

Eye Disease Classification Using EfficientNet-B0 Based on Transfer Learning

I Dewa Ayu Pradnya Pratiwi Tentriajaya^{1*}, I Gusti Ngurah Lanang Wijayakusuma^{2*}

* Matematika, Universitas Udayana

tentriajaya.2208541013@unud.ac.id¹, lanang.wijaya@unud.ac.id²

Article Info

Article history:

Received 2025-06-03

Revised 2025-07-07

Accepted 2025-07-09

Keyword:

CNN,
Classification,
EfficientNet-B0,
Eye Disease,
Fundus Image

ABSTRACT

This study focuses on developing and evaluating a deep learning approach employing EfficientNet-B0 based on transfer learning to classify retinal fundus images into four categories: Cataract, Diabetic Retinopathy, Glaucoma, and Normal. The model was trained using a retinal image dataset and demonstrated stable training performance, indicated by a consistent decrease in both training and validation loss without signs of overfitting. The training accuracy reached 92%, while the validation accuracy ranged between 94–95%. Model performance evaluation using a confusion matrix and classification report showed excellent classification results, particularly for the Diabetic Retinopathy class, with an F1-Score of 0.98. The Cataract and Normal classes also achieved high performance, with F1-Scores of 0.94 and 0.92, respectively. However, classification accuracy slightly declined for the Glaucoma class, which experienced some misclassification with the Normal class. Overall, the model achieved a classification accuracy of 94% on the test dataset, indicating good generalization capability. These findings suggest that the model holds strong potential for implementation in automated medical image-based diagnostic support systems. Nonetheless, performance improvement in classes with relatively higher misclassification rates is still required to ensure model reliability in clinical practice.



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

I. PENDAHULUAN

Mata merupakan organ penting yang membantu manusia melihat dengan menangkap cahaya dan mengubahnya menjadi sinyal yang dipahami oleh otak sebagai gambar. Melalui penglihatan, manusia dapat menjalankan berbagai aktivitas sehari-hari, mulai dari lingkungan sekitar. Namun, fungsi visual yang kompleks ini sangat rentan terhadap gangguan. Isu tersebut menjadi sorotan utama *World Health Organization* (WHO) mengenai tantangan besar dalam meningkatkan kesehatan mata. Tantangan-tantangan ini termasuk jumlah penderita gangguan penglihatan yang besar dan terus bertambah diperkirakan mencapai 1,1 miliar orang pada tahun 2020 [1]. Berdasarkan laporan WHO pada tahun 2012, sebanyak 285 juta orang di seluruh dunia mengalami gangguan penglihatan, dengan rincian 39 juta orang mengalami kebutaan dan 246 juta lainnya mengalami penglihatan rendah (*low vision*) [2]. Kemudian, berdasarkan analisis Global Burden of Disease Study, penyebab utama

kebutaan secara global pada usia 50 tahun ke atas di tahun 2020 meliputi katarak (15,2 juta kasus), degenerasi macula terkait usia atau *age-related macular degeneration* (AMD) (1,8 juta kasus) dan retinopati diabetik (0,86 juta kasus) [3]. Penyakit-penyakit ini tidak hanya menurunkan kualitas hidup penderita tetapi juga dapat menyebabkan kebutaan apabila tidak dideteksi dan ditangani secara dini [4]. Oleh karena itu, adanya deteksi dini menjadi langkah krusial dalam menjaga fungsi penglihatan serta mencegah gangguan mata menjadi kondisi yang lebih parah. Pendekatan deteksi ini sangat penting untuk menghindari kebutaan atau memperlambat kerusakan visual, terutama bagi populasi rentan.

Sistem deteksi penyakit mata telah berevolusi seiring berkembangnya teknologi. Munculnya pendekatan deep learning seperti Convolutional Neural Network (CNN) telah membuka peluang dalam deteksi gambar atau citra khususnya citra medis secara akurat. CNN memiliki beberapa model pre-trained seperti AlexNet, VGG-16, Resnet, DenseNet, MobileNet, EfficientNet dan lain-lain yang digunakan untuk

deteksi ataupun klasifikasi berbagai penyakit khususnya penyakit mata seperti retinopati diabetik, katarak dan glukoma.

Beberapa penelitian sebelumnya telah mengembangkan model CNN untuk klasifikasi penyakit mata dengan berbagai pendekatan dan arsitektur. Pada penelitian tersebut klasifikasi penyakit mata dilakukan dengan model MobileNetV2 dengan modifikasi pada bagian *head* model, hasilnya menunjukkan akurasi sebesar 72% [4], namun pada penelitian ini hanya menggunakan 601 citra fundus retina yang diklasifikasikan menjadi kelas Normal dan Abnormal, sehingga kurang merepresentasikan kompleksitas kondisi penyakit mata dengan model VGG-16 juga pernah dilakukan dengan hasil akurasi sebesar 88% [5], pada penelitian ini klasifikasi dilakukan lebih bervariasi yaitu ke dalam empat kelas (*Cataract*, *Diabetic Retinopathy*, *Glaucoma* dan Normal). Penelitian serupa juga pernah dilakukan dengan menggunakan model Alexnet dengan akurasi sebesar 87% [6]. Selain itu terdapat pula klasifikasi penyakit mata yang menggunakan model YOLO V8 dengan akurasi yang lebih baik yaitu sebesar 92% [7]. Komparasi antar model juga pernah dilakukan yaitu membandingkan model ResNet50, MobileNet dan DenseNet121 untuk mengklasifikasi penyakit mata, hasilnya menunjukkan bahwa model dengan akurasi tertinggi yaitu ResNet sebesar 92,27% [8].

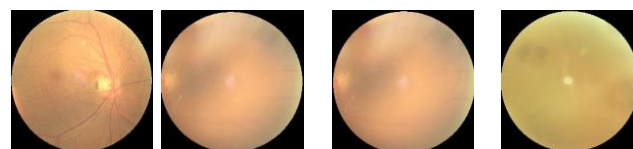
Berdasarkan hasil-hasil tersebut, terlihat bahwa pemilihan arsitektur model berpengaruh terhadap performa klasifikasi penyakit mata, begitu pula dengan ukuran dataset, tetapi tidak menutup kemungkinan ada hal lain pula yang memengaruhi performa klasifikasi. Oleh karena itu, pada penelitian ini model EfficientNet-B0 diusulkan sebagai metode untuk melakukan klasifikasi terhadap penyakit mata dengan berbasis transfer learning. Hal ini karena sebagaimana namanya, EfficientNet dikenal sebagai model yang mampu menyeimbangkan antara akurasi dan efisiensi komputasi secara optimal [9]. Arsitektur ini telah menunjukkan performa terbaik dengan tingkat akurasi mencapai 84,4% pada basis data ImageNet [10] yaitu sebuah basis data benchmark berskala besar yang digunakan secara luas dalam pengujian performa model klasifikasi citra, menjadikannya salah satu model yang efisien dan unggul dalam klasifikasi citra berskala besar.

Adapun tujuan dari penelitian ini adalah untuk mengembangkan model klasifikasi citra penyakit mata menggunakan EfficientNet-B0 dengan pendekatan transfer learning yang mampu mengklasifikasikan berbagai jenis kelainan mata secara akurat. Oleh karena itu, pendekatan yang diusulkan mampu mengidentifikasi kelainan pada mata secara umum tanpa terbatas pada satu jenis gangguan tertentu. Kemudian, adapun manfaat dari penelitian ini adalah untuk menyediakan metode klasifikasi berbasis CNN yang akurat dan efisien untuk mendukung sistem *Computer-Aided Diagnosis* (CAD), serta membantu dan memfasilitasi deteksi awal penyakit mata secara lebih cepat dan tepat sehingga pada akhirnya dapat meningkatkan kualitas hidup masyarakat.

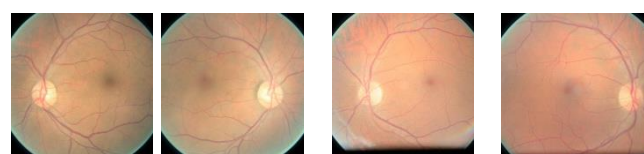
II. METODE

1) Dataset

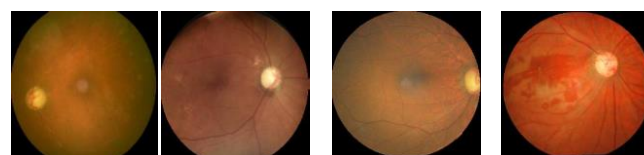
Dataset yang digunakan pada penelitian ini merupakan dataset sekunder yang diperoleh dari platform Kaggle. Dataset ini terdiri dari 4.223 citra fundus retina dengan format JPG yang dikelompokkan ke dalam 4 kelas yaitu *Cataract*, (1038 gambar), *Diabetic Retinopathy* (1104 gambar), *Glaucoma* (1007 gambar) dan Normal (1074 gambar). Adapun beberapa gambar dari dataset ini ditunjukkan pada Gambar 1 hingga Gambar 4.



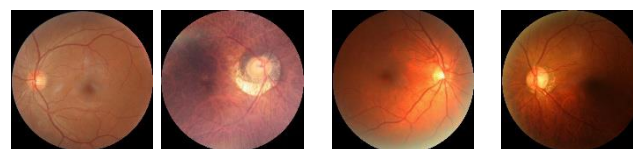
Gambar 1. Kelas *Cataract*



Gambar 2. Kelas *Diabetic Retinopathy*

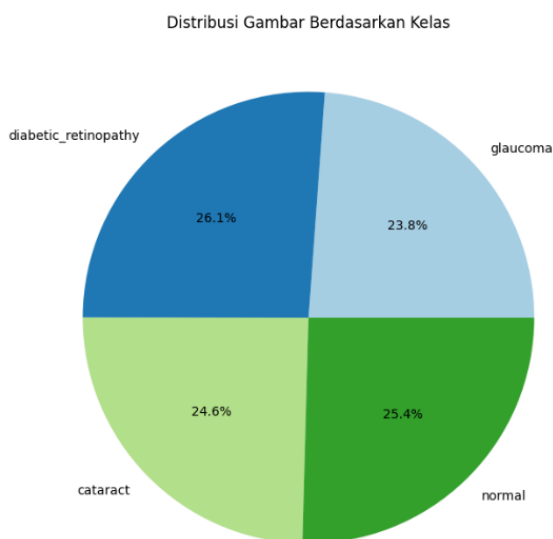


Gambar 3. Kelas *Glaucoma*



Gambar 4. Kelas Normal

Adapun distribusi dari dataset ini disajikan pada Gambar 5. Distribusi antar kelas relatif seimbang, dengan proporsi tiap kelas berada dalam rentang 23–26% dari total data. Semenata, ketidakseimbangan data atau kelas terjadi ketika jumlah satu kelas jauh lebih besar dibandingkan kelas lainnya [11]. Dengan demikian pada penelitian ini tidak diterapkan penyeimbangan dataset tambahan seperti *oversampling* atau *undersampling*. Namun, metrik evaluasi seperti *F1-Score* digunakan untuk mempertimbangkan presisi (berapa banyak prediksi benar dari semua yang diprediksi positif) dan recall (berapa banyak prediksi benar dari semua yang seharusnya positif), sehingga lebih adil dalam menilai performa model pada masing-masing kelas [12].



Gambar 5 Distribusi Data Citra Fundus Retina

2) Pre-processing

Langkah awal dalam tahap *pre-processing* adalah mengubah ukuran (*resize*) seluruh citra menjadi 224x224 piksel, menyesuaikan dengan kebutuhan input dari model pre-trained EfficientNet-B0 [13]. Selanjutnya, dataset dibagi menjadi data latih (*train set*), data validasi (*validation set*), dan data uji (*test set*), dengan proporsi masing-masing sebesar 70%, 15%, dan 15%. Pada data latih, dilakukan augmentasi berupa transformasi acak seperti rotasi hingga 15 derajat, *flipping* horizontal, serta penyesuaian *brightness*, *contrast*, dan *saturation*. Tujuan augmentasi ini adalah untuk memperkaya keragaman data latih dan mengurangi risiko overfitting serta meningkatkan kemampuan generalisasi model terhadap data baru [14]. Sebaliknya, data validasi dan data uji tidak mengalami augmentasi agar hasil evaluasi lebih representatif terhadap performa model pada data nyata [13]. Selain itu, dilakukan normalisasi menggunakan nilai mean dan standar deviasi dari dataset ImageNet dengan mean = [0.485, 0.456, 0.406] dan std = [0.229, 0.224, 0.225] agar distribusi piksel citra sesuai dengan standar pretraining model.

3) Implementasi Efficientnet-B0

Efficientnet adalah arsitektur CNN yang dikembangkan oleh peneliti dari divisi *artificial intelligent* Google [15] yaitu Mingxing Tan dan Quoc V.Le pada tahun 2019 [16]. Arsitektur ini dikenal memiliki akurasi tinggi dan kemampuan komputasi yang lebih efisien [17], [18]. Dalam penelitian ini, digunakan arsitektur EfficientNet-B0 dimana arsitektur ini menjadi dasar pengembangan keluarga EfficientNet dari B1 hingga B7 [10].

Salah satu inovasi utama EfficientNet adalah pendekatan penskalaan model secara seimbang, yaitu dengan:

1. Menambah kedalaman (*depth*) yaitu menambah jumlah lapisan agar model bisa mempelajari pola yang lebih kompleks.
2. Menambah lebar (*width*) yaitu menambah jumlah fitur yang diproses per lapisan.
3. Memperbesar resolusi gambar (*resolution*) yaitu menggunakan gambar input yang lebih besar agar lebih banyak detail bisa ditangkap.

Namun, menambah salah satu dari ketiganya secara berlebihan justru bisa menyebabkan pemborosan sumber daya dan penurunan efisiensi. Karena itu, EfficientNet memperkenalkan teknik yang disebut Teknik Compound Scaling. [19] Compound scaling merupakan metode sistematis yang secara bersamaan menyeimbangkan kedalaman, lebar, dan resolusi dengan menggunakan satu koefisien penskalaan, yaitu ϕ (phi). Rumus matematis penskalaannya ditunjukkan pada persamaan (1) sampai (4).

$$Depth(d) = \alpha^{\phi} \quad (1)$$

$$Width(w) = \beta^{\phi} \quad (2)$$

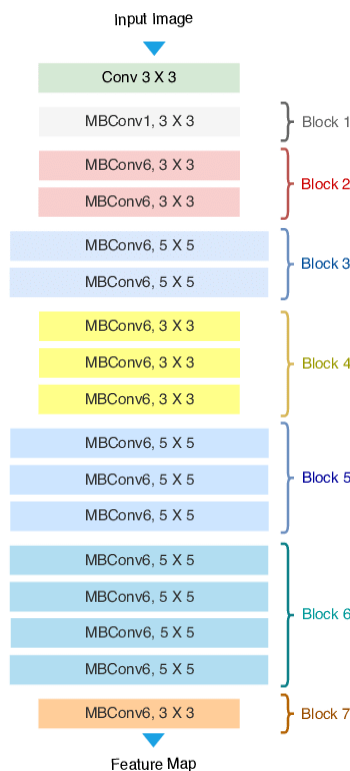
$$Resolution(r) = \gamma^{\phi} \quad (3)$$

dengan syarat:

$$\alpha \times \beta^2 \times \gamma^2 \approx 2, \text{ dan } \alpha \geq 1, \beta \geq 1, \gamma > 1 \quad (4)$$

Koefisien ϕ digunakan untuk menyesuaikan ukuran model berdasarkan kapasitas komputasi yang tersedia. Pendekatan ini juga memperhitungkan kebutuhan komputasi dengan satuan FLOPS (*Floating Point Operations per Second*), yaitu jumlah operasi matematis yang dibutuhkan model. Model awal yang dikembangkan disebut EfficientNet-B0, yang menjadi fondasi untuk varian lainnya (B1 hingga B7). EfficientNet-B0 dirancang menggunakan pendekatan multi-objektif yang mirip dengan MnasNet, dengan fokus pada keseimbangan antara akurasi dan efisiensi komputasi. Model ini kemudian dapat diskalakan menggunakan rumus compound scaling untuk membuat varian EfficientNet yang lebih besar atau lebih kecil, sesuai kebutuhan [19]. Pada penelitian ini digunakan model dasar dari Efficientnet yaitu EfficientNet-B0 sebagai base model dengan arsitektur seperti pada Gambar 6 [15],[16]. Arsitektur ini dimulai dengan layer konvolusi 3x3 yang berfungsi sebagai ekstraksi fitur awal, kemudian diikuti oleh tujuh blok utama MBConv dengan berbagai konfigurasi kernel (3x3 dan 5x5). Blok pertama menggunakan MBConv1 dengan kernel 3x3, sedangkan blok-blok berikutnya menggunakan MBConv6 dengan kernel 3x3 atau 5x5. Setiap blok dibangun untuk menangkap fitur lokal maupun global secara efisien dari citra retina. Seiring bertambahnya kedalaman blok, jumlah saluran juga meningkat, yang memungkinkan model untuk mengekstrak representasi fitur yang lebih kompleks. Setelah melewati semua blok MBConv, arsitektur diakhiri dengan lapisan Global Average Pooling yang merangkum informasi spasial, diikuti oleh dropout sebesar 0.2 untuk mengurangi overfitting. Kemudian, lapisan terakhir adalah dense layer dengan jumlahnya yang disesuaikan yaitu sebanyak empat unit output

yang merepresentasikan jumlah kelas target (*Cataract*, *Diabetic Retinopathy*, *Glaucoma*, dan *Normal*). Untuk fungsi aktivasi digunakan *CrossEntropyLoss*, yang secara internal menggabungkan *log-softmax* dan *negative log-likelihood* [20], dengan demikian fungsi aktivasi *softmax* tidak digunakan pada model karena sudah menggunakan *CrossEntropyLoss*.



Gambar 6 Arsitektur EfficientNet Secara Umum

A. Transfer Learning dan Fine Tuning

Pendekatan transfer learning dilakukan dengan memanfaatkan bobot pretrained EfficientNet-B0. Berbeda dengan pembelajaran mesin konvensional yang mempelajari fitur-fitur dari awal, konsep transfer learning memungkinkan model menggunakan kembali fitur-fitur yang telah dipelajari dari dataset besar sebelumnya untuk diterapkan pada tugas baru, sehingga model tidak mengulang belajar dari awal [21]. Dengan demikian, proses pembelajaran fitur berlangsung jauh lebih cepat dibandingkan dengan tugas pembelajaran mesin tradisional [22].

Pada tahap awal, seluruh layer fitur dibekukan (*frozen*) dan hanya layer klasifikasi yang dilatih. Selanjutnya dilakukan fine-tuning, yaitu membuka lima layer terakhir dari backbone agar ikut dilatih untuk menyesuaikan dengan data domain baru. Pembukaan lima layer ini dipilih karena merupakan bagian akhir dari jaringan yang lebih sensitif terhadap fitur domain spesifik dari dataset retina, sehingga membantu meningkatkan akurasi klasifikasi tanpa mengorbankan efisiensi pelatihan.

B. Konfigurasi Pelatihan

Konfigurasi pelatihan model menggunakan optimizer Adam dengan learning rate sebesar 0.0001 dan batch size sebanyak 32. Proses pelatihan dilakukan dengan maksimum 50 epoch, menggunakan fungsi loss Categorical Crossentropy untuk menangani klasifikasi multikelas. Untuk menghindari pembelajaran yang stagnan, digunakan scheduler ReduceLROnPlateau yang akan menurunkan learning rate sebesar 0.5 jika tidak terjadi peningkatan performa selama 7 epoch (*patience*). Selain itu, diterapkan juga teknik early stopping dengan *patience* sebesar 10 epoch dan delta minimum perubahan loss sebesar 0.0001 untuk mencegah *overfitting* dan mempercepat waktu pelatihan. Untuk memastikan replikasi hasil dan kredibilitas teknis dalam penelitian ini, digunakan berbagai tools dan library open-source yang telah umum digunakan dalam pengembangan model deep learning berbasis citra, yaitu:

- Bahasa pemrograman: Python 3.10
- Framework deep learning: PyTorch 2.0
- Lingkungan pengembangan: Google Colab

4) Metrik Evaluasi

Metrik evaluasi yang digunakan dalam tugas *deep learning* memiliki peran penting dalam mencapai performa klasifikasi yang optimal [20]. Evaluasi kinerja model dimulai dengan membentuk confusion matriks yaitu sebuah tabel yang menunjukkan perbandingan antara hasil prediksi model dengan label sebenarnya. Dari matriks ini, dapat diketahui jumlah prediksi yang benar maupun salah, baik untuk kelas positif maupun negatif. Berikut ini adalah definisi dari istilah-istilah yang akan digunakan:

- True Positive (TP) yaitu kondisi saat model memberikan prediksi positif, dan hasil yang sebenarnya juga menunjukkan kondisi positif (prediksi benar).
- True Negative (TN) yaitu kondisi saat model memberikan prediksi negatif, dan pada kenyataannya memang benar negatif (prediksi benar).
- False Positive (FP) yaitu model memberikan prediksi positif, tetapi dalam kenyataannya hasil tersebut seharusnya negatif (prediksi keliru/salah).
- False Negative (FN) yaitu model memberikan prediksi negatif, tetapi sebenarnya hasil yang benar adalah positif (prediksi salah).

Kemudian, berdasarkan informasi dari matriks ini, dihitung metrik evaluasi seperti, *accuracy*, *precision*, *recall*, dan *F1-Score*.

A. Accuracy

Accuracy dapat dicari dengan menghitung rasio antara jumlah prediksi yang benar (baik kelas positif maupun negatif) dengan total seluruh data yang diuji. *Accuracy* memberikan gambaran umum mengenai kemampuan model secara menyeluruh. Perhitungan *accuracy* dapat dilakukan dengan persamaan (5) berikut :

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

B. Precision

Precision dihitung sebagai perbandingan antara jumlah prediksi positif yang benar (true positif) terhadap seluruh prediksi yang dinyatakan positif (baik benar maupun salah). Metrik ini penting ketika kesalahan positif (false positif) perlu diminimalkan. Perhitungan precision dapat dilakukan dengan persamaan (6) berikut :

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

C. Recall

Recall atau sensitivitas mengukur seberapa baik model mampu mendeteksi seluruh data yang benar-benar termasuk dalam kelas positif. Ini dihitung sebagai perbandingan antara jumlah prediksi positif yang benar terhadap jumlah total data aktual yang benar-benar positif. Metrik ini penting dalam konteks di mana kesalahan negatif (false negatif) sangat merugikan. Persamaan matematis untuk recall yaitu sebagai berikut :

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

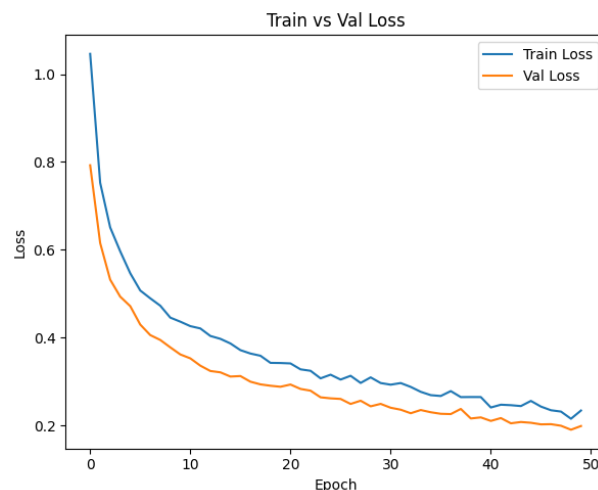
D. F1-Score

F1-Score merupakan ukuran evaluasi yang menggabungkan *precision* dan *recall* dengan mempertimbangkan false positif dan false negatif. F1 Score digunakan untuk menyeimbangkan *precision* dan *recall*, khususnya dalam kondisi distribusi data antar kelas yang tidak seimbang. Nilai F1 Score akan tinggi hanya jika baik *precision* maupun *recall* memiliki nilai tinggi, sehingga memberikan gambaran kinerja model yang lebih seimbang. Adapun rumus untuk menghitung F1-Score yaitu sebagai berikut:

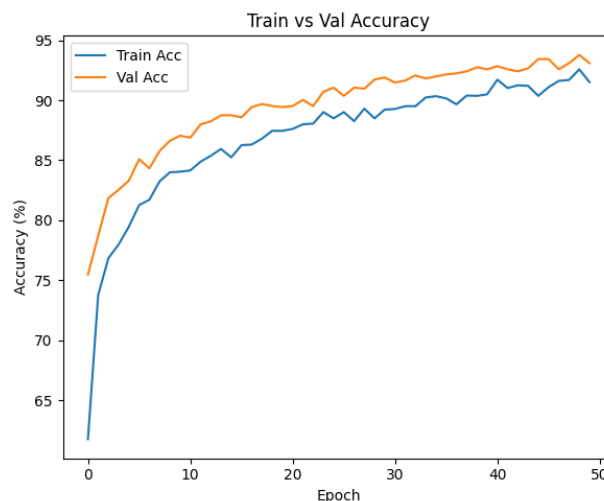
$$F1 - Score = 2x \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

III. HASIL DAN PEMBAHASAN

Gambar 7 menunjukkan bahwa nilai *train loss* dan *validation loss* sama-sama mengalami penurunan yang konsisten dari awal hingga akhir pelatihan. Pada awal pelatihan, *train loss* dimulai dari nilai yang relatif tinggi (>1.0) dan menurun secara stabil hingga mendekati 0.25 di akhir epoch. Sementara itu, *validation loss* juga menunjukkan tren penurunan yang serupa, bahkan sedikit lebih rendah dibanding *train loss* setelah sekitar epoch ke-10. Hal ini menunjukkan bahwa model mampu belajar dari data pelatihan dengan baik, dengan tidak menunjukkan tanda-tanda overfitting yang serius. Perbedaan antara *train loss* dan *validation loss* juga relatif kecil dan stabil, yang berarti model memiliki kemampuan generalisasi yang baik terhadap data yang belum pernah dipelajari sebelumnya. Tren penurunan yang mulus pada kurva *train* dan *validation loss* juga menunjukkan bahwa proses pelatihan berlangsung stabil, begitu pula learning rate yang digunakan cukup optimal. Selain itu, tidak tampak fluktuasi tajam yang biasanya menandakan ketidakstabilan pelatihan.



Gambar 7 Grafik Train dan Validation Loss



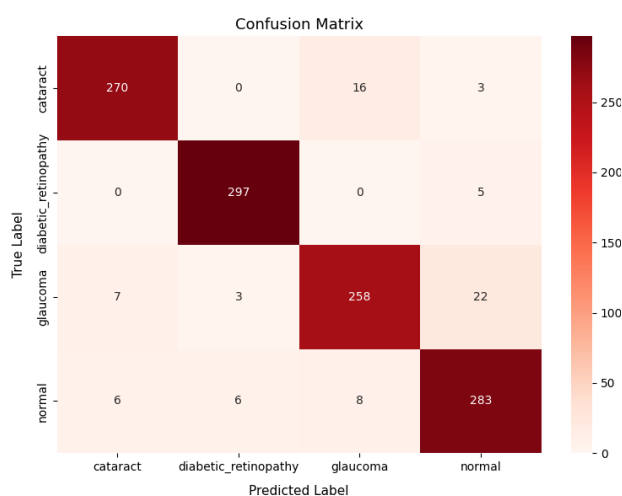
Gambar 8 Grafik Train dan Validation Accuracy

Grafik 8 menunjukkan bahwa akurasi pelatihan (*train accuracy*) dan validasi (*validation accuracy*) meningkat secara konsisten seiring bertambahnya epoch. Di awal pelatihan, akurasi pelatihan dimulai sekitar 65% dan meningkat hingga mendekati 92%. Akurasi validasi bahkan sedikit lebih tinggi, mencapai sekitar 94–95% pada akhir epoch.

Beberapa poin penting dari grafik *loss* dan *accuracy* yaitu sebagai berikut:

- 1) Model tidak overfitting, karena *validation accuracy* lebih tinggi atau setara dengan *train accuracy*, ini mengindikasikan bahwa model tidak hanya mengingat atau menghafal data yang digunakan saat pelatihan, tetapi juga mampu melakukan generalisasi dengan baik terhadap data baru.
- 2) Stabilitas pelatihan. Kurva kedua akurasi menunjukkan peningkatan yang mulus tanpa fluktuasi tajam. Hal ini mencerminkan bahwa model dilatih dengan learning rate yang stabil dan konfigurasi pelatihan yang tepat.

- 3) Performa sangat baik. Dengan akurasi validasi di atas 90%, model menunjukkan performa yang sangat baik dalam klasifikasi empat kelas penyakit mata berbasis citra retina.
- 4) Berdasarkan grafik *Train vs Validation Loss*, terlihat bahwa nilai *training loss* dan *validation loss* secara konsisten menurun seiring dengan bertambahnya jumlah epoch. Hal ini mengindikasikan bahwa model mengalami proses pembelajaran yang baik dan tidak menunjukkan gejala overfitting, karena *validation loss* juga terus menurun hingga akhir pelatihan. Bahkan, *validation loss* cenderung lebih rendah dibandingkan *training loss*.
- 5) Sementara itu, pada grafik *Train vs Validation Accuracy*, akurasi pada data pelatihan dan data validasi terus meningkat dan menunjukkan pola yang stabil. *Validation accuracy* secara konsisten sedikit lebih tinggi dari *training accuracy*, terutama pada akhir pelatihan. Fenomena ini menguatkan asumsi bahwa model tidak mengalami overfitting, dan justru mampu melakukan generalisasi yang efektif terhadap data validasi.
- 6) Secara keseluruhan, kedua grafik ini mencerminkan proses pelatihan yang stabil dan efektif, serta menunjukkan bahwa arsitektur dan parameter yang digunakan telah berhasil menghindari permasalahan umum seperti underfitting ataupun overfitting.



Gambar 9 Confusion Matriks

Evaluasi performa model dilakukan untuk menilai sejauh mana kemampuan model dalam mengklasifikasikan gambar fundus retina ke dalam empat kelas penyakit, yaitu *Cataract*, *Diabetic Retinopathy*, *Glaucoma*, dan Normal. Dua bentuk evaluasi yang digunakan adalah confusion matriks dan *classification report* yang memuat metrik evaluasi seperti *accuracy*, *precision*, *recall* dan *F1-Score*.

Berdasarkan confusion matriks pada Gambar 9, terlihat bahwa model secara umum menunjukkan performa klasifikasi yang baik. Sebagian besar sampel pada masing-masing kelas berhasil diklasifikasikan dengan benar, walaupun masih terdapat beberapa kasus misklasifikasi yang

perlu dicermati lebih lanjut. Hal ini penting, mengingat dalam konteks medis, kesalahan klasifikasi dapat berdampak signifikan terhadap penanganan pasien.

TABEL 1
Classification Report

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
<i>Cataract</i>	0,95	0,93	0,94
<i>Diabetic Retinopathy</i>	0,97	0,98	0,98
<i>Glaucoma</i>	0,91	0,89	0,90
Normal	0,90	0,93	0,92
<i>Accuracy</i>	0,94		

Kinerja terbaik ditunjukkan oleh model pada kelas *diabetic retinopathy*. Dari total 302 citra pada kelas ini, sebanyak 297 berhasil diklasifikasikan dengan benar, hanya lima citra yang salah diklasifikasikan sebagai normal. Hal ini terlihat dari nilai *precision* sebesar 0.97, *recall* sebesar 0.98, dan *F1-Score* yang tinggi sebesar 0.98. Nilai ini menunjukkan bahwa model telah akurat dalam mendeteksi keberadaan *diabetic retinopathy*, sekaligus memiliki sensitivitas yang tinggi terhadap kasus tersebut.

Untuk kelas *Cataract*, model juga menunjukkan performa yang kuat, dengan jumlah klasifikasi benar sebanyak 270 dari 289 sampel. Terdapat beberapa kasus yang salah diklasifikasikan sebagai *Glaucoma* (16 kasus) dan Normal (3 kasus). Nilai *precision*, *recall*, dan *F1-Score* untuk kelas ini masing-masing adalah 0.95, 0.93, dan 0.94. Nilai ini mencerminkan keseimbangan yang baik antara kemampuan model dalam mengenali *Cataract* dan menghindari klasifikasi keliru terhadap kelas tersebut.

Namun demikian, performa model sedikit menurun pada kelas *Glaucoma*. Dari 290 sampel, hanya 258 yang diklasifikasikan dengan benar, sedangkan sisanya diklasifikasikan salah ke kelas *Cataract* (7 kasus), *Diabetic Retinopathy* (3 kasus), dan Normal (22 kasus). Hasil ini menghasilkan *precision* sebesar 0.91, *recall* sebesar 0.89, dan *F1-Score* sebesar 0.90. Cukup tingginya jumlah kesalahan klasifikasi ke kelas Normal menunjukkan bahwa terdapat kemungkinan tumpang tindih fitur antara kedua kelas tersebut, yang menyebabkan model kesulitan dalam membedakan karakteristik visual dari *Glaucoma* dan Normal.

Adapun pada kelas Normal, model berhasil mengklasifikasikan 283 dari 303 citra dengan benar. Meskipun demikian, terdapat sejumlah kecil kesalahan klasifikasi ke kelas penyakit lain, yaitu *Cataract* (6 kasus), *Diabetic Retinopathy* (6 kasus), dan *Glaucoma* (8 kasus). Nilai *precision* dan *recall* masing-masing sebesar 0.90 dan 0.93, dengan *F1-Score* sebesar 0.92. Hal ini menunjukkan bahwa model cukup baik dalam mengidentifikasi kondisi mata yang sehat, walaupun kesalahan klasifikasi ke kelas penyakit perlu diminimalkan lebih lanjut, terutama mengingat pentingnya deteksi dini terhadap kondisi patologis.

Secara keseluruhan, model berhasil mencapai akurasi klasifikasi sebesar 94% pada data uji yang terdiri dari 1.184 citra. Nilai ini mengindikasikan bahwa model memiliki akurasi yang tinggi dan kemampuan generalisasi yang efektif terhadap data baru. Hasil evaluasi ini juga memperkuat potensi implementasi model dalam membantu proses diagnosis berbasis citra medis secara otomatis. Akan tetapi, peningkatan performa pada kelas dengan tingkat misklasifikasi yang relatif lebih tinggi, seperti *Glaucoma*, masih diperlukan. Perbaikan dapat dilakukan melalui teknik augmentasi data, penyesuaian arsitektur model, atau penggunaan pendekatan ensemble untuk meningkatkan kemampuan diskriminatif model terhadap fitur-fitur yang saling tumpang tindih antar kelas.

IV. KESIMPULAN

Berdasarkan hasil evaluasi yang telah dilakukan, model klasifikasi citra fundus retina dengan EfficientNet-B0 menunjukkan performa yang sangat baik dalam mengklasifikasikan empat kategori kondisi mata, yaitu *Cataract*, *Diabetic Retinopathy*, *Glaucoma* dan Normal. Dengan tingkat akurasi keseluruhan sebesar 94%, serta nilai *precision*, *recall*, dan *F1-Score* yang konsisten tinggi pada keempat kelas. Selain itu, berdasarkan grafik performa pelatihan dan validasi, model menunjukkan pola konvergensi yang baik dengan tren *loss* yang menurun dan *accuracy* yang meningkat secara stabil. Hal ini menandakan bahwa model telah berhasil belajar dengan efektif tanpa menunjukkan gejala overfitting, serta memiliki kemampuan generalisasi yang kuat terhadap data yang tidak dikenal. Dengan demikian EfficientNet-B0 terbukti mampu melakukan klasifikasi secara akurat pada data uji.

Performa tertinggi dicapai pada kelas *Diabetic Retinopathy*, dengan *F1-Score* sebesar 0.98, sedangkan kelas *Glaucoma* masih menunjukkan ruang untuk perbaikan akibat misklasifikasi yang relatif lebih tinggi dibandingkan kelas lainnya. Meski demikian, secara umum model telah menunjukkan kapabilitas yang menjanjikan untuk diterapkan dalam sistem pendukung diagnosis berbasis citra medis. Namun, untuk memastikan akurasi dan reliabilitas yang lebih tinggi, penelitian lanjutan dapat diarahkan pada optimalisasi model melalui eksplorasi arsitektur lain, penanganan ketidakseimbangan data, serta validasi eksternal pada dataset klinis yang lebih kompleks dan beragam.

DAFTAR PUSTAKA

- [1] I. McCormick, I. Mactaggart, A. Bastawrous, M. J. Burton, and J. Ramke, "Effective refractive error coverage: an eye health indicator to measure progress towards universal health coverage," Jan. 01, 2020, Blackwell Publishing Ltd. doi: 10.1111/opo.12662.
- [2] L. Fauzi, L. Anggorowati, and C. Heriana, "Skrining Kelainan Refraksi Mata Pada Siswa Sekolah Dasar Menurut Tanda Dan Gejala," *J Health Educ*, vol. 1, no. 1, Aug. 2016, doi: 10.15294.
- [3] R. R. A. Bourne et al., "Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: The Right to Sight: An analysis for the Global Burden of Disease Study," *Lancet Glob Health*, vol. 9, no. 2, pp. e144–e160, Feb. 2021, doi: 10.1016/S2214-109X(20)30489-7.
- [4] R. Indraswari, W. Herulambang, and R. Rokhana, "Deteksi Penyakit Mata Pada Citra Fundus Menggunakan Convolutional Neural Network (CNN) Ocular Disease Detection on Fundus Images Using Convolutional Neural Network (CNN)," *TECHNO.COM*, vol. 21, no. 2, pp. 378–389, May 2022, doi: 10.62411.
- [5] C. A. Putri and S. Rakasiwi, "Diagnosis Dini Penyakit Mata: Klasifikasi Citra Fundus Retina dengan Convolutional Neural Network VGG-16," vol. 9, no. 1, pp. 208–216, Apr. 2025, doi: 10.29408/edumatic.v9i1.29571.
- [6] C. R. Mulyasari, A. I. Hadiana, and A. Komarudin, "Deteksi Penyakit Diabetes, Katarak Dan Glaukoma Pada Citra Fundus Retina Mata Manusia Menggunakan Cnn Dengan Arsitektur Alexnet," *JUMANJI Jurnal Masyarakat Informatika Unjani*, vol. 8, no. 1, pp. 53–68, Apr. 2024, doi: <https://doi.org/10.26874/jumanji.v8i1.341>.
- [7] M. N. I. Muhlashin and A. Stefanie, "Klasifikasi Penyakit Mata Berdasarkan Citra Fundus Menggunakan Yolo V8," *Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 2, p. 1363, Apr. 2023, doi: <https://doi.org/10.36040/jati.v7i2.6927>.
- [8] F. A. Shidik, K. Musthofa, A. P. Kartiningtyas, and T. Agustin, "Seminar Nasional Amikom Surakarta (Semnasa) 2024 Analisis Citra Medis Untuk Identifikasi Penyakit Mata Dengan Teknologi Convolutional Neural Networks," *Prosiding Seminar Nasional AMIKOM Surakarta*, vol. 2, Nov. 2024.
- [9] K. A. Shastry, S. Shastry, A. Shastry, and S. M. Bhat, "A multi-stage efficientnet based framework for Alzheimer's and Parkinson's diseases prediction on magnetic resonance imaging," *Multimed Tools Appl*, Jan. 2025, doi: 10.1007/s11042-025-20597-5.
- [10] V. Ravi, V. Acharya, and M. Alazab, "A multichannel EfficientNet deep learning-based stacking ensemble approach for lung disease detection using chest X-ray images," *Cluster Comput*, vol. 26, no. 2, pp. 1181–1203, Apr. 2023, doi: 10.1007/s10586-022-03664-6.
- [11] M. Dubey, J. Tembhurne, and R. Makhijani, "Improving coronary heart disease prediction with real-life dataset: a stacked generalization framework with maximum clinical attributes and SMOTE balancing for imbalanced data," *Multimed Tools Appl*, Nov. 2024, doi: 10.1007/s11042-024-19429-9.
- [12] N. A. Fajrina, H. Z. Pradana, S. I. Purnama, and S. Romadhona, "Application of the EfficientNet-B0 Architecture in the Classification of Acute Lymphoblastic Leukemia," *Jurnal Riset Rekayasa Elektro*, vol. 6, no. 1, pp. 59–68, Jun. 2024, doi: 10.30595/JRRE.
- [13] P. Gupta, J. Nirmal, and N. Mehendale, "Custom CNN architectures for skin disease classification: binary and multi-class performance," *Multimed Tools Appl*, Dec. 2024, doi: 10.1007/s11042-024-20503-5.
- [14] L. Arora et al., "Ensemble deep learning and EfficientNet for accurate diagnosis of diabetic retinopathy," *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-81132-4.
- [15] A. D. Azzumzumi, M. Hanafi, and W. M. P. Dhuhita, "Klasifikasi Penyakit Paru-Paru Berdasarkan Peningkatan Kualitas Kontras dan EfficientNet Menggunakan Gambar X-Ray," *TEKNIKA*, vol. 13, no. 2, pp. 293–300, Jul. 2024, doi: 10.34148/teknika.v13i2.881.
- [16] A. Irsyad, Islamiyah, and F. Amal, "Klasifikasi COVID 19 dengan Metode EfficientNet Berdasarkan CT Scan Paru-Paru," *Fountain of Informatics Journal*, vol. 8, no. 2, 2024, doi: 10.21111/fij.v8i2.10497.
- [17] N. Gupta, H. Garg, and R. Agarwal, "A robust framework for glaucoma detection using CLAHE and EfficientNet," *Visual Computer*, vol. 38, no. 7, pp. 2315–2328, Jul. 2022, doi: 10.1007/s00371-021-02114-5.
- [18] H. Garg, N. Gupta, R. Agrawal, S. Shivani, and B. Sharma, "A real time cloud-based framework for glaucoma screening using EfficientNet," *Multimed Tools Appl*, vol. 81, no. 24, pp. 34737–34758, Oct. 2022, doi: 10.1007/s11042-021-11559-8.
- [19] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," May 2019, [Online]. Available: <http://arxiv.org/abs/1905.11946>
- [20] L. Alzubaidi et al., "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00444-8.

- [21] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *J Big Data*, vol. 3, no. 1, Dec. 2016, doi: 10.1186/s40537-016-0043-6.
- [22] A. Ghosh, B. Soni, and U. Baruah, "A Fine-Tuned EfficientNet B1 Based Deep Transfer Learning Framework for Multiple Types of Brain Disorder Classification," *Iranian Journal of Science and Technology - Transactions of Electrical Engineering*, vol. 48, no. 3, pp. 1279–1299, Sep. 2024, doi: 10.1007/s40998-024-00726-w.