ML Midsem Project Submission

Anonymous Author(s)
Affiliation
Address
`email`

Abstract

In this report we will explain our methods of creating the final classification labels for the dataset, how we trained it and the novelty that we plan on using in the future for better performance. Further we justify our choice of evaluation metrics and the results we obtained.

1 Introduction

Glaucoma is an eye disease where fluid builds up in the front part of the eye, causing an increase in pressure, which can harm the optic nerve. In this project, we are supposed to tackle 2 problems: a binary classification of referentiable and non-referentiable glaucoma, followed by a multiclass classification of referentiable glaucoma. The images of the Fundus are provided along with their true label, which was decided based on the decision of 2 normal graders and in case of disagreement by the third specialist grader.
Upon literature survey it was found that traditionally SVMs were the best classifier for Glaucoma until Deep Neural Networks were incorporated where specificity and sensitivity both crossed 90. Transfer learning has also been incorporated on models such as VGG16, VGG19, InceptionV3, ResNet50 and Xception to obtain specificity > 85 and sensitivity > 90.
The objective of this project is to build a model that not only successfully classifies Glaucoma but unlike other models gives reasons as to why the model thinks the test input has Glaucoma. Given the large dataset while traditional models will before very well, the true challenge will be to figure out feature extraction techniques that can successfully carry out the multiclass classification.
If made successfully, this model can help doctors in tracking the disease progression and even changing their treatment strategies on the fly

2 Materials and Methods

2.1 Dataset

In the original dataset given on the JustRAIGS website it was given that even if the graders G1 and G2 agree on the main binary classification, conflicts in labels have not been resolved (same goes for the situation where G1 and G2 do not agree with each other and one among G1 and G2 agrees with G3). Therefore, we have prepared a dataframe for our dataset which contains 'AND' values of the 10 labels provided by G1 and G2 in the cases where they agree and the 'AND' values of the 10 labels provided by G3 and the grader among G1 and G2 that agrees with G3 (in the case where G1 and G2 do not agree with each other).

2.2 Methodology

We started off by implementing CNN using tensorflow and then also prepared an alternative version of CNN using Pytorch. High level features are being extracted using CNN. Convolutional layers and maxpool layers have been used in the CNN before applying the flattening layers and the fully connected layers. Binary cross entropy loss function has been used since the first task at hand is binary classification. Adam has been used as the optimizer for the task. The data is split into train and val set in the 80:20 ratio and the CNN is then trained.

2.3 Novelty

We plan on trying to incorporate Meta Learning and creating a MAMS based model (Model-Agnostic Meta-Learning) based model by splitting the dataset to see if we can create a model which can perform better on unseen data. This will be helpful in the future as even on new data that is differnet from the data in the dataset, we are hoping to see better predictions

2.4 Evaluation Metric

Evaluation metric - Accuracy,specificity and sensitivity will be used as evaluation metrics for this task. Specificity and sensitivity are valuable evaluation metrics to be used for this task as the dataset has high class imbalance. Therefore, accuracy cannot be the sole metric that can explain how well the model is performing.

3 Results

So far we have implemented a CNN architecture for binary classification of the images into the two classes i.e "RG" and "NRG". The CNN architecture was trained on a mere 1000 images and as expected, the accuracy on the validation set was good enough but the recall score was poor. Data augmentation was then done in the training data to account for the high class imalance present in the data. Evaluation metrics that the CNN architecture has been tested on are accuracy and recall.

4 Discussion

5 Work to be done

We will be implementing more classifier models like SVM and Neural Networks. After classical ML models, we will move on to Deep Learning models and compare their performances on this dataset; given that this dataset is large, it is worth finding if DL models hold an advantage over ML models.

6 Distribution of work among group members

6.1 Report

Shobhit, Abhay, Chaitanya

6.2 Literature Survey

Shobhit, Abhay, Chaitanya

6.3 novelty

Shobhit, Abhay, Chaitanya

6.4 Training

6.4.1 CNN using pytorch

Abhay

6.4.2 CNN using tensorflow

Chaitanya, Shobhit