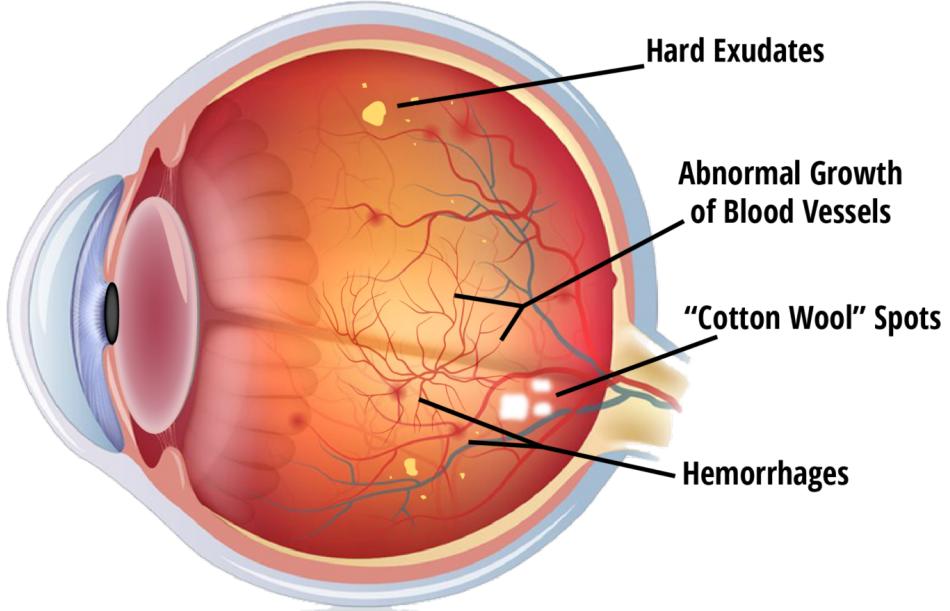


**Healthy Eye**



**Diabetic Retinopathy**

# Diabetic Retinopathy Detection

AUTHOR: SHIPING LI

# Background & Goal

## ► **Background**

Patients living in India rural areas have a higher risk of developing blindness. Even if screening is conduct successfully, the screen images are required professional doctors to review one by one and then provide the diagnosis.

To avoid more people missing the optimal treatment time, Aravind Eye Hospital in India captured screen images from rural areas and hoped to leverage technology to screen images automatically and gain information on the severity of the condition.

## ► **Goal**

Leverage deep learning to accelerate the diabetic retinopathy detection.

# Data Info

- ▶ Source: Captured from the rural areas by Aravind Eye Hospital  
Obtained from Kaggle Platform
- ▶ Total images: 5590 images
  - Training Set: 2930 images
  - Validation Set: 732 images
  - Test Set: 1928 images

Each image comes with a unique id number and is rated on a scale of 0 to 4 by a clinician, based on the severity of diabetic retinopathy. The rest images are test data.

# Target Value

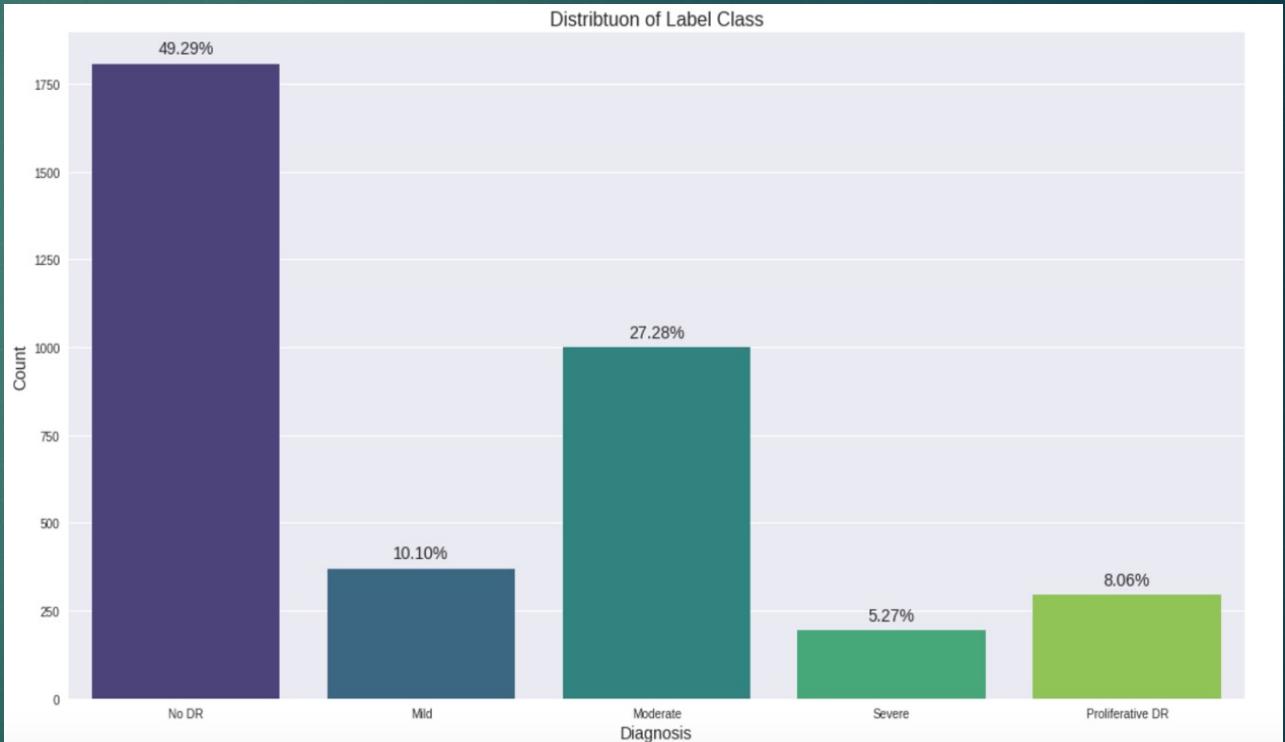
Since our goal is to detect if a fundus image has diabetic retinopathy and its possible condition, our target value is exactly the categorical five stages of the disease.

- ▶ stage 0 – No DR. No DR indicates healthy eyes.
- ▶ stage 1 – Mild. The earliest stage of diabetic retinopathy is often where swelling begins in retina's vessels.
- ▶ stage 2 – Moderate. In this stage, increased swelling of microscopic blood vessels begins to block blood flow to the retina, preventing proper nourishment.
- ▶ stage 3 – Severe. A larger section of blood vessels in the retina become blocked, resulting in insufficient blood flow to this area. The body receives signals to start growing new blood vessels in the retina.
- ▶ stage 4 - Proliferative diabetic retinopathy. This is an advanced stage of the disease in which new blood vessels grow in the retina, causing blurriness, a limited field of vision, and even blindness.

# Data Exploratory analysis

## 1) Distribution of Label Class

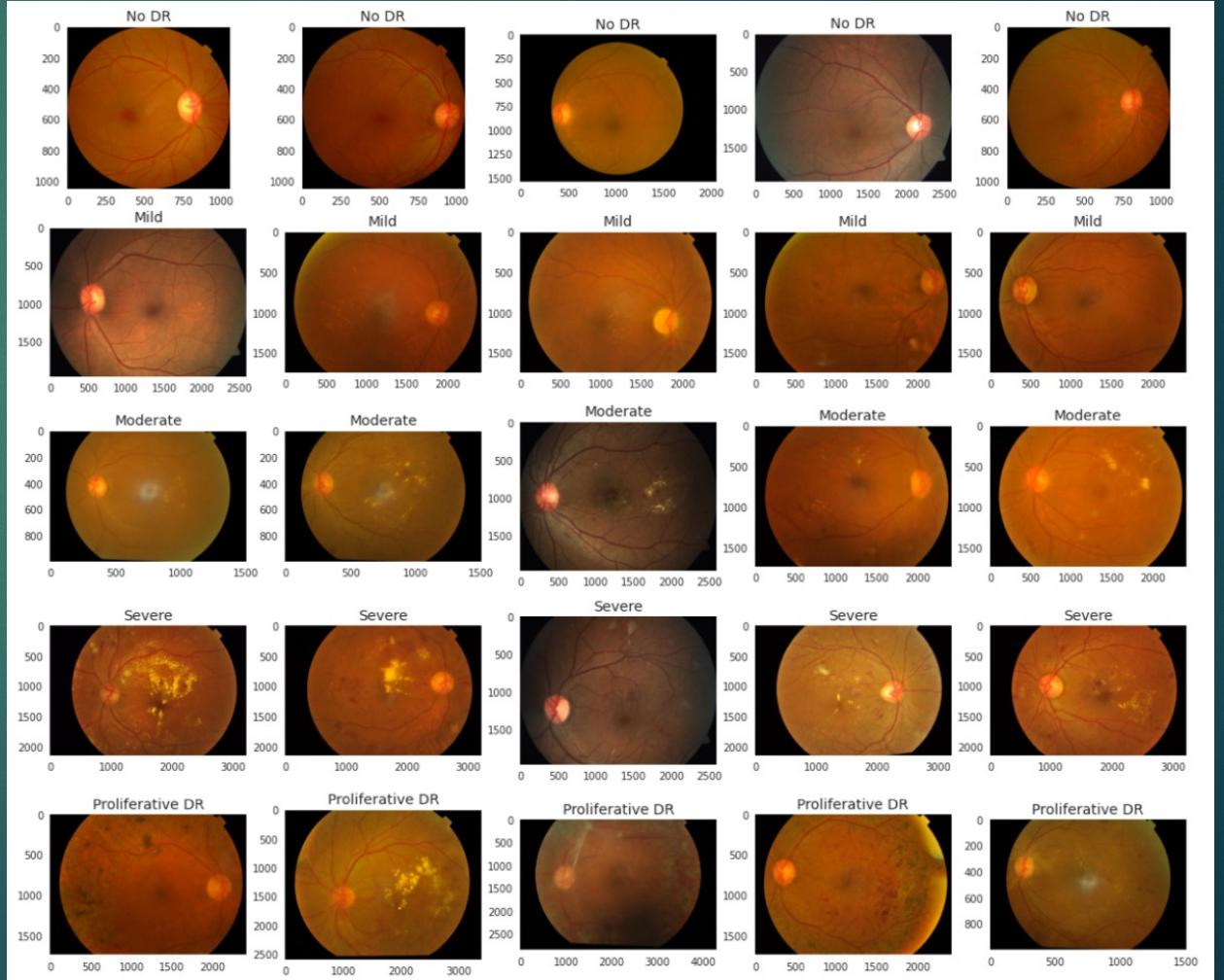
- The label class of severity is distributed unequally.
- No DR and Moderate accounts add up to 77% approximately.
- Problem: People with diabetic retinopathy will be wrongly informed of no disease.



# Data Exploratory analysis

## 2) Samples of fundus images across different severity

- The sizes are varying. Need to resize to the same dimension
- Some images are low-quality eg: low brightness, saturation
- As the severity level increases, the fundus images tends to have more light spots and be cloudy.



# Data Exploratory analysis

## 3) Detect low-quality images and enhance quality

Method : converted the images from RGB to HSV color.

(H): hue – color  
(S) : Saturation - greyness  
(V): Value – Brightness

Mean(Saturation) < threshold1  
Mean(Value) < threshold2



**Bad Quality**

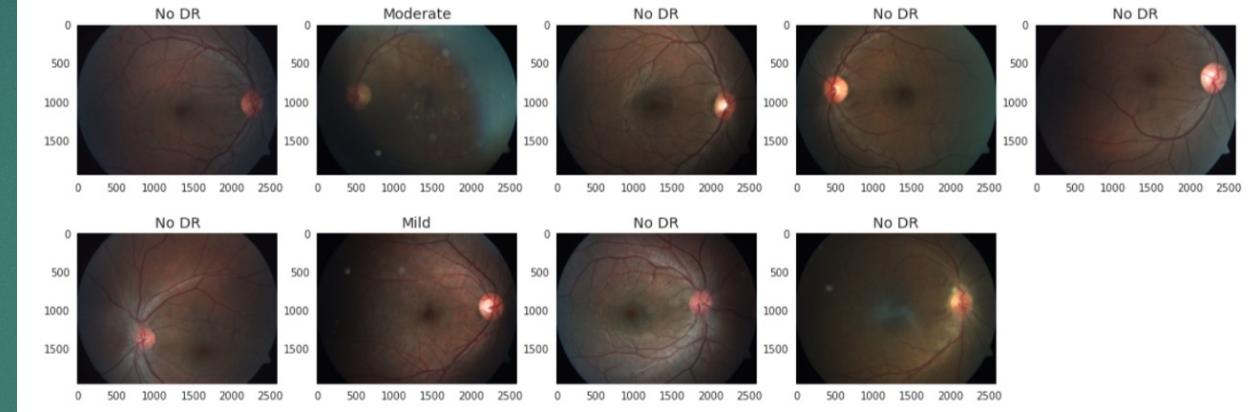


Scale their Values and  
Saturation by a factor > 1

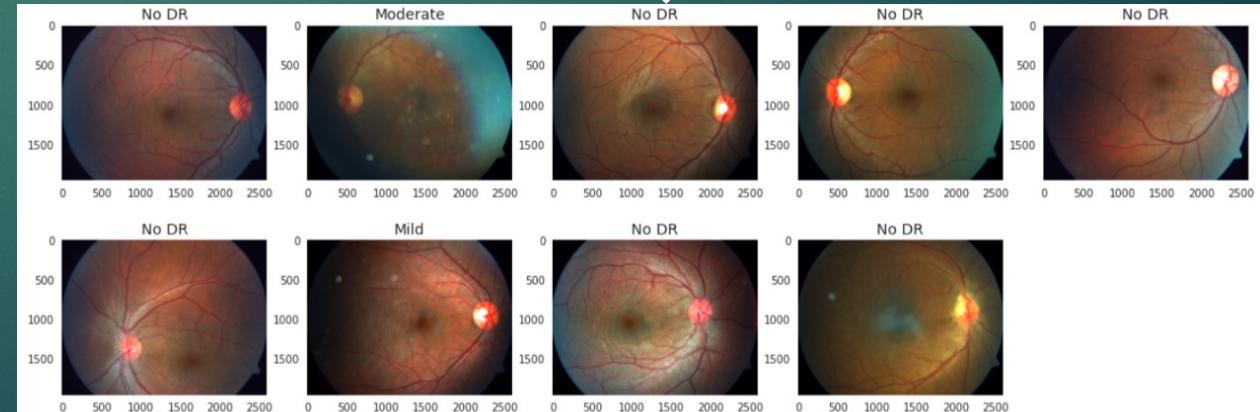


**Enhanced Quality**

Before



After



# MODEL TRAINING AND TUNING

## 1) Basic Model Setups

- ▶ RestNet50 : has the advantages of **increasing the network depth** while using **reducing the complexity** of the model with fewer parameters.
- ▶ Metric: Quadratic Weighted Kappa (QWK)

Why?

Take the similarity of the predicted rating and actual rating into account. The 5 stages of retinopathy are hierarchical, so the gap between the predicted and the actual rating is also meaningful.

Give higher credits to those predictions closer to the actual rating, compared to the predictions further from the actual rating.

eg: actual – Proliferative DR, Predicting Severe vs Predicting No DR

# MODEL TRAINING AND TUNING

## 2) Warmup Training (Freeze layers except the last 5)

Purpose : avoid the model weights fluctuating drastically and deviating from the optima in early training.

What we do?

Start with small learning rate  $1/5e^{-3}$



Increase the learning rate by  $1/5e^{-3}$  at the end of each epoch

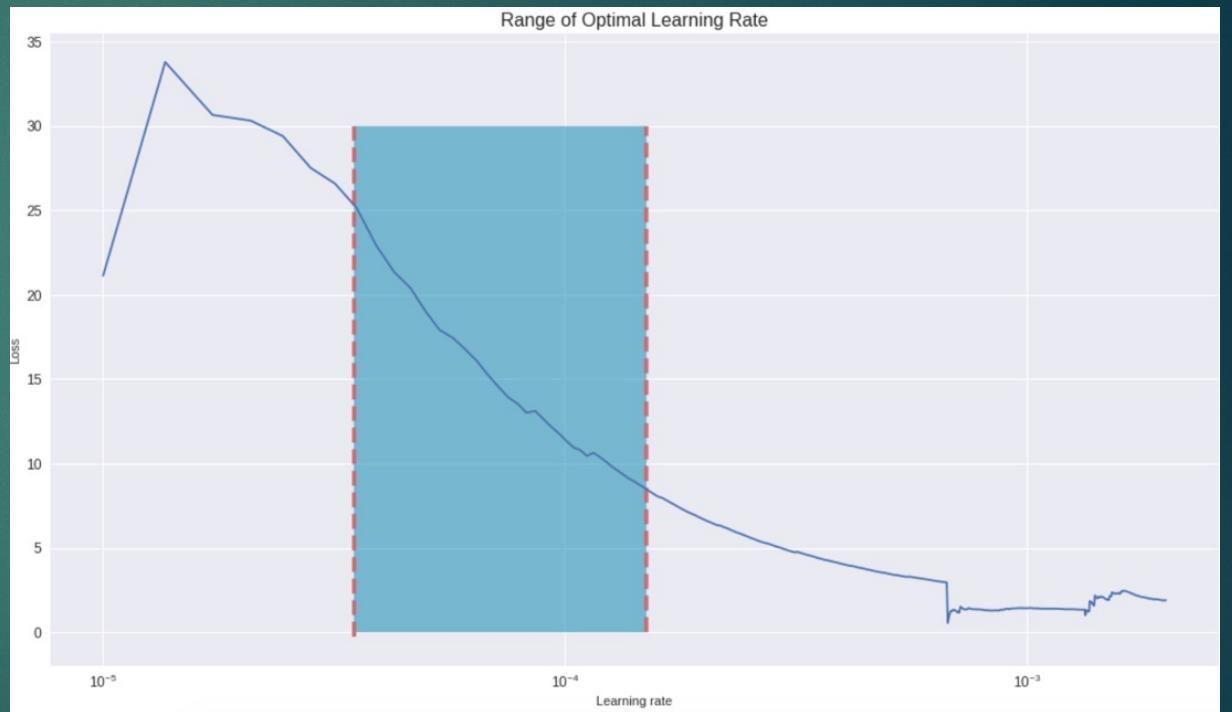


End with the “conventional” learning rate  $1/e^{-3}$  through 5 iterations

# MODEL TRAINING AND TUNING

## 3) Find the Optimal Learning Rate Range (**unfreeze all layers**)

- ▶ Search learning rates between  $1e^{-5}$  and  $1e^{-2}$
- ▶ Begin with  $1e^{-5}$ , increasing along after every batch
- ▶ Optimal learning rate range where loss decreases most rapidly.



# MODEL TRAINING AND TUNING

## 4) Fine Tuning

### Some Setups

- Learning Rate:  $6e^{-5}$
- Decay Rate: 0.5
- Epochs: 30
- Early Stopping Monitor: val\_loss

### Candidate Model 1

Val\_loss achieves the best at the 9th epoch and the corresponding QWK score reaches its first peak after a rapid climb

The gap starts diverging after the 9<sup>th</sup> epoch.



# MODEL TRAINING AND TUNING

## 5) Candidate Model 2

Our dataset was imbalanced as mentioned above, so I was interested in if adding more weights to minorities would lead to better results.

VS

Candidate Model 1:

Averaged Cross-Entropy Loss Function

Candidate Model 2:

Weighted Cross-Entropy Loss Function

Best val\_loss at the 18<sup>th</sup> epoch. Results is getting stable after 11<sup>th</sup> epoch though.



# MODEL EVALUATION

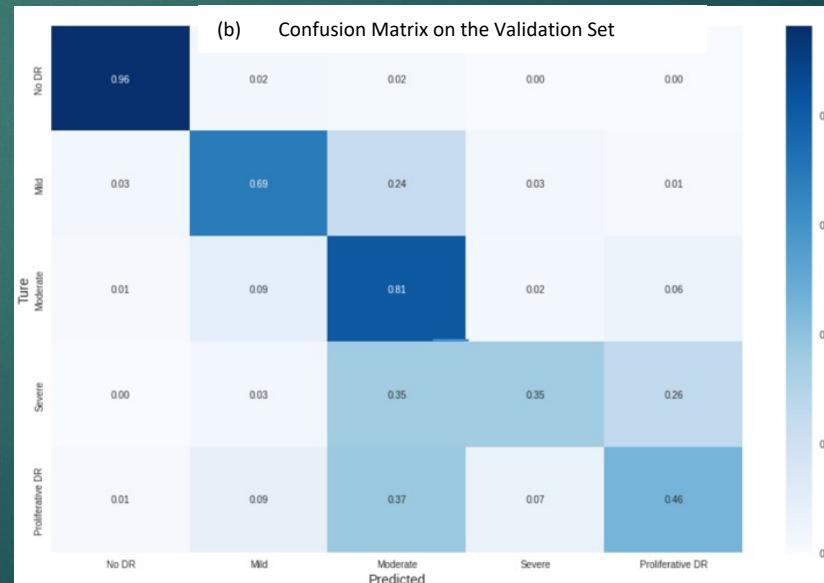
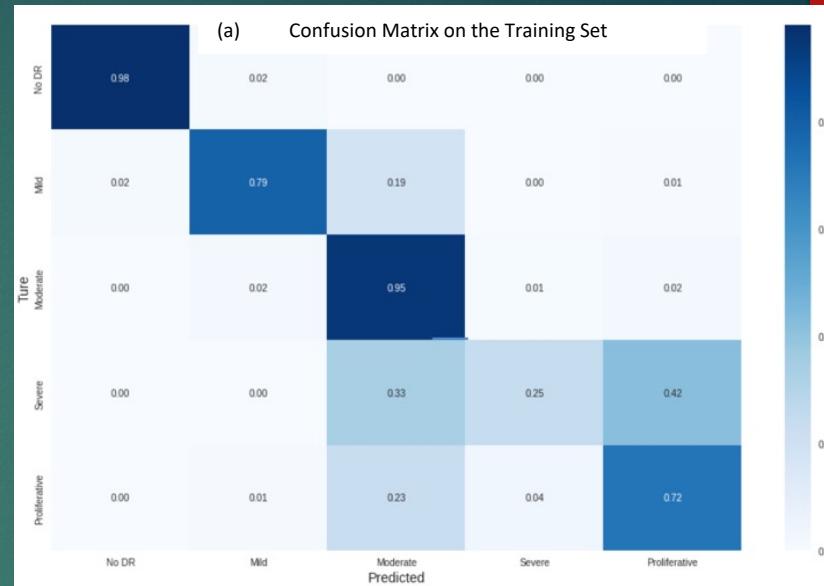
## 1) Candidate Model 1

### Confusion Matrices

- The patterns on the training and validation confusion matrices were very close
- Severe, Proliferative, and Mild DR were mostly misclassified as Moderate.
- Good at majorities

### Quadratic Weighted Kappa Scores

- On the training set: 0.938
- On the Validation set: 0.854
- Just remind that perfect prediction yield a score of 1. We have 0.938 and 0.854 respectively, which are very decent scores.



# MODEL EVALUATION

## 2) Candidate Model 2

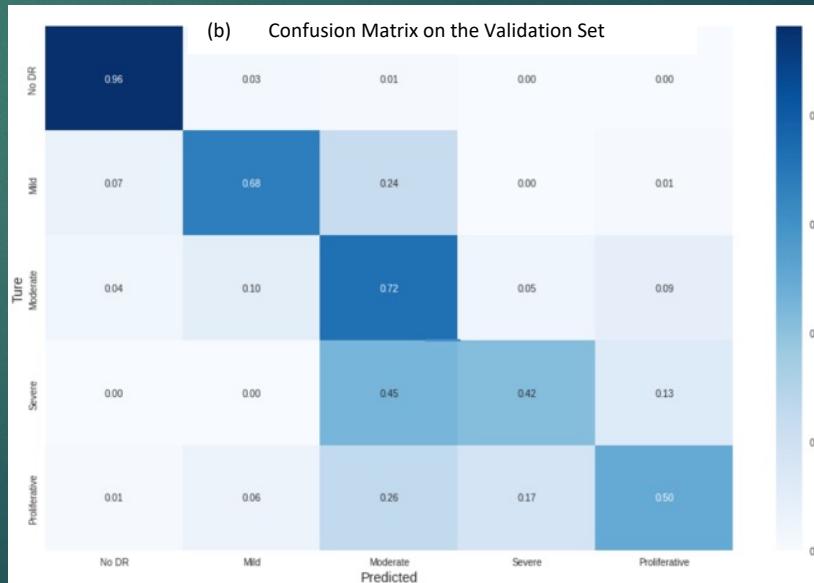
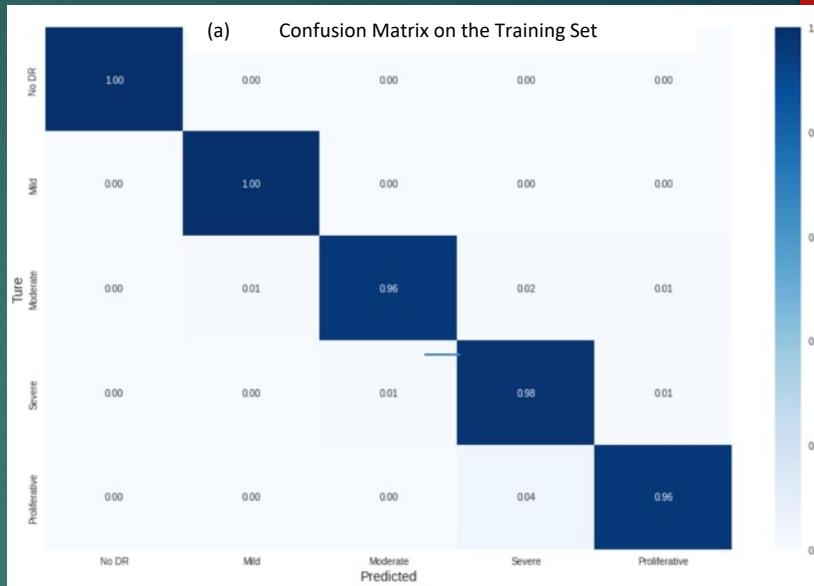
### Confusion Matrices

- ▶ Patterns are different on training and validation set
- ▶ Perfect on Training Data, but not on Validation Data
- ▶ Slightly better Performance on Severe and Proliferative DR

(Recalls: 42% & 50% vs 35% & 46%)

### Quadratic Weighted Kappa Scores

- ▶ On the training set: 0.992
- ▶ On the Validation set: 0.862
- ▶ Large gap between training and validation QWK scores. (overfitting)



# Final Model & Prediction on Test Set

## Final Model

### Candidate Model 1 (Averaged Cross-Entropy)

Even though the Candidate Model 2 has QWK scores slightly higher than the first candidate model, we choose the first candidate model as our final model because it provides more consistent performance on the unseen data.

## Prediction on Test Set

QWK Score: 0.766632

The results is slightly lower than 0.854 on the validation set but is still acceptable.

(Note: Result are generated on Kaggle Platform. The labels of test data is not provided.)

# Summary & Future of Work

- ▶ The performance of the model is good overall.
- ▶ It's most effective on identifying No DR, compared to other classes.
- ▶ The Model may be biased to Moderate DR, given other classes except No DR tend to be identified as Moderate DR.
- ▶ The good news is people with DR can be confirmed with high probability. The bad news is people in worse stage are optimistically diagnosed as moderate, which will delay their treatment.

In the future, we can focus more on the improvement of identification of Severe and Proliferative Classes.

- ▶ We may consider to utilize the techniques of image data augmentation, in purpose of expanding the sizes of these two classes.
- ▶ We can learn from the methods of improving the images quality to reduce the noises or enhance the characteristics of the different categories.
- ▶ Lastly, there are many outstanding algorithms for image classification, such as EfficientNet and DenseNet. It's always good to explore.

# Thank you!

