DRSS Severity Classification on OCT Images

Link to Presentation Video: https://youtu.be/JqCqhoMEpwQ

Link to Github: https://github.com/ziyul893/fml\_final\_project

Michael Meng   
*Department of Electrical and Computer Engineering*  
*Georgia Institute of Technology*Atlanta, Georgia  
mmeng35@gatech.edu  
  
2nd Ziyu Liu  
*Department of Electrical and Computer Engineering*  
*Georgia Institute of Technology*Atlanta, Georgia  
zliu862@gatech.edu

*Abstract*— Clinical diagnosis of eye diseases is a lengthy process that involves many OCT (Optical Coherence Tomography) scans on a patient over a period of many weeks. Therefore, there exists a need to automate or speed up this process. This paper will be using four classification methodologies: Naïve Bayes, Support Vector Machines, AlexNet with new weights, and ResNet50 with transfer learning. Each have been shown to have succeeded on similar datasets. The goal is to correctly identify the DRSS (Diabetic Retinopathy Severity Scale) of the eye using OCT images from OLIVES [1] testing dataset. This paper also proposes a the flashbulb training method of training convolutional neural networks (CNNs) based on the psychological phenomenon of flashbulb memories that has the advantage of decreasing training times for CNNs. The resulting accuracy for the four methods are around 40%, indicating that there is room for improvement with aspects of the classification process outside of the model itself. Although the flashbulb method demonstrated in this paper did not produce accurate results, it displays a promising upward trend, suggesting certain tweaks are needed to point this method in the correct direction.

Keywords—Machine Learning, Transfer Learning, Medical Image Classification

# Introduction

This paper aims to classify OCT frames into 3 DRSS levels 0, 1, and 2 based on the OLIVES dataset. Two example frames from an OCT scan is shown in Figure 1. This classification is for the purpose of diagnosing and thus treating a patient for diabetic retinopathy.

To achieve high accuracy, this paper attempts four distinct methods of classification: Naïve Bayes, Support Vector Machines, training on ResNet50 CNN model with transfer learning, and training on AlexNet CNN model with new weights. Explanations on the methods and how they were used will be explained in section II.

A picture containing worm, screenshot, monochrome, monochrome photography

Description automatically generatedA picture containing screenshot, monochrome, black, monochrome photography

Description automatically generated

*Figure 1. Frames from an OCT scan of the same class*

As no authors of this paper are particularly adept at interpreting OCT scans, especially in evaluating the severity of Diabetic Retinopathy, heuristics that could be used in conventional classification in image datasets such as ImageNet do not apply. Furthermore, a volume is made up of multiple frames with the two examples in Figure 1 being from the same volume. Although these count as the same class, they are obviously different and it cannot be told from a layman’s view whether they represent the same scan taken from a different angle, or differ in some other criteria, which makes any sort of predetermination extra difficult. The questions then arise of 1.How to correctly classify images from a dataset in which the designer of the training model knows very little about, and 2.Due to the size of the OLIVES dataset and its complexity, how to simplify and shorten the training time beyond image processing techniques, as any processing performed on the OCT frames may cause undesired information to be filtered.

To tackle both issues mentioned, this paper makes an attempt to implement a new training method for CNNs mentioned in section III that references the psychological phenomenon of flashbulb memory. Section IV contains the results and comparisons of the four classification and finally section V contains an evaluation of the results as well as detailing any future work.

# Different Methods Used

A total of four methods are used to attempt to classify the OCT frames chosen based on their expected success with the dataset. Each method has been found to have previously succeeded with other medical image classifications, so they are known to be capable of the task that motivates this paper, although more faith is placed into training AlexNet from scratch and training ResNet50 with transfer learning due to their superior computational complexity and CNN’s natural compatibility with images. Therefore, less innovation will be tried with the Naïve Bayes and Support Vectors Machines and only an autoencoder will be used to decrease computational time with those models.

## Naïve Bayes

Naïve Bayes method is a way to predict the probability of a certain class with the feature given the probability of feature in a certain classes. Four parameters used are posterior probability, likelihood, class prior probability, and predictor prior probability, where posterior probability is the prediction of classes based on features provided. There have been successful attempts at using Naïve Bayes classifiers to determine KRAS mutations in colon cancer[6].

## Support Vector Machines (SVM)

Previous successful attempts have been made on classifying medical X-ray images[3]. The authors used VGGnet-16 and AlexNet as feature extractors to train a SVM, and this paper is attempting something similar with the autoencoder that will be described in section III. The primary goal of SVM is to propagate dataset that not linearly separable to high dimension and identify a hyperplane to maximize the classification effect. To maximize the hyperplane distance with any data point, a technique called Maximal Margin Classifier is used to find the Maximal Margin hyperplane (MMH), which id ideally furthest from any training data. Training observations that lies on the margin boundary, or the support vector determines the MMH. Usually, the goal for SVM algorithm is to produce the parameters of the MMH.

## Training a New Alexnet

Alexnet is a revolutionary Convolutional Neural Network (CNN) architecture that was first designed in 2012. It consists of 8 layers, including 5 convolution layers, 2 fully connected layers, and 1 softmax layer. The pre-trained AlexNet is available in the Pytorch package, but it will not be used in this approach. There have been successful attempts utilizing AlexNet for skin hydration and skin solvent penetration measurements[4]. Training from scratch means that the original structure of AlexNet will be retained, but the model uses all the OCT image data to train the network from blank. This method is computation heavy, but the design will be specific to desired dataset.

## Transfer Learning on ResNet50

ResNet50 is a variant of the original ResNet architecture introduced in 2016 that aimed to improve the depth, performance and training speed of the original CNN model[5]. The pretrained weights used in this transfer learning process are the default weights provided by Pytorch, which was chosen based on the criteria of having the best accuracy classifying images in the CIFAR-10 database. The changes made to the network were only in the inputs, as OCT frames are black and white while the network normally expects a 3-channel RGB input, and in the output, where only 3 classes are needed rather than 1000.

# Training Methods

## Convolutional Autoencoder

Due to the size of each frame being 504 x 496 pixels, the input to the Naïve Bayes and SVM models would be millions of dimensions. This would be far too computationally expensive, both in resources and time to achieve over the length of the entire OLIVES dataset. Therefore, an autoencoder is used to decrease the representation of a frame down to the 64 most significant dimensions as inputs to the above models.

## Pre-Processing

Although the size of the frames is not an issue with the Naïve Bayes or SVM algorithms because as mentioned previously they are mostly to test the complexity of the dataset. AlexNet and ResNet50 were trained and designed with a specific image size(224 x 224) in mind and the OCT frames shall be resized to respect this architecture. They will also be normalized with the mean and standard deviation of the grey scale values of the entire dataset.

In addition, through inspection of the OLIVES dataset, the training and testing set data are interpreted to be structured as such: 49 OCT scan frames are generated per visit that the patient makes. Each patient makes a visit every 4 or 8 weeks for a total of 104 weeks, or 2 years. Assuming that the DRSS value of the patient’s eye does not change in the duration of a single visit, it’s reasonable that at least the 49 frames generated in that particular visit can be “bundled” together to train at once, as they are with the same patient in the same visit. This bundling is done via 3D convolution and should speed up the training process. But this is in comparison to training the models frame by frame and may be sacrificing some accuracy. Therefore with the flashbulb training detailed below which is already an attempt at a time saving technique, the two CNN models will be trained frame by frame with their labels rather than through a combination of multiple frames.

## Flashbulb Training

For both Alexnet and ResNet50 architectures, the model will be trained using primarily classes 0 and 2, then late into the training process a combined 3D-convoluted image made of multiple instances (volumes) of class 1 will be introduced to train the model for one epoch at a higher learning rate. This method of training is inspired by the psychological phenomenon of “flashbulb memory”[2] where in which it is observed that the human brain experiencing a dramatic/emotionally significant event tends to remember it in high detail and that the event lasts for a extended period amount of time in the brain compared to everyday occurrences. Thus the term “flashbulb training” is coined for this training method.

Similarly, from the view of an optometrist who is used to only seeing a certain severity label, say of example level 2, then they may recognize a level 0 instance if they had only met it a select few times before, as it would be considered a rare example. In this example, the “everyday occurrence” would be the patients whose OCT scans indicated a DRSS severity level of 2 and the significant event to generate the “flashbulb memory” would be the patient whose OCT scan in a certain eye indicated a DRSS severity of level 2. This can also be generalized into other fields, not just optometry and in extension, image processing. Of course, the assumption is made here that the class introduced late into the training process is rare, which may not be the case.

As a neural network mimics the brain via the neuron, connections and activation, this flashbulb memory training method should in theory “shock” the model with the high learning rate integrated in backpropagation of the introduced new class, leaving its “impression” in the weights of the model. An advantage of this training method is that the training time is decreased significantly, as only select examples of an alternative class is needed to train the network rather than using all training datasets in every epoch.

# Results

First, the accuracy of the four methods mentioned in section II on the testing dataset in OLIVES are listed. Then, the results of the flashbulb training method in AlexNet and ResNet50 are separately compared in figures 2 and 3 below to get an idea of how well this method works.

The table below lists the overall accuracy of the methods to correctly classify a given frame to its target of one of the three DRSS severity levels. For the flashbulb training methods, only the version with the highest accuracy is included.

Table 1. Method and Their Overall Accuracies

|  |  |
| --- | --- |
| Method | Accuracy |
| Naïve Bayes | 0.401 |
| SVM | 0.352 |
| AlexNet | 0.442 |
| ResNet50 | 0.385 |
| AlexNet Flashbulb | 0.362 |
| ResNet50 Flashbulb | 0.348 |

The following two figures compare the distribution of correct predictions between the “regular” – which means the conventional method of training all batches equally - verses the flashbulb training method of training the model exclusively on class 1 at the epoch indicated on the horizontal label in the graph, and only on classes 1 and 2 otherwise.

A picture containing text, screenshot, font, number

Description automatically generated

*Figure 2. Comparison of the Distributions of Correct Predictions for all Training Methods in AlexNet*

A picture containing text, screenshot, font, line

Description automatically generated

*Figure 3. Comparison of the Distributions of Correct Predictions for all Training Methods in ResNet50*

# Discussion

From Table 1 it can be seen that the best performing method of classification is AlexNet trained regularly, with a 44.2% accuracy on the testing dataset of OLIVES. This is to be expected, as this is the most time and resource consuming method. It is followed by Naïve Bayes, which is more unexpected as the authors thought that the two CNNs would perform better than the non-CNNs. This suggests that most likely because the dataset that ResNet50 was trained on is far too different from OLIVES, one is that it’s most likely trained to recognize everyday objects and through RGB channels as well, this may have made it incompatible with the greyscale, medical images of OLIVES and therefore it should not have been applied in the same way. The accuracies of the other methods were as expected, SVM because it wasn’t expected to perform as well as the CNNs, and the models with the flashbulb training method as it was intended to be a training method that saved on computational complexity.

Looking at figures 2 and 3 comparing the distribution of correct predictions for the two CNN models, it can be seen that trained conventionally, class 1 frames tend to get predicted the most out of any. This is most likely due to the fact that class 1 frames are the most prevalent in the training dataset, and thus the models learn to recognize class 1 frames the best. The effect of decreasing the model’s exposure to class 1 frames are immediate: if it is introduced too early, then the model will gradually “forget” class 1 frames and it can be seen that zero class 1 frames are correctly predicted in the case of epoch 5 in AlexNet. However, there is a clear trend: the closer toward the end of the training that the new class in introduced to the model, the better the model gets at recognizing that class. Although even at its highest for the flashbulb training methods, the number of predicted class 1 frames never get close to that of the conventionally trained CNN, this suggests that maybe more epochs need to be tested, or that there may be some other optimizing technique to be used with flashbulb training aside from artificially increasing the learning rate. However, it should be noted that the real benefit of this training method has more to do with the decreased training time; Each session of the flashbulb training method took around half the time to finish compared to the conventionally trained model.

Future works could see many adjustments and improvements being made in this project. The OLIVES dataset contains 16 biomarkers with 4 clinical labels with each OCT scan as described in [1]. These biomarkers could be used in the training process of the model, although due to their normally being used by trained practitioners, the question remains of how to fully efficiently utilize this data in a machine learning algorithm.

While ResNet50 with transfer learning seemed to be compatible with flashbulb training, AlexNet being trained from new weights seemed to be much more sensitive to any changes made in the network. For example, when introducing the novel class to the network during the fifth epoch, it was found that the learning rate could not exceed 0.1, as doing so would cause subsequent epochs to output loss values of NaN (Not a Number). As much as 50 epochs were continued after this to observe if the loss value would drop, but no change occurred. Having a high learning rate is important for the flashbulb training process as the idea is to use the novel class to “shock” the network, and lower learning rates would cause situations such as AlexNet’s flashbulb at epoch 5, which results in the network forgetting about the trained weights and not predicting class 1 frames at all.

##### VI. References

1. M. Prabhushankar, K. Kokilepersaud\*, Y. Logan\*, S. Trejo Corona\*, G. AlRegib, C. Wykoff, "OLIVES Dataset: Ophthalmic Labels for Investigating Visual Eye Semantics," in *Advances in Neural Information Processing Systems (NeurIPS 2022) Track on Datasets and Benchmarks*, New Orleans, LA,, Nov. 29 - Dec. 1 2022
2. Brown, R., & Kulik, J. (1977). Flashbulb memories. Cognition, 5(1), 73–99. [https://doi.org/10.1016/0010-0277(77)90018-X](https://psycnet.apa.org/doi/10.1016/0010-0277(77)90018-X)
3. D. O. Alebiosu and F. P. Muhammad, "Medical Image Classification: A Comparison of Deep Pre-trained Neural Networks," 2019 IEEE Student Conference on Research and Development (SCOReD), Bandar Seri Iskandar, Malaysia, 2019, pp. 306-310, doi: 10.1109/SCORED.2019.8896277.
4. X. Zhang, W. Pan and P. Xiao, "In-Vivo Skin Capacitive Image Classification Using AlexNet Convolution Neural Network," 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC), Chongqing, China, 2018, pp. 439-443, doi: 10.1109/ICIVC.2018.8492860.
5. He, K., Zhang, X., Ren, S., Sun, J. (2016). Identity Mappings in Deep Residual Networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science(), vol 9908. Springer, Cham. https://doi.org/10.1007/978-3-319-46493-0\_38
6. Pershad Y, Govindan S, Hara AK, Borad MJ, Bekaii-Saab T, Wallace A, Albadawi H, Oklu R. Using Naïve Bayesian Analysis to Determine Imaging Characteristics of KRAS Mutations in Metastatic Colon Cancer. Diagnostics (Basel). 2017 Sep 2;7(3):50. doi: 10.3390/diagnostics7030050. PMID: 28869500; PMCID: PMC5617950.