Diagnosis System for Fundus Diseases Based on Deep Learning

By:

Lama Ali - 2112125
Raha Sahly - 2006904
Ghadi Jamal - 2112660
Elaf Almalki - 2111732
Shahad Alzhrani - 2112039
Raneem Alsulami - 2005845

Final Report

Computer Science and Artificial Intelligence Department
College of Computer Science and Engineering
University of Jeddah, SAUDI ARABIA

Supervisor
Dr.Safa Alsfry

(November 2024)

1 Introduction

The human eye is a vital and complex organ, responsible for vision. Human visual perception makes up 85% of total information intake[1], highlighting the crucial role of vision in daily activities. According to the World Health Organization (WHO), more than 2.2 billion individuals worldwide suffer from ocular diseases such as glaucoma, diabetic retinopathy (DR), age-related macular degeneration (AMR) and cataracts, with approximately 1 billion of these cases ultimately leading to blindness[2]. Despite the availability of effective treatments, a large proportion of these diseases remain undiagnosed, especially in areas with poor access to specialized medical care.

In clinical settings, early detection is critical to preventing vision loss. Two of the primary diagnostic tools are color fundus photography (CFP), which involves photographing the back of the eye (the fundus) to identify ocular disorders, and Optical Coherence Tomography (OCT), which creates cross-sectional images of the eyes by measuring wavelengths of infrared light reflected off the retina, allowing for the detection of structural changes inside the eye. However, both tools face challenges, including the percentage of expert doctors is low in small areas with limited capabilities, in addition to the lack of advanced equipment to enable them to obtain accurate diagnoses.

In response to these challenges, there is a growing interest in developing automated systems for diagnosing eye diseases using deep learning techniques[3]. By leveraging deep learning and computer vision, it is possible to build intelligent systems that can analyze large sets of both fundus and OCT images with high accuracy. Such systems can help in the early detection of various fundus diseases in countries with limited resources.

The motivation behind this project is to develop and implement an intelligent system that diagnoses fundus diseases by classifying various conditions in fundus and OCT images using deep learning techniques. This approach aims to improve early detection and treatment of fundus conditions, thereby preserving vision, slowing disease progression, and enhancing access to timely care, particularly in regions with limited medical resources.

2 Problem definition

This study aims to enhance the early detection of fundus diseases to manage them effectively and stop them from progressing. However, diagnosing these conditions using fundus and OCT images may represent a challenge for healthcare providers, especially those in resource-limited areas. Many fundus diseases may not show clear signs in their early stages, making it hard to get an accurate diagnosis[4]. Effective early diagnosis often requires specialist doctors who can analyze fundus images correctly, which may not be accessible in resource-limited areas. Therefore, there is a need for artificial intelligence tools to assist in providing timely and accurate diagnoses of various fundus diseases[5]. Timely detection of these diseases can prevent them from progressing to serious stages, which might result in losing all vision. To address this challenge, diagnostic tools are crucial to speed up and simplify the diagnostic process for medical professionals. This approach will ensure that patients receive prompt care while reducing the risk of long-term visual impairment.

3 Aim of the project

The objective of this study is to improve the early detection of fundus diseases and decrease misdiagnosis rates by ensuring that the system's performance meets or exceeds that of human experts. By employing deep learning methods [6], this project aims to improve overall health care outcomes and provide access to accurate, specialist eye care for people living in poor areas where eye diseases are often diagnosed by general practitioners rather than specialist ophthalmologists. As a result, it will support faster and more precise detection of fundus diseases.

4 Objectives

- **Data collection and preparation:** Integrate the CFP image dataset with the OCT images obtained from the optical coherence tomography (OCT) device. This merger enhances the comprehensiveness of the system, enabling compatibility with various fundus examination devices.

- **Develop fundus disease detection system:** Diagnose fundus diseases such as diabetes, glaucoma, cataracts, and others by developing deep learning algorithms that can capture and detect the disease in fundus and OCT images.

- **Build a personalized treatment recommendation system:** Implement a knowledge-based system that recommends suitable treatments for each diagnosed case based on disease severity and patient history.

- **Real-time integration with diagnostic devices:** Integrate the system with diagnostic devices capable of capturing fundus images in real-time, allowing for immediate diagnosis and facilitating faster medical responses.

5 Literature Review

Artificial intelligence (AI) is becoming increasingly important in ophthalmology, particularly in the diagnosis of fundus diseases [7]. In recent years, deep learning techniques, especially convolutional neural networks (CNNs), have gained attention due to their ability to analyze complex retinal images with high accuracy. AI-based tools are becoming more effective at diagnosing conditions such as diabetic retinopathy (DR), age-related macular degeneration (AMD), and retinal vein occlusions.

AI's role extends beyond retinal diseases, as it is also improving in the detection of other serious conditions, such as glaucoma, cataracts, and viral conjunctivitis. All these improvements are used in the applications of pediatric ophthalmology. Conditions like myopia, strabismus, and ptosis are growing concerns. Early detection is critical in preventing long-term visual damage; AI offers the opportunity of early and reliable screening, even at home, providing a significant advantage over traditional methods that are often limited by cost and accessibility.

The application of AI, especially deep learning models like ResNet, VGG, and DenseNet, has provided new opportunities for increasing the accuracy of the diagnosis. These models show success in interpreting retinal images, enhancing early detection, and aiding clinical decision-making. In particular, automated systems such as the FDA-approved IDx-DR [8] for diabetic retinopathy highlight how AI can reduce the burden on healthcare professionals while improving outcomes for patients.

This literature review shows the growth of AI in ophthalmology, focusing on its potential to transform the diagnosis and treatment of various fundus diseases. Despite challenges related to data quality, algorithm transparency, and generalization [9], the advancements in AI give an overview into the future with more efficient, affordable, and effective eye care.

5.1 Datasets

Table 1: Summary of Datasets Used in the Papers

Dataset Name	**Size**	**Public/Private**	**Year Collected**	**Device Used**	**Disease Types**	**Source**
APTOS-2019	3,662 images	Public	2019	Smartphone and fundus camera	Diabetic retinopathy, diabetic macular edema	Kaggle (APTOS 2019 competition)
Kaggle Diabetic Retinopathy Dataset	88,702 images	Public	2019	Smartphone and fundus camera	Diabetic retinopathy	Kaggle (APTOS 2019 competition)
IDRiD	81 images	Public	2018	Fundus camera	Diabetic retinopathy, diabetic macular edema	Unknown
DRISHTI-GS	1,000 images	Public	2018	Fundus camera	Glaucoma	Unknown
Messidor	1,200 images	Public	2009	Fundus camera	Diabetic retinopathy	Unknown
Retina Dataset	35,000 images	Public	2019	Fundus camera	Cataracts	Unknown
Kaggle Diabetic Retinopathy Dataset	88,702 images	Public	2015	Not specified	Focuses on diabetic retinopathy with five grades: no DR, mild, moderate, severe, and proliferative DR	Collected as part of Kaggle's Diabetic Retinopathy competition
Two Public OCT Datasets (Duke University and Noor Eye Hospital)	12,649 images	Public	Unknown	Unknown	Includes Age-related Macular Degeneration (AMD), Diabetic Macular Edema (DME), and Central Serous Chorioretinopathy (CSR)	Collected from Duke University and Noor Eye Hospital in Tehran
Duke Dataset and Clinical Dataset from Beijing Hospital	110 images	Public	Unknown	Unknown	Diabetic Macular Edema (DME)	Collected from Duke University and Beijing Hospital
Public and Private Retinal Fundus Image Datasets	40 images	Public	2004	Canon CR5 non-mydriatic camera with a 45° field of view	Focuses on diabetic retinopathy	Collected from a diabetic retinopathy screening program in the Netherlands

*5.2 **Methodology***

For the OCT dataset, researchers have explored various methods to process the data while preserving its critical details and quality, as these are essential for accurate diagnosis of eye diseases. These efforts aim to enhance the precision and reliability of models used for retinal disease classification. In one of the studies[10] the focus was on preprocessing techniques like normalization, standardization, rescaling, and converting images from BGR to RGB formats. Leveraging the ResNet-50 model with these methods, the study concluded that rescaling was particularly effective for diagnosing retinal diseases. While most preprocessing techniques improved overall accuracy, the precision metric did not consistently show similar improvements. But another researcher[11] proposed a distinct approach that used the FCN method to segment retinal regions. This method involved steps such as flattening, cropping, and removing background pixels by setting their brightness to zero, which effectively eliminated the need for prior image denoising. The researchers compared the FCN method with SD and RPE techniques, with FCN delivering the highest segmentation accuracy, making it the preferred method. another researcher[12] adopted a different strategy for retinal disease classification by applying preprocessing techniques like denoising, normalization, cropping, and brightness adjustment to refine image quality. The researchers utilized two OCT datasets of varying quality to ensure robust and reliable results. They tested various models, including SVM, GoogLeNet, VGG-16, and multi-scale convolutional networks. CliqueNet was uniquely optimized without using a pre-trained model, while other models like ResNet50, DenseNet121, DPN92, and ResNext101 employed transfer learning with tailored modifications to boost their performance. Each model was fine-tuned with specific learning rates and dropout configurations, achieving high accuracy in disease classification. Among all tested models, CliqueNet emerged as the top performer across both datasets, and the study strongly recommends its use for medical image classification tasks. This collection of studies reflects the diverse and innovative approaches taken to optimize preprocessing and classification techniques for retinal disease diagnosis, providing valuable insights for future research and clinical applications.

As for CFP fundus images, many studies[13] have focused on enhancing the accuracy and reliability of systems that detect retinal diseases. These efforts focus on improving retinal image quality using preprocessing techniques and utilizing state-of-the-art algorithms to more precisely classify diseases. Researchers used CLAHE, a method for enhancing image contrast, along with lighting correction. These methods enhanced the sharpness of the photographs of important retinal components, including the blood vessels and optic nerve. ResNet-50, VGG-16, Xception, and a custom-built model were among the models they tested. The results showed that accuracy was increased by combining these picture improvements with deep learning models.Another study[14] employed a different approach. The strategy employed transfer learning where the pre-trained model known as Xception was modified to determine whether the referred retinal images were normal or showed signs of age related macular degeneration. Rescaling of images was done by the researchers together with eliminating extraneous features so as to normalize the dataset comprising images from different sources. Compared to models such as VGG-16 or ResNet-50 Xception achieved better performance and validation accuracy. Challenges were experienced in this study as well as because the images were taken in different conditions.another research[15] employed the classical image processing techniques for the purposes of recognition of the sick retinas. For this purposes, such techniques as the adaptive histogram equalization (for contrast enhancing) and median filtering (de-noising) were used. These methods facilitated the recognition of such features as the optic disks, blood vessels and exudates associated with degenerative changes in the retina.

Table 2: Summary of Deep Learning Approaches for Fundus Disease Detection

Title	Year	Dataset	Model	Accuracy	Limitations
Development of a Fundus Image-Based Deep Learning Diagnostic Tool for Various Retinal Diseases	2023	Fundus images from Ulsan University Hospital	CNN (ResNet50, VGG19, Inception v3)	Accuracy: 94.7%, Sensitivity: 93.5%, Specificity: 95.2%, AUC-ROC: 0.96	Age distribution imbalances and resolution differences in fundus photos
Automated Detection of Mild and Multi-Class Diabetic Eye Diseases Using Deep Learning	2020	Messidor, DRISHTI-GS, Retina Dataset	CNNs (VGG16, InceptionV3)	VGG16: 88.3% accuracy (multi-class), 85.95% (mild)	Limitations in detecting early-stage features
Detection of Diabetic Eye Disease from Retinal Images Using a Deep Learning Based CenterNet Model	2020	APTOS-2019, IDRiD, EYEPACS, Diaretdb1	CenterNet with DenseNet-100	97.93% accuracy (APTOS-2019), 98.10% (IDRiD)	Excels in lesion localization, challenges with lighting
Detection of Diabetic Retinopathy Using Deep Learning Methodology	2020	Diabetic Retinopathy Detection 2015, APTOS 2019 (Kaggle)	DenseNet-169, Adam Optimization	Training Accuracy: 0.953, Validation Accuracy: 0.9034	Emphasizes need for multiple images for better classification
Combining Artificial Intelligence and Image Processing for Diagnosing Diabetic Retinopathy	2023	Kaggle Diabetic Retinopathy dataset (88,702 images)	DCNN, CPFCM for segmentation	Accuracy: 98.80%, Sensitivity: 99.20%, Specificity: 97.00%	Performance influenced by dataset quality and size
Computer Vision for Eye Diseases Detection Using Pre-trained Deep Learning Techniques and Raspberry Pi	2023	Hospital eye images (details unspecified)	NASNetMobile, VGG19, ResNet50, InceptionV3, MobileNet	Challenges in lighting and artifacts	Limited ability to assess disease severity; generalizability issues
On OCT Image Classification via Deep Learning	2019	Two public OCT datasets from Duke University and Noor Eye Hospital	CliqueNet	Accuracy: Over 0.98	The study struggles with dataset size variability and image quality inconsistencies across public datasets
Automatic Detection of Retinal Regions Using Fully Convolutional Networks	2019	Duke Dataset and a clinical dataset from Beijing Hospital	Fully Convolutional Networks (FCN)	Accuracy: 99.69% (Inception V3 classifier on clinic dataset)	The proposed method may underperform on datasets with severe image distortions or noise due to reliance on segmentation accuracy
Age-Related Macular Degeneration Detection in Retinal Fundus Images by a Deep Convolutional Neural Network	2024	Public and private retinal fundus image datasets, including DRIVE	Xception Convolutional Neural Network (CNN) with transfer learning	Accuracy: Over 80%	Requires high-quality labeled datasets and is resource-intensive on GPUs
Diagnosis of Diabetic Retinopathy	2022	Retinal fundus images (48 images tested)	MATLAB-based algorithm using morphological operations and manual processing	Accuracy: High accuracy for detecting exudates (specific metric not given)	Limitations: Limited to small datasets and dependent on manual pre-processing and segmentation

6 Methodology

This study presents a comprehensive approach to diagnosing ocular diseases using deep learning methodologies. By leveraging two distinct datasets derived from CFP (Color Fundus Photography) and OCT (Optical Coherence Tomography) imaging techniques, we aim to develop a robust diagnostic system capable of classifying fundus images into multiple categories. The methodology is designed to address the challenges posed by class imbalance, image quality variations, and the need for compatibility with diverse imaging modalities. Our approach integrates advanced preprocessing pipelines, a pre-trained VGG19 architecture for feature extraction, and fine-tuning techniques to achieve accurate multi-class classification.

6.1 Data collection

We used two datasets to train the model. The first is CFP images of the fundus found from Kaggle[16]. This data consists of 4 classes (glaucoma, diabetic retinopathy, and cataract) and images of the fundus of healthy eyes. Brief Overview of the Diseases:

CNV (Choroidal Neovascularization): Abnormal growth of blood vessels under the retina, often associated with Age-related Macular Degeneration (AMD). It can cause vision loss.

DME (Diabetic Macular Edema): Swelling in the macula caused by leaking blood vessels in diabetic retinopathy, leading to blurred or distorted vision.
Drusen: Yellow deposits of fat and proteins that accumulate under the retina, often linked to Age-related Macular Degeneration (AMD). They may or may not affect vision.
Glaucoma: A group of eye diseases that damage the optic nerve, often caused by high intraocular pressure. It can lead to vision loss if untreated.
Diabetic Retinopathy: A complication of diabetes that affects the blood vessels in the retina, potentially causing vision impairment or blindness. Cataract: Clouding of the lens in the eye leads to blurred vision and difficulty seeing. It is commonly associated with aging.

The second set contains images coming from the fundus' OCT device. We used the "Large Dataset of Labeled Optical Coherence Tomography (OCT) and Chest X-ray Images" dataset, publicly available on the Mendeley platform [17].
A Brief Overview of the Diseases:

CNV (Choroidal Neovascularization): Abnormal growth of blood vessels under the retina is often associated with Age-related Macular Degeneration (AMD). It could lead to vision loss.

DME (Diabetic Macular Edema): swelling in the macula created by blood vessels leaking in the diabetic retinopathy leading to blurred vision.

Drusen: Yellow deposits of proteins and fats that grow under the retina are often connected with Age-related Macular degeneration (AMD). They could affect vision.

Normal: This class contains healthy, not affected retinal images without any abnormal growths or swelling.

The dataset comprises two folders: a training set and a testing set, and each subfolder contains NORMAL, CNV, DME, and DRUSEN retinal OCT images. The total number of images in the dataset was originally 83,484 but the classes contained duplicates that were removed because they hold no value in terms of training. The final total number of images became 76,497 after duplication removal. The distribution of images across the training and testing sets is shown in Table 3.

Class	**Training**	**Testing**	**Total**
DME	10,929	250	11,179
CNV	31,613	250	31,863
DRUSEN	7,824	250	8,074
NORMAL	26,131	250	26,381

Table 3: Distribution of images across training and testing sets.

The dataset shows a major class imbalance, with more CNV and NORMAL images compared to DME and DRUSEN classes. This imbalance affects the training process, potentially leading to overfitting on the majority classes and reduced accuracy on minority classes during testing.

We used two sets of data from two different types of fundus scanning devices to make the model more comprehensive, enabling it to be used for more than one type of fundus scanning device. As for the third group, it is the Knowledge Dataset. We plan to have it collected personally by a specialist ophthalmologist, and it is what we will use to suggest appropriate treatments for patients.

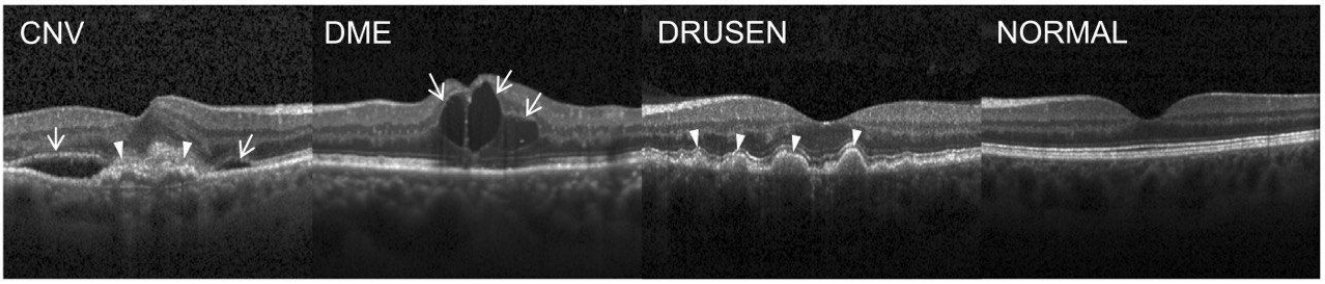


Figure 1: Representative Optical Coherence Tomography Images [18]

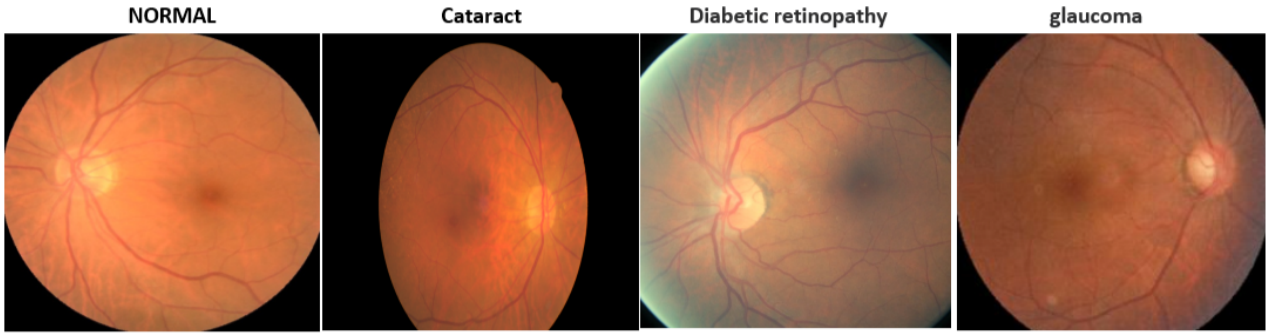


Figure 2: color fundus photography Images

6.2 Data Preprocessing

6.2.1 Preprocessing for CFP Dataset

The preprocessing of CFP (Color Fundus Photography) images was designed to enhance image quality and ensure compatibility with deep learning models. The following steps were applied:

6.3.1.1 Grayscale Conversion CFP images were converted from RGB to grayscale to simplify data representation and reduce computational complexity. This step preserved structural information critical for diagnosis, such as blood vessels and lesions, while removing color information that may introduce noise.

6.3.1.2 Contrast Enhancement Using CLAHE Contrast Limited Adaptive Histogram Equalization (CLAHE) was used to improve local contrast in grayscale images. CLAHE adjusts the intensity histogram of localized regions, enhancing visibility in areas with low contrast. The following parameter combinations were evaluated during preprocessing:

clipLimit values: 2.0, 3.0, 5.0 tileGridSize values: (8, 8), (16, 16), (32, 32).

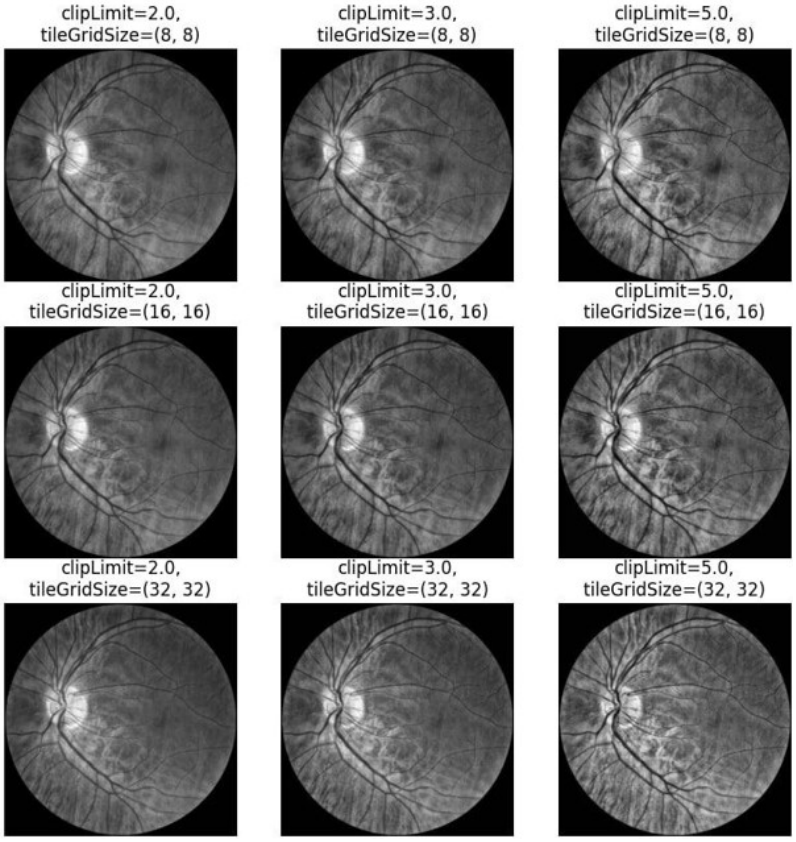


Figure 3: Effect of CLAHE parameter variations on a retinal image contrast enhancement.

6.3.1.3 Global Intensity Normalization Histogram equalization was utilized as a global image normalization technique to adjust pixel intensity values across the entire image. This method ensures that the overall brightness is distributed uniformly, which enhances visibility in underexposed or overly dark regions of the image. By redistributing the pixel intensities, the preprocessing step improved image clarity, making diagnostically relevant features more distinguishable across varying imaging conditions.

6.3.1.4 Preprocessing Pipelines Based on the techniques above, two preprocessing pipelines were developed for evaluation:

Pipeline 1: Grayscale Conversion → CLAHE → Model Training
Pipeline 2: Grayscale Conversion → CLAHE → Histogram Equalization → Model Training
These pipelines were designed to assess the combined effects of local and global intensity adjustments on the dataset.

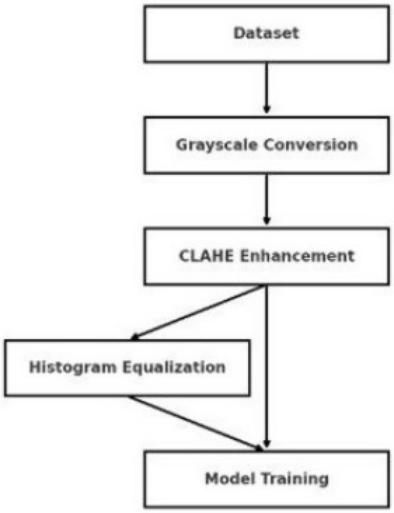


Figure 4: CFP preprocessing pipelines

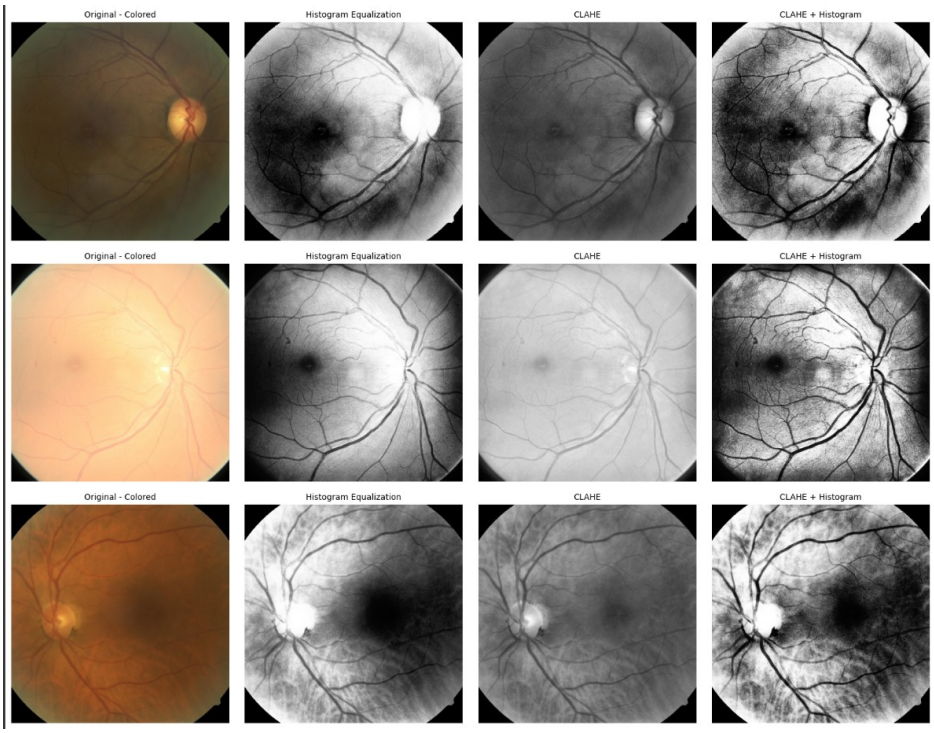


Figure 5: The impact of different preprocessing techniques on 3 images: Column 1 presents the original images, Column 2 normalized images, Column 3 Pipeline 1, and Column 4 Pipeline 2

6.2.2 Preprocessing for OCT Dataset

For the OCT images, the following preprocessing steps were performed:

6.3.2.1 White Border Removal After evaluating the images in the dataset, was observed that many images contained avoidable white borders. A dynamic cropping method was implemented to remove these borders by identifying and cropping

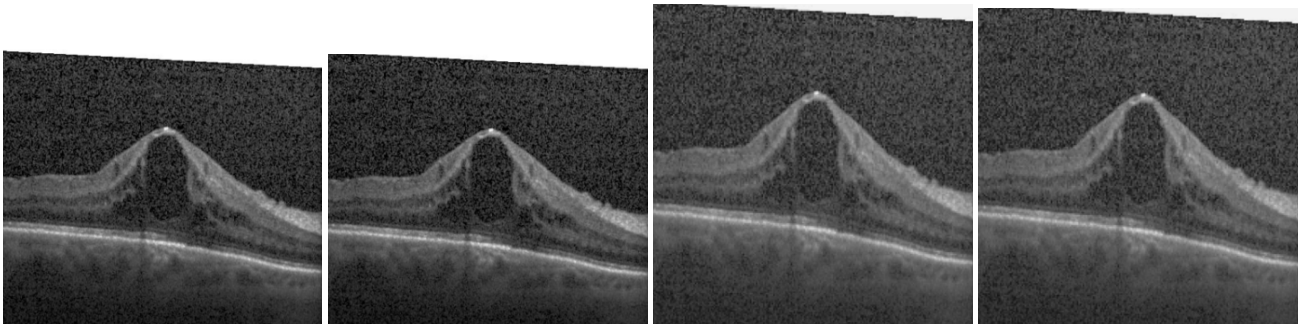
the regions of interest while ensuring consistency across the images.

6.3.2.2 Resizing All images were resized to a fixed dimension of 224×224 pixels using bicubic interpolation. This assures orderly input size for the deep learning model (VGG19).

6.3.2.3 Normalization Pixel intensity values of all images were rescaled to the range $[0, 1]$ using a scaling factor of $1/255$. This scaling ensured numerical balance and enhanced the convergence speed of the training process.

6.3.2.4 Conversion to RGB Grayscale images were converted to three channels, ensuring compatibility with the VGG19 model.

6.3.2.5 Batching and Prefetching The preprocessed dataset was divided into batches of size 16. Prefetching was also applied to optimize data loading during the training process and to enhance computational efficiency. To visualize the preprocessing steps, we applied the same pipeline to a sample image. Figure 6 illustrates the different stages of preprocessing.



(a) Original image (b) Removed white border (c) Resolution transformation (d) Data normalization

Figure 6: Visualization of preprocessing steps applied to the OCT dataset.

6.3 Model Training

In this section, we describe the training process for the CFP model, OCT model, and the OCT-CFP Classifier , including their architectures, compilation, and fine-tuning strategies.

6.3.1 CFP Image Diagnosis Model

The CFP image diagnosis model is designed to classify fundus images into four categories: *Normal*, *Cataract*, *Diabetic Retinopathy*, and *Glaucoma*. This model is based on the pre-trained VGG19 architecture and employs transfer learning to leverage pre-trained weights for feature extraction while adding custom layers for fine-tuning specific to our dataset. The following sections describe the structure and functionality of the model in detail.

Model Architecture

- **Base Model (VGG19):** The VGG19 model, pre-trained on the ImageNet dataset, serves as the backbone of our architecture. It includes 19 layers consisting of convolutional layers, max-pooling layers, and fully connected layers. We exclude the top classification layers (*include_top=False*) to repurpose the model for feature extraction. The input size is set to `(224, 224, 3)` to match the standard dimensions required by VGG19.

- **Convolutional and Pooling Layers:** After the base model, we add a convolutional layer with tunable filters (`filters1`, varying between 64 and 96) and a kernel size of `(3, 3)`. This is followed by a batch normalization layer to stabilize training and a max-pooling layer to downsample the feature maps while retaining important spatial information.

- **Global Average Pooling:** A `GlobalAveragePooling2D` layer reduces the spatial dimensions of the feature maps to a single vector. This enables a compact representation of the extracted features while preventing overfitting.

- **Dense Layers and Dropout:** A fully connected dense layer with tunable units (`256--384`) and a ReLU activation function learns complex relationships between the features. To improve generalization and reduce overfitting, a dropout layer is included with a rate of `0.1`.
- **Output Layer:** The final dense layer contains 4 units corresponding to the four diagnostic categories. A `softmax` activation function generates probabilities for each class, ensuring that the output represents a valid probability distribution.
- **Learning Rate Scheduler:** A learning rate scheduler dynamically adjusts the learning rate during training. After two epochs, the learning rate is reduced by a factor of 10 to enable finer adjustments to the model weights as training progresses.

Model Compilation and Fine-Tuning

The model is compiled with the Adam optimizer and a categorical cross-entropy loss function, suitable for multi-class classification problems. Accuracy is used as the primary evaluation metric. During fine-tuning, the last 10 layers of the VGG19 base model are unfrozen to allow them to learn domain-specific features from the CFP dataset. A smaller learning rate (`1e-5`) is employed to ensure stable training and prevent the weights from diverging.

Implementation Summary

The final architecture of the model can be summarized as:

1. Pre-trained VGG19 as the feature extractor (frozen during initial training).
2. Additional convolutional and pooling layers for feature refinement.
3. Global average pooling to generate feature vectors.
4. Fully connected dense layers with regularization and dropout for classification.
5. Softmax activation in the output layer to predict the probabilities for each diagnostic category.

6.3.2 OCT Model Architecture

The OCT image diagnosis model is based on the VGG19 architecture, utilizing transfer learning to leverage pre-trained weights from ImageNet. This model is customized to classify OCT images into four categories: *Normal*, *CNV*, *DME*, and *Drusen*. Below is a detailed breakdown of the model architecture.

- **Base Model (VGG19):** The VGG19 model, pre-trained on ImageNet, acts as the backbone of the architecture. We exclude the top layers of the model (*include_top=False*) to use it as a feature extractor. The input size is set to `(224, 224, 3)` to match the VGG19 architecture's required input dimensions. The first few convolutional layers from the pre-trained model remain frozen to retain the learned features from ImageNet.
- **Convolutional Layer and Pooling:** AAfter the base model, we add a convolutional layer with tunable filters (`filters1`, varying between 64 and 96) and a kernel size of `(3, 3)`. This is followed by a batch normalization layer to stabilize training and a max-pooling layer to downsample the feature maps while retaining important spatial information.
- **Global Average Pooling:** A `GlobalAveragePooling2D` layer is applied after the convolutional layers to reduce the spatial dimensions of the feature maps to a single vector, providing a compact representation of the extracted features.
- **Dense Layers and Dropout:** A fully connected dense layer with tunable units (`256--384`) and a ReLU activation function learns complex relationships between the features. To improve generalization and reduce overfitting, a dropout layer is included with a rate of `0.1`.
- **Output Layer:** The final dense layer contains `4` units, corresponding to the four diagnostic categories: *Normal*, *Glaucoma*, *Diabetic Retinopathy*, and *Cataract*. A `softmax` activation function is used to produce probabilities for each class.
- **Learning Rate Scheduler:** A learning rate scheduler dynamically adjusts the learning rate during training. After two epochs, the learning rate is reduced by a factor of 10 to enable finer adjustments to the model weights as training progresses.

6.3.3 Model Compilation and Fine-Tuning

The model is compiled using the Adam optimizer and a sparse categorical cross-entropy loss function, which is suitable for multi-class classification tasks. The accuracy metric is used to evaluate the model's performance. During fine-tuning, the last 10 layers of the VGG19 base model are unfrozen to learn domain-specific features from the OCT dataset. A smaller learning rate (`1e-5`) is applied during this phase to prevent weights from diverging while allowing the model to adapt to the new dataset. Class weights was used for the imbalance in the dataset.

6.3.4 Implementation Summary

The architecture of the OCT model can be summarized as:

1. Pre-trained VGG19 as the feature extractor (frozen during initial training).
2. Additional convolutional layers with filters of 64, followed by batch normalization and max-pooling layers.
3. Global average pooling to reduce feature map dimensions.
4. Dense layers with ReLU activation and dropout for regularization.
5. Output layer with softmax activation for classification.
6. Class weights are used to address class imbalance during training

6.3.5 OCT and CFP Model

The model, built using a pre-trained VGG19 architecture, classifies OCT (Optical Coherence Tomography) and CFP (Color Fundus Photography) images.

- **Global Feature Flattening:** After feature extraction, the output of the base model is flattened using a `Flatten` layer, converting the multi-dimensional feature maps into a single feature vector suitable for dense layers.
- **Dense Layers and Dropout:** A dense layer with `256` neurons and a ReLU activation function learns higher-level feature representations. To prevent overfitting, a dropout layer with a rate of `0.6` is added, ensuring robust generalization by randomly deactivating some neurons during training.
- **Output Layer:** The final dense layer contains `2` units, corresponding to the binary classification task ("OCT" vs. "CFP"). A `softmax` activation function is used to output probabilities for each class, ensuring the outputs form a valid probability distribution.
- **Training Configuration:** The model is compiled using the Adam optimizer, which adapts the learning rate during training for efficient convergence. The loss function is `sparse_categorical_crossentropy`, suitable for multi-class classification tasks with integer-encoded labels. The metric used to evaluate model performance is `accuracy`.
- **Training and Validation:** The model is trained for `5 epochs` on a balanced dataset of OCT and CFP images. Training and validation datasets are split with an 80-20 ratio to evaluate the model's ability to generalize to unseen data.
- **Saving the Model:** After training, the model is saved in `HDF5` format for easy reloading and deployment. This ensures reproducibility and simplifies further analysis or fine-tuning tasks.

Model Compilation and Fine-Tuning

The model architecture is designed to leverage transfer learning using the VGG19 pre-trained model as the backbone. This approach takes advantage of the robust feature extraction capabilities of VGG19, trained on the ImageNet dataset. The model compilation and training steps are summarized as follows:

1. **Base Model Initialization:** The VGG19 model, pre-trained on the ImageNet dataset, is used as the feature extractor. The top classification layers are excluded by setting `include_top=False`, and the input size is specified as `(224, 224, 3)`.
2. **Freezing Layers:** All layers in the base model are frozen (`base_model.trainable = False`) during the initial training phase to preserve pre-trained weights and prevent overfitting on the limited dataset.

3. **Custom Classifier:** The classifier includes the following layers:
   - `Flatten()` to convert feature maps into a 1D vector.
   - A dense layer with 256 units and ReLU activation for feature learning.
   - Dropout with a rate of 0.6 to mitigate overfitting.
   - A dense output layer with 2 units (representing two classes) and softmax activation to output probabilities for each class.
4. **Loss Function and Optimizer:** The model is compiled with the Adam optimizer, which adapts the learning rate dynamically, and a sparse categorical cross-entropy loss function, suitable for integer-encoded multi-class classification problems.
5. **Evaluation Metric:** Accuracy is chosen as the primary evaluation metric to monitor model performance during training and validation.

Implementation Summary

The final implementation is as follows:

Training Details

The model is trained for 5 epochs using the specified training and validation datasets. The initial training phase involves freezing the base model to utilize pre-trained features, ensuring efficient learning from the domain-specific CFP dataset. The Dropout layer aids in reducing overfitting, while the Adam optimizer ensures stable convergence.

The final trained model is saved as `oct_cfp_vgg19_model_optimized.h5` and can be reloaded for inference or further fine-tuning. This modular design facilitates easy deployment and reuse for similar classification tasks.

7 Use Case diagram

A use case outlines the interactions and steps required for a user to achieve a specific goal. It focuses on understanding the users objectives and the flow of actions necessary to fulfill those objectives effectively. A use case Diagram for Diagnosis System for Fundus Diseases Based on Deep Learning, as follows:

Table 4: Login

Primary Actors	Health Worker, General Practitioner
Description	Users log into the system using their unique credentials to access shared functionality.
Preconditions	The user (General Practitioner or Health Worker) must have an active account in the system
Postconditions	- The user is authenticated and granted access to the system
Extensions	- Enter Password: Inputting the password as part of the login process. - Error Message: Triggered if the credentials are invalid or the account is locked

Table 5: Insert Information

Primary Actors	Health Worker, General Practitioner
Description	Users input patient-related information. For example: - Health Worker may input patient medical history. - General Practitioner may provide their own data, such as symptoms, personal information, or updates.
Preconditions	- The user must be logged into the system
Postconditions	- The information is successfully saved in the system.
Extensions	- Verify Error Message: Triggered if the data entered is incomplete or invalid.

Table 7: Start Live Diagnosis

Primary Actors	Health Worker, General Practitioner
Description	Users initiate live diagnosis functionality. - Health Worker may use this to conduct real-time consultations with General Practitioners. - Patients may participate in live consultations or self-guided diagnostic processes.
Preconditions	- A diagnosis option must have been selected.
Postconditions	- The live diagnostic session is started
Includes	- Start Camera: Activates the camera for imaging

Table 8: Start Diagnosis by Upload Image

Primary Actors	Health Worker, General Practitioner
Description	Users upload images to assist with diagnosis. Examples: - Health Worker upload medical scans. - General Practitioners upload photos of symptoms for analysis.
Preconditions	- A diagnosis option must have been selected
Postconditions	- The uploaded image is processed by the system.

Table 9: Start Camera

Primary Actors	Health Worker, General Practitioner
Description	The system activates an external integrated camera to capture photos of the symptoms
Preconditions	The live diagnosis process must be initiated
Postconditions	The camera successfully streams or captures the required data

Table 10: Display Diagnosis Result

Primary Actors	System (for generation), Health Worker, General Practitioner (for viewing)
Description	The system generates and displays diagnosis results. - Users can review results
Preconditions	- Diagnosis data must already be processed by the system
Postconditions	- The results are accessible for viewing by any authorized user

Table 11: Take Result

Primary Actors	Health Worker, General Practitioner
Description	Users retrieve diagnosis results for further action. - Health Worker may use the results to prepare treatment plans. - General Practitioner may save the results for reference or share them with other professionals
Preconditions	- Diagnosis results must be generated and available in the system
Postconditions	- Results are successfully accessed or downloaded

Table 6: Choose Diagnosis Options

Primary Actors	Health Worker, General Practitioner
Description	Users select diagnostic methods, such as: - Live diagnosis through real-time interaction. - Uploading images for analysis.
Preconditions	- General Practitioner-related information must already exist in the system.
Postconditions	- The system directs the user to the selected diagnostic process.

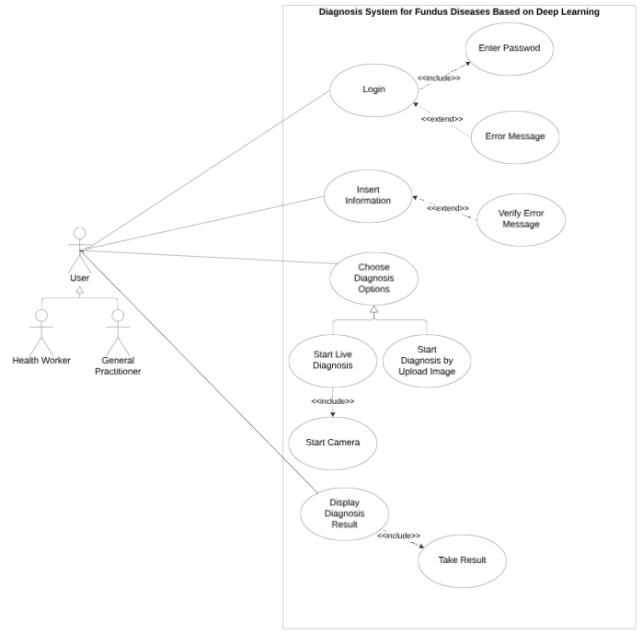


Figure 7: Use case diagram

8 Build interface

We decided to make our project's interface a website using HTML, CSS, and JavaScript. We chose a website because it's always accessible and easy to use for both doctors and regular users.
The website has a simple and clear interface. On the homepage, the user will see two options: they can't start the process without choosing one. The options are either to register as a new user or log in if they already have an account.
After logging in, the user will have two choices: - **Live Examination**: This option uses the camera connected to the fundus device for a real-time diagnosis, with the result displayed immediately. - **Upload an Image**: The user can upload a fundus (CFP) or OCT image. The system will then classify the image to determine whether it's CFP or OCT. Once the type is identified, the image will be sent to the correct model for diagnosis. When the user clicks on "Show Result," the diagnosis will appear.

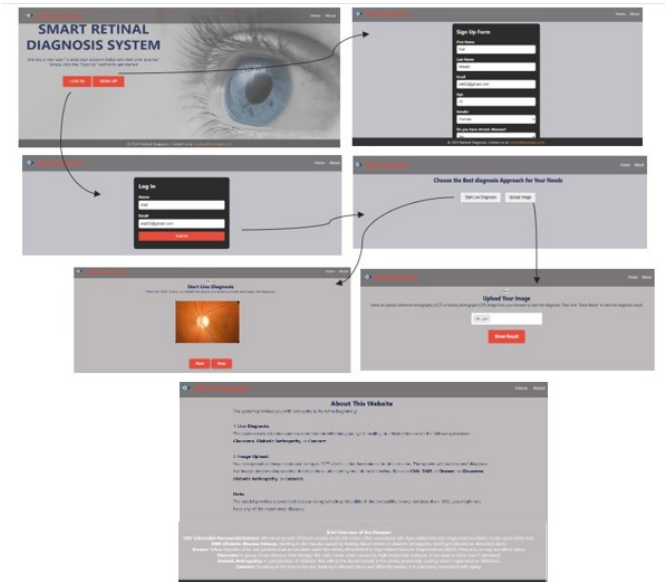


Figure 8: Interface building

9 Result

In the implementation phase, the Google Colab platform was utilized for preprocessing tasks and model development. Two distinct datasets, CFP and OCT, sourced from Kaggle (as discussed in Section 3), were employed in this study. Additionally, a combined dataset was constructed by merging the CFP and OCT datasets to develop a model capable of distinguishing between these two data types.

The VGG19 architecture served as the foundation for the models, with its layers modified and hyperparameters fine-tuned using Keras Tuner to optimize performance. Subsequently, the model was tailored and compiled for three distinct tasks, each corresponding to a specific dataset and classification objective.

The CFP dataset was split into training, validation, and testing subsets with a ratio of 60:20:20, resulting in 7,588 images for training, 2,531 for evaluation, and 2,532 for testing. Similarly, the OCT dataset was divided into a 70:20:10 ratio, yielding 53,547 images for training, 7,650 for evaluation, and 16,300 for testing.

The performance of the developed models was rigorously evaluated using various evaluation metrics to ensure their reliability and effectiveness for the intended classification tasks.

9.0.1 Evaluation Metrics

– **Accuracy:** Measures correct predictions for all predictions.

$$Accuracy = \frac{\text{True Positive} + \text{True Negative}}{\text{Total number of predictions}}$$

Suitable for balanced data, but for unbalanced data, it may reflect something other than the model's actual performance.

– **Precision:** Measures the true positive predictions among all positive predictions made by the model.

$$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

Precision is essential to know the model's ability to avoid false positives.

– **Recall:** Measures positive predictions for all positive cases, including ones mispredicted by a wrong label.

$$Recall = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

Recall is essential to ensure the model can detect all positive cases.

– **F1 Score:** Harmonic mean of precision and recall.

$$F1\,\text{score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Suitable for unbalanced data and reflects the model performance. While both F1-score and accuracy are essential metrics for evaluating classification models, they serve different purposes. They should be used depending on the specific characteristics of the dataset and the task at hand.

– **Macro-average:** The metric is calculated for each class independently, and then the average is calculated for all classes. Macro is suitable for the multi-class problem because it gives a fair assessment of the model's performance across all classes, ensuring that all classes contribute to calculating the final metrics ratios.

$$\text{Macro-precision} = \frac{\text{Precision}_A + \text{Precision}_B + \text{Precision}_C}{3}$$

$$\text{Macro-recall} = \frac{\text{Recall}_A + \text{Recall}_B + \text{Recall}_C}{3}$$

Macro-F1-score is the harmonic mean of macro-precision and macro-recall.

10 Discussion

10.1 Analysis of Model Performance Across Preprocessing Techniques

This analysis evaluates the performance of the CFP diagnosis model under three preprocessing conditions: raw CFP images, images processed with Contrast Limited Adaptive Histogram Equalization (CLAHE), and images processed with both CLAHE and histogram equalization. By comparing metrics such as accuracy, precision, recall, F1-score, loss, and AUC scores, we aim to understand the impact of preprocessing on the model's diagnostic capability.

10.1.1 Performance on Raw CFP Dataset

The model trained on the raw CFP dataset showed reasonable performance. Key observations include:

- **Validation Accuracy:** The model achieved a validation accuracy of approximately 93.95%, demonstrating strong diagnostic ability with the raw data.
- **Precision and Recall:** Precision was highest for 'Diabetic Retinopathy' (97%) but lower for 'Normal' (84%), indicating potential bias toward certain classes.
- **F1-Score:** The overall F1-scores were acceptable, ranging from 0.89 to 0.96, but the scores for 'Glaucoma' (0.89) suggest room for improvement.
- **Confusion Matrix:** The confusion matrix highlights misclassifications, particularly between 'Normal' and 'Glaucoma'.
- **AUC Scores:** High AUC scores (e.g., 0.99 for 'Diabetic Retinopathy') indicate excellent discriminatory ability, but slightly lower scores for 'Normal' (0.9873) suggest sensitivity issues for this class.

Strengths: High overall accuracy and excellent performance for 'Diabetic Retinopathy'.
Weaknesses: Subpar recall for 'Glaucoma' and moderate confusion between 'Normal' and other classes.

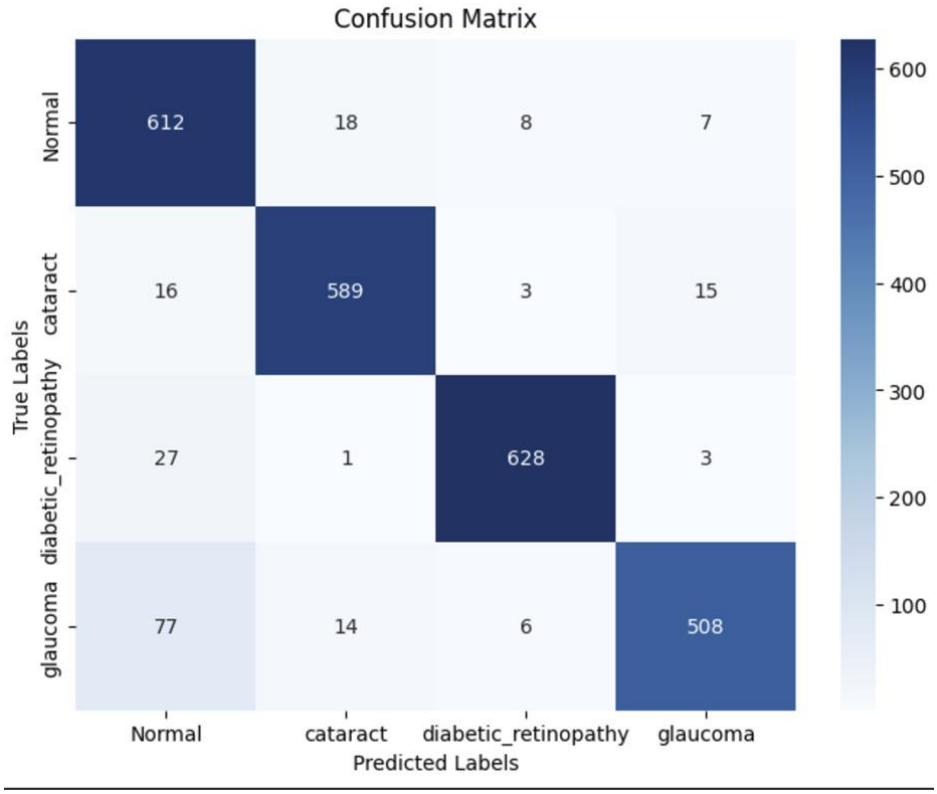


Figure 9: Confusion Matrix for the Raw CFP Dataset

10.1.2 Performance on CLAHE-Processed CFP Dataset

Preprocessing with CLAHE resulted in a noticeable improvement in performance metrics:

- **Validation Accuracy:** Improved to 95.72%, indicating enhanced generalization.
- **Precision and Recall:** Precision improved for 'Normal' (90%) and 'Cataract' (97%), reflecting better handling of underrepresented features.
- **F1-Score:** The F1-scores for 'Cataract' and 'Diabetic Retinopathy' reached near-perfect levels (0.94–0.96), while scores for 'Glaucoma' remained at 0.82.
- **Confusion Matrix:** Misclassifications decreased significantly compared to the raw dataset. 'Glaucoma' instances, however, continued to overlap with 'Normal'.
- **AUC Scores:** AUC improved for all classes, with 'Cataract' achieving a perfect score (1.00), signifying excellent sensitivity and specificity.

Strengths: Enhanced contrast in images likely led to improved feature extraction, especially for 'Cataract'.
Weaknesses: Misclassifications for 'Glaucoma' suggest it remains a challenging class for the model.

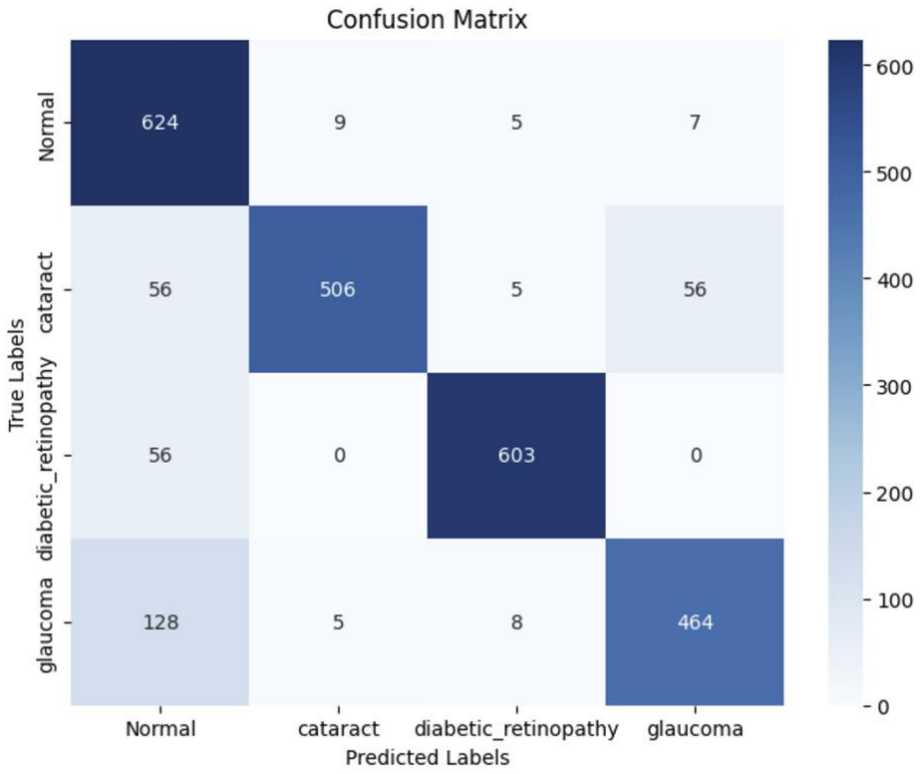


Figure 10: Confusion Matrix for the CLAHE-Processed CFP Dataset

10.1.3 Performance on CLAHE and Histogram-Equalized Dataset

Combining CLAHE with histogram equalization led to the best overall performance:

- **Validation Accuracy:** Reached the highest level at 96.16%, underscoring the value of combined preprocessing techniques.
- **Precision and Recall:** Precision and recall for 'Normal' and 'Glaucoma' improved significantly, reaching 99% and 0.99, respectively, for some classes.
- **F1-Score:** All F1-scores exceeded 0.90, with 'Diabetic Retinopathy' achieving 0.97.

– **Confusion Matrix:** The confusion matrix exhibited minimal misclassifications, with the majority of instances correctly classified.
– **AUC Scores:** AUC scores were uniformly high across all classes (e.g., 0.9940 for 'Cataract'), confirming robust classification performance.

Strengths: The combined preprocessing methods enhanced feature clarity, leading to significant improvements in recall and AUC for all classes.
Weaknesses: None of the classes showed notable weaknesses, though minor confusion persisted in a few borderline cases.

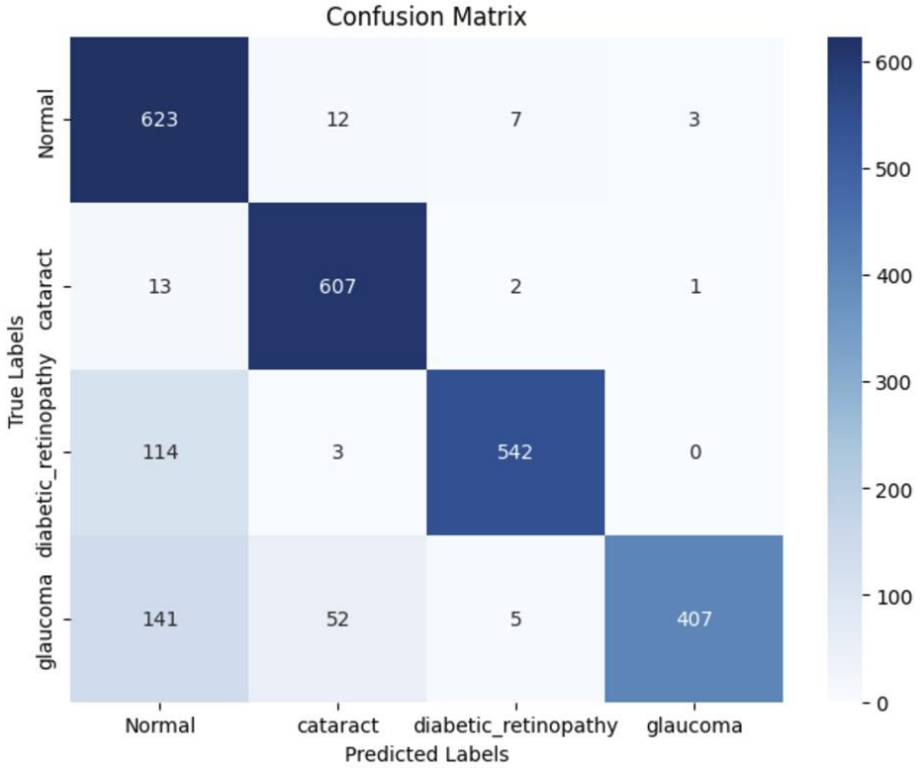


Figure 11: Confusion Matrix for CLAHE and Histogram-Equalized Dataset

10.1.4 Comparative Analysis

– **Accuracy Trends:** Accuracy increased progressively across the three conditions, with combined preprocessing delivering the best results.
– **Class-Specific Insights:** The addition of CLAHE and histogram equalization particularly benefited 'Normal' and 'Glaucoma', two classes previously affected by misclassification.
– **Generalization:** Improved validation and test metrics across preprocessing conditions indicate that the model effectively leveraged enhanced features for better generalization.
– **AUC and Sensitivity:** AUC scores consistently improved, reflecting better sensitivity and specificity for all classes.

10.2 Analysis of Model Performance on OCT Dataset

This analysis evaluates the performance of the CFP diagnosis model when applied to a completely different dataset: the OCT dataset. While the model was originally optimized for CFP images, its performance on OCT images provides insights into its generalization capabilities and limitations when handling data with distinct feature characteristics.

10.2.1 Performance Overview

The key performance metrics of the model on the OCT dataset are summarized below:

- **Validation Accuracy:** The validation accuracy decreased significantly compared to its performance on the CFP dataset, indicating that the model struggled to generalize to the new data type. Accuracy hovered around 86%.
- **Precision and Recall:**
  - Precision for most classes dropped, particularly for classes with overlapping features or fewer examples in the dataset.
  - Recall for certain classes, such as 'DME' and 'Drusen,' showed noticeable declines, suggesting the model's reduced sensitivity to identifying these conditions accurately.
- **F1-Score:** The F1-scores ranged from moderate to low, with certain classes achieving acceptable levels (e.g., 'CNV') while others exhibited poor scores, reflecting imbalances in precision and recall.
- **Confusion Matrix:**
  - Significant misclassifications were observed, particularly between 'Normal' and 'DME' and between 'Drusen' and other classes.
  - This suggests that the model struggled to differentiate between certain OCT-specific features that were not present in the CFP dataset.
- **AUC Scores:**
  - The ROC-AUC scores were lower overall, indicating weaker discriminatory performance on the OCT dataset. Some classes, like 'CNV,' maintained relatively higher AUC scores compared to others.

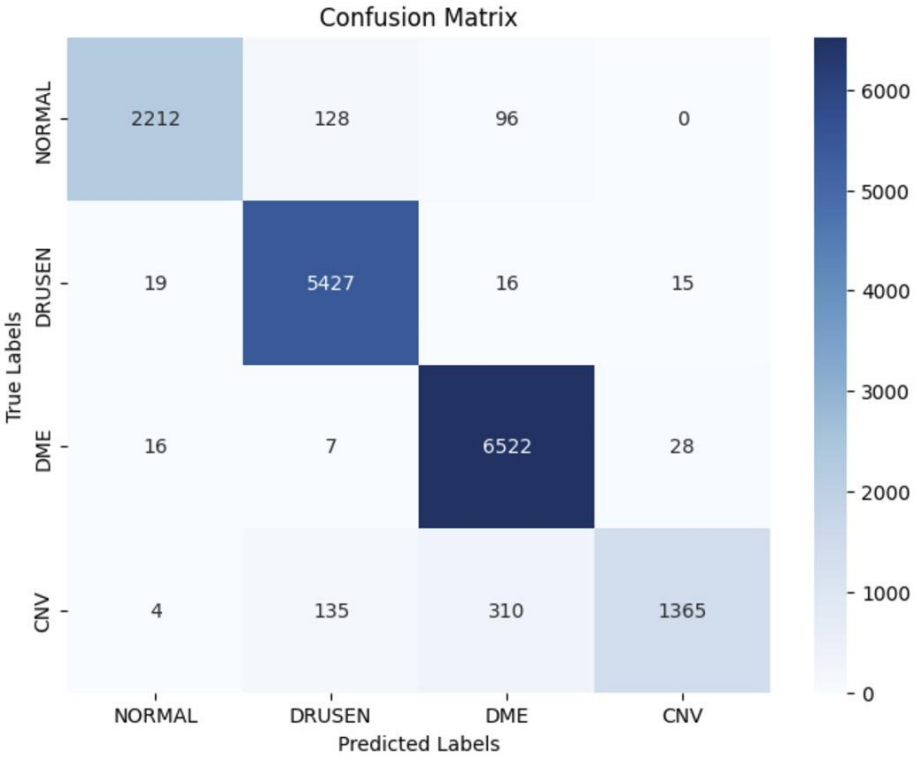


Figure 12: Confusion Matrix for OCT Dataset

10.3 Key Observations

- **Domain Shift Effect:**

    * The OCT dataset represents a significant domain shift from the CFP dataset in terms of image characteristics, including texture, contrast, and underlying features. The model's architecture, trained and tuned on CFP data, was not fully equipped to handle these differences.

- **Feature Mismatch:**

    * The filters and feature maps learned by the model were specialized for CFP images, which may explain the poor transferability to OCT images. OCT images require distinct feature extraction, particularly for detecting subtleties like fluid pockets and retinal layers.

- **Misclassification Trends:**

    * The confusion between 'Normal' and pathological conditions like 'DME' indicates that the model failed to capture OCT-specific distinguishing features.
    * Classes with overlapping visual characteristics in the OCT dataset, such as 'Drusen' and 'CNV,' further exacerbated misclassifications.

- **Preprocessing Limitations:**

    * Unlike the CFP dataset, where preprocessing significantly improved performance, the preprocessing techniques employed for the CFP model may not be suitable for OCT images.

*10.4 **Challenges in pre-processing process***

1. Frangi Filter Integration During the preprocessing phase, the Frangi filter was tested to enhance tubular structures, such as blood vessels, in the CFP dataset. Despite its potential to improve diagnostic accuracy, the implementation faced significant technical difficulties. Each application of the filter required approximately six hours to process before abruptly terminating without completing the operation. No error messages were generated to diagnose the issue, and repeated attempts with varying configurations yielded the same results.

These issues highlight the challenges of incorporating computationally intensive techniques into workflows with large datasets. Future efforts could focus on optimizing the processing pipeline, utilizing high-performance computing resources, or testing the filter on smaller data subsets to enable its successful integration.

2. Vessel Segmentation Vessel segmentation was attempted using a pre-existing repository designed to extract blood vessel structures from fundus images. However, this process encountered multiple failures due to code incompatibilities with modern software libraries. Specific issues included the deprecated use of np.int, which caused crashes even after attempts to update to compatible versions of NumPy.

Additionally, the segmentation process was highly resource-intensive, with runs taking considerable time and often terminating abruptly without clear error messages. Attempts to optimize by reducing data size, managing RAM usage, and testing various library versions failed to resolve the problem.

3. Despite implementing dynamic cropping to remove white borders, some residual white areas remain in the images. This issue continues due to variations in image content and white border sizes, which prevented the complete removal of white borders in most images. Moreover, while various noise removal filters were tested, they led to unnecessary blurring in images, which was in the end not included in the final preprocessing pipeline to keep the image details.

These challenges underscore the need for robust software compatibility and resource management strategies. Future work may involve exploring alternative vessel segmentation techniques, updating outdated code repositories, and leveraging high-performance computing to ensure successful execution. For the OCT dataset, further testing of preprocessing techniques may be necessary, such as improving the removal of residual white borders and exploring more effective noise reduction methods, which may be required to improve image quality without compromising key features.

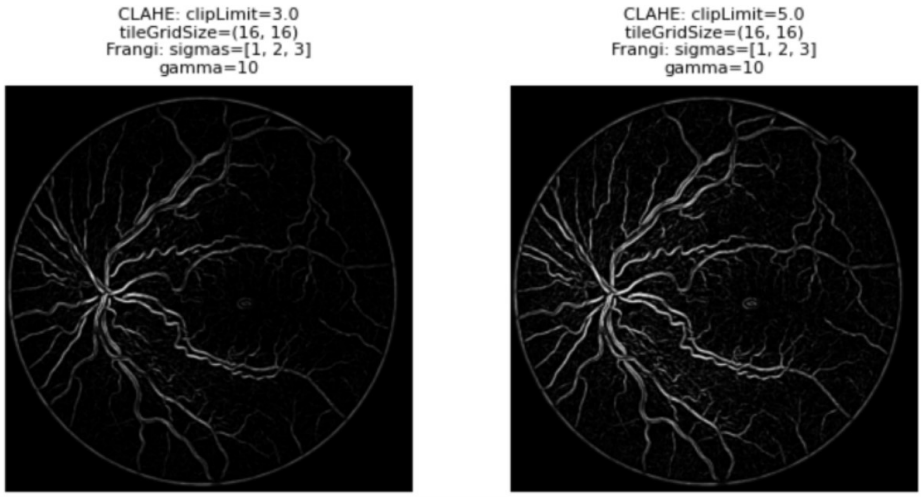


Figure 13: Vessel Segmentation using Frangi

10.5 ***Conclusion***

The model's performance on the OCT dataset highlights the importance of dataset-specific optimization. While the CFP-trained model demonstrated robust performance on its intended dataset, its ability to generalize to OCT images was limited due to significant domain differences. Future work should focus on adapting the model architecture and training pipeline to better handle OCT-specific features, ultimately improving its versatility and applicability to a wider range of ophthalmological datasets

References

1. Akinwumi Fajola, Olayide Olabumuyi, Aloni Alali, Bunmi Adetula, Rebecca Ogbimi, and Suodei Akenge. Beyond community health service provision: Assessing the knowledge attitude and practice of eye care among beneficiaries of an intervention in an inner-city community in lagos metropolis. Asian Journal of Medicine and Health, 22(2):1–11, 2024.
2. World Health Organization. World report on vision. 2019.
3. Lokman Balyen and Tunde Peto. Promising artificial intelligence-machine learning-deep learning algorithms in ophthalmology. Asia-Pacific Journal of Ophthalmology, 8(3):264–272, 2019.
4. Junjun He, Cheng Li, Jin Ye, Yu Qiao, and Lixu Gu. Multi-label ocular disease classification with a dense correlation deep neural network. Biomedical Signal Processing and Control, 63:102167, 2021.
5. Adam Grzybowski, Kai Jin, Jun Zhou, Xiang Pan, Ming Wang, Jun Ye, and Tien Yin Wong. Retina fundus photograph-based artificial intelligence algorithms in medicine: A systematic review. Ophthalmology and Therapy, 13(8):2125–2149, August 2024. Epub 2024 Jun 24.
6. Jaweria Aslam, Muhammad Asad Arshed, Sawaira Iqbal, and Hafiz Muhammad Hasnain. Deep learning based multi-class eye disease classification: Enhancing vision health diagnosis. Technical Journal, 29(01):7–12, 2024.
7. Nicoleta Anton, Bogdan Doroftei, Silvia Curteanu, Lisa Catãlin, Ovidiu-Dumitru Ilie, Filip Târcoveanu, and Camelia Margareta Bogdănici. Comprehensive review on the use of artificial intelligence in ophthalmology and future research directions. Diagnostics, 13(1):100, 2022.
8. Zhixi Li, Stuart Keel, Chi Liu, Yifan He, Wei Meng, Jane Scheetz, Pei Ying Lee, Jonathan Shaw, Daniel Ting, Tien Yin Wong, Hugh Taylor, Robert Chang, and Mingguang He. An Automated Grading System for Detection of Vision-Threatening Referable Diabetic Retinopathy on the Basis of Color Fundus Photographs. Diabetes Care, 41(12):2509–2516, 10 2018.
9. Emma Beede, Elizabeth Baylor, Fred Hersch, Anna Iurchenko, Lauren Wilcox, Paisan Ruamviboonsuk, and Laura M Vardoulakis. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In Proceedings of the 2020 CHI conference on human factors in computing systems, pages 1–12, 2020.
10. Lakshmi Chakka. Analyzing optimal image preprocessing techniques for automated retinal disease diagnosis. 2023.
11. Z. Sun and Y. Sun. Automatic detection of retinal regions using fully convolutional networks for diagnosis of abnormal maculae in optical coherence tomography images. Journal of Biomedical Optics, 24(5):1–9, May 2019.
12. Depeng Wang and Liejun Wang. On oct image classification via deep learning. IEEE Photonics Journal, 11(5):1–14, 2019.

13. R. Sarki, K. Ahmed, H. Wang, Y. Zhang, J. Ma, and K. Wang. Image preprocessing in classification and identification of diabetic eye diseases. Data Science and Engineering, 6:455–471, 2021.
14. A. García-Floriano and E. Ventura-Molina. Age-related macular degeneration detection in retinal fundus images by a deep convolutional neural network. Mathematics, 12:1445, 2024.
15. A. Mukherjee, D. Rathore, S. Shree, and S. Jameel. Diagnosis of diabetic retinopathy. International Journal of Engineering Research and Applications, 5(2):21–24, 2015. Available from: IJERA Website.
16. Tanji Mahamed. Odir5k classification dataset. https://www.kaggle.com/datasets/tanjemahamed/odir5k-classification, 2021. Accessed: 2024-10-06.
17. Daniel S. Kermany, Kang Zhang, and Michael Goldbaum. Large dataset of labeled optical coherence tomography (oct) and chest x-ray images. https://data.mendeley.com/datasets/rscbjbr9sj/3, 2018. Accessed: 2024-11-28.
18. Daniel S. Kermany, Kaifu Zhang, and Michael Goldbaum. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell, 172(5):1122–1131.e9, 2018.