Diabetic Retinopathy

Advanced Machine Learning

Faculty of Sciences of the University of Lisbon

Teachers

Luís Correia

Helena Aidos

Students

Diogo Azevedo - 55773

Lucas Silva - 55922

Table of Contents

1 Introduction

The present work has as objective, the creation of a *Convolutional Neural Network*, in conjunction with image preprocessing, eliminating noise from the dataset and exacerbating pathology features, through the use of *Python* with the dependencies of the *tensorflow* and *keras*. In order to be able to initially distinguish between whether or not you have ocular pathology (Binary Classification), and then distinguish by the different types of existing pathology, the data is on a scale of 0-4, as we will describe later.

1.1 Problem

The diabetic retinopathy, is an ocular pathology that has the possibility to manifest itself in any patient with Diabetes Melitus, both in type 1 and type 2 and its occurrence tends to increase with age, and eventually cause blindness. The process of diagnosing this ocular pathology, requires a doctor to check the retina, and verify if there any abnormalities, like shown on the figure 1.

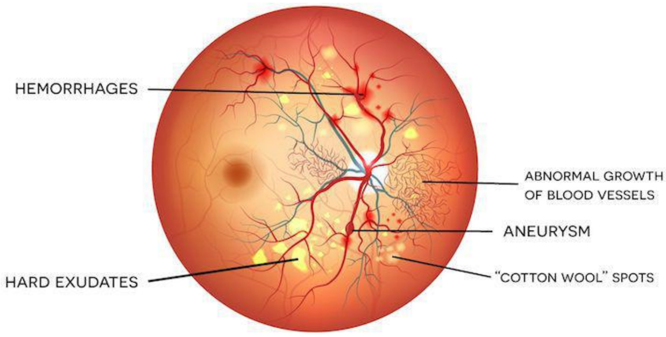

Figure 1: Diabetic Retinopathy Representation

As one can imagine, this process can take a long time to conduct, it usually takes about 2 to 3 weeks for the patient to receive feedback of the exam. And there is always a risk of a doctor misdiagnosing. Therefore, there is a need to expedite the time, it takes to conduct the process of the exam,in order to start treatment to prevent the probability of blindness, while having a way to backup the diagnosis provided to the patient.

1.2 Dataset

The Dataset consists of images of retinas (left and right) of the eyes of several individuals, as well as a csv file that indicates the level of the retinopathy, found in each of the images, in a scale of 0 to 4 (Healthy, Mild, Moderate, Severe, Proliferative respectively), where 0 corresponds to the absence of the pathology, and 4 as the most serious level of pathology. We also added a binary target for the binary classification task, with the values of 0 or 1, zero meaning there was no existing diabetic retinopathy, and one meaning there was.The images have different levels of quality, which translates to a certain level of noise in the dataset. An example of a retina of the dataset used on this project, is shown on the figure 2.

Figure 2: Retina from Dataset

While analyzing the distribution on the dataset, we observed that it was balanced for the binary classification task, this means that the number of retinas that were classified as having diabetic retinopathy, was pretty much the same as those classified as not having it. However, when analyzing the distribution for the categorical classification task, it was certainly not balanced, the Mild, Severe and Proliferative instances were significantly less than the Healthy and Moderate classes.

2 Approach

In order to solve the problem mentioned, we decided to divide the problem into a binary classification and a categorical classification. So initially we started by tackling the binary classification problem, where we distinguish between whether or not the retina provided has an ocular pathology, and after that we take on the categorical classification. But first, there was a need to clear noise that was present on the data, so we proceeded to resize and crop the images of the dataset. After that, there is a need to check the distribution of the classes for the binary and categorical classification tasks, and perform oversample in the case of any unbalanced class. Finally, we just had to create several data sets with different preprocessing parameters, to train and verify which one had the best results.

3 Implementation

As mentioned previously, there is noise on the images that needs to be cleared as much as possible, for that we created three datasets with different preprocessing parameters. Such as, the cropped raw images, where we only crop the image on the circle, to try and avoid most of the black present around the retina. One with cropped gray scaled images, and one with cropped RGB images, both with contrast enhanced by the CV2 Gaussian blur, where on the first the image colors are only shades of gray, and the second the image colors, where the Gaussian blur is typically to reduce image noise and reduce detail. In order to, try and enhance the features of the images, as shown on the figure 3.

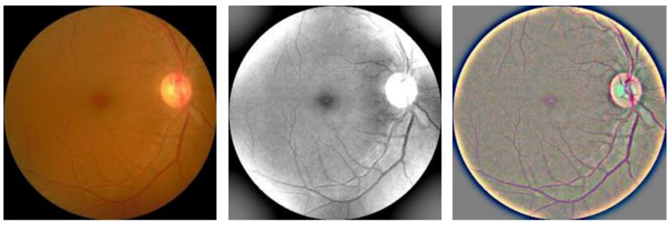

Figure 3: Cropped Raw (Left), Cropped Gray Scale (Middle), Cropped RGB (Right)

In some cases, there can be some loss of information associated with cropping the images, by removing some of the edges of the retina of the image. Nevertheless, in the majority of the cases the advantages far outnumber the disadvantages, so by doing this our models will be more efficient when learning the existing patterns.

As mentioned previously, the level of diabetic retinopathy class was unbalanced, and for that we decided to try to balance as much as possible the class, for the purpose of training the models. When models are trained with unbalanced dataset they do not tend to generalize the predicts made, and only predict as the class that has significantly more instances. Therefore, we proceeded to oversample the levels that has less instances as mentioned before, so we decided to have 800 instances of each one of the classes except for the Healthy class, as one can see on the figure 4.

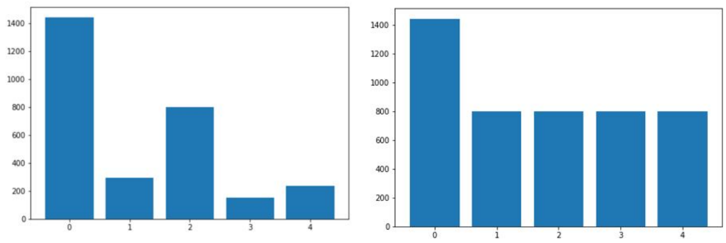

Figure 4: Before Oversampling (Left), After Oversampling (Right)

After the preprocessing previously described, we end up with three different datasets to train the models. Finally, we create several CNN's in order to train and test the data available, and evaluate whether or not the model obtained has performed well. We tested several different models, with up to five convolutional layers, we started by adding a first layer with 32 filters, and up to the five layers with 512 filters, while using max pooling to reduce the dimensionality and allowing for assumptions to be made about features contained in the sub-regions binned. Finally, using flatten on the output, before the final Dense layer. We observed that after the third layer there was no improvement on the performance, with the addition of convolutional layers and for that reason we settled on a three layer model.

4 Results

For the several different models created, firstly on the binary classification models the results obtained were pretty favorable. And for the three datasets, we were able to distinguish between whether or not the retinas had the diabetic retinopathy, with models that were fairly simple, like mentioned previously. As it is shown on the table 1.

Models	Recall (%)	Accuracy (%)
Raw	85.3	85.1
Gray Scale	96.5	96.9
Color Enhanced	98.3	98.6

Table 1: Binary Classification Models

Analyzing the results obtained for the binary classification task, we can observe that the feature enhancement done to the images of the original dataset, can greatly improve the performance of the models as it is proven on the table 1. For the categorical classification models, the results were not as positive as the binary classification, in these the standard model was not able to achieve good results, when we tried more complex models with more layers and data augmentation, achieved by rotating the images on the dataset, by that we did not need to collect new data and simply utilise the original data. However the models obtained were not able to identify the Mild, Severe and Proliferative cases, and the model achieved accuracy around 75%. As it is seen on the figure 5.

Figure 5: Categorical Classification Matrix

We should note, that the classes with less instances were the hardest ones to identify. Lastly, we proceeded to look into the activation maps, to identify what the models are considering on the images to classify the level of retinopathy. And with the observations, we were able to identify that the models can not identify the hard exudates and hemorrhage, on the retinas, as it is shown on the figure 6.

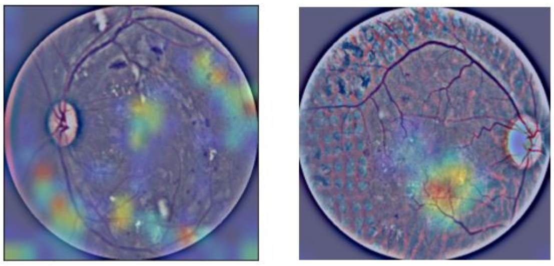

Figure 6: Activation Maps, Severe (Left), Proliferate (Right)

A way to solve this, would probably be by having even more images to train the models, by having 5 different classes, we end up with a small number of instances for each one. These models need a large amount of information to be able to have a high performance.

5 Conclusion

During the realization of this project, we are able to conclude the difficulties associated with these type of problem, where the quantity and quality of the data provided to the models, can greatly impact the outcome, and whether or not the models created are good at predicting the target. However, the data available is balanced and does not have much noise, for that, there was a need to divide the whole problem into smaller tasks. The image cropping and feature enhancement utilized, were able to greatly improve the performance of the models. Because of that we were able to create simpler models that were able to distinguish with high performance the binary classification task. Despite the efforts made, for the categorical classification task, the quantity of images provided to the models was not nearly enough, to obtain a model with high performance able of predicting correctly the level of diabetic retinopathy.
With this, we can identify the need and potential of Convolutional Neural Networks, which these have to be able to solve difficult tasks, that had once to be done by people and take a long time, as in the case of diagnosing diabetic retinopathy.

References

[1] Asia Pacific Tele-Ophthalmology Society (APTOS). *APTOS 2019 Blindness Detection*. 2019. URL: `https://www.kaggle.com/c/aptos2019-blindness-detection/data?select=train.csv`.

[2] M. Antony and S. Brggemann. *Kaggle Diabetic Retinopathy Detection*. 2015. URL: `https://www.kaggle.com/c/diabetic-retinopathy-detection/discussion/15617`.

[3] Zhiguang Wang and Jianbo Yang. "Diabetic Retinopathy Detection via Deep Convolutional Networks for Discriminative Localization and Visual Explanation". In: (Mar. 2017).