RETINAL OCT IMAGE CLASSIFICATON USING CNN

Sumedha Sanjeev Galgali
Master of Data Analytics,
Dept. of Science
The University of Western, Ontario
sgalgali@uwo.ca

Ritika Pandey
Master of Data Analytics,
Dept. of Science
The University of Western, Ontario
rpande6@uwo.ca

Garima Gambhir
Master of Data Analytics,
Dept. of Science
The University of Western, Ontario
ggambhi@uwo.ca

Abstract - This project focuses on developing an automated deep learning-based system for classifying Optical Coherence Tomography (OCT) images into four categories: Normal, Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), and Drusen. OCT imaging plays a critical role in diagnosing retinal conditions; however, manual interpretation is labor-intensive and susceptible to errors. To address these challenges, the study utilized a dataset of 84,495 labeled OCT images, applying preprocessing and data augmentation techniques such as normalization, rotation, and zooming.

Two model architectures were explored: a pre-trained ResNet50 model and a custom Convolutional Neural Network (CNN). The ResNet50 model, leveraging pre-trained features, achieved superior performance with a test accuracy of 99.34% and a test loss of 0.033, outperforming the custom CNN model, which attained a test accuracy of 97.21% and a test loss of 0.106. Class-wise evaluation revealed higher precision, recall, and F1-scores for ResNet50, particularly in categories with overlapping features, such as CNV and DME. ResNet50's robustness was further validated through confusion matrix analysis and ROC curve evaluation, which showed AUC values close to 1.0 across all categories.

The results demonstrate the effectiveness of ResNet50 in delivering accurate and reliable automated diagnoses, supporting ophthalmologists with a second-opinion tool for retinal disease detection. Future work will focus on optimizing hyperparameters, integrating ensemble techniques, and enhancing model scalability to improve diagnostic efficiency and accessibility, especially in resource-constrained settings.

I. INTRODUCTION

Retinal diseases are a significant cause of early vision loss, impacting the retina, a thin, sensitive tissue layer at the back of the eye. Disorders like Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), and Drusen can severely impair eyesight if they are not diagnosed and treated early. This study focuses on leveraging transfer learning to develop automated methods for identifying these conditions.

The task of identifying retinal diseases from Optical Coherence Tomography (OCT) images can be approached as a machine-learning classification problem. This research categorizes OCT images into four distinct classes: CNV, DME, Drusen, and Normal. The growing use of OCT imaging in clinical settings underscores the need for computer-aided diagnostic tools, which can improve the precision and speed of disease detection, ultimately supporting better treatment decisions. Figure 1 presents examples of OCT images for these categories.

Traditional techniques for automating retinal diagnostics often rely on heavy pre-processing steps and shallow neural networks, which can be inefficient and time-consuming. This study adopts transfer learning to address these challenges, employing deep-learning models to streamline the process. By implementing these advanced methods, the research seeks to enhance the accuracy and efficiency of detecting retinal diseases, paving the way for more reliable diagnostic systems.

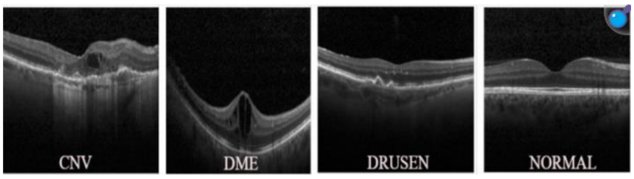


Figure 1

II. BACKGROUND

The integration of machine learning (ML) and artificial intelligence (AI) into medical imaging has revolutionized the field, enabling more efficient analysis and interpretation of complex datasets. AI-powered systems can process large volumes of medical images, identify patterns, and accurately diagnose diseases. Convolutional neural networks (CNNs), which excel at extracting layered features from images, have recently achieved exceptional results in image classification tasks. These strengths make CNNs particularly valuable for medical applications, including the analysis of Optical Coherence Tomography (OCT) images.

Automating diagnostic processes with AI offers several benefits, such as reducing the workload of medical professionals and ensuring consistent, accurate results. These systems are highly scalable, making them ideal for handling extensive datasets and improving access to diagnostic tools in underserved or remote areas. This project explores advanced deep-learning techniques and optimization methods to enhance the performance of automated systems for retinal disease detection, focusing on increasing the reliability and accuracy of disease classification.

The study aims to address the growing demand for more effective retinal disease diagnosis by leveraging AI-driven technologies. By improving diagnostic efficiency, this work has the potential to enhance patient outcomes. Additionally, it represents a step forward in medical imaging research, paving the way for the adoption of AI-assisted diagnostic tools in clinical environments.

III. LITERATURE REVIEW

The use of deep learning in analyzing retinal Optical Coherence Tomography (OCT) images has gained prominence for its potential to enhance diagnostic accuracy in ophthalmology. Transfer learning, particularly with architectures like ResNet-50, has demonstrated effectiveness in leveraging pre-trained models to extract complex features from OCT datasets, even with limited annotated data. Studies highlight the benefits of data preprocessing, such as normalization and augmentation, in improving model generalization and addressing class imbalances.

Custom Convolutional Neural Networks (CNNs) have also been explored for their simplicity and efficiency in smaller datasets, though they often underperform compared to pre-trained models on complex medical imaging tasks. Metrics like accuracy, precision, and recall are standard for evaluating model efficacy, with transfer learning models consistently achieving higher performance.

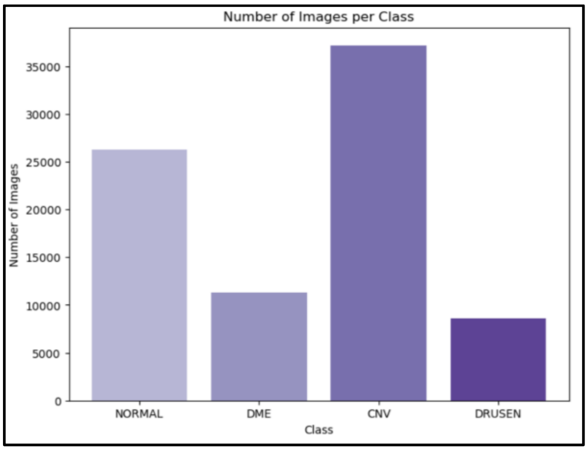
This research builds on these advancements by comparing ResNet-50 and a custom CNN for classifying retinal conditions, demonstrating the superior capability of deeper transfer learning architectures in medical imaging applications.

IV. METHODS

V. DATASET DESCRIPTION

The dataset used in this project consists of 84,495 retinal Optical Coherence Tomography (OCT) images, which are organized into four categories: Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), Drusen, and Normal. These categories represent different retinal conditions, with the Normal category consisting of healthy retinal images that serve as a baseline for comparison against the other conditions.

Four types of retinal diseases are included in the dataset: Normal, Drusen, Diabetic Macular Edema (DME), and Choroidal Neovascularization (CNV). Vision impairment is frequently caused by CNV, which involves aberrant blood vessel development beneath the retina. DME is the term for retinal edema brought on by blood vessel fluid leakage, which is frequently brought on by diabetes. Drusen, which are yellow deposits beneath the retina, are usually linked to the initial phases of macular degeneration. Healthy retinal scans are included in the Normal category to provide a reference point for comparing the other conditions.



The dataset is further divided into three main subsets to support the model training and evaluation process. The Training set includes the largest portion of the data and is used to train the model. This set allows the model to learn the relevant features and patterns associated with each condition. The Validation set contains a separate group of images used during the training process to monitor the model's performance and prevent overfitting. This set helps in fine-tuning hyperparameters and adjusting the model's behavior as needed. Finally, the Testing set comprises unseen images, which are used to assess the model's performance after training. This subset ensures an unbiased evaluation of the model's ability to generalize to new data.

By using a dataset with sufficient representation of each category, the project ensures that the model can accurately classify OCT images, contributing to the development of an automated system for efficient and reliable retinal disease detection.

VI. DATA PREPROCESSING

To optimize the dataset for deep learning and improve model outcomes, a series of preprocessing steps were applied.

The OCT images were grouped into four categories: Normal, Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), and Drusen. These were organized into three subsets: Training, Validation, and Testing. The Training subset contained the largest portion of data and was used to train the model, while the Validation subset provided feedback on performance during training. The Testing subset was set aside exclusively for evaluating how well the model performed on unseen data. To ensure computational efficiency and consistency, all images were resized to 128x128 pixels, a dimension that preserved critical retinal details.

Normalization was employed by scaling pixel intensities to a uniform range, enabling faster and more stable training. To further diversify the training data and enhance the model's ability to generalize, several data augmentation techniques were applied. These included rescaling to standardize image formats, random shifts in width and height (up to 20%) to simulate positional variation in scans, and horizontal flipping to account for different orientations. Rotations up to 10 degrees were introduced to account for variations in imaging angles, and zooming adjustments (up to 20%) helped the model learn features at varying scales. These augmentations effectively expanded the variability within the training dataset, making the model more robust in handling real-world data.

The dataset exhibited class imbalance across its four categories: NORMAL, CNV, DME, and DRUSEN. This means the number of samples in each class was not evenly distributed, with some classes having significantly more samples than others. Such imbalances can lead to biased model predictions, where the model disproportionately favors classes with more data, resulting in poor performance for underrepresented classes. Class weights assign higher importance to underrepresented classes and relatively lower importance to overrepresented ones, ensuring that the model treats all classes equitably.

By leveraging these preprocessing strategies, the dataset was effectively prepared to support the training of accurate and reliable deep learning models, while maintaining computational efficiency.

CLASS	**WEIGHT**
Normal	0.5609
CNV	1.8392
DME	2.4224
DRUSEN	0.7931

VII. RESEARCH OBJECTIVES

The primary objective of this research was to create an automated system capable of classifying retinal OCT images into four categories: CNV, DME, Drusen, and Normal. The significance of this work lies in its potential to revolutionize ophthalmological diagnostics by enabling timely and precise detection of retinal diseases. Early detection is crucial for reducing vision loss and enhancing patient outcomes, particularly in regions with limited access to specialized healthcare.

VIII. RESEARCH METHODOLOGY

The classification of retinal diseases was addressed using a deep learning-based method. Preprocessing the dataset to preserve consistent image dimensions and applying data augmentation strategies to improve model generalization were part of the methodology. Two different model architectures were investigated: a bespoke CNN created from scratch and a ResNet-50 model that used transfer learning. The bespoke CNN offered a more straightforward, portable option, while transfer learning allowed the ResNet-50 model to use pre-trained weights for effective feature extraction. The Training, Validation, and Testing subsets were used to train and assess both models, and hyperparameters were adjusted to maximize their effectiveness. Accuracy, precision, recall, and F1-score were among the evaluation metrics. Class weighting and data augmentation were used to overcome issues including overfitting and dataset bias.

IX. MODEL BUILDING

For this project, two distinct model architectures were employed to classify retinal OCT images: a custom CNN model and a ResNet-50 model.

The baseline model was the ResNet-50 model. The model was adjusted for retinal OCT image classification by unfreezing specific layers after being initialized with pre-trained weights from the ImageNet dataset. This method enabled the model to learn task-specific features pertinent to the retinal circumstances while simultaneously utilizing its general image processing skills acquired from ImageNet.

Additionally, a custom CNN model was designed from scratch to provide a lightweight alternative. The custom CNN focused on capturing essential features from the OCT images using a simpler architecture, which offered a balance between performance and computational efficiency.

Both models were evaluated on their ability to classify the OCT images into four categories: NORMAL, CNV, DME, and DRUSEN, providing a comparison to identify the most effective approach for accurate retinal condition classification.

A.1. BASELINE MODEL: RESNET50

The ResNet-50 model was employed as the baseline model for this project due to its effectiveness in handling deep neural networks through residual learning. This architecture, which consists of 50 layers with skip connections (or residual connections), addresses the vanishing gradient problem and allows gradients to flow more efficiently through the network, enabling the training of very deep networks. In this implementation, the ResNet-50 model was initialized with pre-trained weights from the ImageNet dataset to leverage transfer learning. This approach allowed the model to apply prior knowledge of generic image features, thus improving its efficiency in classifying retinal OCT images.

The model was configured by utilizing the base ResNet-50 architecture while excluding its fully connected layers (via `include_top=False`) to tailor it for the retinal OCT classification task. A GlobalAveragePooling2D layer was added after the convolutional base to reduce the feature map dimensions without compromising spatial information. To prevent overfitting, especially given the smaller dataset, a dropout layer with a rate of 0.5 was incorporated. The final dense layer consisted of 4 neurons with a SoftMax activation function, enabling the model to classify input images into four categories: NORMAL, CNV, DME, and DRUSEN. The layers of the base model were set as trainable, allowing for fine-tuning during the training process. Fine-tuning the ResNet-50 model enabled the pre-trained weights to be adapted to the specific characteristics of OCT images while preserving the general features learned from ImageNet.

The model was compiled using the Adam optimizer with a learning rate of 0.001, and a categorical cross-entropy loss function was chosen for the multi-class classification task. Accuracy was tracked as the primary metric during both training and evaluation to assess the model's performance.

A.2. MAIN MODEL: Custom CNN Model

A custom convolutional neural network (CNN) was developed specifically to classify OCT images, offering a distinct approach compared to the ResNet-50 model by prioritizing the extraction of significant features from the dataset. The architecture includes three convolutional layers, utilizing filters of sizes 32, 64, and 128, with a kernel size of (3, 3) for each layer. To introduce non-linearity, a ReLU activation function is applied after each convolution, enabling the network to identify complex patterns in the data. Each convolutional layer is followed by a max-pooling operation with a pool size of (2, 2), which serves to reduce the spatial dimensions of the feature maps. This step not only minimizes computational overhead but also ensures the model emphasizes critical features for effective classification.

Following the convolutional and pooling layers, the feature maps were flattened into a one-dimensional vector to prepare them as input for the fully connected layers. A dense layer with 128 units and a ReLU activation function was then introduced to capture high-level patterns from the extracted features. A dropout layer with a 50% rate was incorporated to deal with any overfitting that might occur during training. The final output layer includes 4 neurons, representing the four categories (NORMAL, CNV, DME, and DRUSEN), and uses a SoftMax activation function to generate class probabilities.

The custom CNN model was configured using the Adam optimizer with a learning rate of 0.001 and a categorical cross-entropy loss function, suitable for multi-class classification tasks. To enhance the model's generalization and reduce overfitting, data augmentation techniques such as rotation, width/height shifts, zooming, and horizontal flips were applied to the training images, increasing the perspective diversity of the dataset. The model's performance was evaluated based on accuracy, which was tracked during training and evaluation.

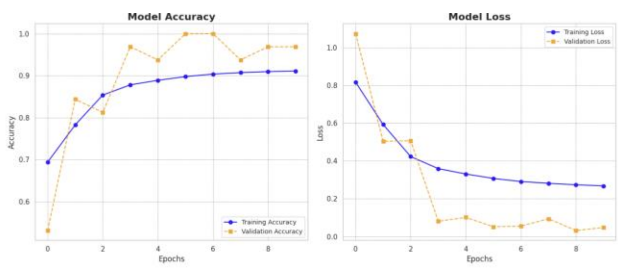
X. RESULTS AND EVALUATION

The ResNet50 model, serving as the baseline for this study, outperformed the custom CNN model in retinal OCT image classification, achieving a test accuracy of 99.34% with a test loss of 0.033, compared to the CNN model's test accuracy of 97.21% and test loss of 0.106. This demonstrates ResNet50's ability to leverage pre-trained features and adapt them effectively to the task.

Class-wise evaluation further emphasized ResNet50’s superiority, with higher precision, recall, and F1-scores across all categories. The confusion matrices indicated that ResNet50 had fewer misclassifications, particularly in categories with overlapping features. For instance, while the CNN model achieved F1-scores of 98%, 96%, 97%, and 99% for CNV, DME, NORMAL, and DRUSEN, respectively, ResNet50 consistently delivered better metrics, showcasing its robustness.

MODEL	**CNN**	**ResNet**
Training Accuracy	91%	93.76%
Training Loss	0.2703	0.1862
Test Accuracy	97%	99.34%
Test Loss	0.0477	0.0330

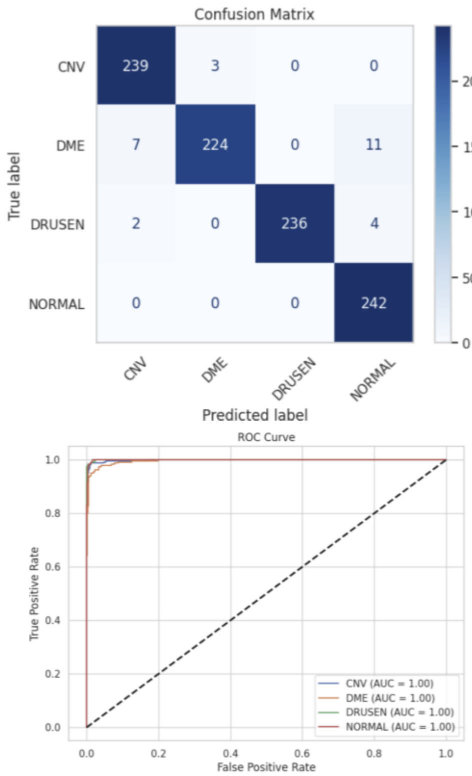
Training dynamics revealed stable improvements in both models, with the ResNet50 model demonstrating a higher alignment between training and validation metrics, suggesting effective generalization and minimal overfitting. The custom CNN model exhibited slightly less pronounced improvements, reflecting its simpler architecture.



For Main Model: Custom CNN Model

The ROC curve analysis supported ResNet50's superior discriminative ability, with AUC values approaching 1.0 for all classes, affirming its reliability in distinguishing between retinal conditions. In contrast, the CNN model had marginally lower AUC values for some classes.

Confusion matrix analysis highlighted the classification performance of CNN model. The CNN model, while effective, showed more misclassifications in complex categories like DME. Overall, the fewer misclassifications in CNN underscored its ability to extract and utilize discriminative features, resulting in robust classification.



For Main Model: Custom CNN Model

XI. MODEL COMPARISON

The ResNet50 model demonstrated significantly superior performance compared to the custom-designed CNN, with a test accuracy of 99.34% versus 97.21%. This marked difference highlights the advantages of using a pre-trained deep learning architecture for complex tasks like retinal OCT image classification. ResNet50's exceptional performance can be attributed to several factors. First, its residual learning framework addresses the vanishing gradient problem commonly faced in deep networks by introducing skip connections. These connections allow gradients to flow more effectively through the network, enabling the model to learn progressively refined feature representations. This is particularly critical for medical imaging tasks, where fine-grained features differentiate disease classes.

Second, the pre-trained weights from ImageNet provided ResNet50 with a robust foundation of generalized image features. This transfer learning approach allowed the model to adapt effectively to the OCT dataset with minimal training, leveraging its existing knowledge of visual patterns such as edges, textures, and shapes. The depth of ResNet50, comprising 50 layers, further enabled it to extract intricate and hierarchical features from the images, capturing subtle variations that differentiate conditions like Choroidal Neovascularization (CNV) and Diabetic Macular Edema (DME).

In contrast, the custom CNN, though computationally lightweight, lacked the architectural complexity to perform at the same level. Its three convolutional layers and limited parameter space constrained its ability to model the nuanced relationships within the data. While data augmentation and normalization improved its generalization to an extent, the custom CNN struggled with overlapping features between similar classes, leading to higher misclassification rates in challenging cases. Additionally, the custom CNN exhibited slightly higher test loss (0.106 compared to 0.033 for ResNet50), indicating less robust learning and generalization.

These findings emphasize the importance of deeper architectures and pre-trained models in achieving high accuracy for medical imaging tasks. ResNet50's ability to balance feature extraction with effective generalization underscores the value of transfer learning and advanced architectural designs in providing reliable automated diagnostic tools.

XII. LIMITATIONS

The "Retinal OCT Image Classification" project, while achieving impressive results, has certain limitations that warrant attention. The dataset, though comprehensive, may not fully capture the diversity of retinal conditions and demographics, which could limit the model's ability to generalize effectively across different populations and clinical environments. Furthermore, despite employing data augmentation, the model exhibited susceptibility to overfitting, suggesting a need for stronger regularization strategies to enhance its robustness.

Another key limitation is the model's lack of interpretability. In clinical settings, transparency and explainability are crucial for building trust and enabling healthcare professionals to understand and verify AI-driven diagnostic recommendations. Additionally, the system's performance on noisy or low-quality images has not been rigorously tested, raising concerns about its reliability in real-world scenarios where such conditions are common. High computational requirements also present a barrier to deployment in resource-constrained settings, such as rural clinics or portable diagnostic tools.

Finally, the model has yet to undergo clinical validation, which is essential for establishing its practical reliability and effectiveness in diagnosing retinal diseases.

Addressing these limitations, including expanding the dataset, enhancing interpretability, and optimizing for efficiency, will be critical for transitioning this research into a reliable and widely usable diagnostic tool.

XIII. FUTURE WORK

In the future, we plan to enhance the performance of our retinal OCT image classification model through several key improvements. First, we intend to fine-tune the pre-trained layers of the ResNet50-based model. This process will allow the model to better adapt to the unique features of OCT images, enabling it to capture more domain-specific patterns and improve classification accuracy. Additionally, we aim to implement more extensive data augmentation techniques to further enhance the model's robustness and reduce the risk of overfitting. By applying transformations such as rotation, zoom, and shifts to the training data, we can simulate a wider range of variations and help the model generalize better to unseen data. Furthermore, we plan to optimize the learning rate to improve the training process. In order to enhance the training process, we also intend to maximize the learning rate. We can guarantee more effective convergence, lessen the possibility that the optimizer will become trapped in local minima, and speed up model training by utilizing learning rate schedules or adaptive learning strategies. It is anticipated that these modifications will greatly improve the model's overall performance and increase its suitability for practical uses in the classification of retinal diseases.

XIV. CONCLUSION

The Retinal OCT Image Classification project demonstrates the powerful potential of Artificial Intelligence (AI) in advancing the diagnosis of retinal diseases, such as Diabetic Macular Edema (DME), Choroidal Neovascularization (CNV), and Drusen. By utilizing high-resolution Optical Coherence Tomography (OCT) images and leveraging advanced deep learning models, including ResNet50 and a customized Convolutional Neural Network (CNN), this study achieved impressive results, with the CNN model reaching a classification accuracy of 97%. This high accuracy underscores the model's robustness in identifying retinal conditions and its applicability in real-world medical scenarios.

The project began with a comprehensive analysis of the dataset, which consisted of 84,495 labeled OCT images. Data preprocessing, including resizing, normalization, and augmentation, ensured the dataset was optimized for model training. Initial experiments with the ResNet50 model established a baseline, but it was the custom CNN model that significantly outperformed, benefiting from architectural improvements and hyperparameter tuning that contributed to its superior performance.

This work addresses key challenges in traditional diagnostic methods by offering a scalable and automated solution, which is particularly crucial in high-volume clinical environments and underserved regions with limited access to specialists. Additionally, the CNN model reduces misclassification errors and maintains balance across all disease categories, ensuring both fairness and reliability in its predictions.

The promising results of this project pave the way for further research and improvements. Future work will focus on incorporating advanced data augmentation techniques, fine-tuning pre-trained layers, and optimizing learning rates to enhance model performance. By building on these developments, this project holds the potential to revolutionize ophthalmological diagnostics, offering timely, accurate detection and ultimately improving patient outcomes and reducing global vision loss.

XV. REFERENCES

[1] A. Choudhary, S. Ahlawat, S. Urooj, N. Pathak, A. Lay-Ekuakille, and N. Sharma, “A Deep Learning-Based Framework for Retinal Disease Classification,” *Healthcare*, vol. 11, no. 2, p. 212, Jan. 2023

[2] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.

[3] 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* (pp. 770-778).

[4] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 60.

[5] Huang, D., Swanson, E. A., Lin, C. P., Schuman, J. S., Stinson, W. G., Chang, W., & Stopyra, G. (1991). Optical coherence tomography. *Science*, 254(5035), 1178-1181.