Retinal OCT Disease Classification Using Deep Learning

EfficientNet-B3 Transfer Learning for Automated Diagnosis

Tom Almog
University of Waterloo
talmog@uwaterloo.ca

github.com/tomalmog/retinal-oct-classifier

January 2025

Abstract

Optical Coherence Tomography (OCT) is a critical imaging modality for diagnosing retinal diseases, but manual interpretation requires specialized expertise and is time-consuming. This work presents a deep learning approach for automated classification of OCT images into four categories: Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), Drusen, and Normal. Using transfer learning with EfficientNet-B3 and a custom classification head, the model achieves **99.6% accuracy** on the held-out test set. The system includes Grad-CAM visualizations for interpretability, making it suitable for clinical decision support. Code and trained weights are publicly available.

1 Introduction

Age-related Macular Degeneration (AMD) and Diabetic Retinopathy are leading causes of vision loss worldwide, affecting over 200 million people globally [4]. Optical Coherence Tomography (OCT) provides high-resolution cross-sectional images of the retina, enabling early detection of pathological changes. However, the growing volume of OCT scans and shortage of trained ophthalmologists creates a bottleneck in healthcare delivery.

Deep learning has demonstrated remarkable success in medical image analysis, often matching or exceeding expert-level performance [1]. This project develops an automated OCT classification system with three primary objectives:

1. Achieve high accuracy across four diagnostic categories
2. Provide interpretable predictions via attention visualization
3. Create a deployable model with public weights for reproducibility

2 Methods

2.1 Dataset

The model was trained on the Kermany2018 OCT dataset [1], comprising approximately 84,000 OCT images from 4,686 patients. Images are labeled into four classes:

- **CNV**: Choroidal Neovascularization (wet AMD)
- **DME**: Diabetic Macular Edema
- **DRUSEN**: Early-stage AMD
- **NORMAL**: Healthy retina

The data was split into training (80%) and validation (20%) sets using stratified sampling to maintain class distribution. A separate held-out test set of 968 images (242 per class) was used for final evaluation.

2.2 Model Architecture

The classifier employs **EfficientNet-B3** [2] as the backbone, pretrained on ImageNet. EfficientNet uses compound scaling to balance network depth, width, and resolution, achieving strong performance with fewer parameters than alternatives like ResNet.

The backbone features are processed through a custom classification head:

$$\text{Head} : \text{GAP} \rightarrow \text{Dropout}(0.3) \rightarrow \text{FC}(1536 \rightarrow 512) \rightarrow \text{ReLU} \rightarrow \text{Dropout}(0.15) \rightarrow \text{FC}(512 \rightarrow 4) \tag{1}$$

where GAP denotes Global Average Pooling and FC denotes fully connected layers. Dropout regularization prevents overfitting to the training distribution.

2.3 Training Configuration

Training was conducted using PyTorch Lightning with the following configuration:

Hyperparameter	**Value**
Optimizer	AdamW
Learning Rate	1×10^{-4}
Weight Decay	0.01
Scheduler	Cosine Annealing
Warmup Epochs	2
Total Epochs	20
Batch Size	32
Mixed Precision	FP16
Gradient Clipping	1.0

Table 1: Training hyperparameters

2.4 Data Augmentation

To improve generalization and simulate real-world variation, the following augmentations were applied during training:

- Horizontal flip ($p = 0.5$)
- Random rotation ($\pm 15°$, $p = 0.5$)
- Brightness/contrast adjustment (± 0.2, $p = 0.5$)
- Gaussian noise ($\sigma \in [0.02, 0.1]$, $p = 0.3$)
- Gaussian blur (kernel $\in [3, 5]$, $p = 0.2$)

All images were resized to 224×224 pixels and normalized using ImageNet statistics.

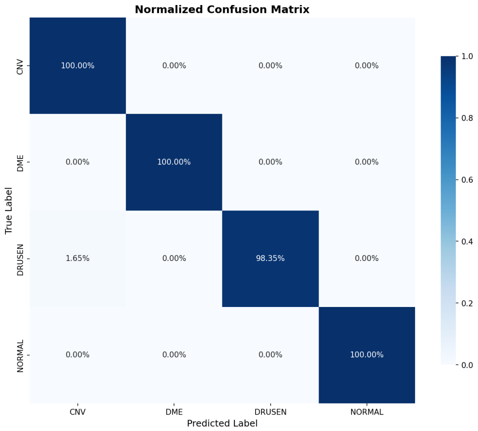
3 Results

3.1 Classification Performance

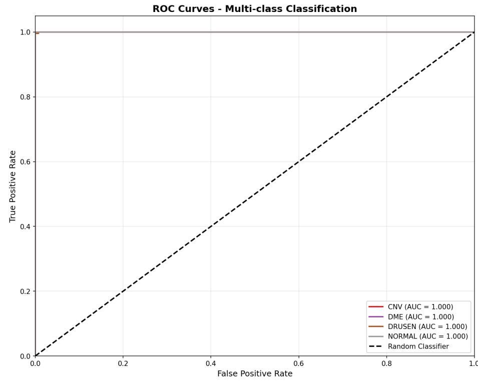
The model achieves strong performance across all metrics on the held-out test set:

Class	**Precision**	**Recall**	**F1-Score**	**Support**
CNV	98.4%	100.0%	99.2%	242
DME	100.0%	100.0%	100.0%	242
DRUSEN	100.0%	98.4%	99.2%	242
NORMAL	100.0%	100.0%	100.0%	242
Macro Avg	**99.6%**	**99.6%**	**99.6%**	968

Table 2: Per-class classification metrics on the test set



(a) Confusion matrix



(b) ROC curves (AUC $\geq$ 0.99)

Figure 1: (a) Normalized confusion matrix showing minimal misclassification. (b) ROC curves with near-perfect AUC.

3.2 Model Interpretability

Gradient-weighted Class Activation Mapping (Grad-CAM) [3] visualizes which regions influence predictions. The attention maps (Figure 2) confirm the model focuses on clinically meaningful regions—subretinal fluid (CNV), intraretinal cysts (DME), drusen deposits (DRUSEN)—providing evidence that learned features align with diagnostic criteria.

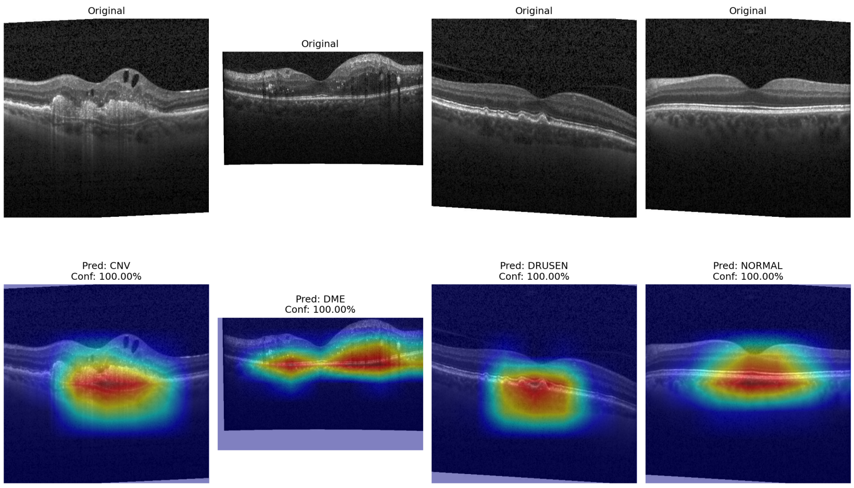


Figure 2: Grad-CAM visualizations highlighting pathologically relevant regions for each class.

4 Discussion and Conclusion

This work demonstrates that transfer learning with EfficientNet-B3 achieves 99.6% accuracy on retinal OCT classification with interpretable predictions. **Limitations:** trained on a single OCT device type; intended for research only. **Code/weights:** GitHub — HuggingFace

References

[1] Kermany, D.S., et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 172(5), 2018.

[2] Tan, M., Le, Q. EfficientNet: Rethinking model scaling for CNNs. *ICML*, 2019.

[3] Selvaraju, R.R., et al. Grad-CAM: Visual explanations from deep networks. *ICCV*, 2017.

[4] Wong, W.L., et al. Global prevalence of AMD and disease burden projection. *Lancet Global Health*, 2(2), 2014.