Jack Mazac

AI In Healthcare

April 6, 2025

TRANSFER LEARNING APPROACH FOR BINARY CLASSIFICATION OF RETINAL FUNDUS IMAGES

## ABSTRACT

This paper presents a deep learning approach for the binary classification of retinal fundus images into normal and abnormal categories. Using transfer learning with a pre-trained EfficientNet-B0 architecture, I developed a model capable of achieving 88.89% validation accuracy and 0.8000 F1 score despite the constraints of a small dataset. The implementation utilized PyTorch with hardware acceleration via Apple Metal Performance Shaders. My findings demonstrate the effectiveness of transfer learning for medical image classification tasks with limited data availability, while also highlighting challenges and opportunities for future improvements.

## INTRODUCTION

Automated classification of retinal fundus images represents an important advancement in ophthalmology, with potential applications in early disease detection and screening programs. Distinguishing between normal and abnormal retinal images is a fundamental binary classification task that serves as a precursor to more specific pathology identification. Deep learning approaches have shown significant promise in this domain, yet training robust models typically requires large datasets that are often unavailable in specialized medical contexts.

This work explores the application of transfer learning using a pre-trained EfficientNet-B0 model to overcome the limitations imposed by a small retinal image dataset. By leveraging knowledge gained from pre-training on the ImageNet dataset, I hypothesized that the model could extract relevant features for retinal image classification with minimal fine-tuning, providing a viable approach for practical implementation in resource-constrained environments.

## MATERIALS AND METHODS

### Dataset

The dataset consisted of 46 retinal fundus images categorized into two classes: normal (16 images) and abnormal (30 images). This class imbalance reflects the typical distribution encountered in clinical settings, where abnormal findings often outnumber normal cases in curated datasets. We employed an 80/20 split for training and validation, resulting in 36 images for training and 9 for validation. While this dataset size is limited, it represents a realistic scenario frequently encountered in specialized medical imaging applications.

## Model Architecture

I utilized an EfficientNet-B0 architecture pre-trained on ImageNet as my base model. This choice was motivated by EfficientNet's demonstrated ability to balance model size and performance through compound scaling. The model was adapted for binary classification by modifying the final classification layer to output two classes (normal and abnormal). Implementation was achieved through the PyTorch framework with the timm (PyTorch Image Models) library providing access to the pre-trained model weights.

## Training Methodology

My approach centered on transfer learning to leverage the feature extraction capabilities of the pre-trained network. The image preprocessing pipeline included resizing all images to 224×224 pixels and normalizing using ImageNet mean and standard deviation values (means: [0.485, 0.456, 0.406], standard deviations: [0.229, 0.224, 0.225]). To address the limited dataset size, I implemented data augmentation in the form of random horizontal flips with 0.5 probability.

Training parameters were configured as follows: Adam optimizer with a learning rate of 1e-4, Cross-Entropy Loss function, batch size of 8, and training for 10 epochs. I utilized Apple Metal Performance Shaders (MPS) for GPU acceleration on macOS, which significantly reduced training time compared to CPU-only execution. Model selection was performed based on validation F1 score, with the best-performing model saved for final evaluation and inference.

## RESULTS

Training progression over 10 epochs revealed several notable patterns. The training loss generally decreased from an initial value of 1.6134 to 0.0001 in the final epoch, indicating successful optimization on the training set. Validation accuracy peaked at 88.89% during epochs 5-9, with a corresponding F1 score of 0.8000. Interestingly, despite reaching its lowest training loss in epoch 10 (0.0001), the model exhibited a significant drop in F1 score to 0.4375, strongly suggesting the onset of overfitting.

The best model performance was achieved at epoch 5 with:

- Validation Accuracy: 88.89%

- F1 Score: 0.8000

- Training Loss: 0.6804

This model was saved as my final model for inference on new retinal images. The performance plateau observed between epochs 5-9 suggests that the model reached its optimal generalization capability around epoch 5, with further training providing minimal benefits and eventually leading to degraded performance on the validation set.

## DISCUSSION

The convergence pattern observed during training provides several insights into the model's learning dynamics. The rapid improvement and subsequent plateau in validation metrics by epoch 5 indicate that the transfer learning approach effectively leveraged pre-trained features, requiring minimal fine-tuning to adapt to retinal image classification. This observation supports the hypothesis that features learned on natural images can transfer effectively to specialized medical imaging tasks.

The sharp decline in F1 score in epoch 10, despite continued reduction in training loss, clearly demonstrates the overfitting phenomenon commonly encountered with small datasets. This highlights the importance of early stopping and proper validation strategies when working with limited data. The consistently high validation accuracy (88.89%) across epochs 5-9 suggests a robust plateau in performance, rather than random fluctuations, indicating that the model found a stable local optimum.

The EfficientNet-B0 architecture proved well-suited for this task, offering a good balance between model complexity and performance. The relatively small model size compared to larger architectures facilitated training on consumer hardware while still achieving competitive results.

## LIMITATIONS

Several limitations should be acknowledged. First, the small dataset size (46 images total) inherently limits the model's ability to generalize to diverse retinal presentations. Second, the class imbalance (30 abnormal vs. 16 normal) may bias the model toward the majority class, potentially affecting its performance on normal cases. Third, the simple 80/20 train-validation split provides only a single validation set evaluation; a more robust k-fold cross-validation approach would yield more reliable performance estimates.

The binary classification approach, while useful as a screening tool, provides limited diagnostic value compared to multi-class classification of specific pathologies. Additionally, the lack of an independent test set separate from the training and validation data means that my reported performance metrics may not fully represent the model's generalization capability to entirely new data.

## FUTURE WORK

Several avenues for improvement warrant exploration. First, expanding the dataset size through additional data collection or collaboration with other institutions would likely improve model performance and generalization. Second, implementing more diverse data augmentation techniques beyond horizontal flips—such as rotations, color shifts, and elastic deformations—could further enhance the model's robustness to variations in image acquisition.

Model ensembling approaches, combining predictions from multiple architectures or training runs, could potentially improve classification reliability. Additionally, incorporating explainability techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM) would enhance the interpretability of model decisions, a critical consideration for clinical applications.

Future work should also explore the development of multi-class models capable of identifying specific retinal pathologies, rather than the current binary normal/abnormal classification. This would increase the diagnostic utility of the system while providing more granular information to clinicians.

## CONCLUSION

This study demonstrates the feasibility of developing an effective retinal image classification system using transfer learning with a pre-trained EfficientNet-B0 model, even with a small dataset. The achieved 88.89% validation accuracy and 0.8000 F1 score represent promising performance for a binary classification task in this domain. The use of transfer learning, coupled with hardware acceleration via Apple's MPS, enabled efficient model development without extensive computational resources.

My findings highlight both the potential and limitations of deep learning approaches in specialized medical imaging contexts with data constraints. While the results are encouraging, they also underscore the importance of addressing dataset limitations, implementing robust validation strategies, and incorporating explainability features for practical clinical deployment. This work provides a foundation for future research into more sophisticated retinal image analysis systems with potential applications in screening and diagnostic support.

## REFERENCES

1. Tan, M., & Le, Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning.

2. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision, 115(3), 211-252.

3. Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., ... & Webster, D. R. (2016). Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA, 316(22), 2402-2410.

4. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.