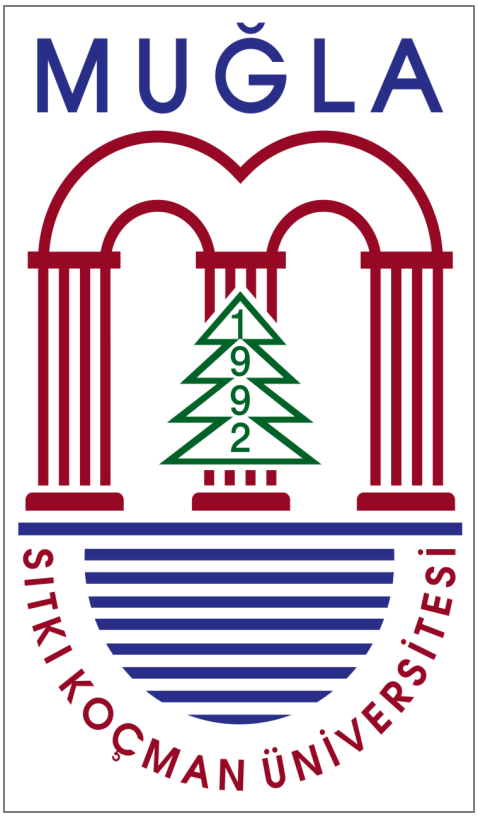
Designing a Decision Support Web Interface for Glaucoma Diagnosis



Computer Engineering
Senior Project Report

Advisor : Tuğba Önal-Süzek, PhD

Group Members : Kıymet Deren TOY,
Görkem SAVRAN

Date : 12/06/2022

Designing a Decision Support Web Interface for Glaucoma Diagnosis

Kıymet Deren TOY, Görkem SAVRAN
Computer Engineer
Muğla Sıtkı Koçman University
12/06/2022

Summary

Glaucoma is defined as a progressive optic neuropathy characterized by retinal ganglion cell (RGH) loss. [1] With this disease, optic nerve fibers are gradually lost. If it is not recognized and treated early, it causes permanent vision loss and is the second leading cause of blindness in the world, after cataract. According to the data of the World Glaucoma Association, it is predicted that there will be 76 million glaucoma patients worldwide in 2020.

It is seen that 50% of people with glaucoma in developing countries and 90% in underdeveloped countries are not aware of their disease, especially since the disease does not cause complaints in the early stages of the disease. Early diagnosis is very important in glaucoma, as the disease does not cause any symptoms and vision loss cannot be reversed. [2] However, a single method is not sufficient for the correct diagnosis and early diagnosis of the disease. Today, many researches are carried out to eliminate this deficiency, various articles are published, and it is desired to help doctors by training artificial intelligence models.

Although there are artificial intelligence models produced for glaucoma testing, the fact that none of them has been turned into a website and made available to doctors makes our project unique. This project aimed to minimize the problems faced by ophthalmologists during the decision-making process and to provide them with a second opinion. The product we will reveal will be a website that every ophthalmologist can easily use and get support while diagnosing glaucoma.

Thanks to this website, ophthalmologists will be able to easily perform glaucoma tests by uploading retinal photos of their patients to the system. A trained artificial intelligence model will be used in the background of the website with photographs of patients previously diagnosed with glaucoma. Thus, a prediction/result will be made by comparing the newly uploaded retina photo with previous patient photos.

Retinal photographs to be used in the testing phase can be obtained through OCT devices, fundus cameras, and digital ophthalmoscopes. This will ensure that the site can be used easily in every health institution.

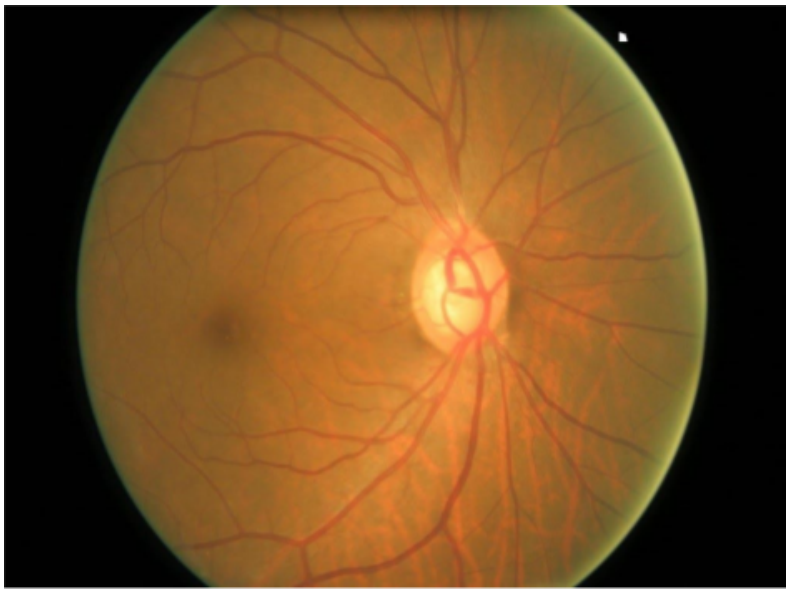


Figure 1: *Retinal of a patient diagnosed with glaucoma*

Table of Contents

1. Introduction

Glaucoma is defined as optic nerve damage resulting from increased intraocular pressure and is the second leading cause of blindness in the world, after cataract. The main forms of glaucoma are open-angle, angle-closure and congenital glaucoma. Open-angle glaucoma is also called primary or chronic glaucoma, as it is the most common type of glaucoma (at least 90% of all glaucoma cases). In this type of glaucoma, many people experience permanent vision loss due to the fact that the symptoms do not manifest themselves much and the disease is detected late.

(a) (b)

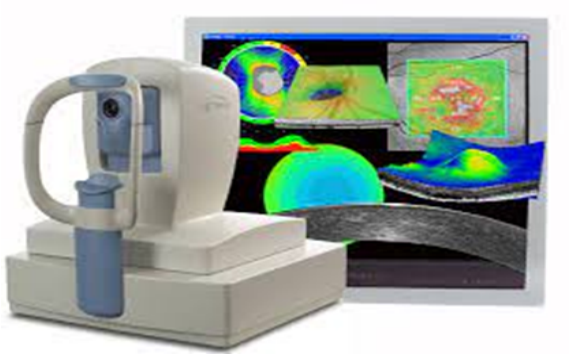
People with glaucoma gradually lose their peripheral (side) vision and may miss objects to the side. In the example, it is possible to see the differences between individuals with normal vision and individuals with glaucoma. Individuals with glaucoma are as if they are looking through a tunnel.

The greatest risk of the disease is that there is no return once this vision loss has occurred. For this reason, early diagnosis/diagnosis is very important. A single method is not sufficient for the correct diagnosis and early diagnosis of glaucoma disease. Therefore, more than one technique is used together in the diagnosis of glaucoma today.

The most commonly used of these methods are;

- measurement of intraocular pressure (IOP measurement),
- measurement of central corneal thickness (Pachymetry) ,
- evaluation of anterior chamber angle patency (Gonioscopy),
- visual field examination,
- ONH (optic nerve head) and RNFL (retinal nerve fiber layer) evaluation

If we consider each one in turn, in the first stage, the patient's intraocular pressure (eye pressure) is measured by the physician with a device called a tonometer. Eye pressure up to 21 mmHg is considered normal, and values above this are considered high eye pressure. However, high eye pressure alone is not enough for the diagnosis of glaucoma, because glaucoma can also be seen in people with normal eye pressure and sensitive optic nerve. Likewise, while pachymetry helps to understand whether the abnormality in eye pressure is due to the thickness of the cornea, gonioscopy allows to examine the anterior chamber angle, where the reticulated structure, which is the area where the eye water empties into the capillaries, is located. However, these methods cannot be used to directly diagnose early glaucoma.



In visual field examination, the patient's visual field is determined by revealing retinal sensitivity based on the patient's expression. However, since this method is based on patient expression and there is no reduction in the visual field in the early stages of the disease, it is not suitable to be used for early diagnosis.

Today, OCT devices are used to detect structural losses before functional losses occur. With OCT, it is possible to diagnose glaucoma before visual field damage develops. However, abnormal tests do not necessarily indicate that the patient has glaucoma. In some eyes, the optic nerve head may differ from the normative data and be perceived as damaged by this device. In these cases, visual field and clinical examination results should be interpreted together to avoid false positive results. Here, physicians have a great responsibility in the decision-making process and it is very important to have experience while making this decision.

In fact, with our project, in cases where doctors need a second opinion; we aim to provide them with a decision-support mechanism (a second opinion) and to create a website that every doctor can easily use on equal terms. On this website, we will test the retina photo uploaded by the clinicians to the system in the artificial intelligence model that we have trained, and we will show on the page what percentage of glaucoma patients have glaucoma.

2. Methods

i. Data Collecting

The retina photos to be used for training the artificial intelligence model were obtained from ready-made data on the internet. Some of the datasets we used are:

1. ORIGA-light fundus photographs dataset [3]

2. Drishti fundus photographs dataset [4]

3. ACRIMADataset [5]

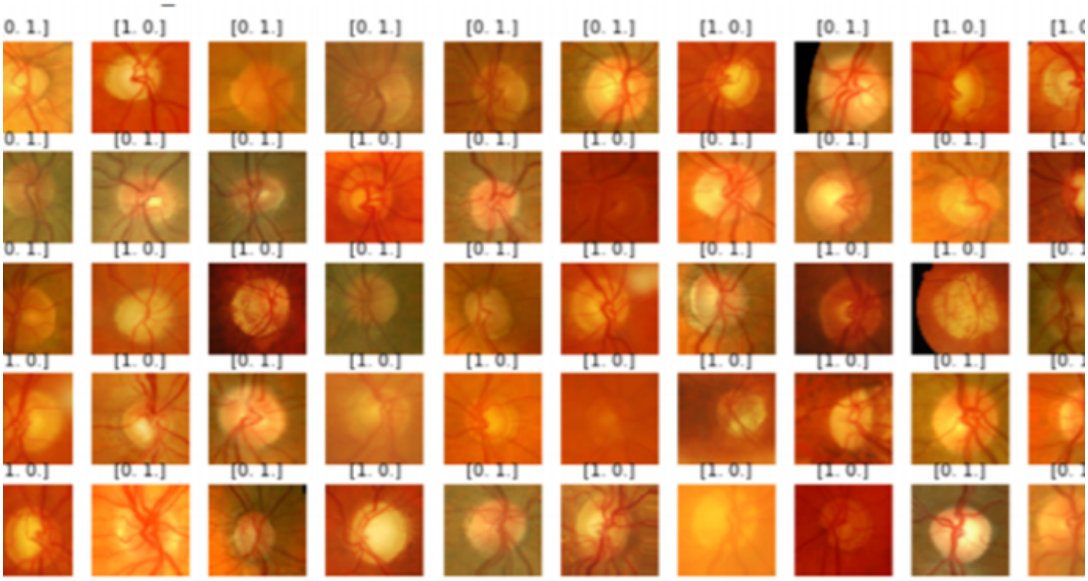


Figure 2: *Example of ACRIMA dataset*

During the project process, it was aimed to increase the learning accuracy by compiling additional data sets from the internet. Care has been taken to ensure that the positive and negative sub-training sets of our data are balanced by running the necessary sampling algorithms to ensure that the positive and negative sets of the compiled retinal images are balanced and to prevent oversampling or undersampling.

Some of the training data were used as test data, since real patient data could not be used. In order to do this, real data parts were removed from the data set and 3-fold cross-validation was applied on the trained data.

ii. Training Artificial Intelligence Model Using Open Source Machine Learning Libraries

We trained a sequential model using python's open source machine learning libraries, keras and tensorflow. The output layer of our model consists of 2 layers and uses the softmax activation function. While compiling our model, we used 'binary_crossentropy' as the loss function and 'accuracy' as the metric. We used the ACRIMA dataset to train our model, which had a total of 705 photos. We used 3-fold cross validation to increase the accuracy of the model and avoid overfitting problems.

Fold 1

```
***Performance on Validation data***
Accuracy  : 0.9197860962566845
Precision : 0.9215225275063004
f1Score : 0.919994474127572
[[95 10]
 [ 5 77]]
```

Fold 2

```
***Performance on Validation data***
Accuracy  : 0.9572192513368984
Precision : 0.9573276232543714
f1Score : 0.9571621109641196
[[101   3]
 [  5  78]]
```

Fold 3

```
***Performance on Validation data***
Accuracy  : 0.9786096256684492
Precision : 0.9795930911549573
f1Score : 0.9786515675789031
[[100   4]
 [  0  83]]
```

iii. Testing the Artificial Intelligence Model

We tested our model with data that we did not use in training, and we achieved an average success of 0.95 as a result of 3 folds. We then tested our model with 134 photos, independent of the dataset. The confusion matrix of the sklearn library was used to evaluate the predictions and then the F1 score was calculated. (Accuracy can be examined from the confusion matrix in the Results section.)

iv. Designing the Web Interface

One of the biggest problems we encountered in the project was KVKK. According to this law, it was a very difficult process to process patient photos on our servers. That's why we decided to process patient photos in users' (ie clinicians) browsers instead of processing them on servers. We made a static website using Angular and Tensorflow.js and embedded our model inside this website. Thus, results can be obtained without patient photos being sent to our servers, that is, without violating the KVKK.

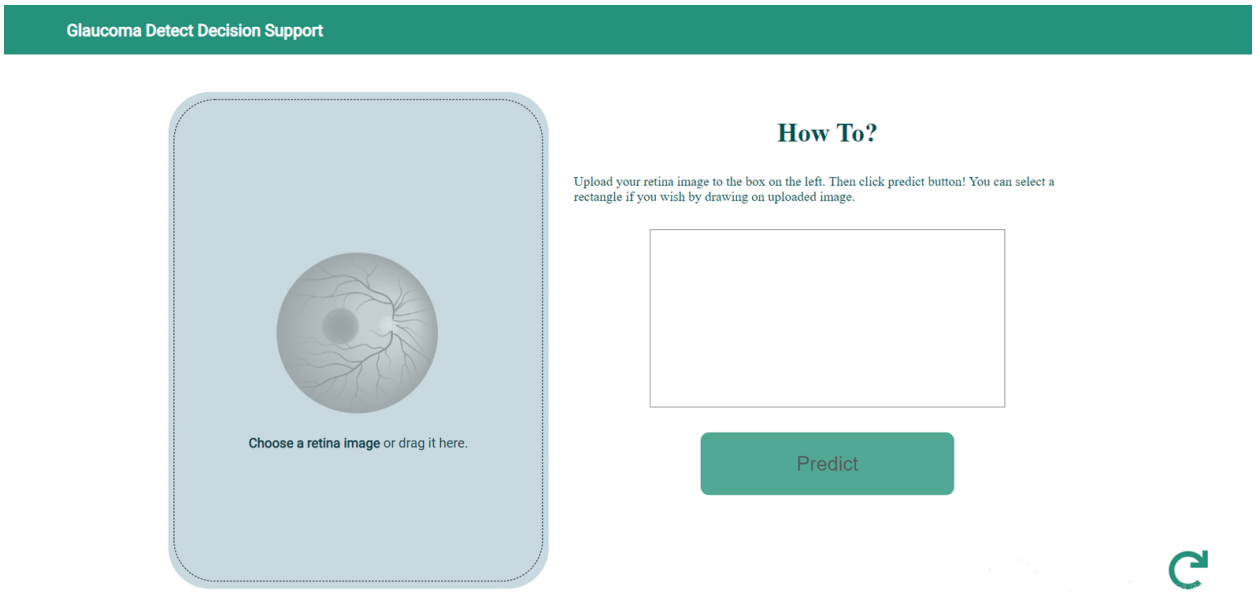


Figure 3: *Web interface of our project*

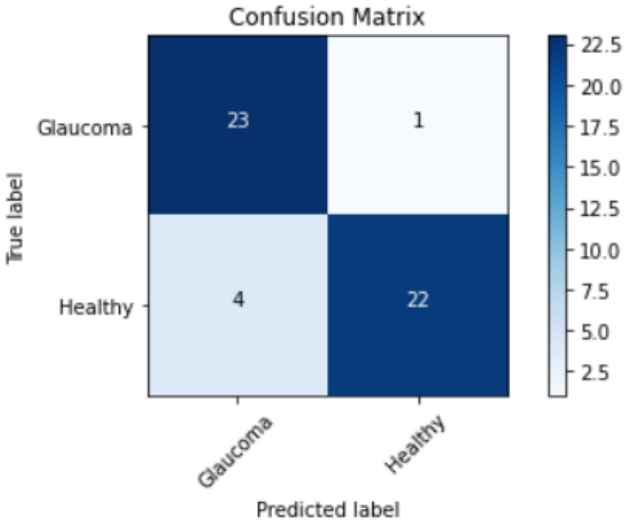
v. Publishing the Web Interface on the Internet

We can publish our website easily and also for free. Because our website is a single-page application, most of the big companies out there give us free hosting. We can host it into Google Firebase, Github Pages or one of the other cloud platforms. Hosting a single-page application is as easy as uploading a file.

3. Results

As a result of testing the data set, the following confusion matrix was obtained. And based on this matrix, the F1 score was calculated and found to be 0.95.

```
==============TEST RESULTS============
Found 139 images belonging to 2 classes.
1/1 [==============================] - 2s
Accuracy  : 0.9568345323741008
Precision : 0.9570662114376295
f1Score : 0.9567480193060741
```



So far, we have been entitled to receive TUBITAK 2209A university students research funds with this project. At the same time, we were selected as one of the 27 finalists in the field of health by participating in TUBITAK's 2242 university students research competitions.

In addition to these, we applied to the Bigg program with our project and passed the 2nd stage. If we can get a grant, we will train an artificial intelligence model with approximately 87 thousand retina photos and try to increase the accuracy to 98%.

4. Conclusion

In order to decide which model should be the most accurate in data science projects, we need to evaluate the demands coming from the business units. If we only make model selection based on accuracy in our project outputs, this will mislead us.

Accuracy is a metric that is widely used to measure the success of a model but does not appear to be sufficient on its own. F1 Score value shows us the harmonic mean of precision and sensitivity values.[6]

The main reason for using F1 Score value instead of Accuracy is not to make an incorrect model selection in unevenly distributed data sets. In addition, F1 Score is very important to us as we need a measurement metric that will include not only False Negative or False Positive but also all error costs.

Here, as can be seen from the findings section, we have achieved a very good F1 score of 0.95 and our accuracy metric is 94.3%. In short, these show that we have an accurate model. The only thing is that the total number of photos in the data set we have is 705. Therefore, more data should be labeled and our model should be retrained with these data. Only in this way can we make a diagnosis close to what a real doctor can.

5. References

[1] Glokopedi, https://www.todnet.org/glokom/oct-deger.htm , Erişim Tarihi:04/09/2020

[2] TOBB ETÜ HASTANESİ TIP FAKÜLTESİ Glokom: Sinsi Görme Hırsızı,

https://hastane.etu.edu.tr/bulletin/3522-glokom , Erişim Tarihi: 04/09/2020

[3] Drishti veri seti, https://cvit.iiit.ac.in/projects/mip/drishti-gs/mip-dataset2/enter.php, Erişim Tarihi: 21/12/2020

[4] ORIGA-light veri seti, https://drive.google.com/drive/folders/1VPCvVsPgrfPNIl932xgU3XC_WFLUsXJR , Erişim Tarihi: 21/12/2020

[5] ACRIMA veri seti, https://figshare.com/articles/dataset/Retinal_Image_Synthesis_and_Semi-supervised_Learning_for_Glaucoma_Assessment/7339511 , Erişim Tarihi: 24/06/2021

[6] F1 score, https://medium.com/@gulcanogundur/do%C4%9Fruluk-accuracy-kesinlik-precision-duyarl%C4%B1l%C4%B1k-recall-ya-da-f1-score-300c925feb38 , Erişim Tarihi: 17/01/2022