



## Multiclass Classification of Retinal Diseases from OCT Images Using ResNet50-Based Transfer Learning

Septia Dini Arifita<sup>\*1</sup>, Tanaya Anindita Irawan<sup>2</sup>, Ade Rizky Darmawan<sup>3</sup>, Aviolla Terza Damaliana<sup>4</sup>, Shindi Shella May Wara<sup>5</sup>

<sup>1,2,3,4,5</sup>Universitas Pembangunan Nasional, Jawa Timur  
[23083010040@student.upnjatim.ac.id](mailto:23083010040@student.upnjatim.ac.id)<sup>1\*</sup>

---

### Abstract

Retinal diseases are among the leading causes of visual impairment and blindness when not detected at an early stage. Advances in artificial intelligence, particularly deep learning, provide opportunities to support automated retinal disease diagnosis using Optical Coherence Tomography (OCT) images. This study aims to develop a multiclass retinal disease classification model using ResNet50-based transfer learning on the Retinal OCT C8 dataset, which consists of eight retinal condition categories: Age-related Macular Degeneration (AMD), Choroidal Neovascularization (CNV), Central Serous Retinopathy (CSR), Diabetic Macular Edema (DME), Diabetic Retinopathy (DR), Drusen, Macular Hole (MH), and Normal. The research stages included data preprocessing, image augmentation, transfer learning-based model training, and fine-tuning of the final network layers. Model performance was evaluated using accuracy, precision, recall, and F1-score metrics. Experimental results demonstrated that the proposed ResNet50 model achieved an accuracy of 93%, precision of 93%, recall of 93%, and F1-score of 93% on the testing dataset. These findings indicate that ResNet50 is effective in identifying multiple retinal diseases from OCT images. The proposed approach has potential applications in computer-aided diagnostic systems to assist clinicians in performing faster and more accurate retinal disease screening and early detection.

**Keywords:** *deep learning, multiclass classification, ResNet50, retinal disease, transfer learning*

---

### 1. Introduction

Retinal diseases are among the major causes of visual impairment and blindness worldwide. Various retinal disorders, including Age-related Macular Degeneration (AMD), Diabetic Retinopathy (DR), Diabetic Macular Edema (DME), Choroidal Neovascularization (CNV), and Drusen, can lead to permanent vision loss if not detected and treated at an early stage. According to global health reports, the prevalence of retinal disorders continues to increase due to population aging and the growing incidence of chronic diseases such as diabetes mellitus [1]. Therefore, accurate and timely diagnosis plays a crucial role in improving treatment outcomes for retinal diseases. One of the most widely used imaging technologies for retinal disease diagnosis is Optical Coherence Tomography (OCT). OCT provides high-resolution cross-sectional images of retinal structures, enabling ophthalmologists to identify various retinal abnormalities more effectively [2]. However, the interpretation of OCT images still relies heavily on the expertise of ophthalmologists and can be time-consuming, particularly when handling large numbers of patients. This challenge has motivated the development of computer-aided diagnostic systems capable of automatically and accurately classifying retinal diseases.

Recent advances in artificial intelligence, particularly deep learning, have demonstrated remarkable performance in medical image analysis tasks. Convolutional Neural Networks (CNNs) have become one of the most widely adopted approaches because they can automatically extract image features without requiring manual feature engineering [3]. Previous studies have successfully applied CNN-based methods for retinal disease classification using OCT images. Nevertheless, training CNN models from scratch often requires large-scale datasets and substantial computational resources. To overcome these limitations, transfer learning has emerged as an effective solution by leveraging knowledge learned from large pre-trained datasets [4]. Among various CNN architectures, ResNet50 is one of the most widely used models for transfer learning applications. ResNet50 employs residual learning mechanisms that effectively address the vanishing gradient problem in deep neural networks, resulting in more stable and efficient training processes. Previous studies have reported that ResNet50 can achieve high image classification performance across various domains, including medical imaging applications [5]. However, most existing studies focus on binary classification tasks or a limited number of disease categories, whereas retinal diseases encountered in clinical practice exhibit considerably greater diversity and complexity.

Based on the literature review, a research gap remains regarding the development of a robust multiclass classification model capable of simultaneously identifying multiple retinal diseases with high accuracy. Furthermore, studies utilizing multiclass retinal OCT datasets containing numerous disease categories remain relatively limited compared to binary classification approaches. Therefore, this study proposes a ResNet50-based transfer learning approach for multiclass retinal disease classification using the Retinal OCT C8 dataset, which consists of eight retinal condition categories: AMD, CNV, CSR, DME, DR, Drusen, MH, and Normal. This study aims to develop and evaluate a multiclass retinal disease classification model based on ResNet50 by utilizing transfer learning and fine-tuning techniques to improve the model's ability to recognize distinctive features of different retinal diseases. The findings are expected to contribute to the development of computer-aided diagnostic systems that can support clinicians in retinal disease screening and early detection processes more efficiently, accurately, and reliably. The research hypothesis states that applying ResNet50-based transfer learning to OCT images can achieve high classification performance in terms of accuracy, precision, recall, and F1-score, making it an effective approach for supporting retinal disease diagnosis.

## 2. Methodology

This study employed a quantitative experimental approach using deep learning techniques for multiclass retinal disease classification based on Optical Coherence Tomography (OCT) images. The classification model was developed using the ResNet50 architecture combined with transfer learning and fine-tuning strategies. The overall research workflow consisted of data collection, image preprocessing, data augmentation, model training, model testing, and performance evaluation. The research subjects consisted of retinal OCT images obtained from the Retinal OCT C8 dataset available on the Kaggle platform. The dataset was specifically developed for retinal disease classification studies and contains eight retinal condition categories: Age-related Macular Degeneration (AMD), Choroidal Neovascularization (CNV), Central Serous Retinopathy (CSR), Diabetic Macular Edema (DME), Diabetic Retinopathy (DR), Drusen, Macular Hole (MH), and Normal. These images were used as the primary data source for model training, validation, and testing.

### 2.1. Data Collection

The dataset used in this study was obtained from the Retinal OCT C8 dataset publicly available on Kaggle at <https://www.kaggle.com/datasets/obulisainaren/retinal-oct-c8>. The dataset consists of 24,000 retinal OCT images labeled into eight categories: Age-related Macular Degeneration (AMD), Choroidal Neovascularization (CNV), Central Serous Retinopathy (CSR), Diabetic Macular Edema (DME), Diabetic Retinopathy (DR), Drusen, Macular Hole (MH), and Normal. Each category contains an equal number of images, with 3,000 images per class. The dataset is divided into 2,300 images for training, 350 images for validation, and 350 images for testing in each category. Consequently, the complete dataset consists of 18,400 training images, 2,800 validation images, and 2,800 testing images. Before model training, all images underwent a preprocessing stage that included resizing the images to  $224 \times 224$  pixels, pixel value normalization using the ResNet50 preprocessing function, and image augmentation to increase data diversity and improve model generalization. The dataset was then organized into training, validation, and testing subsets to evaluate the classification performance and generalization capability of the proposed model.

### 2.2. Research Workflow

This study was conducted through several stages, including the collection of the Retinal OCT C8 dataset, image preprocessing, dataset splitting, development of a ResNet50-based transfer learning model, model training, fine-tuning, testing, and performance evaluation. These stages were designed to develop a model capable of accurately classifying eight categories of retinal diseases using Optical Coherence Tomography (OCT) images. The dataset consisted of OCT images that had been categorized into eight retinal disease classes. Transfer learning was employed to leverage knowledge from a pre-trained model, thereby improving classification performance and reducing training time. Model evaluation was carried out to assess the ability of the proposed model to correctly identify each retinal disease category. The overall research workflow is illustrated in Figure 1.

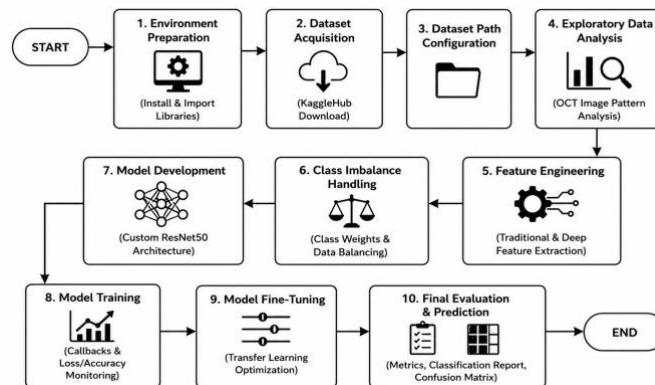


Fig. 1: Research Workflow for Multiclass Retinal Disease Classification Using ResNet50 Transfer Learning

Based on Figure 1, the research began with the collection of the Retinal OCT C8 dataset obtained from the Kaggle platform. The next stage involved image preprocessing, which included resizing all images to  $224 \times 224$  pixels, normalizing pixel values using the ResNet50 preprocessing function, and applying data augmentation techniques to increase sample diversity and reduce the risk of overfitting. After

preprocessing, the dataset was divided into training, validation, and testing sets. The subsequent stage involved developing a classification model based on the ResNet50 architecture using a transfer learning approach with pre-trained weights obtained from the ImageNet dataset. During the feature extraction stage, the convolutional base of ResNet50 was frozen, allowing only the newly added classification layers to be trained. Subsequently, a fine-tuning process was performed by unfreezing several of the final layers of the network to enable the model to learn OCT-specific image characteristics more effectively. The trained model was then evaluated using the testing dataset to generate predictions for each retinal disease category. Finally, model performance was assessed using accuracy, precision, recall, F1-score, and confusion matrix metrics to determine the effectiveness of the proposed approach in performing multiclass retinal disease classification.

### 2.3. ResNet50 Architecture

The classification model used in this study was ResNet50 (Residual Network with 50 Layers), which was pre-trained on the ImageNet dataset and implemented through a transfer learning approach. ResNet50 is a widely used Convolutional Neural Network (CNN) architecture that employs residual learning to address the vanishing gradient problem commonly encountered in very deep neural networks [6]. Through the use of residual or shortcut connections, information from previous layers can be propagated directly to deeper layers, resulting in more stable and efficient model training.

In this study, the convolutional base of ResNet50 was utilized as a feature extractor to capture important representations from Optical Coherence Tomography (OCT) images. The pre-trained ImageNet weights were retained as the initial parameters of the model, while the classification head was modified to accommodate the eight retinal disease categories included in the dataset. The customized classification head consisted of a Global Average Pooling layer, a fully connected Dense layer with ReLU activation, a Dropout layer to reduce overfitting, and an output layer with eight neurons using the Softmax activation function. The architecture of the proposed ResNet50 model is illustrated in Figure 2.

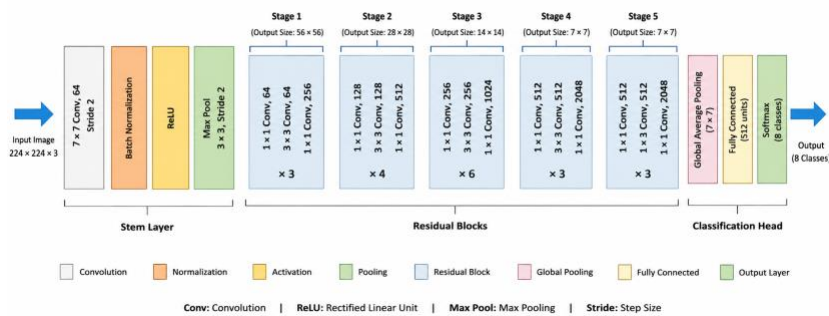


Fig. 2: ResNet50 Architecture for Retinal Disease Classification

As illustrated in Figure 2, OCT images with a resolution of  $224 \times 224 \times 3$  are used as the input to the proposed model. The images are processed through multiple convolutional layers and residual blocks within the ResNet50 architecture to extract discriminative features representing the characteristics of each retinal disease category. After feature extraction, the output of the convolutional base is forwarded to the classification head, which consists of a Global Average Pooling layer, a Dense layer with ReLU activation, a Dropout layer to reduce overfitting, and an output layer containing eight neurons corresponding to the eight retinal disease classes. Global Average Pooling is employed to reduce the dimensionality of the extracted feature maps without introducing a large number of additional parameters. The Dense layer is then used to learn high-level relationships among the extracted features. Furthermore, a Dropout layer is applied to improve model generalization and reduce the risk of overfitting during training[7]. Finally, the Softmax activation function is utilized in the output layer to generate class probabilities for multiclass retinal disease classification.

The probability generated by the Softmax function is calculated using Equation (1).

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}} \dots \dots \dots (1)$$

In Equation (1),  $P(y_i)$  represents the probability that an input image belongs to class (i), while  $(z_i)$  denotes the output value (logit) of neuron (i) in the output layer before normalization. Furthermore, (n) represents the total number of classes involved in the classification task. The Softmax function transforms the output logits into a probability distribution whose probabilities sum to one. The class with the highest probability value is selected as the final prediction of the model. In this study, the classification task involves eight retinal disease categories, namely Age-related Macular Degeneration (AMD), Choroidal Neovascularization (CNV), Central Serous Retinopathy (CSR), Diabetic Macular Edema (DME), Diabetic Retinopathy (DR), Drusen, Macular Hole (MH), and Normal.

### 2.4. Transfer Learning and Fine-Tuning

This study employs a transfer learning approach by utilizing a ResNet50 model pre-trained on the ImageNet dataset. Transfer learning enables the model to leverage knowledge acquired from a large-scale dataset, thereby reducing training time and improving classification performance compared with training a model from scratch[4]. During the initial stage, the convolutional base of ResNet50 was used as a feature extractor while retaining the pre-trained ImageNet weights. All layers within the feature extraction module were frozen, preventing

their parameters from being updated during training. Consequently, only the newly added classification layers were trained using the Retinal OCT C8 dataset. This stage, commonly referred to as feature extraction, allows the model to utilize generic visual features previously learned from a large and diverse image dataset. After the feature extraction stage achieved satisfactory convergence, a fine-tuning process was performed to enhance the model's ability to recognize disease-specific characteristics in Optical Coherence Tomography (OCT) images.

During this stage, several of the final layers of ResNet50 were unfrozen, allowing their weights to be updated through backpropagation. Fine-tuning enables the model to adapt previously learned feature representations to the unique characteristics of the eight retinal disease categories included in this study. Model training was conducted using the Adam optimizer and the categorical cross-entropy loss function, which is appropriate for multiclass classification problems. Validation data were utilized throughout the training process to monitor model performance and minimize the risk of overfitting. The best-performing model was selected based on the highest validation accuracy achieved during training. The combination of transfer learning and fine-tuning is expected to improve the model's generalization capability and enhance classification performance in identifying the eight retinal disease categories from OCT images.

## 2.5. Model Performance Evaluation

Model performance was evaluated using a confusion matrix and several classification metrics, including accuracy, precision, recall, and F1-score. The confusion matrix was employed to visualize the classification results by presenting the number of correct and incorrect predictions for each class. Through the confusion matrix, the ability of the model to distinguish among the retinal disease categories included in this study can be analyzed in greater detail [8]. Accuracy was used to measure the overall correctness of the classification model on the testing dataset. This metric represents the proportion of correctly classified samples relative to the total number of evaluated samples [9]. The accuracy metric is calculated using Equation (2).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \dots \dots \dots (2)$$

In addition to accuracy, this study employed precision to evaluate the correctness of positive predictions generated by the model. Precision measures the proportion of correctly predicted positive samples relative to the total number of samples predicted as positive [9]. A higher precision value indicates that the model produces fewer false positive predictions and is therefore more reliable when identifying a particular class. The precision metric is calculated using Equation (3).

$$Precision = \frac{TP}{TP + FP} \dots \dots \dots (3)$$

Recall was used to measure the model's ability to identify all actual positive samples within a particular class. A high recall value indicates that the model is capable of detecting most of the relevant samples and produces fewer false negative predictions [10]. Therefore, recall is particularly important in medical image classification tasks, where failing to identify a disease case may lead to serious consequences. The recall metric is calculated using Equation (4).

$$Recall = \frac{TP}{TP+FN} \dots \dots \dots (4)$$

Furthermore, the F1-score was used as a balanced evaluation metric that combines both precision and recall into a single measure. This metric is particularly important in multiclass classification problems because it provides a more comprehensive assessment of model performance by considering both false positive and false negative predictions simultaneously [11]. A high F1-score indicates that the model achieves a good balance between precision and recall across different classes. The F1-score metric is calculated using Equation (5).

$$F1-Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \dots \dots \dots (5)$$

In the above equations, (TP) (True Positive) represents the number of positive samples that are correctly classified, (TN) (True Negative) represents the number of negative samples that are correctly classified, (FP) (False Positive) denotes the number of negative samples incorrectly classified as positive, and (FN) (False Negative) denotes the number of positive samples incorrectly classified as negative. All experiments in this study were conducted using the TensorFlow and Keras frameworks. The evaluation results obtained from the testing dataset were subsequently analyzed to assess the capability of the proposed ResNet50 model in classifying eight categories of retinal diseases based on Optical Coherence Tomography (OCT) images.

## 3. Results and Discussion

After the training and fine-tuning processes were completed, the performance of the proposed model was analyzed using training accuracy, validation accuracy, training loss, and validation loss curves. These curves were used to monitor the learning behavior of the model throughout the training process, evaluate the convergence of the network, and identify potential overfitting or underfitting issues. The training and validation performance of the ResNet50 model is presented in Figure 3.

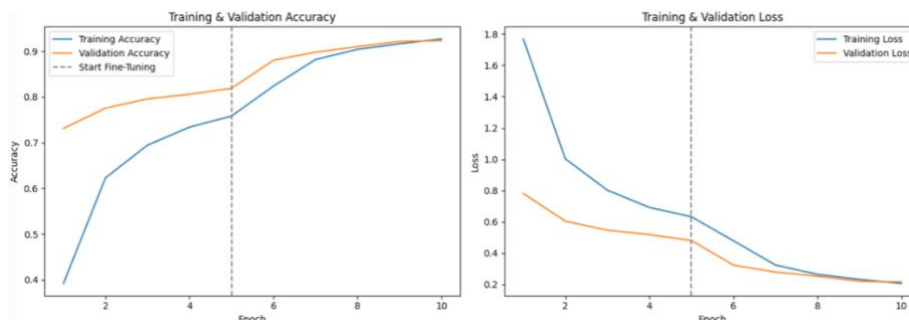


Fig. 3: Training and Validation Accuracy and Loss Curves of the ResNet50 Model

Based on Figure 3, both the training accuracy and validation accuracy exhibited a consistent upward trend throughout the training process. During the feature extraction stage, the model successfully increased the validation accuracy to 81.89%. After the fine-tuning process was initiated at epoch 6, the model performance improved significantly, ultimately achieving a validation accuracy of 92.32% at the end of training. These results indicate that the fine-tuning strategy effectively enhanced the model’s ability to learn OCT-specific retinal disease features. The loss curves also demonstrate a gradual decrease in both training loss and validation loss throughout the training process. The training loss decreased from approximately 1.77 to 0.21, while the validation loss decreased from 0.78 to 0.21. The relatively small gap between the training and validation curves suggests that the model achieved good generalization performance and did not exhibit significant signs of overfitting.

In addition to evaluation metrics such as accuracy, precision, recall, and F1-score, model performance was further analyzed using a confusion matrix. A confusion matrix provides a detailed representation of classification results by displaying the number of correct and incorrect predictions for each retinal disease category. Through this analysis, classification error patterns can be identified, including classes that are frequently misclassified as other categories. This evaluation is important because a high overall accuracy does not necessarily imply that all classes are classified equally well. Therefore, the confusion matrix serves as an effective tool for identifying classes that are easily recognized by the model as well as those that remain difficult to distinguish based on OCT image characteristics. The confusion matrix of the proposed ResNet50 model is presented in Figure 4.

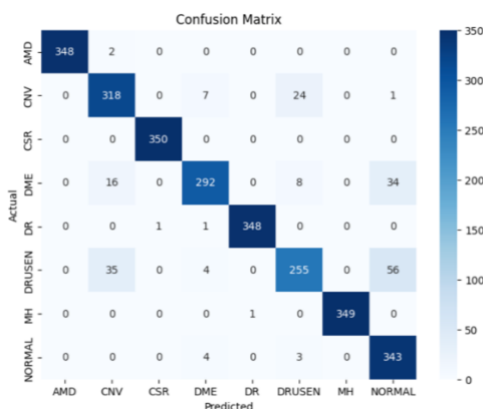


Fig. 4: Confusion Matrix of the ResNet50 Model on the Testing Dataset

Based on Figure 4, the majority of samples were correctly classified by the proposed model. The CSR class achieved perfect classification performance, with all 350 testing samples correctly identified. Similarly, the AMD, DR, and MH classes demonstrated excellent classification results, with 348, 348, and 349 correctly classified samples, respectively. These findings indicate that the ResNet50 model was highly effective in learning the distinctive characteristics of these retinal disease categories from OCT images. However, several misclassifications were observed in certain classes. The Drusen class exhibited the highest number of classification errors, with 56 samples incorrectly classified as NORMAL and 35 samples misclassified as CNV. In addition, the DME class also experienced several misclassifications, particularly to the NORMAL class (34 samples) and the CNV class (16 samples). These results suggest that some retinal disease categories share similar visual characteristics in OCT images, making them more challenging for the model to distinguish accurately. To provide a more detailed analysis of model performance across individual retinal disease categories, additional evaluations were conducted using precision, recall, and F1-score metrics for each class. Furthermore, the average confidence score was calculated to measure the model’s confidence in its predictions. The confidence score was obtained from the highest Softmax probability assigned to each predicted class. The classification performance and confidence scores of the ResNet50 model on the testing dataset are presented in Table 1.

**Table 1 :** Classification Performance and Confidence Scores of the ResNet50 Model on the Testing Dataset

Class	Precision	Recall	F1-Score	Confidence Score
AMD	1.00	0.99	1.00	99.45%
CNV	0.86	0.91	0.88	90.81%
CSR	1.00	1.00	1.00	99.88%
DME	0.95	0.83	0.89	90.71%
DR	1.00	0.99	1.00	97.45%
DRUSEN	0.88	0.73	0.80	82.41%
MH	1.00	1.00	1.00	99.03%
NORMAL	0.79	0.98	0.88	92.95%
<b>Macro Average</b>	<b>0.93</b>	<b>0.93</b>	<b>0.93</b>	<b>94.09%</b>

Based on Table 1, the ResNet50 model demonstrated excellent performance in classifying eight retinal disease categories using Optical Coherence Tomography (OCT) images. Overall, the model achieved a macro precision, macro recall, and macro F1-score of 0.93, indicating balanced classification performance across all classes. The CSR and MH classes achieved the best performance, with precision, recall, and F1-score values of 1.00. These results indicate that the model successfully captured the distinctive characteristics of these classes, resulting in nearly perfect classification performance. Similarly, the AMD and DR classes achieved F1-scores of 1.00, demonstrating the model's strong capability to distinguish these retinal diseases from other categories.

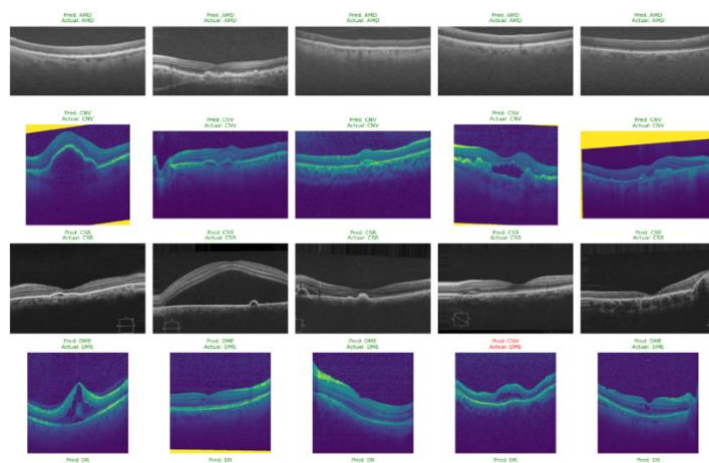
The CNV and DME classes also achieved satisfactory performance, with F1-scores of 0.88 and 0.89, respectively. However, several misclassifications were observed, particularly in the DME class, which resulted in a relatively lower recall value compared to other categories. Meanwhile, the NORMAL class achieved a high recall value of 0.98 but a lower precision value of 0.79. This finding suggests that some samples belonging to other disease categories were incorrectly classified as NORMAL by the model. The lowest performance was observed in the DRUSEN class, which achieved a recall value of 0.73 and an F1-score of 0.80. According to the confusion matrix, a considerable number of DRUSEN samples were misclassified as NORMAL and CNV. This phenomenon indicates that DRUSEN shares similar visual characteristics with these classes in OCT images. Clinically, DRUSEN is characterized by extracellular deposits beneath the retinal pigment epithelium (RPE), which may produce reflectivity patterns resembling normal retinal structures under certain conditions. Furthermore, some morphological features of DRUSEN overlap with retinal structural changes observed in CNV, making accurate discrimination more challenging. As a result, the model experienced greater difficulty in consistently identifying DRUSEN compared with the other retinal disease categories. Nevertheless, the precision value of 0.88 indicates that most samples predicted as DRUSEN were correctly classified. Therefore, the primary limitation lies in detecting all DRUSEN samples rather than in the correctness of DRUSEN predictions.

From the confidence score analysis, the model exhibited high prediction confidence across most classes, with an average confidence score of 94.09%. The CSR class achieved the highest confidence score of 99.88%, followed by AMD (99.45%) and MH (99.03%). In contrast, the DRUSEN class obtained the lowest confidence score of 82.41%, which is consistent with its lower recall and F1-score values. Overall, these results indicate that the proposed ResNet50 model not only achieves high classification accuracy but also provides reliable predictions with strong confidence across multiple retinal disease categories. In addition to the class-wise evaluation, the overall performance of the model was analyzed using accuracy, precision, recall, and F1-score metrics. This evaluation provides a general assessment of the model's ability to classify OCT images into eight retinal disease categories. Furthermore, the average confidence score was calculated to evaluate the model's confidence in its predictions. A summary of the overall performance of the ResNet50 model on the testing dataset is presented in Table 2.

**Table 2 :** Overall Performance of the ResNet50 Model

Metric	Value
Accuracy	93%
Precision (Macro Avg)	93%
Recall (Macro Avg)	93%
F1-Score (Macro Avg)	93%
Average Confidence Score	94.09%

Based on Table 2, the ResNet50 model achieved an overall accuracy of 93%, indicating that the majority of OCT images in the testing dataset were correctly classified. The precision, recall, and F1-score values also reached 93%, demonstrating that the model maintained a balanced performance across all retinal disease categories while effectively minimizing classification errors. Furthermore, the model achieved an average confidence score of 94.09%, indicating a high level of confidence in its predictions. This high confidence score suggests that the probabilities generated by the model were generally consistent and reliable across different classes. Overall, the evaluation results demonstrate that the combination of transfer learning and fine-tuning using ResNet50 provides strong classification performance for identifying eight retinal disease categories from Optical Coherence Tomography (OCT) images. To further illustrate the classification capability of the proposed model, prediction results were visualized using several representative OCT image samples from each retinal disease category. The visualization presents both the ground-truth labels and the predicted labels, allowing a direct comparison between the actual class and the model's prediction. The visualization results of the ResNet50 model are shown in Figure 5.



**Fig. 5:** Visualization of Prediction Results for Representative OCT Images Using the ResNet50 Model

Based on Figure 5, the majority of OCT image samples were correctly classified according to their ground-truth labels. This result indicates that the ResNet50 model was able to extract discriminative features that effectively distinguish among different retinal disease categories. Although several misclassifications were observed, particularly in classes with similar visual characteristics such as DRUSEN, NORMAL, and CNV, the overall prediction visualization demonstrates that the model produced consistent and reliable classifications for most testing samples. Furthermore, the confidence score analysis revealed a high level of prediction certainty across the majority of classes. The CSR class achieved the highest confidence score of 99.88%, followed by AMD (99.45%), MH (99.03%), and DR (97.45%). In contrast, the DRUSEN class obtained the lowest confidence score of 82.41%, which is consistent with its relatively lower recall and F1-score values. Nevertheless, all classes achieved confidence scores above 80%, indicating that the model maintained a satisfactory level of confidence when classifying the eight retinal disease categories from OCT images.

The experimental results demonstrate that the proposed ResNet50 model is capable of performing multiclass retinal disease classification with high effectiveness. The model achieved accuracy, precision, recall, and F1-score values of 93%, indicating balanced classification performance across all classes. These findings suggest that the transfer learning and fine-tuning strategies successfully enabled the model to learn relevant OCT image features associated with different retinal diseases. Based on the confusion matrix and classification report, most retinal disease categories were classified with excellent performance, particularly AMD, CSR, DR, and MH, which achieved F1-score values close to or equal to 1.00. However, the DRUSEN class remained more challenging to classify due to visual similarities with other retinal conditions. In addition, the average confidence score of 94.09% further confirms the reliability of the model's predictions. Overall, the results indicate that ResNet50 has strong potential for implementation in computer-aided diagnostic systems to support retinal disease screening and early detection using Optical Coherence Tomography (OCT) images.

## 4. Conclusion

This study successfully implemented transfer learning and fine-tuning techniques using the ResNet50 architecture for multiclass classification of eight retinal disease categories based on Optical Coherence Tomography (OCT) images. Experimental results demonstrated that the proposed model achieved an accuracy, precision, recall, and F1-score of 93%, along with an average confidence score of 94.09%. These findings indicate that ResNet50 is highly effective in extracting discriminative features and distinguishing among various retinal disease categories. The strengths of the proposed model are reflected in its excellent classification performance across most classes, particularly AMD, CSR, DR, and MH, which achieved near-perfect classification results. However, the model showed relatively lower performance in the DRUSEN class due to its visual similarities with other retinal conditions, particularly NORMAL and CNV. Despite this limitation, the overall results demonstrate that the combination of transfer learning and fine-tuning provides a robust approach for retinal disease classification using OCT images.

Future research may focus on utilizing larger and more diverse retinal OCT datasets, applying advanced data augmentation techniques, or comparing ResNet50 with more recent deep learning architectures to further improve classification performance. In addition, integrating explainable artificial intelligence (XAI) methods may provide better interpretability and increase the clinical applicability of automated retinal disease diagnosis systems.

## References

- [1] S. V. Viraktamath, K. B. Kadakuntla, and S. Begumpur, "Detection of Diabetic Maculopathy," in *Proceedings of the International Conference on Artificial Intelligence and Smart Computing (ICAISC)*, IEEE, 2023. doi: 10.1109/ICAISC58445.2023.10199599.
- [2] N. Eladawi *et al.*, "Optical Coherence Tomography: A Review," in *Artificial Intelligence in Retina*, Academic Press, 2020, pp. 191–221. doi: 10.1016/B978-0-12-817440-1.00007-3.
- [3] A. Awasthi, S. B. Rao, and K. Acharjya, "Efficient Image Feature Extraction Using Convolutional Neural Networks," in *Proceedings of the International Conference on Networking and Advanced Technologies (ICoNAT)*, IEEE, 2024, pp. 1–5. doi: 10.1109/ICoNAT61936.2024.10775088.

- [4] P. Dey and P. Dey, "Transfer Learning: Leveraging Pre-trained Knowledge for Efficient Models in Deep Learning," in *Advances in Artificial Intelligence and Machine Learning*, Springer, 2025, pp. 251–261. doi: 10.1007/978-981-95-3272-8\_17.
- [5] A. Puspitasari, D. S. Salsabila, and D. Roliawati, "Application of ResNet50-CNN for Classification Optimization on Fashion Datasets," *Indones. J. Data Sci.*, vol. 3, no. 1, pp. 1–12, 2025, doi: 10.30989/ijds.v3i1.1533.
- [6] L. Xingyu, "ResNet: Enabling Deep Convolutional Neural Networks Through Residual Learning," *ArXiv Prepr.*, 2025, doi: 10.48550/arXiv.2510.24036.
- [7] H. Çetiner and S. Metlek, "Analysis of Different Pooling Functions on a Convolutional Neural Network-Based Model," *Int. J. 3D Print. Technol. Digit. Ind.*, vol. 8, no. 2, pp. 266–276, 2024, doi: 10.46519/ij3dptdi.1484354.
- [8] I. L. Kharisma, S. Susanti, R. Rustiani, R. A. Pratama, and K. Kamdan, "Implementation of the VGG19 Model with Transfer Learning for Retinal Disease Diagnosis: A Study on Normal Eyes, Diabetic Retinopathy, Cataract, and Glaucoma Datasets," *Eng. Proc.*, vol. 107, no. 1, p. 111, 2025, doi: 10.3390/engproc2025107111.
- [9] Y. Xia and J. Sun, "Basic Performance Metrics for Machine Learning Models," in *Machine Learning Fundamentals and Applications*, CRC Press, 2026, pp. 485–496. doi: 10.1201/9781003610281-15.
- [10] A. Shaykholova, "Model Performance on Domain Test Set," 2026, doi: 10.60797/irj.2026.163.75.3.
- [11] S. Simhadri, M. Ponnamp, R. Rajitha, and R. Balamurugan, "Enhanced Multi-Class Model Evaluation: Analyzing BERT, GPT-2, and LLaMA with Precision, Recall, and F1-Score Metrics," in *Proceedings of the International Conference on Intelligent Machine Intelligence and Applications (ICIMIA)*, IEEE, 2025, pp. 984–989. doi: 10.1109/ICIMIA67127.2025.11200914.