Diabetic Retinopathy Detection

A P Vaideeswaran

IIT Hyderabad

20 November 2023

Domain/Area of the Project

Computer Vision in Healthcare

Motivation/Relevance of the Project

Diabetic retinopathy (DR) is the leading cause of blindness in the working-age population. This project aims to detect and classify DR into four stages if the disease exists. This might help identify the condition and estimate the severity even before visiting a doctor. This is especially important as the disease barely presents any symptoms in the early stages. Early medical attention can prevent loss of eyesight or slow down the disease. Doctors can also use this model to keep track of the patient's condition without even a visit. DR detection is time-consuming and requires a lot of expertise and infrastructure. Therefore, this project would help doctors and patients alike, giving the health sector time to improve the infrastructure required to cater to the needs of an increasing diabetic population, especially in rural areas.

Description of The Dataset

- The dataset includes labeled retina images with a subject ID and either a left or right field. For instance, `1_left.jpeg` is the left eye of patient ID 1.
- A clinician has rated the presence of the disease in each image on a scale of 0 to 4, according to the following scale: 0 - No DR 1 - Mild 2 - Moderate 3 - Severe 4 - Proliferative DR.

Description of The Dataset

- The images in the dataset come from different models and types of cameras, which can affect the visual appearance of left vs. right.
- Some images are shown as one would see the retina anatomically, while others are shown as one would see through a microscope condensing lens.
- An image can be told to be inverted using the following methods: It is inverted if the macula is slightly higher than the midline through the optic nerve or if there is no notch on the side of the image.
- The dataset can be downloaded from the following link: `https://www.kaggle.com/competitions/diabetic-retinopathy-detection/data`

Specific Objectives/Goals of the Project

- Given the image of the eye, the task is to classify it amongst five classes (stage 0 - NO disease to stage 4 - proliferative DR).
- The model should be resilient to noise, artifacts, lack of focus, and inappropriate exposure.

Challenges anticipated

- Effects of noise and artifacts, lack of focus, and inappropriate exposures.
- We have a large dataset comprising several large images (upwards of 5000x5000). The aspect ratios of the images are different as the images might have been taken using different instruments.
- Pre-processing of each image, possibly involving downsampling, must be done. However, this should be done so that the key indicators of the condition, like microaneurysms, are NOT lost.
- In this case, there is a notion of distance between classes. For instance, say the model incorrectly classifies stage 1 as stage 2; it is an incorrect classification. However, if stage 1 is classified as stage 4, it is also incorrect but, in some sense, MORE incorrect. This can be quantified using the quadratic weighted Kappa metric.
- Class imbalance is also expected as there are several images of the early stages of the disease compared to stage 4.

How do you expect to meet those challenges?

- While pre-processing, methods like histogram stretching and Gaussian blurring might improve image quality.
- FC-DenseNet can be used to identify the microaneurysms. The cross-entropy loss function of DenseNet can be replaced with the focal loss to weigh the aneurysms more than the background itself.
- A loss function that considers the distance between the classes has to be chosen.
- To deal with class imbalance (especially for stage 4), data augmentation, self-supervised learning, and focal loss can be employed.
- Multiple cores should be employed to deal with huge datasets efficiently. The following article was used for some initial insights on doing pre-processing: `https://www.nature.com/articles/s41598-021-04750-2`

Initial Approach to the Project

- Adaptive cropping with thresholding has been done on the images. In other words, if a column has more than 95 % (threshold) black pixels, that column can be removed. Each image was then resized down to 200x200.
- Amongst the state-of-the-art classification models like VGG16, Resnet-50, and EfficientNet 3 and 4, the best model for this project is to be identified.
- The ideation phase involves Transfer Learning. We maintain a penultimate hidden layer with 16 nodes for each model. The models shall then be trained on the 200x200 images for around 50 epochs. If the performance is satisfactory, we try 32 nodes and possibly even 64 nodes.
- The top two models can further be experimented with a second layer, and the performance be re-evaluated on the 200x200 images. This would help identify the golden bullet model for this project.

Initial Approach to the Project

- This model shall then be trained on 1000x1000 images. Using 200x200 images saves lots of time, as training would be significantly longer on the 1000x1000 images. Spending lots of time on a model without guaranteed performance would be wasteful.
- If the model performance is sufficiently good, more advanced methods can be used that effectively find the microaneurysms, thereby fine-tuning the performance.
- Hyper-parameter tuning can be performed if the performance is NOT sufficiently good. The model could be trained on a larger number of epochs, or the width and depth of the ANN could be increased.

Updates

The performance of all the models on the Test Images was evaluated before transfer learning. The absolute difference between the predicted and true labels was taken across the entire set and averaged out. The following results were obtained:

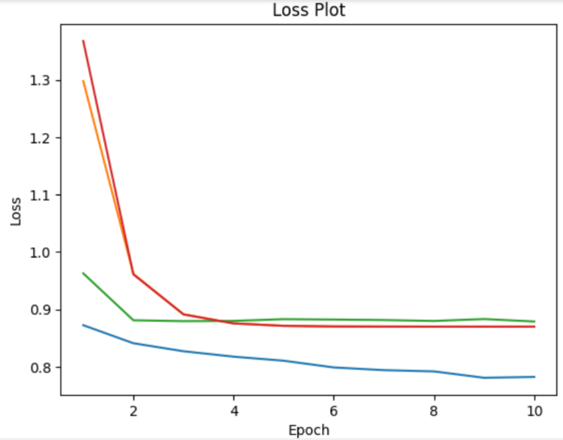
Models - Before Training			
VGG16	ResNet50	EfficientNetB3	EfficientNetB4
2.4172	1.7261	2.8894	2.4217

Comments

- The Transfer Learning was done keeping 16 nodes in the penultimate layer with the ReLU Activation Function.
- The output layer had 5 nodes corresponding to the 5 classes with the Soft Max Activation Function for 10 Epochs.
- The loss function employed was the Categorical Cross Entropy, with the Optimization being done using the Adam Optimizer. The chosen batch size was 32.
- More nodes in the penultimate layer weren't explored for VGG16, ResNet50, and EfficientNetB3 owing to their poorer test performance improvement relative to EfficientNetB4 (provided in the Results Slide).
- **Note:** The train accuracies aren't changing significantly for all the models. This could possibly be a limitation of Transfer Learning - only so much can be learned by training the last set of layers.
- **Note:** EfficientNetB4 might still have a good test performance despite having comparable accuracies to other models, as it might provide a "less" incorrect answer.

Loss Plot

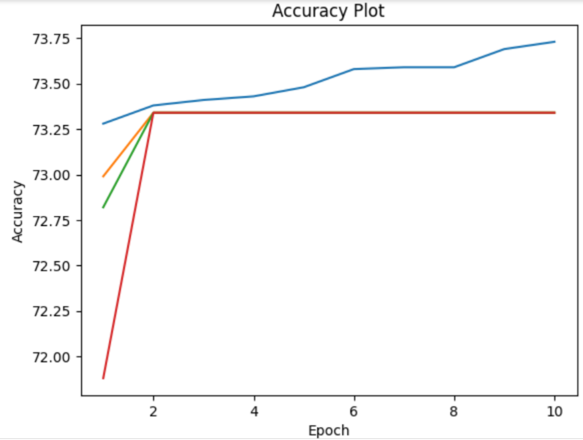
The number of epochs wasn't increased from 10 to 50 as the accuracy and loss weren't significantly improving for the models.



Blue: VGG16, Orange: ResNet50, Green: EfficientNetB3, Red: EfficientNetB4

Accuracy Plot

The number of epochs wasn't increased from 10 to 50 as the accuracy and loss weren't significantly improving for the models.



Blue: VGG16, Orange: ResNet50, Green: EfficientNetB3, Red: EfficientNetB4

Results

The performance of all the models on the Test Images was evaluated after transfer learning. The absolute difference between the predicted and the true labels was taken across the entire set and averaged out. The following results were obtained:

Models - After Transfer Learning			
VGG16	ResNet50	EfficientNetB3	EfficientNetB4
1.0606	3.4211	1.6133	0.5789

Hence, the following change in performance was observed - **EfficientNetB4** has improved the most - **Golden Bullet Model**:

VGG16	ResNet50	EfficientNetB3	EfficientNetB4
1.3567	-1.695	1.2761	1.8428

EfficientNetB4 - How to Improve Further?

- All the model layers must be trained to suit the exact purpose. Weights of **all the models mentioned above** were based on classifying the ImageNet Dataset.
- While we are still dealing with a classification problem, nuances specific to this task can only be learned by training on this specific dataset. This exercise might be more fruitful than simply adding more layers in the front.
- Since we are confident about the learning capacity of EfficienetNetB4, we can train the entire Network on 500x500 Images. These images would also provide the necessary resolution to classify the rarer and serious cases of the disease.

Thank You

I thank Dr. Soumya Jana, for allowing me to work on such an exciting and challenging project.

I would also like to thank Dr. Sumohana Channappayya and Mr. Ambarish Parthasarathy for providing me with GPU access, without which it would have been impossible to train the various models and evaluate their performances on time.

All the codes done so far can be found on the following GitHub Repository:
`https://github.com/Vaideeswaran21/PRML_Project`