

# Retina Image Analysis

Students : Alexandre Jann, Maylis Touya, Paola Zanchi

Teaching Assistant : Michael Beyeler

## Introduction

With an estimated number of 17.9 million death per year, cardiovascular diseases are the first cause of death (WHO). More people die annually from these diseases than from any other cause. Cardiovascular diseases are a group of heart's and blood vessels' disorder. It includes different types of disorders: strokes, heart attack, coronary heart disease, cerebrovascular disease, thromboembolic disease, rheumatic heart disease, cardiomyopathy, and other conditions. Out of all cardiovascular diseases death, 85% are due to heart attacks and strokes. Many factors must be taken into account for the development of cardiovascular disease. High blood pressure is a very predominant factor that accounts for about 13% of deaths. Tobacco and diabetes have also an impact as well as lack of exercise, obesity, and poor diet. Prevention of cardiovascular disease and identification at early stages can prevent premature deaths. These diseases usually take place in low- and middle-income countries (75% of cardiovascular disease death). This is mainly because those countries often do not have the benefit of integrated primary health care programs for early detection and treatment of people with risk factors as there is for the people in high-income countries.

The eye fundus is the interior surface of the eye opposite the lens. It is supplied by two distinct vascular systems: arteries and veins. With fundus photography, a special fundus camera points through the pupil to the back of the eye and takes pictures. Colour images provide documentation of the ocular fundus. The resulting images can be spectacular and help the doctor to find, watch and treat disease. An

eye fundus is a less invasive physical exam, which allows us to see the blood vessels' conditions in a very short amount of time. It seems that it can be a very good exam and can maybe replace the actuals vascular exams that are invasive and take a long time to be made.

This project is quite interesting because it allows us to have a mathematical insight into the determination of diseases and comorbidity risks. Moreover, since an eye fundus is a less invasive physical exam, allowing us to see well the conditions of the blood vessels in a very short amount of time, it seems that it can be a very good exam, and maybe can replace the actuals vascular exams that are invasive and take a long time to be made. As said before, a very high number of cardiovascular diseases can be avoided thanks to prevention and early detection. With this project, detection could be easier for the patients and the doctor and show efficient results.

In this project, we wanted to know if we can predict if someone had a cardiovascular disease by looking at its eye fundus blood vessels' tortuosity and diameter plus other features?

## Material and Method

### Material

#### A/Eye Fundus Snapshot

Eye fundus snapshots are easy to take. Eye fundus examination is used for screening for vision problems and via the health status of the retina, macula and blood vessels. This can be done in people of any age. To carry out this examination, it is necessary to use ophthalmic drops that permit dilating the pupils and increasing the angle of observation and therefore allow better visualization of the eye

fundus with the different structures present: retina, retinal vasculature, optic disc, macula, and posterior pole. Images are produced using a low-power microscope attached to a camera. Let's note that a fundus camera or retinal camera is a specialized low power microscope with an attached camera designed to photograph the interior surface of the eye, including the retina, retinal vasculature, optic disc, macula, and posterior pole (i.e. the fundus). Your eyes will be dilated before the procedure. Widening (dilating) a patient's pupil increases the angle of observation. This allows the technicians to image a much greater area and have a clearer view of the back of the eye. Based on the reflected light effect, it is then possible to obtain an image<sup>1</sup>. In colour fundus photography, the image intensities represent the amount of reflected red (R), green (G), and blue (B) wavebands, as determined by the spectral sensitivity of the sensor. The capture lasts only a few minutes per eye, is not painful and non-invasive. The use of eye drops (eye drops) does not show any health impairment. The only side effects that have been perceived and are that sometimes it is possible to see phenomena of ocular dryness, foreign body sensation, and watery eyes.<sup>2</sup>

#### B/UK BioBank

The dataset that we have come from the UK Biobank. This is a large-scale biomedical database and a very important resource for research. The recruitment has begun in 2006 and new data are regularly added to the database.

The dataset is made of a colossal amount of biological and medical information from about 500 '000 people which made this dataset the largest and richest of its kind. There is no other biobank as detailed and that provide a long-term perspective on health in the world. Across the world, researchers and scientist are allowed to access the database to improve public health and contribute to the discovery of new medicine and treatment. Across the world, academic, commercial or charitable organizations are encouraged to use this

Biobank. More than 90 countries use it. All the participants live in the UK and are between 40 and 69 years old. All the data are anonymized. The information is vast and includes as much medical data as blood, urine or saliva samples but also data on genetics or the lifestyle of individuals. It contains a massive amount of imaging data as well giving an imaginative approach.

Thanks to this biobank, studies can contribute to a better understanding of life-threatening illnesses such as cancer, heart disease and stroke. This can lead to improving human health and different prevention, diagnosis and treatment. The purpose of this project is to reach a better characterization of the diseases that develop in some people but not in others (why, how) to prevent and treat them.

#### C/ Software and Server

For this project, we first wanted to use Matlab but it actually didn't work very well with us, so then we decided to use python, which we know much better. We used python mainly for the analysis of the images and the calculation of the tortuosity and then used R to do the analysis. Python is an interpreted programming language, and very easy to use. It favours structured, functional and object-oriented imperative programming. It is a language that can be likened to math: you have to be structured, so everything flows naturally. It has strong dynamic typing and automatic memory management. On the other hand, R is the language we are the most comfortable with.

We had access to a gigantic server that belongs to the research group. Thanks to its good puissance, we did our calculation, and all the data could stay on it without causing privacy issues for the patients.

#### Method :

##### A/ARIA

It is a software originally used for measuring the tortuosity of a plant's roots. In fact, ARIA stands for Automatic Root Image Analysis<sup>3</sup> This software allows large phenotyping

experiments and can help to establish relationships between two different variables. In our case, it has been used to measure the tortuosity of blood vessels in fundus images.

### B/Tortuosity Measurement

Tortuosity is the property of a curve being tortuous and twisted. The concept of tortuosity is vague with multiple definitions and various evaluation methods introduced in different contexts. It can describe different mechanism depending on the subject of the study (electric, hydraulic, thermal...). These tortuosities are defined differently, and their values can differ.

#### 1. First Method

We used the distance factor (called DF) to calculate the tortuosity of the blood vessels. It consists of the ratio between the length of a line and the length between its first and last point of measurement. The formula is pretty simple and easy to manipulate.

$$DF = \frac{L_t(seg)}{C(seg)}$$

The formula for the distance factor (DF) with the total length of the path of the segment (numerator) divided by the length of the segment between its first and last point (denominator).

#### 2. Second Method

Another method of calculating tortuosity is based on the fact that the vessels are described by points, and that for each point, we can create a circle that contains two other points of the segment. Doing this, we get a lot of measurements with the centres of the circles and their radius. It is thanks to the radius that we will be able to define the tortuosity of the segments: for straight segments, as the circle is very large, the radius is very large as well, and for very tortuous segments, the circle will be very small with small radius. We divide 1 by the sum of all these radiuses and get a result that we then divide by the length of the segment to have a normalization. The result gives the tortuosity score. The higher the score, the more tortuosities the vessel. The lower the

score, the straighter the segment. We didn't really know which of these two methods is the most precise. We used the first one with the distance factor that may be more visual for our research and analysis.

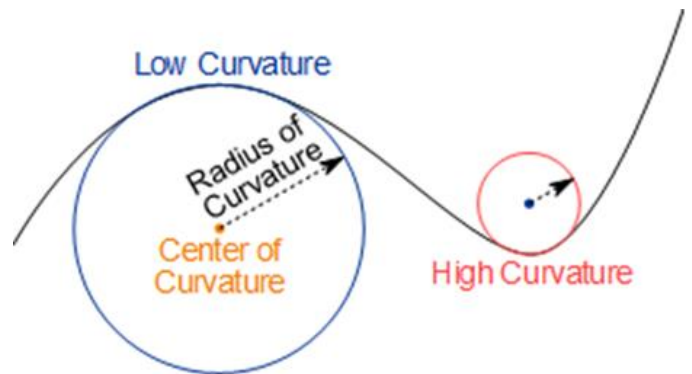


Figure 1 : Visual representation of the tortuosity calculation performed by method 2.

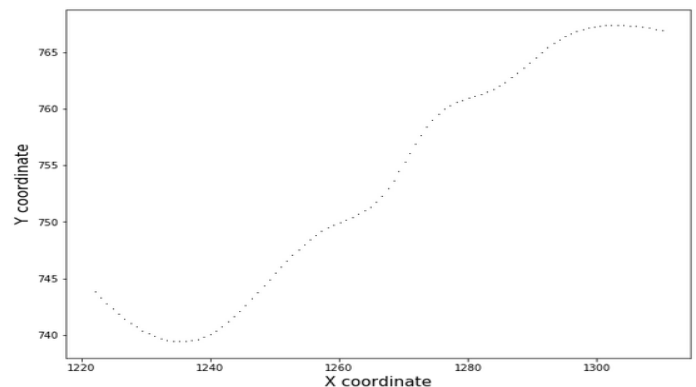


Figure 2 : Plot of tortuosity versus diameter calculated with method 2.

### C/ Statistical tools :

#### 1. Interquantile space method

We used the quantile method to determine the outliers of our DF: all the DFs that were above the 4th quantile were set as irrelevant, and thus deleted from the dataset. As shown on the right, we can again see that the formula of the quantile method is pretty simple and easy to manipulate and understand. As you can see, the calculated value was used as a threshold to remove all of the outliers: as said earlier every value above this one was set as an outlier and thus deleted.

$$Threshold = 3^{rd} \text{ quantile} + (1,5 * (3^{rd} \text{ quantile} - 1^{rst} \text{ quantile}))$$

## 2. Linear regression

For our analysis, we performed a linear regression. A linear regression model is a model whose purpose is to establish a relationship between variables. It includes one explained variable and one or more explanatory variables. If there is only one explanatory variable, it is called a simple linear regression model. As soon as there is more than one explanatory variable, we speak of a multiple linear regression model. The use of linear regression can be differentiated according to two categories: prediction, forecasting and error reduction or for the explanation of variation and quantification of a relationship between variables. In our case, we use it to see if there is a relationship between the tortuosity of the vessels and diseases such as strokes and angina.

## 3. Data Normalization

Normalization allows adjusting the values that have been measured on different scales. A normal distribution is a family of distributions characterized by symmetry and few outliers. Almost all observations are included in the range:  $\mu \pm 2\sigma$ . To quantify normality, it is possible to use a measure of asymmetry (skewness) and a measure of flattening (kurtosis). The purpose of normalization is to avoid transactional anomalies such or data redundancy that could result from poor data modelling. It reduces redundancy and increases data integrity. In our case, we wondered if we had to do normalization of our data, but as for the linear regression we removed the outliers as we will explain later, normalization wouldn't really be useful, so we preferred using the original dataset.

## 4. K-means algorithm

The k-means is a method used for the repartition of the data into groups (clusters) to minimize a certain function. It is a method of vector quantization. Its purpose is to minimize the within-group variances. We usually used the distance between a point and the means of

all the points of its cluster, and the sum of the square of these distances must be minimized.

## Results

Firstly, by plotting on some fundus images with Python, each point found by ARIA and adapting their diameter to the one found by the software, we managed to find that firstly some fundis are way more torturous than others, just by looking at them and by looking at their plotted blood vessels. This means that tortuosity is something unique about one another, therefore, that's something that we can maybe use for medical purpose.

To measure tortuosity, we had to choose between two methods. The first one is pretty simple and is, in fact, used in the paper we read to start this project. We used the distance factor (called DF) to calculate the tortuosity, but we also used a method that is a little bit easier to understand. In fact, it's based on the fact that with three points, we can draw a circle. If the segment is very tortuous, the three points would make a circle with a very small radius. But if the three points are pretty much aligned, the radius will be gigantic because the circle will be immense. After some times, we had abandoned the second method because if the line is "in general strait" but have some very small turns and returns (as you can see on the right), the tortuosity will still be considered as big, even if the blood vessel is pretty strait. By using the first method, we can have a better view and understanding of the tortuosity with a simple formula that is easy to handle.

But, this tortuosity problem can be linked to another problem, and this might be the way ARIA is working. We know that ARIA is a software originally used to find a plant's roots on an image, and that it was used for our blood vessels here. In fact, it's a pretty good idea: blood vessels are wired like roots and have the overall same morphology: some are long and wide, some are thin and short etc... So, at first, it sounded like a good idea to use it. But, when we plotted our blood vessels on the fundus' pictures, we noticed that some vessels had

problems: ARIA did not understand when a vessel was on top of another one (thus making two vessels when there was clearly only one), the software had also issues with measurements at the edge of the images, but also and above all: it found some blood vessels when there was clearly nothing (and very tortuous ones). With this pretty severe issue due to the software, we needed to find a way to discard these outliers. So as to do so, we used a simple yet quite efficient way to do it: we used the interquartile space method. As previously shown, in the preceding section, we used a simple method which is removing all of the outliers above the fourth quantile for our distance factor and our diameter.

Introducing the vessels tortuosity and diameter for each participant and selecting relevant features such as age, (systolic blood pressure), BMI, genetic sex to try to predict the diastolic blood pressure. This would result in fitting the data into a linear model which would then try to predict the diastolic blood pressure from our selected features. For each participant, we attributed one eye fundus to each person. If two eye fundi are available, we choose to assign them their left eye. We designed 3 different models :

1. Outliers participants who have a too high average tortuosity compared to the cohort would be removed.
2. For each eye, the vessels which have too high tortuosity would be removed because we consider them as outliers values, then we compute the average and assign this value to the participant.
3. The median tortuosity of each eye is computed and assigned to respective participants.

This first version gave a similar result for model 2 and 3, and a little bit different for model 1. As this wasn't the most interesting thing to predict (it can be precisely measured by non-invasive means) we got interested in diseases such as strokes and anginas. We decided to select the 2nd model as it was the most accurate and

useful according to us and we implemented a linear model to predict both strokes and anginas from the previously mentioned features. It seemed that a logistic regression model was more suited for this kind of prediction so we eventually switched to such a model.

We also did a k means, to confirm our linear regression model with a machine learning algorithm. As one can see on the left, with the table and the image, it is possible to see that there is no correlation between strokes and DF/diameter, since there are no defined groups with only green triangle.

| true      | cluster   |        |
|-----------|-----------|--------|
|           | No Stroke | Stroke |
| No Stroke | 2081      | 7328   |
| Stroke    | 51        | 122    |

Figure 3 : Truth table of the k-mean clustering

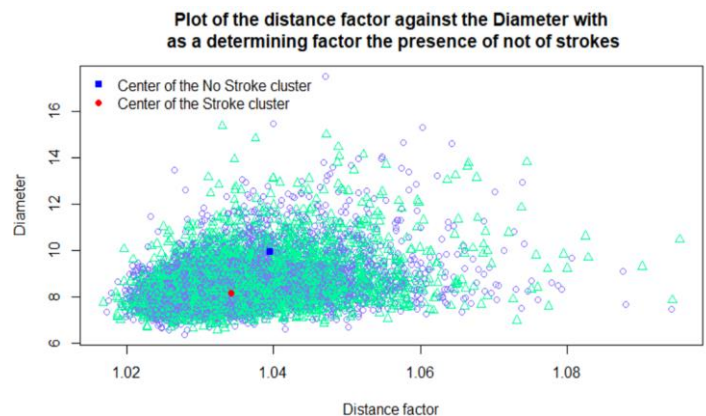


Figure 4 : Plot of the k-means clustering

We calculated the accuracy for this and it was around 0,23.

## Discussion

The dataset we had was not a perfect one: in fact, it was a bit biased in terms of ethnicity with many white people, about 50 times more than any other group. This can be explained by the fact that the Uk biobank is, as its name suggests, based in England a country where most of the population is of Caucasian origins. However, it is still biased in term of gender because there are more women than men, therefore the results we might have found will be female-biased, and therefore can be not

applied to the whole population. The numbers we used for this deduction are from the “Genetic Sex” data field and does only take into account the biological gender of the person. Moreover, the age was also a bias, because the median ages of all of the participant are 58, which is a bit old. And even though we were working on cardiovascular diseases that are commonly found in old people, in our idea of predicting the possibility of having or not having a cardiovascular disease, a perfectly balanced dataset in terms of age would have been better. But, since the concept of a biobank is that it’s based on voluntarism since it’s difficult to obtain and work on real phenotypic and genetic data, we cannot truly blame the fact that there are “too much” or “too little of a type” of people: this biobank is meant to represent the British population.

The fact that ARIA is more of a root software than a blood vessels software, but has still been used is somewhat a good and a bad thing for us. In fact, it gave us a way to find vessels and calculate their tortuosity, but the downsides are the many flaws of this software such as the creation of artefacts, or the split of blood vessels in two when they cross each other. Therefore, it was difficult to find a proper way to get rid of the outliers and the falsely created segments.

We did not find a single useful result other than confirming the literature: even though eye fundus can be used as a proxy on the body vasculature it is difficult to find reliable and useful metrics to predict if people are more or less prone to have cardiovascular disease. We did not find a good way to link eye vessels’ tortuosity to cardiovascular disease, therefore we had to include this parameter in linear or logistic regression models or the k means clustering algorithm taking into account

multiple other variables. The results we found are not statistically conclusive, so it will be hard to use them clinically, therefore as said before we cannot use them for diagnosis of cardiovascular disease. However, because this test is not as invasive as biopsy it would be an interesting alternative to diagnose vision and non-vision related diseases.

Maybe in order to find better results with our data, we would have needed more time to fine-tune the models we have created. It would also have been interesting to have more data such as if the participant had a stroke or angina after taking a picture of their eyes. And in general, we maybe should have looked more in the literature to see what is already known about eye vessels in relation to cardiovascular diseases because our research was too scarce<sup>4,5</sup> to be useful on this subject.

## Conclusion

The results that we get are not really the ones that we expected. In fact, we cannot find a correlation between tortuosity and strokes. Maybe this is due to the conditions of the experiment and the dataset that is not perfect, or this is just that there is nothing. Maybe if we had more precise data on cardiovascular disease and before and after images, we could have more accurate results. Despite that, we think that eye fundi can have potential, and if it’s not in the area of cardiovascular disease, that may be another area. Even if we do not find any correlation, it doesn’t mean that an eye fundus is not interesting, and that tortuosity of the vessels doesn’t reveal any relevant results. Further studies should be done in other domains. We also learnt a lot through this project, especially working as a team of future biologists with everyone’s background, strengths and weaknesses.

## References

1. Michael Abràmoff, Christine N. Kay, Chapter 6 - Image Processing, Editor(s): Stephen J. Ryan, Srinivas R. Sadda, David R. Hinton, Andrew P. Schachat, Srinivas R. Sadda, C.P. Wilkinson, Peter Wiedemann, Andrew P. Schachat, *Retina (Fifth Edition)*, W.B. Saunders, 2013, Pages 151-176, ISBN 9781455707379, <https://doi.org/10.1016/B978-1-4557-0737-9.00006-0>.
2. Oliverio, Giovanni William et al. "Safety and Tolerability of an Eye Drop Based on 0.6% Povidone-Iodine Nanoemulsion in Dry Eye Patients." *Journal of ocular pharmacology and therapeutics: the official journal of the Association for Ocular Pharmacology and Therapeutics* vol. 37,2 (2021): 90-96. doi:10.1089/jop.2020.0085
3. Pace J, Lee N, Naik HS, Ganapathysubramanian B, Lübberstedt T (2014) "Analysis of Maize (*Zea mays* L.) Seedling Roots with the High-Throughput Image Analysis Tool ARIA (Automatic Root Image Analysis)". *PLOS ONE* 9(9): e108255. <https://doi.org/10.1371/journal.pone.0108255>
4. Cheung, Carol Yim-Lui et al. "Retinal vascular tortuosity, blood pressure, and cardiovascular risk factors." *Ophthalmology* vol. 118,5 (2011): 812-8. doi:10.1016/j.ophtha.2010.08.045
5. Strandberg, Timo E, and Kaisu Pitkala. "What is the most important component of blood pressure: systolic, diastolic or pulse pressure?." *Current opinion in nephrology and hypertension* vol. 12,3 (2003): 293-7. doi:10.1097/00041552-200305000-00011
6. Lara Rodríguez, Luis David, and Gonzalo Urcid Serrano. "Exudates and blood vessel segmentation in eye fundus images using the Fourier and cosine discrete transforms." *Computación y Sistemas* 20.4 (2016): 697-708.