Fundus Image-Based Detection of Ocular Diseases Using Deep Convolutional Neural Networks

Derek Huynh and Matthew Biniam
Holy Trinity Catholic Highschool, Colonel By Secondary School
March 2021

Table of Contents

1 Abstract

The following study examines one of the most substantial problems in the medical field, the diagnosis of ocular diseases. According to the WHO, over 2.2 billion people live with a vision impairment. Leading causes of vision complications include conditions such as diabetes, cataracts, glaucoma, and myopia. Unfortunately, many of these ailments only present symptoms at advanced and irreversible stages. Consequently, it is imperative to diagnose these diseases early to prevent long-term health-related burdens. We propose a novel detection system for multiple ocular diseases using a deep convolutional neural network. The model is capable of identifying several diseases from fundus images of the retina, achieving an AUC score of 0.75. This study aims to serve as an effective assistive diagnostic tool for eye-care professionals through the use of a CNN. We believe that our research shows an automated system for eye disease detection is possible for future clinical use.

2 Introduction

2.1 Background Information

According to Health Canada, over 47% of ocular disease cases are diagnosed 4 years late. Ocular diseases are conditions that interfere with a patient's eye health. Although not severely affecting life expectancy, vision complications heavily limit patients' daily lives. Current diagnostic criteria identify ocular diseases based on empirical observations from imaging techniques such as Optical Coherence Tomography. Unfortunately, these methodologies are heavily dependent on human analysis and interpretation, making them susceptible to both inaccuracy and inconsistency. This uncertainty has contributed to late diagnosis and a reduction in the efficacy of treatments.

The financial burden associated with multiple clinical visits is an additional problem with current diagnostic criteria. Present methodologies rely on obtaining numerous eye scans from expensive imaging technologies. Furthermore, retinal ganglion cells can be strenuous to empirically observe in the early-stages. As a result, diagnosis is often delayed, and by the time it is diagnosed, there may already be irreversible long-term effects like permanent vision impairment or complete vision loss.

Fig-1: 6-month increment of retinal ganglion cell destruction in Glaucoma Patient

2.2 Purpose

Thus far, ocular detection methodologies have largely been focused on empirical observations and human interpretation. This study aims to develop an accurate and effective assistive detection tool through the use of a deep convolutional neural network. The primary purpose of this project is to allow an accelerated and accurate method for the detection of the eight most common ocular diseases - Diabetic Retinopathy, Glaucoma, Cataracts, Age-related macular degeneration, Hypertension, and Myopia. These diseases were selected based on their early-detection difficulty and global prevalence.

3 Retinal Diseases

3.1 Disease Selection

The eight diseases we selected to focus on detecting are:

- Diabetic Retinopathy
- Glaucoma
- Cataracts
- Age-related macular degeneration
- Hypertension
- Myopia
- Other Diseases

These specific diseases were selected as the subjects of our early detection technology due to their high prevalence and widespread impact on Canadians. We henceforth believed there was a high demand for us to focus on these specific ocular diseases. Additionally, these are all diseases that have the potential to create massive vision losses if not detected early, and often do not reveal their symptoms until advanced and deadly stages. A notable disease as such is Glaucoma, which is the leading cause of blindness for people over the age of 60. Despite the cruciality that early detection of these diseases may potentially have in maintaining high-quality ocular health, many of their current chief diagnostic methods involve empirical observations of retinal images that are difficult and inefficient to gauge. For instance, these current retinal images do not provide detailed and clear representations of ganglion cell depletion, which obscures our understanding of neuronal activity between the retina and the brain.

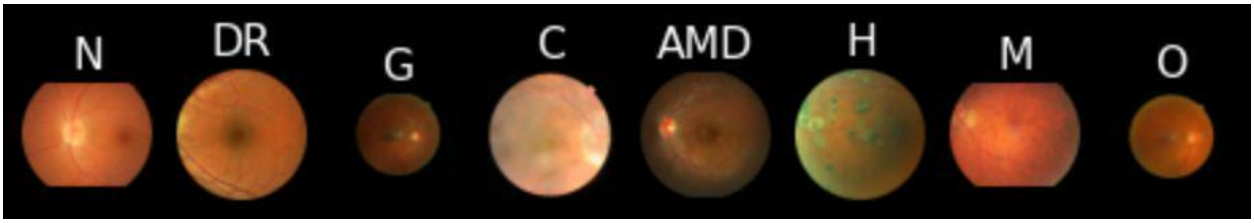

Fig-2: Sample Fundus-Images of 8 exploited retinal diseases

3.2 Diabetic Retinopathy

Diabetic retinopathy is a complication of diabetes that affects the eyes. It is the most common form of vision loss associated with diabetes and is the leading cause of blindness among working-class adults. It involves damages to the blood vessels of the retina. In diabetic retinopathy, an excess of sugar leads to the blockage of blood vessels that nourish the retina, and the subsequent new blood vessels the eye tries to make grow do not develop pro and suffer from leakages. Although it initially may cause minimal to no symptoms, it can eventually lead to blindness. Anyone who has type-1 or type-2 diabetes is susceptible to developing this condition. The most at-risk individuals are people who have had diabetes for longer periods and have less controlled blood sugar. As the condition develops common symptoms occur including floaters, blurred vision, fluctuating vision, impaired colour vision, and vision loss. As this disease shows minimal symptoms at first, it is recommended to complete eye checks regularly for people who suffer from diabetes. These eye checks can use retinal scans to be detected early. As diabetic retinopathy has virtually no symptoms at first, retinal scans can provide the urgent need for early detection to limit the potential of significant vision loss from the condition. This is made even more important since the damaging effects of diabetic retinopathy are irreversible, however, treatment can prevent a diabetic retinopathy patient's vision to worsen.

3.3 Glaucoma

Glaucoma is a group of diseases that damage the optic nerve. As this nerve gradually deteriorates, blind spots appear in a patient's visual field which contributes to vision loss. It is caused by high pressure in the eye and can be genetic. Glaucoma is believed to be the leading cause of blindness for people over the ages of 60. Like diabetic retinopathy, glaucoma begins with minimal indicative factors but can gradually develop and only become apparent once it affects patients in its advanced stages. To compound the deadliness of this disease, it is additionally very difficult to detect glaucoma through a retinal scan in the early stages. Since vision loss due to glaucoma can not be recovered, and the effects of glaucoma can be slowed down if detected early, the need for retinal scan technologies that can quickly and effectively detect glaucoma at the micron level is vast.

3.4 Cataracts

Cataracts refer to the clouding of normally clear lenses and have a high occurrence in people of advanced age. Cataracts often develop slowly and only interfere with vision, later on, making early detection imperative. Although cataract surgery is both safe and effective, early detection of cataracts can help optometrists prescribe the glasses that would allow a patient to have the ideal vision possible. Retina scans can help make cataract surgeries more effective.

3.5 Age-related macular degeneration

Age-related macular degeneration occurs when your macula, the small central portion of the retina, begins to wear down causing severe vision problems. The effects of this disease happen to the majority of individuals as they age and are the leading cause of severe, permanent vision loss for people over the age of 60. Despite this, it is rare for this disease to cause blindness. Since there is no cure for age-related macular degeneration, efficient early detection from advanced retinal scans can be pivotal when it comes to ensuring patients get the required treatment as early as possible and prevent further loss of vision.

3.6 Hypertension

Hypertension refers to the state when the pressure inside someone's eye or the intraocular eye, is higher than normal. Hypertension is important to detect early using effective retinal scan technologies, since it is either a cause or effect for a variety of ocular diseases. The most notable of which is glaucoma. In general, Hypertension gradually increases in individuals as we age.

3.7 Myopia

Myopia, also known as near-sightedness, is an ocular disease that occurs when the eyeball is too long or the cornea is too curved for light entering your eye to be focused properly. Detailed retina scans can effectively detect myopia, which is important since it will allow doctors to quickly and effectively gauge appropriate eyewear for people suffering from myopia so they can live with the best vision possible. It can also help optometrists do surgery for myopia to higher accuracy. Myopia is additionally a genetic disease, so widespread access to retina scans is important in order to uphold optimal eye health for individuals.

3.8 Other Diseases

These were sourced from the RFMID dataset. The ODIR dataset also contributed to the "other" category.

- ➔ Media haze (MZ)
- ➔ Branch retinal vein occlusion (BRVO)
- ➔ Tessellation (TSLN),
- ➔ Epiretinal membrane (ERM)
- ➔ Retinal traction (RT)
- ➔ Central retinal vein occlusion (CRVO)
- ➔ Tortuous vessels (TV)
- ➔ Exudation (EDN)
- ➔ Asteroid hyalosis (AH)

4 Retinal Scan Dataset

4.1 Data Collection

Our first step was data collection, where we collected fundus images from eight different datasets, from which we had to consolidate into a single large dataset. Two of the datasets contained images of multiple diseases while the other six were binary and only contained images of one specific disease. We then performed cursory data analysis on each of the datasets, inspecting their populations and searching through the images provided. Undesirable images (i,e those without annotations) were removed.

The use of eight different datasets(ODIR, RFMiD, IDRiD, Messidor-2, PALM, ADAM, G1020, REFUGE) to obtain our data was done to ensure our model had optimal robustness and was representative of the ocular diseases covered as a whole. These contained images of different conditions and settings, whereas most other studies conducted (such as Diaz-Pinto et al.) used few images (less than 4000) with idyllic conditions. This gave us the ability to generalize the conclusions we derived from our data analysis. Additionally, our cursory data analysis on each data set reduced excessive variance and offset anomalies that may have diminished the accuracy of our results. The various sources of our datasets additionally ensured that our data came from a large variety of people suffering from ocular diseases so that the potential of systematic factors could not diminish the accuracy of our results.

4.2 Data Cleansing and Pipelining

The cleaned datasets were then combined into a single CSV file containing the file name and 8 categorical labels for each image with the Pandas library. The number of images totalled 13,775. Next, we created data pipelines to efficiently preprocess and stream the data to our deep learning models. The images were cropped and resized to 244x244px and saved along with the labels into a TFRecords binary file to make our pipeline as efficient as possible.

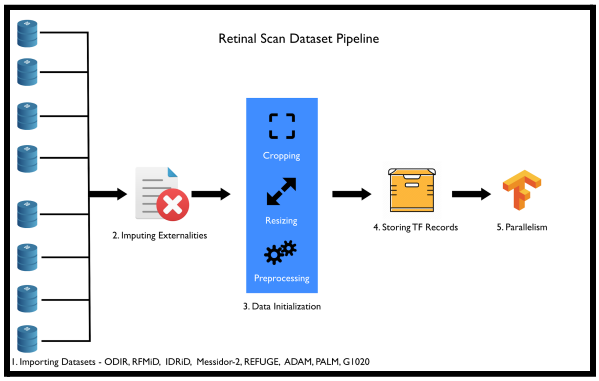

Fig-3: Retinal Scan Dataset Pipeline

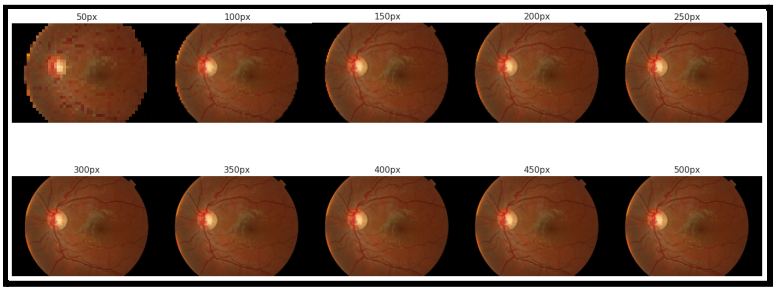

Fig-4: Retinal Scan Cropping and Resizing

5 Convolutional Neural Network Architecture

The performance of CNNs on large and complex image classification tasks has been well-documented, with far lower error rates on the premier computer vision competition the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) being achieved than with classical machine learning algorithms such as Support Vector Machines (SVMs). Given their efficacy with classifying images and their promising results in medical contexts, CNNs were determined to be the most suitable model for this study. We implemented two state-of-the-art CNNs, the first being VGG-19, which placed second in ILSVRC 2015 and the second being SE-Resnet-34, a variation of the Squeeze and Excitation networks which won first place in ILSVRC 2017. We implemented and trained each CNN using the TensorFlow and Keras deep learning libraries. All architectures contained a data augmentation layer prior to the first convolutional or residual block that randomly flipped the inputted image horizontally or vertically to regularize the model and artificially increase the number of training examples. The output layers for all models were an 8 neuron dense layer with the sigmoid activation function in order to give probabilistic outputs for each of the 8 classes. The ReLU (rectified linear unit) activation was used for the convolutional and fully connected layers. Early Stopping was implemented, ending training once validation loss had not decreased after the specified patience of 20 epochs. All models were trained on Nvidia Tesla P100 GPUs.

5.1 VGG-19

The VGG-19 architecture is a simple and classical architecture developed by the Visual Geometry Group for ILSVRC 2014. The main contribution of this architecture is the usage of much deeper networks than ever before, with their largest network proposing 19 sequential layers of convolutional and fully connected layers. Each convolutional block contains at least two convolutional layers with varying numbers of filters increasing by a factor of two after each block and all with a kernel size of 3×3. This small kernel size allows the network to better learn the low-level features of the images which it will then use in the top layers to construct high-level representations. Batch Normalization layers were placed after each Conv2D and FC layers to preserve the variance of the inputs and outputs. There is some debate surrounding the placement of the Batch Normalization layer and the ReLU nonlinearity activation function, with the original Batch Normalization paper recommending that they be placed before. We decided to measure the performance of both approaches in the VGG-19 architecture with no significant differences in the results. Additionally, a Normalization layer was placed before the first convolutional layer which simply standardizes the inputs by performing the following:

$$x = \frac{x-\mu}{\sigma}$$

Where x is the RGB image inputs scaled from 0 to 1.0, μis the mean and σis the standard deviation, both calculated from a sample of the data. Finally, dropout layers with a keep rate of $p = 0.5$were placed before the FC layers to regularize the model and prevent overfitting.

5.2 SE-Resnet-34

Our second model is a modification to the 2015 ResNet-34 architecture using the Squeeze and Excitation blocks proposed by Hu et, al. which they called SE-ResNet-34. The typical Residual Unit in a normal ResNet contains two main convolutional layers each using a kernel size of 3×3. Each Residual Unit has a skip connection where the inputs are added to the outputs of the main layer (they are also passed through a single convolutional layer if the stride is greater than 1). A consequence of adding the residuals to the outputs of the layers after the nonlinearities are applied is that it alleviates the issue of exploding and vanishing gradients which is prevalent in very deep neural networks. The main modification to this original architecture is the addition of SE blocks. They are essentially mini-neural networks that perform feature recalibration and are able to enhance the performance of preexisting architectures. They each consist of a GlobalAveragePooling layer (*squeeze*) which embeds the dependencies between the image channels to retain important global information. It then passes this information to two fully connected layers, one with the ReLU activation and the last with the Sigmoid activation (*excitation*). These are then multiplied with the outputs of the main Residual block and added with the output of the skip connection.

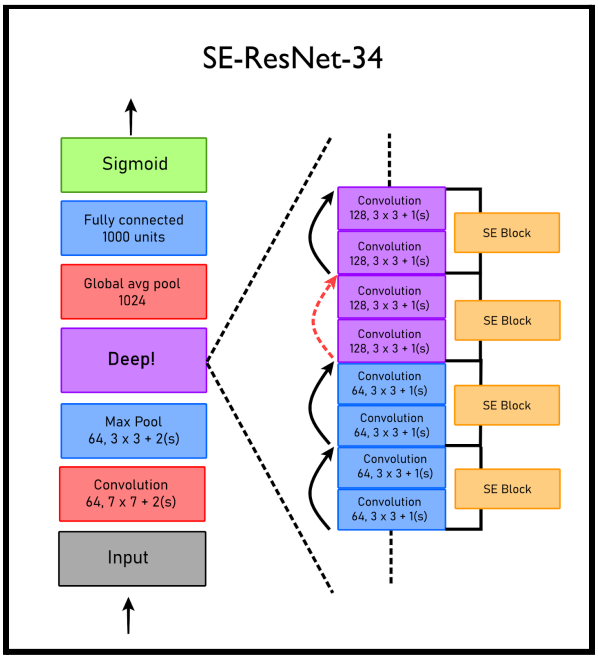

Fig-5: SE-ResNet 34

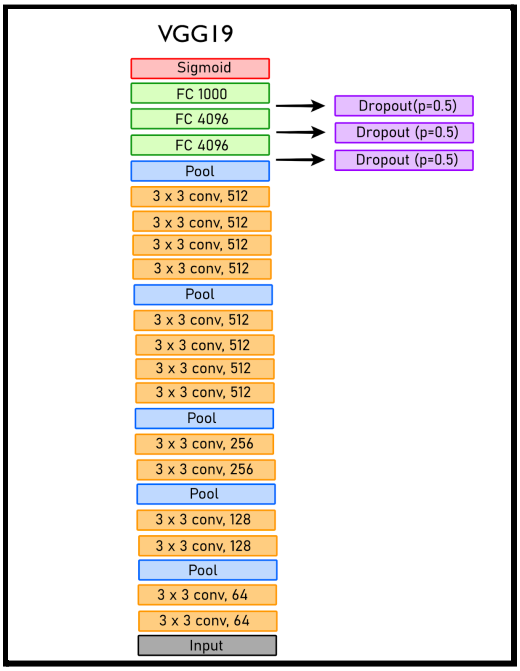

Fig-6: VGG19

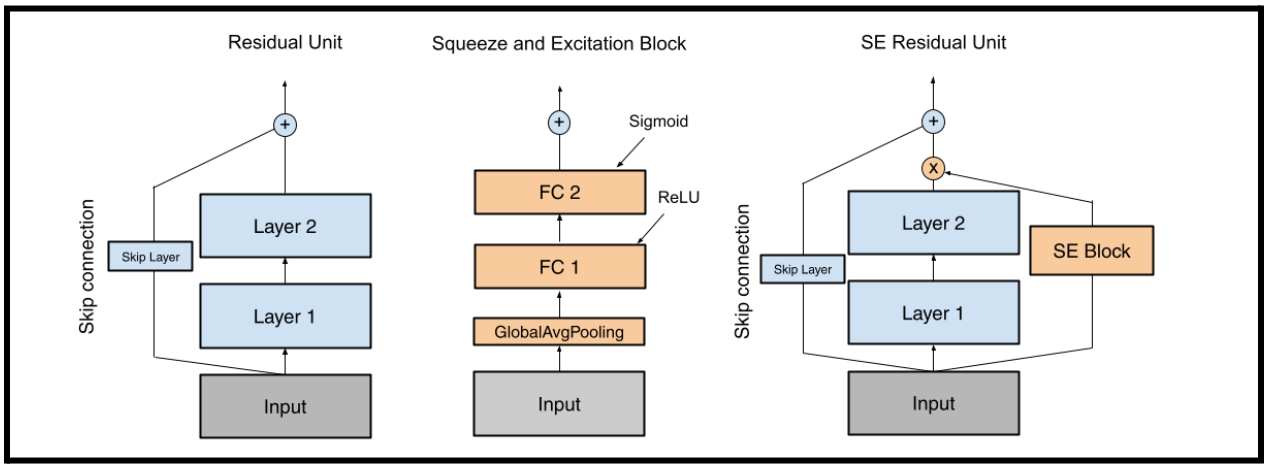

Fig-7: Residual Unit, Squeeze and Excitation Block, SE Residual Unit

5.3 Hyperparameters

5.3.1 Optimizer

Nesterov Adaptive moment estimation (Nadam) with a learning rate of 3e-4 was used along with *Stochastic Gradient Descent* (SGD) as optimizers for our model. The advantages of using Nadam are its fast convergence speed and ability to adapt its learning rate and ease of use. A normal Gradient Descent update would be described as such:

$$\theta_t = \theta_{t-1} - \alpha \nabla J(\theta_{t-1})$$

Where θ_t is the updated parameter, and $\alpha \nabla J(\theta)$is the learning rate times the gradient of the cost function w.r.t to the parameter θ_{t-1} . Usage of SGD can be cumbersome when considering the need to tune hyperparameters. Much of the optimizer's performance relies on selecting an optimal learning rate or the usage of a learning rate scheduler which can be time-consuming. Adam reduces the need for hyperparameter tuning by employing an adaptive learning rate algorithm which is as follows:

$$m \leftarrow \beta_1 m - (1 - \beta_1) \nabla_\theta J(\theta)$$

$$s \leftarrow \beta_2 s + (1 - \beta_2) \nabla_\theta J(\theta) \otimes \nabla_\theta J(\theta)$$

$$\hat{m} \leftarrow \frac{m}{1 - \beta_1^{\,t}}$$

$$\hat{s} \leftarrow \frac{s}{1-\beta_2^{\,t}}$$

$$\theta^t \leftarrow \theta^{t-1} + \alpha \hat{m} \oslash \sqrt{\hat{s} + \epsilon}$$

Adam works by storing an exponentially decaying average of past gradients and squared gradients $m$and s respectively. The parameters are then updated, scaling the learning rate based on m and s. Nadam is simply the Nesterov trick added to the Adam optimizer which is able to converge even faster. It on average reaches convergence faster than SGD and can even edge out SGD using momentum optimization with the Nesterov trick. However, SGD has been shown to have better generalization errors than adaptive optimizers. A cyclical learning rate scheduler called OneCycle described by Smith et, al. was used with the SGD optimizer to speed up convergence. The process involved training the model for one epoch while increasing the learning rate exponentially by a constant factor and plotting the loss. The optimal learning rate is slightly lower than when the algorithm diverges (the loss shoots up) and is chosen as the highest learning rate for the scheduler. At each epoch the loss is linearly increased from a learning rate 10 times less than the max to the maximum learning rate halfway through, after which it gets decreased back until the end of the epoch.

5.3.2 Batch Size

When examining the proper batch size for each iteration while training a neural network, there is some discussion on whether or not to use large (in the thousands) or small (no more than 32) batch sizes. In order to make use of parallelization as well as to keep gradient updates as recent as possible, we opted to use batch sizes of 32.

5.4 Model Comparison

* All calculated with optimized decision thresholds unless specified otherwise

Model	AUC	Precision	Recall	F1-Score
SGD-VGG-19	0.75	0.57	0.78	0.66
SGD-VGG-19 (No thresh.)	0.72	0.68	0.66	0.67
Nadam-VGG-19	0.74	0.59	0.76	0.66
SGD-SE-ResNet-34	0.70	0.53	0.69	0.60
Nadam-SE-ResNet-34	0.71	0.54	0.71	0.62
Class-Weight-SE-ResNet-34	0.70	0.49	0.73	0.59

No-ES-SE-ResNet-34	0.72	0.50	0.75	0.60
Oversampled-SE-Resnet-34	0.60	0.42	0.70	0.53

Fig-8: Comparison of the various models trained

Through developing and training our CNNs we were able to derive multiple important insights that will be crucial in any future application of clinically viable disease detection systems using fundus images of the retina.

Our models were trained on an 80/10/10 train/validation/test split, with the training dataset containing 11,020 images and the validation and test set containing 1377 and 1378 images respectively. We chose these splits to maximize the amount of training data available due to the relatively small size of our dataset. The validation set was used to monitor the generalization error while training and to tune the hyperparameters. Of the 7 model architectures and configurations that we tested, the VGG-19 model trained with the stochastic gradient descent optimizer achieved the performance, with an AUC score of 0.75, recall of 0.78, precision of 0.57 and an F1-score of 0.66 on the test set. These metrics were chosen based on their effectiveness in evaluating the performance of a model on imbalanced datasets, which is the case for this scenario.

Fig 9: Normal 48%, Diabetic Retinopathy 23% , Glaucoma 5%, Cataracts 2%, AMD 3%, Hypertension 1%, Myopia 5%, Other 17%.

To illustrate the difference between a biased and unbiased metric, the binary accuracy reported by the model is 91.3%, far more optimistic than the AUC and F1-score. The main consequence of the dataset's imbalance that causes this behaviour is that it incentivizes the model to predict the majority class, biasing the results towards the majority class to the detriment of predicting the minority classes. The metrics chosen eliminates this by considering what types of misclassifications the model is making (i,e false positives and false negatives).

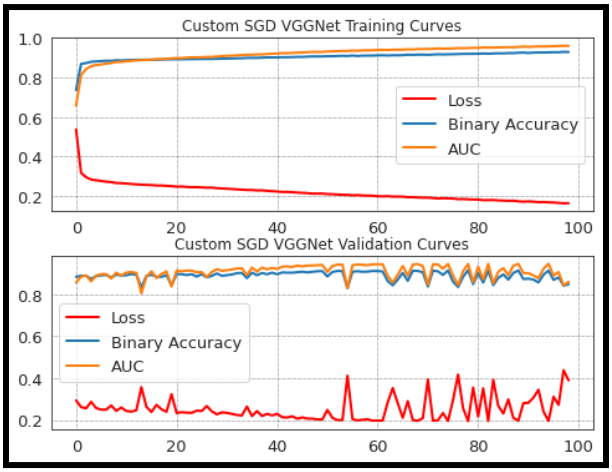

Fig-10: VGGNet Training Curve

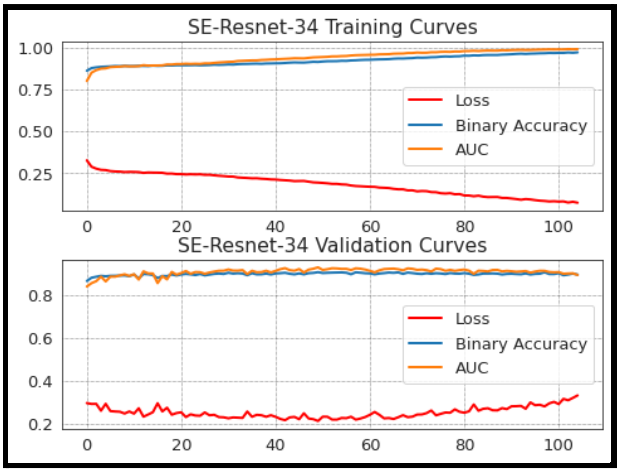

Fig-11: SE-Resnet-34 Training Curves

To address the class imbalance we used multiple techniques and compared their performance. These included balanced class weights, undersampling (reducing the number of examples from the majority class), oversampling (artificially increasing the number of examples from the minority classes) and adjusting the decision thresholds for each class. The latter proved to be the most effective, increasing the performance of the SGD-VGG-19 model from an AUC of 0.72 to 0.75. The thresholds were approximated using the validation set by iteratively searching for the threshold that produced the best F1-score for each class. It is worthwhile to note that the F1-score on the test set decreased slightly (0.67 vs 0.66) when using the new thresholds. It is expected for there to be slight variation with the evaluations across the test and validation set due to different images and distributions. The thresholds also changed the Precision-Recall trade-off, increasing the recall of the model (0.66 vs 0.78) while degrading the precision (0.68 vs 0.57). This is advantageous for detecting ocular diseases as a higher recall indicates a lower false-negative rate. As such, fewer patients are left undiagnosed, while conversely increasing false diagnoses. This achieves our goal of creating a model that addresses the early diagnosis problem concerning ocular diseases. Regardless, the higher false-positive rate still indicates that a careful final evaluation performed by trained eye-care professionals would still be required. Computing and applying balanced class weights to the model turned out to be too aggressive and

greatly degraded the performance of the model (an F1-score of only 0.59). Oversampling and undersampling were used as well to attempt to alleviate the class imbalance (randomly discard examples from the "normal" class and use data augmentation to create artificial examples of the minority classes) though this also proved to be ineffective. There is also a sizable gap in performance between the VGG-19 and SE-ResNet-34 architectures (F1-scores of 0.66 and 0.60 for the worst VGG-19 model and best SE-ResNet-34 respectively). It is possible that the complexity of the ResNet contributed to overfitting on the training set and a more diversified dataset would see significant performance increases. One thing to consider is that the SE-ResNet architecture is far more efficient with the parameters, only having 23,028,550 trainable parameters compared to VGG-19's 143,754,816 (about 84% less) making it faster at inference time as well as during training (128s/epoch vs 28s/epoch).

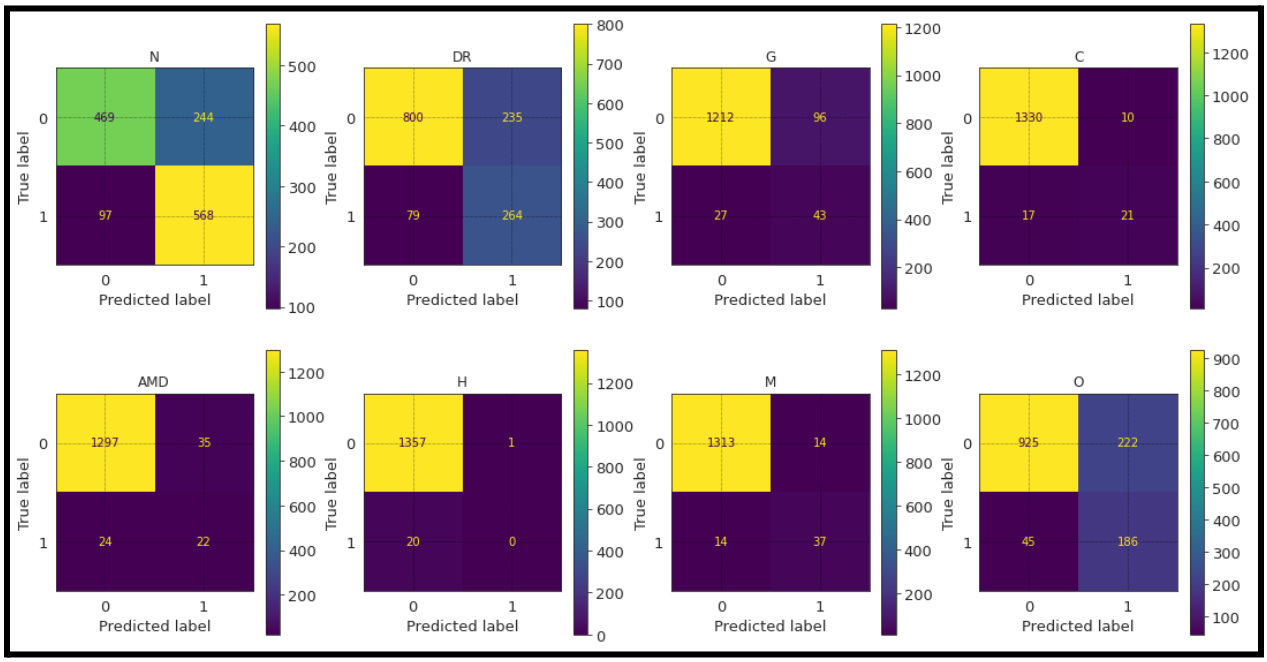

Fig-12: SGD-VGG-19 confusion matrix.

With regards to predictions for the individual classes, it appears that the model is most proficient at identifying fundus images with cataracts, myopia, and healthy images. Hypertension was the most difficult class, with none of the images in the test set being properly identified after adjusting the decision threshold. This likely indicates that it may be especially difficult to extract meaningful features from these images, and considering the small proportion of these classes collecting more samples would be necessary.

6 Conclusion

With the results and insights gained from developing our model, we can assert that an implementation of a predictive convolutional neural network for ocular disease detection is possible. The high-risk nature of the task necessitates more definitive positive findings on a more substantial population before a large-scale system can be considered as a reliable and credible clinical tool in a diagnostic setting, however, our study lends credence to the future adoption of such an application. We show that even with modest dataset size, a model can be produced to complement the prognosis of a trained eye-care professional and can be used to detect anomalous retinal fundus images as well as give confidence to their diagnosis. Our efforts to address the class imbalances in the dataset also underlines the need for more data in order to create an unbiased and generalizable model. Attempts to rectify the imbalances were insufficient in addressing the high variance of the data and the lack of training examples for the minority classes. We found that most techniques for imbalanced datasets failed to achieve sufficient performance and thus the most effective solution would be to gather additional data for the minority classes. Unfortunately many large fundus image datasets are proprietary and can only be accessed by paying a large fee. Thus our research indicates that advances in development with this domain would be greatly accelerated if more data with quality images and annotations were to be released publicly and free of charge.

References

1. Arjun, S., Saluja, K., & Biswas, P. (2021, January 04). Analysing ocular parameters for web browsing and graph visualization. Retrieved March 4, 2021, from https://arxiv.org/abs/2101.00794

2. Author Bill Holton, R., & Holton, B. (2017, April 12). Vision tech: Earlier eye disease detection may be possible thanks to new research. Retrieved March 27, 2021, from https://www.afb.org/aw/17/11/15390

3. Discovery Eye Foundation, D. (2016, March 24). The costs of eye care. Retrieved March 2, 2021, from https://discoveryeye.org/the-costs-of-eye-care/

4. Fu, Y., Li, F., Wang, W., Tang, H., Qian, X., Gu, M., & Xue, X. (2020, September 04). A new screening method FOR COVID-19 based ON OCULAR feature recognition by machine learning tools. Retrieved March 4, 2021, from https://arxiv.org/abs/2009.03184

5. He, K., Zhang, X., Ren, S., & Sun, J. (2015, December 10). Deep residual learning for image recognition. Retrieved March 2, 2021, from https://arxiv.org/abs/1512.03385

6. Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. (2019, May 16). Squeeze-and-excitation networks. Retrieved March 3, 2021, from https://arxiv.org/abs/1709.01507

7. Iskander, J., & Hossny, M. (2020, August 12). An ocular biomechanics environment for reinforcement learning. Retrieved March 27, 2021, from https://arxiv.org/abs/2008.05088

8. Jung, Y., Park, J., Low, C., Tiong, L., & Teoh, A. (2020, December 12). Periocular in the Wild Embedding learning with cross-modal Consistent Knowledge Distillation. Retrieved March 5, 2021, from https://arxiv.org/abs/2012.06746

9. Krishnan, A., Almadan, A., & Rattani, A. (2020, November 17). Probing fairness of mobile ocular biometrics methods across gender on visob 2.0 dataset. Retrieved March 4, 2021, from https://arxiv.org/abs/2011.08898

10. Langholz, E. (2019, May 22). Oculum AFFICIT: Ocular Affect Recognition. Retrieved March 6, 2021, from https://arxiv.org/abs/1905.09240

11. Li, N., Li, T., Hu, C., Wang, K., & Kang, H. (2021, February 16). A benchmark of Ocular DISEASE INTELLIGENT recognition: One shot for MULTI-DISEASE DETECTION. Retrieved March 4, 2021, from https://arxiv.org/abs/2102.07978

12. Macantosh, H. (2015, April 12). Multiple eye disease detection using deep neural network. Retrieved March 27, 2021, from https://ieeexplore.ieee.org/document/8929666

13. Meller, G. (2020, December 21). Ocular disease recognition using convolutional neural networks. Retrieved March 4, 2021, from https://towardsdatascience.com/ocular-disease-recognition-using-convolutional-neural-networks-c04d63a7a2da

14. Nations, U. (2019, March 12). Eye care, vision impairment and blindness. Retrieved March 2, 2021, from https://www.who.int/health-topics/blindness-and-vision-loss#tab=tab_1

15. Pantanowitz, A., Kim, K., Chewins, C., Tollman, I., & Rubin, D. (2020, November 24). Addressing the eye-fixation problem in gaze tracking for human computer interface using the Vestibulo-ocular reflex. Retrieved March 4, 2021, from https://arxiv.org/abs/2009.02132

16. Simonyan, K., & Zisserman, A. (2015, April 10). Very deep convolutional networks for large-scale image recognition. Retrieved March 2, 2021, from https://arxiv.org/abs/1409.1556
17. University of British Columbia, U. (2018, March 12). Color fundus Photography. Retrieved March 2, 2021, from https://ophthalmology.med.ubc.ca/patient-care/ophthalmic-photography/color-fundus-photography/
18. Zanlorensi, L., Laroca, R., Luz, E., Britto Jr., A., Oliveira, L., & Menotti, D. (2019, November 21). Ocular recognition databases AND Competitions: A survey. Retrieved March 3, 2021, from https://arxiv.org/abs/1911.09646
19. Zanlorensi, L., Lucio, D., Britto Jr., A., Proença, H., & Menotti, D. (2019, November 21). Deep representations FOR Cross-spectral Ocular Biometrics. Retrieved March 4, 2021, from https://arxiv.org/abs/1911.09509
20. Zhang, Z., Srivastava, R., Liu, H., Chen, X., Duan, L., Kee Wong, D., . . . Liu, J. (2014, August 31). A survey on computer aided diagnosis for ocular diseases. Retrieved March 7, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4163681/
21. H. He and E. A. Garcia, "Learning from Imbalanced Data," in IEEE Transactions on Knowledge and Data Engineering, vol. 21, no. 9, pp. 1263-1284, Sept. 2009, Retrieved March 6th 2021 from doi: 10.1109/TKDE.2008.239.
22. Fan, R., & Lin, C. (2007). A Study on Threshold Selection for Multi-label Classification.
23. Smith, L. N. (2018). A disciplined approach to neural network hyper-parameters: Part 1 -- learning rate, batch size, momentum, and weight decay. ArXiv:1803.09820 [Cs, Stat]. Retrieved March 6th 2021 from http://arxiv.org/abs/1803.09820

24. Kingma, D. P., & Ba, J. (2017). Adam: A method for stochastic optimization. ArXiv:1412.6980 [Cs]. Retrieved March 6th 2021 from http://arxiv.org/abs/1412.6980

25. Hoffer, E., Hubara, I., & Soudry, D. (2018). Train longer, generalize better: Closing the generalization gap in large batch training of neural networks. ArXiv:1705.08741 [Cs, Stat]. Retrieved March 6th 2021 from http://arxiv.org/abs/1705.08741

26. Krause, J. et al. Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy. Ophthalmology (2018). Retrieved March 6th 2021 from doi:10.1016/j.ophtha.2018.01.034

27. DuEE, B. (2017, March 12). Baidu research Open-Access dataset - Introduction. Retrieved March 2, 2021, from http://ai.baidu.com/broad/introduction

28. Guillaume PATRY, G. (2021, February 18). Messidor-2. Retrieved March 2, 2021, from https://www.adcis.net/en/third-party/messidor2/

29. None, I. (2018, April 12). IDRiD (Indian diabetic RETINOPATHY Image Dataset). Retrieved March 2, 2021, from https://academictorrents.com/details/3bb974ffdad31f9df9d26a63ed2aea2f1d789405

30. Challenge, G. (2107, November 12). Riadd (isbi-2021) - grand challenge. Retrieved March 2, 2021, from https://riadd.grand-challenge.org/Home/

31. Diaz-Pinto, A., Morales, S., Naranjo, V. et al. CNNs for automatic glaucoma assessment using fundus images: an extensive validation. BioMed Eng OnLine 18, 29 (2019). https://doi.org/10.1186/s12938-019-0649-y