**Abstract**

Convolutional neural networks can play a critical role in the clinical diagnosis of conditions that can be discovered through medical imaging. Our project will dive into the world of retinal optical coherence tomography in order to develop a framework that can serve as a support tool in the diagnosis of common retinal issues that can lead to blindness.

Our framework will involve a dataset of OCT images of patients retinas that will be split into train, test, and validation sets. The CNN architecture will evaluate our our dataset, resulting in some preliminary results. Implementation of REeLu activation, will be done with loss calculated using softmax classifier.

Applying this approach to a dataset of optical coherence tomography images, we demonstrate how CNN structures such as AlexNet and VGGNet can be employed in medical image analysis. We also provide some the different stages of initial performance on CNN layers and how we can optimize our hyperparameters to obtain our desired result.

# **Introduction**

Age-related macular degeneration (AMD) and diabetic macular edema (DME) are some of the leading causes of blindness worldwide. In the United States alone, almost 10 million individuals suffer from AMD. Additionally, each year more than 200,000 people develop choroidal neovascularization (CNV) and nearly 750,000 individuals over the age of 40 suffer from DME. Currently, anti-vascular endothelial growth factor (anti-VEGF) medications are used to treat such retinal diseases. In order to administer these medications, optical coherence tomography (OCT) is needed to provide a clear cross-sectional representation of the retina. OCT uses light to capture high resolution cross sectional images of the retina, and is one of the most commonly performed medical imaging procedures worldwide. OCT is critical to the treatment of these diseases, though diagnosis through OCT can be slow and difficult.

The diagnosis of diseases using the development of machine learning algorithms are becoming more prevalent. These algorithms are capable of classifying image data very quickly, and could potentially be useful for expediting diagnosis of retinal diseases. A popular model that is used for image analysis is convolutional neural networks.

Convolutional neural networks (CNNs) are made up of neurons that have learnable weights and biases. We have a neuron that takes in some input, performs a dot product, and then optionally follows it with non-linearity. We then express the whole network through a single differentiable score function moving from a raw image pixel on one end to class scores at the other.

**Problem Statement**

In this paper, we present an approach to diagnose CNV, DME, or AMD using a convolutional neural network architecture trained on the images produced by OCT. The images are provided in a dataset made publicly available by Kermany et al [1]. The dataset consists of 207,130 grayscale OCT images of the retinas of 4,686 patients, split into train, test, and validation sets. There are 37,206 images with CNV, 11,349 with DME, 8,617 with DRUSEN, and 51,140 that are normal. All images were verified by clinical experts that are trained to diagnoses these diseases from OCT. The past study was able to successfully implement a CNN that achieve accuracies of 96.6%, so we expect to achieve similar accuracies. Our success will be evaluated by whether our selected architectures perform similarly to the one presented in the study and can perform as well as if not better than human experts.

**Technical Approach**

We aim to train several types of CNN architectures. Two that we plan to use are the AlexNet and VGGNet, which are commonly employed in medical image analysis [6]. Before we test our data, we will be be normalizing our training data.

Alexnet is a Deep Convolutional Neural Network for image classification [4]. Alexnet is composed of 8 layers where the first 5 are convolutional and the last 3 are fully connected layers [4]. Some pooling and activation layers also exist in between these layers [4].

VGGNet is a 19-layer model referred to as VGG19. It is a popular model used for medical data [6].

We are planning to fragment our technical approach into stages in order to best evaluate progress and compare methods against one another. We will describe these stages below.

First, we will train a simple three layer CNN on the dataset and evaluate its initial performance. We will set this performance as the benchmark to compare against when we adjust the CNN.

Our second stage will work on further exploring the AlexNet architecture and integrating it optimally on top of our 3-layer CNN. We are also planning to spend a portion of our time during the second stage working on optimizing the hyperparameters involved. Once we achieve a notable difference from the initial benchmark, we will move onto the third stage of our implementation.

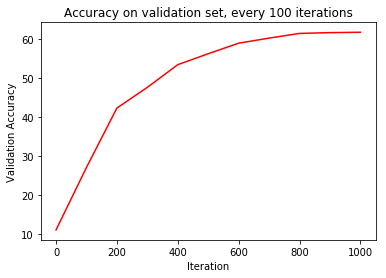
During our third stage of implementation, we aim to explore a different model to evaluate the medical imaging, VGGNet. If we achieve similar results to the AlexNet, we are planning to see whether there are characteristics of the VGGNet model which can further lift our second former AlexNet model.

Our fourth and final stage of implementation will involve heavy experimentation based off results and findings from using AlexNet and VGGNet in order to achieve incremental or significant improvement to our accuracy results. Our experimentation will involve priming our dataset with RNNs, utilizing batch normalization and dropout layers, and different optimization algorithms.

**Preliminary Results**

Currently, we are in the beginning stages of implementing a three-layer CNN and AlexNet. The three-layer network implemented two convolutional layers, with 32 5x5 filters and 16 3x3 filters, respectively. ReLU activation was implemented after each convolutional layer. One fully connected layer with four output nodes was included after the convolutional layers. The loss was calculated using softmax cross entropy loss.

The performance of the the three-layer network was not optimal, most likely because it is not deep enough to pick-up nuanced features in the OCT images. Additionally, we are currently not using dropout or batch normalization, which can be important in improving performance.



**Figure 2:** Performance of three-layer CNN on validation set after training

AlexNet was also implemented on this dataset. It employed three convolutional layers with increasing filter size, but was counterbalanced with downsampling from maxpooling layers. Batch normalization and dropout with keep probability of 0.8 were also used after each convolutional/maxpool layer. The last two layers were fully connected with ReLU activation. The largest class score was taken as the predicted disease, with the loss calculated through softmax cross-entropy. This is currently under implementation, and we hope to evaluate its performance as our immediate next steps.

**References**

[1] Friedman, D.S., O’Colmain, B.J., Mun˜ oz, B., Tomany, S.C., McCarty, C., de Jong, P.T., Nemesure, B., Mitchell, P., and Kempen, J.; Eye Diseases Prevalence Research Group (2004). Prevalence of age-related macular degeneration in the United States. Arch. Ophthalmol. 122, 564–572.

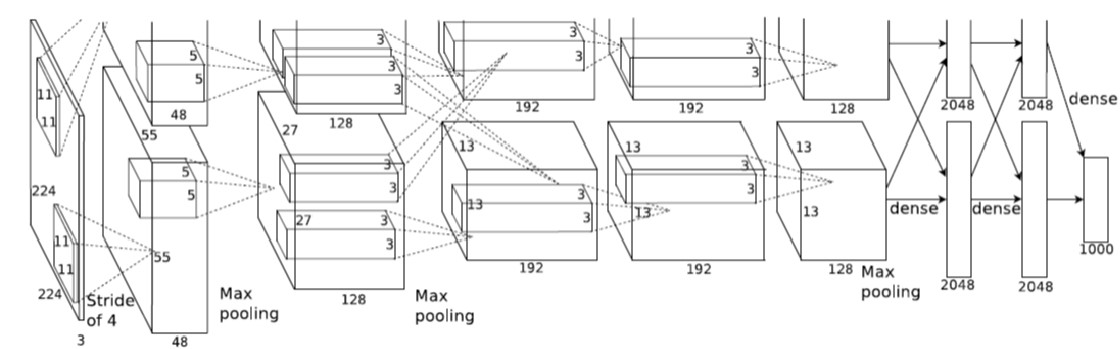
[2] Ferrara, N. (2010). Vascular endothelial growth factor and age-related macular degeneration: from basic science to therapy. Nat. Med. 16, 1107–1111.

[3] Wong, W.L., Su, X., Li, X., Cheung, C.M., Klein, R., Cheng, C.Y., and Wong, T.Y. (2014). Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. Lancet Glob. Health 2, e106–e116.

[4] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84-90.

[5] Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., … Zhang, K. (2018). Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell*, *172*(5), 1122-1131.e9.

[6] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A., Ciompi, F., Ghafoorian, M., … Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, *42*, 60-88. doi:10.1016/j.media.2017.07.005



The diagram above is a sequence of layers in Alexnet. Note that the diagram is divided into two parts where the first half is executing on GPU 1 and the other half on GPU 2 [4].