

Detection of Age - Related Macular Degeneration using Deep Learning

Deepika S.N Vemuri

Department of Computer Science and Engineering
BVRIT Hyderabad, College of Engineering for Women
Hyderabad, India

Nagaveni. B

Assistant Professor

Department of Computer Science and Engineering
BVRIT Hyderabad, College of Engineering for Women
Hyderabad, India

Abstract—Age-related macular degeneration is a disease that leads to loss of vision in the central field of the eye. It is known to affect nearly a million people in India alone every year. This condition occurs in 2 stages, although not always guaranteed to be consecutive: dry AMD and wet AMD. The presence of dry AMD does not have a significant effect on the patient's vision. So the patient generally would not even know that they have the disease until the wet AMD stage where their vision starts getting blurred. However, upon reaching the wet AMD stage, the loss of vision becomes irreversible.

The usage of AI and deep learning in fields like medicine is already proving to be revolutionary. Even though in the current state of things, the suggestions given by an AI cannot entirely replace a doctor's, such systems nonetheless could reduce the overall work that has to be done by the doctor.

Dry AMD is difficult to detect. By the time the disease progresses into the wet AMD stage, the damage done becomes irreversible. Hence for such a disease, early detection is crucial. In rural areas, where people may not have access to ophthalmologists, an early stage screening could prove to be vision saving. People of a certain age could have their OCT scans taken and undergo a screening test. If a positive result is output they would then understand that there is some sort of risk associated with the condition of their eyes and seek further medical help. The work presented in this paper aims to produce a deep learning model to act as this screening test. The model would take an OCT scan as the input and predict the stage of the disease.

Keywords—Dry/Wet AMD, OCT scans, vision loss, screening test

I. INTRODUCTION

The macula is the central part of the retina. It is responsible for the central, high-resolution, color vision that is possible in good light. As the body ages and possibly due to genetics, it could wear down. Age – related macular degeneration is a disease that leads to loss of vision in the central field of the eye. It's a common disease with more than a million cases per year in India alone.

It occurs in 2 stages: dry AMD, wet AMD. Some work has already been done in this area using time series data (Hwang, et al., 2019). The two stages, however, are not guaranteed to be consecutive which makes the disease all the more difficult to detect and diagnose. The presence of yellowish fatty deposits on the macula, technically known as drusen, could lead to dry AMD. A few small drusen, however, do not lead to

any significant changes in vision. As the condition gets worse, the light-sensitive cells in the macula get thinner and eventually die. The wet AMD stage, which is much more serious, is indicated by the presence of abnormal leaky blood vessels under the retina. These blood vessels leak blood and fluid into the retina causing the scarring of the macula. Vision becomes distorted such that straight lines look wavy.



Figure 1 (a): The vision of a person with wet AMD.

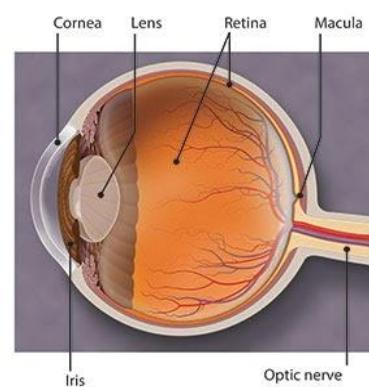


Figure 1 (b): The anatomy of a human eye. The macula is at the back in the center.

I.I METHODOLOGY

Several options currently exist for the detection of AMD including OCT scans, Amsler grids and near visual acuity ([Swartz, et al., 2015](#)). In addition to this, some work on the detection of AMD has been carried out by the usage of fundus images ([Phan, et al., 2016](#)). OCT scans (Optical Coherence Tomography) give a cross-sectional view of the various layers in the eye and their thicknesses. It is a quick, painless and contactless procedure provided the right machine is available. Such a scan could prove to be useful for the identification of the stage of disease in the eye ([Soichiro, et al., 2019, Ruyu Qi, 2017](#)). Under this work, a dataset of OCT scans were taken from Kaggle's 'Retinal OCT Images' dataset containing images under 4 categories: CNV (Choroidal Neovascularization / wet AMD), drusen, normal and DME (Diabetic Macular Edema). From this dataset, the images under the DME category were removed. A few sample images from the dataset used are given in Figure 2.

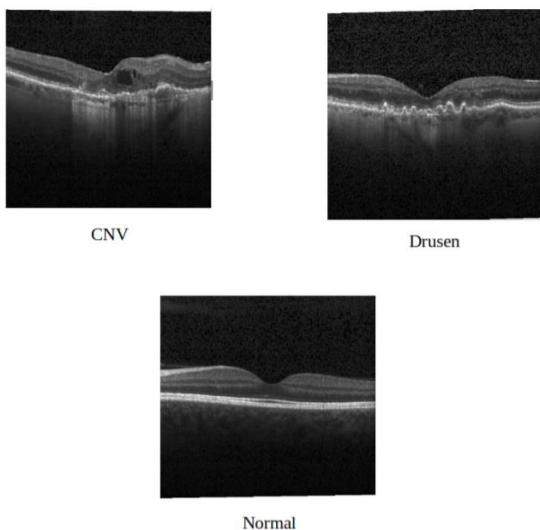


Figure 2: (a) CNV OCT scan. (b) An OCT scan with drusen deposits on the retina. (c) OCT scan of an eye without AMD.

II. HANDLING DATASET IMBALANCE

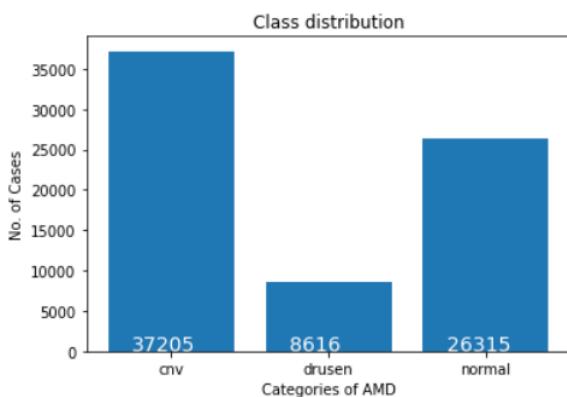
The dataset was initially skewed. There were comparatively

fewer drusen images and more CNV images. Handling this imbalance was necessary as the model would naturally try to overfit towards the majority class. There were three approaches used to reduce this imbalance ([Seguro, 2018](#)).

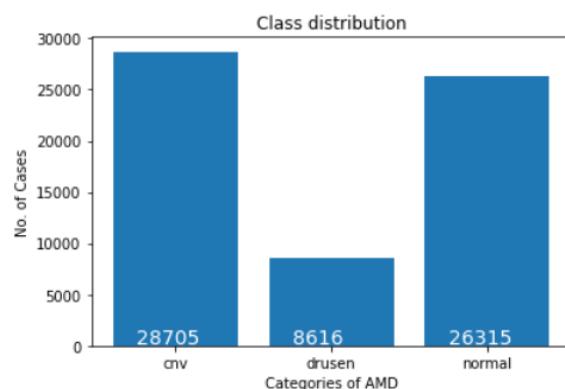
- **Undersampling:** Undersampling methods delete or merge examples in the majority class. Oversampling on the other hand, duplicates or augments the examples in the minority class ([Brownlee, 2020](#)). In this paper, for undersampling, a random sample of about 8500 images were removed from the CNV category. Since the dataset was already comparatively large, the negative impact of the loss of data on the effectiveness of the model was relatively less.
- **Adjusting the weights of the classes:** Although reducing a few images from the dominating class was somewhat effective, it was at the cost of the loss of some valuable data. Instead of this, the classes can be weighted. The ratio of each class is computed with respect to the other classes. These ratios are used as a parameter in the training phase. By setting weights for the classes, a constraint is imposed such that the model is penalized more when a wrong prediction is made on the majority class.
- **Dropout:** Deep neural networks tend to overfit. So the main purpose of a dropout layer is to combat this overfitting. Dropout has one hyperparameter which specifies the probability at which outputs of the layer are dropped out, or inversely the probability at which the outputs are retained. For example, a value of 0.6 indicates that there is a 60% chance for a node to get dropped out.

III. TRANSFER LEARNING

In transfer learning, the knowledge that has been gained in one setting is used to improve optimization in another setting. In other words, a pre-trained model is trained on another dataset. This results in a significant decrease in the training time for a neural network model and can result in lower



(a) Original dataset



generalization error.

(b) Undersampled dataset

Figure 3: Bar graphs showing the dataset class distribution (a) Class distribution of the original dataset consisting of three classes: CNV Drusen Normal OCT scans. (b) Class distribution of the dataset after undersampling

In this paper, the base model used is ResNet50 which was initially trained on the ImageNet dataset. This model is then trained on another dataset (Retinal OCT Images). So, an already existing model is tweaked and adjusted to suit the specific problem being solved. It would take much longer to start the training procedure from scratch.

III.I. THE VANISHING GRADIENTS PROBLEM

In deep neural networks as the gradient is back-propagated to earlier layers, repeated multiplication may make the gradient extremely small. Certain activation functions, like the sigmoid function, scale down a large input space into a small input space between 0 and 1. So a large change in the input of the sigmoid function will cause a small change in the output. Hence, the derivative becomes small. For a shallow network, this is not a significant problem. But when more layers are used, it can cause the gradient to be too small for training the network effectively. This was a problem that consistently occurred with the early CNN architectures. Residual networks were one of the first ConvNets to solve this problem. These networks introduced the concept of skip connections.

As the name implies they provide residual connections straight to earlier layers. More specifically, the activation unit from one layer could be fed directly to a deeper layer of the network. This is a skip connection. These skip connections enable the building of very deep neural networks.

The ResNet50 model consists of a series of convolutional layers + skip connections, then average pooling and then an output fully connected (dense) layer. The training images were fed into the ConvNet by means of an image generator. This sent the training images in batches of 32 and handled the preprocessing of the images as well. The flow of the work presented in this paper can be seen in Figure 4.

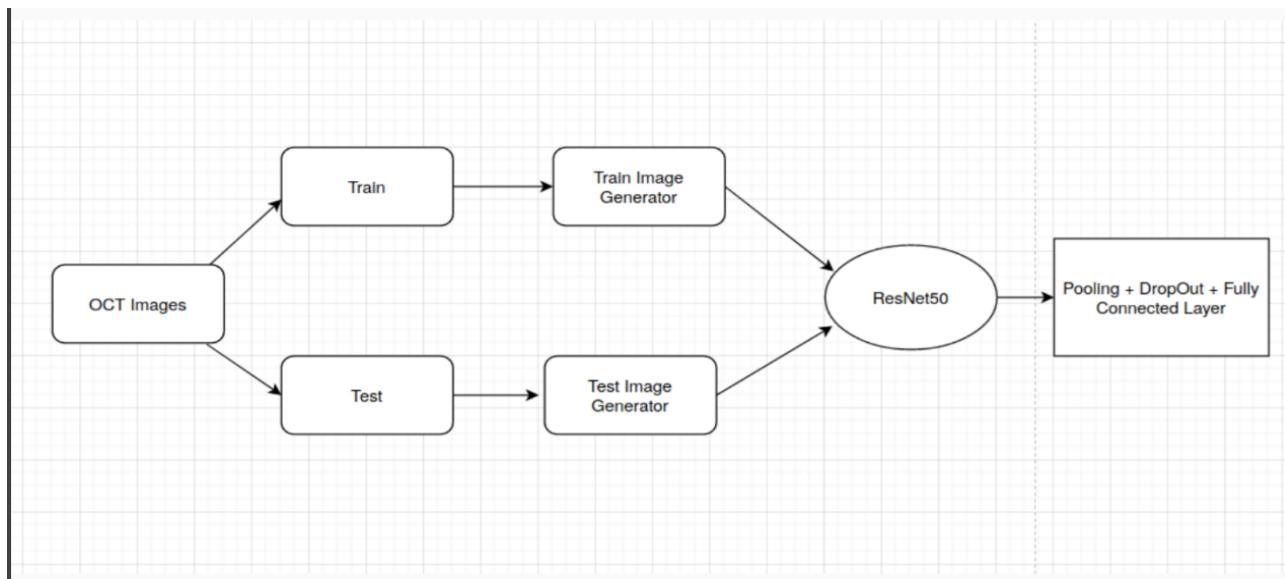


Figure 4: Architecture of the project. The OCT images are split into train and test sets and are converted into image generators (Brownlee, 2019). The ResNet50 model (+ a few additional layers) are then trained using transfer learning.

IV. RESULTS

The parameters that were taken into account for potentially influencing the performance of the model were: dataset imbalance, choice of optimizer and loss function and the dropout hyperparameter. Accuracies were computed and compared for different cases as summarized in Table 1. Hyperparameter tuning was performed for three variations of the same dataset: the original dataset, undersampled dataset and the weighted classes dataset.

CONCLUSION

From this work, the following conclusions can be drawn. On the whole, the test cases which used the SGD optimizers offered lower accuracies as compared to the test cases that used the Adam optimizer. Secondly, the accuracies got comparatively better for the undersampled dataset when compared to the original dataset. These results go well with the hypothesis of the negative impact that dataset imbalance could have.

The undersampled case clearly loses valuable data. The weighted class neural network on the other hand doesn't lose any of the data but simply adjusts priorities or rather the weights on the nodes. This is done depending on their relative population ratio in the original dataset. From the results it can be seen that most cases perform better in the weighted class neural network case when compared to the original and undersampled cases.

Finally, by comparing the results tabulated in the above section, it can be observed that a maximum accuracy of around 86.55% is obtained. This is in the case where a dropout layer with a hyperparameter of 0.7 and the Adam optimizer are used on the weighted class neural network.

Dataset	Optimizer	Dropout	Epochs	Accuracy
Unbalanced	Adam	0.7	5	83.78
Unbalanced	Adam	0.7	10	70.38
Unbalanced	Adam	0.6	10	84.30
Unbalanced	SGD	0.7	10	36.6
Unbalanced	SGD	0.6	10	46.33
Reduced CNV	Adam	0.7	10	79.62
Reduced CNV	Adam	0.6	10	70.54
Reduced CNV	SGD	0.7	10	43.54
Reduced CNV	SGD	0.6	10	44.68
Weighted classes	Adam	0.7	10	86.55
Weighted classes	Adam	0.6	10	82.41
Weighted classes	SGD	0.7	10	46.33
Weighted classes	SGD	0.6	10	47.47
Pre trained weights	Adam	0.5	10	Overfitting

Table 1: Accuracies obtained on three variations of the dataset and by using the ResNet50 model's pre-trained weights.

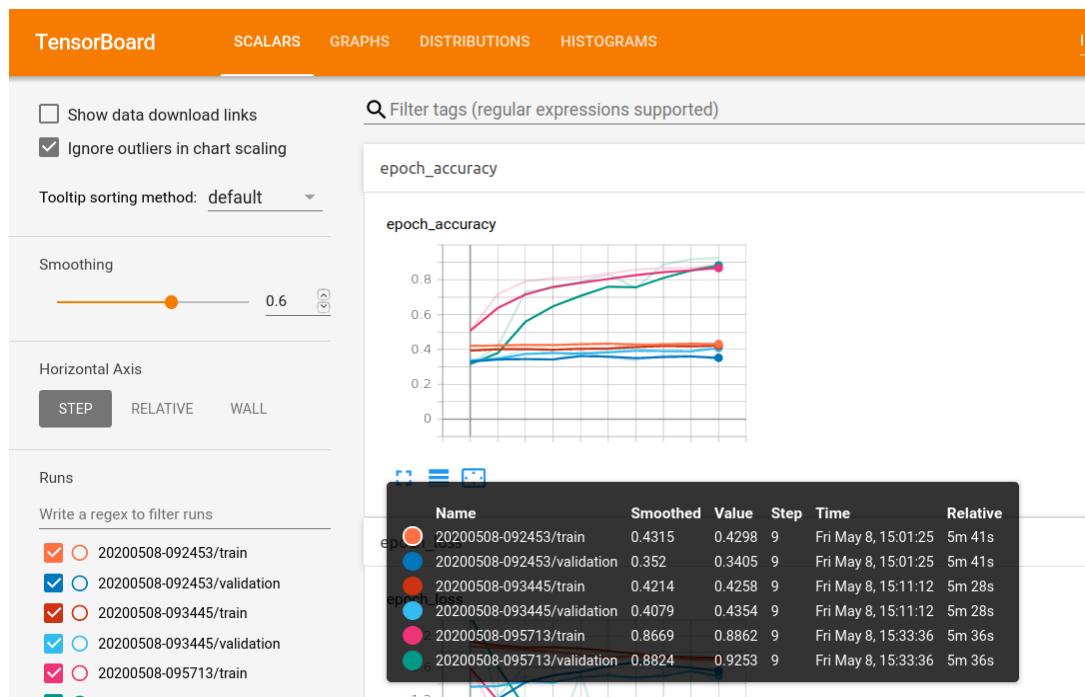


Figure 5: The epoch accuracy plot obtained by logging the results and displaying them using Tensorboard.

V.I. FUTURE WORK

This work has been limited to the usage of OCT scans for the estimation of the stage of disease in a person's retina. This can be extended further and tested on other forms of data (1.2). In addition to this, the motivation of making up for the lack of ophthalmologists in rural areas can be further realized by increasing accessibility. This can be done by means of a web or android application.

REFERENCES

- [1] Hwang DK, Hsu CC, Chang KJ, Chao D, Sun CH, Jheng YC, Yarmishyn AA, Wu JC, Tsai CY, Wang ML, Peng CH, Chien KH, Kao CL, Lin TC, Woung LC, Chen SJ, Chiou SH. *Artificial intelligence-based decision-making for age-related macular degeneration*. Theranostics, 2019
- [2] Soichiro Kuwayama, Yuji Ayatsuka, Daisuke Yanagisone, Takaki Uta, Hideaki Usui, Aki Kato, Noriaki Takase, Yuichiro Ogura, Tsutomu Yasukawa, Automated Detection of Macular Diseases by Optical Coherence Tomography and Artificial Intelligence Machine Learning of Optical Coherence Tomography Images, Thno.org, 2019.
- [3] Jason Brownlee, Image Augmentation for Deep Learning With Keras in Deep Learning, Machine Learning Mastery, September 13, 2019.
- [4] Susan Ruyu Qi, Machine Learning and OCT Images - the Future of Ophthalmology, Medium, Dec 8, 2017
- [5] Porto Seguro, Resampling strategies for imbalanced datasets, Kaggle, 2018.
- [6] Roy Swartz, Anat Loewenstein, Early detection of age related macular degeneration, International Journal of Retina and Vitreous, Article Number 20, December 01 2015.
- [7] Jason Brownlee, Random Oversampling and Undersampling for Imbalanced Classification, Machine Learning Mastery, January 15, 2020.
- [8] Jason Brownlee, How to Choose Loss Functions When Training Deep Learning Neural Networks in Deep Learning Performance, Machine Learning Mastery, January 30 2019.
- [9] Thanh Vân Phan, Lama Seoud, Hadi Chakor, Farida Cheriet, Automatic Screening and Grading of Age-Related Macular Degeneration from Texture Analysis of Fundus Images, Journal of Ophthalmology, Volume 2016, Article ID 5893601, March 21 2016.