Scientific Journal of Informatics
Vol. 12, No. 3, August 2025
p-ISSN 2407-7658 https://journal.unnes.ac.id/journals/sji/index e-ISSN 2460-0040

Deep Learning-Based Eye Disorder Classification: A K-Fold Evaluation of EfficientNetB and VGG16 Models

Cinantya Paramita[1*], Sindhu Rakasiwi[2], Pulung Nurtantio Andono[3], Guruh Fajar Shidik[4], Shier Nee Saw[5], Muhammad Ivan Rafsanjani[6]
[1,2,3,4,6]Dinus Research Group for Al in Medical Science (DREAMS), Universitas Dian Nuswantoro, Indonesia
[5]Department of Artificial Intelligence, Universiti Malaya, Malaysia

Abstract.
Purpose: The study evaluates EfficientNetB3 and VGG16 deep learning architectures for image classification, focusing on stability, accuracy, and interpretability. It uses Gradient-weighted Class Activation Mapping to improve transparency and robustness. The research aims to create reliable AI-based diagnostic tools.
Methods: The study used a dataset of 4,217 color retinal fundus images divided into four classes: cataract, diabetic retinopathy, glaucoma, and normal. The dataset was divided into 70% for training, 10% for validation, and 20% for testing. The researchers used a transfer learning approach with EfficientNetB3 and VGG16 models, pretrained on ImageNet. Real-time augmentation was applied to prevent overfitting and improve generalization. The models were compiled with the Adam optimizer and trained with categorical cross-entropy loss. Early stopping was implemented to allocate computational resources efficiently and reduce overfitting. A learning rate scheduler (ReduceLROnPlateau) was added to adjust the learning rate if no significant improvement was made concerning validation loss. EfficientNetB3 was more efficient in model size, possessing only 12 million parameters compared to VGG16's 138 million, making it suitable for resource-constrained mobile or embedded systems. The final evaluation was done on the held-out test set.
Result: The EfficientNetB3 architecture outperforms VGG16 in classification accuracy and loss value stability, with an average accuracy of 93%. It also exhibits better transparency and predicted accuracy, making it a reliable model for medical image categorization.
Novelty: This work introduces a novel framework integrating EfficientNetB3 architecture, stratified cross-valuation, L2 regularization, and Grad-CAM-based interpretability, focusing on openness and explainability in model evaluation.

Keywords: CNN modern, EfficientNetB, Color fundus photography, K-Fold, VGG16, Machine learning, Grad-Cam
Received May 2025 / **Revised** July 2025 / **Accepted** August 2025

INTRODUCTION

With the progression of modern lifestyles, digital devices like computers, cell phones, and handheld tablets have become indispensable in daily activities. However, prolonged usage of these devices can result in eye-related disorders, most notably Computer Vision Syndrome (CVS) [1], [2], which is characterized by symptoms such as eye fatigue, dryness, and blurred vision. Extended screen exposure and suboptimal ergonomic postures further exacerbate these symptoms, particularly among individuals experiencing age-related declines in visual function. Concurrently, retinal diseases such as diabetic retinopathy, glaucoma, and cataracts [3] are among the leading causes of irreversible visual impairment, primarily due to the retina's complex and delicate structure. These diseases taken together emphasize the need of disseminating knowledge about eye health and implementing preventative actions to maintain optimal visual function in the face of fast advancing digital technology.

A number of imaging modalities, such as fundus fluorescein angiography (FFA), color fundus photography (CFP) [4], and optical coherence tomography (OCT), have evolved into devices that are crucial in the field of current ophthalmology for the purpose of identifying and monitoring retinal abnormalities [5]. Among these methods, FFA provides helpful data on tissue perfusion and vascular leakage; CFP is usually regarded

*Corresponding author.
Email addresses: cinantya.paramita@dsn.dinus.ac.id (Paramita)
DOI: 10.15294/sji.v12i3.26257

as the gold standard for non-invasive visual documenting of retinal architecture and vascularity. OCT's sensitivity rises still further when combined with color fundus photographs. Its high-resolution features have shown potential for recognizing choroidal neovascularization and help to provide comprehensive view of the retinal layers. Retinal diseases are typically asymptomatic in their early stages; however, the damage to retinal tissue progresses over time.

The visualization of characteristic lesions through fundus images facilitates a more objective and accurate diagnosis of these conditions, such as figure 1 (a) the presence of bright yellow deposits in the retina resulting from the leakage of fat and protein from damaged capillary blood vessels [6], [7]; (b) the absence of typical retinal abnormalities in some patients with diabetic retinopathy or other retinal disorders; (c) optic nerve damage, often indicate by an increased cup-to-disc ratio, with the optic disc appearing more prominent and brighter in the image [8]; and (d) cloudiness or opacity of the eye lens, particularly in the central area, causing light diffusion an indicative feature of cataracts [9].

(a) diabetic retinopathy (b) normal (c) glaucoma (d) cataract

Figure 1. Illustrative images showcasing multiple eye conditions

Current research in retinopathy emphasizes early diagnosis to prevent serious complications. Image analysis is employed to extract critical features from fundus images, while machine learning techniques are leveraged to develop reliable classification models due to their strong generalization capabilities. To enhance model performance, image augmentation techniques such as rotation, shifting, zooming, and lighting variations are applied to expand training data and mitigate overfitting risks [10]. In classifier development, two primary approaches are utilized: classical image processing and pattern recognition methods, as well as artificial neural network-based approaches. Classical methods require manual feature extraction, whereas neural networks perform more effectively with large datasets but are less optimal when data is limited. Since diabetic retinopathy fundus datasets are typically domain-specific and limited in size, classical methods combined with machine learning are deemed more suitable in this context. An ensemble method emerges as a key strategy in this study. The approach combines transfer learning with deep learning-based CNN architectures [11], [12], including foundational CNN models [13], modern CNNs (EfficientNetB3) [14], and pre-designed CNNs (VGG16) [9], [15].

The comparative analysis of deep learning models for early disease detection demonstrates the superior performance of modern CNN architectures, particularly the EfficientNet family. This superiority is further corroborated by studies across medical domains: EfficientNetB attained 93.47% accuracy in lung cancer detection from CT scans [14], 92.07% [16] in breast histopathology classification, 90,8% for canine cataracts [17], latest research for fundus diseases had result about 87% [18] and EfficientNetB3 with squeeze-and-excitation block detects diabetic retinopathy from fundus images with 88.74% accuracy [19]. The consistent high performance across diverse imaging modalities (fundus photography, CT scans, and histopathology slides) establishes EfficientNet architectures as the optimal choice for automated medical diagnosis systems, combining high accuracy with computational efficiency for clinical deployment.

Additionally, data augmentation is applied to enrich the training dataset, alongside fine-tuning neural networks pre-trained on large-scale datasets such as ImageNet. This combined methodology is expected to improve both accuracy and generalization capabilities in diabetic retinopathy fundus image classification. This study aims to enhance the accuracy of eye disease diagnosis by implementing the EfficientNetB3 deep learning model. The novelty of the research lies in its use of a lightweight yet effective architecture for detecting retinal medical conditions. The developed model is expected to assist medical professionals in early detection [14] and automated diagnosis, while also serving as a reference for advancing machine learning applications in healthcare.

METHODS

The use of appropriate research instruments is essential for ensuring the effective implementation and successful outcomes of a study, enabling accurate data collection and analysis through suitable techniques. This study employed both hardware and software tools. The hardware utilized included a system equipped with an RTX 3070 GPU, AMD Ryzen 7 5500 processor, 16 GB of RAM, a 512 GB SSD, and the Windows 11 operating system. For the software components, the study made use of Python along with several supporting libraries, including OS, NumPy, pandas, computer vision library, matplotlib, TensorFlow, and Scikit-learn. The methodological approach focused on image classification using the EfficientNetB3 and VGG16 models, incorporating cross-validation and image augmentation techniques to classify eye diseases based on the provided image dataset.

This study adopts an experimental approach by applying deep learning methods to classify 4217 fundus images into various disease categories [18]. Two model architectures are compared in this research: EfficientNetB3 and VGG16. Both are transfer learning-based models pre-trained on the ImageNet dataset. The dataset consists of retinal fundus images, which are divided into three subsets: 70% for training, 10% for validation, and 20% for testing. Each subset contains subfolders representing specific eye disease classes. For cross-validation purposes, data from the training and validation folders were merged into a single Data Frame, which was used throughout the training and evaluation process. Data preprocessing involved image augmentation to increase data diversity and prevent overfitting. The augmentation process included random rotations, horizontal and vertical shifts, zooming, shearing, lighting variations, and horizontal flipping. These transformations aim to enable the model to recognize disease patterns in a generalized manner, independent of specific imaging conditions. Model construction was carried out by leveraging the EfficientNetB3 and VGG16 architectures without their top classification layers. The base convolutional layers of both models were initially frozen to retain the learned features from ImageNet. New custom classification layers were added, including a dropout layer with a rate of 0.5 and a fully connected dense layer with L2 regularization where (λ = 0.001), aimed at reducing overfitting and improving generalization. These were then extended with several fully-connected (dense) layers to perform final classification. During initial training, some of the early layers in each model were frozen to retain their pre-trained weights, while the remaining layers were fine-tuned using the research dataset. The models were compiled using the Adam optimizer with a learning rate of 0.0001 [20], loss was calculated using the categorical cross-entropy approach, and model performance was assessed based on accuracy metrics. An overview of the research workflow is illustrated in the following figure 2.

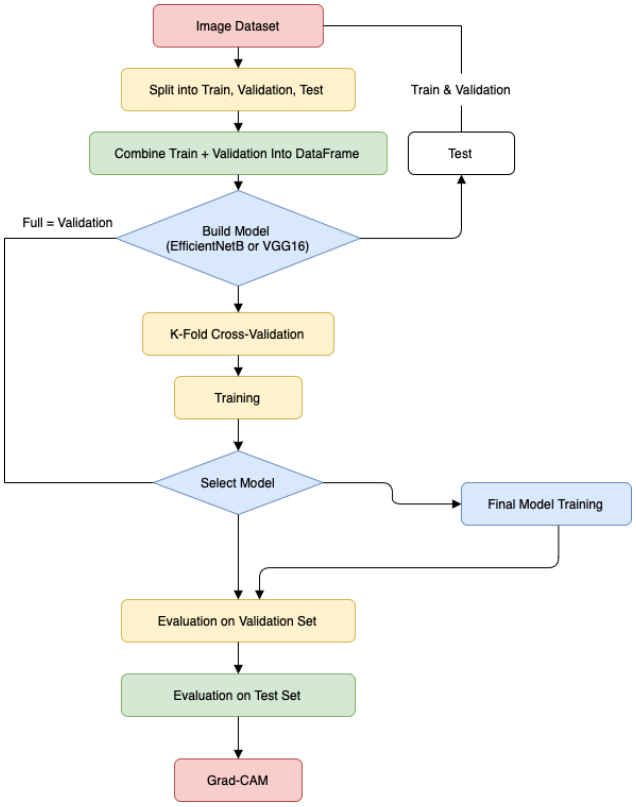


Figure 2. Flow method EfficientNetB3 and VGG16

The process begins with data preparation and model definition, followed by cross-validation using the Stratified K-Fold method [21]. Once the best-performing model is identified, it is retrained using the combined training and validation datasets, then evaluated on the held-out test dataset, which was never used during training or validation. The evaluation includes measuring accuracy and loss, and generating a confusion matrix and classification report with precision, recall, and F1-score for each class, reported both individually and as macro and weighted averages. Model predictions are further interpreted using Grad-CAM (gradient-weighted Class Activation Mapping), which highlights regions of the retinal fundus image that influence the model's decisions. This visualization overlays a heatmap on the original image, helping to understand how the model classifies each case.

Collecting dataset image

The Kaggle platform offers freely available datasets for the aim of machine learning research. Photographs of color fundus photography (CFP) taken under a range of conditions were gathered from there. Kaggle's choice was based on its ability to offer well-structured pre-labeled datasets that allow effective data organization for the aim of model development. Four main categories define the photographs in this collection: cataracts, glaucoma, diabetic retinopathy, and normal eyes. Every category consists of enough variation in terms of image angles, lighting conditions, and image quality to guarantee the most efficient model training and evaluation. Four categories were equally distributed out of the about 4217 images gathered. The variety of the dataset allows the model to be trained to precisely recognize and distinguish between different eye diseases, so enabling a diagnosis that is both more dependable and more swiftly in medical uses.

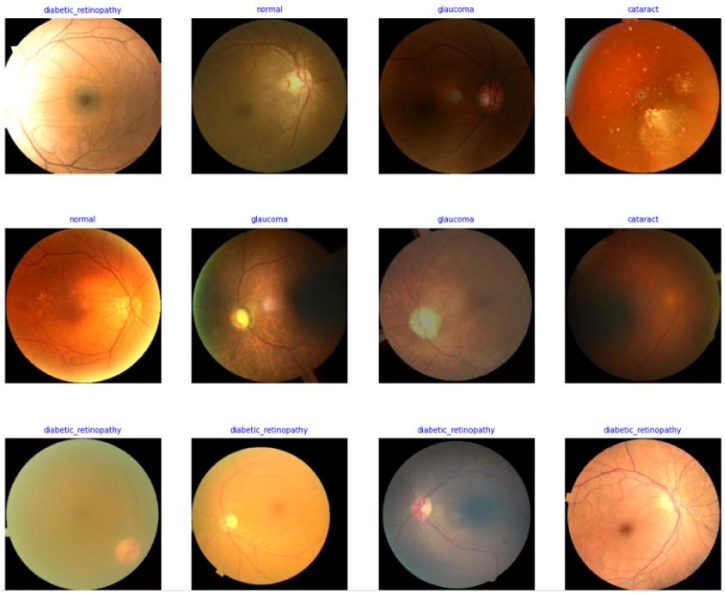


Figure 3. Labeling of color fundus photography for disease identification

Preprocessing

Several cautious phases of data preparation were carried out in this work to ensure that the data were ready enough for the phase of deep learning model training. Retinal fundus photos in the collection have been divided into subfolders based on their labels for eye illnesses. Figure 4 presents a pictorial representation of the used methods of data preparation. Beginning the processing procedure was combining data from the training and validation folders into a single Data Frame structure. Using K-Fold Cross Validation [22] needed this integration since stratified splitting depending on class labels rely on consolidated data kept in a single structure. Then there are steps in picture augmentation and preprocessing. All retinal fundus photos were reduced to 256 × 256 pixels in order to provide constant input dimensions for the model design and normalized to ensure consistent input scale and improve training efficiency. For multi-classification, labels were one-hot encoded. To prevent overfitting and improve generalization, real-time augmentation was applied including rotation, translation, zoom, shear, brightness change, and horizontal flipping.

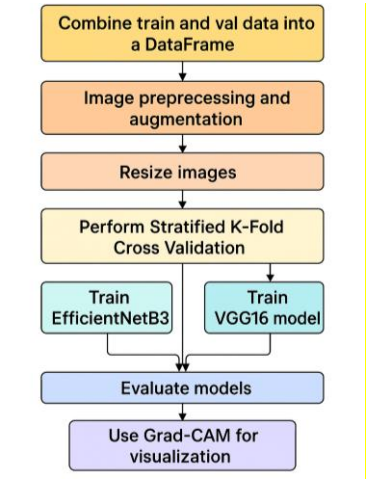


Figure 4. CFP processing stage

Additionally changed were pixel values to suit a realistic range for neural network performance. Variance in the training set was increased and overfitting risk was reduced by random data augmentation. Among the augmenting techniques were zooming, shearing, brightness change, horizontal flipping, random rotation, horizontal and vertical shifting.

Label encoding then transformed image labels into numerical form to enable training for multi-class classification. The stratified K-fold method was then used to separate the dataset so that the labels remained distributed proportionately across all of the many folds. As part of the transfer learning strategy, the training method used two distinct deep learning architectures: VGG16 and EfficientNetB3, both of which used pretrained ImageNet. Accuracy, loss values, confusion matrix, and classification reports were among all the records on training and evaluation outcomes. Additionally, to improve model interpretability, the Grad-CAM (gradient-weighted Class Activation Mapping) technique was used to identify the precise regions of the retinal fundus image that the model concentrated on during prediction.

Training

Prior to model training, the preprocessed and augmented retinal fundus images were structured in a format compatible with deep learning frameworks. Each image was resized to 256 × 256 pixels with three RGB color channels and categorized according to its respective class label [23]. The training process was conducted in two main stages. The first stage involved the application of Stratified K-Fold Cross Validation with five folds (5-fold), aimed at comprehensively evaluating the model's performance across various data distributions and mitigating the risk of overfitting. In each fold, the model was trained on a subset of the training data and validated on a separate subset, maintaining class distribution balance throughout. The accuracy and loss from each fold were recorded and averaged to determine the overall model performance. Upon completion of the cross-validation phase, the model demonstrating the highest performance was selected for retraining using the combined training and validation datasets. The final model was then evaluated on the test dataset, which had been previously set aside. Training was performed using the Adam optimization algorithm configured with a learning rate of 0.0001, a batch size of 32, and a maximum of 15 epochs. Early stopping was implemented with a patience of 3 epochs and a minimum delta of 0.0001 to prevent overfitting. Additionally, the ReduceLROnPlateau learning rate scheduler was applied, reducing the learning rate by a factor of 0.1 when no improvement in validation loss was observed for 5 consecutive epochs, with a minimum learning rate set at 1e−6, employing categorical cross-entropy as the loss function and accuracy as the performance evaluation metric. Early stopping tracked the validation loss (val_loss) during training. Should no significant change be observed over consecutive epochs, the training process was automatically ended and the weights with best performance were maintained. Among the training process outputs were test accuracy, test loss, a confusion matrix, a classification report containing accuracy, recall, and F1-score for every class. In addition employed to show the areas of the retinal fundus pictures most influencing the predictions of the model was Grad-CAM (gradient-weighted class activation mapping), so enhancing the interpretability of the automated diagnostic system.

Evaluation

Determining the effectiveness of the created system and its dependability in classifying color fundus images depends on first model evaluation. Two approaches of evaluation are used: qualitatively and quantitatively. Performance of the model was evaluated quantitatively using metrics including classification accuracy, F1-score, mean matrix values, accuracy and loss curves, and the confusion matrix. Concurrently for a qualitative evaluation, gradient-weighted Class Activation Mapping (Grad-CAM), graphically depicting the Convolutional Neural Network (CNN) decision-making process, was applied. By stressing the parts of the input image most affected on the classification output, this method offers information on the capacity of the model to discover clinically significant characteristics.

The quantitative technique is implemented by computing generally used classification assessment metrics derived from the confusion matrix. The confusion matrix is a tabular showing of every class in the dataset's accurate and false prediction count. Several formulas obtained from this matrix let one evaluate the model's performance.

Table 1. Parameters on a confusion matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive	T_p	F_p
Predictive Negative	F_n	T_n

Table 1 lists the derived projected results of the confusion matrix. T pertains to the properly categorized normal occurrences; the value I show shows the number of anomalous events that are appropriately identified. Conversely, F denotes anomalous events mistakenly labeled as normal and F indicates normal events that have been mistakenly categorized as anomalies. Every deep learning classifier is extensively evaluated using several significant criteria: recall [24], accuracy [25], precision, and the F1-score [26].

Usually calculated based on the confusion matrix, these assessment measures are absolutely important for assessing the performance of binary classification algorithms.

Accuracy is given by

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

Precision is given by

$$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$

Recall is given by

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3}$$

F1-score is given by

$$F1\text{-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

According to the F1 score, which is sometimes referred to as the F-measure, the accuracy of a model is determined by a balance between recall and precision. One can find it by computing the proportion of true positive [27] forecasts versus the overall count of true and false positive outcomes.

Recall simultaneously divides the total number of actual positive samples by the number of true positives to properly identify all relevant positive cases. In the case of multi-class classification as investigated in this work, macro average and weighted average methods [28], [29] are used to combine the evaluation measures amongst all classes. The macro average, which gives equal weight to every class, computes the mean of every metric regardless of the overall case count in any class. This approach ensures that every class contributes equally to assess the overall performance. One can create the macro average mathematically as follows:

$$\text{Macro-F1} = \frac{1}{n}\sum_{i=1}^{n} F1_i \tag{5}$$

Unlike the macro average, the weighted average provides weights based on the number of instances per class thereby include the proportion of samples in every class. This method assures that classes in more samples contribute more importantly generally to the total determine. The weighted average is mathematically formulated as follows:

$$\text{Weighted-F1} = \sum_{i=1}^{n} \frac{N_i}{N_{\text{total}}} \times F1_i \tag{6}$$

Reviewing accuracy and loss graphs during the training period helped to guarantee the stability of the learning process and lower the overfitting risk. Usually, accuracy should increase as epochs go while loss should drop. Training and validation accuracy differ noticeably enough to indicate overfitting is occurring.

To further understand the model's classification process for making decisions from an interpretability perspective, a qualitative method was employed in addition to the quantitative evaluation. One of the methods applied for this goal is gradient-weighted Class Activation Mapping (Grad-CAM) [30], it is a visualizing tool stressing the parts of an input image regarded important by the model while making specific classification decisions. Grad-CAM relates the gradients of the target class in relation to the generated feature maps from the last convolutional layer. The Grad-CAM [31], [32] activation map is the last result of this process; one is able to obtain it through using the following equation :

$$L^{c}_{\text{Grad-CAM}} = \text{ReLU}(\sum_k \alpha_k^c A^k) \tag{7}$$

Let A_k denote the activation of the k^{th} feature map, and α_k^c represent the average gradient weight of the class score with respect to the feature map, mathematically formulated as follows:

$$\alpha_k^c = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \tag{8}$$

where y^c is the model output for class c and Z denotes the total number of elements in the feature map. The ReLU activation function is applied to emphasize features that positively contribute to the prediction outcome.

Through the resulting Grad-CAM activation maps, it becomes possible to visually identify specific regions of the retinal fundus images that are primarily attended to by the model during the classification process. These highlighted areas commonly correspond to key anatomical structures such as the optic disc, macula, or blood vessels, which are also critical regions in clinical diagnoses made by ophthalmology professionals.

RESULTS AND DISCUSSIONS

The performance comparison between the EfficientNetB3 and VGG16 models was evaluated using K-Fold cross-validation, focusing on mean accuracy and mean loss as primary indicators. As illustrated in Figure 5, EfficientNetB3 achieved a slightly higher mean accuracy compared to VGG16. Despite its slower training time (58 seconds per epoch compared to 41 seconds for VGG16), EfficientNetB3 was significantly more efficient in terms of model size, containing only 12 million parameters compared to VGG16's 138 million. This makes EfficientNetB3 better suited for deployment in mobile or embedded systems with limited resources, indicating its superior ability to correctly classify fundus images across various eye disorders. Meanwhile, VGG16 demonstrated a lower mean loss, suggesting fewer extreme prediction errors. Overall, the training results demonstrated that both models exhibited competitive performance. However, EfficientNetB3 consistently outperformed VGG16 in terms of accuracy and reliability, particularly in the multi-class classification of eye diseases with high visual similarity. These findings highlight each model's generalization capability and lay the groundwork for further evaluation of class-wise metrics and interpretability in the subsequent analysis.

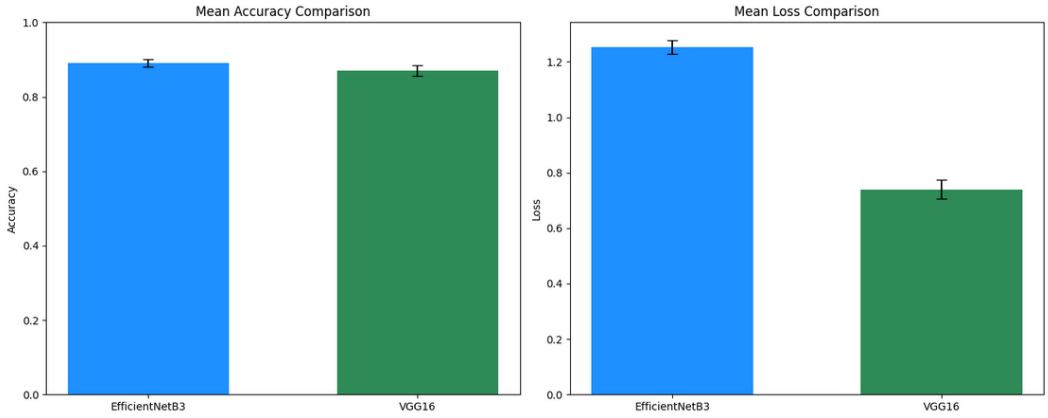


Figure 5. Cross-validation performance comparison

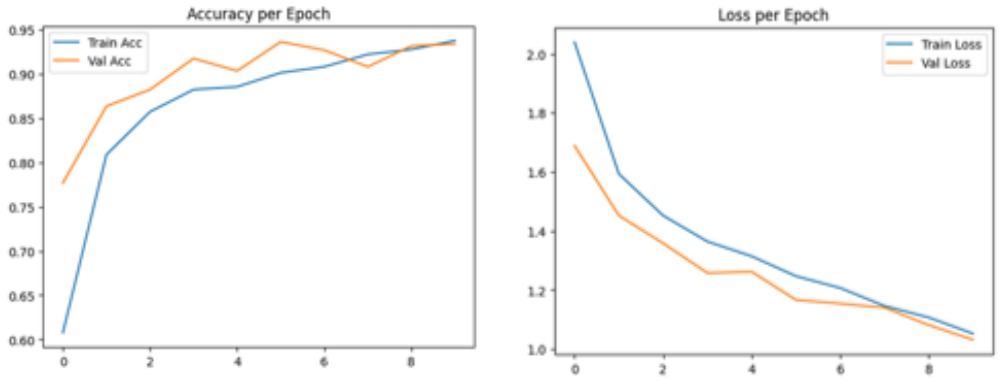
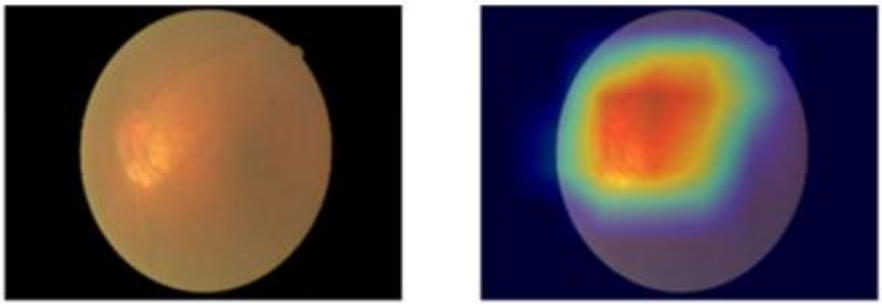


Figure 6. Model learning performance

Quantitative results revealed that the classification model based on the EfficientNetB3 architecture achieved high and stable performance. Overall, the model reached an accuracy of 93%, indicating a high proportion of correct predictions on the test data. Additionally, the precision, recall, and F1-score metrics

offered detailed insights into the performance of each class. Among the four classified categories of cataract, diabetic retinopathy, glaucoma, and normal. The diabetic retinopathy class achieved the highest performance with an F1-score of 0.98, followed by the cataract class with 0.95. Although the glaucoma class obtained the lowest F1-score of 0.89, this still reflects the model's good capability in recognizing its corresponding visual patterns. The normal class recorded an F1-score of 0.91. Furthermore, the macro average and weighted average values of all metrics indicated the model's consistency in handling both balanced and imbalanced class distributions, with an overall F1-score of 0.93.

Figure 6 illustrates the trend of increasing accuracy and decreasing loss during the model training process. The accuracies for both training and validation sets progressively increase, reaching near-optimal levels by epoch 10, indicating that the model is learning progressively. No significant difference was observed between training and validation accuracy, suggesting that the model did not experience overfitting. Similarly, the loss curves show a consistent downward trend for both datasets, indicating that the model successfully achieved optimal convergence.



(a) Original (b) Grad-Cam

Figure 7. Grad-CAM Visualization of Model Attention on each color fundus photography

Figure 7 presents a comparison between the original fundus image (a) and its corresponding Grad-CAM visualization (b). The red regions indicate areas that most strongly influenced the model's classification decision, while the blue regions represent areas with minimal contribution. In this example, the model concentrates its attention on the central region of the retina, likely encompassing the optic disc or macula, structures commonly associated with clinical indicators of ocular diseases. This visualization confirms that the model's predictions are not generated randomly but are based on clinically meaningful visual features, thereby reinforcing the interpretability of the model.

Table 2. Classification evaluation results by class

Class	Precision	Recall	F1-Score	Support
Cataract	0.94	0.97	0.95	105
Diabetic Retinopathy	0.98	0.97	0.98	111
Glaucoma	0.93	0.86	0.89	102
Normal	0.89	0.93	0.91	108
Macro Avg	0.93	0.93	0.93	426
Weighted Avg	0.93	0.93	0.93	426

The highest performance was achieved by the diabetic retinopathy class as shown in Table 2, with an F1-score of 0.98. In contrast, the glaucoma class had the lowest F1-score (0.89), though it remained within a reasonably high range. Precision and recall values were balanced across all classes, indicating that the model was unbiased toward specific classes and maintained stable classification performance

Figure 8, illustrates the performance of the proposed model in classifying four retinal fundus images conditions, namely cataract, diabetic retinopathy, glaucoma, and normal. The model demonstrates high accuracy in identifying diabetic retinopathy, with 108 out of 111 instances correctly classified. This result indicates the model's strong capability in detecting distinct pathological features such as hemorrhages, microaneurysms, and exudates. Similarly, cataract is accurately classified in 102 instances, with only minor misclassifications, suggesting the model's effectiveness in recognizing characteristic signs like lens opacity or reduced retinal clarity.

On the other hand, the classification of glaucoma presents some challenges. A total of 14 samples were misclassified, with seven labeled as cataract and seven as normal. This confusion is likely caused by overlapping visual patterns in the optic disc region or subtle changes in the retinal nerve fiber layer, which may not be distinctly captured. The normal class was classified correctly in 100 cases, although a small number of images were misinterpreted as having disease characteristics. In general, the model has strong cross-class generalizability with little bias. Confusion analysis shows that the model is accurate in identifying serious retinal abnormalities, but it also shows where the model may need some improvement, especially in glaucoma identification.

By providing the formulae for accuracy, precision, recall, and F1-score, we can ensure that the assessment criteria are legitimate. In order to evaluate the model's categorization performance objectively, some standard calculations are used.

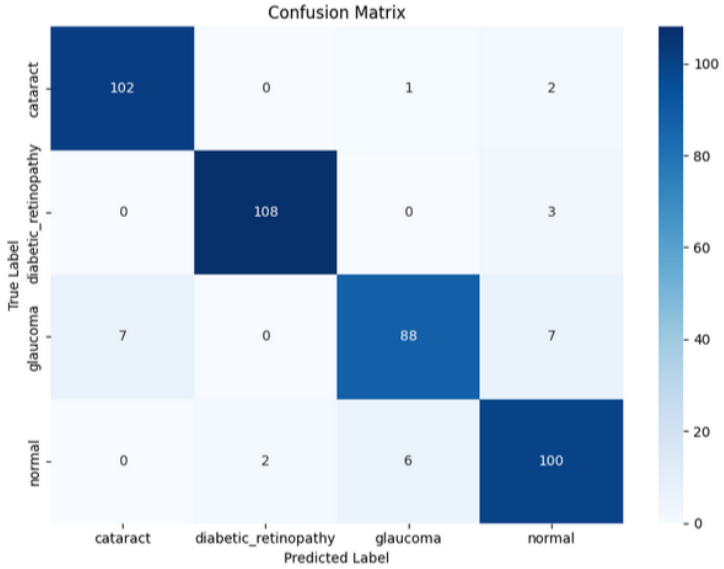


Figure 6. Confusion matrix efficientNetB3 model

CONCLUSION

This study conducted a thorough assessment of the EfficientNetB3 and VGG16 architectures for classifying retinal fundus images, focusing on both their prediction accuracy and the interpretability of the models. The EfficientNetB3 model consistently outperformed VGG16 in terms of accuracy and stability, achieving a 93% accuracy on the test dataset. Visualization using Grad-CAM confirmed the model's focus on clinically relevant regions such as the optic disc and macula, reinforcing its reliability in detecting key pathological features. The integration of stratified cross-validation, transfer learning, L2 regularization, and Grad-CAM provided a robust and explainable framework for medical image classification, with strong generalization across diabetic retinopathy, cataract, glaucoma, and normal classes.

From a medical standpoint, it is important to acknowledge that such analyses are only valid when based on clinically acquired fundus images, particularly those obtained through Color Fundus Photography (CFP). While the model shows potential for early screening, its outputs should not be considered final diagnostic results. The expertise of medical professionals, especially ophthalmologists, remains essential to validate and interpret findings comprehensively. AI-based predictions can augment but not replace clinical judgment. Therefore, the collaborative integration of AI tools and healthcare expertise is key to delivering safe, effective, and accountable medical diagnostics.

REFERENCES

[1] A. Alamri *et al.*, "Computer vision syndrome: Symptoms, risk factors, and practices," *J. Fam. Med. Prim. Care*, vol. 11, no. 9, pp. 5110–5115, Sep. 2022, doi: 10.4103/jfmpc.jfmpc_1627_21.

[2] F. Bahkir and S. Grandee, "Impact of the COVID-19 lockdown on digital device-related ocular health," *Indian J. Ophthalmol.*, vol. 68, no. 11, p. 2378, 2020, doi: 10.4103/ijo.IJO_2306_20.

[3] W. N. Ismail and H. A. Alsalamah, "A novel CatractNetDetect deep learning model for effective cataract classification through data fusion of fundus images," *Discov. Artif. Intell.*, vol. 4, no. 1, p. 54, Aug. 2024, doi: 10.1007/s44163-024-00155-y.

[4] B. Haj Najeeb, B. S. Gerendas, A. Montuoro, C. Simader, G. G. Deák, and U. M. Schmidt-Erfurth, "A Novel Effect of Microaneurysms and Retinal Cysts on Capillary Perfusion in Diabetic Macular Edema: A Multimodal Imaging Study," *J. Clin. Med.*, vol. 14, no. 9, p. 2985, Apr. 2025, doi: 10.3390/jcm14092985.

[5] V. Chaikitmongkol *et al.*, "Color Fundus Photography, Optical Coherence Tomography, and Fluorescein Angiography in Diagnosing Polypoidal Choroidal Vasculopathy," *Am. J. Ophthalmol.*, vol. 192, pp. 77–83, Aug. 2018, doi: 10.1016/j.ajo.2018.05.005.

[6] M. Odio-Herrera, G. Orozco-Loaiza, and L. Wu, "Gene Therapy in Diabetic Retinopathy and Diabetic Macular Edema: An Update," *J. Clin. Med.*, vol. 14, no. 9, p. 3205, May 2025, doi: 10.3390/jcm14093205.

[7] J. Guo, X. Li, W. Zhang, J. Zhong, and S. Liu, "Validation of Automatic Diabetic Retinopathy Screening and Diagnosis via Deep Neural Networks on Multi-modal Retinal Fundus Image Datasets," in *2023 International Annual Conference on Complex Systems and Intelligent Science (CSIS-IAC)*, IEEE, Oct. 2023, pp. 834–840. doi: 10.1109/CSIS-IAC60628.2023.10363900.

[8] M. F. Karim, M. A. Hossain, A. Y. Srizon, and N. Haque, "Early Detection of Glaucoma from Cropped Fundus Images Using Transfer-Learned Convolutional Neural Network," in *2024 IEEE International Conference on Power, Electrical, Electronics and Industrial Applications (PEEIACON)*, IEEE, Sep. 2024, pp. 1–6. doi: 10.1109/PEEIACON63629.2024.10800166.

[9] K. S. Gill, V. Anand, and R. Gupta, "Cataract Detection using optimized VGG19 Model by Transfer Learning perspective and its Social Benefits," in *2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, IEEE, Aug. 2023, pp. 593–596. doi: 10.1109/ICAISS58487.2023.10250513.

[10] P. Thanapol, K. Lavangnananda, P. Bouvry, F. Pinel, and F. Leprevost, "Reducing Overfitting and Improving Generalization in Training Convolutional Neural Network (CNN) under Limited Sample Sizes in Image Recognition," in *2020 - 5th International Conference on Information Technology (InCIT)*, IEEE, Oct. 2020, pp. 300–305. doi: 10.1109/InCIT50588.2020.9310787.

[11] C. Supriyanto *et al.*, "A Bibliometric Review of Deep Learning Approaches in Skin Cancer Research," Dec. 17, 2024. doi: 10.20944/preprints202412.1296.v1.

[12] E. R. Subhiyakto *et al.*, "Evaluation of Resampling Techniques in CNN-Based Heartbeat Classification," *Ingénierie des systèmes d Inf.*, vol. 29, no. 4, pp. 1323–1332, Aug. 2024, doi: 10.18280/isi.290408.

[13] S. Liu, W. Wang, L. Deng, and H. Xu, "Cnn-trans model: A parallel dual-branch network for fundus image classification," *Biomed. Signal Process. Control*, vol. 96, p. 106621, Oct. 2024, doi: 10.1016/j.bspc.2024.106621.

[14] N. Tawfik, H. M. Emara, W. El-Shafai, N. F. Soliman, A. D. Algarni, and F. E. A. El-Samie, "Enhancing Early Detection of Lung Cancer through Advanced Image Processing Techniques and Deep Learning Architectures for CT Scans," *Comput. Mater. Contin.*, vol. 81, no. 1, pp. 271–307, 2024, doi: 10.32604/cmc.2024.052404.

[15] A. S. Arnob, A. K. Kausik, Z. Islam, R. Khan, and A. Bin Rashid, "Comparative result analysis of cauliflower disease classification based on deep learning approach VGG16, inception v3, ResNet, and a custom CNN model," *Hybrid Adv.*, vol. 10, p. 100440, Sep. 2025, doi: 10.1016/j.hybadv.2025.100440.

[16] M. Liu, Y. Pei, M. Wu, and J. Wang, "Focal Cosine-Enhanced EfficientNetB0: A Novel Approach to Classifying Breast Histopathological Images," *Information*, vol. 16, no. 6, p. 444, May 2025, doi: 10.3390/info16060444.

[17] S. Park, S. Go, S. Kim, and J. Shim, "Deep Learning-Based Classification of Canine Cataracts from Ocular B-Mode Ultrasound Images," *Animals*, vol. 15, no. 9, p. 1327, May 2025, doi: 10.3390/ani15091327.

[18] A. Fitriatuzzahra, N. Rosmawarni, and P. Hamonangan Kinantan, "Implementation of Efficientnet-B0 CNN Architecture for Classification of Eye Diseases Based on Fundus Image: Normal, Cataract, Diabetic Retinopathy, and Glaucoma," in *2024 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, IEEE, Nov. 2024, pp. 543–546. doi: 10.1109/ICIMCIS63449.2024.10956479.

[19] R. B. Dixit and C. K. Jha, "Fundus image based diabetic retinopathy detection using EfficientNetB3 with squeeze and excitation block," *Med. Eng. Phys.*, vol. 140, p. 104350, Jun. 2025, doi:

10.1016/j.medengphy.2025.104350.
[20] C. Ma, L. Wu, and W. E, "A Qualitative Study of the Dynamic Behavior for Adaptive Gradient Algorithms," Sep. 2021, doi: https://doi.org/10.48550/arXiv.2009.06125.
[21] K. Pal and B. V. Patel, "Data Classification with k-fold Cross Validation and Holdout Accuracy Estimation Methods with 5 Different Machine Learning Techniques," in *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, IEEE, Mar. 2020, pp. 83–87. doi: 10.1109/ICCMC48092.2020.ICCMC-00016.
[22] A. A. Shujaaddeen, F. M. Ba-Alwi, A. T. Zahary, G. Al-Gaphari, A. M. Al-Badani, and A. Alsabry, "Enhancing a Random Forest Model Based on Single Rule Reduction for Tax Evasion Depends on the Values of K in K-Fold Validation Technique," in *2024 1st International Conference on Emerging Technologies for Dependable Internet of Things (ICETI)*, IEEE, Nov. 2024, pp. 1–9. doi: 10.1109/ICETI63946.2024.10777271.
[23] B. K. Triwijoyo, A. Adil, and M. Zulfikri, "Detection and classification of hypertensive retinopathy based on retinal image analysis using a deep learning approach," *Comput. Methods Programs Biomed. Updat.*, vol. 7, p. 100191, 2025, doi: 10.1016/j.cmpbup.2025.100191.
[24] A. Gupta, A. Anand, and Y. Hasija, "Recall-based Machine Learning approach for early detection of Cervical Cancer," in *2021 6th International Conference for Convergence in Technology (I2CT)*, IEEE, Apr. 2021, pp. 1–5. doi: 10.1109/I2CT51068.2021.9418099.
[25] Cinantya Paramita, Catur Supriyanto, Amalia, and Khalivio Rahmyanto Putra, "Comparative Analysis of YOLOv5 and YOLOv8 Cigarette Detection in Social Media Content," *Sci. J. Informatics*, vol. 11, no. 2, pp. 341–352, May 2024, doi: 10.15294/sji.v11i2.2808.
[26] M. Olek, "About Evaluation of F1 Score for RECENT Relation Extraction System," May 2023, doi: 10.48550/arXiv.2305.09410.
[27] B. J. Erickson and F. Kitamura, "Magician's Corner: 9. Performance Metrics for Machine Learning Models," *Radiol. Artif. Intell.*, vol. 3, no. 3, p. e200126, May 2021, doi: 10.1148/ryai.2021200126.
[28] M. C. Hinojosa Lee, J. Braet, and J. Springael, "Performance Metrics for Multilabel Emotion Classification: Comparing Micro, Macro, and Weighted F1-Scores," *Appl. Sci.*, vol. 14, no. 21, p. 9863, Oct. 2024, doi: 10.3390/app14219863.
[29] K. Takahashi, K. Yamamoto, A. Kuchiba, and T. Koyama, "Confidence interval for micro-averaged F1 and macro-averaged F1 scores," *Appl. Intell.*, vol. 52, no. 5, pp. 4961–4972, Mar. 2022, doi: 10.1007/s10489-021-02635-5.
[30] T. Ahad, H. B. Kibria, and M. Y. Mehemud, "MultiClass Classification of Chest Diseases using CXR Images with DenseNet201+CNN and Grad CAM Visualization," in *2024 IEEE International Conference on Power, Electrical, Electronics and Industrial Applications (PEEIACON)*, IEEE, Sep. 2024, pp. 368–372. doi: 10.1109/PEEIACON63629.2024.10800227.
[31] Y. Liang, M. Li, and C. Jiang, "Generating self-attention activation maps for visual interpretations of convolutional neural networks," *Neurocomputing*, vol. 490, pp. 206–216, Jun. 2022, doi: 10.1016/j.neucom.2021.11.084.
[32] S. F. Dipto and M. O. F. Goni, "Classification of X-Ray Images for the Automated Severity Grading of Knee Osteoarthritis by Ensemble Learning Through EfficientNet Architectures with Grad-CAM Visualization," in *2024 IEEE International Conference on Power, Electrical, Electronics and Industrial Applications (PEEIACON)*, IEEE, Sep. 2024, pp. 108–113. doi: 10.1109/PEEIACON63629.2024.10800349.