Detecting the Signs of Papilloedema

R. Fields, M. Gomes, G. Baron, A. Kumarvel.

The University of Sheffield

Abstract. Papilloedema is the abnormal swelling of the optic nerve, a common symptom of other diseases of the brain. This project aims to design and create a mobile application that can detect the signs of papilloedema in patients by taking a picture of the back of eye, applying machine learning and image processing techniques to the picture, and predicting the likelihood of papilloedema in the picture taken; hoping to increase the efficiency of diagnosis, speed up treatment and save time and money for hospitals, general practitioners and opticians.

Table of contents

Detecting the Signs of Papilloedema ... 1
R. Fields, M. Gomes, G. Baron, A. Kumarvel.
Introduction ... 3
Motivation ... 3
Project's Usefulness ... 3
Overview of Chapters ... 3
1 State of the Art ... 4
1.1 Diagnosing Eye Diseases ... 4
1.2 Products ... 4
Improvements on existing products: Proposed Solution ... 5
1.3 Image Processing ... 6
Current Techniques ... 6
1.4 Machine Learning ... 8
Feature Extraction ... 8
Neural Networks ... 10
1.5 MHRA Regulations ... 11
2 Requirements and Analysis ... 11
2.1 Introduction ... 11
Research and Interviews ... 11
2.2 Functional Requirements ... 11
2.3 Non-Functional Requirements ... 12
3D Printing ... 13
2.4 The Dataset ... 13
Backup Dataset ... 13
2.5 App Design ... 13
2.6 Image Processing ... 14
Problems with the Dataset ... 14
The Combination of Processing Techniques ... 14
2.7 Machine Learning ... 15
Chosen Techniques ... 15
Time Constraints ... 15
2.8 Ethics ... 15
3 Design ... 15
3.1 Implementation ... 15
3.2 Work Plan ... 16
4 Conclusions ... 16

References 16
5 Appendix 19

List of Figures

1 Fourier Transform Example 6
2 Histogram Equalization Example 7
3 Edge Detection Example 7
4 Haar-like feature 9
5 Image Integral 9
6 Neural Network visual exmaple 10
7 Data flow between mobile phone and server 12
8 Dataset Images 13
9 Flow chart showing how a user could interact with the application 19
10 Ethics Process Flowchart 20
11 Image showing how to use an Opthalmascope, taken from [9] 21
12 Opthalmascope 21

Introduction

This Darwin Project aims to create a mobile application that, through taking a picture of the back of the eye, will be able to detect the signs of papilloedema in a patient. The project will require the design and development of an application in accordance to the requirements identified, along with machine learning techniques to be able to detect and analyse important features in the images and return a suggestion. Image processing on the data set of photographs and the images collected from the app will also be used, to aid the machine learning process.

Motivation

Papilloedema is the swelling of the optic disc in the eye that is the result of raised intracranial pressure [13]; that is, the build-up of pressure inside the skull and brain tissue. The swelling can be bilateral, meaning it occurs in both of the eyes, or unilateral, although this case is extremely rare. It can be present for a period of hours or weeks and is mostly likely a symptom of another disease, but manifests in the form of headaches and their associated symptoms.

Not all optic disc swelling is papilloedema though, so careful examination needs to be carried out to be able to distinguish it from other causes. Checks can be carried out to determine whether the diagnosis is correct. These take the form of fundus photography or slit lamp examination and are performed by an optician or general practitioner; these techniques will be discussed later in the report. The abnormalities that are looked for are disc swelling and venous engorgement; since it is almost always bilateral, it is important that checks are carried out on both of the eyes.

If papilloedema is found, further investigations need to be carried out as vision loss can occur if left untreated. These further investigations can take the form of MRI scans and usually involve a hospital referral.

This project is being carried out in association with the Department of Ophthalmology at the Royal Hallamshire Hospital, who are helping with the ethical status that is needed for the research and acquiring a data set of images; both of which will be discussed further in the report.

The Project's Impact

By creating a simple device that can be used by physicians to give a reliable estimation as to whether a patient is likely to have papilloedema or not, many problems that have been identified could be solved.

As mentioned, the current way to check for papilloedema is inspection by eye and referral to hospital for further investigations. This is an unreliable method for checking and, if a false positive emerges, time and money is wasted on these further investigations. Hopefully, this application will be able to reduce these unnecessary costs as the only costs involved in the detection process are costs related to the mobile phone used to take the pictures and the lens needed; which are much more affordable than the current equipment being used to take the pictures.

If papilloedema is not treated in its early stages it can develop and cause the patient to go blind, or if it is a symptom of something going wrong in the brain, much worse could occur. Therefore getting patients help quicker is a priority, if the number of false positives is reduced and the number of earlier positive referrals is increased, patients can be treated sooner and not suffer as much as if they had to wait to be seen by someone senior.

A key element that needs to be maintained in the app is the ease of use, so someone with no training would be able to take the photos just as well as a medically trained professional. This aspect would then allow the general public to carry out the tests themselves at home. The proposed solution will be more available to the community e.g. to care homes, where the residents would struggle to get to the doctors and hospitals; or even allow the app to be made available to less developed countries with no access to the large and expensive cameras used, subsequently meaning that people there receive help is necessary.

Overview of Chapters

The rest of this report will now discuss the details, technicalities and design of the project.

Section 1 of this report will discuss the state of the art technologies in relation to how eye diseases (such as papilloedema) are currently being diagnosed, how the products that already exist are currently in use and the available processes for image transformations and machine learning. This report also discusses the MHRA regulations that the solution of this project needs to abide by. It is important to ensure that these regulations are considered at an early stage, since any changes further along this project's time line may be harder to implement.

Each aspect of the project and the ethics process that the project is subject to, along with the categorisation of functional and non-functional requirements is detailed in Section 2. The key aspects are the app design, the acquiring and handling of the data set, including processing of the images, and the machine learning algorithm that had been decided.

Then Section 3 discusses the design of our project with regards to implementation and group work, with the last section concluding the report with the plan for the next semester.

1 State of the Art

In this section, the current techniques and problems for diagnosing eye diseases will be discussed. Applying machine learning to medical images, specifically diagnosing eye diseases, has been a subject of study for many years. Therefore, the aim of this section is to highlight relevant systems previously introduced by researchers, the methods entailed in these researches, issues and further improvements that could be made. In addition, important image processing techniques, which apply to this project's dataset, and a quick overview of diagnosis made in hospitals will also be explained in this section.

1.1 Diagnosing Eye Diseases

The process currently used for diagnosing papilloedema is reliant on a thorough examination of the fundus; in an examination a physician should be able to detect signs of papilloedema by viewing enlarged and tortuous vessels. [17] As mentioned previously papilloedema is simply a symptom of other diseases with the brain, and it is for this reason that early detection of papilloedema is particularly important. If detected early enough appropriate measures can be taken to aid the patient in overcoming the disease. If a physician should detect signs of papilloedema they would be required to send the patient for an MRI scan. There are many products currently available to physicians to aid them in the detection of papilloedema. The following section will discuss these products and their advantages and disadvantages.

1.2 Products

Historically the methods used to view the fundus of the eye have relied on making use of and building on the technology already introduced to the world by photography. Following the first attempts of fundus capture, there were many problems discovered which needed to be accounted for in future designs, e.g. exposure time, controlling the beam of light, eye movement etc. Many years later there exist more complex devices that have accounted for these problems and are used on a regular basis, this section will discuss in greater detail the technologies available and their advantages and disadvantages.

Following the introduction of photography 1826, there have been advances in capturing the fundus of the eye. In the 1850s the ophthalmoscope was introduced by Hermann Von Helmholtz's, this was the first device which allowed for an examination of the fundus of the eye. [31] The best view of the retina will be when the pupils are dilated, and the ophthalmoscope makes use of this by redirecting light rays and bouncing them off tilted mirrors within the ophthalmoscope. The ophthalmoscope has two wheels, one of these wheels can be used to control the beam of light shone into the patient's eye. [32] The second wheel can be used to control the examiners view of the fundus, it is used adjust to the examiners eyesight. The figure 11 shows the different views and the parts of an opthalmascope.

The ophthalmoscope is one of the first devices used by ophthalmologists and optometrists to view the fundus of the eye. One of the reasons it is so widely used in the medical community, is because of its portability, and its cheap manufacturing cost. However as one can imagine the ophthalmoscope provides a very restricted view of the retina; this requires the examiner to have a basic knowledge of the different sections of the retina. Moreover the ophthalmoscope does not capture the retina, this increases the risk of misdiagnosis.

Another device used more commonly in Optometrists is the Slit lamp, this lamp involves controlling a beam of light which enters the patient's eye after being redirected off a mirror. [22] Slit lamps have previously been used in optometrists and hospitals, however they are now mostly found in optometrists offices, and hospitals employ a different product which will be discussed shortly. The paper by MD Alan M.Mindlin et al. entitled Slit Lamp use, [22] discusses the procedures followed as part of the slit lamp examination. The paper also briefly discusses advantages and disadvantages to the slit lamp. Despite being able to provide a better view of the fundus than the ophthalmoscope, the slit lamp still only illuminates a small section of the retina. Moreover, during an examination with a slit lamp, a patient will have a beam of light shone into their eyes. While the slit lamp has been designed to ensure that this light is safe, this still causes discomfort to the patient. As with the ophthalmoscope, the slit lamp also doesn't allow for capture of the fundus, and therefore this technique also relies on the physicians notes/diagrams. Despite the product's shortcoming's, this is still widely used in optometrists offices, this is due to the ease in maintenance. The slit lamp is quite robust and therefore if it needs repair it will be to simply fix a fused lamp. This is quite an easy repair, and this means that when broken these machines are not out of use for a long time.

As discussed before, there is another product which is used by physicians to view the fundus of the eye, and this is the Non-mydriatic Fundus Camera. [11] Mydriatic is a phenomenon in which the pupil is dilated. [21] Therefore the non mydriatic fundus camera does not require the patient's pupil to be dilated. The fundus camera is far superior to the other products discussed this is because the examination conducted with a fundus camera doesn't put the patient under any distress. [11] It also allows for capture of the fundus. The high quality images taken complement the ophthalmologists notes and together these are used to provide better diagnosis and ultimately better care for the patients.[11] Despite its superiority, the fundus camera does have disadvantages, one of these is the size of the product. Since this is quite an expensive product it is not very accessible and therefore a hospital will only have a few of these (depending on the size of the hospital and the number of patients it tends to). The fundus camera is also unable to detect anomalies outside the photographic view, therefore a physician may still miss abnormalities. [11] The figure 12 shows a non mydriatic fundus camera, this image can be used to see how large the equipment is.

The following table (table 1) shows an overview of all the products discussed in this section. It can be used to compare the products that already exist.

Product	**Where is it used**	**Advantages**	**Disadvantages**	**Price**
Opthalmascope	Mostly used by GPs	Portable	Provides restricted view of retina, no image capture	£380
Slit Lamp	Mostly used by Optometrists	Easy to maintain	Provides restricted view of retina, does not capture fundus of the eye, potential discomfort to patients and it's not portable	£3,555
Non mydriatic fundus camera	Mostly used by Opthamologists	Easy to maintain, allows for image capture of fundus	Very expensive, not portable	£5,000

Table 1: Overview of existing products

Improvements on existing products: Proposed Solution

There are many options available to physicians to aid in the diagnosis of papilloedema, however even the best product available to the physician can still lead to misdiagnosis. This project intends to provide a solution which should reduce the cases of misdiagnosis. The proposed solution (which will be discussed further detail in a later section), will make use of the advances in technology to provide a suggestive diagnosis to aid a physician. It will improve upon the existing products as it will be more portable, cheaper than the fundus camera and captures the fundus of the eye while placing the patient under as little stress (from the examination) as possible. The proposed solution will also be easier to use, therefore allowing it to be used by any non medical person; making it more accessible as it can be taken to diagnose people in the community.

The closest existing ancestor to the proposed solution is the Volk iNView. [35] This is a modern device available on the market for fundus photography. The Volk iNView has a large fundus lens attachment that clips onto the back of a mobile phone. [35] To take a photograph, a free application is provided that is designed for use with the lens. The application does not provide any image analysis and the lens attachment is more expensive to purchase than the proposed solution. The solution will improve on the Volk iNView by providing a basic image analysis. [35] Despite its portability, the Volk iNView will still need to be used by a physician to examine the fundus. [35] This makes the product less accessible to the public, this product is also quite expensive making it harder to acquire. This project's proposed solution will overcome these shortcomings by making an app which is easy to use and intuitive in design. The app will also provide the patient with advice, e.g. "We recommend you see your GP for a more thorough exam." Any extra components required for the proposed solution will be cheap; making mass production cheaper.

1.3 Image Processing

The purpose of image processing is to alter an image, or images, in any number of different ways to suit their purpose. A selection of image processing techniques will be detailed here and later in the report they will be related to this project.

Current Techniques

- **Fourier Transforms** can be used to calculate the frequencies present in a photo and later to sharpen an image. The values of the frequencies being low or high is decided by a given threshold and the ratio can be is assessed. A low count of high frequencies would indicate a blurry photo. The transform decomposes an image into its sine and cosine components and then outputs an image in the Fourier or frequency domain, with the original image being in the spatial domain[27]; Figure 1 shows two images and their corresponding frequency domains. Each point in the frequency domain image represents a particular frequency contained in the original spatial domain image [34]; so the number of frequencies corresponds to the number of pixels in the original image, meaning the two images in different domains are still the same size. The Fourier image is shifted in such a way that the centre is the image mean F(0,0), meaning then that the further away from the centre an image point is, the higher its corresponding frequency is. The ratio is now assessed and altered according to the ratio that has been deemed appropriate for the images and the frequencies are transformed back to the original domain; for Figure 1 the transformation would be from the right image to the left.

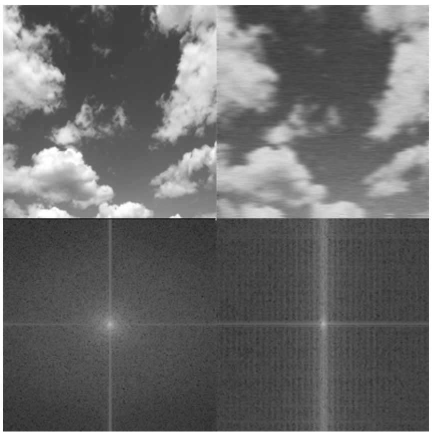

Fig. 1: The fourier transform applied to an image, the frequency domain representation of each image is below it. The image on the right has had noise added and been blurred, with the one on the left being the image after the transform.

- **The Iterative Closest Point** algorithm creates two point clouds from a standard image and the image being altered. Key features will be identified in both images and the algorithm then tries to match the points at the key features through rotation and translation transformations. The aim of the matching process is to minimize the square errors being corresponding points [18], by continuously checking them.

- **Histogram Equalisation** can be used to alter the contrast in the image [29]. It is used to generate better views of bone structure in x-rays and can give better details in photographs that are over or under-exposed. Equalisation is generally used on greyscale images so that each pixel only has one value, the intensity of the pixel. The histogram is plotted using the pixel intensity against the frequency of the pixel intensity and the method transforms the image so the histogram of the resulting image is constant.

  A matrix of pixel intensity is firstly created, the dimensions being the size of the image in pixels. Then the frequency of each intensity present in the matrix is counted, and the probability of each calculated. The cumulative probability is found; that being the probability that the intensity falls within a specified range, and this is multiplied by the maximum value in that range. Any decimal numbers are floored and placed back into matrix form, hopefully giving a new image with better contrast [15].

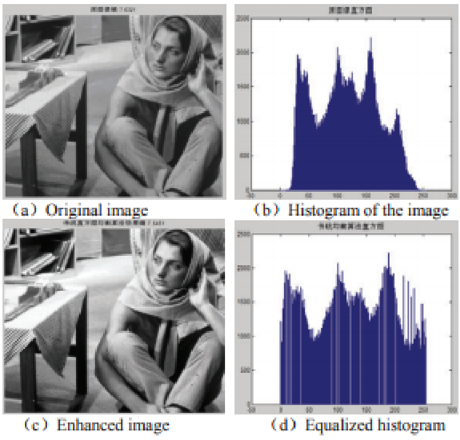

Fig. 2: An image that has been enhanced by histogram equalisation, the contrast in the enhanced image is much more prominent that the original [36]

- **Feature Removal** is a process that can being used to disregard unnecessary features from an image, leaving the important parts. Canny Edge Detection [25] can identify the edges present in an image and then provide an image containing only these, Figure 3 shows the before and after of using this technique. It works by firstly filtering out noise and calculating the intensity gradient in the image. Defining two thresholds, an upper and a lower, non-maximum suppression is used to remove everything that isn't considered to be part of an edge; that being that if a pixel intensity isn't part of the upper limit then it is set to zero. Then the edges are found using the logic that; if a pixel gradient is higher than the upper it will be an edge, if it is less than the lower it is rejected, and if it is in between the two then it is accepted as an edge if it is connected to a pixel that is accepted [12].

Fig. 3: The image on the left is the original image and the one to the right is after the edges of feature objects had been returned.

1.4 Machine Learning

Since this project's objective is to provide users with a mobile application that can reliably detect signs of papilloedema, it is crucial that the correct machine learning techniques are implemented. In this section, the relevant machine learning techniques will be discussed, as well as the methods which are currently being used for medical imaging and detection of eye diseases.

Firstly, it is important to consider the different properties needed for the detection component of the application being developed. As the underlying problem is to detect patterns in medical images in order to produce a recognition percentage, image classification/recognition algorithms will be considered and implemented in the development of this project. For this reason, list **??** summarises necessary properties which the machine learning techniques being implemented in this project must have.

1. **Rotation Invariance** - Pictures will be taken using mobile phones, hence the orientation of the images will possibly vary. Using rotation invariant algorithms will ensure that the detection component of the app works equally well for all the users.

2. **Efficient when Evaluating** - Since most of the computation will be done on the server side (see section 2.3), it is crucial that the recognition algorithm runs smoothly and efficiently. This allows multiple users to access the application's features at the same time without crashes or noticeable lags.

3. **Shape, Brightness and Contrast Invariance** - As aforementioned, different users will have access to this application, for this reason, the images will be taken with different cameras and lighting conditions. Therefore, having an algorithm which performs well in different lighting conditions is idea. In addition, the shape of the images may also vary with the cameras being used and the angles of the images.

Although the ideal machine learning technique will have the properties above, image processing techniques will be applied to the dataset in order to normalise the images. Sections 1.3 and 2.6 highlight the techniques being used and how they relate to this project.

In the upcoming subsections, the main approaches that are currently being used to detect patterns in images will be discussed. Note that advantages and disadvantages of each approach will be highlighted, as well as the underlying method of each approach and their properties. The main image pattern recognition methods are reviewed in the following subsections.

Feature Extraction

Introduced by Viola and Jones [26], Haar-Cascade-like algorithms extract features from image datasets using the concept of image integral (or summed-up area table) and AdaBoost in order to obtain rectangular features which can identify an object in an image (see figure 4). Image integral is the process of summing up the greyscale values of rectangular areas in an image with the purpose of identifying how light/dark a portion of an image is - this is done in linear time. AdaBoost refers to the boosting technique used to create strong classifiers given a set of weak classifiers. Feature extraction algorithms tend to be extremely fast when evaluating, however, the process of training the classifier can take days depending on the resolution and complexity of the images, and the classifier generated is not rotation invariant. Note that the underlying methods used to obtain the final object classifier are explained in the upcoming text.

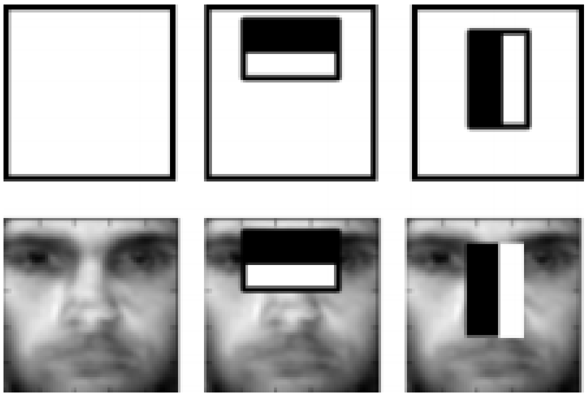

Fig. 4: Example of haar-like feature being evaluated on a face classifier.

As described by Bradley, D. and Roth, G.[8], it is possible to calculate the summed up area of a rectangular portion of an image in linear time, using the expression $I(x, y) = f(x, y) + I(x-1, y) + I(x, y-1) - I(x-1, y-1)$. $I(x, y)$ is the summed up area of rectangular area from the origin (top-left of the image) to the pixel located at (x, y), and $f(x, y)$ is the greyscale value of the pixel at (x, y). Once the image integral has been stored for all the pixels, the summed up area of an arbitrary rectangle (top left at (x_1, y_1) and bottom right at (x_2, y_2)) can be calculated in constant time using the relation $D = I(x_2, y_2) - I(x_1 - 1, y_2) - I(x_2, y_1 - 1) + I(x_1 - 1, y_2 - 1)$, where D is the summed up area of the said rectangle. Figure 5 demonstrates this process, with the relevant pixels labelled in the figure.

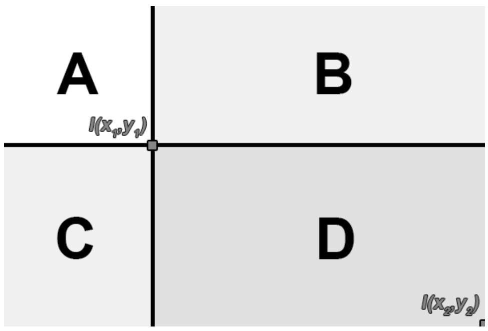

Fig. 5: Image integral example. The summed up area of D can be calculated using $D = I(x_2, y_x) - B - C - A$, which is equivalent to the expression previously mentioned.

In machine learning, the term "boosting" simply means the process of generating strong classifiers from a series of weak classifiers – classifiers that, alone, cannot fully detect the desired pattern. AdaBoost, or Adaptive Boosting, which was introduced by Freund, Y. and Schapire, R.E[14], is an adaptive approach to the boosting problem. Freund and Schapire's boosting technique consists of tweaking the weights of the weak classifiers with respect to the previous false-positives and false-negatives in order to improve the subsequent set of tweaked weak classifiers. In addition to adapting the weights of the weak classifiers, AdaBoost plays an important part in selecting the order in which the weak classifiers are applied – in the Haar-Cascade algorithm, features which reject the highest number of false-positives are applied first, and the subsequent levels of features to be checked are more complex. Therefore, this "cascading" will reject most of the false positives within the first few levels of features, increasing the overall evaluation performance.

Neural Networks

The human brain consists of neurons (brain cells) and neuron connections (or synapses), which send/receive electro-chemical signals to/from other neurons. Roughly speaking, the incoming signals to a neuron are combined in some way, and if the combined input exceeds a threshold, the neuron then "fires" an output signal to other neurons. Neural Networks model the way human brains work – they consist of layers of nodes (which represent neurons) and connections from each layer (representing the synapses). The description of Neural Networks provided in this report is merely a short summary of such an extensive topic; a more detailed description can be found in the first chapter of Kevin Gurney's book[16].

Each layer in a Neural Network – NN for short – is fully connected to the neighbouring layers, and each node connection has a weight, which, broadly speaking, represent how much of the information in a node is carried to a different node in the subsequent layer. The way in which neural networks learn is comparable to learning process of the human brain – it adjusts the weights of each connection in order to minimise the overall error. In supervised learning, the error function, usually denoted as $\mathbf{E}$, is simply defined as the sum difference between the model outputs of all the samples (the predictions) and the actual value. This relation is mathematically modelled as $\mathbf{E} = \sum_{x \in X}(y_x - f(x))^2$, where each x is an individual sample, y_x is the corresponding actual value and $f(x)$ is the prediction value. It is worth noting that this definition of the error function is only valid when the predictions and the actual values are represented numerically. For cases where the model of the values are not naturally represented with number, the error for an individual prediction can be assigned to a meaningful value, such as $E(x) = 1$ if $f(x) \neq y_x$, and $E(x) = 0$ if $f(x) = y_x$, where $E(x)$ represents the error for a single sample.

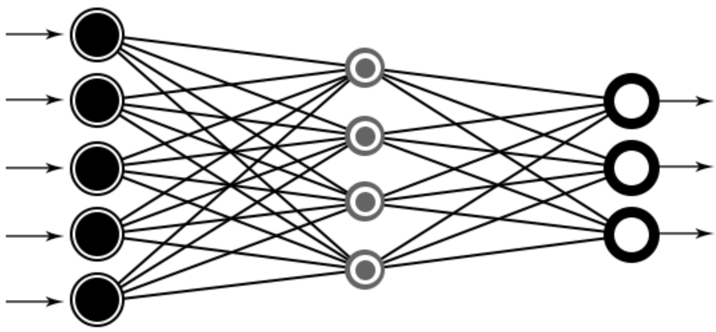

Fig. 6: Visual example of a NN with 3 layers. The input layer and the output layer are on the left and right respectively. The inner layer is called the "hiddel layer".

Without loss of generality, assuming N_i represents the number of nodes in the i^{th} layer, the number of weights (or connections) contained inside a NN – that is, not counting the number of input, output and constant weights – can be calculated with the expression $\mathbf{W} = N_1N_2 + N_2N_3 + \ldots + N_{m-1}N_m$, where $\mathbf{W}$ is the total number of weights, $m \in \mathbb{N}$ is the total number o layers in the network. In figure 6, there are 3 layers, containing 5, 4 and 3 nodes in each respective layer. Using the model for the number inner weights, the network in figure 6 contains $5 * 4 + 4 * 3 = 32$ inner weights. It is clear that the number of layers affect the number of weights in a NN. Fortunately, the algorithms used to adjust the weights tend to scale well with data and input size – for example, Stochastic Gradient Descent[7], only updates weights which are relevant for a particular training sample, hence not all the weights are updated at once.

Although conventional NNs seem to model the human brain quite well, it is not normally used for image pattern recognition, as it does not encode the spatial property of images. For this reason, **Convolutional Neural Networks**, or CNNs, are used in the field of image processing. The main difference of conventional NNs to CNNs is that CNNs contain extra layers (hence being considered a deep learning technique[1]) which encode the architecture of images quite well. The main layer in CNNs is the convolutional layer, which uses the convolution operator to apply a kernel, or convolution matrix, to the images. This operation essentially applies filters (such as blur and sharpening) which might be used for feature

[1] Deep learning simply refers to NN architechtures which contain multiple hidden layers (usually more than one).

extraction, while keeping the spatial relation between pixels. A more in depth description of such CNNs can be found at Ujjwal Karn's post about Convolutional Neural Networks[33].

Although the features obtained in CNNs are not themselves rotation invariant, there are adaptations made to neural networks, such as Rowley et. al's face detection NN [28], which are rotation invariant. The disadvantage of using CNNs is that the training step usually requires a very large dataset.

1.5 MHRA Regulations

The Medicines and Healthcare Products Regulatory Agency (MHRA) is a regulation body which censors any new medicines and medical devices that enter the field of operation.[20] In the proposed solution the mobile application and the 3D printed arm (which will attach the 20 Dioptre lens to the mobile phone) will be scrutinised by the MHRA. Therefore this project needs to ensure from an early stage that these regulations are not violated. The MHRA regulations require the proposed solution to have a small margin of error and the proposed device must include fail safe safety protocols which can be activated. [19] The regulations also require that any failing on the part of the medical device should not have adverse affects on the clinical condition of the patient, moreover the safety of the patient should never be compromised over the lifetime of the device. [19]

The proposed solution will never provide a user with a definitive diagnosis, this project merely aims to provide a suggestive diagnosis and it will only recommend a user to consult a doctor. In the borderline cases where the application is unable to diagnose the patient it will likely recommend the user to visit the GP. In this way, the solution will avoid causing any adverse affects on the condition of the user. At every stage of development, there will be regular discussions to ensure that there are no violations of the regulations.

2 Requirements and Analysis

2.1 Introduction

This section of the report will relate to what has been discussed previously to the implementation of the project. It will detail the functional requirements and the non-functional ones that would be beneficial but due to the limitations surrounding the project, may not be achieved. These will then be expanded to explain how the project will be designed and implemented as a whole for each major section, drawing on the different techniques researched. Considerations, constraints and the methodology used are also going to be addressed.

Research and Interviews

As mentioned before, this project is a collaboration with experts at the Royal Hallamshire Hospital. This is beneficial to the developers on this project as they will be able to conduct regular interviews with one expert in particular. During these interviews there will be clear agendas made which will aid the discussions to ensure that the pertinent topics are discussed. The developers on this team are looking forward to further discussions with the medical expert in order to gain a fuller understanding of the problem. There will also be a significant amount of desk research done around these meetings/interviews. From this research the developers hope to better understand the difficulties faced by other researchers. As all the collaborators work side by side on this project, the shared knowledge of the respective fields can and will be put forward towards the progression of this research.

2.2 Functional Requirements

The final product will need to meet certain standards and have certain capabilities in order to be deemed successful. These capabilities are:

- **Application** - the app created should be able to take a photograph of the back of the eye. There should be a guideline on how to take the photos and steps to help with the image processing. It should display the suggested diagnosis on the mobile screen.

- **Data** - the data set of images for training should be thoroughly processed through relevant transformations. The images taken when using the app should also have to ability to be transformed before undergoing the detection process.
- **Detection** - the machine learning algorithm chosen should be adequately developed and trained to detect the swollen optic nerve and vessels in the eye. It should then be able to process the images passed from the app and return a suggested diagnosis to the user.

The details and methods needed for each of these requirements are now going to be explained in the rest of this section.

2.3 Non-Functional Requirements

In section 1.3 and section 1.4, state of the art image processing and machine learning techniques were discussed. To apply these techniques effectively, capable-enough hardware would be required. Unfortunately, a mobile phone has much more limited processing power than a typical desktop computer; for instance, the iPhone X has an A11 Bionic Chip with 2.39Ghz clock rate and an integrated GPU [4]. This is a result of mobile phones not being designed for large computational work. To address this issue, a dedicated server can be used and a connection can be established between it and the phone.

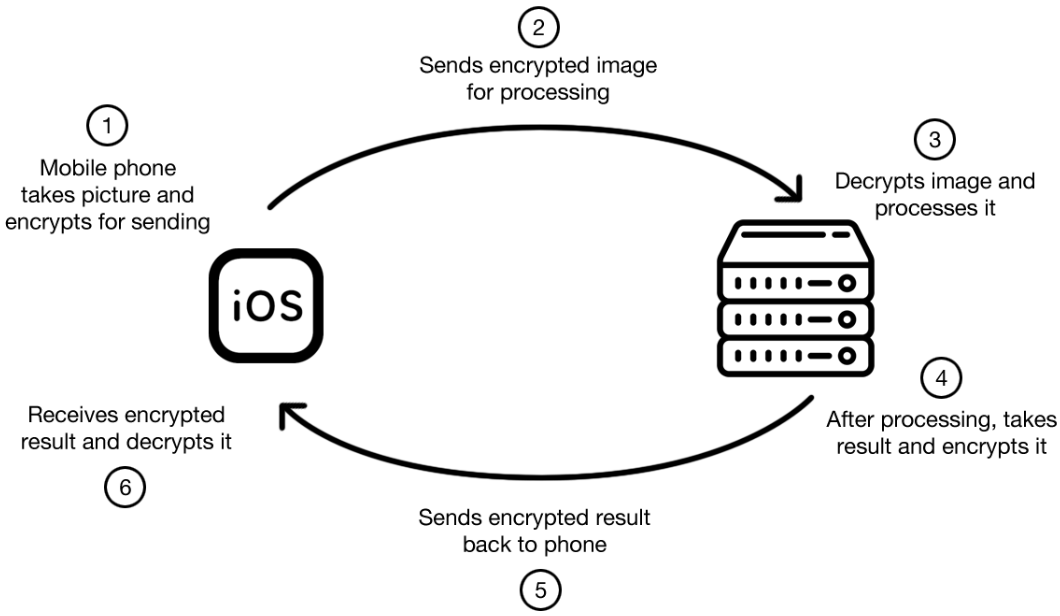

Fig. 7: Data flow between mobile phone and server

Figure 7 outlines the flow of data between the mobile phone application and the server. The user begins the process by taking a photo of the retina using the mobile application then agreeing to send the image to the server for processing. The server, with its higher computational power, then applies the various machine learning and image processing algorithms to derive a result. The result, for ethical purposes, will be a percentage likelihood of the image containing papilloedema.

Once a result has been obtained it is then sent back to the mobile phone application. This transfer of information will happen asynchronously to allow for further use of the mobile phone while server-side processing occurs.

For further ethical purposes, the data channels between the mobile phone and server would have to be encrypted. This is due to the sensitive nature of the data being transferred. Therefore, an extra step is required before sending and receiving data; once the user has verified the image to be sent, it must be encrypted. Once the server receives the image, it must then be decrypted before processing. These same steps are mirrored for the sending and receiving of the result.

3D Printing

3D printing is still a fairly new technology, but it can be used to create a wide array of objects for many uses. The University of Sheffield owns 3D printers that can be accessed for the purpose of the project if needed. The use for such technology here is the possibility of an arm being developed that will hold the phone and the 20 dioptre lens used to magnify the eye at a set distance and remove some of the human error that will be present.

Using 3D printing is a quick and relatively low cost production method, with hopefully the pay off of better diagnosis. The difficult aspect would be the designing of the arm, as the length of the arm should be the focal length of the lens used. And this would have to be calculated correctly in order to achieve the best image being captured. However, this is an additional feature to the project as the product should be able to work without the use of the arm.

2.4 The Dataset

Since the core features of the app being developed in this project heavily rely on machine learning, the quality of the dataset plays a huge role in the completion of this project. When discussed with Dr. Imran (from the Hallamshire Hospital), it was established that, if granted the permission to use images taken from the Hallamshire Hospital, it would be possible to acquire approximately 2 thousand images in the space of 3 months. Therefore, if this period is extended, it would be possible to obtain roughly 4 thousand images (taken with mobile phones) by the end of this project. This gives a lower bound of 2.5 thousand images in the dataset. Depending on the complexity of the pattern being analysed on the images, the said dataset size could suffice for the purposes of this project.

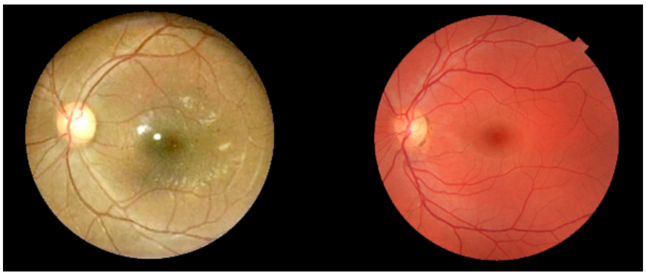

Fig. 8: Dataset images examples. Left image was taken with a phone, image on the right was taken with a fundus camera (Wikipedia).

Backup Dataset

If, for some reason, the number of images acquired for the dataset is relatively small, the NHS currently has a dataset of annotated images of the fundus of the eye, taken with the fundus camera. This dataset contains 5k+ images, and this dataset size should be enough to train the machine learning algorithms discussed in section 2.7. In addition to the NHS dataset, the MESSIDOR dataset[30] (which is free to use) contains thousands of images of the back of the eye. These are annotated images that are not necessarily images of unhealthy eyes, but enough to test some of the techniques implemented in this project.

2.5 App Design

The application is being designed for iOS devices and as such must follow the Apple human interface design principles. The guidelines value consistency and ease of use, specifying particularly designed sets of components such as buttons and tabs along with various layout and sizing considerations [3].

To adhere to the guidelines the decision to keep the interface simplistic with a reduced number of sections was made. This allows for a simple and easy to understand program flow. It also considers the possibility of frequent usage by reducing the number of steps between core functions of the application. The following flow diagram demonstrates this.

Figure 9 in the appendices shows the high level flow of the application for a user standpoint. Upon opening the application, the user can browse the images in the image gallery view. From here, the user may either select a particular image with the intention of processing it or take a new photograph. The actual image processing happens asynchronously and is not controlled by the user and was therefore excluded from the diagram.

2.6 Image Processing

Problems with the Dataset

The purpose of the image processing stage is to be able to deliver the best set of data to the machine learning stage; in order to aid the feature learning and detection stages. This can be done through a combination of numerous transforming aspects that were explained in section 1.3, with each aspect being the solution to a particular problem identified in the data collection process.

The possible problems and their solutions identified were:

- Greyscale - most machine learning algorithms work best with a greyscale image, so the images in the data set will be converted to greyscale before they are passed to any more transformations. Some of the transformations used will also require the images to be in greyscale.
- Focus - A photo is likely to be out of focus due to human error in being able to take a perfectly still photo. Photos returned from the professional machinery may still be blurry, possibly not to the extent of one from a phone, but due to the uncomfortable nature of having light shone into the eye the eye could move and blur the photo in this way. The Fourier transform can be used here to alter the frequency domain and sharpen the images.
- Orientation - All the photos not having the same orientation with respect to key features in the photo, such as the optic nerve, is desired when passing the data set into the machine learning algorithms. This would be more of a problem with the photos taken by hand rather than machine, as the user can rotate a phone to get a photo but the machinery depends on the patient being in a certain place so their photos should always be the same.

  The iterative closest point algorithm described in section 1.3 could be useful here, but as no two pictures of eyes are going to be the same, matching feature points between them could turn out to be very difficult. A simple alternative could be to store the coordinate position of the optic nerve in a photo and continuously rotate the image by smalls amounts until this coordinate matches to a standard value set.
- Contrast - The brightness or contrast of photos is not always consistent, even when using machinery, so there needs to be a way to normalise this feature of the images in order to accurately assess them all equally. Histogram equalisation would be relevant to this problem, and the standard contrast level can be set, with each individual image histogram being altered to match that level.
- Unnecessary Data - Passing in purely the key features that will be used to detect the health of the eye for the machine learning algorithm would speed up the process, so a method that extracts the optic nerve and veins from each image and then passes a data set of just these would be beneficial. Canny edge detection (as discussed in section 1.3) is the most straight forward method to use here as it would hopefully detect the veins and optic nerve, therefore removing the back wall of the eye. The images could also be cropped around the optic nerve using a set radius to further reduce the amount of data being sent.

The Combination of Processing Techniques

The methods described above can improve the data set; but the application order of the methods could potentially derive alternate outputs. A way to combat this issue would be to test the methods needed on a small sample set in a variety of combinations, prioritising the most important features that need to be corrected. The outputs could then be compared across different starting images to see which combination give the best results.

2.7 Machine Learning

In section 1.4, some of the most popular approaches to image processing using machine learning were discussed. In this section the dataset, time constraints and the general complexity of the problem will be considered in order to justify the approach chosen for the machine learning component of this project.

Chosen Techniques

Given the dataset proposed in section 2.4 and the complexity of the problem – which is related to the functional requirements of the machine learning component discussed in section 2.2 – it can be deduced that there are two main procedures when detecting signs of papilloedema: detecting the optic disc and the extraction of vascular patterns in the images of the fundus of the eye. Once these two procedures have been applied to a given input image, a simple pattern recognition algorithm can be applied to the resulting information in order to produce a prediction.

Aquino, A. et al.[6] have introduced a method for detecting the boundaries of the optic disk using edge detection (discussed in section 1.3) and feature extraction techniques. The said method was evaluated using 1200 images from the MESIDOR database, and the average time taken for the computation was 1.67 seconds. This article is an example of how each of the procedures could be enacted; using image processing techniques and feature extraction to separate the information from the image.

In addition to using feature extraction methods, the actual detection of the signs of papilloedema will be done using CNNs. Given the size of the database that will be obtained, it has been agreed that CNNs would be the best choice for the pattern recognition feature of the app. This is because it has rotation, shape and lighting invariance properties.

Time Constraints

Unfortunately, there is a limited amount of time to complete the project. From when this report was written, the deadline was only four months away. For this reason, it is absolutely necessary to make sure that all the machine learning algorithms are implemented, trained and ready to be tested at least 4 weeks before the final deadline. In addition to the implementation constraints, it is also worth noting that the data for the database has to collected before the training and testing.

2.8 Ethics

Every research project needs an ethics approval, this is done to ensure that any data collected over the course of this project is handle with sensitivity. This project has already submitted an IRAS application and following approval data sets can start to be formed and provided to the developers. Figure 10 has been taken form the Sheffield Clinical Research Office [10] and it details the steps that have been taken as part of ethics application process.

Following the ethics approval process the collaborators at the Royal Hallamshire hospital will be able to share any images that have already been taken and will be able to collect images and form a data set. As part of the ethics process one of the things to consider will be what needs to be done in the case that an abnormality is detected. This is an ongoing discussion amongst the members of this project. However the first priority at this stage is securing an approval to begin a data set which can then be used to begin the image processing.

3 Design

3.1 Implementation

The mobile application will be developed using the modern Apple development work flow. This involves the use of XCode as the Integrated Development Environment (IDE) for the purposes of deploying to Apple devices [2]. Further to this, the application will be written using the Swift programming language [5]. Swift, unlike the previously used language for iOS development, Objective-C, is designed to be readable and maintainable. This is achieved by abstracting various concepts such as memory management and C-style headers away from the user and simplifying the overall language structure.

Image processing and machine learning implementations on the server require a separate development strategy. For the core program structure, Python will be used. Python is well accustomed and designed for highly mathematical situations and is therefore perfect for image processing applications. To greatly improve the speed and efficiency of the processing, CUDA [24] will also be incorporated. This lends the computational power of General Purpose Graphics Processing Units (GPGPUs), ideal for highly parallel situations such as image processing. CUDA is a C-style parallel computing platform that is separate to Python - therefore to interface with CUDA the PyCUDA library [1] will be used. This library provides various wrappers and methods to maintain the Python language structure as opposed to a C language structure.

3.2 Work Plan

The development team consists of four members and therefore work tasks should be distributed evenly depending on the particular skill sets of the individuals. However, due to the small size of the team, it would also be acceptable to help each other for various, more complex tasks.

The main development focus in the initial stages will be to prototype various image processing and machine learning algorithms to decide on the best approach. Concurrently, the application will also be in development. As the application is likely to be a smaller part of the project, once it has been completed resources can be shifted onto mobile-server communication and the optimisation and improvement of the image processing algorithms with GPU acceleration.

4 Conclusions

The broad aim of the project is to develop a mobile application primarily for doctors that, through the use of fundus photography, image processing techniques and machine learning, will be able to detect the signs of papilloedema in a patient. In particular, processes such as Fourier Transformations, Histogram Equalisation and Feature Extraction as discussed in section 1.3 can be used to simplify the image data for a more successful analysis by a neural network (as detailed in section 1.4). Further to this, the application will require an interface designed for ease of use and efficiency with the ability to encrypt image data to abide by the ethical standards outlined in section 2.8.

Section 1.3 highlighted the main image processing techniques which would be useful when normalising the data for the machine learning component of the project. More specifically, how the orientation of images could be solved using automated methods (ICP), contrast and brightness correction with Histogram Normalization and Fourier transforms to correct blur and noise. Ideally, all the methods discussed would solve the problems mentioned in section 2.6, and prepare the images for the machine learning part.

Each of the machine learning techniques researched for this project has its own advantages and disadvantages. As seen in sections 1.4 and 2.7, modern methods tend to use a combination of techniques for feature extraction and pattern recognition. In addition, given the data proposed in 2.4, the machine learning techniques chosen comply with the complexity and time constraints of the overall aim of the project. Hopefully, the final application will be able to use the approaches chosen in order to quickly detect the signs of papilloedema.

Currently, some image processing techniques and boiler code for applying filters to datasets have been implemented, as well as acquiring images from the MESSIDOR dataset. In addition, the mobile application development has also been started. However, in the upcoming semester, the remaining image processing techniques must be implemented (such as the Histogram Equalization), neural networks and feature extraction for the fundus images also need to be completed. Lastly, once these techniques have been implemented locally and tested, the integration with the servers and mobile phones will be done – these details are discussed in sections 2.3 and 3.1.

References

[1] Andreas Klöckner (2017). PyCUDA. https://mathema.tician.de/software/pycuda/.

[2] Apple Inc. (2017a). Apple Developer Program. https://developer.apple.com/programs/.

[3] Apple Inc. (2017b). Human Interface Guidelines. https://developer.apple.com/design/.

[4] Apple Inc. (2017c). iPhone X Specifications. https://www.apple.com/iphone-x/specs/.

[5] Apple Inc. (2017d). Swift. https://developer.apple.com/swift/.

[6] Aquino, A., Gegúndez-Arias, M.E. and Marín, D (2010). Detecting the Optic Disc Boundary in Digital Fundus Images Using Morphological, Edge Detection, and Feature Extraction Techniques. *IEEE transactions on medical imaging*, pages 1860–1869.

[7] Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. *Proceedings of COMPSTAT'2010*, pages 177–186.

[8] Bradley, D. and Roth, G. (2007). Adaptive Thresholding Using Integral Image. *Journal of Graphics Tools*, 1:13–21.

[9] Clara Hoi Ka Wu (2017). Using an Ophthalmoscope. https://www.fastbleep.com/biology-notes/20/286.

[10] Clinical Research Office (2017). Clinical Research Office: For Researchers. https://www.sheffieldclinicalresearch.org/for-researchers/.

[11] Coburn Technologies (2017). Advantages and Disadvantages of a Non-Mydriatic Fundus Camera. http://www.coburntechnologies.com/2016/04/12/non-mydriatic-fundus-camera/.

[12] doxygen (2017). Canny Edge Detection. https://docs.opencv.org/2.4/doc/tutorials/imgproc/imgtrans/canny_detector/canny_detector.html.

[13] Dr Mary Lowth (2016). Optic Disc Swelling, Including Papilloedema. https://patient.info/doctor/optic-disc-swelling-including-papilloedema.

[14] Freund, Yoav, and Robert E. Schapire (1995). A desicion-theoretic generalization of on-line learning and an application to boosting. *European conference on computational learning theory*, pages 23–37.

[15] Gelar Budiman (2015). Histogram Equalisation. https://www.youtube.com/watch?v=eNBZI-qYhpg.

[16] Gurney, K. (The University of Sheffield) (1997). An Introduction to Neural Networks.

[17] James Garrity (2017). Papilledema - Eye Disorders. http://www.msdmanuals.com/en-gb/professional/eye-disorders/optic-nerve-disorders/papilledema.

[18] Jose Luis Blanco (2013). Iterative Closest Point (ICP) and Other Matching ALgorithms. https://www.mrpt.org/Iterative_Closest_Point_%28ICP%29_and_other_matching_algorithms.

[19] Medicines and Healthcare products Regulatory Agency (1993). COUNCIL DIRECTIVE 93/42/EEC of 14 June 1993 concerning medical devices. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:1993L0042:20071011:EN:PDF.

[20] Medicines and Healthcare products Regulatory Agency (2013). Medical devices: how to comply with the legal requirements. https://www.gov.uk/guidance/medical-devices-how-to-comply-with-the-legal-requirements.

[21] Miller-Keane Encyclopedia and Dictionary of Medicine, Nursing, and Allied Health, Seventh Edition (2003). mydriatic. (n.d.). https://medical-dictionary.thefreedictionary.com/mydriatic.

[22] Mindlin, Alan M and Lamberts, David W and Grandon, Stanley (1975). Slit lamp use. *Journal of the American College of Emergency Physicians*, 4(4):336–339.

[23] Nidek Co Ltd (2017). Non-Mydriatic Auto Fundus Camera AFC-330. https://www.nidek-intl.com/product/ophthaloptom/diagnostic/dia_retina/afc-330.html.

[24] Nvidia Corporation (2017). CUDA. https://developer.nvidia.com/cuda-zone.

[25] OpenCV (2017). Canny Edge Detector. https://docs.opencv.org/3.3.0/da/d22/tutorial_py_canny.html.

[26] Paul Viola, Michael Jones (2001). Rapid Object Detection using a Boosted Cascade of Simple Features. *Proceedings of the 2001 IEEE*, 1.

[27] R. Fisher, S. Perkins, A. Walker, E. Wolfart (2003). Fourier Transform. http://homepages.inf.ed.ac.uk/rbf/HIPR2/fourier.htm.

[28] Rowley, H.A., Baluja, S. and Kanade, T. (1998). Rotation invariant neural network-based face detection. *IEEE Computer Society Conference*, pages 38–44.

[29] Ruye Wang (2016). Histogram Equalisation. http://fourier.eng.hmc.edu/e161/lectures/contrast_transform/node2.html.

[30] TECHNO-VISION program (2004). MESSIDOR database. accessed on 10/12/2017.

[31] The College of Optometrists (2017). Ophthalmoscopes (part 1). https://www.college-optometrists.org/the-college/museum/online-exhibitions/virtual-ophthalmic-instrument-gallery/ophthalmoscopes.html.

[32] Timberlake, George T and Kennedy, Michael (2005). The Direct Ophthalmoscope How it Works and How to Use It.

[33] Ujjwal Karn (2016). An Intuitive Explanation of Convolutional Neural Networks. https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets. accessed on 10/12/2017.

[34] University of New Mexico (2017). Introduction to Fourier Transforms for Image Processing. https://www.cs.unm.edu/~brayer/vision/fourier.html.

[35] Volk (2017). Volk iNview | iPhone Fundus Camera. https://volk.com/index.php/volk-products/ophthalmic-cameras/volk-inview.html.

[36] Youlian Zhu, Cheng Huang (2012). An Adaptive Histogram Equalization Algorithm on the Image Gray Level Mapping. *Physicis Procedia 25*, pages 601–608.

5 Appendix

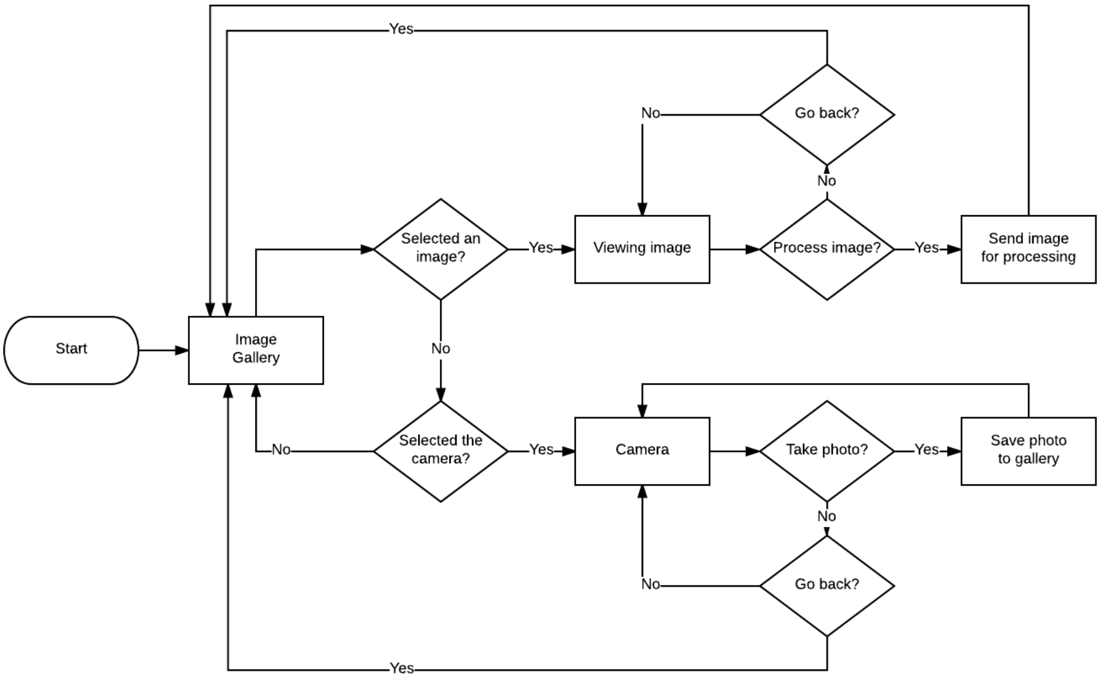

Fig. 9: Flow chart showing how a user could interact with the application

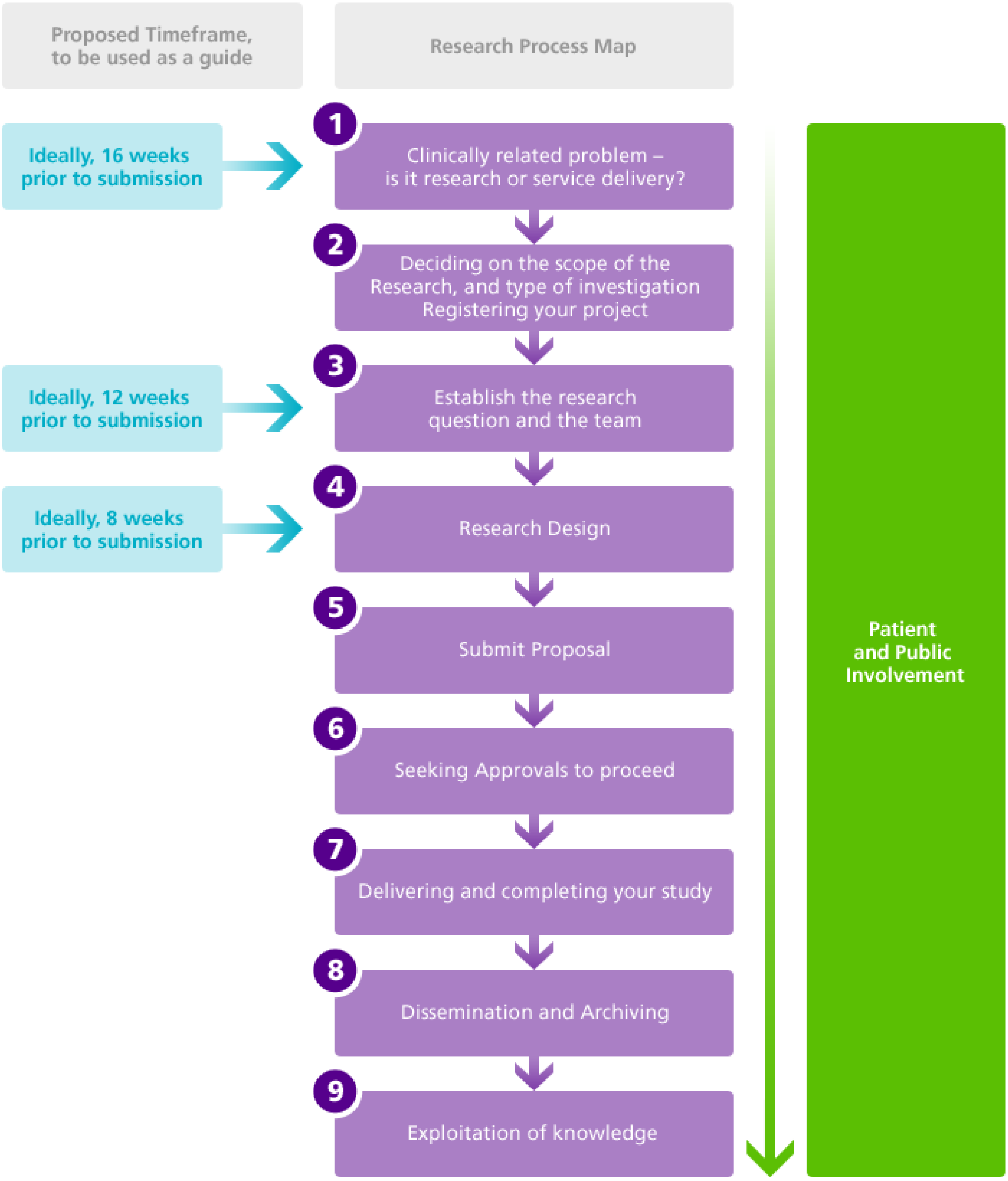

Fig. 10: Flow chart showing the ethics process, taken from the Clinical Research Office [10]

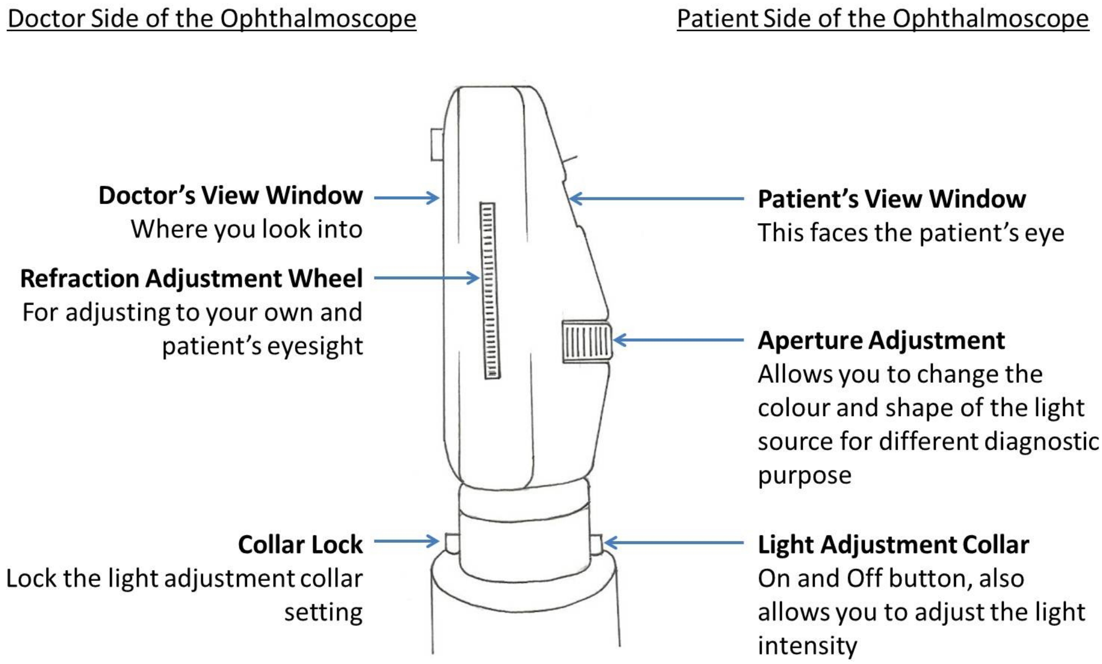

Fig. 11: Image showing how to use an Opthalmascope, taken from [9]

Fig. 12: Non-Mydriatic Fundus Camera, taken from [23]