

# Multi-Disease Retinal Classification Using EfficientNet-B3 and Targeted Albuementations: A Benchmark on Kaggle Retinal Fundus Images Dataset

Kurniawan Aji Saputra<sup>1)</sup>, Farrikh Al Zami<sup>2)\*</sup>, Defri Kurniawan<sup>3)</sup>, Muhammad Naufal<sup>4)</sup>, Muslih<sup>5)</sup>, Rama Aria Megantara<sup>6)</sup>, Ricardus Anggi Pramunendar<sup>7)</sup>

<sup>1,2,3,4,5,6,7)</sup> Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia

<sup>1)</sup>[111202214211@mhs.dinus.ac.id](mailto:111202214211@mhs.dinus.ac.id), <sup>2)</sup>[alzami@dsn.dinus.ac.id](mailto:alzami@dsn.dinus.ac.id), <sup>3)</sup>[defri.kurniawan@dsn.dinus.ac.id](mailto:defri.kurniawan@dsn.dinus.ac.id), <sup>4)</sup>[m.naufal@dsn.dinus.ac.id](mailto:m.naufal@dsn.dinus.ac.id), <sup>5)</sup>[muslih@dsn.dinus.ac.id](mailto:muslih@dsn.dinus.ac.id), <sup>6)</sup>[aria@dsn.dinus.ac.id](mailto:aria@dsn.dinus.ac.id), <sup>7)</sup>[ricardus.anggi@dsn.dinus.ac.id](mailto:ricardus.anggi@dsn.dinus.ac.id)

**Submitted** : Oct 27, 2025 | **Accepted** : Nov 13, 2025 | **Published** : Jan 02, 2026

**Abstract:** Retinal diseases remain one of the leading causes of blindness worldwide. This study develops a deep learning pipeline for multiclass retinal disease classification using EfficientNet-B3 combined with Albuementations to improve generalization. We target four classes: cataract, diabetic retinopathy, glaucoma, and normal. We use the Kaggle Retinal Disease dataset (4,217 fundus images) divided into 70% training, 10% validation, and 20% testing. Images are resized to 224×224 and augmented with horizontal flip, random brightness contrast, CLAHE, shiftscale rotate, crop, gamma correction, and elastic transformation. The EfficientNet-B3 backbone is refined after head training with warm-up and learning rate regularization (batch normalization, dropout). After 50 epochs, the best validation performance reaches 0.9526, and on the hold-out test set, the model achieves 95.38% overall accuracy. The F1 scores per class were 1.0000 (diabetic retinopathy), 0.9685 (cataract), 0.9255 (normal), and 0.9184 (glaucoma). Confusion analysis showed that most errors involved glaucoma being misclassified as normal, likely due to optic disc similarities. These results demonstrate that EfficientNet-B3 with targeted augmentation provides accurate and reliable multi-disease screening of fundus images, with the potential to support faster and more consistent triage in clinical workflows. Future research should expand clinical validation and explore attention mechanisms or multimodal input to reduce glaucoma-normal ambiguity.

**Keywords:** albuementations; cataract; diabetic retinopathy; EfficientNet-B3; fundus images; glaucoma; retinal disease classification

## INTRODUCTION

Retinal diseases are a leading cause of blindness and vision impairment worldwide, significantly impacting millions of people (Alzami et al., 2019). Among these, diabetic retinopathy (DR), glaucoma, cataracts, and age-related macular degeneration (AMD) are the most common, with diabetic retinopathy being a major cause of blindness in working-age adults (Cahya et al., 2021). DR results from long-term damage to the retina due to diabetes, often leading to vision loss if not detected and treated early (Mellor et al., 2023). As the global prevalence of diabetes increases, so does the incidence of DR and other retinal diseases, making early detection essential (Mouhafid et al., 2025).

Traditionally, diagnosing retinal diseases has relied on expert ophthalmologists who manually analyze retinal images using techniques like fundus photography. While effective, this process is time-consuming and dependent on the clinician's experience (Meedeniya et al., 2025). Additionally, it can be prone to subjectivity and inconsistencies, which may lead to delayed or incorrect diagnoses (Tiwari et al., 2025). To address these challenges, there has been growing interest in automating retinal disease diagnosis using deep learning (DL), particularly convolutional neural networks (CNNs). These models have shown considerable promise in medical image analysis by learning and extracting relevant features from images without human intervention (Bilal et al., 2025).

\*name of corresponding author



Recent advancements in deep learning have led to the development of more efficient and accurate models for retinal disease classification. One such model, EfficientNet-B3, has become a prominent solution due to its balance of high accuracy and computational efficiency (Hussain et al., 2024). EfficientNet-B3 achieves superior performance with fewer parameters compared to traditional CNN architectures, making it suitable for real-time clinical applications where computational resources may be limited (Jatmoko et al., 2024).

In addition to the use of efficient architectures like EfficientNet-B3, data augmentation techniques are vital for improving the generalization of deep learning models. Albumentations, a popular data augmentation library, provides a wide range of image transformations that help expand the training dataset by introducing variability, such as rotations, flips, and color adjustments (Qulub et al., 2024). This increased diversity in the data reduces overfitting and improves the model's ability to generalize to unseen data, which is especially important in medical image classification where annotated datasets are often limited and imbalanced (Jatmoko et al., 2024).

The integration of EfficientNet-B3 with Albumentations augmentation holds great potential for enhancing the accuracy and reliability of automated retinal disease detection systems. Such systems can provide faster, more consistent diagnoses, easing the burden on ophthalmologists and enabling early intervention for retinal diseases (Hussain et al., 2024). This is particularly beneficial in underserved regions where access to specialists may be limited (Xu et al., 2021). Automated systems could also support cost-effective screening programs, offering a more accessible diagnostic tool for early detection of retinal diseases, particularly in developing countries (Basit et al., 2025).

Despite the promising potential of deep learning for retinal disease classification, several challenges remain. One major challenge is the lack of large, high-quality annotated datasets, which are essential for training deep learning models (Anwar et al., 2025). While public datasets are available, they are often small, unbalanced, or may not fully represent the diversity of retinal diseases encountered in real-world clinical settings (Giammarco et al., 2025). Another challenge is model interpretability. Medical professionals need transparent models that can explain their decisions, especially in critical cases (Mohan et al., 2025). To overcome these issues, further research is needed to improve model generalization and ensure that automated diagnosis systems are both reliable and interpretable (Aldrees et al., 2025).

Despite notable progress, existing studies still leave several important gaps, including the absence of multi-disease classification using EfficientNet-B3, limited augmentation strategies evaluated for improving model generalization, and the lack of assessments that represent diverse real-world imaging conditions. These limitations indicate the need for an approach that is more robust across multiple retinal diseases. To address this gap, this study develops a classification model based on EfficientNet-B3 combined with Albumentations to increase data variability and reduce overfitting. The research objectives are to build an efficient multi-disease classifier and evaluate its performance using accuracy, precision, recall, and F1-score as standardized metrics, providing a stronger baseline that can be expanded in future clinical-focused research.

This study contributes by developing an EfficientNet-B3-based multiclass classifier trained on publicly available fundus images, integrating Albumentations augmentation to improve generalization, and providing an empirical baseline for evaluating model performance across four retinal disease categories. These contributions offer a foundation that can be expanded in future work related to clinical validation and interpretability-oriented methods.

## LITERATURE REVIEW

This section reviews the current literature on retinal disease classification by examining the methods, datasets, and limitations of previous studies to provide context for this research. Although many studies have successfully detected specific retinal diseases, challenges still persist in developing models that can classify multiple diseases simultaneously, handle limited and varied datasets, and maintain consistent performance without relying on clinical validation or interpretability techniques.

Table 1. Stage of Art Literature Review

Authors	Year	Datasets	Methods	Augmentation	Accuracy	Limitations
(Jatmoko et al., 2024a)	2024	Kaggle Retinal Disease dataset	EfficientNet-B3 with DenseNet	Augmentation was not reported	94.33%	Further improvements are needed to improve the accuracy of the classification process.
(Basit et al., 2025)	2025	Qatari Biobank retinal dataset	Deep Learning with CNN-based transfer learning	Augmentation was not reported	99.76%	High computational cost, not applicable to all retinal diseases.
(Mouhafid et al., 2025)	2024	Glaucoma dataset (ORIGA, ACRIMA)	Multi-branch CNN with Bayesian optimization	Augmentation was not reported	95.09%	Limited to glaucoma detection, not generalizable to other diseases.

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

(Tiwari et al., 2025)	2025	Diabetic Retinopathy dataset (MESSIDOR)	SVM vs CNN for DR detection	Using augmentation in general without technical details.	87.5%	Just focused on DR, lacking multi-disease detection.
(Hussain et al., 2024)	2024	Kaggle Retinal Disease dataset	EfficientNet-B3 + Data Augmentation	Using simple augmentation in the form of horizontal flipping	94%	Small dataset, overfitting issues, lacks clinical real-time validation.
(Jatmoko et al., 2024b)	2024	Eye Disease dataset from local hospitals	CNN with Image Preprocessing (CLAHE)	Augmentation was not reported	94.5%	High reliance on manual annotations, limited to single institutions.
(Giammarco et al., 2025)	2025	OCT Image Dataset from local clinics	Explainable AI, CNN with CAM for localization	Using K-Fold Cross Validation instead of augmentation	92.5%	Focused on localization, not comprehensive disease classification.
(Meedeniya et al., 2025)	2025	ORIGA, ACRIMA datasets	CNN-based glaucoma detection	Using augmentation in general without technical details.	94.5%	Not applicable for multi-disease classification.
(Qulub et al., 2024)	2024	Fundus images from multiple hospitals	CNN-based detection with image preprocessing	Using augmentation in general without technical details.	88.9%	Overfitting on small dataset, dependency on manual image labelling.

The research gap table reveals several challenges faced by existing studies. One significant limitation is the focus on single-disease detection. For example, (Mouhafid et al., 2025) and (Basit et al., 2025) achieve high accuracy in detecting these diseases, they reaching 95.09% and 99.76% accuracy. However, these studies typically focus on one disease at a time, leaving a gap in multi-disease classification, which is essential in clinical settings. This research, using EfficientNet-B3 and Albumentations augmentation, aims to explore multi-disease classification by evaluating four retinal conditions (cataract, diabetic retinopathy, glaucoma, and normal) thus providing an initial benchmark for models that handle multiple classes within a unified framework.

Another major challenge is dataset limitations. (Hussain et al., 2024) and (Qulub et al., 2024) use small or imbalanced datasets, which result in overfitting and poor generalization. They report 94% and 88,9% accuracy, but acknowledge the overfitting issues. To overcome this, we research applies Albumentations data augmentation, increasing dataset diversity through transformations like rotations, flips, and color adjustments. These augmentations help improve the model’s robustness during training, although they do not fully resolve the broader challenges of dataset diversity found in real clinical environments.

The lack of clinical validation is another gap in existing studies. (Jatmoko et al., 2024a) and (Hussain et al., 2024) report high accuracy in controlled environments, but their models lack testing in real-world clinical settings. (Basit et al., 2025) also highlight high computational costs, which limit applicability. This study focuses on improving model generalization using data augmentation and transfer learning. The evaluation is conducted entirely on publicly available Kaggle retinal datasets; therefore, clinical validation is left for future work.

Interpretability remains a significant challenge in retinal disease classification. Previous studies such as (Giammarco et al., 2025) and (Jatmoko et al., 2024b) employed Class Activation Mapping (CAM) and CLAHE to highlight disease-related regions, although their focus was limited to specific conditions. This study focuses on improving classification performance across multiple retinal diseases using EfficientNet-B3 and Albumentations augmentation. While interpretability techniques such as Grad-CAM or saliency maps are not implemented in this work, they represent an important direction for future research to enhance transparency and clinical acceptance.

Although several studies report issues related to dataset imbalance in retinal disease classification, such as in (Tiwari et al., 2025) and (Meedeniya et al., 2025). This study applies Albumentations augmentation to increase visual diversity during training. The purpose of this augmentation is not specifically to balance class proportions, but to improve the model’s ability to generalize to various retinal image variations. This approach helps enhance robustness and reduce overfitting, particularly when working with datasets that may contain subtle differences between classes.

In conclusion, the research gaps identified in the existing literature indicate the need for multi-disease classification models, stronger generalization across diverse datasets, and approaches that offer better transparency in the decision-making process. This study utilizes EfficientNet-B3 combined with Albumentations augmentation to improve robustness and achieve high classification performance on fundus images. Although the experiments are limited to a public dataset, the proposed approach demonstrates promising results and provides a foundation that can be extended toward broader clinical applications in future studies. This work contributes to existing research by presenting a model that effectively classifies multiple retinal diseases and offers a baseline for further improvement in clinical validation and interpretability-oriented methods.

\*name of corresponding author



## METHOD

This study uses an experimental approach to develop a deep learning-based retinal disease classification system utilizing the EfficientNet-B3 architecture combined with the Albumentations augmentation technique. The research methodology is systematically structured, starting from data collection, pre-processing, augmentation, model training, and performance evaluation using appropriate classification metrics.

### Dataset and Data Sharing

The dataset used in this study comes from the Kaggle Retinal Disease Dataset, which contains 4,217 retinal fundus images divided into four main categories: Cataract, Diabetic Retinopathy, Glaucoma, and Normal, which are visualized in the image below. All images are in RGB format with resolutions varying between 256x256 and 512x512 pixels. Before being used in the training process, all images were resized to 224x224 pixels to match the standard input of the EfficientNet-B3 architecture.

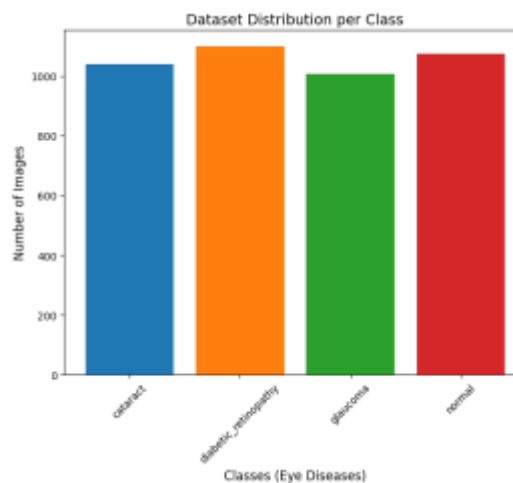


Fig. 1 Dataset Distribution

The distribution of the number of images for each class can be seen in the figure 1 above, which visualizes the relatively balanced number of datasets between classes. The dataset was then divided into three subsets using a stratified sampling method: 70% for training (2,951 images) were used as the training set, which served to train the model and adjust the neural network weights, 10% for validation (422 images) were used as the validation set to monitor model performance during the training process and prevent overfitting, and 20% for testing (844 images). Then, for the test data, which amounted to 844 images, the class distribution is as follows: Cataract with 208 images, Diabetic Retinopathy with 220 images, Glaucoma with 201 images, and Normal with 215 images. This distribution is used to support all evaluation metrics per class reported in the Results Chapter. The 70–10–20 ratio was chosen from a similar reference journal because it yields better results in deep learning-based research and provides sufficient data for the learning process. The data distribution proportions are visualized in Figure 2, while Figure 3 shows some examples of retinal images from each disease class used in this study.

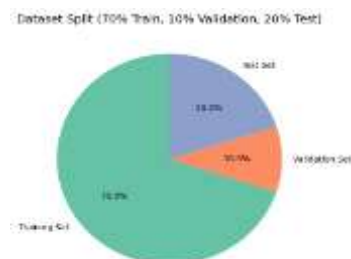


Fig. 2 Dataset Splitting

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

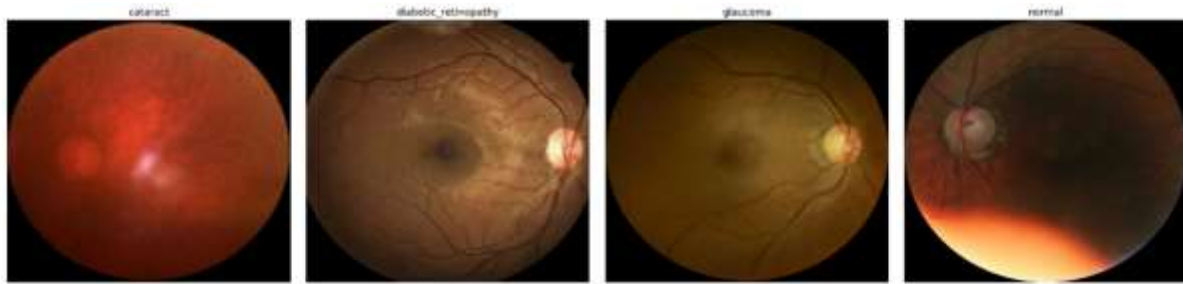


Fig. 3 Dataset Sample Images

**Preprocessing and Data Augmentation**

The preprocessing stage involves image augmentation using the Albumentations library, which expands visual variation and reduces the risk of overfitting. This technique helps the model recognize retinal images under varying conditions, such as lighting, orientation, and texture differences commonly found in medical images. For details, see the table below.

Table 2. Pipeline Augmentation

Transformation (Variabel)	Parameter / Value	Description
HorizontalFlip	p = 0.5	Flips the image horizontally with 50% probability to simulate left-right retinal orientations and improve directional adaptability.
RandomBrightnessContrast	p = 0.2	Randomly adjusts brightness and contrast (20%) to handle illumination variations in retinal images.
CLAHE ( <i>Contrast Limited Adaptive Histogram Equalization</i> )	p = 0.3	Enhances local contrast (30%) in dark retinal regions, making vessels and fine structures more visible.
ShiftScaleRotate	shift_limit=0.02, scale_limit=0.05, rotate_limit=5, p=0.4	Randomly shifts the image position (up to 2%), scales it ( $\pm 5\%$ ), and rotates it ( $\pm 5^\circ$ ) with 40% probability to simulate minor camera movements and improve model robustness to positional variations.
RandomSizedCrop	min_max_height = (150, 224), size = (224,224), p=0.3	Randomly crops image regions with heights between 150-224 pixels, then resizes them to 224x224. This helps the model in 30% focus on key retinal areas such as the macula and optic disc.
RandomGamma	p = 0.2	Randomly in 20% applies gamma correction to adjust non-linear lighting and enrich color intensity variations.
ElasticTransform	p = 0.2	Applies in 20% random elastic deformation to simulate natural distortions from eye or camera movement without altering retinal anatomy.

The results of applying this augmentation pipeline yield a wider variety of retinal images without altering key biological characteristics. An example of the augmentation results is shown in Figure 4, which shows the original retinal image along with the various transformations applied.

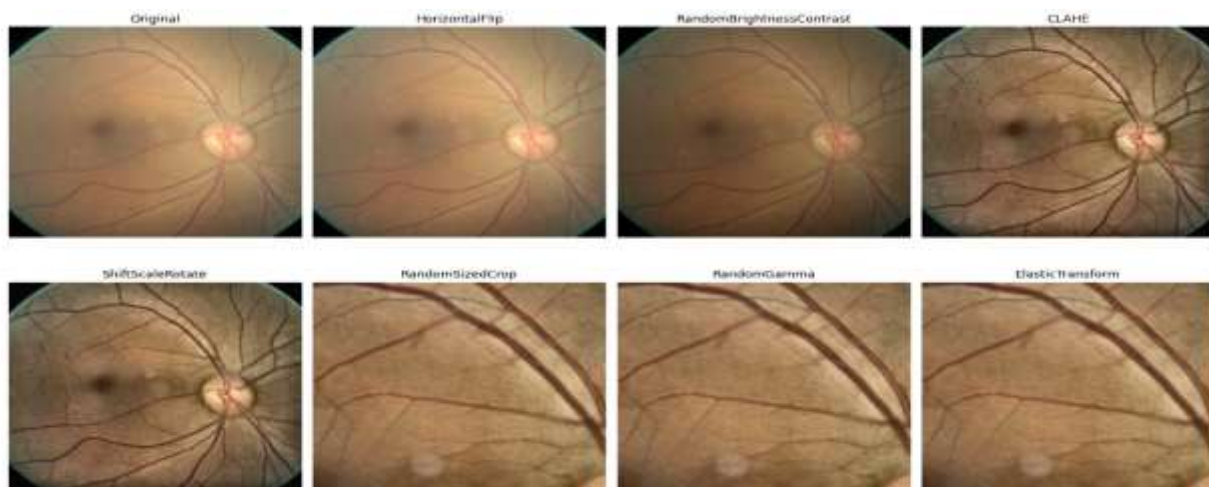


Fig. 4 Albumentations Augmentation Results

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

### Model Architecture Design

The model used in this study is EfficientNet-B3, which consists of two main components: a feature extractor (backbone) and a classifier head. The backbone extracts feature from 224x224x3 pixel retinal images through several convolution blocks. The process begins with a preprocessing stage, followed by an initial convolution layer (3x3 Conv) with a stride of 2, resulting in an output size of 150x150 pixels. The features are then processed through several Mobile Inverted Bottleneck Convolution (MBConv) blocks with 3x3 and 5x5 kernel configurations, varying the number of iterations and strides, gradually reducing the spatial dimensionality from 150x150 to 10x10. Each MBConv block aims to enrich the image feature representation while maintaining parameter efficiency.

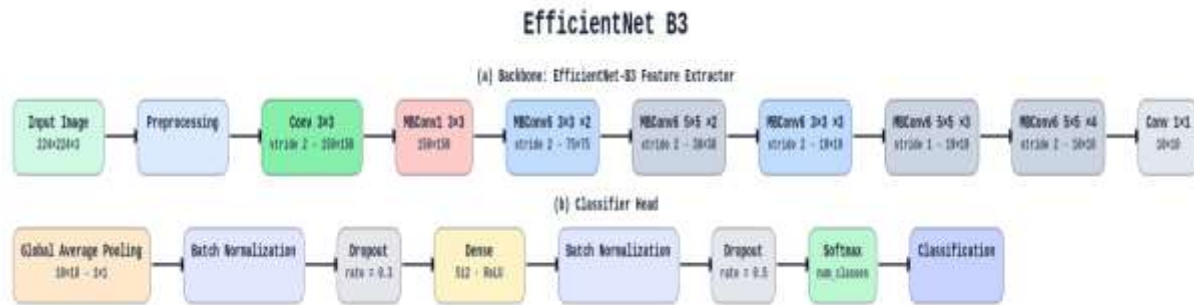


Fig. 5 EfficientNet-B3 Model Architecture

The classifier head receives the output from the backbone in the form of a 10x10 pixel tensor, which is then reduced to 1x1 using Global Average Pooling. Afterward, a Batch Normalization process is performed to stabilize the activation values before entering the Dropout layer with a removal rate of 0.3 to reduce the risk of overfitting. Next, the features are processed by a 512-neuron Dense layer with a ReLU activation function, followed by a second Batch Normalization and Dropout with a rate of 0.5. The final stage uses a Softmax layer to convert the output results into probabilities for each target class, resulting in a retinal disease classification prediction. The complete model structure can be seen in Figure 5 above.

### Evaluation Model

The best model from the training results was evaluated using 844 retinal images from the test dataset, which were not involved in the training or validation process. The evaluation was conducted by measuring classification performance based on several key metrics: Accuracy, Precision, Recall, F1-Score, and two further derivatives: Macro Avg and Weighted Avg. All metrics were calculated based on the Confusion Matrix for each retinal disease class.

Accuracy: the proportion of correctly classified samples out of the total.

$$Accuracy = \frac{\sum_{i=1}^N 1(\hat{y}_i = y_i)}{N} \quad (1)$$

Precision: the fraction of correct predictions within a given class.

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

Recall: the fraction of actual class samples correctly identified.

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

F1-score: harmonic mean of Precision and Recall.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Macro-average F1 (Macro-F1): the unweighted average of F1 across all classes, reflecting balanced performance.

$$Macro\ Average\ (Macro-Avg) = \frac{1}{K} \sum_{k=1}^K M_k \quad (5)$$

Weighted-average F1 (weighted-F1): the class-size-weighted average F1, accounting for class imbalance.

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

$$\text{Weighted Avg} = \frac{\sum_{k=1}^K n_k \times M_k}{\sum_{k=1}^K n_k} \quad (6)$$

Conceptually, Accuracy indicates the proportion of correctly classified samples out of the total. Precision represents the accuracy of predicting a class, Recall indicates the extent to which the model successfully recognizes all actual samples from that class, and the F1-Score is the harmonic mean between the two. Macro Avg provides an unweighted average of all values per class, reflecting the balance of performance between classes without considering the data size of each class. Meanwhile, Weighted Avg considers class size as a weight, providing a fairer picture in datasets with imbalanced distributions.

This study, all of the above metrics were used to assess the model's ability to recognize four classes of retinal diseases: Cataract (208 images), Diabetic Retinopathy (220 images), Glaucoma (201 images), and Normal (215 images). The evaluation was conducted to ensure that the model not only achieved high overall accuracy but also consistently performed across each retinal disease class.

## RESULT

### Training and Validation Performance

The EfficientNet-B3 model was trained over 50 epochs. In the first 10 epochs, all base layers were frozen to ensure only the classification layer was trained, using a learning rate of 1e-3. The training process used callbacks such as EarlyStopping, ReduceLROnPlateau, and ModelCheckpoint to optimize the val\_accuracy and val\_loss. In the first epoch, the training accuracy was 0.5918, with a val\_accuracy of 0.7583. In epoch 5, the val\_accuracy increased to 0.8981, and in epoch 10, the validation accuracy was recorded at 0.8886, with the validation loss decreasing from 0.6036 to 0.3096.

From epoch 11 onward, the learning rate was initially set to 1e-6 and gradually increased. In fine-tuning, LearningRateScheduler is used to set a warm-up at the beginning of each epoch with lower learning rates, namely 1e-6, 5e-6, and 1e-5 for the first to third epochs. This aims to ensure more stable training and eliminate gaps when switching from head to fine. More specifically, the LR-Warm-Up logic aims to allow the model to gradually adapt to weight updates, minimizing initial fluctuations that can disrupt the optimization process.

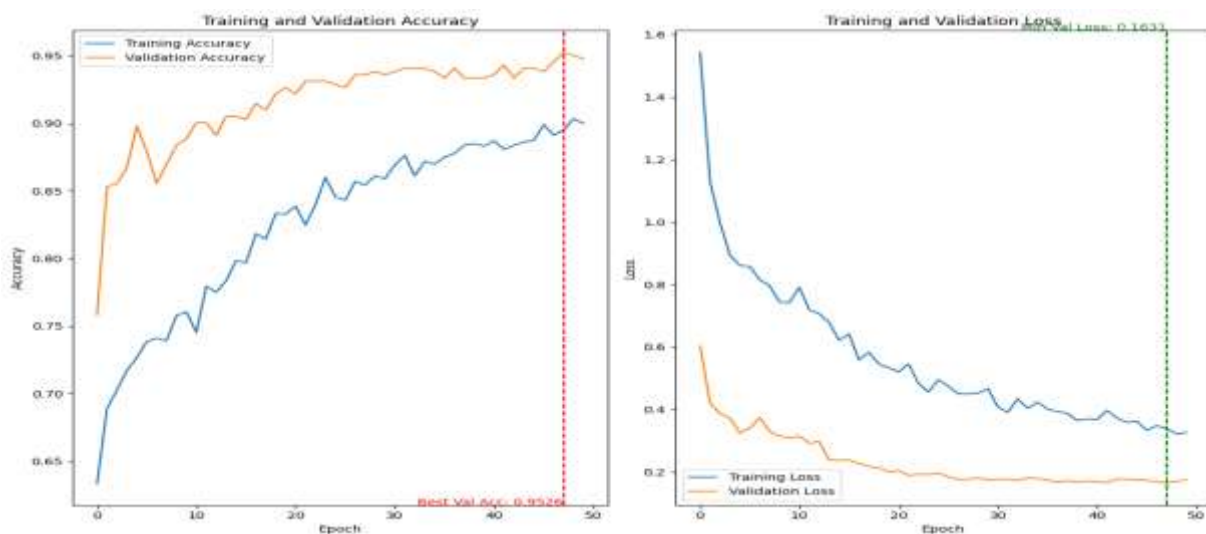


Fig. 6 Training Result Visualization

The image above shows the results of this training process. In this training process, the highest accuracy reached 0.9121 and the val\_accuracy increased to 0.9526, with the validation loss decreasing to 0.1631. Despite continued improvement, the training results stagnated after the 48th epoch, indicating that the model had achieved optimal performance on the existing dataset. Overall, Fine-Tuning increased the val\_accuracy to 0.9526, while Head Training achieved a val\_accuracy of 0.8981. The model shows stable performance, with the learning rate adjusted gradually to ensure good optimization without any indication of overfitting.

### Test Set Performance

After the training phase, the EfficientNet-B3 model was tested on 844 test data images consisting of four classes: Cataract (208 images), Diabetic Retinopathy (220 images), Glaucoma (201 images), and Normal (215 images). The evaluation was carried out using Accuracy, Precision, Recall, and F1-Score metrics. The evaluation

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

results showed an overall accuracy of 95.38%. Below is a table showing the classification report of the model on the test data.

Table 3. Classification Report for Test Set

Class	Precision	Recall	F1-Score	Support
Cataract	0.9756	0.9615	0.9685	208
Diabetic Retinopathy	1.0000	1.0000	1.0000	220
Glaucoma	0.9424	0.8955	0.9184	201
Normal	0.8991	0.9534	0.9255	215
Accuracy	0.9538	0.9538	0.9538	0.9538
Macro Avg	0.9543	0.9526	0.9531	844
Weighted Avg	0.9546	0.9538	0.9538	844

The table above shows that Diabetic Retinopathy performed perfectly, with precision, recall, and an F1-score reaching 1.0000. Cataract also performed very well, with an F1-score of 0.9685. Although Glaucoma had the lowest F1-score at 0.9184, this still represents a fairly good result. Meanwhile, Normal had an F1-score of 0.9255.

Overall, this model demonstrated excellent identification capabilities for retinal diseases. The Diabetic Retinopathy class performed best, followed by the Cataract class. Meanwhile, the Glaucoma and Normal classes ranked third and fourth, with slightly lower results.

**Confusion Matrix and Error Distribution**

Confusion matrix in figure 7 below is analysis of the test data showed that the model distribution successfully classified retinal images. In the Cataract class, the model correctly classified 200 images, although 5 images were incorrectly classified as Normal and 3 images as Glaucoma. For the Diabetic Retinopathy class, the model correctly classified 220 retinal images. In the Glaucoma class, although 180 images were correctly classified, 18 images were incorrectly classified as Normal and 3 images as Cataract. Meanwhile, in the Normal class, 205 images were correctly classified, but 8 images were incorrectly classified as Glaucoma and 2 images as Cataract.

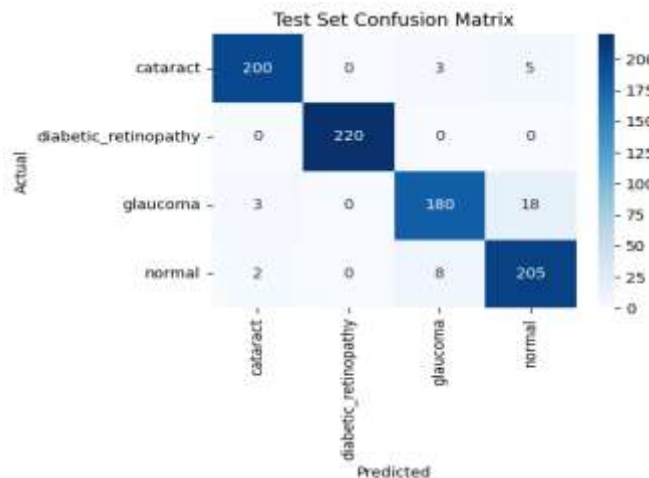


Fig. 7 Confusion matrix

As seen in the figure above, the model demonstrated high performance with excellent accuracy in the Diabetic Retinopathy and Cataract classes. However, misclassifications were more frequent in Glaucoma, which is difficult to distinguish from Normal due to the visual similarities in the retinal fundus images, particularly in the optic disc area. Nevertheless, the model still shows stable performance with high accuracy, with overall accuracy reaching 95.38% and high F1-score in almost all classes, which indicates its ability to classify retinal diseases effectively.

**Summary of Results**

All experimental results demonstrate that the EfficientNet-B3 model trained with Albumentations augmentation achieves high performance in multi-class retinal disease classification. The training process demonstrated good stability across training and validation data, with a minimum loss of 0.16 and a peak accuracy of 95.38%. These results indicate that the model successfully learned while avoiding overfitting. Based on testing on 844 test images, the model achieved an overall accuracy of 95.38%. The Diabetic Retinopathy class performed

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

best with a perfect F1-score of 1.0000, demonstrating its ability to recognize the hallmarks of this disease. Conversely, the Glaucoma class had the lowest F1-score of 91.84%, due to the similarity of visual patterns to the Normal class, particularly in the optic disc area.

All results presented above are factual data from the experiments conducted, without any modifications or assumptions beyond the actual results. This reflects the model's performance.

### DISCUSSIONS

Experimental results show that the EfficientNet-B3 model trained with Albumentations augmentation successfully achieved high performance in retinal disease classification. With 95.38% accuracy on the test data, this model demonstrated excellent ability to identify retinal diseases, especially in the Diabetic Retinopathy class, which had an F1-score of 1.0000. The Cataract class also performed very well with an F1-score of 0.9685. However, the Glaucoma class had the lowest F1-score of 91.82%, indicating challenges in distinguishing between Glaucoma and Normal due to the visual similarity of the optic disc. This is consistent with the confusion matrix results, where some Glaucoma images were incorrectly classified as Normal.

Compared to previous research, the EfficientNet-B3 model showed a slight improvement in accuracy and training speed. For example, research by (Hussain et al., 2024) using the same model achieved 94% accuracy, while this model achieved 95.38% accuracy on a similar dataset. Despite this, distinguishing between glaucoma and normal classes remains a challenge, primarily due to the similarity of optic disc patterns, as reported in previous studies. This suggests that dataset bias could be a significant factor affecting model performance, as the dataset may not fully capture the diversity of clinical conditions, particularly subtle structural changes in glaucoma.

One reason for the model's performance is its use of Mobile Inverted Bottleneck Convolution (MBConv) blocks, which reduce computational costs while maintaining performance, making it suitable for real-time applications in medical image classification tasks. However, the model still struggles with distinguishing between classes with similar features, such as glaucoma and normal retinal images, highlighting the need for further optimization. In conclusion, while EfficientNet-B3 shows strong potential for retinal disease classification, challenges remain, particularly in distinguishing visually similar conditions. Future research should focus on incorporating more diverse and clinically validated datasets to reduce bias and improve generalizability, along with further analysis of feature representation and model optimization to enhance robustness and accuracy.

### CONCLUSION

This study successfully developed the EfficientNet-B3 model with Albumentations augmentation for retinal disease classification, demonstrating superior accuracy and stability. The resulting model achieved 95.38% accuracy on the test data, with the highest F1-score in the Diabetic Retinopathy class, reaching 1.0000. However, the greatest challenge remained with the Glaucoma class, which faced classification difficulties due to its visual similarity to the Normal class, particularly in the optic disc area.

The main contribution of this study is the application of augmented data techniques to improve model performance, particularly in addressing class imbalance in medical datasets. The effectiveness of this augmentation strategy in improving model generalization and robustness is a key contribution of the study. However, the model's limitations include challenges in distinguishing between visually similar classes like Glaucoma and Normal, and the dataset's lack of diversity, which may impact the model's generalizability to a wider range of clinical conditions.

Suggestions for future research include using more diverse datasets to reduce bias and improve generalization, and incorporating multimodal imaging or transformer-based networks to handle more subtle visual differences between classes. Furthermore, the use of attention mechanisms in the model could be a crucial step in improving classification accuracy between classes with higher feature similarities, such as Glaucoma and Normal.

### ACKNOWLEDGMENT

The authors gratefully acknowledges the Intelligent Distributed Surveillance and Security (IDSS) Laboratory for providing essential facilities, guidance, and technical support that made this research possible.

### REFERENCES

- Aldees, A., Min, H., Dutta, A. K., Daradkeh, Y. I., & Anjum, M. (2025). Improving Fundus Detection Precision in Diabetic Retinopathy Using Derivative-Based Deep Neural Networks. *CMES - Computer Modeling in Engineering and Sciences*, 142(3), 2487–2511. <https://doi.org/10.32604/cmcs.2025.061103>
- Alzami, F., Abdussalam, Megantara, R. A., Fanani, A. Z., & Purwanto. (2019). Diabetic Retinopathy Grade Classification based on Fractal Analysis and Random Forest. *2019 International Seminar on Application for Technology of Information and Communication (ISEMANTIC)*, 272–276. <https://doi.org/10.1109/ISEMANTIC.2019.8884217>

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

- Anwar, M., Farhan, S., Ul Haq, Y., Azeem, W., Ilyas, M., Voicu, R. C., & Tanveer, M. H. (2025). E-GlauNet: A CNN-Based Ensemble Deep Learning Model for Glaucoma Detection and Staging Using Retinal Fundus Images. *Computers, Materials and Continua*, 84(2), 3477–3502. <https://doi.org/10.32604/cmc.2025.065141>
- Arshad Hussain, M. S., Babu, S., Sri Satya Sai, M. K., Siddartha, K., & Naik, K. B. (2024). Retinal Disease Classification using EfficientNet-B3. *Proceedings - 2024 4th International Conference on Soft Computing for Security Applications, ICSCSA 2024*, 337–344. <https://doi.org/10.1109/ICSCSA64454.2024.00060>
- Basit, S. A., Al-Absi, H. R. H., Musleh, S., & Alam, T. (2025). Cardiometabolic biomarker prediction based on retinal fundus image. *Engineering Applications of Artificial Intelligence*, 160. <https://doi.org/10.1016/j.engappai.2025.111734>
- Bilal, H., Keles, A., & Bendeche, M. (2025). Advances in disease detection through retinal imaging: A systematic review. In *Computers in Biology and Medicine* (Vol. 194). Elsevier Ltd. <https://doi.org/10.1016/j.compbiomed.2025.110412>
- Cahaya, F. N., Hardi, N., Riana, D., & Hadiani, S. (2021). Klasifikasi Penyakit Mata Menggunakan Convolutional Neural Network (CNN). *SISTEMASI: Jurnal Sistem Informasi*, 10(3), 618–626. <https://doi.org/https://doi.org/10.32520/stmsi.v10i3.1248>
- Di Giammarco, M., Santone, A., Cesarelli, M., Martinelli, F., & Mercaldo, F. (2025). Explainable retinal disease classification and localization through Convolutional Neural Networks. *Image and Vision Computing*, 162. <https://doi.org/10.1016/j.imavis.2025.105667>
- Hasan, M. N., Pial, M. E. R., Das, S., Siddique, N., & Wang, H. (2025). DIA-VXNET: A framework for automated diabetic eye disease detection using transfer learning with feature fusion network. *Biomedical Signal Processing and Control*, 100. <https://doi.org/10.1016/j.bspc.2024.106907>
- Jatmoko, C., & Lestiawan, H. (2024). PREDIKSI PENYAKIT MATA MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK. (SEMNAS RISTEK) *Seminar Nasional Riset Dan Inovasi Teknologi*. <https://doi.org/https://doi.org/10.30998/semnasristek.v8i01.7129>
- Jatmoko, C., Lestiawan, H., Agustina, F., & Erawan, L. (2024). Comparative Study of Classification of Eye Disease Types Using DenseNet and EfficientNetB3. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*. <https://doi.org/10.22219/kinetik.v9i3.1931>
- Meedeniya, D., Shyamalee, T., Lim, G., & Yogarajah, P. (2025). Glaucoma identification with retinal fundus images using deep learning: Systematic review. In *Informatics in Medicine Unlocked* (Vol. 56). Elsevier Ltd. <https://doi.org/10.1016/j.imu.2025.101644>
- Mellor, J., Jiang, W., Fleming, A., McGurnaghan, S. J., Blackbourn, L., Styles, C., Storkey, A. J., McKeigue, P. M., & Colhoun, H. M. (2023). Can deep learning on retinal images augment known risk factors for cardiovascular disease prediction in diabetes? A prospective cohort study from the national screening programme in Scotland. *International Journal of Medical Informatics*, 175. <https://doi.org/10.1016/j.ijmedinf.2023.105072>
- Mohan, R., Kadry, S., Yassine, S., & Rajinikanth, V. (2025). Healthy/Glaucoma Fundus Retinal Image Classification using Butterfly Algorithm Optimized ResNet-Features. *Procedia Computer Science*, 258, 1804–1812. <https://doi.org/10.1016/j.procs.2025.04.432>
- Mouhafid, M., Zhou, Y., Shan, C., & Xiao, Z. (2025). A Robust Approach to Early Glaucoma Identification from Retinal Fundus Images using Dirichlet-based Weighted Average Ensemble and Bayesian Optimization. *Current Medical Imaging Formerly Current Medical Imaging Reviews*, 21. <https://doi.org/10.2174/0115734056335762250128095107>
- Nur, M., Muhlashin, I., Stefanie, A., Universitas, S., Karawang, J. H., Ronggo, W., & Karawang, I. (2023). KLASIFIKASI PENYAKIT MATA BERDASARKAN CITRA FUNDUS MENGGUNAKAN YOLO V8. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 2).
- Putri, C. A., & Rakasiwi, S. (2025). Diagnosis Dini Penyakit Mata: Klasifikasi Citra Fundus Retina dengan Convolutional Neural Network VGG-16. *Edumatic: Jurnal Pendidikan Informatika*, 9(1), 208–216. <https://doi.org/10.29408/edumatic.v9i1.29571>
- Qulub, M. S., & Agustin, S. (2024). IDENTIFIKASI PENYAKIT MATA DENGAN KLASIFIKASI CITRA FOTO FUNDUS MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK (CNN). In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 5).
- Tiwari, S., Sahu, K., Vijh, S., & Awasthi, C. (2025). Deep Learning-Based Classification of Ocular Toxoplasmosis Fundus Images: A Comparative Study of CNN and SVM Models. *Procedia Computer Science*, 259, 1189–1197. <https://doi.org/10.1016/j.procs.2025.04.074>
- Xu, X., Li, J., Guan, Y., Zhao, L., Zhao, Q., Zhang, L., & Li, L. (2021). GLA-Net: A global-local attention network for automatic cataract classification. *Journal of Biomedical Informatics*, 124. <https://doi.org/10.1016/j.jbi.2021.103939>

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.