

# Interview

In this stakeholder interview I discussed the different datasets with the stakeholder and their impact if successful. There were some general tips on the logistical working of the eye healthcare as well.

## General tips:

De severity of diabetic retinopathy is usually a reflection of the blood veins in the rest of the body for a diabetes patient. When moderate or worse is present usually other eye related issues can appear such as fluid in the eye that has to be drained by an eye specialist doctor.

Even if the data is not present in the dataset the application should be able to notice if there is something wrong with the eye that is different from a normal eye. An optician/human would be able to notice this as well and would be a requirement if we want to automate the diagnosing process. If this isn't possible a regular check is needed by a specialized eye doctor (every 2/4 years based on doctor's discretion)

## Considered Datasets

### Dataset 1

Dataset of eyes with and without diabetic retinopathy:

<https://www.kaggle.com/donkeys/retinopathy-train-2015?select=trainLabels.csv>

There are 3 places where screenings are done for eyes:

1. Hospital
2. clinic (private hospital)
3. optician shop
  - Create a picture of the retina
  - Check the patients ability to see

If the severity is moderate or worse the picture should always be sent to a hospital or clinic for a specialist to look at.

Business case: Pictures could be taken by unspecialised personnel and judged by an algorithm. This will save time and money for both the patients en doctors in the eye care and diabetic care.

### Dataset 2

Smaller dataset with multiple eye disease:

<https://www.kaggle.com/andrewmvd/ocular-disease-recognition-odir5k>

Most interesting data set to work with because it could have a large impact on the quality of the countries healthcare. Think of referring patients to the right specialist/doctor with no time in between.

Low data per category.

Not all images are correctly judged(stakeholder would've diagnosed at least 1 eye differently).

### **Dataset 3**

Dataset with images that are taken with a different more expensive machine that is not normally found outside of hospitals, but shows more interesting data:

<https://www.kaggle.com/paultimothymooney/kermany2018>

Important with these images is that they are used as progression photo's of 1 patient for a specific doctor.

A doctor needs to precisely investigate what kind of medication is applicable. This is outside of the projects scope and possibly too difficult for current technology.

Additionally the risk and damage if something goes wrong is very high.

### **Dataset 4**

Dataset of images with and without Glaucoma. This dataset has images that are very similar and might be more consistent for a ML algorithm to categorize:

<https://www.kaggle.com/sshikamaru/glaucoma-detection>

Dataset too small (less than 1k images) and normally a different technique is used to detect this illness.

Patient would often not notice this disease and would benefit from this technology since it could be found early.

Choice

We decided that dataset 1 has the highest value and feasibility combination of the data sets.