Deep Learning For Early Diabetic Retinopathy Detection

Anurag J. Vaidya[1], Joshua V. Stough[2]
Department of Biomedical Engineering, Bucknell University, Lewisburg, Pa [1]
Department of Computer Science Bucknell University, Lewisburg, Pa [2]

Introduction

- Diabetic retinopathy (DR) is the damage to retinal blood vessels that eventually leads to blindness if not treated.
- Of an estimated 285 million people with diabetes mellitus worldwide, approximately one third have signs of DR. [1]
- While robust algorithms exist to diagnose late stage DR, early detection is still unsolved problem because it is difficult to detect micro-aneurysms (Figure 1). [1]

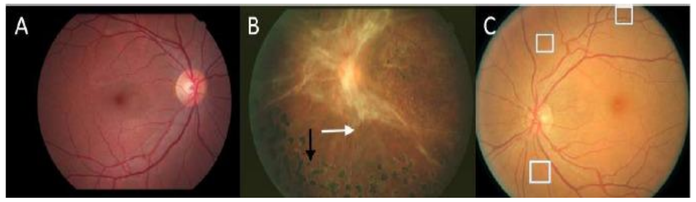

Figure 1: (A) Normal fundus photograph (B) Severe DR with white arrows pointing towards flame shaped hemorrhages (C) Early stage DR with white boxes showing micro-aneurysms

Project goal: Develop machine learning algorithms for early DR detection

Previous Work

- Previous literature models have been able to perform binary classification of diabetic retinopathy at accuracies upwards of 90%. [2]
- Literature methods have not shown impressive results in the graded classification of DR, especially in terms with early diagnosis. [3]
- The first neural network that tried to differentiate between normal patches of retina from patches with micro-aneurysms achieved an accuracy of 74%. [4, 5]

Dataset

- The APTOS 2019 blindness dataset was used to train all the models in this study. [6]
- A clinician has rated each image for the severity of diabetic retinopathy on a scale of 0 (no DR) to 4 (proliferative DR).
- The APTOS dataset was used to run three experiments:

1. **Binary classification**: labels changed to “0” if the image had no DR, or “1” otherwise.
2. **Early detection**: images with no DR were labelled “0,” “1” if original label was “1” or “2”, or “2” if original label was “3” or“4”.
3. **Severity classification**: original labels maintained.

Model Architecture and Hyperparameters

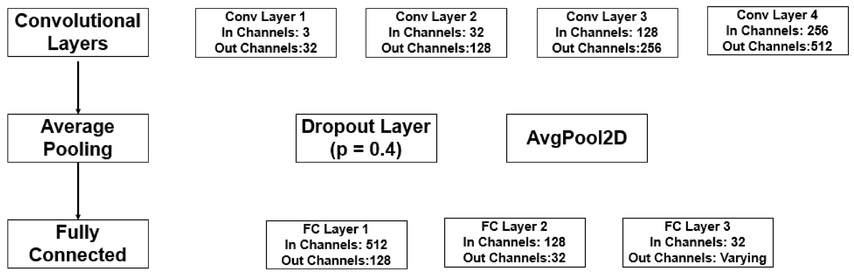

Figure 2: Convolutional neural network architecture. Last fully connected layer had variable out channels depending on experiment

- Batch size = 8
- Epochs: 25 (binary classifier), 40 (early detector and severity classifier)
- Learning rate: 0.0001
- Optimizer: Adam optimizer
- Kernel size: 3x3 for all convolutional layers

Training and validation protocol

- Five-fold stratified cross validation study performed for all experiments. Four splits for training and one split for validation.
- For binary classification, the true positive and false positive rates were calculated for each split. For the early detection and severity classifying experiments, confusion matrices were created for each split.

Results

- The binary classifier model gave very strong validation accuracies (average 97%) and strong ROC curves (Figure 3).

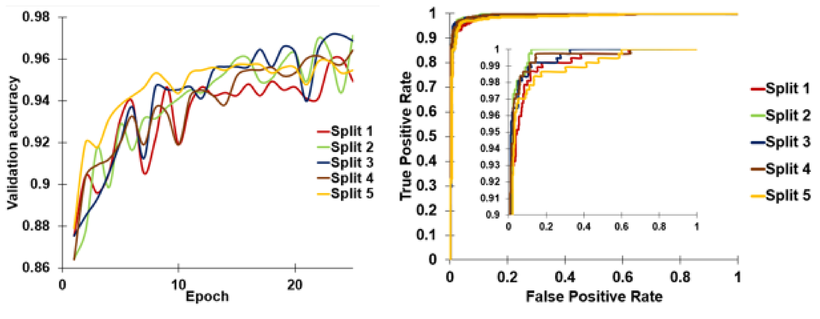

Figure 3: (A) Validation curves and (B) ROC Curves for the five splits from cross-validation study of binary classification experiment.

- The early detector and severity classifier gave validation accuracies surpassing literature results (Table 1 and 2). [4, 5]

Table 1: Validation accuracy for early detector

Split	Validation Accuracy
1	0.86
2	0.87
3	0.83
4	0.84
5	0.86
Average	0.852

Table 2: Validation accuracy for severity classifier

Split	Validation Accuracy
1	0.79
2	0.78
3	0.83
4	0.80
5	0.77
Average	0.794

Discussion

- The binary classifier’s sensitivity was calculated as 0.932 and the specificity as 0.989, which are comparable to literature models [2].
- The specificity being higher than sensitivity means that the model is better at predicting healthier cases than DR cases.
- Confusion matrices (Figure 4) show that model good at predicting early DR, but also confuses it with late DR.

A

True Label \ Model Prediction	0	1	2
0	350	8	2
1	12	243	17
2	0	53	44

B

True Label \ Model Prediction	0	1	2	3	4
0	354	2	5	0	0
1	7	40	24	2	0
2	3	10	183	3	1
3	0	0	26	11	1
4	0	8	31	3	17

Figure 4: Confusion matrices for **(A)** early detector **(B)** severity classifier

- This could happen because deeper in the model, the aperture is larger than the micro-aneurysms.

Future Work

- Create residual connections so that learning earlier in the model can be connected with deeper layers.
- Move to a 3-1-1 cross validation study.
- To avoid overfitting, stop training the model when training and validation loss diverge.

References

[1] Lee, R., et al. Eye and vision 2.1 (2015): 17.
[2] Lam, Carson, et al. AMIA 2018 (2018): 147.
[3] Gardner G., et al. British journal of Ophthal. 1996;80(11):940–944.
[4] Gargeya R, et al. Elsevier. 2017
[5] Gulshan V., et al. JAMA, 2016;316(22):2402–2410.
[6] https://www.kaggle.com/c/aptos2019-blindness-detection/data