Optical Coherence Tomography (OCT) Image Processing for Eye Diseases

Mid-Term Progress Report

Akash Chowdhury
210192109

October 20, 2024

Abstract

Optical Coherence Tomography (OCT) is a critical imaging technique in ophthalmology, providing high-resolution images for diagnosing retinal diseases. This project focuses on the development of an automated image processing tool aimed at improving the analysis of OCT images. The tool employs advanced deep learning methods to segment retinal layers and classify diseases such as choroidal neovascularization (CNV), diabetic macular edema (DME), and drusen. The dataset used is sourced from Mendeley, containing thousands of OCT images labeled into distinct disease categories. Thus far, significant progress has been made in preprocessing the data and implementing an Inception-based convolutional neural network (CNN) for classification tasks. The ultimate goal is to refine the model's accuracy,develop a segmentation tool, and integrate it into a clinical-grade software solution that can assist ophthalmologists in providing faster and more accurate diagnoses.

1 Introduction

Optical Coherence Tomography (OCT) is a non-invasive imaging modality widely used in ophthalmology to capture detailed, high-resolution cross-sectional images of the retina. These images provide crucial information about the structural integrity of the retina, which aids in the early detection and monitoring of various eye conditions, including age-related macular degeneration (AMD), diabetic retinopathy, and glaucoma. OCT allows clinicians to visualize the retinal layers, identify abnormalities, and assess disease progression with high precision.

However, the manual analysis of OCT images can be time-consuming, subjective, and prone to inter-observer variability. This project aims to develop an automated image processing tool that leverages machine learning techniques to assist in the analysis of OCT images. By implementing advanced deep learning models, the tool will focus on two critical tasks: the segmentation of retinal layers and the classification of retinal diseases.

The primary objective of the project is to enhance diagnostic accuracy and efficiency in clinical settings by automating these tasks. The project will employ a convolutional neural network (CNN) architecture, specifically an Inception model, to process the OCT images and classify them into distinct disease categories such as CNV (choroidal neovascularization), DME (diabetic macular edema), DRUSEN, and NORMAL. Additionally, the project will explore the potential of transfer learning to address the challenges posed by limited annotated datasets. Ultimately, the goal is to develop a robust tool that can be integrated into clinical workflows, providing ophthalmologists with faster and more reliable diagnostic insights.

2 Literature Review

Several studies have focused on the application of deep learning models for OCT image analysis. In [**?**], a comprehensive review of deep learning in OCT highlights the advancements in segmentation and classification techniques. Similarly, Kermany et al. [3] used convolutional neural networks (CNNs) to achieve high accuracy in classifying retinal diseases, laying the groundwork for many subsequent developments.

2.1 Segmentation Techniques

The segmentation of retinal layers is critical for diagnosing diseases like AMD and diabetic macular edema (DME). Methods such as thresholding, edge detection, and graph-based algorithms have been widely used in the past, but the advent of CNNs has allowed for more accurate and efficient segmentation. In this project, a CNN-based approach will be employed for segmenting retinal layers, utilizing transfer learning to deal with the limited size of annotated datasets.

2.2 Classification Techniques

The classification of retinal pathologies involves distinguishing between diseased and healthy images. Support Vector Machines (SVM), random forests, and CNNs have been applied with varying degrees of success. For this project, an Inception model was chosen due to its ability to extract detailed features from OCT images, making it ideal for distinguishing subtle pathological differences.

3 Dataset

The dataset used for this project is sourced from Mendeley, titled "Large Dataset of Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images" [4]. The dataset contains thousands of validated OCT images split into four categories: CNV (choroidal neovascularization), DME (diabetic macular edema),

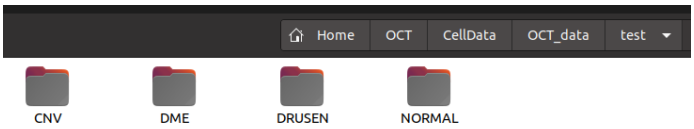


Figure 1: Dataset Structure

DRUSEN, and NORMAL. Images are labeled with disease type, randomized patient ID, and image number.

3.1 Dataset Structure

The dataset is organized as follows:

- **CNV**: Images showing choroidal neovascularization.
- **DME**: Images showing diabetic macular edema.
- **DRUSEN**: Images showing drusen deposits.
- **NORMAL**: Healthy retinal images.

The images are divided into a training set and a test set based on independent patients to avoid data leakage.

4 Methodology

4.1 Preprocessing

Preprocessing is a critical step in preparing Optical Coherence Tomography (OCT) images for analysis. OCT images typically suffer from artifacts like speckle noise, low contrast, and intensity variations, which need to be addressed for accurate segmentation and classification.

4.1.1 Noise Reduction

OCT images often contain speckle noise, which can obscure important structures. To address this, median filtering was applied. The median filter works by replacing each pixel value with the median of its surrounding neighbors, effectively reducing salt-and-pepper noise while preserving edges. This technique ensures that the fine details of the retinal layers remain intact.

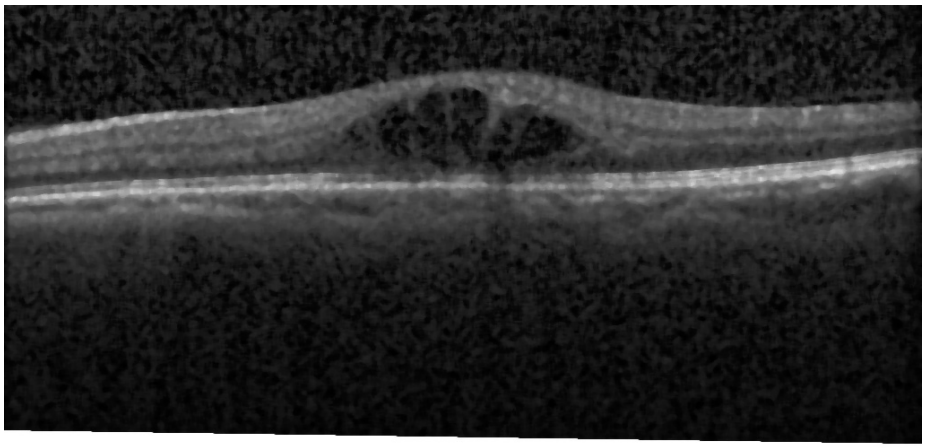


Figure 2: Median Filtering

4.1.2 Contrast Enhancement

Due to the low contrast between different retinal layers, enhancing contrast is necessary for better feature extraction. Histogram Equalization was initially used to adjust the intensity distribution of the images, but a more advanced method, Contrast Limited Adaptive Histogram Equalization (CLAHE), was later applied. CLAHE enhances local contrast by operating on small regions of the image, preventing noise over-amplification while improving the visibility of retinal structures.

4.1.3 Normalization

Normalization is essential to standardize the intensity values of the images, especially when they come from different sources or devices. Two techniques were employed:

- **Min-Max Normalization**: Rescales the pixel values to a fixed range, typically between 0 and 1, using the formula:

$$\text{scaled\_value} = \frac{\text{pixel} - \text{min}}{\text{max} - \text{min}}$$

- **Z-score Normalization**: Standardizes the pixel values based on the mean and standard deviation, ensuring a consistent distribution with mean 0 and standard deviation 1.

4.1.4 Artifact Removal

OCT images can also contain artifacts such as shadows or reflections caused by blood vessels or other obstructions. Shadow compensation techniques were

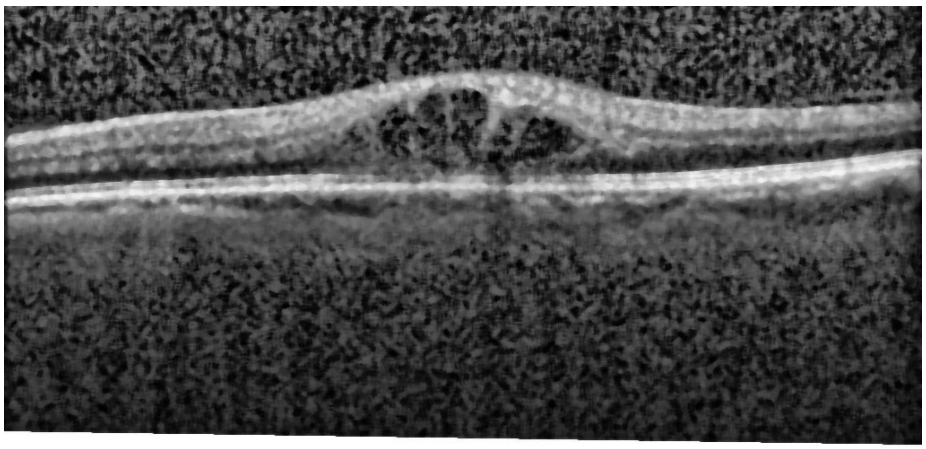


Figure 3: CLAHE:Contrast Enhancement

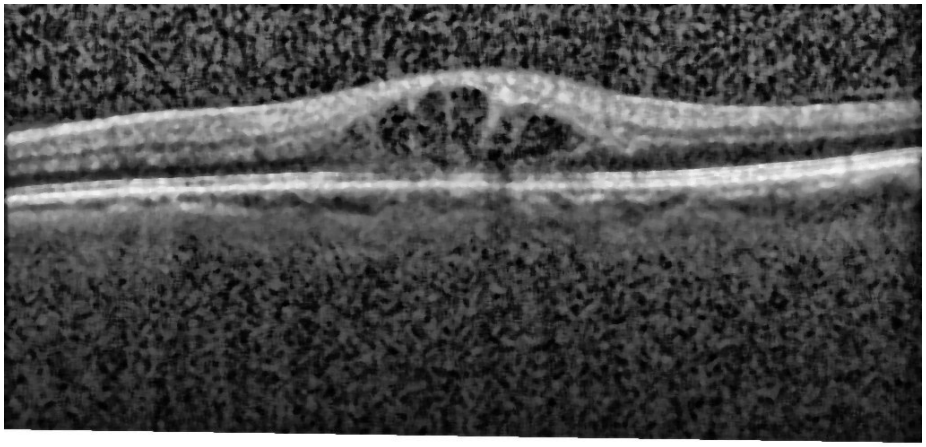


Figure 4: Minmax Normalisation

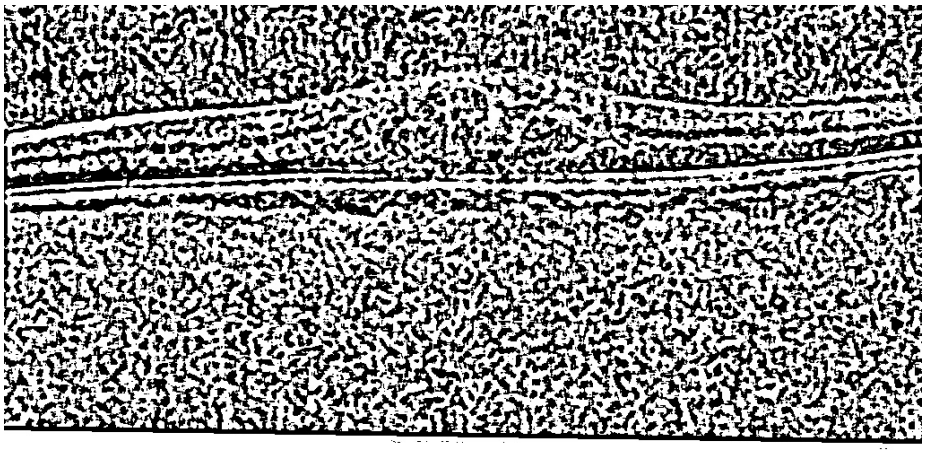


Figure 5: Adaptive Thresholding

used to correct these artifacts, particularly using histogram equalization and thresholding methods. This improves the clarity of the retinal structures by reducing noise and enhancing the overall image quality.

4.1.5 Image Registration

In cases where multiple OCT images are captured from different angles or at different time points, image registration was necessary to align the images into a common coordinate system. Both rigid and non-rigid registration techniques were employed to account for motion artifacts and variations in acquisition angles, ensuring consistency in the analysis.

4.1.6 Segmentation Preprocessing

To prepare the images for segmentation, further feature enhancement steps were applied:

- **Edge Detection**: Sobel and Canny edge detection algorithms were used to identify boundaries between retinal layers. These methods compute intensity gradients to highlight the transitions between different tissue types.
- **Binarization**: The grayscale images were converted into binary form using Otsu's method to emphasize key structures, making the retinal layers easier to segment.

4.1.7 Resampling

Resampling was necessary to standardize the spatial resolution of the OCT images, especially since they were acquired from various devices. Bilinear and

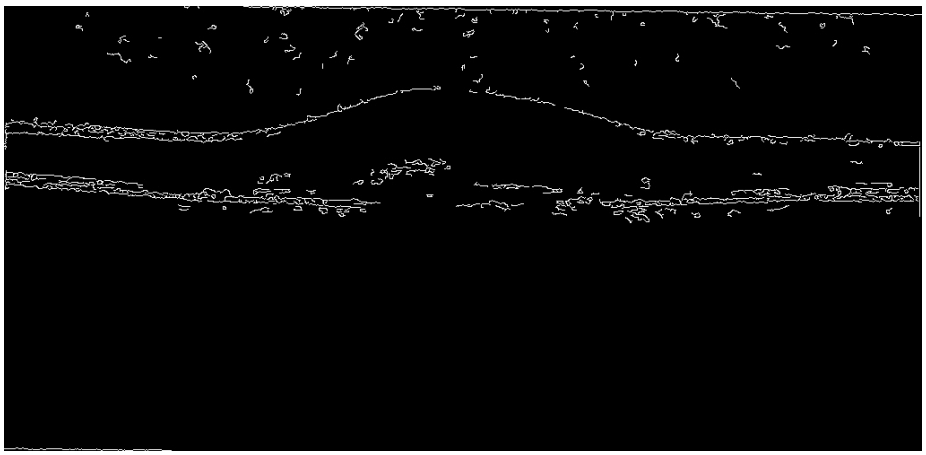


Figure 6: Canny Edge Detection

Figure 7: Otsu Thresholding

Figure 8: Global Thresholding

bicubic interpolation methods were used to resize the images to a uniform resolution, ensuring consistency across the dataset.

4.1.8 Summary of Preprocessing Algorithms

- **Median Filtering**: For noise reduction.
- **CLAHE**: For contrast enhancement.
- **Min-Max and Z-Score Normalization**: For intensity standardization.
- **Shadow Compensation and Thresholding**: For artifact removal.
- **Edge Detection and Binarization**: For feature enhancement.
- **Interpolation**: For image resampling.

4.2 Inception Model Implementation

To classify retinal diseases from OCT images, the InceptionV3 model was employed due to its ability to capture spatial hierarchies and multi-scale features, which are particularly beneficial for detailed medical images like those from OCT.

4.2.1 Model Architecture

InceptionV3 is a convolutional neural network (CNN) architecture designed for deep learning tasks involving image classification. It uses multiple types of filters at various scales to capture different features within the same convolutional layer, enhancing the model's ability to distinguish between retinal pathologies such as CNV, DME, DRUSEN, and NORMAL.

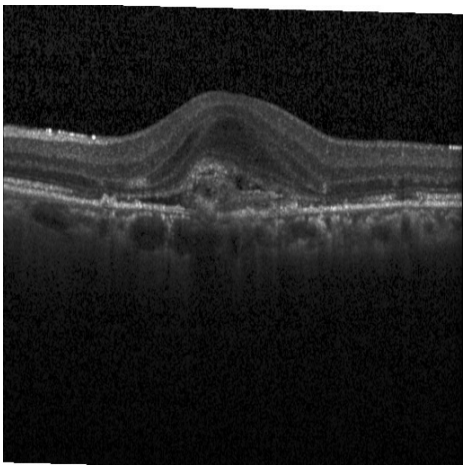


Figure 9: Bilinear Resized Image

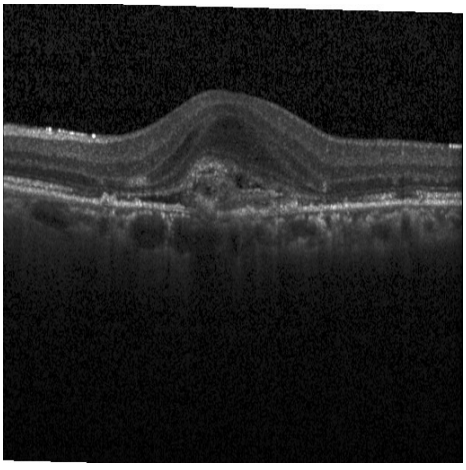


Figure 10: Bicubic Resized Image

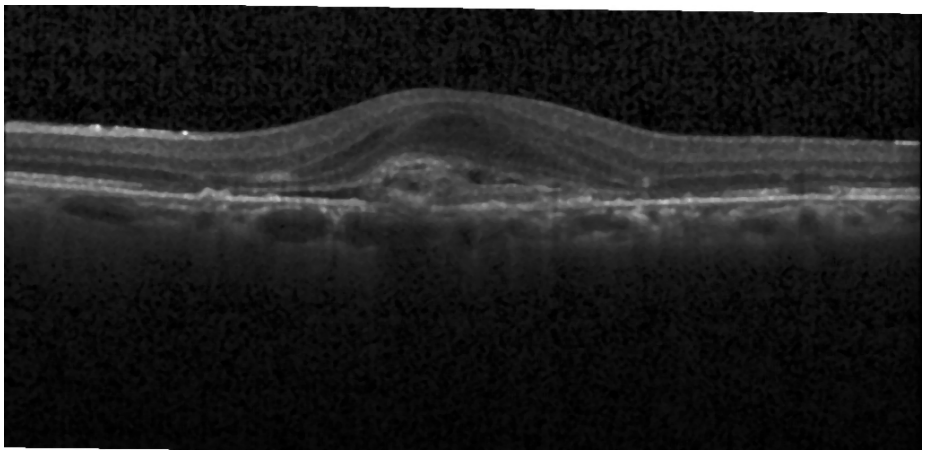


Figure 11: Denoised Image

For this project, the InceptionV3 architecture was initialized with pre-trained weights from the ImageNet dataset. These weights helped the model leverage prior knowledge of general image features, which is crucial given the relatively small size of the OCT dataset. The model's architecture, with its pre-trained weights, allows for transfer learning, speeding up the training process while improving accuracy on medical data.

The model's configuration is as follows: - ***Input Shape***: $224 \times 224 \times 3$, corresponding to the size of the OCT images. - ***Pre-trained Weights***: The weights were loaded from a Kaggle directory containing the pre-trained weights for InceptionV3. - ***Frozen Layers***: To prevent overfitting and reduce computational cost, the layers from the InceptionV3 base model were frozen during training. This ensured that only the top layers, which were added for the OCT classification task, were trained.

The output of the InceptionV3 model was then flattened and connected to a dense layer with four output units corresponding to the four categories of retinal diseases (CNV, DME, DRUSEN, NORMAL). A softmax activation function was used at the output layer to provide probabilities for each class.

4.2.2 Model Training

The model was compiled with categorical cross-entropy loss, given the multi-class classification nature of the problem, and Adam optimizer was used for faster convergence. Several metrics were tracked during training, including accuracy, AUC, Cohen's Kappa, F1 Score, Precision, and Recall. These metrics provided a holistic view of the model's performance, particularly in a medical context where misclassifications can have significant consequences.

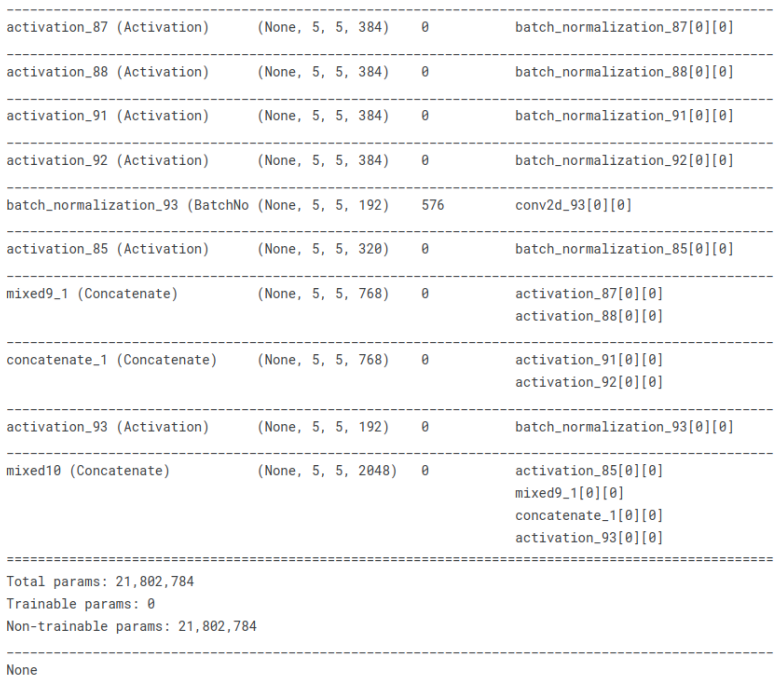


Figure 12: InceptionV3 Model Architecture for OCT Classification

$$\text{Loss} = -\sum_{i=1}^{4} y_i \log(\hat{y}_i) \tag{1}$$

where y_i is the true label, and $\hat{y}_i$ is the predicted probability for each class.

Data augmentation was applied to the training images to reduce overfitting. The images were rescaled, randomly rotated, zoomed, and flipped to enhance the model's robustness. The training generator used the following configuration:

- **Image Size**: 224×224
- **Batch Size**: 500
- **Steps per Epoch**: $\frac{76515}{500}$
- **Number of Epochs**: 30

Validation was conducted using a separate generator for the validation dataset, ensuring that the model's performance was evaluated on unseen data. The training accuracy and validation accuracy over 30 epochs were visualized, showing improvements with each epoch.

```
153/153 [==============================] - 302s 2s/step - loss: 0.0384 - accuracy: 0.9870 - auc: 0.9995 - cohen_kappa: 0.9798
- f1_score: 0.9772 - precision: 0.9875 - recall: 0.9867 - val_loss: 0.5644 - val_accuracy: 0.9049 - val_auc: 0.9691 - val_cohe
n_kappa: 0.8490 - val_f1_score: 0.8206 - val_precision: 0.9068 - val_recall: 0.9036
Epoch 27/30
153/153 [==============================] - 300s 2s/step - loss: 0.0261 - accuracy: 0.9913 - auc: 0.9998 - cohen_kappa: 0.9865
- f1_score: 0.9847 - precision: 0.9918 - recall: 0.9908 - val_loss: 0.4564 - val_accuracy: 0.9160 - val_auc: 0.9752 - val_cohe
n_kappa: 0.8679 - val_f1_score: 0.8503 - val_precision: 0.9175 - val_recall: 0.9150
Epoch 28/30
153/153 [==============================] - 300s 2s/step - loss: 0.0175 - accuracy: 0.9951 - auc: 0.9999 - cohen_kappa: 0.9925
- f1_score: 0.9916 - precision: 0.9954 - recall: 0.9949 - val_loss: 0.4476 - val_accuracy: 0.9125 - val_auc: 0.9751 - val_cohe
n_kappa: 0.8633 - val_f1_score: 0.8433 - val_precision: 0.9137 - val_recall: 0.9121
Epoch 29/30
153/153 [==============================] - 304s 2s/step - loss: 0.0166 - accuracy: 0.9958 - auc: 0.9999 - cohen_kappa: 0.9934
- f1_score: 0.9925 - precision: 0.9958 - recall: 0.9957 - val_loss: 0.4543 - val_accuracy: 0.9085 - val_auc: 0.9743 - val_cohe
n_kappa: 0.8581 - val_f1_score: 0.8407 - val_precision: 0.9096 - val_recall: 0.9075
Epoch 30/30
153/153 [==============================] - 302s 2s/step - loss: 0.0173 - accuracy: 0.9954 - auc: 0.9999 - cohen_kappa: 0.9929
- f1_score: 0.9919 - precision: 0.9957 - recall: 0.9953 - val_loss: 0.4453 - val_accuracy: 0.9145 - val_auc: 0.9750 - val_cohe
n_kappa: 0.8664 - val_f1_score: 0.8493 - val_precision: 0.9168 - val_recall: 0.9136
```

Figure 13: Training and Validation Accuracy Over Epochs

4.2.3 Model Evaluation

After training, the model was evaluated on the test set, which was kept separate from the training and validation data to provide an unbiased estimate of the model's performance. The confusion matrix (Figure 16) showed the model's ability to correctly classify the four retinal disease categories, with high accuracy across CNV, DME, DRUSEN, and NORMAL.

The evaluation metrics were calculated as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

where TP, TN, FP, and FN represent the true positives, true negatives, false positives, and false negatives, respectively. In addition to accuracy, other performance metrics such as F1 Score and Cohen's Kappa were also used to assess the reliability of the model in distinguishing between closely related retinal conditions.

4.2.4 Model Prediction and Results

The trained InceptionV3 model was used to generate predictions on the test set, providing probabilities for each class. The predicted classes were compared to the true labels, and a detailed classification report was generated, outlining the precision, recall, and F1 score for each class.

ROC curves were plotted to evaluate the model's ability to distinguish between the different retinal diseases. The area under the curve (AUC) values further confirmed the model's high performance.

5 Results

The model was trained on the OCT dataset for 30 epochs using a batch size of 32. The following metrics were used to evaluate the model's performance:

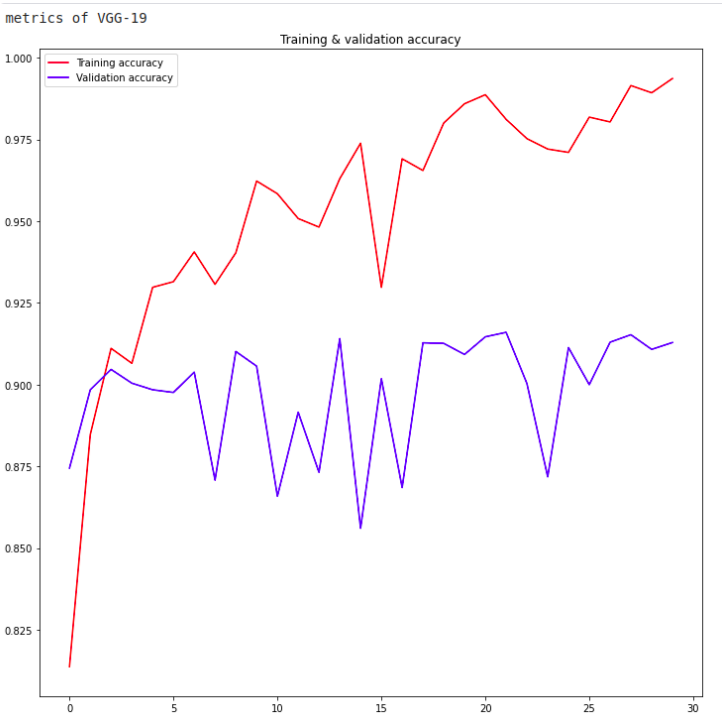


Figure 14: VCG Parameters

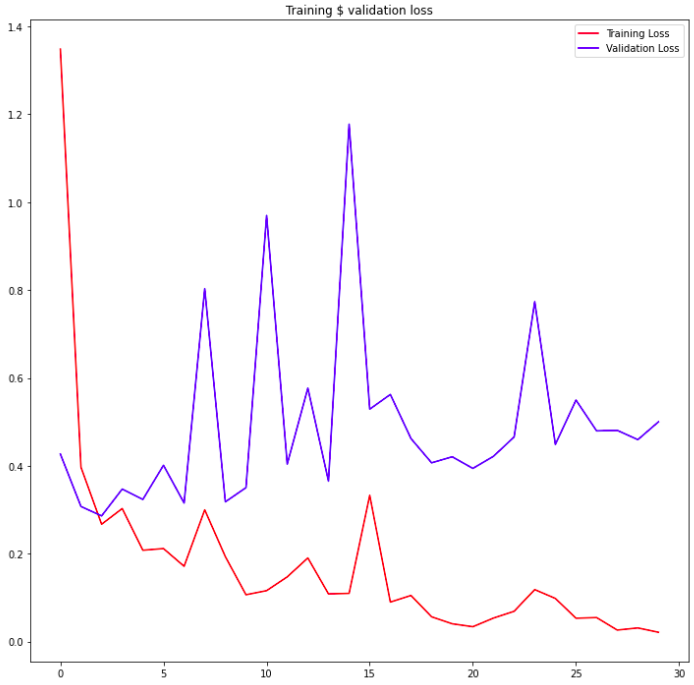
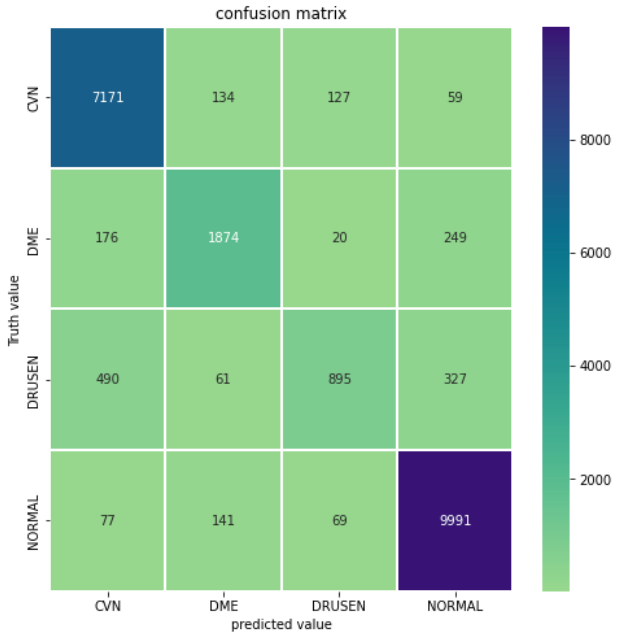


Figure 15: VCG Parameters2




```
[[7171  134  127   59]
 [ 176 1874   20  249]
 [ 490   61  895  327]
 [  77  141   69 9991]]
```

Figure 16: Confusion Matrix of Model Predictions on Test Data

- **Accuracy**: Overall accuracy of the classification.
- **Precision**: Ability of the model to correctly identify positive samples.
- **Recall**: Ability of the model to retrieve all relevant instances.
- **F1-Score**: Harmonic mean of precision and recall.

The confusion matrix (Figure 2) shows that the model performed well in distinguishing between CNV, DME, and NORMAL, though it struggled slightly with DRUSEN.

5.1 Quantitative Analysis

The performance of the model was quantified using the following equation for accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

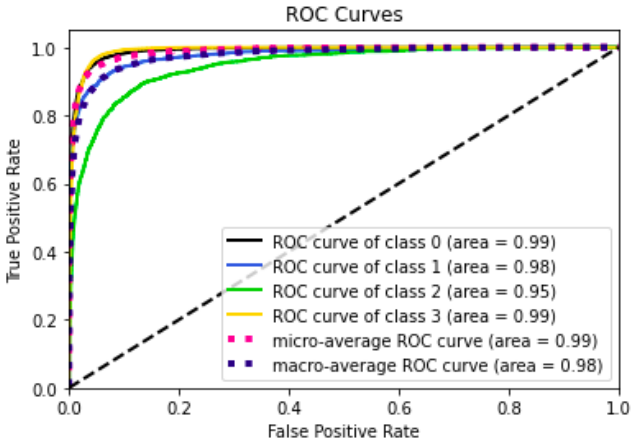


Figure 17: ROC Curve for OCT Disease Classification

	precision	recall	f1-score	support
CNV	0.91	0.96	0.93	7491
DME	0.85	0.81	0.83	2319
DRUSEN	0.81	0.50	0.62	1773
NORMAL	0.94	0.97	0.96	10278
accuracy			0.91	21861
macro avg	0.87	0.81	0.83	21861
weighted avg	0.91	0.91	0.91	21861

Figure 18: Accuracy Results

where TP is the true positives, TN is the true negatives, FP is the false positives, and FN is the false negatives. ***Net accuracy obtained over the test set is :0.911***

5.2 Challenges

One of the major challenges encountered was the class imbalance in the dataset, as certain pathologies (e.g., DME) had fewer images compared to others. Techniques such as class weighting and oversampling were applied to mitigate this issue.

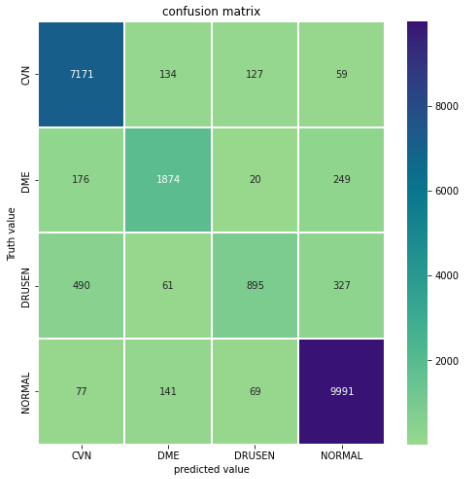


Figure 19: Confusion Matrix of Model Predictions

```
print("Accuracy", sklearn.metrics.accuracy_score(true_classes, predicted_classes))
```

```
Accuracy 0.9117149261241481
```

Figure 20: Accuracy

6 Future Work

Moving forward, several improvements are planned for the project:

- Fine-tuning the model for better generalization on unseen data.
- Incorporating more sophisticated data augmentation techniques.
- Expanding the dataset to include more diverse pathologies.
- Use of more advanced model for both classification as well as segmentation tasks
- Integrating the model into a clinical-grade software tool for real-time OCT analysis.

7 Conclusion

The results obtained indicate that deep learning models, particularly CNNs like the Inception model, hold great promise for automating the classification of retinal diseases from OCT images. Future work will focus on refining the model,developing segmentation tools and integrating it into a user-friendly tool for clinical use.

References

[1] K. Yojana and L. Thillai Rani, *OCT layer segmentation using U-NET semantic segmentation and RESNET34 encoder-decoder*, Department of Electronics and Instrumentation Engineering, Annamalai University, Chidambaram, Tamilnadu, India.

[2] S. Mukherjee, T. De Silva, P. Grisso, H. Wiley, D. L. K. Tiarnan, A. T. Thavikulwat, E. Chew, and C. Cukras, *Retinal layer segmentation in optical coherence tomography (OCT) using a 3D deep-convolutional regression network for patients with age-related macular degeneration*, n.d.

[3] D. S. Kermany, M. Goldbaum, W. Cai, et al., *Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning*, Cell, 2018.

[4] ***Mendeley Data, Large Dataset of Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images, https://data.mendeley.com/datasets/rscbjbr9sj/3.***