Progress Report: OCT Scans Disease Classification

Kristen Lee

1 Introduction

Age-related macular degeneration (AMD) and other retinal diseases are leading causes of vision impairment worldwide (1). Routine eye exams for older adults often include Optical Coherence Tomography (OCT) scans, which provide high-resolution images of retinal layers to detect signs such as swelling, abnormal growths, or thinning (2). However, interpreting OCT scans is labor-intensive and prone to variability as features that differentiate diseases are often subtle and difficult to discern even for experienced clinicians(3).

This project aims to develop a deep learning model to classify OCT scans into six categories: AMD, DME, DR, CSR, MH, and Normal. As shown in Figure 1, an OCT image is provided as the input, and the model outputs a prediction for one of 4 conditions. Automated classification can accelerate screening, improve diagnostic accuracy, and reduce manual workload for ophthalmologists. Deep learning is particularly suitable because it can learn hierarchical features from high-dimensional OCT data, capturing subtle patterns that traditional methods may miss, and generalizing to new patients regardless of race or gender, or new imaging conditions (4).

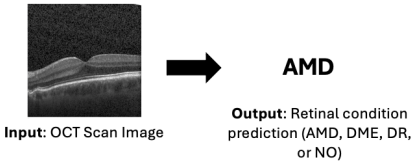

Figure 1: Cleaned and augmented sample from each class

2 Data Processing

The data were sourced from three publicly available Optical Coherence Tomography (OCT) datasets: the UCSD OCT dataset (5), the Optical Coherence Tomography (OCTDL) dataset (6), and the Retinal OCT Image Classification dataset (7). An effort was made to balance the number of samples from each source within each class; however, some datasets lacked certain disease categories or contained limited samples. Despite this, the overall number of images per class was kept balanced across the combined dataset.

A total of 6000 images were used after cleaning and preprocessing. Each of the four categories (AMD, DME, DR, and Normal) contained 1500 samples each to maintain class balance. The dataset was randomly divided into training, validation, and testing subsets with a 70% (4200 images), 15% (900 images), and 15% (900 images) split, respectively, using a fixed random seed to ensure reproducibility. An example of the cleaned, augmented, and preprocessed training samples from each class is shown in Figure 2.

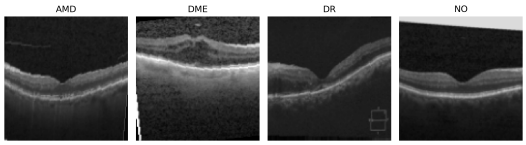

Figure 2: Cleaned and augmented sample from each class

Data Cleaning and Preprocessing
The following preprocessing and cleaning steps were applied to ensure data consistency, quality, and reproducibility:

- **Category Cleanse & Renaming:** Diseases not relevant to this project were removed from the dataset, as rare conditions would require extensive augmentation to achieve comparable sample sizes. Disease category names were also standardized across datasets. For example, CNV and Drusen were merged under the broader category AMD to maintain consistent labeling.
- **Quality Filtering:** Poor quality images and OCT scans outside the foveal region (the darker, central depression in the retina) were excluded to ensure reliable feature extraction.
- **Data Augmentation:** To improve model generalization, random rotations ($\pm 15^\circ$), horizontal flips, random resized crops (scale range 0.8–1.0), and color jittering (brightness, contrast, and saturation ± 0.2) were applied to each training sample.
- **Normalization & Resizing :** All images were converted to RGB and normalized to a $[0, 1]$ range before being transformed into tensors for PyTorch processing. All images were resized to 128×128 pixels to standardize input dimensions while preserving detail.

Plan for Testing on Unseen Data
To evaluate model generalization, new, unseen OCT images not included in the original datasets will be used. This will be achieved by either collecting new OCT scans from an open-access medical imaging repository that were not part of the three original datasets, or using a different OCT dataset that has not been previously utilized during training or validation to ensure the evaluation reflects true model robustness.

Challenges Encountered
During data processing, several challenges were encountered. First, some diseases, such as AMD, had substantially more samples than others, requiring downsampling to ensure all four classes were represented equally (1500 samples each). Second, not all datasets included the same disease categories; for example, only the Retinal OCT Kaggle dataset contained DME and DR, making perfect balance across sources difficult, though overall class representation was maintained.

3 Baseline Model: Random Forest

As a baseline for OCT image classification, a Random Forest (RF) classifier was implemented using features extracted from a pretrained ResNet-18 model (Architecture in Figure 3). The ResNet-18 network, with its final classification layer removed, was used purely as a feature extractor to generate fixed-length embeddings for each image. These embeddings were then fed into the RF classifier, which required minimal tuning (100 trees, default hyperparameters) and was straightforward to implement.

On the test set, the Random Forest achieved an accuracy of 0.75. Class-wise performance (from the classification report) shows that DR was the easiest to classify (F1-score 0.90), while DME was the most challenging (F1-score 0.66). Precision and recall values indicate that AMD and NO were moderately well classified, with some misclassifications occurring primarily between DME and NO (shown in Figure 4) Confusion Matrix. Qualitatively, the model correctly identifies clear and typical images for each class. These results demonstrate that features from a pretrained CNN contain sufficient information to separate the classes reasonably well, providing a reference point to compare the primary neural network model. A key challenge in this task was ensuring that the extracted features were sufficiently representative across all classes, especially given the class imbalance and subtle differences between some retinal diseases, which occasionally led to misclassifications.

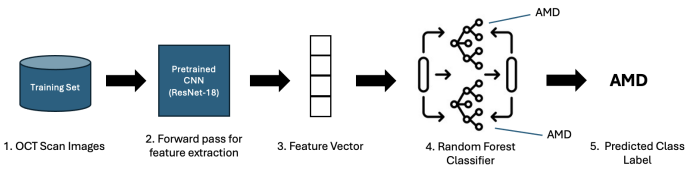

Figure 3: Random Forest Model Architecture

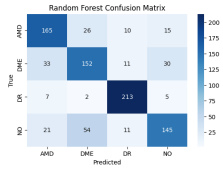

Figure 4: Random Forest confusion matrix

4 Architecture

The primary model used for OCT classification is a custom convolutional neural network (CNN), `OCTCNN`, designed to classify OCT images into four categories: AMD, DME, DR, and Normal. The network takes RGB images of size $128 \times 128 \times 3$ as input and passes them through two convolutional layers. The first layer has 32 filters of size 3×3 with padding=1, followed by ReLU activation and a 2×2 max-pooling layer, reducing the spatial dimensions from 128 to 64. The second layer has 48 filters of size 3×3, followed by ReLU activation and two successive 2×2 max-pooling layers, reducing the spatial size from 64 to 16. The feature maps are then flattened and passed through a fully connected layer with 256 units and ReLU activation, followed by a dropout layer (p=0.5) for regularization, and finally through a fully connected layer producing 4 outputs corresponding to the classes. See architecture diagram below (Figure 5).

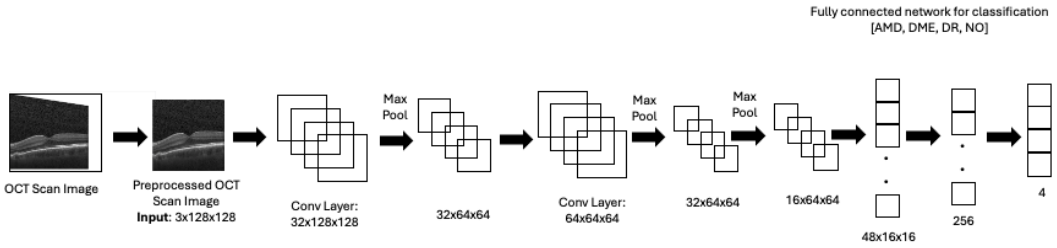

Figure 5: Proposed Model for Disease Classification From OCT Scans

The model has approximately 3.1 million trainable parameters, making it moderately complex and appropriate for the dataset and image resolution. Training was performed using SGD with momentum 0.9, a learning rate of 0.0005, batch size 128, and cross-entropy loss over 15 epochs. After training, the model achieved a final training accuracy of 78%, validation accuracy of 76%, and test accuracy of 76%, with an average training loss of 1.114. Qualitatively, the model distinguished DR images more effectively than DME or AMD early in training, and misclassifications were most common between AMD and Normal classes, reflecting subtle differences in retinal features. Key challenges included balancing the dataset to prevent bias, avoiding overfitting given the moderate dataset size, and tuning the architecture to achieve reasonable accuracy without excessive computational cost.

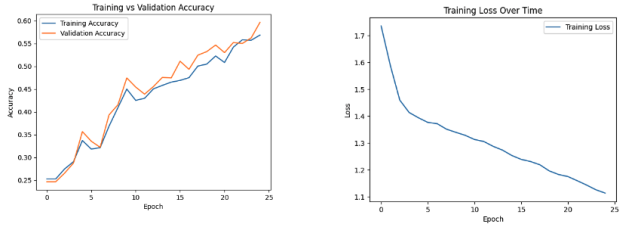

Figure 6: Training vs Validation Accuracy, and Training Loss over 15 epochs

References

[1] Y. D. Jeong *et al.*, "Global burden of vision impairment due to age-related macular degeneration, 1990–2021, with forecasts to 2050: a systematic analysis for the global burden of disease study 2021," *The Lancet Global Health*, vol. 13, pp. e1175–e1190, July 2025.

[2] E. S. Enaholo, M. J. Musa, and M. Zeppieri, *Optical Coherence Tomography*. StatPearls Publishing, 2024. Last Update: October 6, 2024.

[3] M. Dahrouj and J. B. Miller, "Artificial intelligence (ai) and retinal optical coherence tomography (oct)," *Seminars in Ophthalmology*, vol. 36, no. 5-6, pp. 341–345, 2021.

[4] C. S. Lee, D. M. Baughman, and A. Y. Lee, "Deep learning is effective for the classification of oct images of normal versus age-related macular degeneration," *Ophthalmology Retina*, vol. 1, pp. 322–327, Jul-Aug 2017.

[5] Mmazizi, "Ucsd 3-class labeled retinal oct images," 2021. Accessed: 2025-09-26.

[6] S. M. Shuvo, "Octdl: Optical coherence tomography dataset for image-based deep learning method," 2021. Accessed: 2025-09-26.

[7] O. Inaren, "Retinal oct image classification - c8," 2021. Accessed: 2025-09-26.