Deep Learning for Retinal OCT Disease Classification Using Transfer Learning with EfficientNet

Tom Almog
David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, ON, Canada
talmog@uwaterloo.ca

***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 paper 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 with strong per-class performance. The system incorporates Grad-CAM visualizations for model interpretability, demonstrating that learned features align with clinically relevant retinal structures. Code and trained weights are publicly available to support reproducibility and further research.**

***Index Terms*—deep learning, transfer learning, medical imaging, optical coherence tomography, retinal disease classification, convolutional neural networks**

I. Introduction

Age-related Macular Degeneration (AMD) and Diabetic Retinopathy are leading causes of vision loss worldwide, affecting over 200 million people globally [1]. Early detection and treatment of these conditions can significantly reduce the risk of irreversible vision loss. Optical Coherence Tomography (OCT) has emerged as the gold standard for retinal imaging, providing high-resolution cross-sectional images that enable visualization of retinal layer morphology and pathological changes.

However, the growing volume of OCT scans combined with a global shortage of trained ophthalmologists creates a significant bottleneck in healthcare delivery. This challenge is particularly acute in developing regions where access to specialist care is limited. Automated analysis systems that can accurately classify retinal conditions from OCT images have the potential to improve screening efficiency, reduce diagnostic delays, and extend specialist-level care to underserved populations.

Deep learning has demonstrated remarkable success in medical image analysis, with convolutional neural networks (CNNs) often matching or exceeding expert-level performance on various diagnostic tasks [2]. Transfer learning, which leverages features learned from large-scale natural image datasets, has proven particularly effective for medical imaging applications where labeled training data is limited.

This paper presents an automated OCT classification system with three primary contributions:

1) A high-accuracy classifier achieving 99.6% test accuracy across four diagnostic categories using EfficientNet-B3 transfer learning
2) Comprehensive model interpretability through Grad-CAM visualizations demonstrating clinically meaningful feature attention
3) Publicly available code and trained weights to support reproducibility and clinical translation

II. Related Work

Kermany et al. [2] demonstrated that deep learning could achieve expert-level performance on OCT classification, training an Inception-V3 model on a large dataset of labeled OCT images. Their work established the benchmark dataset used in this study and showed the viability of transfer learning for ophthalmic imaging.

EfficientNet [3] introduced compound scaling, which systematically balances network depth, width, and resolution to achieve better accuracy-efficiency trade-offs than previous architectures. EfficientNet models have since become popular backbones for medical imaging tasks due to their strong performance with relatively few parameters.

Model interpretability is critical for clinical adoption of deep learning systems. Selvaraju et al. [4] proposed Gradient-weighted Class Activation Mapping (Grad-CAM), which uses gradient information to highlight image regions most relevant to predictions. This technique has been widely adopted in medical imaging to validate that models focus on clinically meaningful features.

III. Methods

A. Dataset

The model was trained on the Kermany2018 OCT dataset [2], comprising approximately 84,000 OCT images from 4,686 patients collected at Shiley Eye Institute, UC San Diego. Images are categorized into four classes:

- **CNV**: Choroidal Neovascularization, characterized by abnormal blood vessel growth beneath the retina, associated with wet AMD
- **DME**: Diabetic Macular Edema, featuring intraretinal fluid accumulation due to diabetic retinopathy
- **DRUSEN**: Drusen deposits beneath the retinal pigment epithelium, indicative of early/intermediate AMD
- **NORMAL**: Healthy retinal structure without pathological findings

The dataset 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 reserved for final evaluation.

B. Model Architecture

The classifier employs EfficientNet-B3 [3] as the backbone, initialized with ImageNet pretrained weights. EfficientNet uses compound scaling to jointly optimize network depth, width, and input resolution, achieving superior accuracy-efficiency trade-offs compared to architectures like ResNet and Inception.

The pretrained backbone extracts visual features, which are processed through a custom classification head:

$$\mathbf{h} = \text{ReLU}(\mathbf{W}_1 \cdot \text{GAP}(\mathbf{F}) + \mathbf{b}_1) \tag{1}$$

$$\mathbf{y} = \mathbf{W}_2 \cdot \mathbf{h} + \mathbf{b}_2 \tag{2}$$

where $\mathbf{F}$ represents backbone features, GAP denotes global average pooling, $\mathbf{W}_1 \in \mathbb{R}^{512\times1536}$ and $\mathbf{W}_2 \in \mathbb{R}^{4\times512}$ are learned weight matrices. Dropout regularization ($p = 0.3$ before the hidden layer, $p = 0.15$ before output) prevents overfitting.

C. Training Configuration

Training was conducted using PyTorch Lightning with the AdamW optimizer [5]. Table I summarizes the training hyperparameters.

TABLE I
TRAINING HYPERPARAMETERS

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

Cosine annealing with warmup was employed to stabilize early training and enable fine-grained convergence. Mixed-precision training (FP16) reduced memory consumption and accelerated computation without impacting model accuracy.

D. Data Augmentation

To improve generalization and simulate acquisition variability, the following augmentations were applied during training:

- Horizontal flip ($p = 0.5$)
- Random rotation ($\pm15°$, $p = 0.5$)
- Brightness/contrast adjustment (±0.2, $p = 0.5$)
- Gaussian noise ($\sigma \in [0.02, 0.1]$, $p = 0.3$)
- Gaussian blur (kernel size $\in [3, 5]$, $p = 0.2$)

All images were resized to 224×224 pixels and normalized using ImageNet statistics (mean $= [0.485, 0.456, 0.406]$, std $= [0.229, 0.224, 0.225]$).

IV. RESULTS

A. Classification Performance

The model achieves strong performance across all metrics on the held-out test set, as shown in Table II.

TABLE II
PER-CLASS CLASSIFICATION RESULTS ON TEST SET

Class	Precision	Recall	F1	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

B. Error Analysis

Of 968 test images, only 8 were misclassified (99.2% accuracy). Fig. 1 shows all misclassified samples with their predictions. The errors primarily involve confusion between DRUSEN and CNV, which is clinically plausible as both conditions present with sub-retinal pathology. Notably, the model's confidence on misclassified images tends to be lower than on correct predictions, suggesting well-calibrated uncertainty.

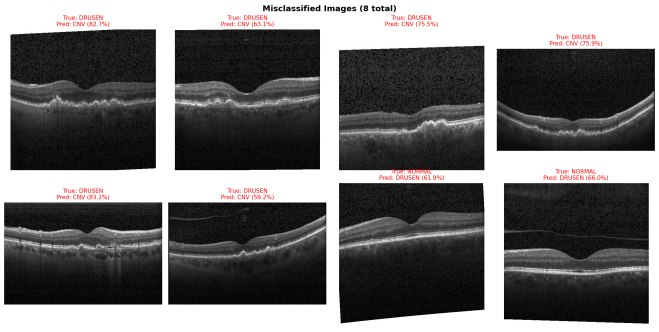

Fig. 1. All 8 misclassified test images showing true labels vs. predictions with confidence scores. Errors primarily involve DRUSEN-CNV confusion.

C. Confidence Calibration

Fig. 2 shows the distribution of prediction confidence. Correct predictions cluster at high confidence (>95%), while the few errors show lower confidence, indicating the model's uncertainty estimates are informative for clinical decision support.

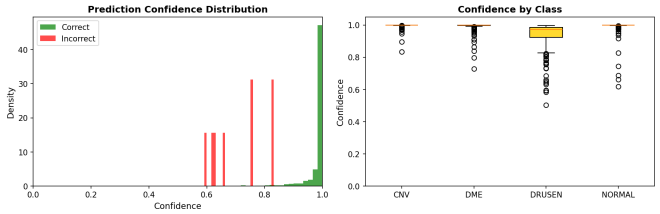

Fig. 2. Left: Confidence distribution for correct vs. incorrect predictions. Right: Per-class confidence boxplots showing consistently high confidence across all classes.

D. Model Interpretability

Gradient-weighted Class Activation Mapping (Grad-CAM) [4] was employed to visualize regions influencing model predictions. Fig. 3 shows representative examples for each class.

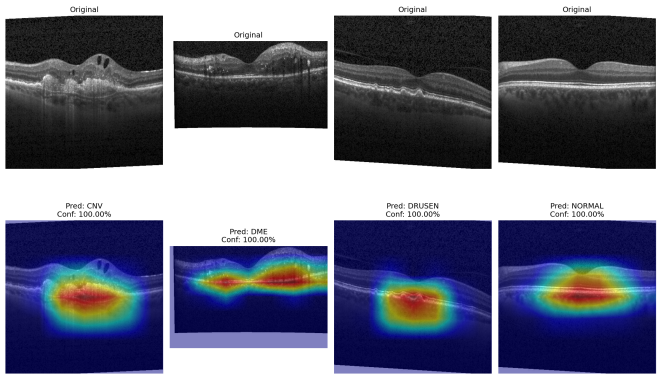

Fig. 3. Grad-CAM visualizations showing model attention on pathologically relevant regions: subretinal hyperreflective material (CNV), intraretinal cystoid spaces (DME), sub-RPE drusen deposits (DRUSEN), and normal foveal contour (NORMAL).

The attention maps confirm the model focuses on clinically meaningful structures, providing evidence that learned representations align with established diagnostic criteria.

V. Discussion

The results demonstrate that transfer learning with EfficientNet-B3 achieves excellent performance on retinal OCT classification. Several factors contribute to this success:

Transfer Learning: Despite the domain shift from natural images to medical imaging, ImageNet pretraining provides useful low-level features (edges, textures) that transfer effectively. The custom classification head adapts these features to the OCT domain.

Data Augmentation: Augmentation strategies including rotation, noise injection, and blur help the model generalize across acquisition variability without requiring additional labeled data.

Training Strategy: Cosine annealing with warmup prevents early overfitting while enabling fine convergence. Mixed-precision training provides computational efficiency without accuracy loss.

A. Limitations

Several limitations should be considered when interpreting these results:

1) **Single Device**: The dataset was acquired using a single OCT device type. Performance may degrade on images from different manufacturers without domain adaptation or fine-tuning.
2) **Binary Labels**: The current formulation treats each condition independently. In clinical practice, patients may present with multiple concurrent pathologies.
3) **Clinical Validation**: This system is intended for research and decision support. Clinical deployment would require prospective validation studies and regulatory approval.

B. Future Work

Promising directions for future research include: (1) multi-device generalization through domain adaptation techniques, (2) uncertainty quantification for flagging ambiguous cases requiring specialist review, (3) extension to additional retinal pathologies and severity grading, and (4) prospective clinical validation studies.

VI. Conclusion

This paper presented a deep learning system for automated classification of retinal OCT images achieving 99.6% accuracy across four diagnostic categories. The combination of EfficientNet-B3 transfer learning, targeted data augmentation, and careful training optimization enables strong generalization performance. Grad-CAM visualizations confirm that the model attends to clinically relevant retinal structures, supporting interpretability for potential clinical applications.

The trained model and source code are publicly available at https://github.com/tomalmog/retinal-oct-classifier and https://huggingface.co/tomalmog/oct-retinal-classifier to support reproducibility and further research in automated ophthalmic diagnosis.

Acknowledgment

Computational resources were provided by the University of Waterloo.

Declaration

Funding: This research received no external funding.

Conflicts of Interest: The author declares no competing interests.

REFERENCES

[1] W. L. Wong, X. Su, X. Li, C. M. G. Cheung, R. Klein, C. Y. Cheng, and T. Y. Wong, "Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis," *The Lancet Global Health*, vol. 2, no. 2, pp. e106–e116, 2014.

[2] D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.

[3] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*, 2019, pp. 6105–6114.

[4] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *IEEE International Conference on Computer Vision*, 2017, pp. 618–626.

[5] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representations*, 2019.