[Team 17] Final Project Eye Cataract Glucoma Retina-Disease (ECGR) Classifier

Pramodini Karwande
M.S. in Statistics
North Carolina State University
Durhum, North Carolina
pkarwan@ncsu.edu

Md Farhamdur Reza
PhD in Electrical Engineering
North Carolina State University
Raleigh, North Carolina
mreza2@ncsu.edu

Rageeni Sah
M.S. in Electrical Engineering
North Carolina State University
Raleigh, North Carolina
rsah@ncsu.edu

***Index Terms*—Eye Diseases classifier, Convolution Neural Network, Imbalance Classes, Predictive modeling**

I. Motivation

Eye diseases like cataract, glaucoma, and retina diseases causes more of blindness to people. Cataract, is a clouding or dulling of the lens inside the eye, which leads to a decrease in vision and is considered as the most common cause of blindness. An example of caratact eye is shown in Fig. 1. Thereby, such eye diseases cause blurred vision, reduced color intensity, increased sensitivity to glare from lights while driving and reduces night time vision. It is generally found among people above age 55 but may occasionally occur in infants and kids. When an eye disease progresses to the point that it affects a person's ability to do normal everyday tasks, surgery may be needed.

Cataract, glaucoma, and retina diseases can not be diagnosed visually and are confirmed only through progressive examination of the eyes by the experienced ophthalmologists. Given that there are only 300,000 optometrists globally, where half of them are located in developing countries, and we are in need of more than 1 million optometrists [1]. Moreover, the World Health Organization estimates that around 18 million people suffering from blindness due to cataracts [2]. Cataract patients in under-developed area hardly get a chance to receive treatment in hospital, because of the limited healthcare resources and economic consideration. The less number of ophthalmologists availability adds to the existing challenges as the large scale screening of the eye disease demands early diagnosis.

There are mainly four categories of cataract detection and grading: Light-focus method, Iris image projection, Slit lamp examination, and ophthalmoscopic transillumination. Manual assessment is still not as effective as it is subjective and these methods consumes more time, energy and cost. So the main application of this work is to provide an easy diagnosis method for cataract. Hence to cost effective and save time and energy of highly knowledgeable ophthalmologists, our Neural Network model will help to classify state of the eyes having eye disease.

Fig. 1: An example of a normal eye and a cataract eye. A cataract eye has blurred vision. (Image is taken from Google).

To achieve our goal, if we deploy our generated Convolution Neural Network model using as software or make it available through the phone app, people can easily take a picture and get the status of the eye in the initial stage. This will help to identify Cataract, glaucoma, retina diseases at early stage and will help to take precautions, This will be also be cost effective.

In this work, the model classifies whether a person's eye is normal or having cataract, glaucoma, retina diseases. This implies that the classifier classify, if a person's eye is not normal, the type of problem in the eye. Cataract dataset [3] from Kaggle will be used to train our model. To achieve this task, we have used two-dimension Convolution Neural Network model and evaluated the performance using the given dataset.

The main technical difficulty that is due to the number of available samples of the cataract dataset. Moreover, the dataset is imbalanced in terms of the number of normal images and the number of images for each type of eye problem. Besides, the images are of varying sizes and have been captured with varying illumination.

II. Dataset

The dataset used in this project is available on Kaggle - a popular online platform for machine learning project and tagged as caratact dataset. The cataract dataset consists of 601 retina images, and are classified into four categories: normal, cataract, glaucoma, and retina disease shown in Fig. 2. We found this dataset is highly imbalance. The datasets contain number of normal retina images is 300, and the number of images with cataract, glaucoma and retina diseases are 100, 101 and 100, respectively. Furthermore, it is found that the dimension of the given dataset varies from around 1848×1728 pixels to 2592×1728 pixels. The dataset not only contains less number of images but also the images within the same classes are found highly anomalous. Fig. 3 demonstrate the anomaly of the dataset within the same class. From this figures, one can observe that the retina images are highly differ form one another as some images contains dark spot and some are not. Moreover, some images are clear and some are opaque.

To overcome the issue with data, different approaches has been applied which will be discussed in the subsequent sections. Image has resized to 150 X 150 pixels. Moreover weights has been assigned to each classes which helps classes with lower number of samples having more weight and vice versa.

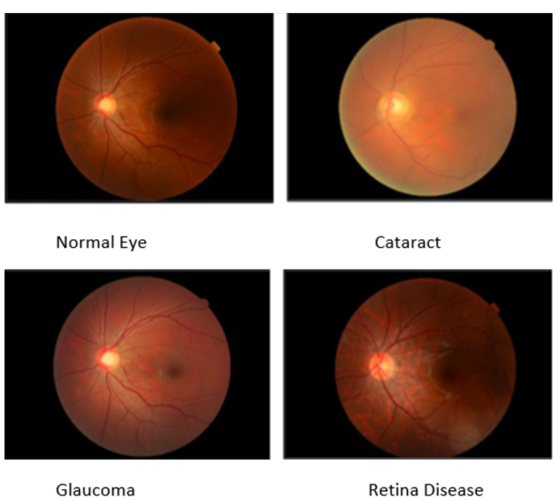

Fig. 2: Example of each of the classes from cataract dataset.

III. Methodology

In this project the goal is to build a multi-class supervised model using convolution neural network Fig. 4 that can classify an eye retina image into either of the classes: normal, glaucoma, cataract or diseased retina with higher accuracy and precision. The dataset is trained on different neural networks architectures (difference in terms of hyper-parameters). The baseline model is taken from Kaggle.com [4]. The baseline models are trained using sklearn framework, but our working framework will be PyTorch. By comparing the performance of the trained neural network model with the baseline model, a neural network model is proposed for the Cataract dataset.

Fig.: Training samples for normal class

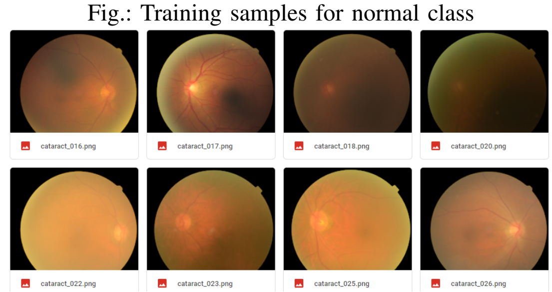

Fig.: Training samples for cataract class

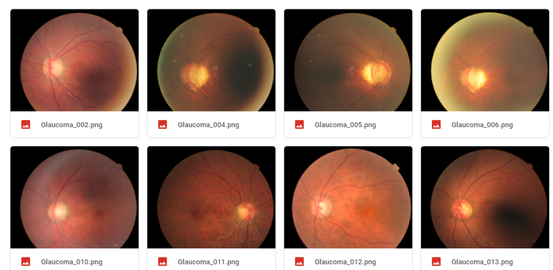

Fig.: Training samples for glaucoma class

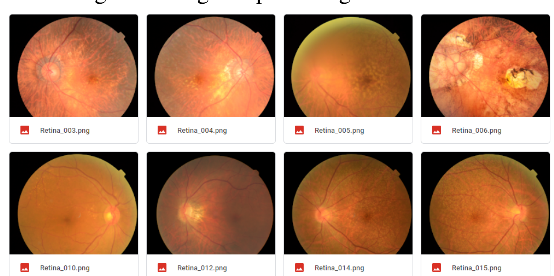

Fig.: Training samples for retina disease class

Fig. 3: Anomaly of the samples under same class. Images with varying illumination.

The key challenges working with this dataset are: the number of samples of the dataset is quite less to train a model, the four classes of the dataset are highly imbalance and the images within the same classes have varying illumination. To improve the performance with the current dataset, different techniques have been applied.

A. Imbalance Data handling

The imbalance data is handled using different weights of each of the classes in calculating loss. The lower the number of images in a particular class, higher the weight is applied. For a K class dataset, the weight for a particular class with n_i samples can be calculated as,

$$W_i = \frac{\sum_{k=1}^{K} n_k}{n_i} \tag{1}$$

where W_i is the weight of the $i-th$ class.

B. Regularization

As the size of the dataset is quite low, there is a chance of getting over-fitted model easily. So, to get rid of the over-fitted model, different regularization techniques have applied and observed the performance. First we have started with batch-normalization, and it has been observed that with the application of batch normalization technique, the prediction accuracy of the model has improved. Dropout, another regularization technique, was also applied to observe the performance of the model. However, application of the dropout technique didn't improve the performance of the model. Finally, to resolve the over-fitting issue, l_2 regularization has been used to get better accuracy from the model. The cost function using l_2 regularization can be expressed as

$$C(\theta) = L_y(\hat{y}) + \lambda \sum_j \theta_j^2 \tag{2}$$

where, is called the regularization parameter and λ is manually tuned. Also, $\lambda = 0$ then the above loss function acts as Ordinary Least Square where the high range value push the coefficients (weights) 0 and hence make it not suitable.

C. Model Selection

In selecting a model, the prime criterion is the model with highest validation accuracy. All the baseline models trained with cataract dataset are classical models: Logistic Regression, Random Forest, Gradient Boosting and Support Vector Machine. The highest validation accuracy was obtained using the Gradient Boosting model and it is about 56%. To improve the validation accuracy of the Cataract dataset, we have mainly implemented neural network model. First, fully connected neural network is used to evaluate the performance. Later, Convolution Neural Network (CNN) model is used to further ameliorate the validation accuracy.

Two convolution neural layer followed by three fully connected layers excluding the input layer has been taken as a final model to evaluate the performance. The number of kernel for the first and second convolution layer is 5 and 4 respectively, where the number of input channels is 3. In-between two convolution operation, max-pooling operation was conducted with the max-pooing size of (2,2). The parameters of the model was selected after a number of experiments by varying different parameters.

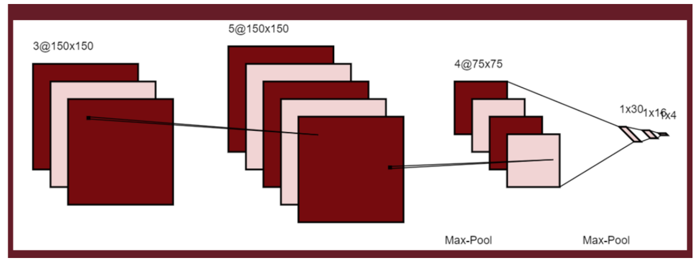

Fig. 4: Proposed CNN model

D. Model Training

To train the model, the whole dataset is divided into training and validation set with the ratio of 7:3. The images for training is used to train different neural network model and the validation images are used to validate the performance. First, we trained the model without using weight to handle imbalance data and regularization to handle the over-fitting issue. Then different parameters like number of neurons per layer or even number of layer were varied to get optimal performance. Then we come up with a model that detect the images better than the other models.

So, to further improve the performance, class weight was estimated and incorporated with the model. Moreover, different hyper-parameter tuning like batch-normalization, dropout, maxpooling and l_2 regularization was done. After performing all the experiments, it was found the proposed model's validation accuracy is about 89% which is much higher than the baseline model with highest validation accuracy.

IV. Evaluation

Model performance is evaluated using validation dataset which is 30% of the total training dataset. Accuracy is the primary metric used for baseline model evaluation. Considering dataset has imbalanced classes, F1-score is additionally used as performance metric. Hence, accuracy and F1-score both are used for model evaluation during training, however, results are compared only using accuracy with baseline work. Table I compares training and validation accuracy of the existing and proposed models.

TABLE I: Results Comparison: baseline models and proposed model training and validation accuracy

Classical Model	Training Accuracy	Validation Accuracy
Logistic Regression	0.604	0.607
Random Forest	0.604	0.607
Gradient Boosting	0.690	0.646
Support Vector Machine	0.531	0.569
Proposed Model (CNN2D)	1.000	0.890

Among the baseline models highest validation accuracy is 64.4% achieved using gradient boosting algorithm. The proposed model has validation accuracy of 89%. The best model parameters consists of training model for 80 epoch with hundred images in a batch along with adam optimizer.

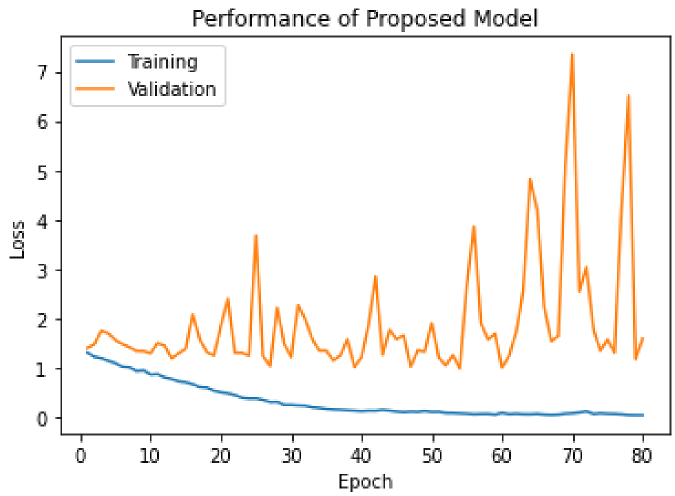

Fig. 5: Loss versus Epochs curve for the proposed model

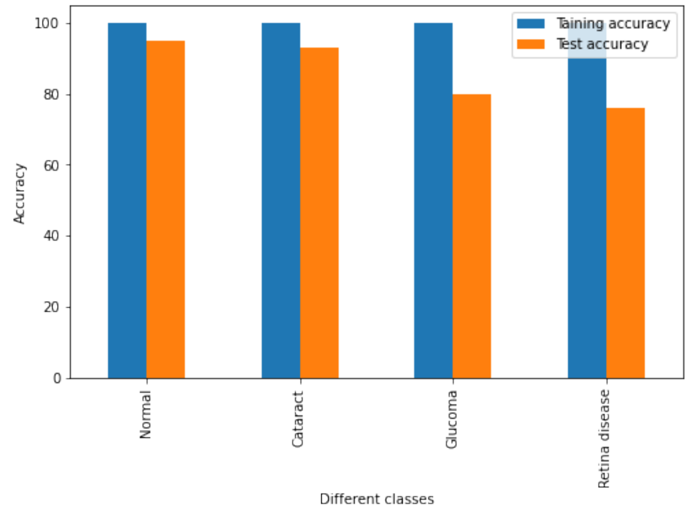

Fig. 6: Training versus validation accuracy

TABLE II: Experiments performed to obtained the best result

Model Parameters	Training Accuracy (%)	Validation Accuracy (%)
Optimizer: SGD lr: 0.0001 epochs: 80 batch size: 130 momentum: 0.9	Normal: 81 Cataract: 82 Glaucoma: 75 Retina disease: 42 Overall: 70	Normal: 77 Cataract: 53 Glaucoma: 35 Retina Disease: 20 Overall: 56
Optimizer: Adam lr: 0.01 epochs: 80 batch size: 130	Normal: 79 Cataract: 81 Glaucoma: 72 Retina disease: 65 Overall: 68	Normal: 50 Cataract: 86 Glaucoma: 45 Retina Disease: 0 Overall: 46
Optimizer: Adam lr: 0.001 epochs: 80 batch size: 100 Batch norm.	Normal: 95 Cataract: 100 Glaucoma: 68 Retina disease: 44 Overall: 79	Normal: 87 Cataract: 76 Glaucoma: 12 Retina Disease: 3 Overall: 59
Optimizer: Adan batch size: 100 epochs: 80 dropout Batch norm.	Normal: 89 Cataract: 87 Glaucoma: 20 Retina disease: 42 Overall: 63	Normal: 88 Cataract: 76 Glaucoma: 3 Retina Disease: 3 Overall: 58

As a part of improvement of accuracy, we have experimented with different optimizers, regularization techniques and parameter variation. Table II demonstrate some of the experiments and the corresponding results. By using Adam optimizer though the overall validation accuracy is 46%, in-contrast to using sochastic gradient descent optimizer with 56% overall validation accuracy, the performance of using Adam optimizer is increased to 59% by incorporating normalization technique. However, with the application of dropout in our proposed model, it is found the the performance is deteriorating to 58%. Finally, when we applied the l_2 regularization technique the best desired performance is obtained. The best validation accuracy found from the proposed model is 89%. All the experiments are conducted with 80 epochs and different batch size. At the time of training the best model corresponding to the validation set was saved.

Fig. 5 demonstrate the number of epochs versus Loss curve. From this figure, one can observe that the amount of loss for the training samples is decreasing with the number of epochs and it is showing steady behavior. On the other hand, for the validation samples, though the loss is decreased with epochs initially, after certain number of epochs it is showing abnormal behavior. The main reason of this behavior is the type of dataset. If we closely inspect the dataset, for any given classes of image there is a huge variation. Some images are clear and some are not. Moreover, the color and light intensity of those images are abrupt.

Fig. 6 depicts the the comparison between training and test accuracy of the proposed model for the four classes. The best obtained training accuracy for all the classes is 100% and the validation accuracy for the normal, cataract, glaucoma and retina disease are 95%, 93%, 80% and 76%, respectively, and that is quite high than the baseline models.

V. Future Scope

The proposed model performance can further be improved by handing known challenges explored in the project. For any deep learning project, dataset size is presumed to be large enough but the current dataset has 601 images which is highly lower than expected. An intensive work can be done to replicate or generate more images by using available images without negative impact and model parameters can be re-tuned with new datasets.The further improvement can be made using transfer learning. A number of models pre-trained on images datasets are available. Existing pre-trained architectures such as ResNet, VGGNet and Inception are trained on Imagenet Large Scale Visual Recognition Challenge (ILSVRC) dataset

[5]. Current dataset can be trained using these existing architecture to avoid existing overfitting problem. The proposed architecture challenges can be explored further using Bayesian Neural Networks (BNN). BNN [6] can be used to correct the impacts of model uncertainty and data uncertainty and, thereby, improving model robustness towards uncertainty.

References

[1] World Health Organization. Global initiative for the elimination of avoidable blindness. WHO/PBL/97.61 Rev 2.2006. Available from : *http://www.who.int/blindness/Vision2020_report.pdf*.

[2] Vision Loss Expert Group of the Global Burden of Disease Study. Causes of blindness and vision impairment in 2020 and trends over 30 years: evaluating the prevalence of avoidable blindness in relation to "VISION 2020: the Right to Sight". Lancet Global Health 2020. doi.org/10.1016/S2214-109X(20)30489-7

[3] Jr2ngb. (2019, August) Cataract Dataset, Version 2. Retrieved Sep 25, 2021, from *https://www.kaggle.com/jr2ngb/cataractdataset*.

[4] John Doel. (2019 October). Cataract Data Exploration with ML models, Version 1. Retrieved Sep 25, 2021, *www.kaggle.com/tsantra/cataract-data-exploration-with-ml-models*

[5] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248–255)

[6] Bishop, C. M. (1997). Bayesian Neural Networks. Journal of the Brazilian Computer Society, 4(1), 61–68. *https://doi.org/10.1590/S0104-65001997000200006*