan prediksi pada setiap kelas stadium. Tahap berikutnya adalah melakukan perhitungan *Performance Metrics* yang meliputi akurasi, *precision*, *recall*, dan F1-*score*. Sebagai tahap akhir, evaluasi

dilakukan secara kualitatif menggunakan teknik Grad-CAM (*Gradient-weighted Class Activation Mapping*). Teknik ini menghasilkan representasi visual berupa *heatmap* yang menunjukkan area fokus spasial model pada citra retina, sehingga dapat divalidasi secara medis apakah keputusan model didasarkan pada fitur klinis yang relevan. Apabila hasil evaluasi belum mencapai indikator performa yang optimal, proses akan beriterasi kembali ke tahap *Modeling* untuk penyesuaian skenario eksperimen.

6. *Deployment*

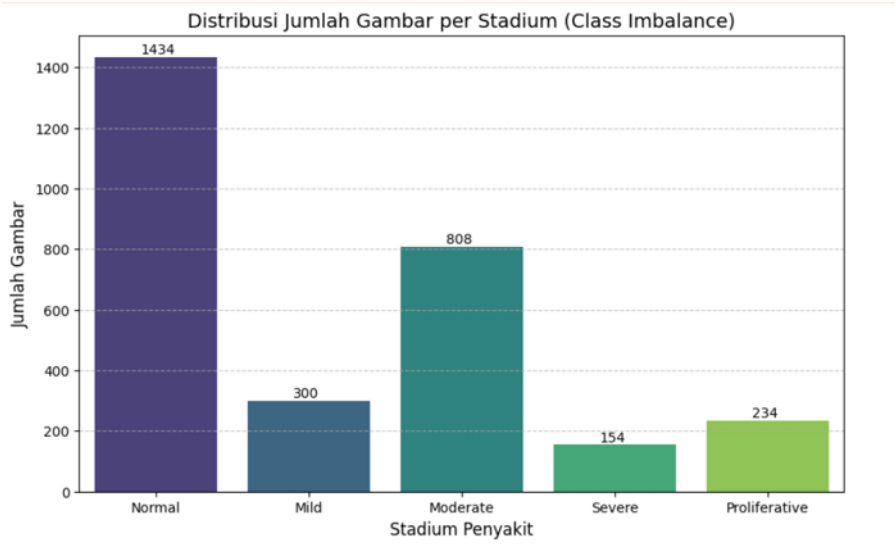
Tahap ini difokuskan pada perancangan antarmuka berbasis web berupa *dashboard* interaktif menggunakan *framework* Shiny for Python. Mengingat penelitian ini berfokus pada optimasi arsitektur model, sistem tidak dirancang untuk menerima input citra baru guna prediksi medis secara real-time. Sebaliknya, dashboard ini difungsikan secara khusus untuk mendokumentasikan dan memvisualisasikan seluruh alur eksperimen secara terpusat.

3 HASIL DAN PEMBAHASAN

Pada bagian ini, disajikan hasil implementasi sistem dan evaluasi kinerja model arsitektur DenseNet-121 yang diuji melalui 8 skenario eksperimen. Pengujian difokuskan pada analisis perbandingan performa antara model *Baseline* dan model Optimasi.

3.1 *Data Understanding*

Penelitian ini menggunakan subset dari dataset publik "APTOS-2019" yang diperoleh dari Kaggle, berisi 2.930 citra fundus retina. Analisis awal memastikan tidak terdapat *missing value* pada dataset, namun ditemukan masalah ketidakseimbangan kelas (*class imbalance*) yang sangat ekstrem, di mana citra mata normal mendominasi secara signifikan dibandingkan stadium kritis



Gambar 2. Visualisasi Distribusi Kelas Subset Dataset APTOS-2019

Tabel 1. Rincian Kelas Dataset APTOS-2019

No.	Tingkat Keparahan (label)	Jumlah Citra	Keterangan Visual
1.	Normal (0)	1.434	Retina sehat, tidak ada kelainan.
2.	Mild (1)	300	Mulai muncul *microaneurysms* (titik merah).
3.	Moderate (2)	808	Terdapat bercak kuning (*Exudates).*
4.	Servere (3)	154	Pendarahan meluas, eksudar terlihat jelas.
5.	Proliferative (4)	234	Pendarahan hebat, pertumbuhan pembuluh abnormal.

3.2 *Data Preparation*

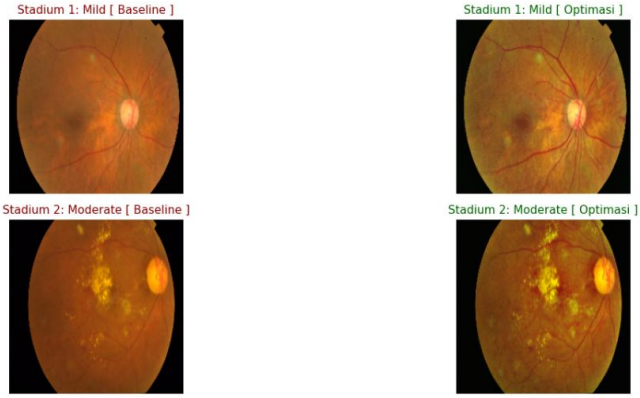
Untuk mempersiapkan data sebelum tahap pelatihan, dilakukan serangkaian proses prapemrosesan guna meningkatkan kualitas citra dan menangani masalah *class imbalance*.

3.2.1 *Split Data*

Data dibagi menggunakan teknik *stratified split* untuk menjaga proporsi jumlah kelas. Komposisi akhir dataset terbagi menjadi 80% data latih (2.344 citra), 10% data validasi (293 citra), dan 10% data uji (293 citra).

3.2.2 Peningkatan Kualitas Citra (*Image Enhancement*) dan *Resize*

Untuk memperjelas fitur klinis retina yang sering kali tersamar akibat pencahayaan tidak merata, diterapkan kombinasi metode *Green Channel Extraction* (GCE) dan *Contrast Limited Adaptive Histogram Equalization* (CLAHE). GCE bertugas mengisolasi saluran warna hijau yang memiliki kontras tertinggi terhadap lesi patologis, sedangkan CLAHE mempertajam detail kontras secara lokal. Setelah kualitas visual ditingkatkan, seluruh dimensi citra diseragamkan (di-*resize*) menjadi 224x224 piksel. Standarisasi dimensi ini merupakan syarat mutlak secara komputasi untuk memenuhi spesifikasi ukuran masukan (*input shape*) yang diwajibkan oleh arsitektur DenseNet-121.



Gambar 3. Perbandingan Citra Asli dan Hasil Optimasi GCE+CLAHE

3.2.3 Penanganan Ketidakseimbangan Data (*Class Weight*) dan Augmentasi

Penanganan *imbalance* dilakukan melalui perhitungan *Class Weight* untuk memberikan nilai bobot penalti yang lebih besar pada kelas minoritas (seperti *Severe* dan *Proliferative*) saat pelatihan model. Selanjutnya, diterapkan teknik augmentasi citra khusus pada data latih meliputi *horizontal flip*, rotasi acak hingga 15 derajat, dan variasi kecerahan/kontras (*color jitter*) sebesar 10% guna menambah variasi data buatan dan mencegah *overfitting*.

3.3 Pemodelan (*Modeling*)

Tahap pemodelan membandingkan dua pendekatan komputasi berbasis arsitektur DenseNet-121. Pelatihan model pada masing-masing pendekatan dieksekusi secara berurutan melalui penyesuaian *hyperparameter* yang terbagi ke dalam 4 skenario konfigurasi (sehingga menghasilkan total 8 skenario pengujian secara keseluruhan). Rincian konfigurasi untuk setiap skenario adalah sebagai berikut:

1. Skenario 1: Menggunakan *batch size* 16 dengan durasi pelatihan sebanyak 50 *epoch*.
2. Skenario 2: Menggunakan *batch size* 32 dengan durasi pelatihan sebanyak 50 *epoch*.
3. Skenario 3: Menggunakan *batch size* 16 dengan durasi pelatihan sebanyak 100 *epoch*.
4. Skenario 4: Menggunakan *batch size* 32 dengan durasi pelatihan sebanyak 100 *epoch*.

Untuk mengevaluasi efektivitas arsitektur pada keempat skenario konfigurasi di atas, digunakan dua pendekatan jalur pemodelan, yaitu:

a. Jalur *Baseline* : Menggunakan Arsitektur dasar dengan lapisan klasifikasi *Linear* Tunggal dan *Cross-Entropy Loss* standar.
b. Jalur Optimasi : Mengganti lapisan akhir dengan *Sequential Deep Classifier* (penambahan *hidden layer* 512 *neuron*, aktivasi ReLU, dan *Dropout* 30%) serta mengintegrasikan *Focal Loss*berbasis *Class Weight.*

3.4 Evaluasi Performa Model

Evaluasi dilakukan untuk membedah efektivitas Skenario *Baseline* dibandingkan dengan Skenario Optimasi. Fokus analisis dititikberatkan pada Skenario 4 yang memberikan performa paling optimal.

3.4.1 Rekapitulasi Metrik Evaluasi

Hasil pengujian performa secara kuantitatif dirangkum pada Tabel 2 untuk menentukan konfigurasi model paling optimal.

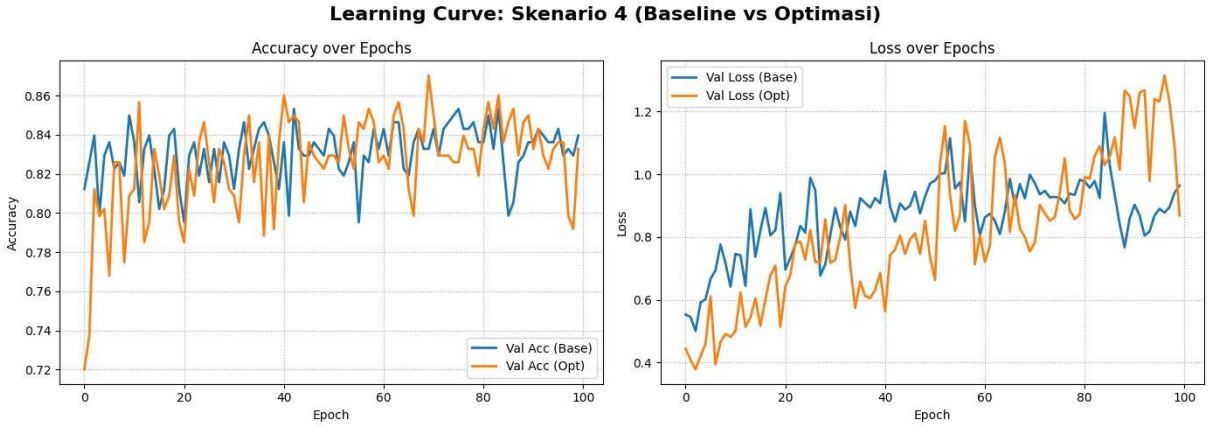
Tabel 2. Rekapitulasi Metrik Evaluasi Model

	Skenario	Varian Model	*Acuracy*	*Precision (Macro)*	*Recall (Macro)*	*F1-score (Macro)*
1	Skenario 1	*Baseline*	0.8532	0.7534	0.7061	0.7261
2	Skenario 1	Optimasi	0.8737	0.8352	0.7449	0.7786
3	Skenario 2	*Baseline*	0.8362	0.7672	0.6691	0.7014
4	Skenario 2	Optimasi	0.8635	0.8037	0.7385	0.7637
5	Skenario 3	*Baseline*	0.8532	0.7631	0.7141	0.7331
6	Skenario 3	Optimasi	0.8635	0.7730	0.7311	0.7352
7	Skenario 4	*Baseline*	0.8532	0.7589	0.6696	0.6904
8	**Skenario 4**	**Optimasi**	**0.8703**	**0.8008**	**0.7658**	**0.7752**

Tabel 2 menegaskan bahwa Skenario 4 dengan Varian Optimasi (*Batch size* 32, durasi 100 *Epochs*) adalah konfigurasi paling stabil dan unggul dengan tingkat Akurasi 87.03%. Oleh karena itu, analisis mendalam secara visual difokuskan pada Skenario 4.

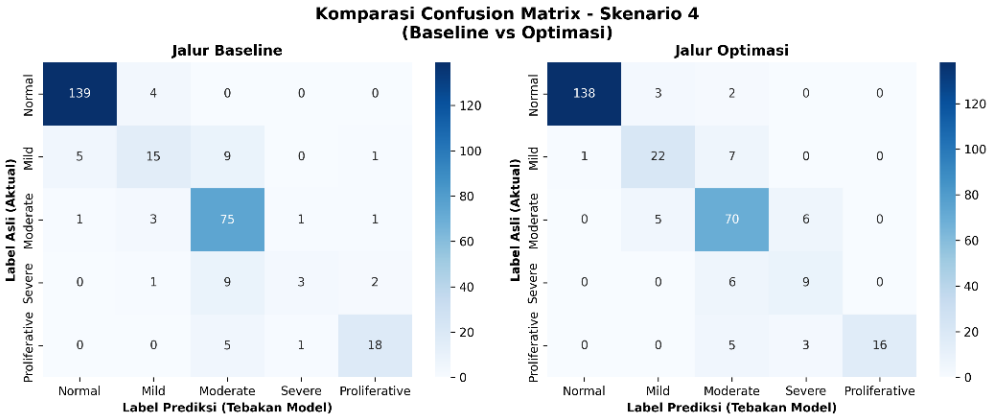
3.4.2 Analisis *Learning Curve* dan *Confusion Matrix*

Dinamika proses pelatihan pada Skenario 4 dapat dilihat melalui grafik *Learning Curve* berikut.



Gambar 4. Grafik *Learning Curve* Skenario 4 (Akurasi dan Loss)

Grafik pada Gambar 4 mengilustrasikan perbandingan fluktuasi akurasi dan *loss* validasi selama 100 *epoch*. Meskipun grafik *loss* (kanan) menunjukkan adanya tren kenaikan pada fase akhir pelatihan yang mengindikasikan gejala *overfitting* pada kedua model, model Optimasi (garis oranye) terbukti mampu mencapai titik puncak akurasi validasi (kiri) yang lebih tinggi dibandingkan model *Baseline*. Tren kenaikan *validation loss* ini terjadi karena akumulasi penalti besar dari fungsi *Focal Loss* terhadap kesalahan prediksi sampel minoritas pada *epoch* tinggi, serta penggunaan *learning rate* konstan yang membuat model sulit konvergen secara stabil di akhir pelatihan. Guna mengatasi kendala *overfitting* ini di masa mendatang, strategi penanganan yang dapat diterapkan meliputi penggunaan *Early Stopping* untuk menghentikan pelatihan secara otomatis saat *validation loss* mulai naik, penerapan *Learning Rate Scheduler* (seperti *ReduceLROnPlateau*) untuk menurunkan *learning rate* secara bertahap, serta penambahan regularisasi *Weight Decay*. Selanjutnya, pembedahan kemampuan model dalam mengklasifikasikan kelas minoritas dianalisis secara mendalam menggunakan *Confusion Matrix*.



Gambar 5. *Confusion Matrix* Skenario 4

Berdasarkan Gambar 5, terbukti bahwa model *Baseline* mengalami bias terhadap kelas mayoritas; misalnya pada kelas *Severe*, model *Baseline* hanya mampu menebak 3 citra dengan benar dan salah memprediksi 9 citra ke dalam kelas *Moderate*.

Sebaliknya, penerapan *Focal Loss*berbasis *Class Weight* pada model Optimasi berhasil mendistribusikan tebakan secara lebih merata. Hal ini dibuktikan dengan lonjakan tebakan benar pada kelas minoritas, di mana *True Positives* (TP) untuk kelas *Mild* meningkat dari 15 menjadi 22 citra, dan kelas *Severe* meningkat signifikan dari 3 menjadi 9 citra. Hasil ini mengonfirmasi bahwa jalur optimasi jauh lebih andal untuk mendeteksi stadium kritis.

3.4.3 Perbandingan dengan Penelitian Terdahulu

Untuk mengetahui posisi performa model yang dikembangkan dibandingkan dengan penelitian terdahulu, dilakukan perbandingan berdasarkan nilai akurasi dan *F1-score* pada dataset APTOS 2019. Perbandingan ini bertujuan untuk memberikan gambaran mengenai kinerja model yang diusulkan terhadap model-model yang telah digunakan pada penelitian sebelumnya. Hasil perbandingan tersebut disajikan pada Tabel 3.

Table 3. Perbandingan dengan Penelitian Terdahulu

No	Peneliti / Sumber (Tahun)	Arsitektur / Metode	Dataset	Akurasi (%)	*F1-score* (%)
1.	**Applied Sciences (2024)** [16] *"An Automated Deep Learning Model to Categorize Stages of Diabetic Retinopathy"*	CNN + CLAHE + DWT (Adam Optimizer)	APTOS 2019	71,85%	71,00%
2.	**EEE Access (2023)** [17] *"A Federated Improved DenseNet for Classification of Diabetic Retinopathy"*	Federated Improved DenseNet (FedDDR)	APTOS 2019	94,27%	90,00%
3.	**C. B. Chandrakala et al. (2026)** [18] *"A Hybrid Deep Learning Approach With Explainable AI for Diabetic Retinopathy Classification"*	Hibrida (EfficientNetB0 + DenseNet121)+ SMOTE	APTOS 2019	84,02%	82,91%
4.	**IEEE Access (2022)** [19] *"Automatic Severity Classification of Diabetic Retinopathy Based on DenseNet and Convolutional Block Attention Module"*	DenseNet169 + CBAM + Weighted Loss	APTOS 2019	82,00%	68,00%
5.	**Peneltian Saat ini (2026) Optimasi Model CNN untuk Klasifikasi *Diabetic Retinopathy* pada Citra Fundus**	**DenseNet-121 + GCE + CLAHE + Deep Classifier + Focal Loss + Class Weights**	**APTOS 2019**	**87,03%**	**77,52%**

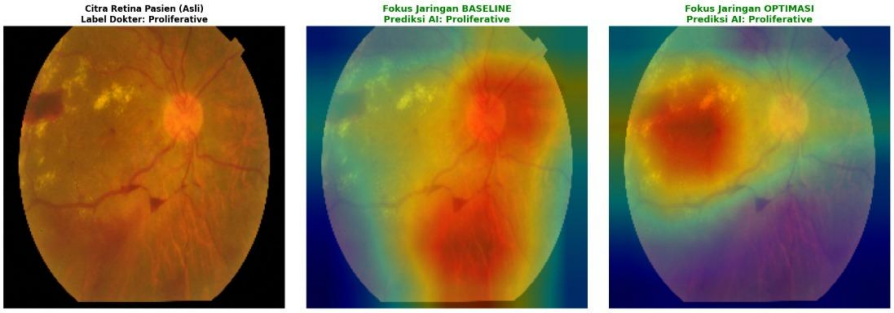
Berdasarkan Tabel 3, penelitian terdahulu menunjukkan variasi kinerja dalam klasifikasi retinopati diabetik. Penelitian Applied Sciences (2024) menggunakan CNN, CLAHE, dan DWT dengan akurasi 71,85% dan *F1-score* 71,00%. Penelitian FedDDR pada IEEE Access (2023) memperoleh akurasi 94,27% dan *F1-score*

90,00%, sedangkan Chandrakala et al. (2026) menggunakan model hibrida EfficientNetB0 dan DenseNet121 dengan SMOTE, menghasilkan akurasi 84,02% dan *F1-score* 82,91%. Sementara itu, Farag et al. (2022) menerapkan DenseNet169 dengan CBAM dan *Weighted Loss*, dengan akurasi 82,00% dan *F1-score* 68,00%.

Penelitian ini mengusulkan DenseNet-121 yang dikombinasikan dengan *Green Channel Extraction* (GCE), CLAHE, *Deep Classifier*, *Focal Loss*, dan *Class Weights*, serta memperoleh akurasi 87,03% dan *F1-score* 77,52%. Hasil tersebut menunjukkan bahwa metode yang diusulkan memiliki akurasi lebih tinggi dibandingkan penelitian nomor 1, 3, dan 4, serta *F1-score* lebih tinggi dibandingkan penelitian nomor 1 dan 4. Meskipun belum melampaui FedDDR dan metode hibrida Chandrakala et al. dalam kedua metrik, hasil ini menunjukkan potensi kombinasi *preprocessing* dan penanganan ketidakseimbangan kelas dalam meningkatkan kinerja klasifikasi retinopati diabetik.

3.4.4 Visualisasi Keputusan Model (Grad-CAM)

Evaluasi metrik secara kuantitatif saja belum cukup untuk membuktikan keandalan model kecerdasan buatan, terutama dalam ranah diagnosis medis. Oleh karena itu, pengujian kualitatif menggunakan algoritma Grad-CAM (*Gradient-weighted Class Activation Mapping*) diterapkan guna menyingkap proses pengambilan keputusan (*black-box*) dari arsitektur DenseNet-121. Grad-CAM memproyeksikan sebuah peta panas (*heatmap*) yang menyoroti area piksel paling berpengaruh dalam menentukan kelas prediksi. Hal ini sangat krusial agar hasil diagnosis model dapat dipertanggungjawabkan secara klinis, yakni dengan memastikan bahwa model benar-benar mendeteksi tanda patologis nyata seperti eksudat atau pendarahan.



Gambar 6. Visualisasi Grad-CAM

Berdasarkan Gambar 6, hasil visualisasi Grad-CAM membuktikan adanya perbedaan pola atensi yang sangat signifikan antara model *Baseline* dan model Optimasi dalam mengambil keputusan.

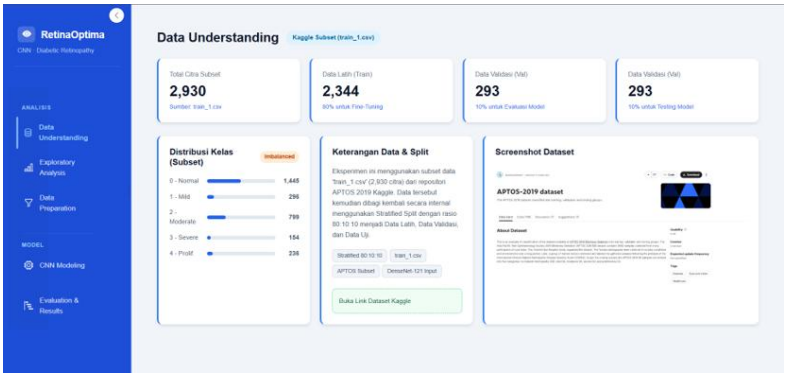
Pada kolom tengah (Grad-CAM *Baseline*), terlihat bahwa peta panas (*heatmap*) cenderung menyebar secara tidak beraturan dan tidak fokus. Model *Baseline* sering kali memberikan bobot atensi yang tinggi (warna merah/kuning) pada area yang tidak memiliki nilai diagnosis klinis, seperti pada area tepi luar retina atau bahkan melebar hingga ke area latar belakang (*background*) yang gelap. Hal ini menunjukkan bahwa tebakan model *Baseline* rentan dipengaruhi oleh piksel yang tidak relevan.

Sebaliknya, pada kolom kanan (Grad-CAM *Optimization*), area fokus komputasi terlihat sangat presisi dan terlokalisasi dengan baik. Model Optimasi terbukti mampu memusatkan perhatiannya secara spesifik hanya pada area yang mengandung *lesi patologis*, seperti bercak eksudat, *mikroaneurisma*, maupun area pendarahan yang terlihat pada citra asli.

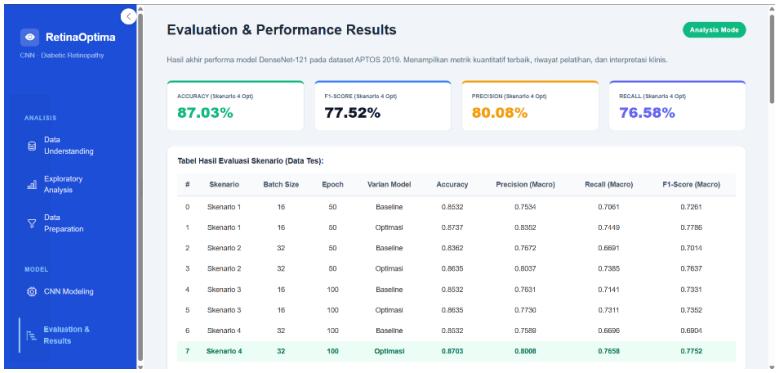
Hasil visualisasi ini secara meyakinkan mengonfirmasi bahwa integrasi metode prapemrosesan (GCE dan CLAHE) serta penerapan *Focal Loss*tidak hanya berhasil meningkatkan nilai metrik akurasi, tetapi juga sukses membimbing model untuk "melihat" dan mendiagnosis penyakit *Diabetic Retinopathy* selayaknya standar observasi medis, bukan sekadar mengenali pola gambar secara acak.

3.5 *Deployment* Sitem

Sebagai fase akhir dari siklus pengembangan model, dilakukan integrasi hasil eksperimen ke dalam sebuah antarmuka aplikasi berbasis web yang diberi nama "RetinaOptima". Sistem ini dibangun menggunakan *framework* Shiny for Python, sebuah pustaka yang memungkinkan penyusunan *dashboard* interaktif dengan logika pemrosesan yang efisien.



*G*ambar 7 Antarmuka Sistem RetinaOptima halaman *Data Understanding*



Gambar 8 Antarmuka Sistem RetinaOptima halaman Evaluasi

Implementasi sistem pada *deployment* ini tidak ditujukan untuk melakukan prediksi citra medis baru secara real-time di lingkungan klinis, melainkan berfungsi sebagai *instrumen analitik* untuk meninjau secara transparan seluruh alur penelitian yang telah dilakukan. Melalui *dashboard* ini, pengguna dapat menelusuri secara interaktif tahapan Data Understanding (seperti distribusi kelas dan statistik *dataset*), memantau perbandingan performa delapan skenario eksperimen pada bagian CNN Modeling, serta melihat hasil evaluasi metrik akhir secara komprehensif. Dengan demikian, sistem ini berperan sebagai bentuk dokumentasi digital yang intuitif dan akuntabel atas validitas hasil penelitian terhadap dataset APTOS-2019.

4 KESIMPULAN

Penelitian ini telah berhasil mengembangkan dan mengoptimasi model *deep learning* berbasis arsitektur DenseNet-121 untuk mengklasifikasikan lima tingkat keparahan penyakit *Diabetic Retinopathy*. Penerapan prapemrosesan *Green Channel Extraction* (GCE) dan CLAHE digunakan untuk meningkatkan kualitas citra fundus, sedangkan modifikasi arsitektur menggunakan *Sequential Deep Classifier* yang dilengkapi *Dropout*, serta penerapan *Focal Loss* dan *Class Weights*, ditujukan untuk membantu menangani ketidakseimbangan kelas pada dataset APTOS 2019. Pengujian pada skenario optimal dengan *batch size* 32 dan 100 *epoch* menghasilkan akurasi sebesar 87,03% atau 0,8703 dan *F1-score* sebesar 77,52% atau 0,7752. Interpretabilitas model juga dievaluasi melalui visualisasi Grad-CAM untuk mengamati area citra fundus yang berkontribusi terhadap keputusan klasifikasi.

Penelitian ini memiliki keterbatasan, antara lain penggunaan satu dataset publik dengan variasi citra dan kondisi pengambilan gambar yang terbatas, serta belum dilakukannya pengujian secara *real-time* di fasilitas kesehatan. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan dataset multisenter yang berasal dari berbagai fasilitas kesehatan dan perangkat kamera fundus guna mengevaluasi kemampuan generalisasi model pada kondisi yang lebih beragam. Pengujian *cross-dataset* juga perlu dilakukan dengan melatih model pada APTOS 2019 dan mengujinya pada dataset lain untuk menilai ketahanan model terhadap perbedaan karakteristik citra. Selain itu, penerapan *Early Stopping* berdasarkan nilai *validation loss* dapat dipertimbangkan untuk menghentikan pelatihan ketika performa validasi tidak lagi membaik, sehingga membantu mengurangi risiko *overfitting* dan ketidakstabilan pelatihan. Penelitian selanjutnya juga diharapkan melakukan validasi pada lingkungan klinis dengan melibatkan tenaga medis yang kompeten sebelum model dipertimbangkan untuk mendukung proses pengambilan keputusan klinis.

DAFTAR PUSTAKA

[1] International Diabetes Federation, IDF Diabetes Atlas, 11th ed. Brussels, Belgium: International Diabetes Federation, 2025. [Online]. Available: https://diabetesatlas.org/resources/idf-diabetes-atlas-2025/

[2] V. Ardelia *et al.*, "Perbandingan Algoritma Canny Edge Detection Dan Prewitt Pada Deteksi Stadium Diabetik Retinopati," Jurnal Ilmiah Informatika, vol. 9, no. 2, pp. 87–94, 2021, doi: 10.33884/jif.v9i02.3762.

[3] M. A. Ardyansyah and Gunawansyah, "Sistem Deteksi Level Diabetic Retinopathy Melalui Citra Fundus Mata dengan Menggunakan Metode CNN (Convolutional Neural Network)," *G-Tech: Jurnal Teknologi Terapan*, vol. 7, no. 4, pp. 1673–1682, Oct. 2023, doi: 10.33379/gtech.v7i4.3332.

[4] K. A. H. Wibowo and Alamsyah, "Klasifikasi Diabetic Retinopathy Menggunakan CNN dengan Arsitektur yang Dimodifikasi," Indonesian Journal of Mathematics and Natural Sciences, vol. 46, no. 1, pp. 45–52, 2023, doi: 10.15294/ijmns.v46i1.46172.

[5] F. H. Syahrul and P. S. Sasongko, "Penerapan Convolutional Neural Network untuk Klasifikasi Tingkat Keparahan Retinopati Diabetik pada Penderita Diabetes Melitus," Jurnal Masyarakat Informatika, vol. 13, no. 1, pp. 1–14, 2022, doi: 10.14710/jmasif.13.1.42354.

[6] S. Vallukappully, I. van der Linde, and A. Chakraborty, "Early detection and classification of diabetic retinopathy by transfer learning of NASNet-large and ResNet-50 convolutional neural networks," *Inform. Med. Unlocked*, vol. 58, Jan. 2025, doi: 10.1016/j.imu.2025.101688.

[7] V. Anand, D. Koundal, W. Y. Alghamdi, and B. M. Alsharbi, "Smart grading of diabetic retinopathy: an intelligent recommendation-based fine-tuned EfficientNetB0 framework," *Front. Artif. Intell.*, vol. 7, 2024, doi: 10.3389/frai.2024.1396160.

[8] G. Alwakid, W. Gouda, M. Humayun, and N. Z. Jhanjhi, "Deep learning-enhanced diabetic retinopathy image classification," *Digit. Health*, vol. 9, Jan. 2023, doi: 10.1177/20552076231194942.

[9] A. M. Moustari, Y. Brik, B. Attallah, and R. Bouaouina, "Two-stage deep learning classification for diabetic retinopathy using gradient weighted class activation mapping," *Automatika*, vol. 65, no. 3, pp. 1284–1299, 2024, doi: 10.1080/00051144.2024.2363692.

[10] H. Shakibania, S. Raoufi, B. Pourafkham, H. Khotanlou, and M. Mansoorizadeh, "Dual Branch Deep Learning Network for Detection and Stage Grading of Diabetic Retinopathy," Mar. 2024, doi: 10.1016/j.bspc.2024.106168.

[11] V. S. Sravya, P. N. Srinivasu, J. Shafi, W. Hołubowski, and A. Zielonka, "Advanced Diabetic Retinopathy Detection with the R-CNN: A Unified Visual Health Solution," *International Journal of Applied Mathematics and Computer Science*, vol. 34, no. 4, pp. 661–678, Sep. 2024, doi: 10.61822/amcs-2024-0044.

[12] S. A. Karthik, M. N. Geetha, K. Prabhavathi, D. Shashank, K. P. Suhaas, and M. Narender, "Early Detection and Severity Classification of Diabetic Retinopathy Using Convolutional Neural Networks," *SN Comput. Sci.*, vol. 6, no. 7, Oct. 2025, doi: 10.1007/s42979-025-04361-y.

[13] F. Ahmed, "HOG-CNN: Integrating Histogram of Oriented Gradients with Convolutional Neural Networks for Retinal Image Classification," Jul. 2025, [Online]. Available: http://arxiv.org/abs/2507.22274

[14] B. C. Anupama, S. N. Rao, M. B. Malini, and V. V. Athreya, "Comparative analysis of novel preprocessing techniques and deep learning based multi modal feature fusion for diabetic retinopathy grading," *Sci. Rep.*, vol. 16, no. 1, Dec. 2026, doi: 10.1038/s41598-025-31339-w.

[15] M. I. Zulfa, A. S. Aryanto, B. A. Wijonarko, and W. A. Ahmed, "Addressing Overfitting in Dermatological Image Analysis with Bayesian Convolutional Neural Network," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 10, no. 2, pp. 463–475, Jul. 2024, doi: 10.26555/jiteki.v10i2.29177.

[16] A. M. Mutawa, K. Al-Sabti, S. Raizada, and S. Sruthi, "A Deep Learning Model for Detecting Diabetic Retinopathy Stages with Discrete Wavelet Transform," Applied Sciences, vol. 14, no. 11, Art. no. 4428, 2024, doi: 10.3390/app14114428.

[17] A. Singh and K. K. Singh, "FedDDR: A Federated Improved DenseNet for Classification of Diabetic Retinopathy," 2023. [Online]. Available: https://www.eyeops.com/]

[18] G. P. Reddy, P. D. Javali, R. Ramyashree, S. Raghavendra, C. B. Chandrakala, and P. S. Venugopala, "A Hybrid Deep Learning Approach With Explainable AI for Diabetic Retinopathy Classification," *IEEE Access*, vol. 14, pp. 28819–28839, 2026, doi: 10.1109/ACCESS.2026.3665564.

[19] M. M. Farag, M. Fouad, and A. T. Abdel-Hamid, "Automatic Severity Classification of Diabetic Retinopathy Based on DenseNet and Convolutional Block Attention Module," *IEEE Access*, vol. 10, pp. 38299–38308, 2022, doi: 10.1109/ACCESS.2022.3165193.