THE UNIVERSITY *of* EDINBURGH

Assessing neurodegeneration of the retina and brain with ultra-widefield retinal imaging

Emma Pead

Doctor of Philosophy

The University of Edinburgh

2019

Declaration

I hereby declare that this thesis submission has been composed by myself and that where reference is made to the published work of others it is explicitly cited. This work has not been submitted for any other degree or professional qualification.

Emma Pead

Abstract

The eye is embryologically, physiologically and anatomically linked to the brain. Emerging evidence suggests that neurodegenerative diseases, such as Alzheimer's disease (AD), manifest in the retina. Retinal imaging is a quick, non-invasive method to view the retina and its microvasculature. Features such as blood vessel calibre, tortuosity and complexity of the vascular structure (measured through fractal analysis) are thought to reflect microvascular health and have been found to associate with clinical signs of hypertension, diabetes, cardiovascular disease and cognitive decline. Small deposits of acellular debris called drusen in the peripheral retina have also been linked with AD where histological studies show they can contain amyloid beta, a hallmark of AD. Age-related macular degeneration (AMD) is a neurodegenerative disorder of the retina and a leading cause of irreversible vision loss in the ageing population. Increasing number and size of drusen is a characteristic of AMD disease progression. Ultra-widefield (UWF) retinal imaging with a scanning laser ophthalmoscope captures up to 80% of the retina in a single acquisition allowing a larger area of the retina to be assessed for signs of neurodegeneration than is possible with a conventional fundus camera, particularly the periphery. Quantification of changes to the microvasculature and drusen load could be used to derive early biomarkers of diseases that have vascular and neurodegenerative components such as AD and other forms of dementia.

Manually grading drusen in UWF images is a difficult, subjective and a time-consuming process because the area imaged is large (around 700mm^2) and drusen appear as small spots (< 125μm). An automatic approach to detecting drusen would overcome these challenges and facilitate investigations into drusen as a biomarker of neurodegeneration. In this thesis, an automatic system inspired by the recent successes of deep learning in medical image analysis was developed. As drusen are abundant in the retinas of people with AMD, a neural network was trained to classify patches in such a dataset of UWF images. This was compared to the manual gradings of two human observers. There was only a moderate agreement between observers (Kappa = 0.53, Average Dice Similarity Coefficient (DSC) = 0.38), reflecting the challenging and difficult nature of manually grading drusen in UWF images. Performances achieved for the automatic system (assessed using the area under curve (AUC) performance statistic) were 0.55-0.59, 0.62- 0.65 and 0.65-0.66 in the central, perimacular and peripheral regions of the retina, respectively. Highest performance was observed in a subset 8 images where observer agreement was at its highest (DSC > 0.8 and < 0.9), achieving AUC 0.55-0.59, 0.78-0.82 and 0.82-0.85 in the central, perimacular and peripheral zones, respectively.

Measurements of the retinal vasculature appearing in UWF images of cognitively healthy (CH) individuals and patients diagnosed with mild cognitive impairment (MCI) and AD were obtained using a previously established pipeline. Following data cleaning, vascular measures were compared using multivariate generalised estimation equations (GEE), which accounts for the correlation between eyes of

individuals with correction for confounders (e.g. age). The vascular measures were repeated for a subset of images and analysed using GEE to assess the repeatability of the results. When comparing AD with CH, the analysis showed a statistically significant difference between measurements of arterioles in the inferonasal quadrant, but fractal analysis produced inconsistent results due to differences in the area sampled in which the fractal dimension was calculated.

When looking at drusen load, there was a higher abundance of drusen in the inferonasal region of the peripheral retina in the CH and AD compared to the MCI group. Using GEE analysis, there was evidence of a significant difference in drusen count when comparing MCI to CH ($p = 0.02$) and MCI to AD ($p = 0.03$), but no evidence of a difference when comparing AD to CH. However, given the low sensitivity of the system (partly the result of only moderate agreement between human observers), there will be a large proportion of drusen that are not detected giving an under estimation of the true amount of drusen present in an image. Overcoming this limitation will involve training the system using larger datasets and annotations from additional observers to create a more consistent reference standard. Further validation could then be performed in the future to determine if these promising pilot results persist, leading to candidate retinal biomarkers of AD.

Lay Summary

During a routine optometrist appointment to check our eye health, we commonly have a picture taken of the back of our eye, called the retina. The retina can be imaged quickly and easily using a machine called a scanning laser ophthalmoscope. As the retina is connected to the brain, it has been discovered that changes that occur in the brain in relation to small blood vessels and nerves can be seen in the retina, which is more accessible to image. Alzheimer's disease is the most common form of dementia and is a progressive disease, which means it gets worse over time. It causes deterioration of brain function such as memory loss (e.g. forgetting names of family members), problems with speech (e.g. mixing up words), disorientation (e.g. getting lost) and hallucinations (e.g. seeing and hearing things that are not there). The early sign of Alzheimer's, called mild cognitive impairment, has symptoms of memory problems (e.g. forgetting to go to appointments) that is worse than you might expect with normal ageing. It is estimated that 850,000 people are living with dementia in the UK and this is predicted to increase to 2 million people by 2050. This places a heavy burden on healthcare providers as well as the families of patients. My research has looked at how the retina changes with Alzheimer's to investigate what some of these changes are and whether they occur in an earlier stage of the disease. If we know how the retina changes in Alzheimer's, in the future, an early warning could potentially be flagged at a routine optometrist appointment.

In an image of the retina, we can see the small blood vessels that carry blood in and out of the tissue. Sometimes there are small lesions, called drusen, that accumulate in the retina with age. These are more common in Alzheimer's and other sight threatening conditions, such as age-related macular degeneration. Properties of blood vessels that are thought to reflect a person's health can be measured using computerised techniques such as how their widths taper or how they spread out to deliver blood and nutrients to the retina. Drusen are difficult to count by hand because they are small and there are sometimes lots of them. This means that if we were to assess the retina for drusen changes in Alzheimer's, a computerised technique is needed. Deep learning is a type of computerised detection or categorisation and is similar to how humans learn new things through examples. By teaching a deep learning system what drusen looks like and what normal retina looks like we can ask it to detect and count them. In my thesis, I developed such a system and compared its ability to identify drusen against two human observers performing the same task manually. There were limitations to my system. It failed to detect drusen that were very small which would result in an underestimation of drusen counts in my analysis, but it did provide for the first time a tool to automatically detect drusen in scanning laser ophthalmoscope images.

Measurements of the retinal vessels from individuals that were cognitively healthy as well as those diagnosed with mild cognitive impairment and Alzheimer's disease were compared. The analysis pointed to subtle differences in the retinal vessels in Alzheimer's. Furthermore, drusen presence in mild cognitive impairment and

Alzheimer's disease was different. My research contributed to the growing evidence that points to retinal changes in Alzheimer's disease. Further work is now needed with larger studies and more images from individuals that have been captured at various time points in their lives to discover when retinal changes occur in the course of the disease and whether these markers can be used to make an early prediction of disease.

Acknowledgements

Throughout this degree, I have received a great deal of support for which I am immeasurably grateful. Foremost, I would like to thank my principle supervisor Dr Tom MacGillivray whose expertise, encouragement, good-humour and patience has never failed to guide me. He is an outstanding mentor and I thank him for his endless support towards completing this thesis and inspiring me to continue on this career path. Thank you to Prof Baljean Dhillon for his approachable demeanour, clinical insight and continuous support throughout this work. I look forward to working with the both of them in the future.

I would like to thank Optos for not only funding the project, but for providing a warm and welcoming environment for me to participate in constructive discussions and lively events. I would especially like to thank Dr Alan Fleming whose exceptional technical insight has been invaluable to my learning and development. Thanks to Dr Jano van Hemert whose approachable demeanour has always been a source of friendly rapport and has always made me feel part of the Optos team. I also gratefully acknowledge the support of NVIDIA Corporation with the donation of the Quadro P6000 GPU used in this research.

It still is, and has been a pleasure to be part of the VAMPIRE group. I would especially like to thank Prof Emanuel Trucco whose brilliant mind and wonderful character has not only been an invaluable source of knowledge but an inspiration for an aspiring academic.

I would like to acknowledge Lajos Csincsik, Nicola Quinn, Tune Peto and Imre Lengyel at Queen's University for being excellent collaborators during the project and I thank them for hosting me on numerous occasions. I would like to thank Prof Sharon Fekrat and Dr Dilraj Grewal at Duke University for their clinical expertise, warmth and friendship during this collaboration. A special thanks to Dr Atalie Thompson for her statistical prowess, patience and guidance throughout the analysis of the data presented in this thesis.

On a personal note, I would like to thank my mum and dad for their love and encouragement and for always believing in me throughout the years. I especially thank them for nurturing my interest in science from an early age, for without them I would not be here today. I thank Melanie, my sister, who has always been there when I have needed some family comfort and silliness. I would like to thank James for his endless love, support and kindness and for never failing to make me laugh when I have needed it the most. Finally, special thanks to our dog Kodee, for being very helpful during the writing of this thesis by sitting on my lap and constantly pawing me for attention.

Publications and Presentations

E Pead, R Megaw, J Cameron, A Fleming, B Dhillon, E Trucco, T MacGillivray. Automated detection of drusen in fundus imaging: A systematic review. Survey of Ophthalmology. 2019 Jul-Aug;64(4):498-511

E Pead, D Grewal, S P Yoon, B Polascik, A C Thompson, J R Burke, C Dunn, E Trucco, T MacGillivray, S Fekrat (2019) A pilot study using ultra-widefield imaging to study

vascular changes in Mild Cognitive Impairment and Alzheimer's disease. The Association for Research in Vision and Ophthalmology (ARVO), Canada– *Poster*

P K Bhullar, E Pead, A C Thompson, S P Yoon, D Grewal, B Polascik, T MacGillivray, E Trucco, S Fekrat (2019) Evaluation of potential biomarkers in multimodal retinal images for diagnosis of Parkinson's disease: a
pilot study. The Association for Research in Vision and Ophthalmology (ARVO), Canada – *Poster*

E Pead, E Trucco, R Megaw, B Dhillon, A Fleming, T MacGillivray (2018) Drusen detection in ultra-widefield retinal imaging using transfer learning of pre-trained deep neural networks. The transformative potential of data and image analysis for eye care. Royal Society London – *Poster*

E Pead, B Dhillon, A Fleming, T MacGillivray (2017) Assessing neurodegeneration of the retina and brain with ultra-widefield retinal imaging. Eye Development & Degeneration. IGMM Edinburgh - *Poster and Presentation*

Contents

List of abbreviations

A	**A**rteriole
ACC	**Acc**uracy
AD	**A**lzheimer's **D**isease
AI	**A**rtificial **I**ntelligence
AMD	**A**ge **R**elated **M**acular **D**egeneration
APD	**A**valanche **P**hoto**d**iode
APOE	**A**polipoprotein **E**
APP	**A**myloid **P**recursor **P**rotein
AREDS	**A**ge-**R**elated **E**ye **D**isease **S**tudy
ARIA	**A**utomated **R**etinal **I**mage **A**nalyser
AUC	**A**rea **U**nder **C**urve
AVR	**A**rteriole **V**enule **R**atio
Aβ	**β** –**a**myloid
CBR	**C**ase **B**ased **R**easoning
CFAR	**C**onstant **F**alse **A**larm **R**ate
CH	**C**ognitively **H**ealthy
CIRCLE	**C**ologne **I**mage **R**eading **C**entre and **L**aboratory
CLAHE	**C**ontrast **L**imited **A**daptive **H**istogram **E**qualisation
CNN	**C**onvolutional **N**eural **N**etwork
CNS	**C**entral **N**ervous **S**ystem
CRAE	**C**entral **R**etinal **A**rteriolar **E**quivalent

CRVE	**C**entral **R**etinal **V**enular **E**quivalent
CSF	**C**erebrospinal **F**luid
DD	**D**rusen **D**etector
DSC	**D**ice **S**imilarity **C**oefficient
E/MIPC	**E**qual/**M**aximum **n**umber of **i**mages **p**er **c**lass
FC	**F**ully **C**onnected
FD	**F**ractal **D**imension
FDa	**A**rteriole **F**ractal **D**imension
FDv	**V**enule **F**ractal **D**imension
fMRI	**F**unctional **M**agnetic **R**esonance **I**maging
FN	**F**alse **N**egative
FOV	**F**ield **o**f **v**iew
FP	**F**alse **P**ositive
GA	**G**eographic **A**trophy
GEE	**G**eneralised **E**stimation **E**quation
HSI	**H**ue **S**aturation **I**ntensity
I-data	**I**nitial **data**
ICC	**I**ntraclass **c**orrelation **c**oefficient
IN	**I**nfero**n**asal
ISLRVC	**I**mage**N**et **L**arge **S**cale **V**isual **R**ecognition **C**hallenge
IT	**I**nfero**t**emporal
KMC	**K**arstuba **M**edical **C**ollege
KNN	**K**-**n**earest **n**eighbours
MCI	**M**ild **C**ognitive **I**mpairment
MGEE	**M**ultivariate **G**eneralised **E**stimation **E**quation

MMSE	**M**ini **M**ental **S**tate **E**xam
MoCA	**M**ontreal **Co**gnitive **A**ssessment
MRI	**M**agnetic **R**esonance **I**maging
MS	**M**anually **S**elected
NFT	**N**euro**f**ibrillary **T**angles
NPV	**N**egative **P**redicative **V**alue
Obs1	**Obs**erver **1**
Obs2	**Obs**erver **2**
OCT	**O**ptical **C**oherence **T**omography
OCTA	**O**ptical **C**oherence **T**omography **A**ngiography
OD	**O**ptic **D**isc
PET	**P**ositron **E**mission **T**omography
PHOG	**P**yramid **H**istogram of **O**riented **G**radients
PPV	**P**ositive **P**redictive **V**alue
R-data	**R**epeat **data**
RGB	**R**ed **G**reen **B**lue
RGC	**R**etinal **G**anglion **C**ells
RNFL	**R**etinal **N**erve **F**ibre **L**ayer
ROC	**R**eceiver **O**perator **C**haracteristic
ROI	**R**egion **o**f **i**nterest
RPE	**R**etinal **P**igment **E**pithelium
SD	**S**tandard **d**eviation
SEN	**Sen**sitivity
SIFT	**S**cale **I**nvariant **F**eature **T**ransform
SIVA	**S**ingapore **'I' V**essel **A**ssessment

SLO	**S**canning **L**aser **O**phthalmoscope
SN	**S**upero**n**asal
SPEC	**Spec**ificity
ST	**S**upero**t**emporal
STARE	**S**tructured **A**nalysis of the **RE**tina
SURF	**S**peeded **U**p **R**obust **F**eatures
SVM	**S**upport **v**ector **m**achine
TN	**T**rue **N**egative
TortA	**A**rteriolar **Tort**uosity
TortV	**V**enular **T**ortuosity
TP	**T**rue **P**ositive
UGEE	**U**nivariate **G**eneralised **E**stimation **E**quation
UObs	**U**nion of **Obs**ervers
UWF	**U**ltra-**w**ide**f**ield
V	**V**enule
VAMPIRE	**V**ascular **A**ssessment and **M**easurement **P**latform for **I**mages of the **Re**tina
WG	Width Gradient
WG(**a**/**v**) **IN**	**I**nfero**n**asal (**a**rteriole/**v**enule) **W**idth **G**radient
WG(**a**/**v**) **IT**	**I**nfero**t**emporal (**a**rteriole/**v**enule) **W**idth **G**radient
WG(**a**/**v**) **SN**	**S**upero**n**asal (**a**rteriole/**v**enule) **W**idth **G**radient
WG(**a**/**v**) **ST**	**S**upero**t**emporal (**a**rteriole/**v**enule) **W**idth **G**radient
WGa	**A**rteriole **W**idth **G**radient
WGv	**V**enule **W**idth **G**radient
WRS	**W**ilcoxon **R**ank **S**um

Chapter 1 Introduction

1.1 Background and motivation

Over 100 years ago Alois Alzheimer presented a lecture on a patient with “a peculiar disorder of the cerebral cortex” [1]. She had suffered from dementia and her brain was discovered to have protein plaques, neurofibrillary tangles and atherosclerotic changes that have now become defining features of Alzheimer’s disease (AD) [2]. The study of the molecular composition and mechanism of protein plaques and neurofibrillary tangles have since been identified as protein aggregate amyloid β (Aβ) and microtubule-associated protein tau [3][4][5]. These are now considered the hallmarks of AD. Atherosclerotic changes have also been defined as risk factors for AD, the most common form of dementia [6].

The aforementioned processes are not exclusive to dementia and there are parallels to other neurodegenerative diseases. Age-related macular degeneration (AMD) is a late-onset neurodegenerative disorder of the retina where progression to later stages results in irreversible vision loss. AMD and AD share many clinical and pathological features such as Aβ deposition in drusen and risk factors such as age and smoking. As the retina is anatomically, physiologically and embryologically linked to the brain there has been an increasing interest in utilising the retina to investigate diseases of the brain and central nervous system (CNS) [7]. The retina is in fact the only area of the body where microvasculature can be imaged non-invasively and has

been utilised to derive biomarkers of systemic conditions such as diabetes, cardio vasculature disease and renal diseases [7][8].

Imaging the retina is simple, non-invasive and inexpensive compared to magnetic resonance imaging (MRI) and positron emission tomography (PET), often used to help diagnose and research AD. With the advent of ultra-widefield (UWF) imaging more of the retina can be captured in a single image, including views of the periphery. This has created the opportunity to conduct a more extensive assessment of the retina for potential biomarkers of neurodegeneration.

Currently there are few treatments and no cure for AMD or AD. Early diagnosis would improve patient care and aide in the development of preventative therapies that could prevent or slow down devastating vision and memory loss in the elderly. In an ageing population, AMD and AD become increasingly prevalent, placing pressure on healthcare providers by depleting resources and increasing financial strain. The burden is also increased for families of the patients who are affected. Non-invasive retinal imaging could help alleviate such pressures by providing an early diagnostic strategy and progression monitoring that can be achieved readily and easily.

1.2 Thesis aims

My study contributes to the need to investigate, characterise and assess changes in the retina in neurodegeneration. Automatic segmentation and analysis of the retinal

vasculature has previously been developed in UWF, however, automatic drusen detection has yet to be established. In this thesis I:

1) Investigate methods for automatically detecting and analysing drusen in UWF retinal images.
2) Implement and validate a computerised drusen detection system in UWF utilising images from people with AMD.
3) Apply existing retinal vasculature analysis methods to a new cohort featuring patients with AD and Mild Cognitive Impairment (MCI) as well as cognitively healthy subjects.
4) Apply novel drusen analysis to the above cohort to quantify drusen and characterise any differences in drusen load and location between the groups.

1.3 Thesis outline

In Chapter 2, I will consider the anatomy of the retina and its link to the brain. I will describe how the brain and retina are affected by neurodegeneration. Next, I will describe UWF imaging, how drusen may be a potential biomarker of AD in the retina with particular attention to the similarities between AMD and AD and highlight the need for drusen detection. In Chapter 3, I review the literature in the field of computerised drusen detection in order to inspire the direction of development for my system that I present in Chapter 4. Analysis of the patient cohort is reported in Chapter 5, including cohort characteristics and the

methodology for measuring the retinal vasculature in UWF images. In Chapter 6, I apply my drusen detection system to the patient cohort and present an exploratory analysis of drusen load and location. In Chapter 7, I discuss the conclusions of the thesis and future work.

Chapter 2 The retina as the window to the brain

2.1 The retina and brain

The human eye is a unique and complex structure. Thought to have evolved around 550 million years ago, it is the fastest muscle in the human body with 2 million working parts that allows us to see 500 shades of grey and 2.7 million colours. The eye is broadly comprised of three distinguishable layers (Figure 2.1). The inner most layer is composed of a complex structure of neurons, called the retina and is responsible for our ability to see. The outer layers consist of the sclera and cornea. The sclera protects and maintains shape of the eye. The cornea refracts and focuses light to the lens and retina, also protecting the eye from infection. The middle layer comprises the iris that controls the amount of light entering the eye and the ciliary body that provides aqueous production. Also, part of the middle layer, the choroid is the vascular network that provides nutrients to outer layers of the retina [9].

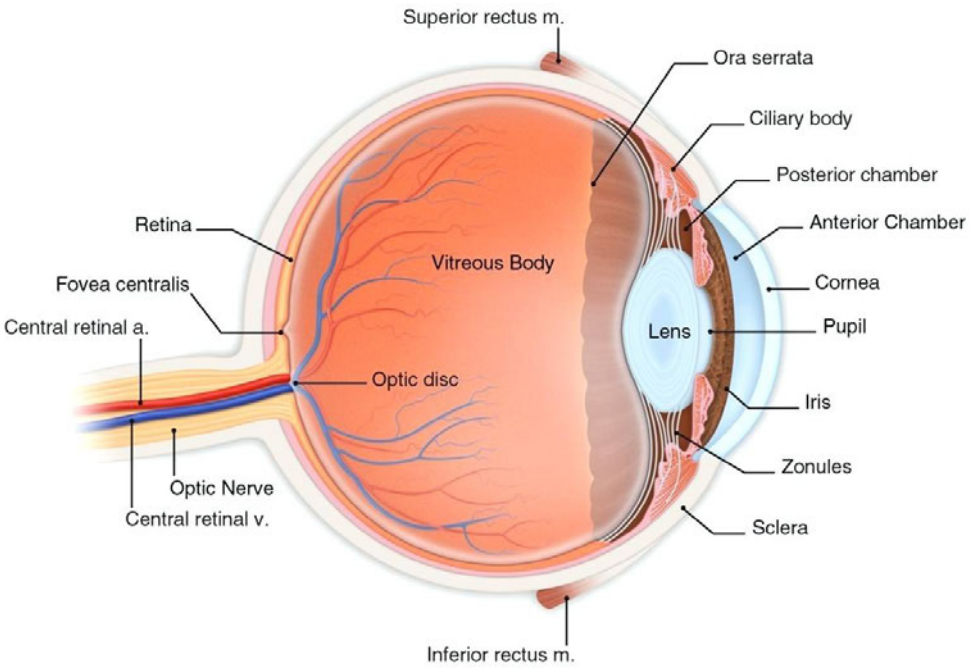

Figure 2.1 Anatomy of the retina [10]. Here we can see the three main layers including the retina.

Light enters the eye through the cornea and is directed to the lens by the iris. The lens then focuses the light to the retina. The retina is composed of highly specialised cells called cones and rods that are responsible for allowing us to see colour (Figure 2.2). The light excites these cells converting the light signal to a neural impulse. This signal is passed through bipolar cells to the retinal ganglion cells (RGC) and eventually to the retinal nerve fibre layer (RNFL). The RNFL increases in thickness towards the optic disc (OD) that is connected to the optic nerve. The electrical signal then travels to the brain though the OD and along the optic nerve (Figure 2.3).

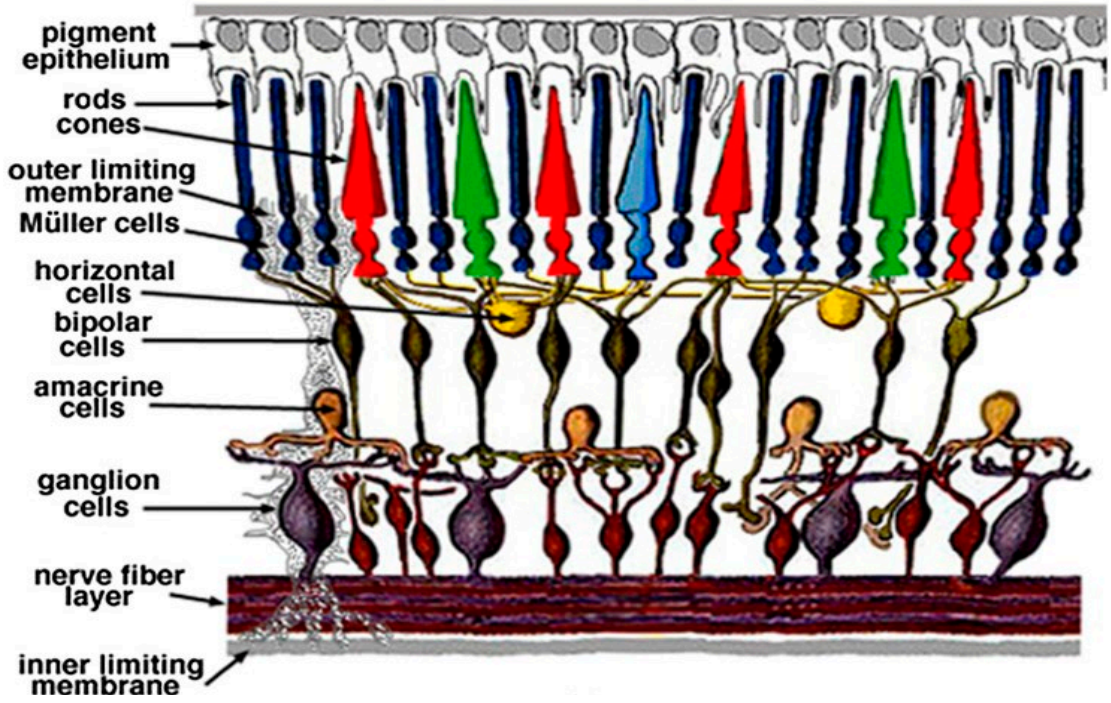

Figure 2.2 The cellular layers of the retina. The RNFL is closest to the surface of the retina [7].

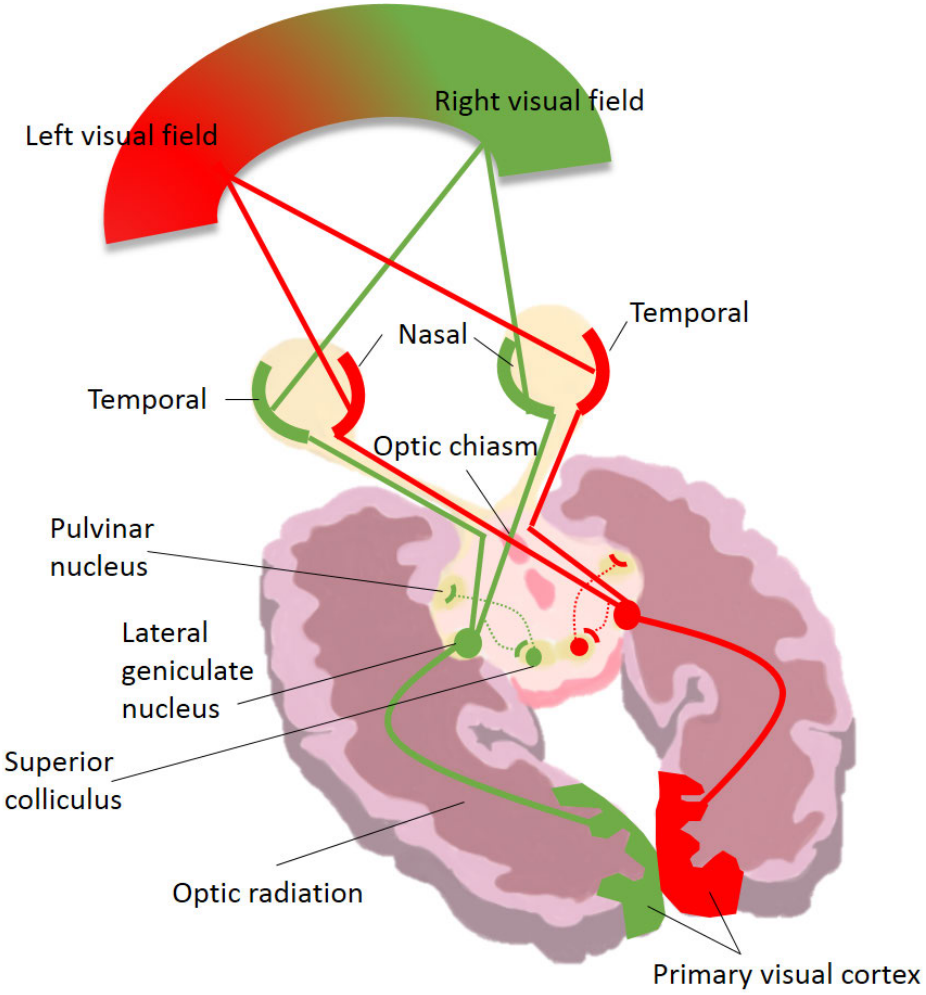

Figure 2.3 Diagram of the eye-brain connection. The lens in our eye is much smaller than most of the objects we see. Light always travels in a straight line, so that when light is focused through the lens the image is upside down. The optic nerve goes through the optic chiasma to the primary visual cortex, or occipital cortex, at the back of the brain. This crossing of nerves allows us to register the image as right side up rather than upside down.

During embryological development the retina is formed from an outpocketing of the primordium of the brain and spinal cord, called the optic vesicle. The optic vesicle then invaginates to form the optic cup, where its inner wall becomes the neural retina and the outer wall, the retinal pigment epithelium (RPE). The retina connects to the brain through the optic nerve to the visual cortex in the brain. The retina is therefore a direct extension of the CNS.

2.1.1 The progression of Alzheimer's disease

Neurodegeneration refers to any pathological process that affects neurons that degenerate by losing structure or function and can occur anywhere in the body. AD is neurodegenerative disease of the brain and the leading cause of dementia [11]. In 2014, it was estimated that there were 850,000 people in the UK were living with dementia. This is projected to rise to 2 million people by 2050 [12]. Converging evidence from genetic at-risk cohorts and healthy ageing subjects have indicated that the pathogenesis of AD begins decades before diagnosis of clinical dementia [13]. Progression of AD is often thought to have three main stages: (1) presymptomatic, (2) prodromal and (3) symptomatic [14][15].

Presymptomatic stages of AD involve abnormal accumulation of Aβ due to abnormal metabolism of a large transmembrane protein called the amyloid precursor protein (APP) [16]. The exact physiological function of APP is not known but it has been shown to aide in neuronal growth and repair and is highly expressed and metabolised

in the brain [17]. Figure 2.4 shows the APP protein spanning a cell membrane. The left-hand side (blue) of Figure 2.4 is the normal pathway of APP processing while the right is the abnormal. Normal APP metabolism starts by α-secretase enzyme cleaving APP into a soluble protein that does not generate Aβ peptide [18]. There are genetic mutations that have been associated with the alteration of the APP cleavage site such as a mutation in the apolipoprotein E (APOE4) gene that has been shown to inhibit α-secretase and increase preferential cleavage by β-secretase (Figure 2.4A) [19]. Variation of the APOE gene (i.e. homozygous or heterozygous) determines the risk of developing sporadic (late onset) AD that accounts for 90-95% of AD cases [20]. PSEN-1 (chromosome 14) or PSEN-2 (chromosome 1) that encode presenlin 1 and presenlin 2 and are subunits of γ-secretase have been shown to increase APP cleavage by γ-secretase (Figure 2.4B) [21]. PSEN-1 and PSEN-2 mutations are linked to familial (early onset) AD that accounts for 5-10% of AD cases [22]. Both APOE and PSEN mutations change the cleavage site of APP producing different length Aβ peptides. The incorrectly cleaved APP form Aβ oligomers that aggregate extracellularly and form diffuse or dense Aβ plaques (so-called amyloid plaques) between neuronal cells which can disrupt synaptic signalling and impair brain function (amyloidosis) [23][24]. Amyloid plaques and synaptic dysfunction can be identified in cerebrospinal fluid (CSF), PET imaging and functional MRI (fMRI) in presymtomatic stages of AD [25].

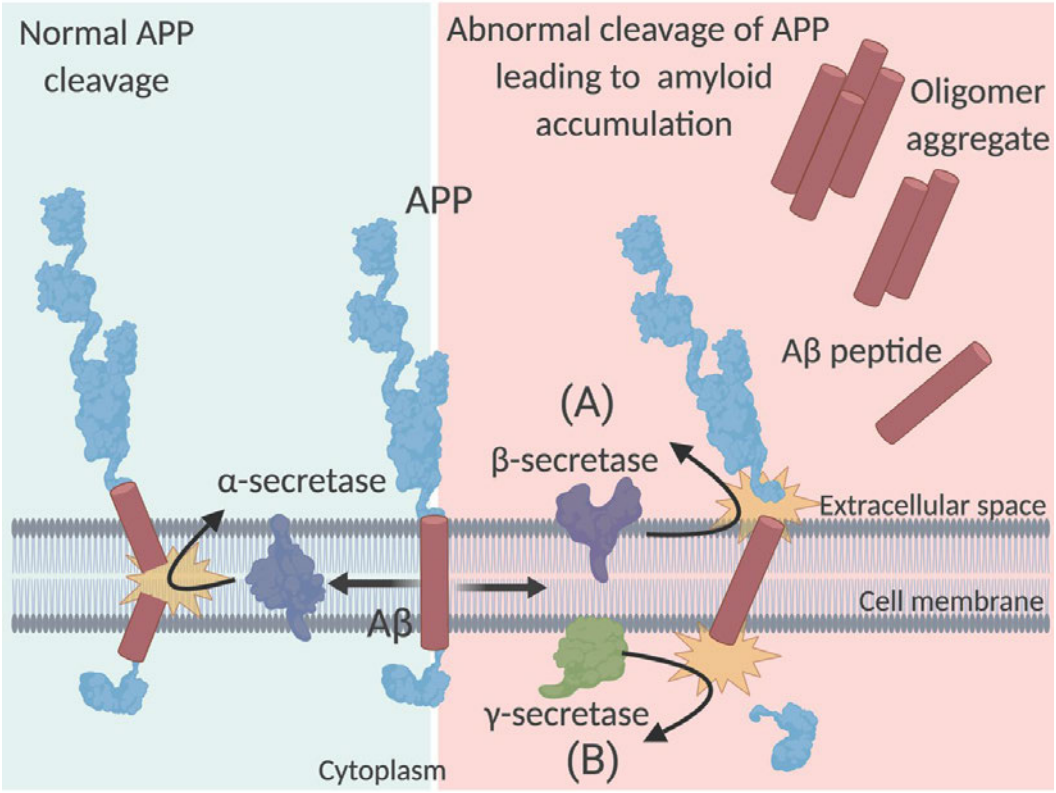

Figure 2.4 APP is a transmembrane protein and when cleaved at the Aβ domain by α-secretase results in a soluble protein that has been shown to have neuroprotective properties [17]. **A)** Mutations in the APP (APOE4) have shown to inhibit α-secretase cleavage and preferential cleavage by β-secretase (Sporadic AD))[19]. **B)** PRESEN1/PRESEN2 mutations (Familial AD) increase γ-secretase enzyme [21]. Both **(A)** and **(B)** lead to excess Aβ peptide that accumulates in the extracellular space to form insoluble Aβ oligomers which can aggregate. Overtime the oxidative stress and consequential biochemical changes leads to neuronal cell death and the hallmark of AD, amyloid plaques [24]. Figure created at biorender.com.

A neuronal cell contains a cytoskeleton comprised of microtubules stabilised by the protein tau [26]. It is thought that the accumulation of extracellular Aβ oligomers activates pathways responsible for phosphorylation of tau, which then loses its ability to stabilise microtubules and promotes self-aggregation within the cell, called neurofibrillary tangles (NFT) [27]. Neurons with non-functioning microtubules cannot signal optimally which leads to apoptosis and atrophy of the brain, narrowing gyri (the folds of the brain), widening sulci (spaces between the folds) and enlarging of ventricles (cavities filled with CSF) [28]. Figure 2.5 shows a comparison of a healthy brain and neurons (*left*) and a brain with AD and neurodegeneration (*right*). AD starts from the hippocampus, the region of the brain responsible for making new memories, and progresses centrifugally [29]. Tau can be identified in CSF, MRI and blood in presymptomatic and prodromal stages of AD [30]. Prodromal AD is typically observed as cerebral amyloidosis with or without evidence of neurodegeneration and subtle cognitive decline mainly with episodic memory disturbances [14]. Cognitive testing in the tail end of prodromal stage of AD, such as mini mental state exam (MMSE) [31] or Montreal cognitive assessment (MOCA) [32], are often conducted to determine cognitive decline.

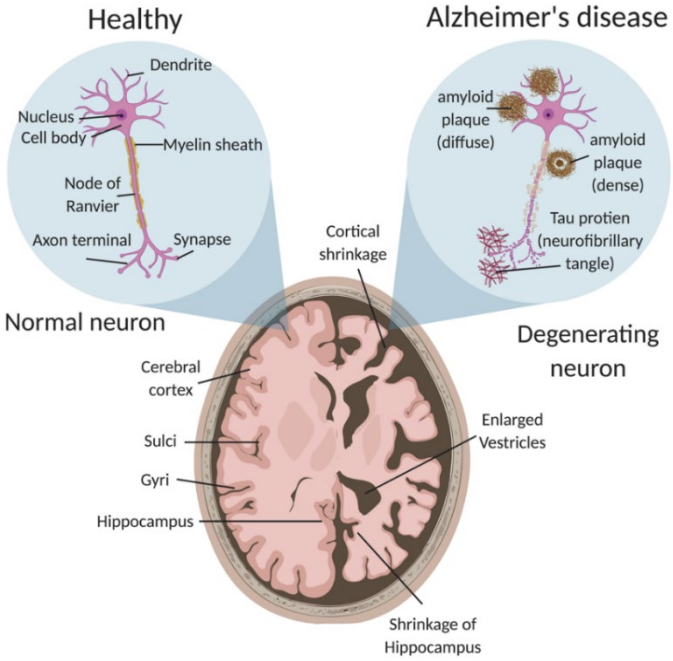

Figure 2.5 Summary of brain biomarkers of neurodegeneration in AD. On the left is a healthy brain and a healthy neuronal cell. On the right is a brain with atrophy, cortical and hippocampal shrinkage. Left shows a healthy neuronal cell which is responsible for communication and integration in the nervous system. The axon is the output of the neuron which branch to connect to other neurons. The dendrites are the input of the neuron and are covered in synapses that connect to other neurons. The axon is covered in a myelin sheath with periodic gaps (node of Ranvier) that serves as a rapid conductor of nerve impulses. On the right is a degenerating neuronal cell in an AD brain. Amyloid plaques and tau proteins impair neuronal cell input and output that block the dendrites and synapses. It is thought that an inflammation response causes apoptosis resulting in global atrophy in the brain [28]. Figure created at biorender.com.

Prodromal to symptomatic AD diagnosis usually occurs when a relative notices a cognitive problem, typically a memory problem, in a family member leading to that person seeking primary care such as from their GP and subsequent referral to secondary care which is often a memory clinic [33]. Early symptomatic stages of AD manifest as loss in short term memory (e.g. a person may not remember what they had for breakfast) and progresses to loss of motor skills (e.g. eating) and language (e.g. word recall) [34]. Later stages of symptomatic AD manifest as loss of long term memory (e.g. name of spouse), increasing disorientation (e.g. getting lost) and eventually individuals becoming bedridden [34]. The final symptomatic stage of AD can be difficult to diagnose from other dementias where cognitive testing is often used in the clinic [35]. MCI in some cases can be considered as the transitional stage of age-associated cognitive decline to AD [36]. Individuals are considered to have MCI if there is evidence that their cognitive function has declined but can still function independently in their daily lives. Such evidence is usually reported by a family member (e.g. a person used to be good at remembering their shopping list in the past but presently requires assistance to remember what to buy) [37].

Figure 2.6 shows the hypothetical model of the succession of AD biomarkers in the brain [25]. Pre-clinical AD is hypothesised to occur as a succession of asymptomatic Aβ accumulation which can be identified in CSF and PET [38]. Amyloid plaques accumulate throughout the parenchyma as diffuse fibrils or more dense plaques later leading to synaptic dysfunction observed in prodromal AD and is identified using fMRI

[28]. Tau-mediated neuronal death is closely followed by atrophy of the brain, imaged using MRI [39]. This leads to symptoms of MCI and eventually dementia that is identified using cognitive testing. The risk of developing each disease stage increases with age. The preclinical period of succession of biomarkers therefore provides a window for preventative therapies or risk factor reduction [40].

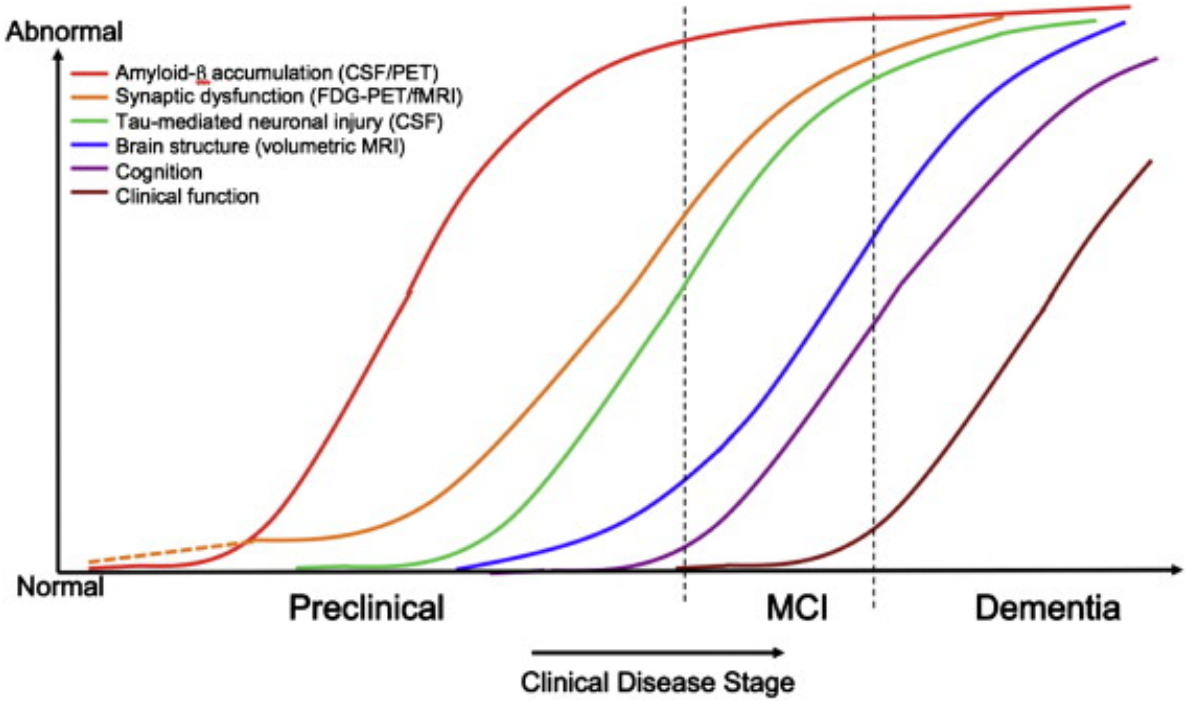

Figure 2.6 A hypothetical model and of the succession of AD biomarkers proposed in [25]. The dashed line indicates carriers of APOE4 allele may have detectable tau deposition before Aβ deposition [40]. Figure from [25].

This raises the key question as to whether retinal changes occur prior to some of the aforementioned disease processes and would therefore translate to potential early

biomarkers or whether they are secondary manifestations of disease that might translate to biomarkers of AD progression. Advances in treatment's for AD is largely symptom driven such as anti-Aβ antibodies for treating mild to moderate AD [41][42] which showed no improvement of clinical outcomes but a sub analysis of mild AD found a slowed cognitive deficit [43]. This highlights the importance of early treatment and how early diagnosis would identify patients in the initial AD stages that could benefit from such interventions. Assessment and characterisation of the retinal changes in AD and MCI compared to cognitively healthy (CH) individuals may therefore help to identify when neuro-retinal degeneration occurs and what some of the measurable signs might be.

2.1.2 Neuro-retinal degeneration

As the retina is anatomically, physiologically and embryologically linked to the CNS it has been suggested that neurodegenerative diseases, such as AD, Parkinson's disease and Huntington disease [44][45], may have pathology in RGCs (where APP is also synthesised) and the optic nerve in the retina [46]. Changes in the retina have also been observed in stroke [47][48], Parkinson's disease [49][50] and Multiple sclerosis [51]. At a molecular level, Aβ and tau have been found within the retina and optic nerve [52][53]. There have been increasing reports of Aβ deposition in the retina [54][55][56] and this has been shown to trigger the breakdown of the RPE [57] and vascular integrity of the retina [58]. Tau deposition reduces axonal transport in retinal neurons that causes a toxic affect to RGC's [59], ultimately leading to RGC apoptosis and thinning of the RNFL [60][61][62][63] and optic nerve abnormalities [64].

Furthermore, previous studies using animal models have reported a correlation between the deposition of senile plaques in the retina (composed of Aβ and Tau) and senile plaque load in the brain [65]. Retinal senile plaques in mouse models have also been detected prior to plaque deposition in the brain, suggesting Aβ deposition in the retina may be an early sign of AD [52]. In histological studies in humans, Aβ deposition have been found in extracellular deposits in the retina, called drusen [55].

The development of the retinal microvasculature is a highly coordinated process that involves internal cell-cell signalling and external stimuli such as oxygen and nutrients to optimise blood flow in the retina [66][67]. Therefore, deviations from this optimum state can lead to damage within the retina. There have been reports of associations between vascular changes and the neurodegenerative process of AD [68] such as Aβ deposition in the retinal vasculature that has been shown to increase stress on the blood vessel walls leading to morphological and functional changes [69].

Neurodegenerative disorders also have overlaps with diseases unique to the eye, such as axonal atrophy observed in glaucoma and Parkinson's disease [70][71] and drusen in AMD and AD [55]. Given that there are shared features between neurodegenerative disorders and other ocular pathologies, precise characterisation of retinal changes in each condition would be required to distinguish early AD biomarkers from sight threatening conditions. The retina can be assessed through ophthalmic examination using retinal imaging, which is a quick, non-invasive procedure and causes hardly any discomfort to the patient. Retinal changes may be extremely subtle and so precise measurements are needed in order to investigate

associations with AD. In this thesis, drusen deposition (that may indicate Aβ deposition) and changes in the microvasculature are assessed using UWF retinal imaging in order to investigate retinal changes that might occur in AD.

2.2 Ultra-widefield retinal imaging to study AD

2.2.1 UWF retinal imaging

Retinal imaging has evolved over the past 40 years as a technology. In parallel, our understanding of many diseases has increased. The oldest and most widely used retinal imaging modality is fundus photography, which has been used to view the macular, optic disc, retinal abnormalities occurring close to the surface of the retina (e.g. geographic atrophy (GA), haemorrhages), the retinal vasculature, RPE and lesions (such as drusen and diabetic retinopathy lesions). Fundus photography typically provides a 35-50° field of view (FOV) of the retina using white light to illuminate the scene and capture a colour image (with a typical resolution of about 7μm). Optimum results are achieved through induced dilation of the pupil where more light can enter the eye to illuminate the retina. Multiple fundus images can be captured by various gaze positions and then stitched together to achieve a larger FOV. However, montaged images often have small misalignments that could lead to erroneous measurements plus a patient can become fatigued from multiple flashes from the camera system. Figure 2.7 shows fundus photographs and the optical system used to acquire the image.

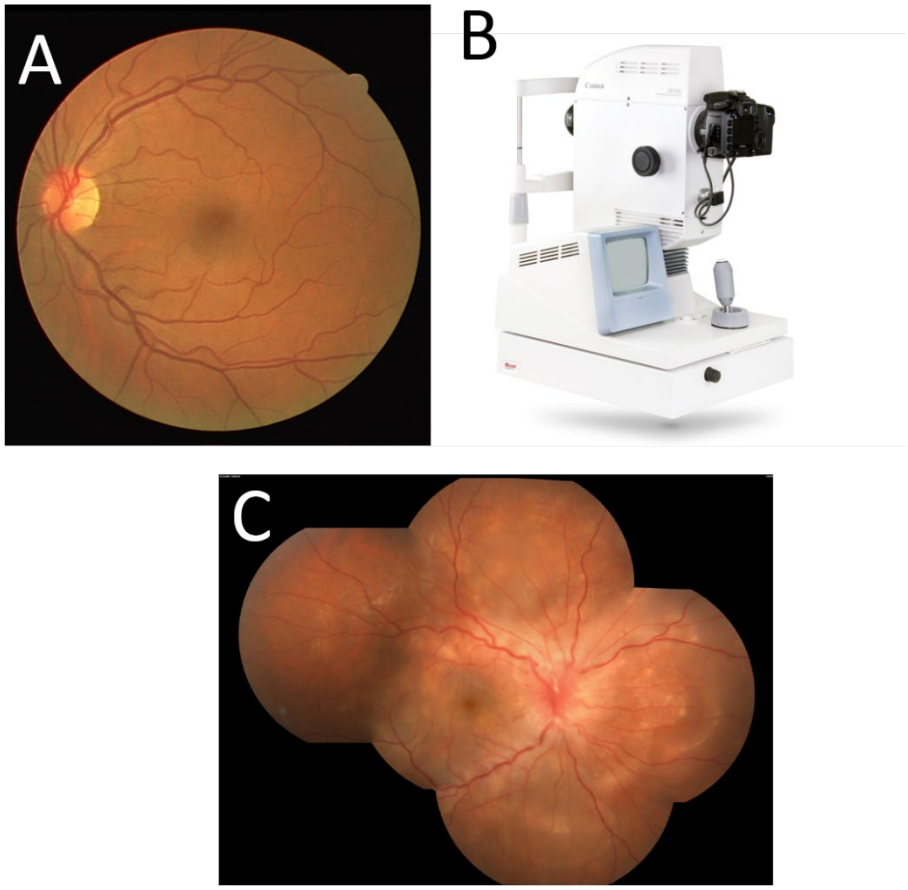

Figure 2.7 Example fundus photographs that can be captured using a fundus camera. **A)** A fundus photograph has a 35-50° FOV. This photograph is macula centered with the OD to the left. **B)** Example of a fundus camera. This camera is a Canon CR-DGi and has a 45° FOV. **C)** Example of a montage of fundus photographs where 5 fundus photographs have been acquired at different gazes and stitched together. Montage image from [72].

Scanning laser ophthalmoscopy (SLO) is a method where a laser scans across the fundus illuminating the retina point-by-point. Combining SLO technology with a

series of lenses and mirrors, a UWF image can be acquired. The retinal imaging modality used in this thesis is the Optos UWF-SLO and can capture 200° of the retina (approximately 80%). Figure 2.8 shows an Optos UWF-SLO instrument and an example UWF image. Optos SLO is recognised for delivering fast, high resolution imaging in terms of sharpness and contrast without the need for pupil dilation and is less susceptible than fundus cameras to any media opacities such as cataracts, amounting to real advantages when imaging an elderly population [73]. The UWF-SLO camera uses red (633nm) and green (532nm) lasers to scan across the retina where the reflected light is collected by lenses and mirrors onto a photodiode and returned to its original red and green components creating a digital image (approximately 20µm resolution).

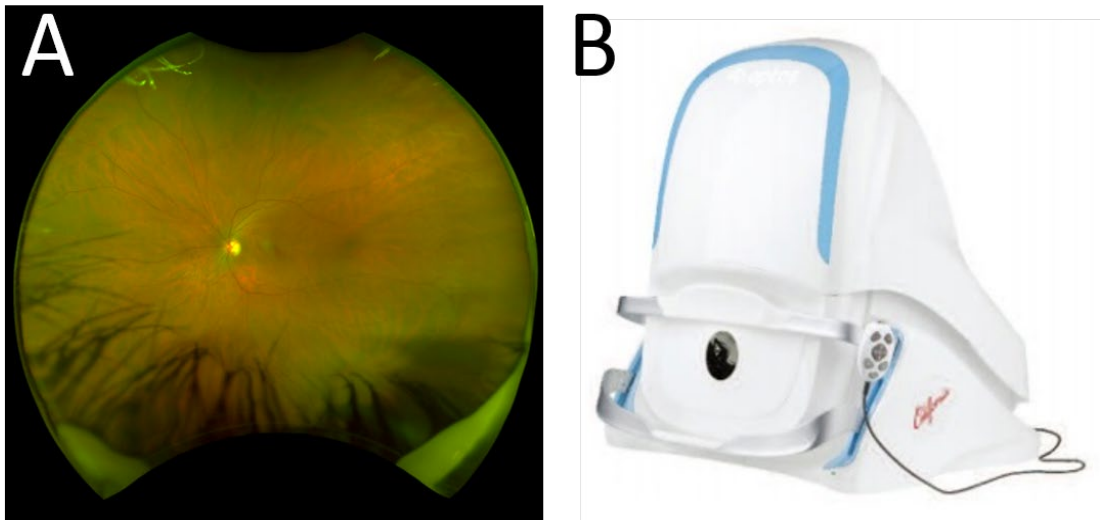

Figure 2.8 Example of an Optos UWF image and an Optos UWF-SLO instrument. **A)** An Optos UWF image (called an optomap) that is used in this thesis. The image consists of a red channel that is obtained by the red laser (633nm) and a green channel that is obtained by the green laser (532nm). In this thesis the images acquired do not have a blue channel but the camera system contains a blue laser (488nm) that is used for fluorescein angiography procedures and an infrared laser (802nm) used for indocyanine green angiography procedures. **B)** An Optos UWF-SLO California model that has a 200° FOV. Earlier models include the P200Tx and Daytona.

Given the large FOV of a UWF instrument, the curvature of the retina must also be taken into account when analysing the image to measure features. Compensating for distortions, due to the retinal curvature, is achieved through stereographic projection [74]. The image is mapped computationally to a standard model of a human eye using

the position of the OD and the fovea as reference points. This image is then flattened, much like projecting a map from a globe, to enable more accurate measurements of distance. This process also allows measurements on the image to be converted from pixels to millimetre equivalents [74]. Figure 2.9 shows an example of a stereographically projected UWF image and the anatomical landmarks.

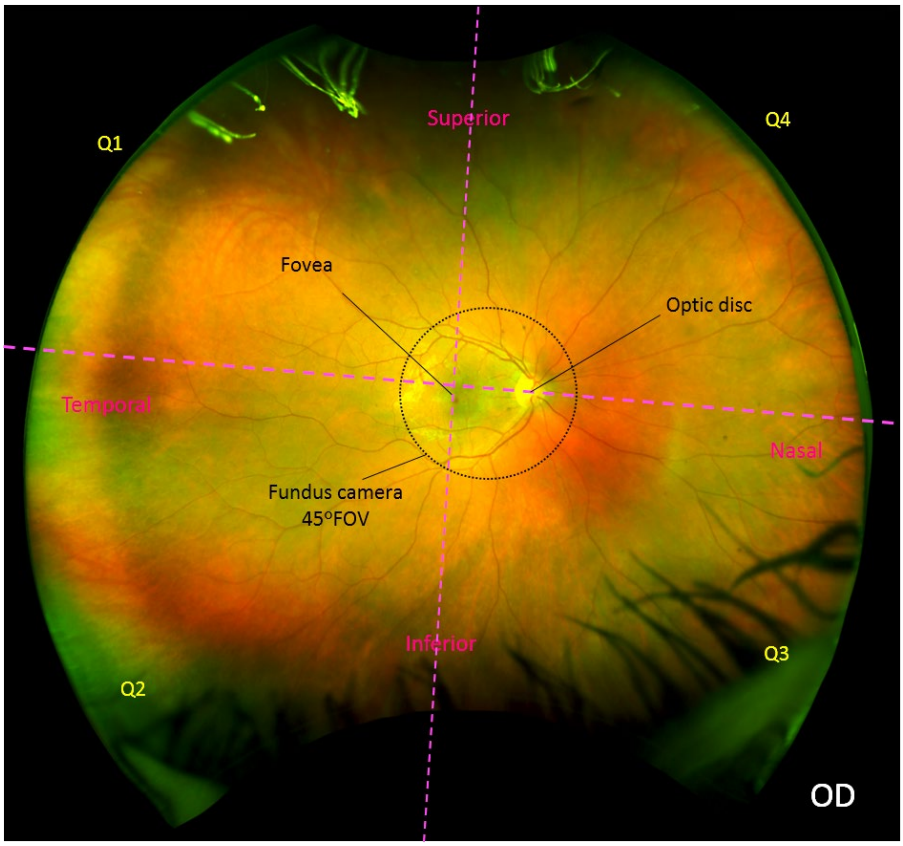

Figure 2.9 Projected UWF image of a healthy individual. The OD is the bright centre of the image and the fovea can be seen as a dark spot. Images are acquired centred on the fovea. In the image we can see lashes in the inferior nasal quadrant (Q3). Lashes also appear in superior temporal quadrant (Q1), as they create bright refection's. In this thesis, I consider any area outside a typical fundus camera FOV (marked here as the dotted black circle) to be the peripheral retina.

UWF imaging allows the periphery of the retina to be investigated for changes occurring with disease. There are anatomical differences between the central and peripheral retina, such as more cone photoreceptors in the former and more rods in the latter[75]. Manifestations of disease in the retina may vary between central and peripheral zones. Imaging a large area of the retina could prove to be beneficial for increasing our understanding of disease mechanisms or deriving biomarkers of eye and brain conditions. Indeed, UWF has utility for monitoring peripheral post-operative tears from cataract surgery that may result in retinal detachment [76]. UWF has also been used to derive additional retinal abnormalities in diabetic retinopathy. The outcome was that diabetic retinopathy severity was graded more severe in UWF images than in the corresponding fundus photograph and therefore the larger FOV provided more information about the extent of the disease and resulted in a more accurate diabetic retinopathy grading [77]. This suggests that there is additional information in the periphery that could have clinical utility that is not accessible with a conventional fundus camera. UWF imaging therefore provides an opportunity to assess a larger area for potential biomarkers of AD.

2.2.2 Drusen as a potential biomarker of AD

Drusen are small deposits of acellular polymorphous debris that gathers between Bruch's membrane and the RPE, which is shown in Figure 2.10, and are considered the hallmark of AMD. AMD is a neurodegenerative disorder of the retina and is one

of the leading causes of vision loss in the elderly, costing the UK economy £155million in 2011 [78]. Reporting to have affected 6 million people worldwide in 2014 [79], AMD is projected to increase to 288 million globally by 2040 [79]. Early AMD is observed as presence of (asymptomatic) macular drusen, often found incidentally through ophthalmic examination, such as from routine optometric examination. Although presence of drusen alone is not diagnostic of AMD, as drusen occur in normal ageing, increasing number and size increases the risk of progression to symptomatic AMD [80]. Late AMD presents as pigmentary changes of the RPE prior to the development of GA (Dry AMD). As neurodegeneration of the RPE progresses, abnormal blood vessels grow (neovascularisation) that are fragile and can leak fluid and blood. This results in scarring of the macula and exudative abnormalities (Wet AMD) and ultimately leads to irreversible vision loss. However, the exact pathogenesis of drusen progression in AMD is incompletely understood and unpredictable [81][82][83].

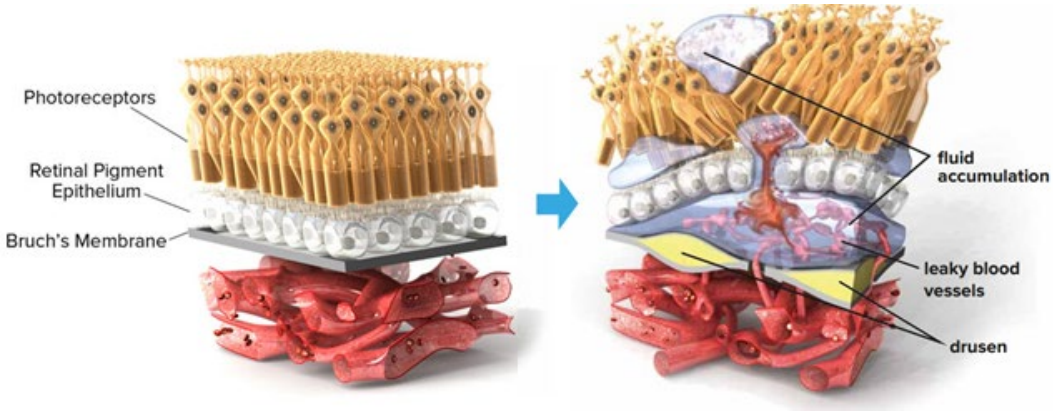

Figure 2.10 AMD pathogenesis in respect to the layers of the retina. Drusen accumulates between the RPE and Bruch's membrane. Fluid leaks within the layers as a result of neovascularisation and can result in irreversible vision loss. Image from [84].

Drusen are broadly categorised into two main types, hard and soft, where AMD severity is usually graded according to the size, number, type of drusen and or the presence of GA and exudative abnormalities [85][86]. Drusen are also characterised on a variety of attributes, where in the context of retinal imaging they are subtyped based on their borders. Red and green wavelengths pass through the retinal layers at unique depths so that each channel within the returned image contains different information (Figure 2.11A). Red light wavelengths travel further to the choroid layers whereas shorter green light wavelengths, where drusen is situated, travels to photoreceptor and RPE layers. This makes drusen more clearly observed in the green channel (Figure 2.11B). Hard drusen are round discrete lesions (< 125μm) with defined borders and tend to be brighter in appearance (Figure 2.11C). The largest

component of hard drusen is cholesterol (40% of drusen composition) where at the core there is a high abundance of non-esterified cholesterol that gives drusen its round and bright appearance in a retinal image [87][88]. Soft drusen have less defined borders (>125μm) and are often confluent with the RPE or decrease in density from centre outwards with fuzzy edges (Figure 2.11D). Soft drusen are often larger than hard drusen (diameter greater than 1000μm) and are composed of loose granulated membranous debris that gives its diffuse and punctate appearance in a retinal image and is often described as a 'starry-sky' pattern [89].

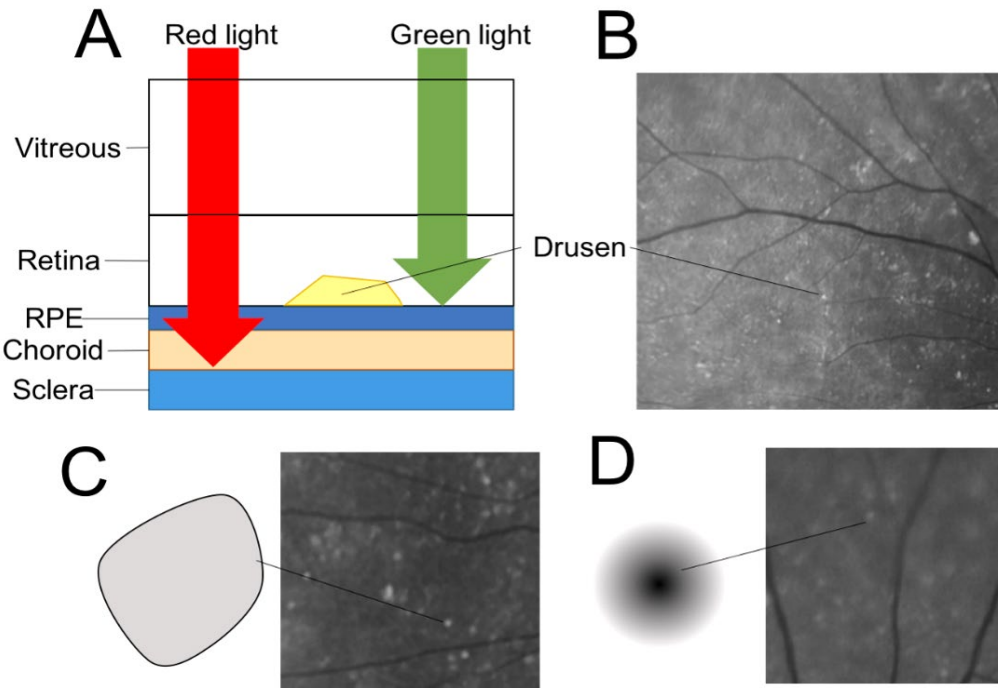

Figure 2.11 Types of drusen. **A)** Schematic of light from the lasers in the UWF instrument reaching different layers to capture different information. The red light reaches lower depths of the retina such as the choroid, where the green light captures more shallow information at the RPE where drusen is situated. **B)** Drusen is most clearly observed in the green channel of a UWF image. This patch from a UWF image contains many hard drusen that are brighter than the surrounding RPE. **C)** Hard drusen are assessed on the basis of their distinct borders. Hard drusen are small discrete lesions that are often round and bright due to their high cholesterol content. **D)** Soft drusen have more diffuse edges with decreasing brightness from the centre outwards. Drusen are much harder to distinguish from the texture of the RPE because they contain loose membranous debris that makes them more confluent with the

RPE.

Figure 2.12 shows what drusen looks like in a UWF image. They are observed as bright yellow or white spots on the retina where hard and soft drusen can appear simultaneously and often cluster. Drusen can be found near blood vessels, the macular and the OD and in various other locations. Drusen can also be similar appearance to dust artefacts from the mirror of the imaging instrument.

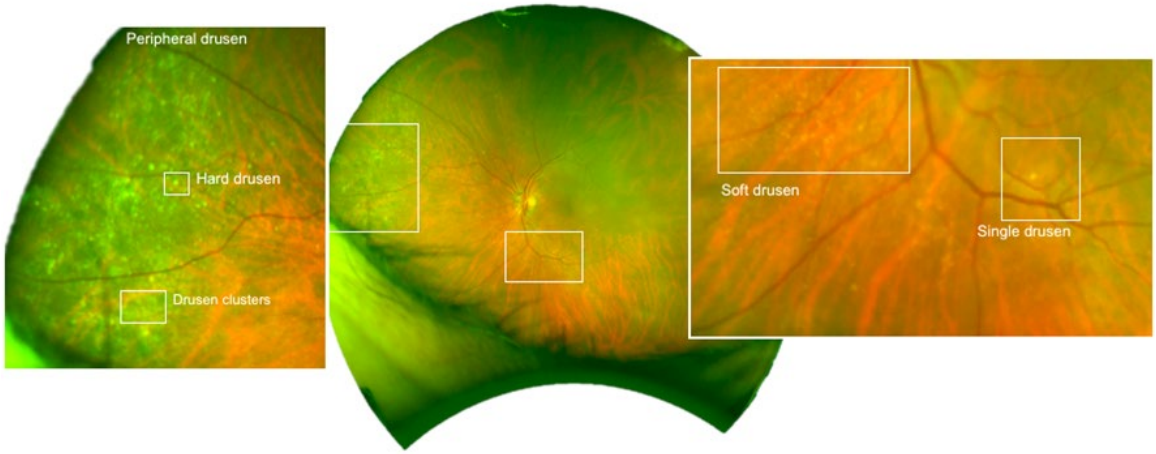

Figure 2.12 Drusen in a UWF image. Drusen are heterogeneous in composition and exists as many types. Soft drusen and hard drusen can appear simultaneously. Soft drusen often form clusters and can be described as a 'starry-sky' pattern.

In this thesis, drusen is an interesting potential biomarker of AD because there have been reports of a higher abundance of hard drusen in the peripheral retina of AD in post mortem tissues [90] and observed in UWF images [91]. AMD and AD have similar risk factors (such as obesity, smoking, hypertension and ageing), share pathways of

inflammation and share a decreased capacity of ageing cells to remove damaged proteins that are also regulated by specific protein pathways [92][93]. There have also been reports that Aβ deposition is specific to drusen from eyes of patients with AMD compared to drusen in normal ageing eyes where increased Aβ assemblies may be associated with advanced AMD [94][55][56][95]. As there are many similarities in the pathogenesis of AMD and AD this suggests that Aβ-targeting therapies being tested for AMD could be applied to AD (and vice versa) [93]. However, it is important to note that the presence of drusen in AMD and AD in a retinal image does not immediately indicate Aβ deposition which could only be confirmed using post mortem biochemical analysis.

There has been recent efforts using curcumin, that binds to Aβ fibrils, to image Aβ load *in vivo* [96]. This study used an automated method to quantify curcumin fluorescence spots in the retina (using an autofluorescent camera system) that was suggested to indicate Aβ deposition where they found more fluorescent spots in AD participants [96]. The limitations of this study were that the participants were required to drink shakes that contained high doses of curcumin where the fluorescent spots were automatically quantified using a method that was not validated against a reference standard (i.e. a human observer) which is needed to evaluate algorithm accuracy [97].

Optical coherence tomography (OCT) is a non-invasive, high resolution retinal imaging technique which uses light waves to form cross-sectional images of the retina

(typically 30° (height) 21° (volume) FOV; up to 5µm axial resolution; 2mm A-scan depth; up to 6 x 6mm scan area). Measurements of the retinal architecture can be obtained and has been used extensively to study changes in mild cognitive impairment and AD. Two meta-analysis [98][99] have shown decreased retinal RNFL thickness is correlated with MMSE scores and cognitive decline. Thinning of the ganglion cell-inner plexiform layer and RGC complex has been observed in AD [100][101]. In an OCT image, soft drusen appear as white mound like elevations of deposit under the RPE whereas hard drusen have a prolate or blunted triangular shape (sometimes described as a 'saw tooth' pattern). Recently, OCT has been used to shed light on the clinical entity of reticular pseudodrusen that accumulate above the RPE that was not previously observed in fundus photographs [102]. OCT is currently proving to be a valuable method for drusen grading and allows drusen types to be assessed within the layers of the retina [103]. OCT angiography (OCTA) is a non-invasive high-resolution retinal imaging technique used to visualise the microvasculature (typically 3x3mm scan area) and has also been used to assess retinal changes in AD. A loss of retinal vessel density, that may indicate accumulation of Aβ, in AD and MCI compared to controls has been reported [104][105].

To the best of my knowledge there has only been one study that has used UWF imaging to fully characterise, drusen load and location in AD where there was significantly more hard drusen in the peripheral retina in AD [91]. In this study, the UWF image was assessed for drusen by overlaying a grid onto the image (see Chapter 4) and manually annotating each grid cell for drusen types. Although the use of UWF

imaging provided a larger area of the retina to be assessed than could be achieved by a fundus, OCT or OCT-A device, manual grading for drusen in the large FOV is a time-consuming process and would be challenging to apply this method of drusen grading to a larger dataset.

2.2.3 The need for automatic drusen detection

In the clinic, ophthalmologists use their expertise to identify types of drusen in a retinal image based on their appearance, with the aim to identify any pathology that may be sight-threatening or causing disturbances in vision. Hard drusen would not need to be treated as they are common in normal ageing but would require regular follow up to make sure they do not develop into soft drusen that may indicate early signs of AMD [106][107]. Such manual grading and or quantification of drusen is a time-consuming task where automatic computerised drusen detection to analyse retinal images would alleviate this challenge and aide the ophthalmologist in stratifying patients for regular follow up. In a research context, finding new associations between drusen and disease would require studies involving the quantification of drusen. However, manual quantification and grading of drusen is difficult and subjective due to their varying nature and similar texture to the ageing retina. In a UWF image there is a large FOV where manually quantifying drusen would be especially difficult and time-consuming. An automatic approach could therefore provide a solution to these challenges and a method to investigate drusen and disease in large datasets as well as monitoring progression in longitudinal studies. In

this thesis, a computerised automatic drusen detection was developed to allow for the quantification of drusen to discover new associations between AD, MCI and CH individuals. It is envisaged that through quantifying drusen in AD and MCI and comparing to CH individuals, there may be unique characteristics between the groups that could indicate early biomarkers of AD in the retina.

2.2.4 Analysing the retinal vasculature to study AD

There have been numerous studies employing vessel width calibres, vessel tortuosity and global branching complexity (fractal dimension) to investigate retinal changes in the context of microvascular health and disease [7]. In a large metanalysis (10,229 participants) vessel width calibres measured from fundus imaging and expressed in the form of common summative parameters central retinal vein equivalent (CRVE) and central retinal artery equivalent (CRAE) have clinical relevance for predicting hypertension [108]. The ratio of CRAE and CRVE, AVR has also been shown to have associations with a decrease in AVR in diabetes mellitus [109] and stroke [110]. Higher fractal dimensions (FD) have been observed in proliferative diabetic retinopathy [111] whilst lower FD has been observed in AD [112][113] in fundus imaging and UWF [91]. Associations between increased retinal tortuosity's and cardiovascular disease as well as with high-density lipoprotein cholesterol have previously been observed in fundus images [114]. To exploit the large FOV of the UWF image and investigate vessel width changes in the peripheral regions a novel parameter, vessel width gradient (WG), was extracted from the largest vessel path in

each quadrant. Venular WG (WGv) has previously been observed to decrease in AD [91].

Such quantitative measures have been extracted from fundus images using semi-automatic specialist software such as SIVA (Singapore 'I' vessel assessment) and VAMPIRE (Vessel assessment and measurement platform for images of the retina) [115][116]. In this thesis the retinal vasculature was analysed using an extension of VAMPIRE for semiautomatic analysis of the retinal vasculature in UWF images [117] (Chapter 5).

2.3 Summary

Associations between retinal changes and neurodegeneration is a longitudinal question to be addressed. In AD for example, it is unclear as to whether retinal changes occur prior to clinical symptoms such as marked cognitive decline and could therefore be used to identify preclinical AD, or whether these changes appear later in the disease process. In order to accurately and reliably quantify and monitor changes in the retina (both in the posterior pole and periphery), novel methods need to be developed. Automatic drusen detection could be used to provide quantitative measures to investigate associations with drusen and disease as well as monitoring progression in longitudinal studies. While the retinal vasculature can be measured in UWF images using existing software, we have no system for automatic drusen detection in UWF images. In part, this is because drusen are ambiguous, difficult and

time-consuming to label in UWF images due to their varying nature. The imaging characteristics of drusen (i.e. bright, yellow and round) are similar in fundus photographs and UWF images of AMD and AD. Therefore, a system that detects drusen would have the potential to detect drusen in any disease where the quantification of drusen load and location may provide insight into associations between drusen and the disease of interest (i.e. AD). This compels an investigation into previously developed automatic methods for detecting in fundus images in order to inspire the development of drusen detection in UWF images.

Chapter 3 Investigating automatic detection of drusen

3.1 Introduction

The focus of this chapter is the investigation into developing automated drusen analysis in UWF-SLO. Automatic drusen detection is an under-developed area for this imaging modality but is comparable to drusen detection in conventional fundus photography. This compels an investigation into which strategies have been utilised in previous fundus imaging work and to identify recent state-of-the-art techniques that are of relevance to my thesis, with the aim of assessing techniques that are applicable to UWF-SLO. The investigation was conducted as a systematic review [118]. The objective was to identify and critique the methods adopted in various algorithms and report their performance against a reference standard. Methods identified from the literature were then tested on UWF images and are also described in this chapter. This information was used to guide the development of drusen detection in UWF in Chapter 4.

3.2 The literature search

3.2.1 Search strategy

To filter articles featuring automatic drusen detection a computerised search of EMBASE, PubMed, Web of Knowledge, ScienceDirect, ACM Digital Library and IEEE Xplore was conducted using search terms "*drusen"* and in combination with *"detection"* or *"classification"* or *"identification"* or *"segmentation"* or *"quantification"* or *"measurement"* or *"algorithm"*. Titles and abstracts were filtered on whether they contain "*age-related macular degeneration"* or *"AMD"* or "*Alzheimer's disease*" or "*AD*" to identify articles featuring only AMD and AD.

3.2.2 Inclusion and exclusion criteria

Articles were included only if they were an original study, written in English and evaluated performance of the algorithm against at least one human grader. Articles were excluded if they were a review, non-human research, non-English language, did not use fundus photography (e.g. OCT) and if the performance of the algorithm was not validated against a human grader.

The process of algorithm validation consists of showing that the algorithm performs correctly through comparison of output to a reference standard [97]. The reference standard is defined here as the manual grading of an image by a human. Performance is usually reported as a statistic such as sensitivity, specificity, accuracy and ROC (see Chapter 4 for definitions of these terms). Another aspect is the size of the dataset used to validate the algorithm. If a small number of images had been used to test the algorithm this might not be representative of a larger target population. Fundus photographs vary between individuals in the extent of disease manifestation or image quality. This means that an algorithm tried out on a small dataset may perform well in the testing phase but fail to generalise to the larger population where more complex data is likely. Articles that validated their proposed system on less than 50 images were thus excluded.

3.2.3 Results

Figure 3.1 shows the results of the literature search as a flow diagram. The initial search returned 2,236 articles in 2017. A further 1,318 articles were excluded after filtering for AMD and AD, such as articles featuring glaucoma (42 articles) and diabetic retinopathy (45 articles). 834 from the remaining 918 articles were excluded as they did not use fundus photography (18 articles), were not an imaging study (770 articles) or were reviews (34 articles). 73 articles did not meet the inclusion criteria where 9 articles did not report a performance to a reference standard, 14 articles validated on fewer than 50 fundus photographs, 3 articles featured software optimisation and

2 articles reported hardware. Following bibliography search of the remaining articles, 1 additional article was identified that met the inclusion criteria. 5 other articles were included following a hand search for the year 2018 and 2 articles were included for the current year (2019). This totalled 16 articles where 11 articles used machine learning techniques and are discussed in this chapter. 5 articles reported deep learning techniques and are discussed in section 4.2.3.

During the search, an earlier review of drusen detection in fundus imaging was also identified [119]. At the time this article was published the techniques considered conventional image processing operations as opposed to machine learning techniques. None of the articles discussed met my exclusion and inclusion criteria and were not included in the systematic review [118] but were considered in this thesis for the investigation into drusen detection in UWF. Notably, these articles did not report performance to a reference standard. In order for such systems to be deployed in studies involving drusen quantification, knowledge of their accuracy would be required to determine if the automatic measurements can be interpreted with a degree of confidence. To the best of my knowledge, there are no publicly available datasets that contain manual annotations of individual drusen to validate a computerised segmentation approach in either fundus or UWF images.

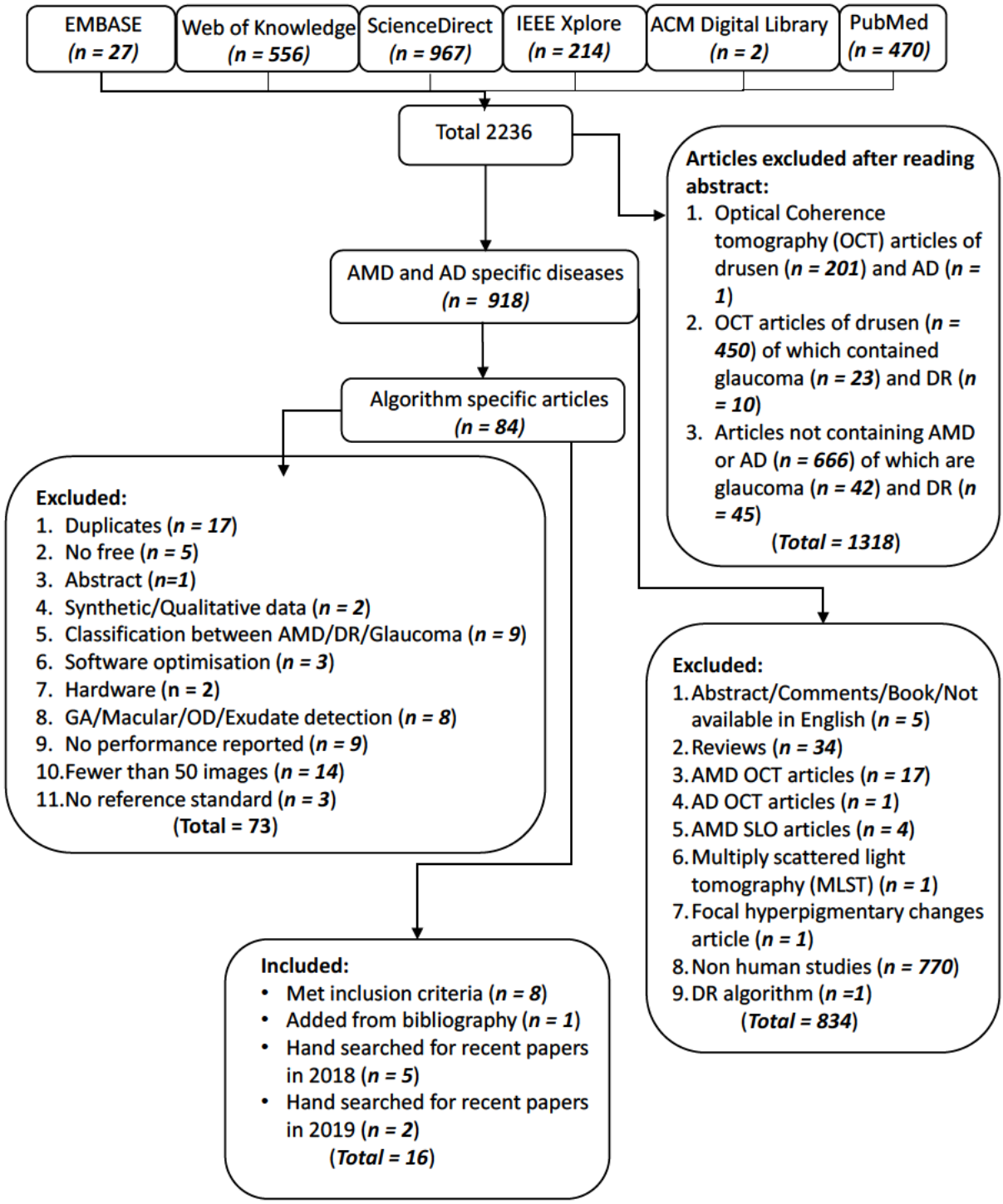

Figure 3.1 Flow diagram of the literature search to identify relevant articles on automatic drusen detection in fundus imaging.

3.3 Conventional image processing

Conventional image processing is a term used in this thesis to refer to methods of extracting information from images using digital processing that requires no learning stage. Such methods involve performing operations on an image to either enhance the image or to extract information from it. Drusen is a characteristic sign of early AMD and grading protocols use size, number and type [120][121][85] to determine AMD severity and monitor progression where earlier efforts to quantify drusen involved conventional image processing techniques to segment drusen [119]. Application of such techniques often need a pre-processing step to address noise that might be present in a fundus image and non-uniform illumination. In a fundus photograph the shape of the retina, vignetting at the edge of the pupil and media opacity contributes to non-uniform illumination across the FOV when illuminating light from the imaging system enters the eye at particular angles. Fundus photographs also contain a black border that needs to be removed as this does not contain any relevant information. Objects that appear bright in a fundus photograph, such as the OD, need to be considered during image processing. These challenges are also present in UWF imaging.

3.3.1 Drusen segmentation in fundus photographs

Histogram based approaches have previously been applied to fundus photographs to segment drusen [122][123]. Pixel values in an image range from 0 (black) to 255

(white) in each of the red, green and blue (RGB) colour channels. An image histogram is the frequency of occurrence of intensity across all the pixels. This allows the number of pixels that have a particular intensity value to be visually inspected and where the shape of the histogram provides information pertaining to the overall quality of the image, characteristics of objects and the potential for enhancement. Rapantzikos et al [123] proposed a histogram based algorithm for segmenting local regions of drusen in fundus photographs called histogram adaptive thresholding (HALT). The authors observed that histograms of windows containing drusen had a positively skewed distribution with a peak of bright pixels. To obtain a segmentation from a histogram, Otsu's thresholding approach, a well-established thresholding method that finds the threshold that maximises the variance between peaks in a histogram [124], was used to segment drusen regions from the background. The result was that there was no need to correct for non-uniform illumination or remove border pixels. The algorithm worked well on clearly defined drusen but failed on drusen that were vague, small or located inside bright regions, within noise and near vessels. Checco et al [122] used a variety of noise removal (random variation of brightness or colour in an image) and histogram normalisation that equalises pixel intensities across the intensity range to correct for non-uniform illumination and to enhance drusen in a manually defined local region in a fundus photograph. A threshold was applied to the enhanced image and drusen were segmented.

Figure 3.2A shows the pixel intensity profiles of selected regions in a UWF image extracted over a set distance of 110µm, which is approximately the same size as the smallest window in [123]. The regions and corresponding signal of intensity profiles contain abrupt changes related to objects in the image. The profile plot shows three large peaks of bright pixel intensities that were sampled from hard (Figure 3.2D) and soft drusen (Figure 3.2E) as well as objects that are thought to be dust inside the imaging system (Figure 3.2G). The lower intensities of these profiles are similar in behaviour to the profiles pertaining to RPE of early AMD (Figure 3.1B) and RPE of an older eye (Figure 3.2C) and a blood vessel (Figure 3.2G). This figure shows that although drusen in a UWF image have a characteristic pixel intensity distribution, certain features within the drusen profile can be similar to those found in other objects such as the intensity of the RPE and can occur simultaneously with intensity of other objects such as blood vessels. This highlights how using a histogram method in a UWF image to segment drusen would be challenging.

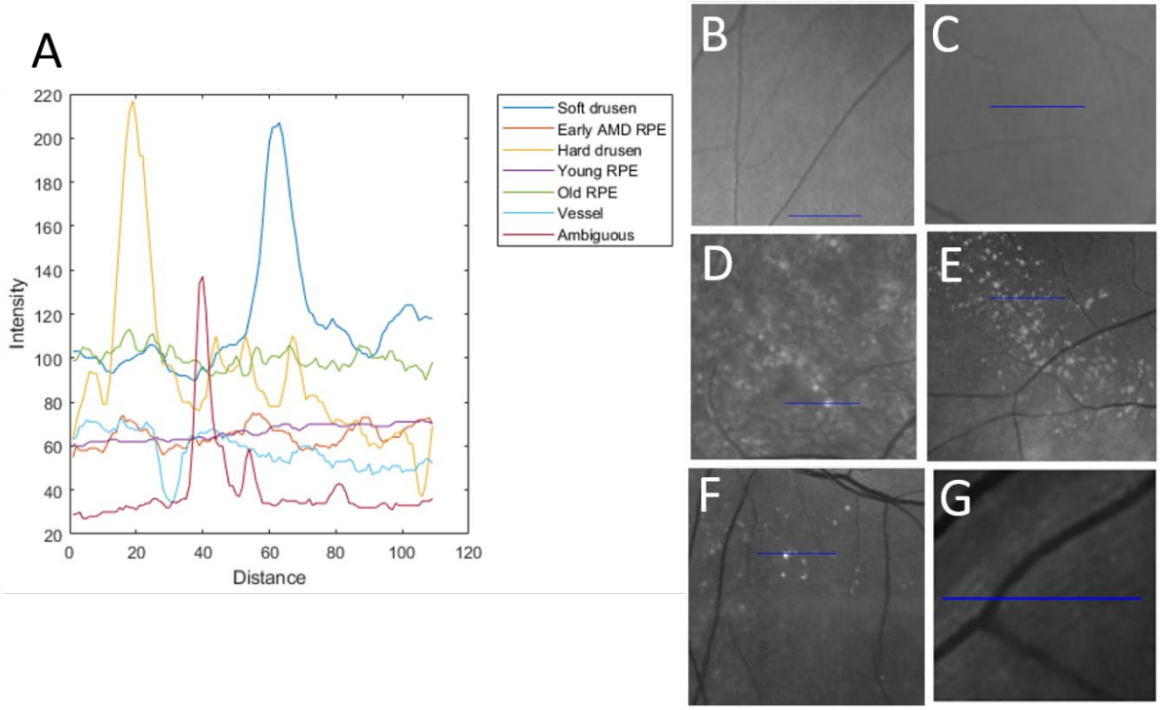

Figure 3.2 A) Intensity profiles of different regions in a UWF image sampled over a distance of 110µm. **B)** sample of early AMD RPE, **(C)** sample of older healthy RPE, **(D)** sample of hard drusen, **(E)** sample of soft drusen, **(F)** sample of ambiguous object that could be either drusen or dust, and **(G)** sample across a blood vessel.

Segmentation of drusen in fundus images using texture has previously been proposed [125]. Texture is a nebulous concept where its interpretation sometimes depends on human perception but mathematical definitions for quantifying it are also available. Drusen are different in texture and colour than the normal appearing retinal surface. For a computer to understand this and use texture content to segment drusen, we need to describe it in a numerical form. There are many methods to extract texture information from an image such as using histogram properties, morphological

operations and filter based techniques, and it is a rich and long standing field of computer vision [126][127]. In the previous figure, the intensity profiles reflected, in part, the texture of different objects in the UWF image. These signals can therefore be utilised to describe texture.

A standard method uses a Fourier transform that convolves a signal into its sine and cosine wave equivalents and into the frequency domain [128]. In an image, each pixel in the Fourier transform will correspond to the intensity of the wave and the position of the pixel corresponds to the frequency and orientation of the wave. The aim of the Fourier transform in image processing is to select certain waves that have a specific frequency and orientation that represents the object of interest. A wavelet is a rapid decaying oscillation that has 0 mean and exists for a finite duration in different sizes and shapes. A wavelet can be scaled to a larger factor that stretches the wavelet and corresponds to a lower frequency or can be scaled by a smaller factor that shrinks the wavelet and corresponds to a higher frequency. A low frequency wavelet when passed over a Fourier transform signal will capture slowly varying changes in a signal and a high frequency wavelet will capture abrupt changes. The wavelet can be shifted along the signal so that it is in line with the feature or object of interest. This type of processing in the frequency domain is called filtering and is used to smooth, sharpen and enhance images. Filters can be applied to the Fourier transform in order to manipulate a specific frequency, such as by amplifying and attenuating the signal, so that by inverting the transformation the desired effect will be applied to the signal [126][127].

Figure 3.3A shows pixel intensity values sampled across a UWF image where the arrows indicate the peaks and troughs corresponding to whether this signal is from a blood vessel, drusen or the retinal surface (Figure 3.3B). A Fourier transform was applied to the signal transforming it into the frequency domain where high frequency spikes correspond to high intensity values (Figure 3.3C). This operation was applied to a whole image and Figure 3.3D illustrates its effect on a small image patch that contains drusen. A Fourier transform of the image is represented in the frequency domain which consists of a signal with magnitude (the intensity of the wave) and the phase (the position of the wave). The image is reconstructed using the phase and magnitude to re-transform the image back into the spatial domain. This results in an enhanced image where drusen appear better defined as do the vessels but there is still some high frequency noise. A Gaussian filter is a filter that has a Gaussian shape that can be used to remove high frequency components whilst preserving high frequency edges (i.e. it is a low pass filter) [129]. In Figure 3.3D, a Gaussian filter is applied to the Fourier transform that results in a reduction of noise and an enhancement of the edges of drusen and blood vessels. Parvathi et al [125] proposed using a Gabor filter (a Gaussian wavelet modulated by a sinusoidal plane) to filter a manually defined region of interest (ROI) in a fundus photograph for high frequency components that correspond to edges of drusen to obtain a segmentation. This method required OD and blood vessel suppression by masking the OD and replacing the dark pixels from blood vessels in the green channel with the local mean of the pixels in the image. Since a Gabor filter can have multiple shapes and scales, this

method required parameter tuning (i.e. selecting the optimum shapes, size of the filter) and could only be applied in local regions of a fundus photograph.

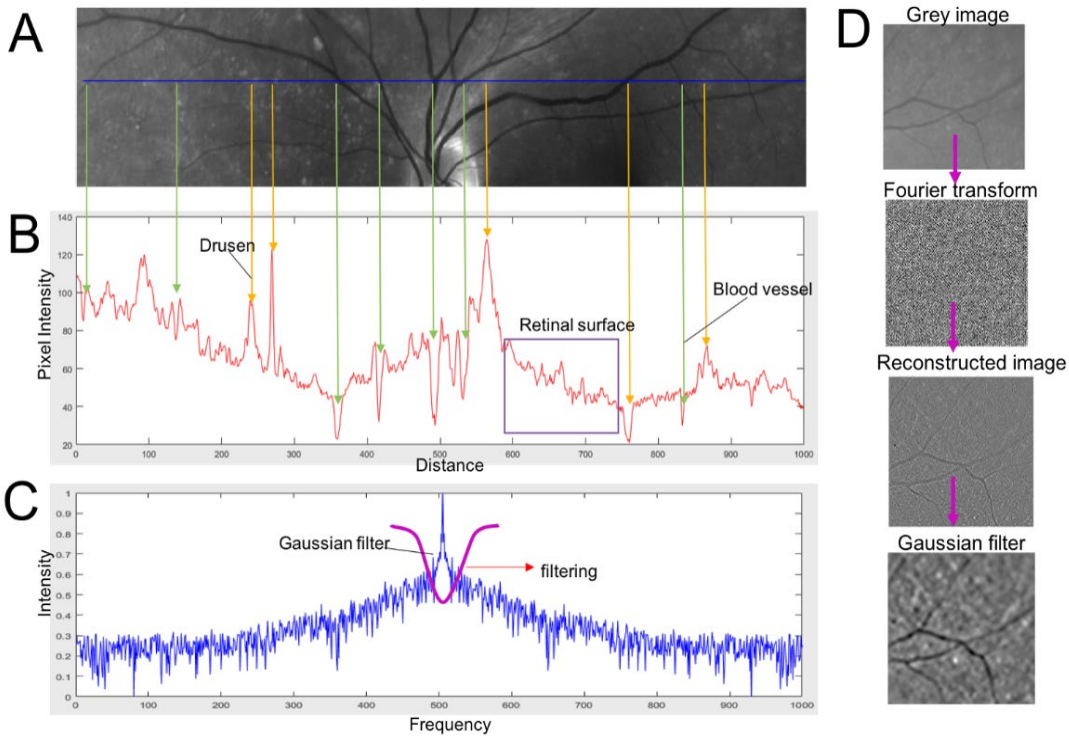

Figure 3.3 Applying a Fourier Transform to a UWF image to enhance drusen. **A)** Shows a blue profile line indicating where pixel intensity values were sampled. **B)** Displays the distance along the sample line plotted against pixel intensity. **C)** The Fourier transform of the signal into the frequency domain. Hypothetical Gaussian filter is overlaid where it can be passed along the signal to filter high frequency noise. In this example the Gaussian filter is placed over a region of high frequency that would be removed to reduce noise. **D)** Each processing step applied to an image patch containing drusen.

Morphological operations have previously been applied to fundus photographs to segment drusen [130]. Morphological processing involves defining a structuring element such as a binary circle or a square and placing it at every location over the input image and comparing it to the surrounding pixels. This can switch on the pixels that are black (i.e. replace with 1) within the neighbourhood defined by the structuring element (dilation) or remove pixels at the boundary (erosion). The result from a dilation is that objects in the binary image become larger in shape and in an erosion, they become smaller with some detail removed. Sbeh et al [130] proposed a structuring element derived from the region maxima of the pixel intensity profile of drusen. This method applied the structuring element to a manually defined ROI that was smoothed using a median filter. The challenge with this method was that the structuring element could only be applied to local regions of drusen at the fovea and gives coarse results when extending to the image as a whole. Figure 3.4 shows a structuring element applied to the processed image patch from the previous figure. Firstly, the background is defined as the Gaussian filtered image and is subtracted from the green channel resulting in an enhancement of drusen pixels (Figure 3.4A). As drusen are circular in appearance, a circle or disk structuring element was defined with a diameter of 10 pixels (Figure 3.4B). The structuring element is applied over the image as a dilation to extend bright objects into the shape of a circle (Figure 3.4C) and the border pixels are removed (Figure 3.4D). The result is a grey image that can be thresholded by selecting all grey values above 0 and the boundary of the resulting mask is traced onto the original grey image (Figure 3.4.E). Bright drusen are

segmented in the centre of the image and the use of a morphological operation creates a smooth outline of drusen.

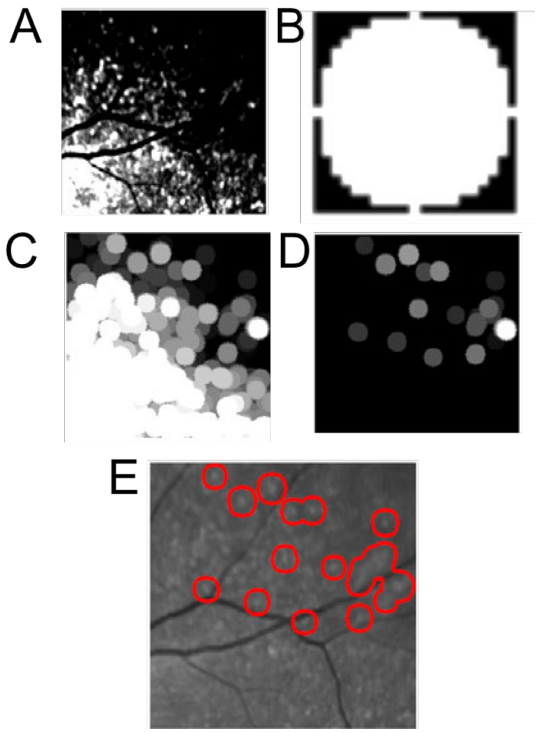

Figure 3.4 Application of morphological processing using a disk structuring element to the patch of drusen in a UWF image from Figure 3.3D after gaussian filtering. The purpose is to obtain segmentation of drusen in the image. **A)** Gaussian filtered image that is subtracted from the green channel and results in an enhancement of drusen pixels. **B)** Binary disk structuring element with a diameter of 10 pixels. **C)** The structuring element is applied over the greyscale image as a dilation to extend bright objects into the shape of a circle. **D)** The border pixels are removed using connected components. **E)** The image is thresholded by selecting all pixels that are above the value of 1 and the outline of the mask is overlaid onto the original image.

3.3.2 Drusen segmentation in UWF-SLO

The methods described in Section 3.2.1 for fundus camera imaging were applied to UWF imaging in order to assess whether they might be suitable for segmenting drusen in these types of images. Conventional image processing methods posed four main challenges. The first was the presence of non-uniform illumination, which would make drusen appearing in the centre of the image brighter but darker in the periphery, and so a segmentation method would need to be robust to such variations. Second, the presence of other bright appearing objects such as the OD might need to be suppressed to reduce the number of false positives. Thirdly, traditional methods seem to perform well on image patches but often failed when applied to the full image in one go. Finally, segmentation methods work well on hard drusen that have clearly defined edges as opposed to soft drusen that are blurrier in appearance. What these methods had in common is the use of the green channel on which segmentation of drusen was performed as they are most visible against the retinal tissue in this colour channel (see chapter 2). With these challenges in mind, a combination of histogram, texture-based and morphological based methods were tested on a fundus photograph and then applied to a UWF image.

Figure 3.5 shows a fundus photograph (green channel only; 2336x3504 pixels) containing drusen and the corresponding image histograms as the input image is put

through the initial processing steps. Figure 3.5A shows the input image and Figure 3.5B shows the same photograph but the OD region suppressed, which was achieved by manually outlining the OD boundary and setting pixels within this region to black. In the corresponding image histogram there is a non-uniform, left-skewed distribution of pixel intensities where removal of the OD decreases a small peak in bright pixels (denoted by the magenta arrows). Non-uniform illumination correction was applied to both images using a Top-hat filter that is a form of morphological filtering using a structuring element (in this method a disk of radius 20 pixels) to perform morphological opening that is subtracted from the original image [131]. The result is an image where the larger bright objects are removed whilst keeping the smaller ones (i.e. possible drusen) (Figure 3.5 C and D). This results in a narrower and more uniform grey level intensity distribution.

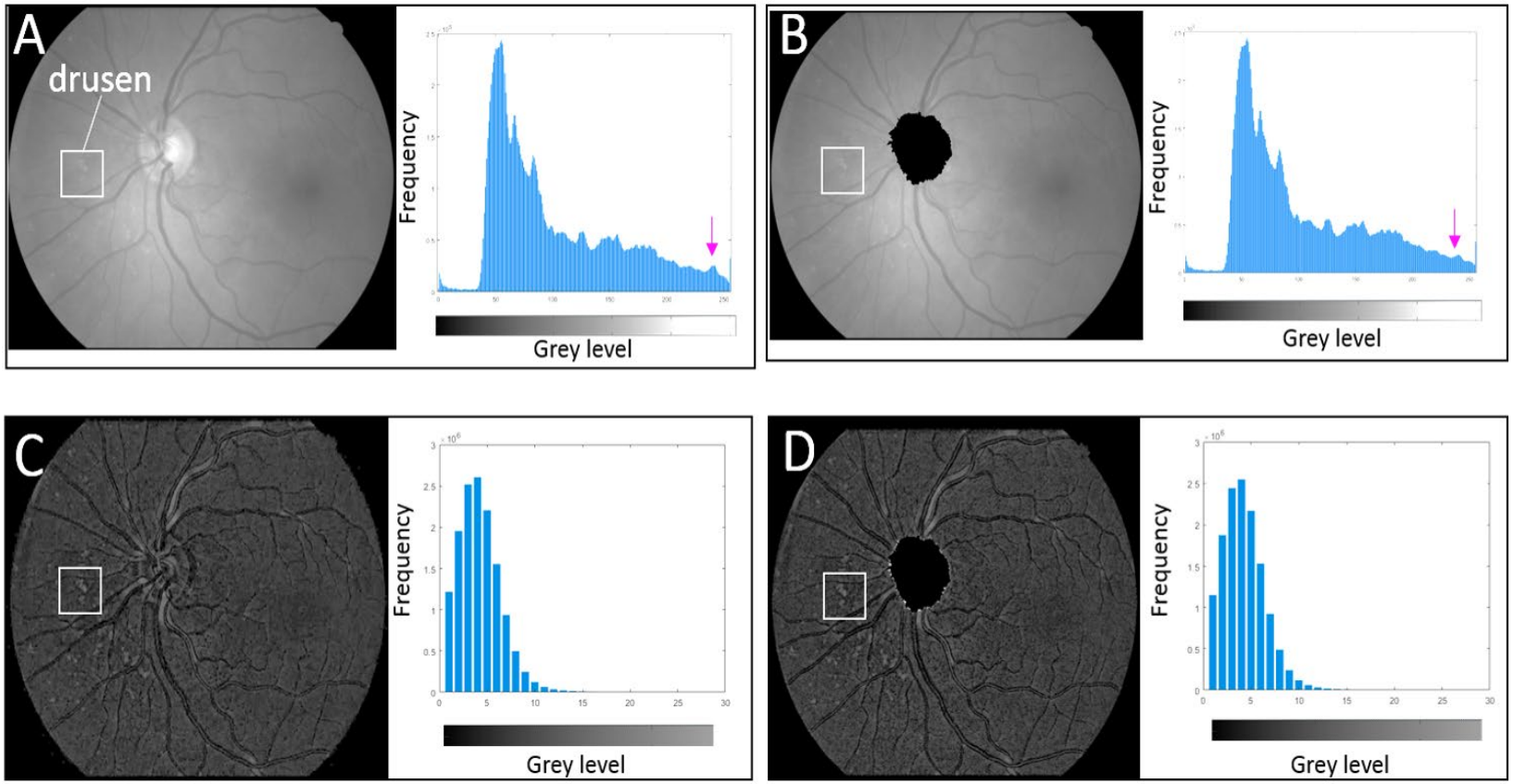

Figure 3.5 Correcting non-uniform illumination in a fundus photograph using a Top-hat filter. **A)** The green channel of the fundus photograph as a grey scale image with an area where there is a mixture of soft and hard drusen marked (white box). It can be seen that

the centre of the image is lighter than peripheral areas of the image. **B)** The region corresponding to the OD has been suppressed manually which has reduced a small peak of bright pixels (*magenta arrow*). **C)** Top-hat filtering of the green channel leads to an enhancement of drusen as well as an image histogram that has a normal distribution between a narrower range of pixel intensities. **D)** The same Top-hat filter applied to the OD suppressed fundus photograph.

To extract texture content from the image, a Fourier transform with a Gaussian filter was applied and subtracted from the filtered green channel, as in Section 3.3.1. Figure 3.6A shows the output with (*left column*) and without OD suppression (*right column*). The result is that smaller brighter pixels that are different in texture to the rest of the image have been enhanced, but there is still some noise within the image from small bright regions. To extract the larger objects that could be drusen and reduce the noise, a second disk structuring element (10 pixels) is applied to the image (see Figure 3.6B) that reduces noise while the larger brighter objects that may be drusen are enhanced. However, this operation has also enhanced the bright regions around the OD even with suppression. To remove this, the image was binarized by selecting all pixels above the value of 0 (black) and the largest object, identified using connected components whereby an object in an image is identified as being part of the same object by scanning neighbouring pixel values, was removed. Figure 3.6 C shows the removal of this region (magenta) in the images without and with OD suppression. This results in the binary images shown in Figure 3.6 D.

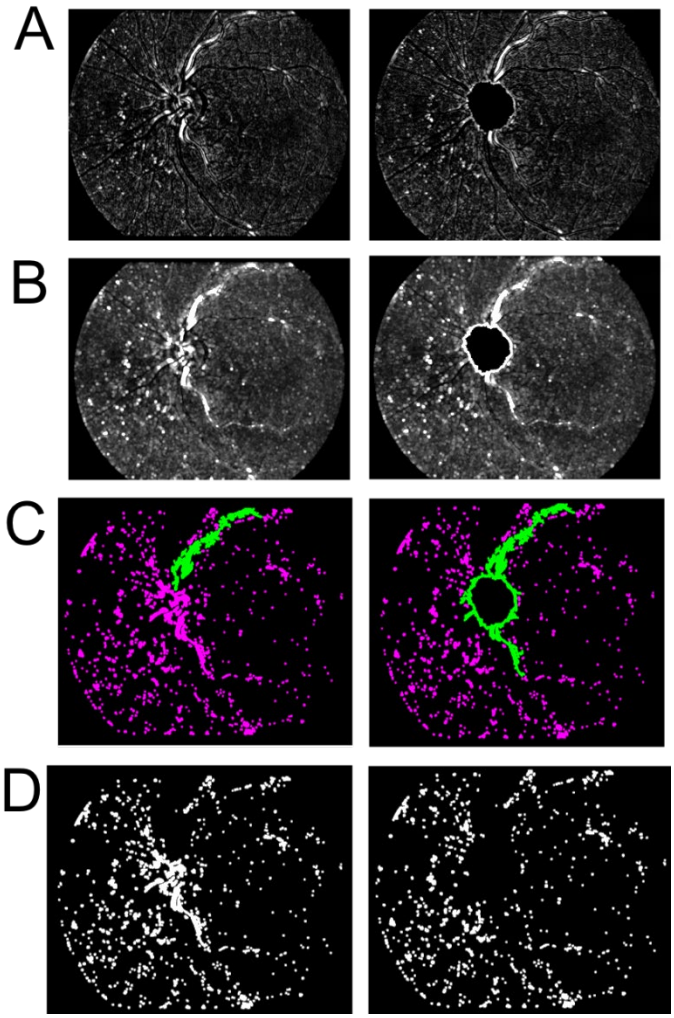

Figure 3.6 Texture based segmentation with morphological filtering of a fundus photograph without OD suppression (*left*) and with OD suppression (*right*). **A)** image segmented by texture. **B)** Morphological filtering using a disk structuring element of 10 pixels in diameter. **(C)** Removal of the object with the largest area (g*reen*) from the smaller objects (*magenta*). **(D)** Final binarized image.

Figure 3.7 shows the final binary segmentations overlaid onto the original fundus photograph where it can be seen that there is a high number of false positives in

segmentation when the OD was not suppressed. Also shown is a close up of the green channel with outlines of candidate drusen overlaid. By visual inspection, the black arrows indicate drusen that has been correctly segmented while the red arrows indicate false positives that occur around blood vessels. The green arrows indicate where drusen was not successfully segmented.

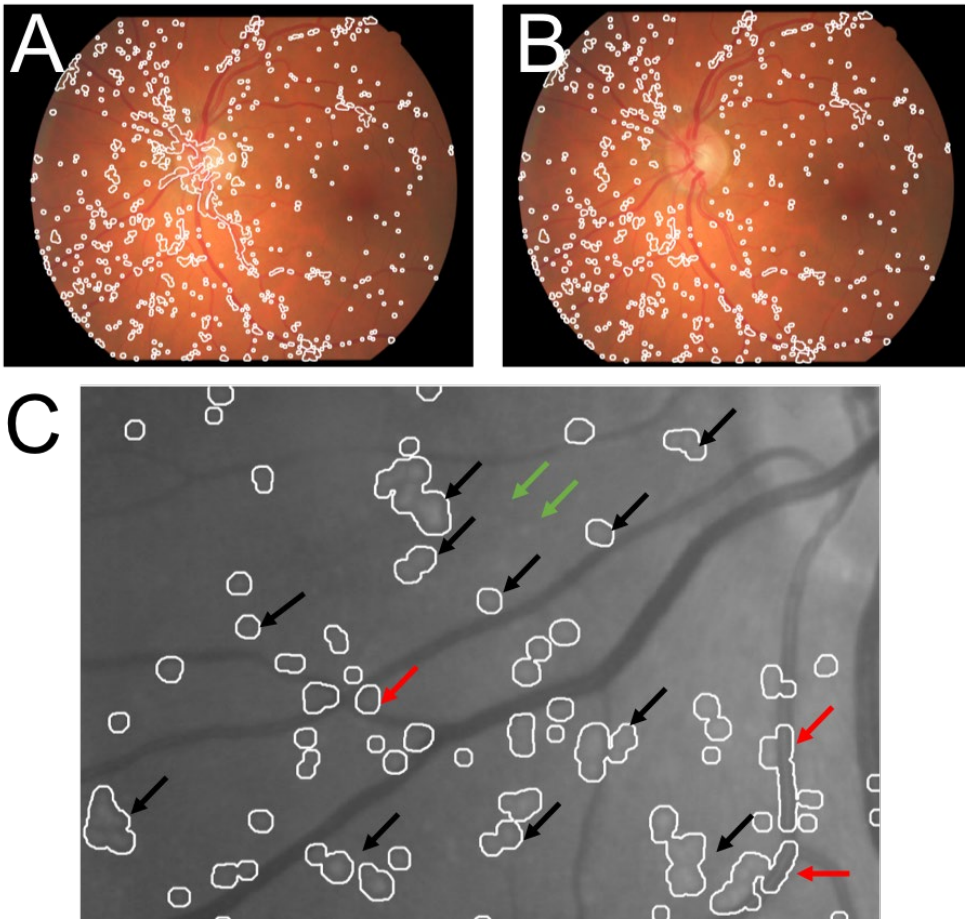

Figure 3.7 Result of texture-based segmentation overlaid onto a fundus photograph without OD suppression **(A)** and with **(B)**. **C)** A close up showing a section of the green channel overlaid with outlines of the segmentation output. Black arrows indicate drusen that have been correctly segmented while red arrows indicate false positives near blood vessels and green arrows indicate false negatives where soft drusen with less defined edges were unsuccessfully detected

Figure 3.8A illustrates the result of the method applied to a UWF image (with the OD suppressed). The image contains clusters of hard and soft drusen as well as eye lashes

that have covered some of the FOV. The face plate is also apparent in the image and contained dust particulates over its surface, and there is a dark spot near the OD. This is a representative example of a challenging UWF image because it contains large amounts of drusen, non-uniform illumination and artefacts (some of which have a similar appearance to drusen, i.e. dust). Visually, the segmentation method has worked well to identify drusen clusters but has produced a false positive segmentation of the face plate. Given that previous work performed well on image patches, a cropped section of the image was segmented separately and is shown in Figure 3.8B where it can be seen that some hard drusen have been identified but also a region that containing both hard and soft drusen as well as normal retinal tissue has been incorrectly selected.

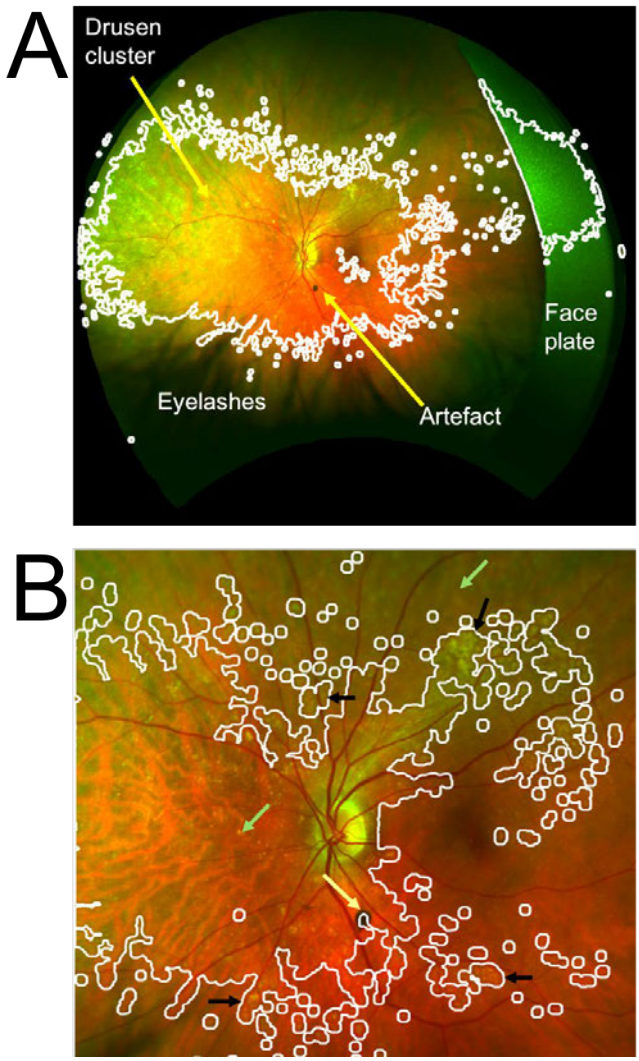

Figure 3.8 Application of traditional image processing technique to segment drusen in a UWF. **A)** UWF image containing large clusters of hard and soft drusen. Eyelashes are present in the bottom of the image as well as a portion of the face plate that appears due to sub-optimal patient positioning at image acquisition. There is a dark artefact near the OD that could be vitreous material. **B)** Segmentation of a local region in the central area of the UWF image. Black arrows indicate drusen that have been correctly segmented while the green arrows indicate false negatives and the

yellow arrow indicates how the segmentation has been unduly influenced by artefact.

As previous grading protocols have used drusen size, type and number to monitor drusen progression and severity [120][121][85], the segmentation result would not be able to return this type of information accurately. In addition, one of the major challenges with some of the previously discussed work discussed is that there was no reporting of segmentation performance against a reference standard. This is because manually segmenting drusen by hand is labour intensive and subjective. Although conventional image processing may produce promising results in a visual sense, without validation to a reference standard their performance cannot be assessed quantitatively.

3.4 Machine learning

The results from the aforementioned literature search showed a trend moving away from traditional image processing techniques (reviewed in [119]) towards novel machine learning approaches where the overall aim is to classify an image into a particular class. Machine learning is a subfield of artificial intelligence (AI) and is based on how humans learn from past experiences. For example, an individual can decide whether they like an item of clothing or not based on the colour and brightness or intensity. In Figure 3.9 this shown as a plot where the points indicate whether an individual likes an item of clothing (green dot) or not (red dot). From this

data we can deduce that this individual does not like dark red clothing but likes bright blue clothing as it is linearly separable. Based on this previous knowledge, when we are presented with a new item of clothing that is bright blue, we would be able to predict that this individual will like this item of clothing. However, if we have another item of clothing that has a medium intensity and is green in colour, it is unclear as to whether they will like or dislike the item. This is where machine learning can be used to model scenarios where the features that define the data are more complex than colour and intensity but can be used to predict the group that new data belongs to. In this example, one could decide that item 2 is closer to more 'likes' than 'dislikes' and therefore it is a predicted to be a 'like'. Such an algorithm or technique is called k-nearest neighbours (KNN) and is an early form of machine learning. Just as an individual can learn what they "like" and "dislike" about an object and use this to decide what they might think about new object, so too can a machine learning algorithm use previously learned features from the data to predict a classification for new, unseen data.

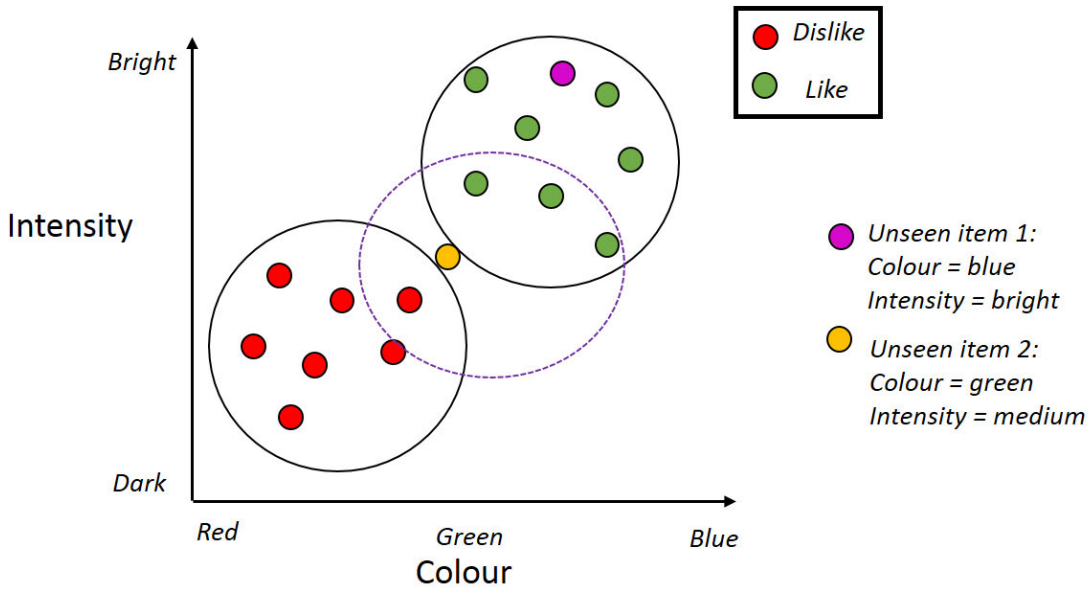

Figure 3.9 Illustrating the concept of classification. The example shows two features that an individual might use to decide whether they like an item of clothing or not. The known information is plotted and could be separated using linear regression, where this individual seems to like bright blue clothing but dislikes dark red clothing. If we have a new item of clothing that is bright blue, it is plotted (*magenta data point*), and we can see that the individual will most probably like the item. When presented with a more difficult decision that is an item of medium intensity and green in colour (*yellow* data point), the data is no longer linearly separable and would require a different method to predict which group it belongs to. In this example we could use the neighbouring data points to decide the label for the new item of clothing. This is called KNN classifier, where k is the number of neighbouring points used.

When the data is an image, features from the data need to be extracted. This is often referred to as handcrafted features and is manually designed by identifying distinguishing features in an image relating to the object of interest. An ophthalmologist can distinguish features of a disease such as GA or drusen, but an AI algorithm will need to extract features from pixels that pertain to an object such as drusen. In Section 3.3 a technique that exploited the brightness, colour and texture of drusen was used to segment drusen from the retinal background. This can be extended to more complex features derived from properties within the image to build a classification algorithm that separates the data into the desired classes. These features can either be learned from previous examples (supervised learning) or determined by the algorithm (unsupervised learning). Figure 3.10 [118] shows an example of a supervised learning pipeline. First, the image is pre-processed to enhance image features or to reduce noise. Next, features such as measures of entropy (properties from the pixel intensity levels in an image), energy (properties from histogram distribution), texture and geometric properties are extracted from the image. This is transformed into a numerical vector that typically undergoes a feature selection stage, whereby the best features that describe the data are selected. At the so-called training stage, a classification algorithm is applied to use patterns learned in the data to distinguish subgroups. In order to validate the algorithm, a testing set is then input into the trained classifier and used to classify new images into subgroups. The results from the literature search (see Section 3.2.3) will be discussed in terms of these machine learning steps and are subsequently

applied to a set of UWF images to classify them into classes - early AMD, older eye and younger eye.

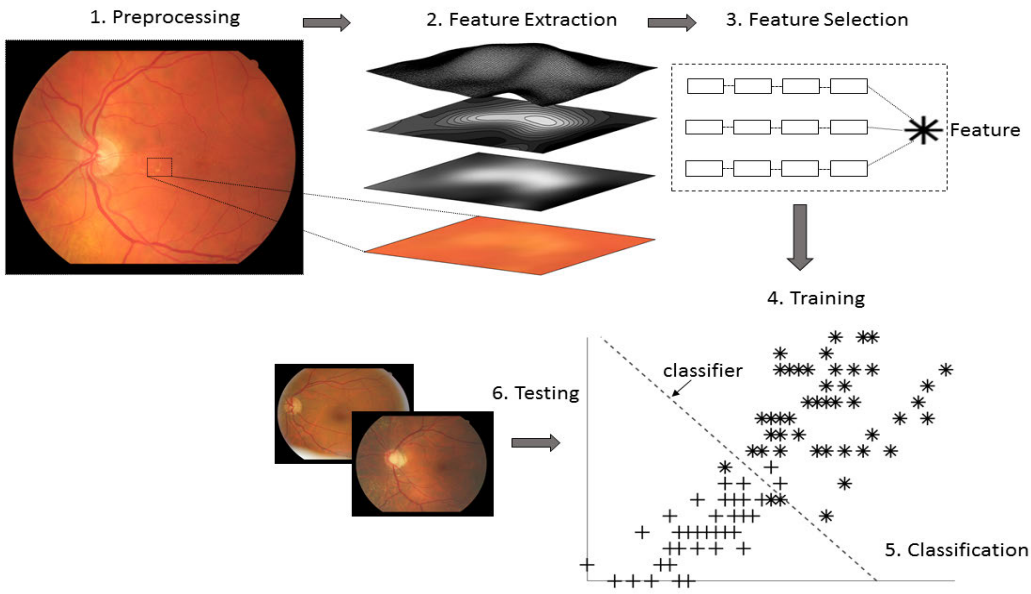

Figure 3.10 Example of a supervised machine learning pipeline from [118]. **1)** The image is preprocessed either to reduce noise or enhance features of interest in the image. **2)** Features such as pixel intensities, geometric properties and textures are extracted. **3)** Features are converted into a numerical vector and undergoes a feature selection stage where the best features that distinguish classes are selected. **4)** A classifier is trained that separates the data into the subgroups. **5)** Training generates a mathematical function that separates the classes. **6)** Testing images are input into the trained classifier to identify which group the image belongs to and this is validated against a reference standard.

3.4.1 Study designs and populations

The 14 studies that met the inclusion and exclusion criteria used 4 publicly available datasets (ARIA [132], STARE [133], AREDS [134], RetinaGallery [135]), 3 used private datasets, 1 article sourced fundus images from a telemedicine platform and 1 article used images from an independent study [136]. Although each article used different methods some articles used the same datasets. Four articles aimed to classify fundus photographs as *disease* or *no disease,* 6 articles aimed to classify AMD into severities using the AREDS grading protocol [86] or an in-house grading protocol (Cologne Image Reading Centre and Laboratory (CIRCLE)). 2 articles aimed to classify *Dry AMD* vs. *Normal* images and 1 *Wet AMD* vs. *Dry AMD* or *Normal*. Table 3.2 summaries the dataset and cameras used for each of the 11 articles [118].

3.4.2 Preprocessing and feature extraction

Similar to conventional image processing, properties from the image histogram (e.g. energy, entropy and intensity) have commonly been used for classifying regions that contain drusen or not. To improve contrast in an image Contrast Limited Adaptive Histogram Equalisation [137] has been used frequently [138][139][139][140][141]. This technique flattens the image histogram to make the image equal in colour intensity that results in an enhanced image in the sense that the image appears to have more definition between objects in the image (i.e. more contrast) Following removal of the black border, median filters have been used to smooth high frequency

noise [142][143]. Grivinsen et al [144] employed intensity and contrast features from manually defined boundaries of drusen that to train a classifier. Burlina et al [145] used conventional image processing techniques such as median filtering, morphological dilation and thresholding to obtain regions of no drusen (background) and drusen. Garcia-Floriano et al [146] exploited mathematical morphology to segment regions of drusen and the macular that was subsequently used to extract a quantitative measure of pixel intensity shape called a Hu moment. The type of preprocessing applied depended on the particular features to be extracted. These studies all exploited drusen colour and intensity from an enhanced green channel to derive a mathematical definition of drusen that could be used to distinguish them from other objects present in the image (i.e. background, optic disc, macular). Hijazi et al [138] replaced blood vessels with null values in their proposed system where removal of black border was not required. Table 3.1 summarises the preprocessing and features used in each study [118].

REFERENCE	DATASET	FUNDUS CAMERA (IMAGE SIZE IN PIXELS)	PRE-PROCESSING	FEATURE	OUTPUT
HIJAZI ET AL 2010 [25]	144 (ARIA)	Not reported	CLAHE Retinal vessels segmented by thresholding and OD segmented using intensity peaks of image (identified by sliding window)	RGB and HSI histogram of each image conceptualised to set of curves (time series)	Disease/No Disease
BURLINA ET AL 2011 [7]	66 (private)	Zeiss FF4 40° FOV (pupils dilated) Images resized to 1000 x 1000	Pyramid decomposition of green channel for regions of high gradient magnitude to create logical masks for training and testing. Areas of high gradient magnitude indicate artefacts and vessels where low gradient magnitude indicate normal retinal tissue	Intensity, colour and gradient features of background (normal retina) and candidate abnormal areas	Disease/No Disease
ZHENG ET AL 2012 [61]	101(ARIA) 97(STARE)	TOPCON TRV-50 fundus camera 35 ° field of view (700 x 605)	Mask of whole image to capture circular fundus ROI. Colour normalisation and uneven illumination is applied. CLAHE to enhance contrast. Blood vessels identified using wavelet features.	Image represented as quadtree, separated by their homogeny, defined by similar pixel values. Image mining algorithm returns features	Disease/No Disease

KANKANABALLI ET AL 2013 [31]	2772(NIH AREDS)	Not reported	Green channel smoothed by large median filter. Median filtered image subtracted from original green channel and the result multiplied to increase contrast	SIFT/SURF features of L*a*b colour channel	AMD severity
GRIVINSEN ET AL 2013 [20]	407(EUGENDA)	TOPCON TRC 501X 50° field of view Canon CR-DGi (non-mydriatic) 45° field of view	Drusen manually outlined	Each pixel in image assigned probability that it belongs to drusen candidate. Boundary of the candidate extracted using intensity and contrast characteristics	AMD severity
MOOKIAH ET AL 2014 [43]	161 (ARIA) 83 (STARE) 540 (KMC)	Carl Zeiss Meditec fundus camera 50 ° field of view (748 x 576) TOPCON TRV-50 fundus camera 35 ° field of view (700 x 605) TOPCON non-mydriatic retinal camera (TRC-NW200) (480 x 364)	CLAHE	Entropy features – Shannon, Kapur, Renyi, Yager Higher Order Spectra (HOS)	Wet/Dry/No Disease
MOOKIAH ET AL 2014 [42]	540 (KMC)	TOPCON non-mydriatic	CLAHE	Features for whole image obtained by discrete wavelet transform (DWT)	Wet/Dry/No Disease

		retinal camera (TRC-NW200) (480 x 364)		decomposition. Linear features extracted from wavelet coefficients (mean, variance, skewness, kurtosis, Shannon entropy, Renyi entropy, Kapur entropy, relative energy, relative entropy, entropy, Gini index).	
BURLINA ET AL 2016 [8]	5500 (NIH AREDS)	Not reported	Resizing and cropping images to conform to expected OverFeat input network	SURF, SIFT, wavelet features	AMD severity
PHAN ET AL [47] 2016	279 (Telemedicine Platform)	Zeiss, DRS, Topcon models 45° FOV (1400, 2,200,3240 pixels along diameter of image)	Pre-processing from [31]	Colour Histograms (RGB, L*a*b colour spaces) Texture - Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), SURF	AMD severity
ACHARYA ET AL 2017 [1]	945 (KMC)	Zeiss FF450 plus mydriatic fundus camera (resized to 480 x 360 from 2588 x 1958	CLAHE	Pyramid of histograms of Orientated Gradients (PHOG) to describe shape and pattern. Features from descriptor: Energy – uniformity of image Entropy features – approximate, fuzzy, Kolmogorov-Sinai, modified multiscale, Permutation, Renyi, Sample, Shannon, Tsallis and wavelet Nonliner features- fractal dimension (D), Hjorth (activity, complexity, mobility parameters), Kolmogorov	Wet/Dry/No Disease

				complexity, largest Lyapunov exponent, Lempel Ziv complexity, relative qualitative analysis (parameters entropy, transitivity, trapping time, recurrence of the 1st type and 2nd type, longest vertical line), Entropy, determinism, laminarity, maximal diagonal line length, averaged diagonal line length, recurrence rate, recurrence time of RQA parameters	
GARCIA-FLORIANO ET AL 2017 [18]	397 (STARE) 70 (RetinaGallery)	Not reported	OD located using [17]. Green channel.	Hu moments were used to describe each object as a measurable quantity calculated from the shape of a set of points	Disease/No Disease

Table 3.1 Summary of the articles determined as meeting the inclusion criteria for review. Table from [118]

3.4.3 Feature selection

As it is possible to extract many features from an image, it is necessary to select the best descriptors of the object(s) of interest that achieve the best classification performance. The feature selection stage involves removing potentially noisy or irrelevant features and identifies salient features that can be used to best distinguish between classes. Six of the articles from the literature search employed a feature selection stage. Zheng et al [140] used a L2 loss function to reduce noisy pixel intensity features. The output is a list of top features that were best for distinguishing images of *disease* and *no disease*. Garcia-Floriano [146] used a software package [147] to identify highly correlated features to *disease* and *no disease*. Mookiah et al [139][148] used parametric (e.g. t-test) and non-parametric tests (e.g. Wilcoxon ranking) to determine the top features for distinguishing *disease* from *no disease*. They reported a texture feature from a Gabor filter as the best performing feature. Acharya et al [141] used an algorithm inspired by ant behaviour and crossover operators in genetics to identify top features for classifying dry AMD and wet AMD from no AMD. They reported energy and entropy features as the top-ranking features.

3.4.4 Classification

The classification stage is the process of using the best features that represent the data to identify a model that separates the data into the desired classes. A dataset of images is typically separated into a training set used to develop the model and a testing set used to validate the performance against a reference standard. The evaluation of an algorithm's performance is commonly reported in terms of statistics such as receiver operating characteristic (ROC), accuracy, sensitivity and specificity (see section 4.3.4 for a definition of these terms).

3.4.4.1 Disease/no disease

Classification is the process of taking the input features and mapping to a discrete label such as *disease* or *no disease*. There are many classifiers available and the choice of which one to use depends on the nature of the data and its application. Table 3.2 summarises previous work used for classifying disease or no disease in fundus photographs [118]. Hijazi et al [138] used a case-based reasoning (CBR) technique to classify images containing disease (AMD) or no disease from histogram-based features. CBR is an automated reasoning and decision-making process whereby new problems are solved using experience accumulated from solving previous problems. Hijazi et al [138] used a 2-step CBR system whereby histogram features from the green channel with the blood vessels removed was the first case and the second case consisted of the same image but with the removal of the OD. The model was built by

storing the training image features for classes of AMD and no AMD so that when an image was passed into the CBR system it would compute the similarity between the learned AMD and no AMD features. The output was whether the input image was similar to AMD or no AMD where the highest similarity would be the prediction for the image. They reported a 75% accuracy for correctly classifying AMD images. Burlina et al [145] proposed a constant false alarm rate (CFAR) algorithm that is used in radar systems to distinguish true signals from noise to determine the origin of the signal. The CFAR system used colour features to train a support vector machine (SVM) classifier. An SVM classifier is a form of regression whereby features are projected into a higher dimensional space so that classes are linearly separable. They reported a positive predictive value of 97% and a negative predictive value of 92% at classifying images that contain *disease* (i.e. pathology that may indicate an unhealthy retina) or *no disease*. Zheng et al [140] proposed an algorithm that represented training images as a hierarchical tree where the child and parent nodes were derived from colour and intensity features of the image. They reported best classification using an SVM classifier that achieved a 99.4% sensitivity. An SVM classifier was also used by Garcia-Floriano et al [146] to classify images as *disease* or *no disease* from pixel intensity features and reported a 92% accuracy.

Reference	Images with disease (dataset)	Images with no disease (dataset)	Classifier	Reference Standard	Performance
Hijazi et al [138]	86 (ARIA)	56 (ARIA)	Case Based Reasoning (CBR)	Labels from ARIA project	ACC = 75% SEN = 82.00% SPEC = 65.00%
Burlina et al [145]	39 (private)	27 (private)	Constant False Alarm Rate (CFAR)	Graders from JHU Wilmer Eye Institute	SEN = 95% SPEC = 96% PPV (positive predictive value)= 97% NPV (negative predictive value) = 92%
Zheng et al [140]	101 (ARIA) 59 (STARE)	60 (ARIA) 38 (STARE)	Naïve Bayes, SVM	Labels from dataset	SPEC = 100% SENS = 99.4% ACC = 99.6%
Garcia-Floriano et al [146]	34 (STARE) 33 (RetinaGallery)	41 (STARE) 37 (RetinaGallery)	SVM	Labels from STARE and RetinaGallery	ACC = 92.1569% Precision = 0.904 Recall = 0.922 F-measure = 0.921

Table 3.2 Summary of previous work that has applied machine learning techniques to classify fundus photographs as disease (AMD) or no disease (no AMD) [118]. ACC, accuracy; SEN, sensitivity; SPEC, specificity; PPV, positive predictive value; NPV, negative predictive value. Table from [118].

3.4.4.2 AMD severity

The AREDS grading protocol used drusen size, count and type to define AMD severity into four classes [134]. Table 3.3 summarises classification of AMD severity in fundus photographs [118]. Phan et al [143] proposed a machine learning system to classify a dataset of fundus photographs into AREDS classes consisting of class 1 {1} for healthy images, class 2 {2} mild AMD (containing small drusen <63μm), class 3 {3} moderate AMD (containing drusen between 63μm and 125μm) and class 4 {4} advanced AMD (containing many drusen with geographic atrophy). They employed a "bag of words" algorithm that used the most salient features in the image and counts the frequencies and bins them into a histogram. The so-called visual vocabulary consisted of colour features where they used an SVM to classify the images into the AREDS classes. They reported a best performance for classifying {1} vs {2} vs {3} vs {4} at an accuracy of 62.7%. Similarly, Kankanaballi et al [142] exploited colour features to generate a visual vocabulary from images of high quality and low quality (i.e. images containing artefacts or poor lighting) and different numbers of images in the training set of each AMD class. They reported a highest accuracy of 98.9% for classifying images into AMD classes from a classifier that was trained from images of high quality with a balanced number of images per class. This accuracy decreased to 96.1% as low-quality images were included. Grivensen et al [144] proposed a system to quantify drusen area, size and location to classify images as low risk or high risk AMD. This involved segmenting drusen using Gaussian filters of the green channel and K-nearest

neighbours to determine whether a pixel belonged to drusen or not. From the segmentation they derived drusen area and diameter and compared this to the manual measurements from two human observers. From the segmentation, colour and texture features were extracted to train a random forest classifier that classifies a high or low risk images. They evaluated the algorithm according to its agreeability between the observers using intraclass correlation coefficients (ICC) and reported an ICC of 0.69 for drusen area and diameter and an ICC of 0.95 for AMD image classification.

Reference	Number of images in AMD severity category	Classifier	Reference Standard	AMD category Test	Performance
Kankanaballi et al [142]	EIPC: • 626 (category 1) • 89 (category 2) • 715 (category 3) • 715 (category 4) MIPC: • 626 (category 1) • 89 (category 2) • 1107 (category 3) • 950(category 4) MS: • 180 (category 1) • 13 (category 2) • 114 (category 3) • 78 (category 4)	Random Forest	Expert Grader	(1) {1 & 2} vs {3 & 4} (2) {1 & 2} vs {3} (3) {1} vs {3} (4) {1} vs {3 &4}	EIPC: 95.4% (SPEC) 95.5% (SEN) 95.5% (ACC) MIPC: 91.6% (SPEC) 97.2% (SEN) 98.9% (ACC) MS: 98.4% (SPEC) 99.5% (SEN) 98.9% (ACC) EIPC: 96.1% (SPEC) 96.1% (SEN) 96.1% (ACC) MIPC: 95.7% (SPEC) 96.0% (SEN) 95.9% (ACC) EIPC: 98.6% (SPEC) 95.7% (SEN) 97.1% (ACC) MIPC: 96.3% (SPEC) 96.8% (SEN) 96.7% (ACC) EIPC: 96.0% (SPEC) 94.7% (SEN) 95.4% (ACC) MIPC: 95.4% (SPEC) 97.7% (SEN) 97.1% (ACC)
Grivinsen et al [144]	Set A: • 17 Observer 1 , 20 Observer 2 (No AMD) • 13 Observer 1 , 9 Observer 2 (Early AMD) • 22 Observer 1 , 23 Observer 2 (Intermediate AMD)	K-nearest Neighbour Linear discriminant classifier Random Forest	2 Observers	Drusen Area: Observer 1 vs Algorithm Observer 2 vs Algorithm Interobserver	0.91 (ICC) 0.86 (ICC) 0.87 (ICC) 0.66 (ICC) 0.69 (ICC)

	Set B: • 216 Observer 1 , 218 Observer 2 (No AMD) • 64 Observer 1 , 64 Observer 2 (Early AMD) • 75 Observer 1 , 76 Observer 2 (Intermediate AMD) Average number of drusen: • 130.4 ± 178.1 (Observer 1), 198.5 ± 243.1 (Observer 2) Average size of drusen (μm^2): • 5,873 ± 10,027 (Observer 1), 5115 ± 8257 (Observer 2)			Drusen Diameter: Observer 1 vs Algorithm Observer 2 vs Algorithm Interobserver Risk Assessment: Observer 1 vs Algorithm Observer 2 vs Algorithm	0.79 (ICC) 0.84 (Observer SEN) 0.96 (Observer SPEC) 0.948 (Algorithm AUC) 0.765 (Kappa) 0.85 (Observer SEN) 0.954 (Observer SPEC) 0.954 (Algorithm AUC) 0.760 (Kappa)
Phan et al [143]	Good Quality: • 50 (category 1) • 43 (category 2) • 24 (category 3) • 22 (category 4) Poor Quality: • 29 (category 1) • 36 (category 2) • 41 (category 3) • 34 (category 4)	SVM & Random Forest	2 graders	{1} vs {2} vs {3} vs {4} {1&2} vs {3} vs {4} {1} vs {2&3} vs {4}	SVM: 62.7% (ACC) Random Forest: 61.7% (ACC) SVM: 75.6% (ACC) Random Forest: 74.2% (ACC) SVM: 72.4% (ACC) Random Forest: 69.9% (ACC)

Table 3.3 Summary of applied machine learning techniques to classify fundus photographs into AMD severity categories [118]. ACC, accuracy; SEN, sensitivity; SPEC, specificity; PPV, positive predictive value; NPV, negative predictive value; AREDS, Age Related Eye Disease Study; EIPC, equal number of images; MIPC, maximum number of images per class MS, manually selected images.

3.4.4.3 Wet/dry/no disease

Table 3.4 summarises previous work used for classifying *wet* or *dry AMD* from *normal* in fundus photographs [118]. Mookiah et al [139][148] used multiple features derived from the CLACHE enhanced images pixel intensities to train an SVM classifier to classify between fundus photographs of dry AMD and no AMD. They reported best performance of 95%. Similarly, a SVM was used by Acharya et al [141] to detect wet AMD from dry AMD using gradient features. There was an imbalance in the number of wet AMD image (21 dry to 1 wet) and used oversampling of the minority class to balance the data. They reported an 85.1% accuracy for detecting wet AMD from dry and normal images. Both of these proposed systems did not require any retinal landmark suppression.

Reference	Images with No-disease (dataset)	Images with AMD(dataset)	Classifier	Reference Standard	Performance
Mookiah et al [43]	101 (ARIA) 36 (STARE) 270 (KMC)	60 (ARIA) 47(STARE) 270 (KMC)	Naïve Bayes, K-nearest Neighbours, Decision Tree, Probabilistic neural network, SVM	Ophthalmologist Group	ACC (ARIA) = 95.07% ACC (STARE) = 95.00% ACC (KMC) = 90.19%
Mookiah et al [42]	270 (KMC)	270 (KMC)	Naïve Bayes, K-nearest Neighbours, Probabilistic neural network, SVM	Ophthalmologist Group	ACC = 93.70% SEN = 91.11% SPEC = 96.30%
Acharya [1]	404 (KMC)	517 Dry AMD (KMC) 24 Wet AMD (KMC)	SVM	Ophthalmologist Group	ACC = 85.12% SENS = 87.2% SPEC = 80%

Table 3.4. Summary applied machine learning techniques to classify fundus photographs into AMD severity categories [1]. ACC, accuracy; SEN, sensitivity; SPEC, specificity.

3.4.5 Machine learning for UWF-SLO classification

3.4.5.1 Materials and methods

22 early AMD (17 right, 5 left), 62 older eyes (30 right, 32 left) and 40 younger eyes (19 right, 21 left) UWF images were used in this thesis to train a classifier to distinguish the images into their categories. Images were captured using an Optos UWF-SLO (Daytona, Optos plc, Dunfermline, UK) and shared by Optos. This represents a challenging dataset to classify for three reasons. The first is that the appearance or texture of the retina looks similar for older eyes and early AMD. In the ageing retina there is a decrease in the abundance of RGC's and photoreceptors [149]. In a UWF image and fundus photograph this is perceived as varying levels of colour and texture. As AMD risk increases with age this texture would also be present therefore it would be challenging to distinguish between early AMD and a retina that is ageing healthily. Second, in a younger eye there is an abundance of RGC's in the posterior retina that are bright in appearance and could resemble drusen. Finally, the variation in image quality is often taxing as an older individual may not be able to open their eye sufficiently wide enough at image acquisition leading to eyelash and eyelid artefacts in the image. Additionally, older participants may not be able to sit in an optimal position at the device which causes bright reflection artefacts that occur in the centre of the image. In this dataset there is a variation of image qualities in each group (Figure 3.11).

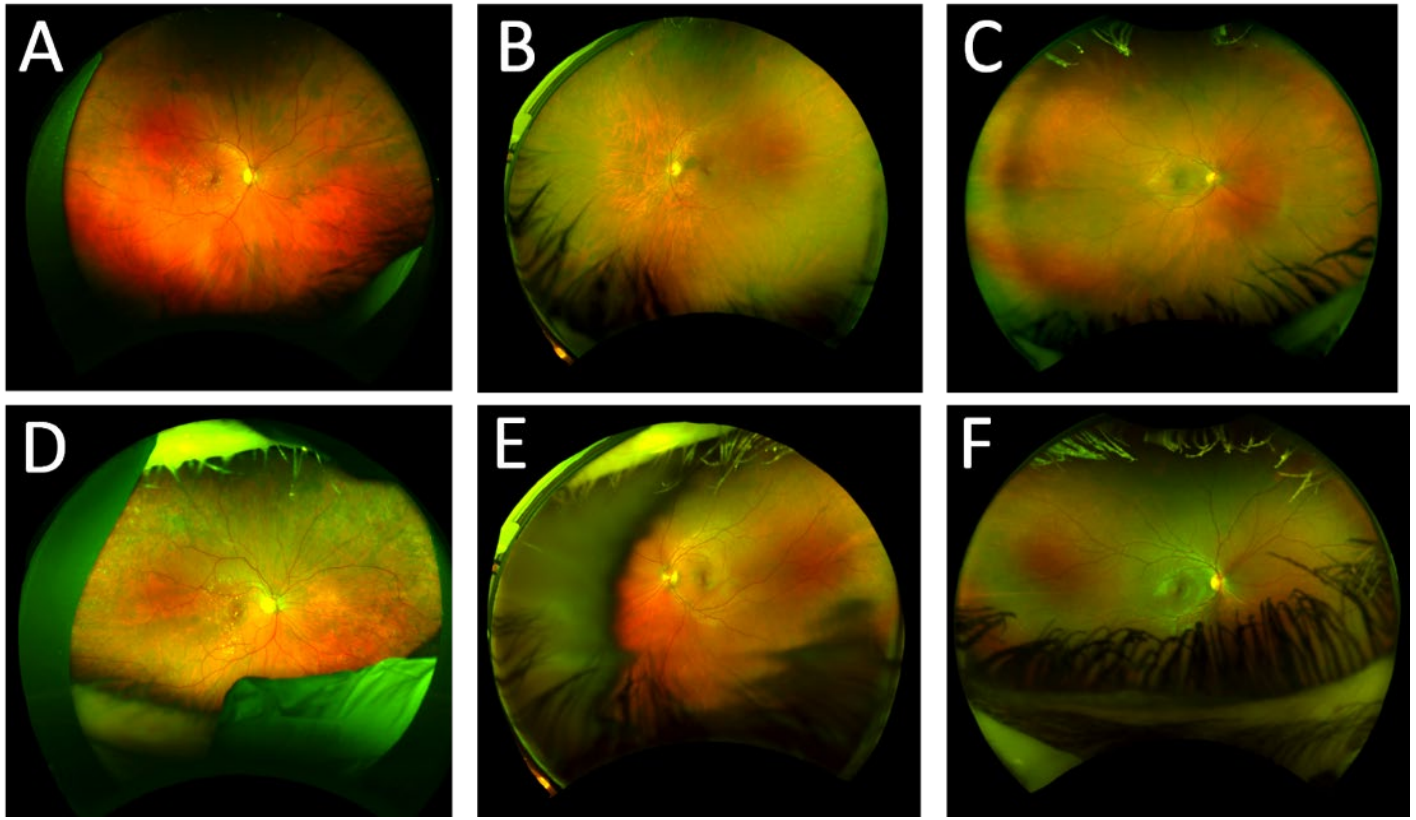

Figure 3.11 Representative examples of UWF images in each group. The top row of images is considered to be good quality as there is a large FOV with minimal eyelash and eyelid obstruction. The bottom row of images is considered to be poor quality because of the eyelash and eyelid obstruction. **A)** Right eye of individual with early AMD, clusters of hard and soft drusen are present near the macular. **B)** Left

eye of an older individual without AMD that contains temporal tessellations (thin choroid). **C)** Right eye of a younger individual that does not contain drusen but bright healthy RGC 's surrounding the macula. **D)** Right eye of an individual with early AMD and of poor quality. There are numerous clusters of hard and soft drusen from the central retina towards the periphery. **E)** Left eye of an older eye of poor quality, much of the image is obscured by artefact. **F)** Right eye of a young individual that is poor quality due to the obstruction of the eyelashes and lid that covers a large portion of the inferior retina.

The dataset in Phan et al [143] consisted of four severities of AMD (see section 3.4.4.2 for definitions of the AREDS class) - where class 1 and class 2 is a similar problem to older eye vs early AMD eyes, and trained the classifier using high quality and low quality images. This gave good results using a bag of words feature extraction method with a Gaussian kernel SVM classifier.

In section 3.3.2 an image histogram was defined using the frequency of pixel intensities in the image, likewise a histogram can be defined using the features contained within an image, called a visual vocabulary. First, the training image was divided into a grid (with overlap) and Speeded Up Robust Features (SURF) was extracted from the image within each grid cell [150]. Next, the extracted features were clustered using K-means (a method used to group data points into a specified number of clusters) whereby the centre of each cluster is a visual word. The frequency of occurrence of each visual word was obtained for each training image and defined each image histogram (i.e. the visual vocabulary). A KNN (see section 3.4) and SVM classifier was trained using the visual vocabulary as features. An SVM classifier is a well-established technique that projects features into a higher dimensional space so that classes are separable (i.e. finds a decision boundary that maximises the space between data points closest to the boundary). The function that transforms the features into the new space and constructs the SVM is called a kernel where there exist many types, such as linear, radial basis function kernel (or gaussian kernel), polynomial and sigmoid, where choice of kernel depends on the nature of the data.

Given that Phan et al [143] had good results with a similar dataset, a bag of features algorithm was applied to extract features from the UWF images. Next, a SVM and KNN classifier was trained and evaluated on UWF images of 8 early AMD (4 right, 4 left), 15 older eyes (7 right, 8 left) and 10 younger eyes (5 right, 5 left). The value of the hyperparameters grid size, overlap, number of K-means clusters and kernel method can be determined by iteratively tuning the hyperparameters and evaluating the classifiers performance. In this thesis, a high performance was obtained in the first iteration using a grid size 32 x 32-pixel, 8 x 8-pixel overlap, 500 clusters for K-means and a Gaussian kernel.

3.4.5.2 Results

The evaluation metrics of accuracy, sensitivity and specificity (see section 4.3.4 for a definition of these terms) were used to assess the performance of the classifiers. Table 3.5 displays the results for the testing set. SVM achieved the highest sensitivity (88.9 %) and specificity (100%) compared to the KNN classifier. The KNN classifier had 1 false positive that predicted an older eye to be early AMD and 1 false negative that predicted 1 image as an older eye when the true label was early AMD. Figure 3.11 shows the false positive and false negative UWF images for the KNN classifier. The false negative could have arisen due to the low abundance of drusen present in the image. The false positive could have occurred due to the bright eyelashes in the image that could be interpreted as bright drusen. Although the image was not labelled as early AMD the image contained many hard drusen in the central retina.

Classifier	Sensitivity	Specificity	Accuracy
KNN	77.8%	95.8%	90.9%
SVM	100%	100%	100%

Table 3.5 Sensitivity, specificity and accuracy of the KNN and SVM classifiers (see section 4.3.4 for definitions of these metrics).

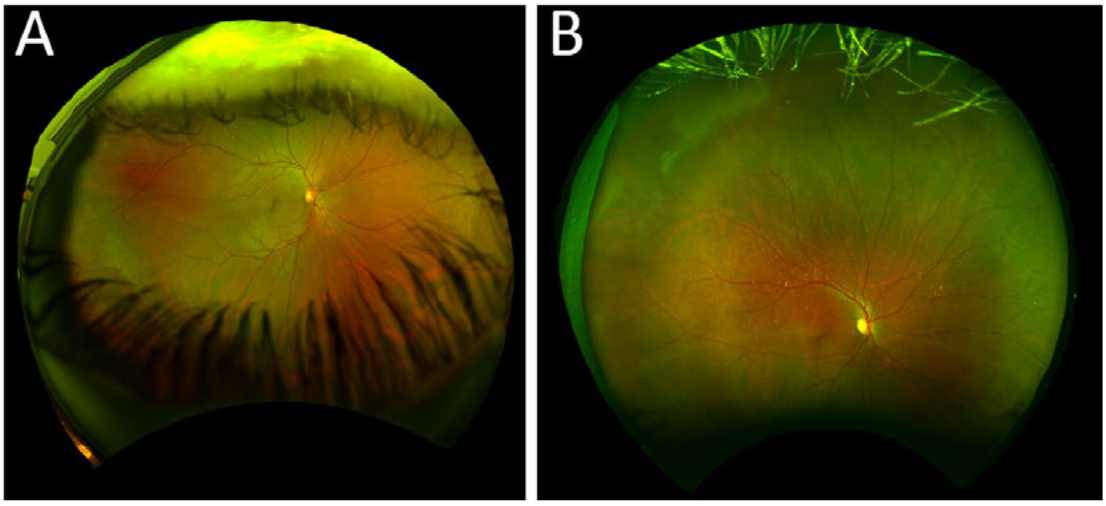

Figure 3.11 UWF images incorrectly classified by the KNN classier. **A)** True image label was older eye but predicted label was early AMD (false positive). **(B)** True image label was early AMD but predicted label was older eye (false negative).

3.5 Discussion

The machine learning methods presented in this chapter showed a tendency to use the green channel of a fundus photograph to extract features for training a classifier to discriminate images into the applicable categories. The most consistent features were texture-based and colour-based where multiple classifiers were often tested to determine the best model [143] [42] [42] [29] [140] . In some articles, retinal landmark suppression such as removal of the blood vessels or OD was required to improve performance [138] [143] [146]. In the articles that aimed to classify AMD severity the most difficult classes to distinguish were mild to moderate AMD. In the clinic, distinguishing between these two classes requires drusen size or load, measurements that could not be obtained without segmentation of drusen. There was a common trend of image-level classification. Only one article quantified individual drusen [144]. Image-level classification of fundus photographs could be used for large scale screening to identify individuals who may be at risk of developing severe AMD. Whereas segmentation of individual drusen could be used to monitor subtle longitudinal changes. However, without a sufficiently large dataset containing hand drawn annotations of drusen boundaries, it would be difficult to train and validate a system with an acceptable performance.

In this thesis, KNN and SVM classifiers achieved good performance for classifying images into early *AMD*, *older eye* and *younger eye* categories. False positives occurred due to the objects appearing in images which looked similar to drusen as

well as the resemblance between older eyes and eyes with early AMD. A false negative with the KNN classifier identified an image that contained drusen. Machine learning techniques are trained on examples that are assumed to be representative of the population. However, new images may not be similar to the training images and may result in misclassification. Machine learning algorithms require hand crafted features which may not generalise to the whole population (so-called overfitting). The results of the literature search identified a recent trend towards deep learning techniques that have been developed to address the challenges that conventional image processing and machine learning (e.g. overfitting) present and is described in Chapter 4.

3.6 Conclusions

Conventional image processing can be used to segment drusen but there is a shortage of a reference standard to validate the algorithm against. Without a quantitative measure of an algorithms accuracy, interpretation of any subsequent measures may not be reliable. Machine learning is a valuable method for image-level classification and yields good results on the UWF image testing set. However, these methods often fail to generalise to datasets that are different and or more complex. A trend towards deep learning methods after 2015 was observed in my evaluation of the literature, which is reflected in the current state-of-the-art for many areas of medical image processing. Deep learning and its application to drusen detection is the subject of the next chapter.

Chapter 4 Automatic drusen detection in UWF images

4.1 Introduction

In this chapter, the development and testing of automatic drusen detection in UWF retinal images is described. First, I introduce the concepts of deep learning and transfer learning before a drusen detector is trained using transfer learning. This detector is then tested on its ability to classify image patches as containing *drusen* or *no drusen* and the results compared to the annotations of two human observers. The aim of this chapter is to develop a novel method to detect drusen in UWF images.

4.2 Deep learning and transfer learning

Deep learning has become state-of-the-art for computer aided detection in medical imaging [151][152]. There is a demand for quick, reliable and accurate interpretations

of medical data from a medical expert. In a healthcare setting, these interpretations can be limited by the subjective nature of the process, image complexity, high interpatient variability and human error. Deep learning is an advancement that has potential to provide solutions to these limitations by aiding the clinician with particular tasks with high accuracy, reliability and repeatability [153]. Unlike machine learning (as discussed in chapter 3), deep learning does not require any feature engineering but rather extracts features directly from data. First, I will discuss the building blocks of a convolutional neural network (CNN) and then I will describe how a CNN "*learns"*. Next, I will discuss the concept of transfer learning and the previous work that has been reported on fundus imaging (identified from the literature search presented in Chapter 3).

4.2.1 Deep learning

Deep learning was inspired by how humans process vision. Hubel and Weissel conducted pioneering studies that revealed how cells in the visual cortex of the brain are organised in layers where neuronal cells filter and respond (or fire) in the presence of different sensory information [154]. This process of accepting an input and applying an operation to output a response is the basis of a neuron that is the building block of an artificial neural network. The neurons are connected together (forming a network) and apply different operations to return an output that is a probability of the input belonging to a certain class. Figure 4.1 shows a mathematical neuron, labelled to illustrate the conceptual derivation from biology [155]. The value of a neuron is determined by the sum of the input weights with the addition of a bias

followed by a non-linear activation function which returns an output weight. An activation function is applied because the input could be any value. A sigmoid function is often used that scales this value between 0 and 1 [156]. Another popular method is an activation function called rectified linear units (Relu) that removes all negative values and changes them to zero [157]. The bias is a constant value that ensures the value of the neuron is never 0. All the weights, bias and outputs are learned during training of the neural network.

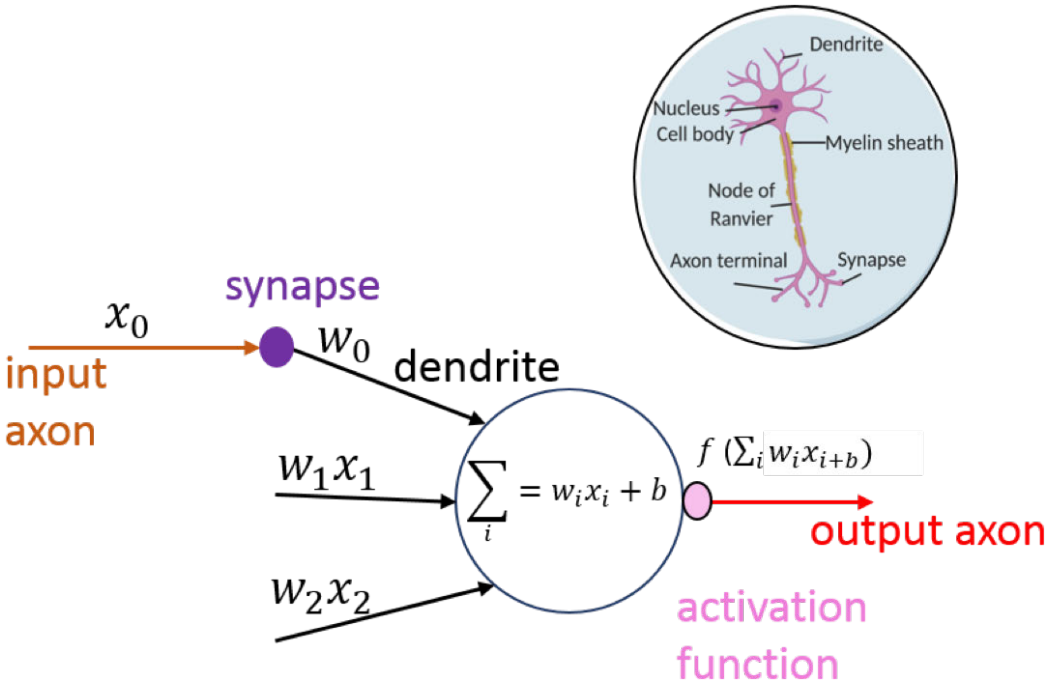

Figure 4.1 A mathematical neuron that accepts multiple inputs (w, weights) that are summed with the addition of a bias (b). An activation function is applied to scale the output. The weights, bias and outputs are learned during training.

The first layer in a neural network is convolution that gives a neuron its initial input weight. Convolution works by comparing images piece by piece (called a filter).

Convolution can be thought of as shining a spotlight of a certain size (i.e. a kernel) over an image (i.e. a sliding window) where the area that is illuminated (called the receptive field) is used to obtain information of that area. Figure 4.2 shows, as an example, classification of *noughts* and *crosses* [158] where a filter (i.e. features) already learned from the data (Figure 4.2B) are applied over the image that multiples the pixel in the filter with the corresponding pixel in the image (Figure 4.2B and C). The value is divided by the number of pixels in the filter returning a weighted sum that is stored in an activation map (Figure 4.2D). In this example, a diagonal kernel detects features with similar shape, and we can see this feature in the activation map. Figure 4.2E shows this process in terms of a neuron where the value of the filter response for each pixel is the weighted sum of inputs into a neuron. For every position of the filter over the image a value is returned. The value of the weights is learned during training.

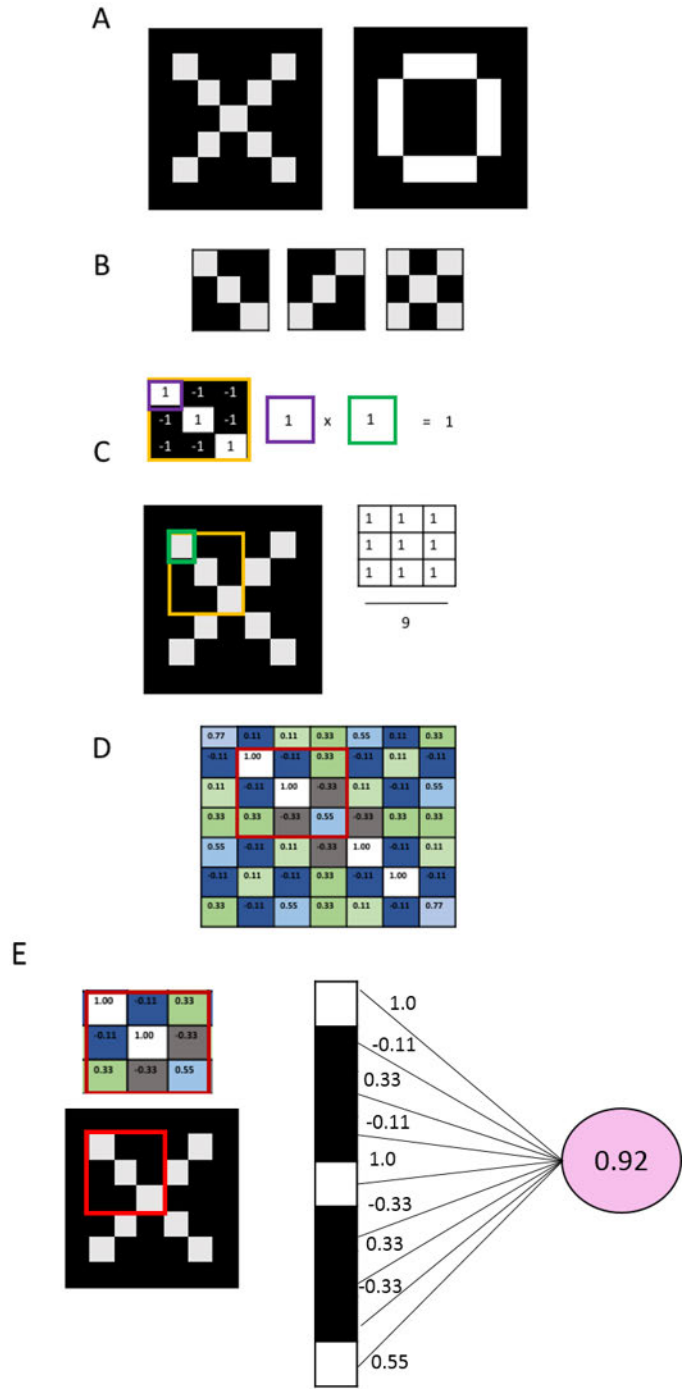
A
B
C
1 -1 -1
-1 1 -1
-1 -1 1
1 x 1 = 1
1 1 1
1 1 1
1 1 1
9
D
E
1.00 -0.11 0.33
-0.11 1.00 -0.33
0.33 -0.33 0.55
1.0
-0.11
0.33
-0.11
1.0
-0.33
0.33
-0.33
0.55
0.92

Figure 4.2 Example of convolution on *noughts* and *crosses*. **A)** Image of either a *naught* or *cross* is input into the network that **B)** has learned features for classifying between the two. **C)** For example, the diagonal feature is compared to a section of the input image. The value in the same position in the feature (purple) is compared to the value in the same position on the input image (green). These corresponding feature pixels are multiplied over the whole feature and divided by the size of the feature. **D)** In this case, the value of the diagonal feature equals 1.00 and is recorded (filtering). This is then applied over the whole input image (convolution) producing a feature map. The activation of the diagonal feature can be seen in the activation map. **E)** This process is visualised in terms of a neuron. The value of the weights is from the weighted sum of the input with an activation applied and are stored as a neuron. In this case, we can see that white pixels have a large weight and results in the high activation value of the neuron. Effectively, there is a single neuron for every position of the filter on the input image. Concept and illustrations adapted from [158].

Usually the aim of a neural network is to classify inputs into predetermined classes. This may be assigned from a probability of an input belonging to a certain class. To progressively (i.e. through several convolutional layers) reduce the size of the output of a layer to a probability, pooling can be applied [159]. This involves walking a window of a specified size over the output of a layer and applying an operation. There are many operations that can be used, such as averaging the value of the window (average pooling) or taking the maximum value of that window (max pooling) [160].

This results in a stack of smaller activation maps that still summarises the features of the data. By using repeating combinations of convolution, activation and pooling layers the output of each layer gets smaller and smaller [161].

The final layers are usually fully connected (FC). The output of the final layer is an N-dimensional feature vector, where N is the number of classes the network is classifying [161]. In the example of *noughts* and *crosses* N would be 2 since there are 2 classes we are trying to classify. In Chapter 3, hand crafted features from the image (e.g. histogram-based features) were input into a classifier (such as an SVM) to categorise images into classes. Convolutional layers, in combination with activations and pooling layers, serve the same purpose as a feature extractor where the purpose of the FC layer is to interpret all these features into an output probability of an input belonging to a certain class. The FC layer introduces non-linearity by combining every output of the previous layer (i.e. “fully connected”) and applies an activation function to return an output probability. Many types of functions exist, but most common are sigmoid (binary classification) or softmax (multi-class classification), each aim to redistribute the output of the FC layer into a probability distribution (i.e. values between 0 and 1).

Figure 4.3 summarises all of the layers and components of a basic neural network that have been discussed so far. In this example, the input is a *cross* that is passed through the convolution layer where filters are applied resulting in a stack of activation maps. In this simple network the activations are input into a pooling layer that results in stacks of smaller activation maps and this is repeated for another

convolutional and pooling layer. In the first convolutional layer, low level features are learned such as contrast or colour and edges. In the next convolutional layer high level features such as shape and contextual information is learned. This is because the activation maps of the input are progressively convolved to smaller images where more detail is preserved in earlier layers and become coarser in later layers. Finally, the activations of the previous pooling layer are all connected to the neurons in the FC layer that flattens the 3D volume into a 1D feature vector. In this example, the desired output is two classes, the value of which will be the weighted sum of the previous layer and is input into a softmax activation function to output a probability. When a *cross* is fed into the network, in the final FC layer there will be certain values that will be high and would contribute to a higher output probability for the input belonging to the *cross* class.

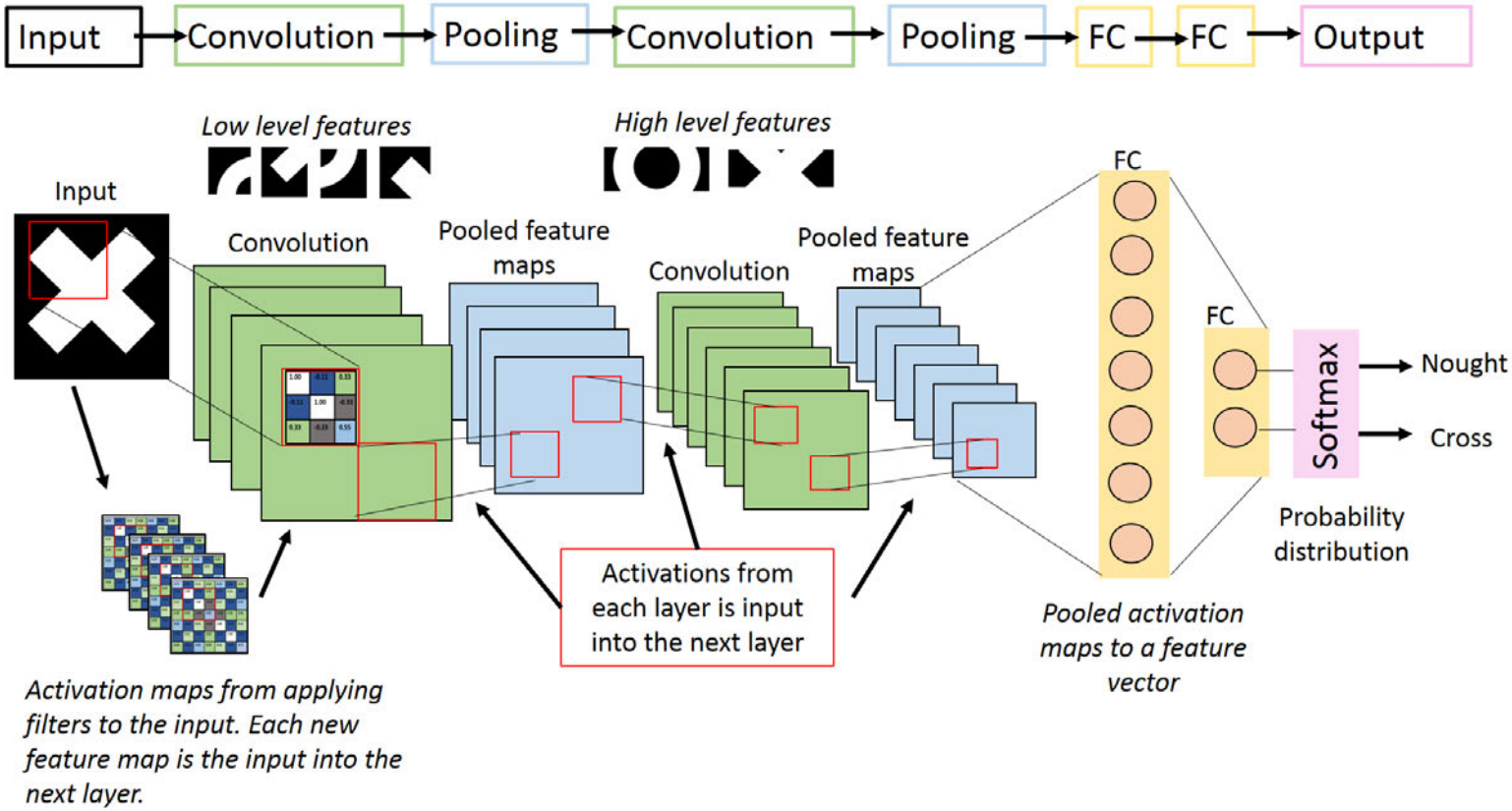

Figure 4.3 Components of a CNN for classifying *noughts* and *crosses*. A *cross* is input into the network and convolution is applied. The activations of a filter are output as an activation map and input into the next layer. This first layer learns low level features such as

colours and edges. The activations from this layer are input into a pooling layer followed by another convolutional layer. This is repeated where the output of the last pooling layer is flattened to a vector of neurons. The final FC layer consists of two neurons (corresponding to the tow classes). Finally, a softmax function is applied to redistribute the values to probabilities. This represents the likelihood (between 0 or 1) that the input image belongs to each of the classes.

How a neural network learns the values of the filters, weights and biases is called backpropagation and can be subdivided into 4 processes, the forward pass, the loss function, the backwards pass and weight update [162].

When a network has been constructed, all values of the weights, biases and filters are set to random values (or initialised using a pretrained network, see Section 4.2.2). During the forward pass a training image is passed through all of the network layers and will output a probability. For example, if we placed a *cross* as the input image, the output of the FC layer could be 0.01 when we know that the output should be 1 (i.e. a high probability that it is a *cross*). This gives an error and how it is computed is called the loss function. This function computes how well the model makes predictions given the current input compared to the correct output [163]. There are many loss functions and the choice of method largely depends on the data, presence of outliers and choice of learning algorithm [164]. Cross entropy is a popular loss function where the cross-entropy loss of prediction scores Y and training target T for number of observations (N) and number of classes (K) is given by,

$$L = -\frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{K} T_{ni}\log(Y_{ni}). \qquad \textbf{(4.1)}$$

This calculates the performance of the network and returns a probability between 0 and 1 for each of the classes. Also called the log-loss, the value of the error increases as the predicted probability (Y) diverges from the true label (T). For example, a predicted probability of 0.01 when the true label is 1 would output a high loss value (L). A model with a perfect score would output a loss of 0. During training the loss is monitored, the higher the loss (i.e. the error) the further we are from the correct

answer. A decrease in loss during training is an assurance that the model is fitting and the network is learning.

Once the error is computed a backwards pass is performed where the weights that contributed to the error are adjusted so that the error is minimised. This is achieved using gradient decent [165], an optimisation strategy that finds the best value of each weight that minimises the error and subsequently updates the value. Figure 4.4 explains the concept of gradient descent. As images are fed through the network, features and weights are constantly adjusted until patterns that occur in the data are learned. The size of this adjustment is called the learning rate. If you have many images this pattern is stabilised and works across a variety of images.

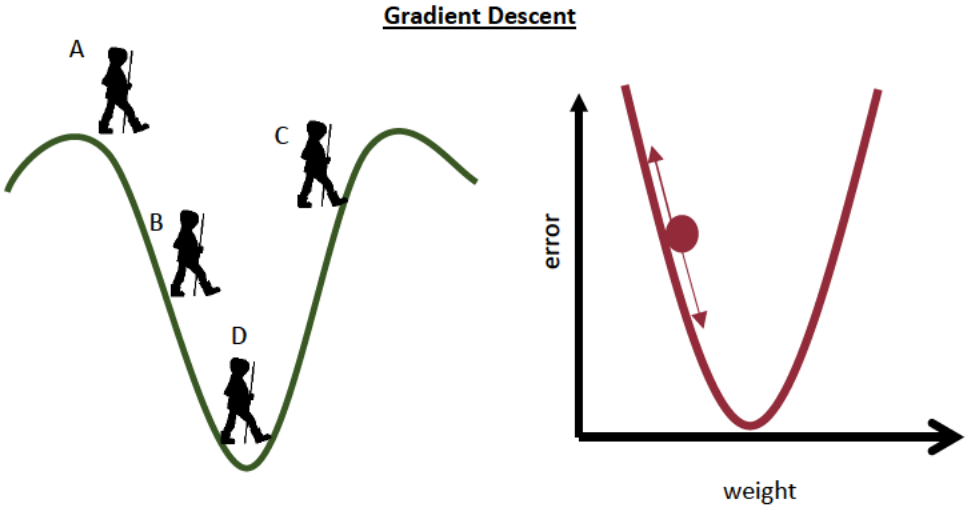

Figure 4.4 The concept of gradient descent [165]. **A)** Imagine a backpacker on a hill who's aim is to reach the bottom. They do not have a map so need to look in all directions (left, right, forward, back) to find where they can descend. **B)** They take a step in the direction they think will take them lower and see that have indeed descended. **C)** They repeat this step continuously until they begin to ascend again. **D)** Realising their previous step was lower, they return to this step. Just as the backpacker adjusts their position to minimise height on the hill so too does a loss function to the value of the weight to minimise the error. The size of the backpacker's steps down the hill and the size of the adjustments of the weight is called the learning rate. The value of the weight is therefore as close to the correct answer as possible and patterns in the data are learned.

Combinations of convolutions, pooling, activations, learning parameters and optimisation strategies form the architecture of a deep learning neural network and

exist in a variety of forms. Over the years, more complex operations have been developed to model larger and more complicated datasets as they become available. Arguably, the driver of new network architectures is the benchmark ImageNet Large Scale Visual Recognition Challenge (ISLRVC) [166]. The challenge was to build a network that correctly classifies millions of natural images into 1000 categories. Early ImageNet winners produced networks such as AlexNet [161] and VGG [167] that were comprised of repeating convolution, normalisation and activation architecture. However, efforts to make these styles of networks *deeper* resulted in a new style of networks; Inception V2 and Inception V3 [168]. So-called due to their inception layers that deploy multiple convolutions, filters and pooling layers simultaneously and in parallel to the same layer. Recently, architectures containing residual layers have been developed whereby outputs from blocks of convolutions and normalisations are re-routed to the previous layer. This allows deeper networks to be trained such as ResNet [169] and can be combined with inception modules such as in Inception-ResNet [168]. Xception is a newer architecture that is similar to an Inception model but maps multiple convolutional layers using residual connections [170]. More recently, networks have been used for semantic segmentation (assignment of every pixel in an image to a class) that follow an encoder (the pre-trained network) and decoder structure. A decoder semantically projects the features learned by the encoder back into the pixel space to obtain a per pixel classification. SegNet is a popular network [171] for semantic segmentation and is a VGG network (the encoder) followed with a decoder layer. The caveat is that these networks require large amounts of detailed pixel level annotations, which is difficult to obtain

especially in a healthcare setting due to time-constraints and data governance issues such as patient anonymity.

4.2.2 Transfer learning

Transfer learning has become a popular method, particularity in medical image analysis, whereby a pre-trained model is reused for a new task, requiring only small amounts of data [151]. In early layers of a neural network basic features such as colour and shapes are learned (see Figure 4.3). In transfer learning the early layers and their weights can be used as a feature extractor and new FC layers are trained for a new purpose (so-called fine tuning). This is on the basis that early learned features will be similar across imaging domains. For example, the colour green is still the colour green in the training set and the transfer learning set. The number of layers that are used as feature extractors (i.e. the number of trainable layers) can also be experimented with to optimise the network.

As neural networks apply operations with a specific size parameter, input images must also be the same size. Typically, pre-trained networks have input sizes of 224×224 pixels, 299×299 pixels and 512×512 pixels. Such predefined sizes have arisen due to the memory and computing power that is required to perform multiple and complex operations on an image. The larger the image, the more computational power is needed to learn features, as well as a deeper network. The first step in training a network involves resizing all images. Image normalisation is performed to make all images have the same data distribution. There are three common

techniques that involve either normalising values of the image (1 to 255) from 0 to 1, or values from -1 to 1 or to centre values around mean. The normalisation technique and the input size adopted depends on the data used to originally train the network. In Figure 4.4, the gradient decent algorithm is *searching* to minimise a loss function to find the optimum value of the weight that minimises the error and was shown in 2-dimensions because this can be easily visualised. In practice, the value of all weights must be made in the many dimensional space of weights and biases. This can be visualised in 3-dimensons where the x-axis is the value of weight 1 in layer 2 and the y-axis is the value of weight 2 in layer 3. The z-axis is the value of the loss for a value of the two weights. The aim is to find the values of the weights at the minimum of the loss function and uses gradient descent to search for the minima, as before. However, if there are large amounts of variation within the data this can lead to differently varying shape of the loss surface between dimensions. For example, weight 1 on the x-axis could vary on a scale of 0 to 1, while weight 2 on the y-axis could vary on a scale from 0 to 0.001. Figure 4.5 shows the loss surface shape as a contour plot next to the visual representation of the gradient descent search along a loss function to visualise the reasoning for image normalisation upon input into a neural network. It shows an uneven loss topology due to the different scales on the x -axis and y-axis and the arrows indicate the search direction to reach the optimal weight value (*asterisk*). In Figure 4.5B we can see how this looks in relation to updating the weight to minimise the error. The drastic updates in the weight value due to the large differences between the weights has caused an overshooting of the optimal weight value. When the x-axis and y-axis are normalised to a range from 0 to

1, the weight parameter is updated in even proportions (Figure 4.5C) and the search along the loss function is less drastic and leading to a better convergence of the weight to optimum value (Figure 4.5D).

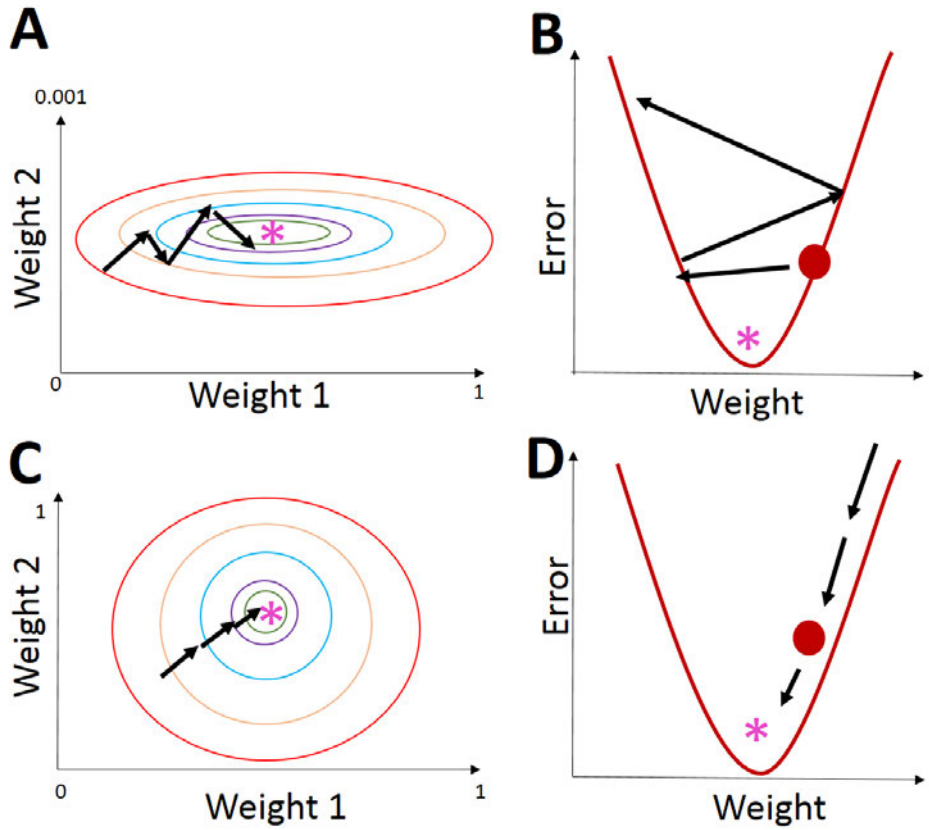

Figure 4.5 Visual representation of the concept of image normalisation upon input into a neural network in relation to the loss function, gradient descent and converging weight values to minimise the error in the loss function. The asterisk denotes the optimum weight value. **A)** Loss surface as a contour plot with the search path of gradient descent (i.e. learning rate). The topology is elongated due to the different scales on the x-axis (0-1) and y-axis (0- 0.001). **B)** Visualisation of the gradient descent search path along the loss function. The drastic weight updates due to the uneven loss topology cause the optimum weight value to be overshot. **C)** Loss surface following image normalisation creates an even loss surface. **D)** Weights are updated in even proportions leading to convergence to the optimum weight value.

To utilise the already learned features from a pre-trained network the learned weights must be fixed, termed frozen. Any number of layers can be frozen but traditionally all weights in the convolutional layers are frozen and the weights of the FC layers are initialised from their trained weights. The ImageNet dataset is a popular dataset used to train many networks where the final FC layer contains a classifier for 1,000 categories [166]. This can be replaced with an FC layer of a specific size. Drop out layers are commonly placed between FC layers which randomly sets inputs to 0 with a given probability and prevents the network from overfitting [172].

In transfer learning the whole dataset is passed through the network and only the weights of the final FC layers are adjusted. Each pass of the dataset through the network is called an epoch and the dataset can be passed in batches to minimise computing time. The processing of a single batch is called an iteration. A portion of the dataset is usually held out as a validation set to test over a certain number of iterations or epochs. During training, the loss and other performance metrics such as accuracy can be monitored to assess neural network performance. Neural network design, training and optimisation is a cyclic process. This involves training the network, adjusting architectures, loss functions, learning rates and classification functions along with their hyperparameters and retraining the network. This process is used to achieve a network design that optimises learning on the dataset.

4.2.3 Previous work

Deep learning and transfer learning have proved successful in many medical images including fundus imaging [173][174][48][175] and OCT [176][177] and have rapidly become key methods for medical image analysis [152][151]. The following articles discussed in the following section were identified in the literature search described in section 3.2.3.

Tan et al [173] implemented a 14-layer CNN to classify 1,110 fundus photographs as *disease* or *no disease* and achieved an accuracy of 95.25%. Burlina et al [178] used transfer learning of the Overfeat CNN [179] to classify 5,600 fundus photographs into their AREDS categories [134]. They achieved a 92% to 95% accuracy where best performance was between class 1 and class 4 AMD severities. The same experiment was performed in their later work where the Overfeat CNN features were used to fine-tune an SVM classifier to classify 5,664 images [174]. They reported best performances of 79.4%, 81.5% and 83.4% to classify images into AREDS classes 4, 3 and 2 respectively. Grassman et al [175] used transfer learning of multiple ImageNet CNN's [180][181][161][168][170] to classify 120,656 fundus photographs into AREDS categories [134]. They built a so-called ensemble of CNN's where the predictions from each network were combined into a mean prediction. They achieved an accuracy of 92.1% for predicting each AMD class. They tested their ensemble on an independent dataset of 5,555 fundus photographs [136] and achieved an accuracy of 34%. The low accuracy was attributed to the presence of younger eyes in this dataset that had dominant macular reflexes (brightly appearing artefact from the flash of the

fundus camera) that was misclassified as stage 4 AMD (i.e. misclassified as containing GA). Peng et al [182] used an ensemble of 3 neural networks (called DeepSeeNet) designed to first classify 3 categories of drusen (*small/none*, *medium* and *large*) then a sub network to classify pigment abnormalities (*hypopigmentation* or *hypopigmentation*) and a final network to detect the GA (*neovascular AMD* or *central GA*). Their training and testing sets contained 58,402 and 900 fundus photographs respectively, with annotations for each of the classes from two retinal specialists. The overall aim was to address the challenge of classifying fundus images into the AREDS categories by detecting each pathology that defines the AMD classes. They achieved a 72% accuracy (71% sensitivity, 71% specificity) for classifying drusen subtypes, 89% accuracy (73% sensitivity, 95% specificity) for classifying pigmentary changes and 96% accuracy (62% sensitivity, 98% specificity) for classifying GA subtypes. The source of misclassifications came from image quality such as bright artefacts that were classified as GA.

Inspired by the direction of the field described here along with recent advances in deep learning and transfer learning and its success for solving medical image analysis tasks, a neural network approach for detecting drusen in UWF images was adopted for this thesis.

4.3 Materials and methods

4.3.1 Training dataset

For a human grader, manually identifying drusen is a subjective and challenging process, particularly in UWF images, as drusen are difficult to distinguish from the background texture of the retina. Training a neural network using ambiguous examples or single drusen (which could be one pixel in size) would be difficult because there would not be enough examples of distinguishing features making the problem too complex and near impossible for the network to solve. To simplify the problem, the network was trained using clearly recognisable examples of patches *positive* for drusen and *negative* for drusen that I selected. Such patches were obtained from images of people with early AMD where drusen are often clearer and higher in abundance than elderly people with AMD or those without disease.

The training set consisted of 25 patients with early AMD (14 left eyes, 15 right eyes) and 20 patients with age related drusen all acquired with a UWF SLO (Daytona, Optos plc, Dunfermline, UK) and diagnosed by ophthalmologists. Patient information such as gender, age and ethnicity were not available. All images were converted to a stereographically projected format to allow the curved retina to be displayed as a flat image using a tool provided by Optos [74]. From the training set, I selected *positive* image patches (513 total). A patch was considered *positive* if it contained unmistakable drusen. In order to obtain a balanced training set of *positive* and *negative* examples, 407 *negative* image patches were randomly sampled from 16

patients with age-related drusen and a further 106 manually selected from the 25 UWF images with early AMD (513 total). Figure 4.6 shows examples of *positive* and *negative* patches.

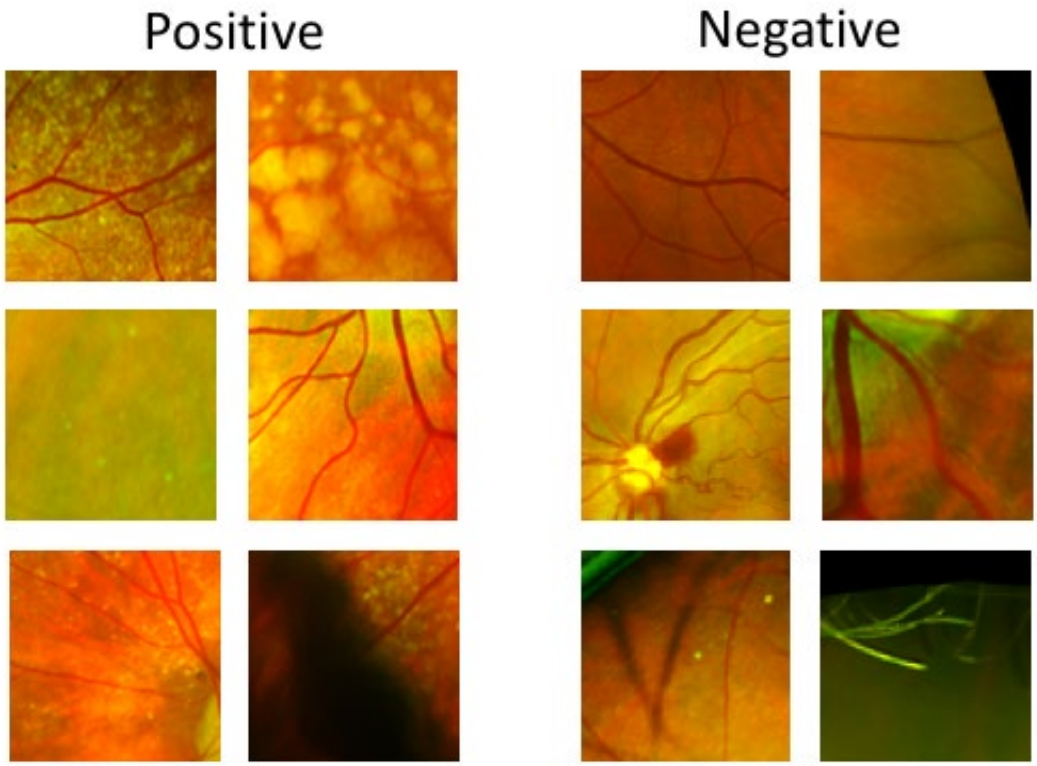

Figure 4.6 Examples of *positive* and *negative* patches. Patches were 122×122 pixels, 224×224 pixels or 299×299 pixels in size depending on the object being sampled (i.e. smaller single drusen in smaller patch or drusen clusters in larger patches) *Positive* patches included drusen clusters (top left) and larger forms of drusen (top left). Single druse (middle left) and single druse near vessels (middle right). Drusen also appear near the optic disc (OD) (bottom left) or are sometimes obscured by eyelashes (bottom right). *Negative* patches included healthy retina (top left) and healthy retina near the image border (top right). Regions near the OD where bright nerve fibre layer

(NFL) could also look like drusen (middle). Potential false positives such as specs of dust (bottom left) and eyelashes that have inadvertently been captured in the image (bottom right) were also sampled.

4.3.2 Training

VGG16 architecture, a well-established architecture for image classification [167], was fine-tuned using transfer learning (ImageNet weights [166]) by retraining the final FC layer. The VGG16 architecture consists of 13 convolutional layers (3×3 pixels), 5 max pooling layers (2×2 pixels) and 3 FC layers and was originally trained to classify non-medical images into 1,000 classes (such as cats, dogs, flowers etc) [167]. To avoid overfitting, a drop out layer (p = 0.5) [172] was inserted following the 4,096 neuron fully connected layers. I replaced the final 1,000 neuron fully connected layer with a 2-class softmax classifier (2 neurons) and trained using stochastic gradient descent with a cross entropy loss and a learning rate of 0.01. Figure 4.7 shows the VGG16 architecture that was modified for this thesis.

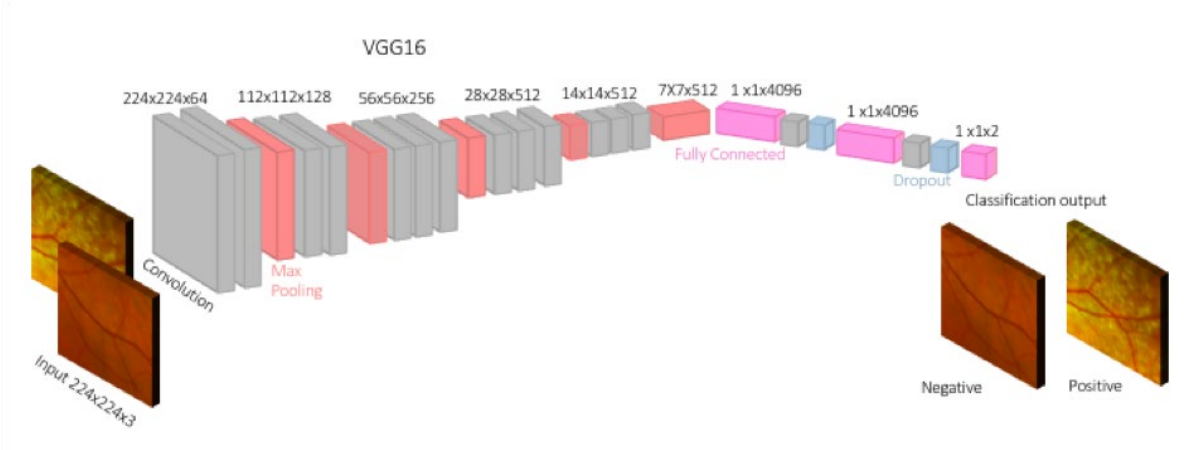

Figure 4.7 VGG16 Network Architecture. VGG16 accepts 224×224 pixels size images into the network. Convolutional layers (*grey layers*) transform the image into stacks of features. Max pooling shrinks the convolution by taking the maximum value in a 2×2 pixel sliding window on the convolution output. Fully connected layers (*pink layers*) outputs a two-dimensional vector that is the probability that the input was either *negative* or *positive* for drusen. To avoid overfitting, a drop out layer (*blue layers*) was inserted following the 4096 neuron fully connected layers.

Each image was resized to 224×224 pixels to conform with VGG16 required input size and normalised by subtracting the mean intensity of the training set from each pixel [161]. The network was trained for 100 epochs and was stopped when the loss no longer decreased. To increase training set size and variability, data augmentation was performed during training by applying rotation to the training patches at random intervals of 0 to 45° and shearing by 45°, separately (see Figure 4.8). Rotation augmentations are important especially in the case of patches containing vessels. If

there are a high number of patches containing horizontal vessels the network may learn vessels as a distinguishing feature. To minimise such scenarios, rotations provide different orientation of objects and shearing provide different distortions of the image and creates more examples for the network to learn from. This increased the training data to a total of 5000 patches. During training, the dataset set was randomly shuffled leaving 25% of the dataset (including augmented images) out for validation every 179 iterations.

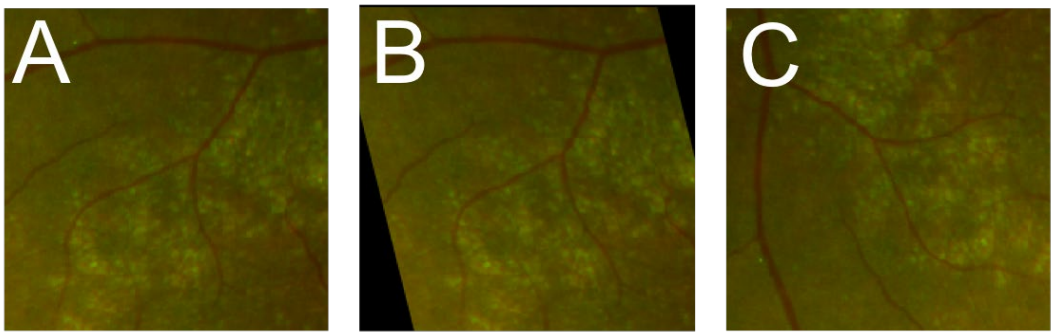

Figure 4.8 Example of different augmentations applied to an image patch containing drusen. Shear **(B)** and random rotations **(C)**.

4.3.3 Testing dataset

The validation dataset consisted of 244 images (199 left eyes, 120 right eyes) of individuals graded for AD, Posterior Cortical Atrophy (an AD subtype) and healthy controls. Images were acquired on an Optos P200Tx (Optos plc, Dunfermline, UK), provided by Dementia Research Centre, University College London with drusen annotations from the Queen's University Belfast, Belfast. Individual image disease

status as well as patient information was not available. All images were stereographically projected [74].

4.3.4 Testing

Two trained observers with expertise in UWF retinal image annotation and histological imaging (referred to as Observer 1 and Observer 2) had manually graded the testing set independently, subdividing each image using the Manchester grid overlaid onto the UWF images [183] (see Figure 3.9). This approach divides the image into cells with each cell containing an area equivalent to that occupied by the OD (i.e. 1.77mm^2). This results in 764 cells of equal area on the retina but with different proportions when viewed on the flattened image. The cells were indexed with x- and y- coordinates and an origin at the fovea (0,0). In order for each cell to be compatible with the neural network input layer, patches were geometrically transformed by linear interpolation to 244×244 pixels using a tool provided by Optos. Some peripheral cells were too elongated for this transform and had to be discarded (416 cells outside the periphery zone (blue) in Figure 4.9). This left 348 cells per image but which adequately covered the posterior pole and peripheral retina. These were the cells annotated as either drusen *positive* or drusen *negative* by the two observers. To assess the performance of the drusen detector in different regions of the image, cells were grouped into three zones; *periphery*, *perimacular* and *central* - according to a simplified grid previously used for grading AMD [184], as highlighted by the different colours in Figure 4.9. The first zone defines the central region of the retina and consists of 16 cells (magenta; 28.32mm^2). This zone typically has a high proportion of

pathology in AMD and covers the early treatment diabetic retinopathy study (ETDRS) grid commonly used to report retinal thickness in neurodegeneration [185]. The second zone defines the perimacular region of the retina and consists of 76 cells (green; 134.52mm^2). The final zone defines the peripheral region of the retina and consists of 256 cells (purple; 453.12mm^2). Areas outside the peripheral zone were often labelled as ungradable by the observers due to the appearance of eye lashes and eye lids in the images. Lashes and lids are difficult to avoid at image acquisition, especially in elderly participants who may not be able to open their eyes as wide as younger patients.

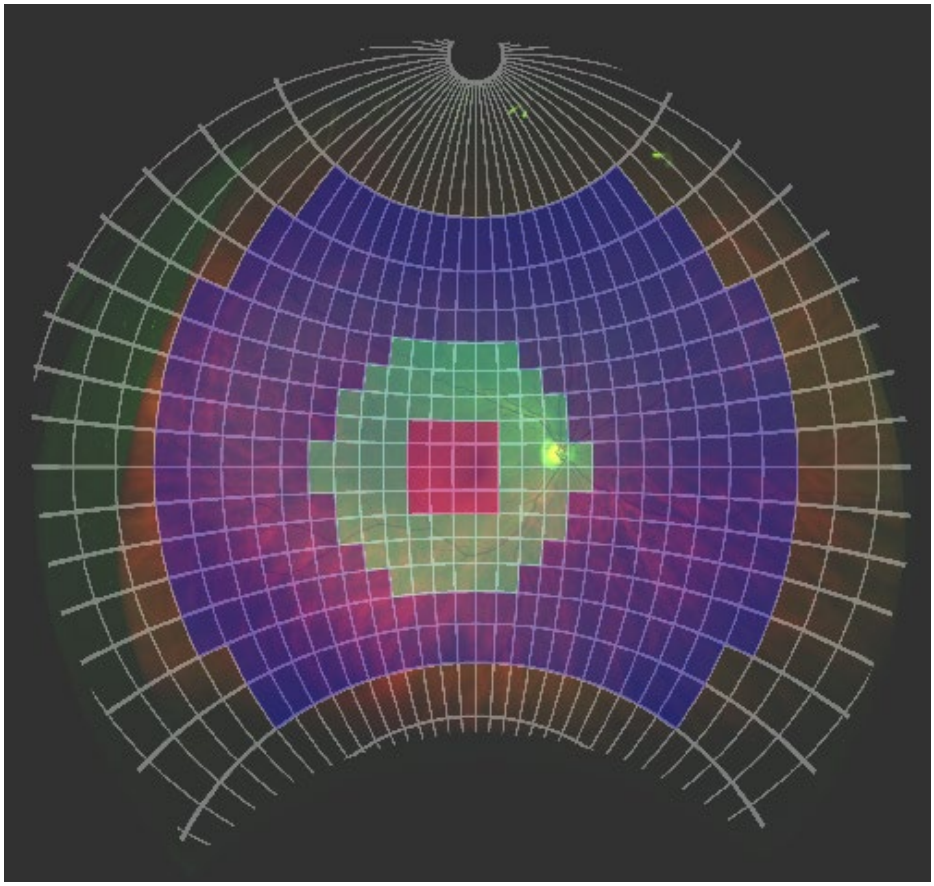

Figure 4.9 The Manchester grid (white lines) overlaid onto a UWF image showing cells grouped into zones. Zones are defined as 3 circles centred on the fovea.

The output of the classification by the network was then compared to the manual annotations by two human observers. As manually grading drusen in UWF images is a challenging process, the union of the 2 observers (UObs) was also considered as a means of assessing performance in the context of low inter-observer agreement. Cells were binarized for drusen labels and used to compare the agreement between the two observers by calculating a Dice Similarity Coefficient (DSC). The DSC is calculated using the binarized image of annotations and is calculated as 2 times the

area (in pixels) overlap (or intersection) of the two observers divided by the total number of white pixels in both of the binary images and is a quantitative measure of performance. Figure 4.10 shows the binarized images, the union and intersection of manual annotations by the human observers.

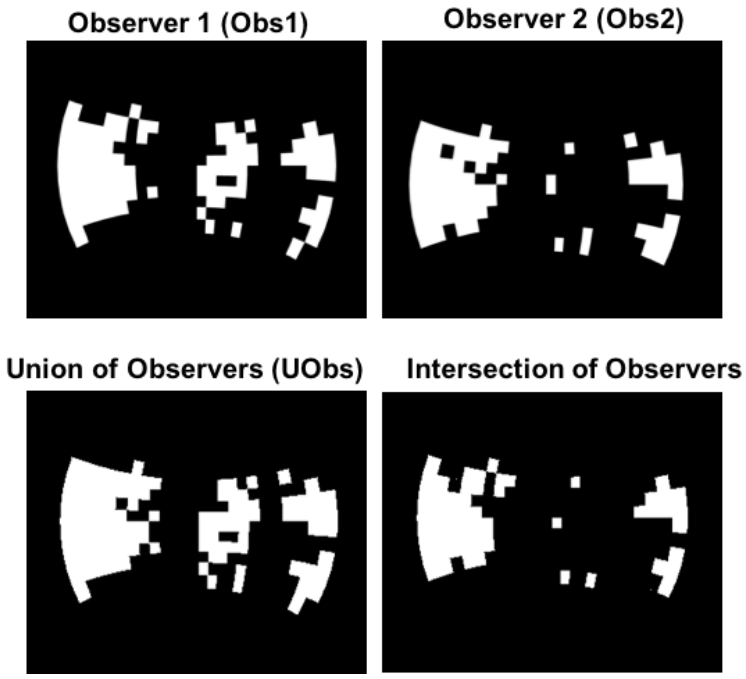

Figure 4.10 Examples of binarised annotations on the MG by Observer 1, Observer 2, union of the Observers (UObs) and intersection of Observers, used to assess performance. The union is defined as all cells labelled positive for drusen, whereas the intersection is defined as only the cells the observers agreed to contain drusen.

A true positive (TP) was an instance or cell predicted as *positive* for drusen by the system and annotated as drusen by the human observer. A true negative (TN) was an instance predicted as *negative* for drusen and annotated as not drusen by the human

observer. A false positive (FP) was an instance predicted *positive* for drusen but annotated manually as not drusen. A false negative (FN) was an instance predicted as *negative* for drusen but is annotated manually as drusen. According to these definitions, the following formulae can be applied to obtain metrics that quantify system performance - sensitivity, specificity and accuracy.

- Sensitivity = (TP/(TP+FN)) x 100 **(4.2)**
- Specificity = (TN/(TN+FP)) x 100 **(4.3)**
- Accuracy = ((TP+TN)/(TP+TN+FP+FN)) x 100 **(4.4)**

Additionally, the area under curve (AUC) receiver operator characteristic (ROC) can be used to measure the relationship between sensitivity and specificity and is calculated by thresholding using the probability returned by the system. The probability indicates the degree at which the instance or cell belongs to a certain class. A curve plotting *1-specifity* against *sensitivity* is created by varying this threshold and summing the area under the curve. Sensitivity and specificity were calculated for each image using both Observer 2 as the reference and Observer 1 as reference and compared using non-parametric Mann-Whitney U test.

Observer agreement was assessed using Cohens Kappa coefficient (κ) a well-established technique for comparing binary gradings [186]. Figure 4.11 shows how κ is calculated and the standard terms used to describe the level of agreement of the calculated score.

		Observer 2	
		Positive	Negative
Observer 1	Positive	a	b
	Negative	c	d

Kappa score:

- 0 = agreement equivalent to chance.
- 0.1 – 0.20 = slight agreement.
- 0.21 – 0.40 = fair agreement.
- 0.41 – 0.60 = moderate agreement.
- 0.61 – 0.80 = substantial agreement.
- 0.81 – 0.99 = near perfect agreement
- 1 = perfect agreement.

Figure 4.11 Interrater agreeability assessment using Cohens kappa coefficient (κ). The counts of agreed annotations (*a* and *d*) and the counts of disagreed annotations (*b* and *c*) are totalled for each image over the set to obtain kappa score, standard error and 95% confidence intervals. A perfect Kappa score is 1 whilst an agreement by chance is 0 [186].

4.4 Results

4.4.1 Inter-observer agreement

The calculated Kappa score indicated a moderate agreement between the observers (κ = 0.53). Figure 4.12 shows the number of cells annotated for drusen in each zone by each observer, where a lower abundance of drusen was observed in the central zone compared to the periphery and perimacular zones. Observer 1 annotated a total

of 262 cells as continuing drusen in the central zone compared to Observer 2 who annotated a total of 213 cells with drusen. In the perimacular zone, Observer 1 annotated 1,234 drusen cells while Observer 2 annotated 1,216 drusen cells. Observer 1 annotated a total of 3,216 drusen cells in the periphery zone compared to Observer 2 who annotated 3,406 cells as drusen. Thus, it appears as if there was a good inter-observer agreement. However, the DSC calculation points to a low level of agreement (average DSC 0.38, SD± 0.27) suggesting that the observers do not agree particularly well on exact cell locations for drusen.

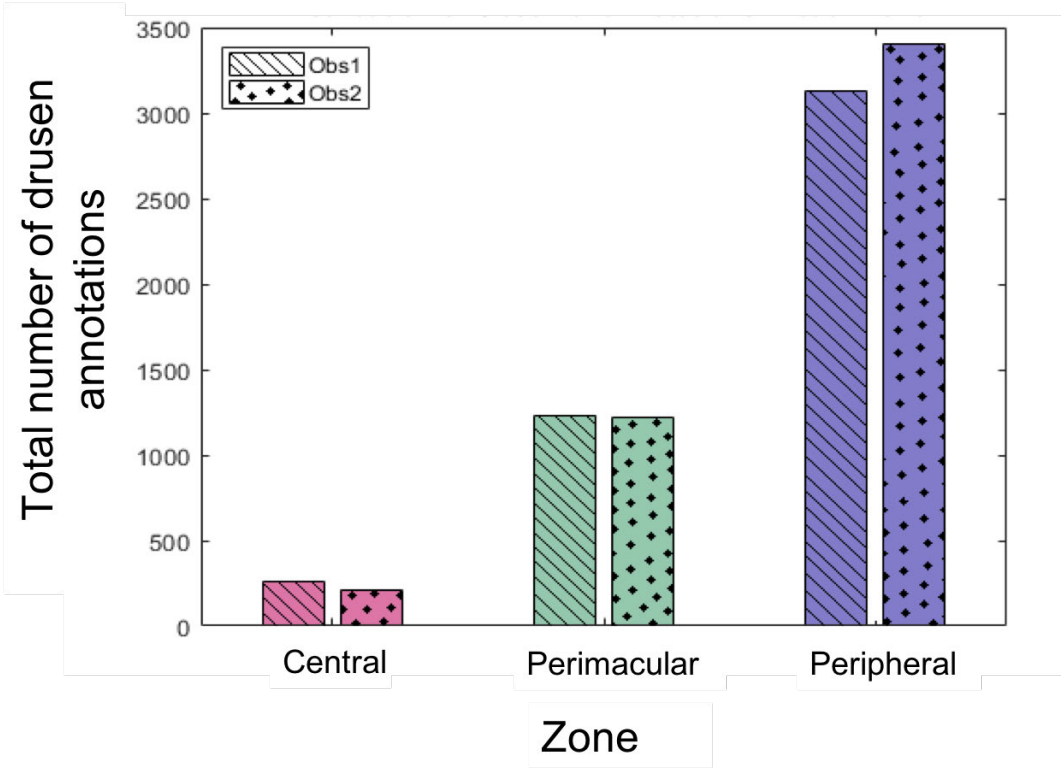

Figure 4.12 Histogram of the total number of cells annotated for drusen by Observer 1 (Obs1) and Observer 2 (Obs2).

Cells where observers disagreed on the presence of drusen were visually assessed and were placed by me into 2 categories: *ambiguous* and *low quality*. Figure 4.13 shows examples of cells from both categories. Ambiguous cells contained subjective drusen, dust or non-obvious lesions in the retina. The low quality contained cells that were difficult to discern where the drusen were located due to poor image quality and low contrast.

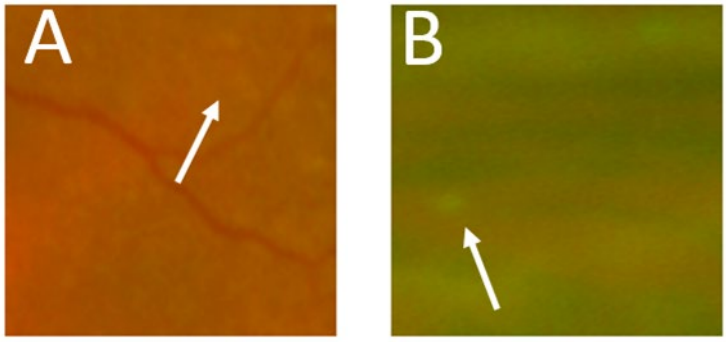

Figure 4.13 Examples cells where observers disagreed. In each category the white arrows indicate possible drusen **A)** *Ambiguous* patch that contains areas that could be subtle drusen or the texture of the retina (*white arrow*). This would be a subjective decision as to whether to manually label as drusen or not. **B)** *Low quality* cell that contains potential drusen or artefact (*white arrow*).

Figure 4.14 shows an example of the gradings from the 2 observers overlaid onto an UFW image. This revealed that where drusen were prevalent there was a higher level of agreement between observers on location. Disagreement seems to occur when drusen overlap into neighbouring cells with one observer annotating the neighbours as containing drusen while the other observer may not. In images where drusen are

sparsely located there was less of an agreement between observers on location. This corresponded to *ambiguous* cells.

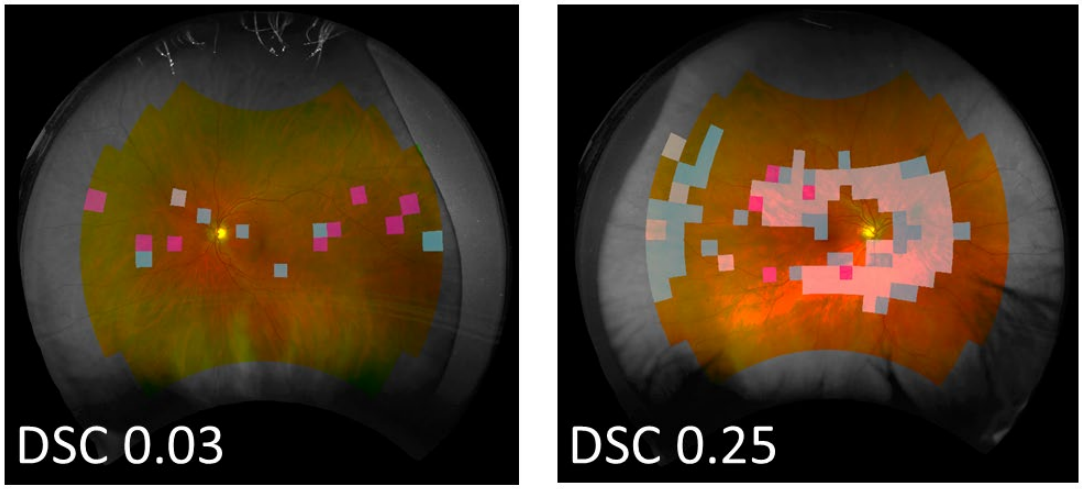

Figure 4.14 Examples of gradings by Obs1 (blue) and Obs2 (pink) and where they intersect (white) in an image with sparse drusen and low agreement (left) and an image with an abundance of drusen and high location agreement (right). It can be seen that much of the disagreement occurs around the main body of drusen clusters where overlaps of drusen into neighbouring cells occurred.

4.4.2 Neural network performance

The neural network, called the drusen detector (DD), trained to a 95% validation accuracy. The drusen detector returned a *slight* agreement ($\kappa = 0.17$) to Observer 1, a *fair* agreement to Observer 2 ($\kappa = 0.21$) and a *slight* agreement ($\kappa = 0.17$) to the union of observers (see Table 4.1).

Comparison	Kappa	SE	95% CI
DD vs Obs1	0.17	0.009	[0.14, 0.19]
DD vs Obs2	0.21	0.009	[0.19, 0.23]
DD vs UObs	0.18	0.009	[0.16, 0.20]
Obs1 vs Obs2	0.53	0.007	[0.52, 0.56]

Table 4.1. Kappa scores, Standard Error (SE), 95% CI of the drusen detector (DD) compared to Observer 1 (Obs1), Observer 2 (Obs2), UObs and between the two observers.

The results in Table 4.2 show in all three zones, the specificity of the drusen detector (96%) was similar to the overall specificity of Observer 1 (using Observer 2 as a reference) (97.3%, p = 0.01). The drusen detector was less sensitive (18-21.8%) than the observers. It achieved the highest sensitivity in the perimacular zone.

	Sensitivity (%)				Specificity (%)			
Zone	DD vs Obs1	DD vs Obs2	DD vs UObs	Obs1 vs Obs2	DD vs Obs1	DD vs Obs2	DD vs UObs	Obs1 vs Obs2
Central	20.9	17.9	16.0	41.2	94.8	94.8	94.9	97.2
Perimacular	24.7	31.4	25.5	55.0	91.4	91.9	92.0	96.9
Periphery	16.4	18.4	14.9	59.0	98.0	98.6	98.3	97.5
All Zones [95% CI]	18.7 [17.6 to 19.83]	21.8 [20.7 to 23.0]	17.7 [16.8 to 18.7]	57.0 [55.6 to 58.5]	96.5 [96.3 to 96.6]	96.7 [96.7 to 96.9]	96.8 [96.7 to 96.9]	97.3 [97.2, 97.46]

Table 4.2. Sensitivity and specificity of the drusen detector (DD) compared to Observer 1 (Obs1), Observer 2 (Obs2) and UObs in all three zones. Observers are compared using Obs1 as reference (Obs1 vs Obs2).

Using Observer 1 as the reference, the area under a ROC curve (AUC) for detecting grid cells with drusen was 0.59, 0.62 and 0.65 in the central, perimacular and peripheral zones respectively. Using Observer 2 as the reference, AUC was 0.58, 0.65 and 0.66 in the central, perimacular and peripheral zones respectively. Figures 4.15 and 4.16 give a graphical representation of this data.

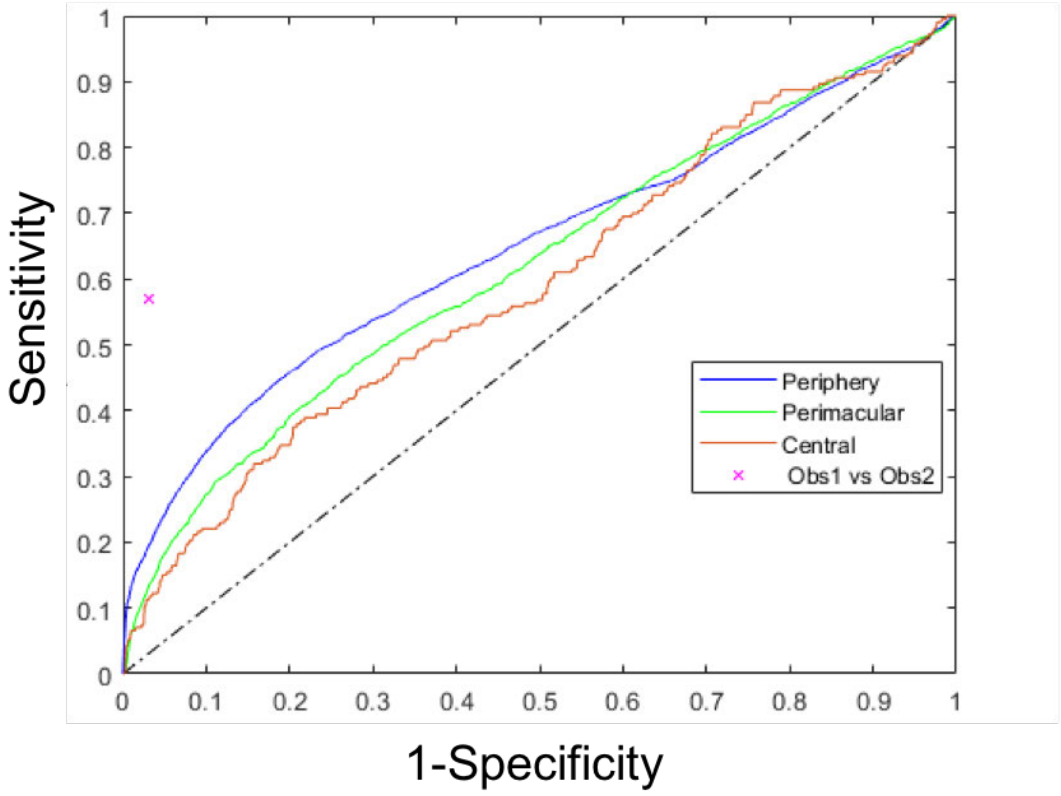

Figure 4.15 ROC curves within each zone for the drusen detector using Obs1 as reference.

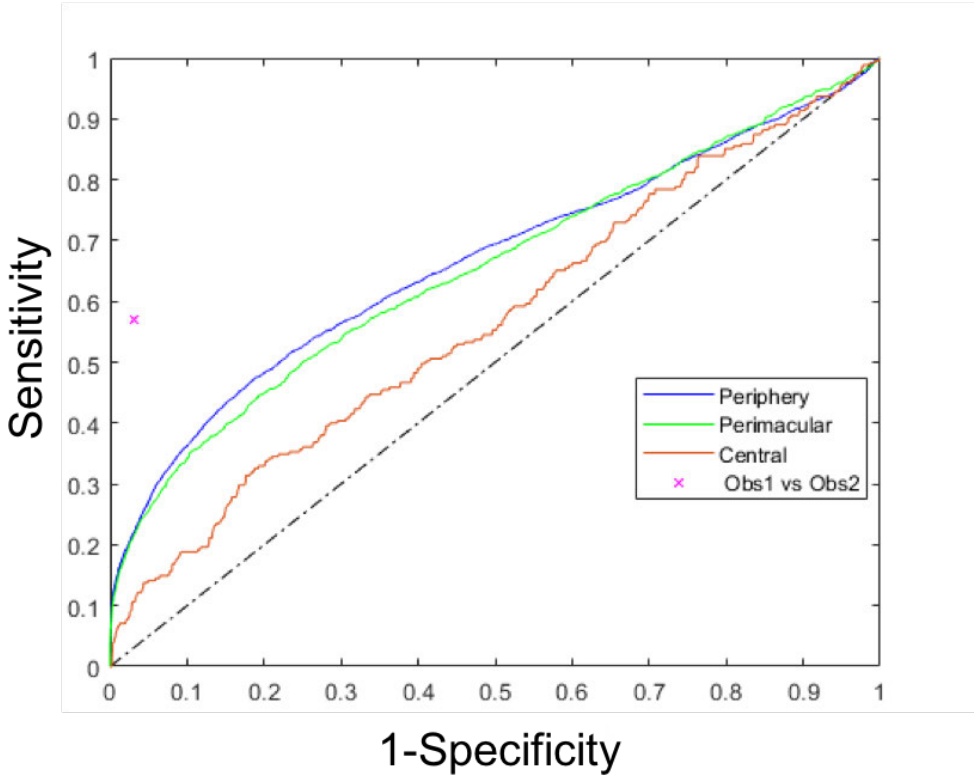

Figure 4.16 ROC curves within each zone for the drusen detector using Obs2 as reference.

I visually inspected false negatives and false positives for the DD system to assess the source of these types of errors and some examples are shown in Figure 4.17. False positives seemed to occur in bright areas near blood vessels and the OD, as the contrast or brightness seems to resemble that of drusen. This was also a source of false negatives, as subtle drusen can look like these types of bright areas near blood vessels and the OD. Areas of low quality, subtle drusen and ambiguous patches were a source of false negatives as drusen were difficult to discern from the background retina. Artefacts at the image border such as the reflection from the faceplate and dust were a source of false positives (specs of dust are bright in appearance and look

like drusen). However, some cells identified as false positives contained drusen and indicated some human error in the manual grading process.

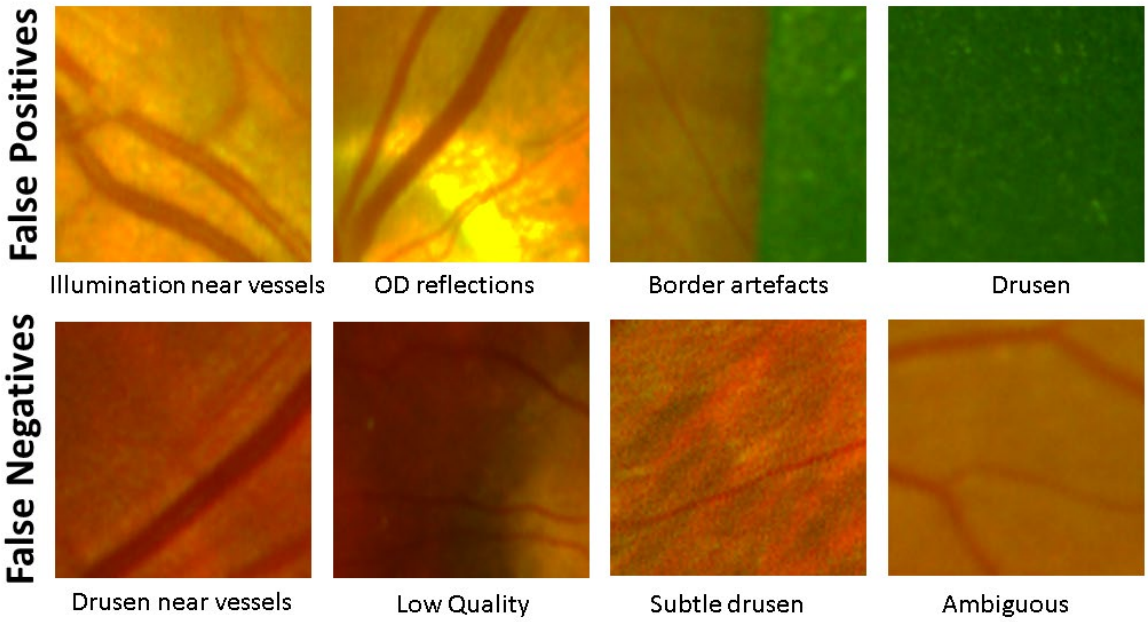

Figure 4.17 Examples of cells that were false positives and false negative returned by the drusen detection system.

To assess how performance of the drusen detector varied with increasing agreement between the two observers, each image was binned according to the DSC of Observer 1 (using Observer 2 as reference) and Observer 2 (using Observer 1 as reference) at 0.1 intervals and plotted against AUC (see Figure 4.18). The best performance for the drusen detector was observed in 8 images (DSC > 0.8 and < 0.9) achieving AUC 0.84 for all zones 45.5% and 46.3% sensitivity (in relation to Observer 1 and Observer 2 respectively) and 99.9% and 97.2% specificity (in relation to Observer 1 and Observer 2, respectively). An AUC of 0.85 and 0.82 was achieved in the peripheral zone (in

relation to Observer 1 and Observer 2 respectively), an AUC of 0.82 and 0.78 in the perimacular zone (in relation to Observer 1 and Observer 2, respectively) and an AUC of 0.55 and 0.59 in the central zone (in relation to Observer 1 and Observer 2, respectively).

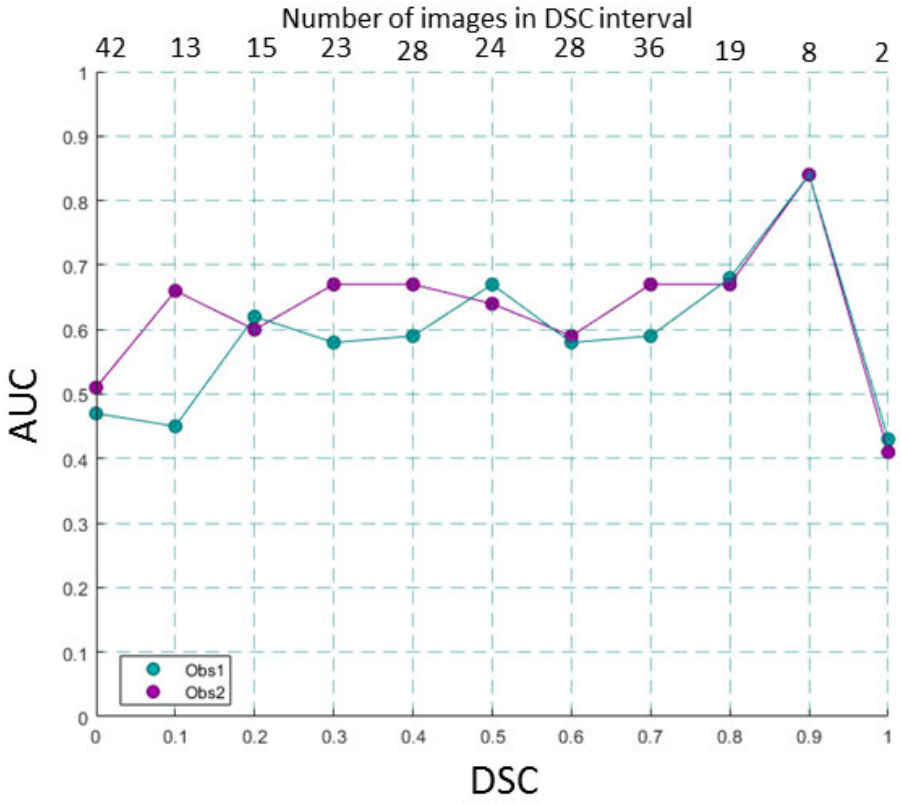

Figure 4.18 Dice Similarity Coefficient (DSC) from Observer 1 (using Observer 2 as reference (blue; Obs1)) and Observer 2 (using Observer 1 as reference (purple; Obs2)) of each image binned into intervals of 0.1, against the drusen detector performance (AUC) of each bin. The number of images within each DSC interval is displayed on the top x-axis. Best performance of 0.84 AUC was achieved between DSC >0.8 and DSC < 0.9.

Figure 4.19 shows the output of the drusen detection in the 8 UWF images with the highest AUC. The predictions are colour coded as a 'heatmap' and overlaid onto the UWF image. Observer 1 annotations (pink) and Observer 2 (blue) and their agreement (white) are also displayed. Cells with a high probability for containing drusen and good observer agreement were visually inspected and observed to contain hard and soft drusen clusters. These cells had a similar appearance to the *positive* image patches used to train the network. Visual inspection of cells with observer disagreement (either blue or pink), which appear around the edges of high agreement, were observed to contain drusen close to the border of the cell.

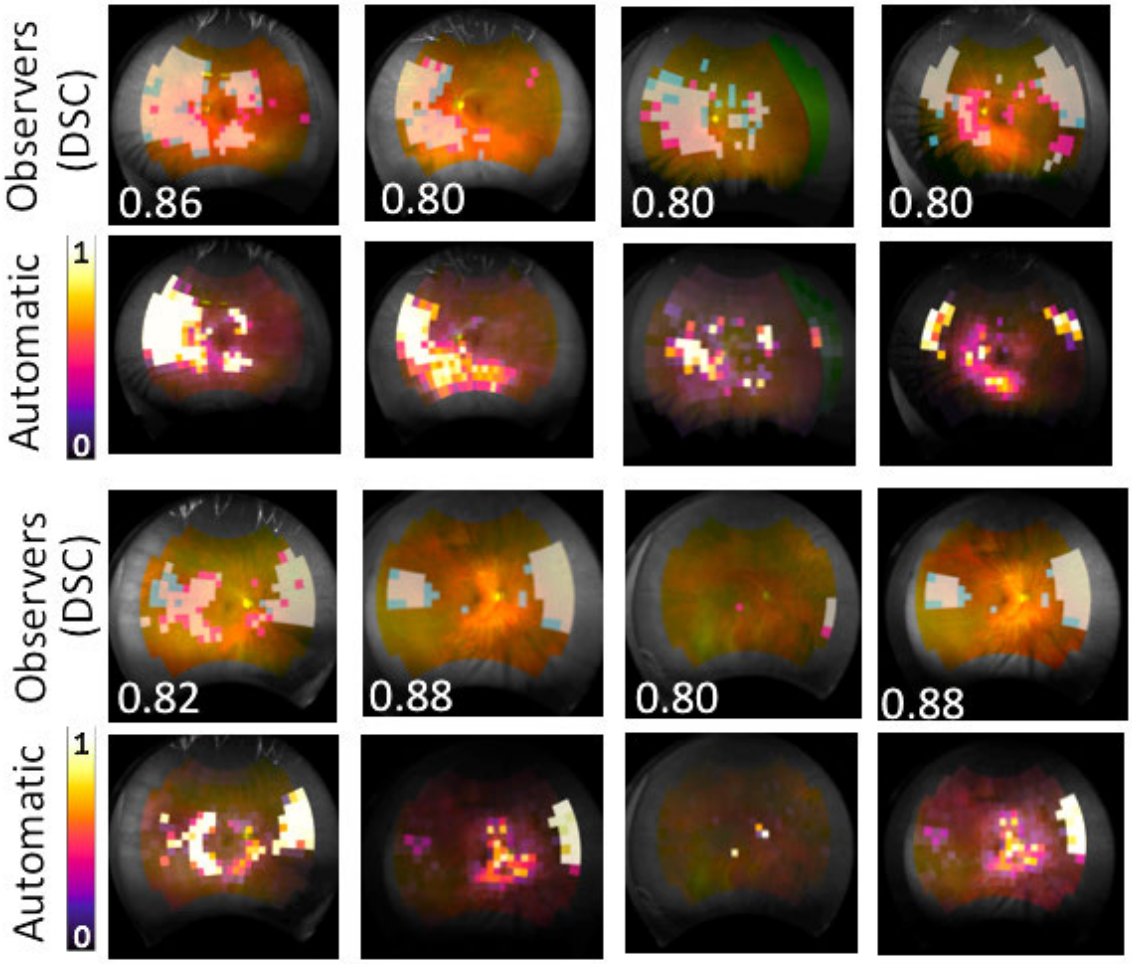

Figure 4.19 Comparison of Observer annotations. White denotes agreement between observer, blue denotes the annotations of Observer 1 and pink the annotations of Observer 2 Dice Similarity Score (DSC) for the observers is also given where a higher number denotes better agreement. The output of the automatic drusen detector is a probability map (white high probability drusen (1) to black low probability drusen (0)) which has been colour coded and overlaid onto the images.

Two images had a DSC of 0.98 and 1.0 where the drusen detector performed at 0.41 AUC. These images were visually inspected and contained exceptionally sparse and subtle lesions surrounding the OD and in areas of low quality due to shadows from lashes. This suggests that the reason for the low AUC was due to the drusen detector failing to identify small and subtle lesions, whereas both human observers were able to classify these instances. One false positive was observed on the OD in one of the images (DSC = 1.0) which contained round and bright drusen like shapes and is shown in Figure 4.20. In order to assess the nature of the potential pathology the accompanying autofluorescent image was inspected to identify any possible OD drusen. No evidence of OD drusen was found.

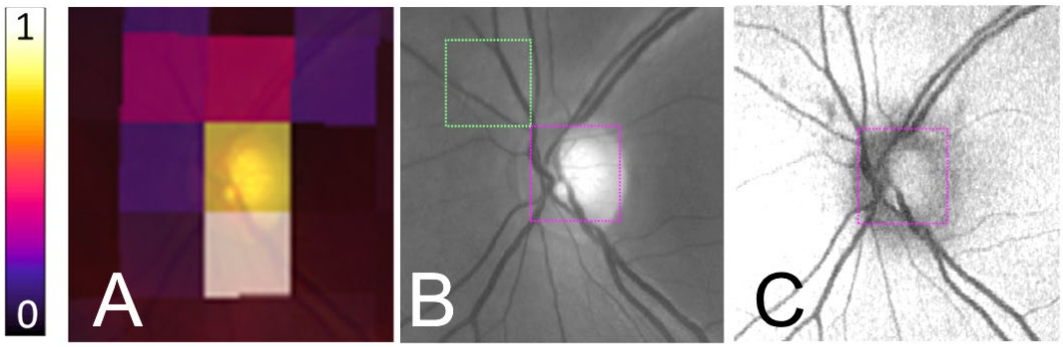

Figure 4.20 Visual inspection of the source of low performance in an Image with DSC of 1.0 and AUC 0.41. **A)** Output of the drusen detector shows high probability cells for drusen near the OD but were false positives (*yellow and white*) and also contains false negatives near the vessels (*pink*). **B)** Visual inspection of the false positives and false negatives in the green channel. The false negative (*green box*) shows regions of very subtle drusen around the vessels. The false positive (*magenta box*) contains bright drusen like shapes in the OD that looks like it could contain OD drusen. **C)** Accompanying autofluorescent image shows no evidence of OD drusen and is a false positive.

4.5 Discussion

I have presented a transfer learning-based system for the automatic detection of drusen in UWF images. Key aspects of the system included: the training of a neural network using a small number of images; performance assessed against two human observers; network specificity similar to that of the observers.

Specificity remained high in all zones indicating the network's ability to classify negative instances. This would be beneficial in a clinical environment, particularly for screening for signs of neurodegeneration, as the number of false positives would be low which would reduce the number of patients being incorrectly referred for further examination. Sensitivity was higher when inter- observer agreement was high and suggests that when the reference standard is reliable the system performs similar to a human observer. However, this was with the exception of two images which had a low AUC for the automatic detection when agreement was highest and was due to the difficult appearance of drusen that were only several pixels in size. This would be a challenge for a neural network to learn/detect but annotated readily by the two graders who had expertise in histological analysis of drusen and would have the ability to assess for drusen in a UWF image in the context of molecular composition and how drusen may be contorted within the layers of the retina.

One image had perfect agreement between observers (DSC = 1.0) and second image near perfect agreement (DSC = 0.98). The DD performed at 0.41 AUC for these images. After visual inspection it was seen that the images contained a low abundance of subtle and small drusen, which represents a difficult classification task for the proposed system. However, the average DSC was low for the whole dataset, suggesting that the observers disagreed frequently on exact locations of drusen. A common source of disagreement occurs when drusen overlap into cells, where one observer might annotate the neighbouring cell as containing drusen but the other does not. The cell annotated within the overlap might only contain drusen a few pixels in size, which would also be a difficult instance for the drusen detector to

classify. Another source of disagreement was from regions of soft drusen that are inherently ambiguous due to their similarity in appearance to the texture of the surrounding retinal tissue. Standardisation of the conditions in which grading is conducted (i.e. training, background room lighting and computer screen size) may help to improve observer agreement. Additionally, dust and vitreous material were also present in some images and can resemble drusen, further contributing to observer disagreement and drusen detector misclassification. To minimise dust artefacts at least, the UWF optics can be cleaned in line with the manufacturer's guidelines for device care in order to remove particulates prior to image acquisition.

The use of heatmaps to visually inspect the predictions were insightful for visualising the output of the neural network and could have application for monitoring drusen progression and changes in longitudinal images. Drusen progression could then be quantified by change in drusen distribution and area from the probability map and used to investigate new biomarkers or associations between drusen and neurodegeneration.

The drusen detection system is likely to be improved by training the network with more image data and also additional observers to create a more consistent reference standard. In previous work using deep learning (discussed in section 4.2.3) for classifying AMD severities involved training neural networks with approximately 5000 - 120,000 images achieving accuracies above 70% where performances varied depending on the dataset (i.e. presence of artefact or low-quality images). Comparable to the work in this thesis, Peng et al [182] used 58,402 fundus

photographs for training and 900 for testing, achieving 72% accuracy (71% sensitivity, 71% specificity) for classifying drusen subtypes, 89% accuracy (73% sensitivity, 95% specificity) for classifying pigmentary changes and 96% accuracy (62% sensitivity, 98% specificity) for classifying GA subtypes from whole fundus photographs. The performance was compared to the image gradings obtained from 88 readers in the AREDS study for classifying AMD severity [134]. Observer agreement in the AREDS dataset for drusen presence, type and area gradings in 1,230 eyes gave kappa scores of 0.50, 0.61 and 0.56, respectively [187]. This is comparable to the observer agreement in this chapter (kappa = 0.53) further highlighting the challenging nature of manually grading for drusen. It also suggests that while more training data would improve the system a more reliable reference standard is still required. Drusen are subjective to grade and is the main reason for the low observer agreement. This highlights the need for precise drusen grading protocols that minimise the subjectivity to provide a reliable reference standard to validate (or train) an automatic system against. Such grading protocols would need to be designed so that annotations can be input into a neural network (e.g. coordinates or patches) and follow grids outlined in this chapter. Such grading methods would be manually intensive, time consuming and expensive. Given the presence of these challenges, the recommended number of additional training data that would most likely see an improvement in performance is more than 500 images annotated by 20 observers.

4.6 Conclusions

The aim of this chapter was to develop a method to automatically detect drusen in UWF images, validated to a reference standard and assess its utility for application for future assessment of neurodegeneration in UWF images. Bearing in mind the low sensitivity, this system will be used in an exploratory analysis of the dataset described in the next Chapter to investigate associations between drusen presence and AD.

Chapter 5 Retinal changes in Alzheimer's disease

5.1 Introduction

In this chapter the analysis of a cohort featuring patients with MCI and AD as well as CH individuals is reported. Vessel analysis was performed using a previously established pipeline and is described. Objective features of the retinal vasculature including vessel widths, vessel tortuosity and branching complexity (measured via fractal analysis) are analysed using statistical modelling that accounts for the correlation between two eyes of an individual in order to assess evidence of differences between the cohort sub-groups. The aim is to investigate potential biomarkers of pre-clinical AD derived from the retinal vasculature in a UWF image.

5.2 Materials and methods

5.2.1 Dataset

This cross-sectional study featuring patients with MCI and AD as well as CH individuals was accessed via Duke University School of Medicine (Durham, North Carolina, USA). Participants were aged 50 or older and were enrolled in a study from the Duke Memory Disorders Clinic. All participants were appropriately consented following local rules and procedures pertaining to ethical approval and clinical research

governance, which included permission to share anonymised data for analysis in this thesis. AD and MCI participants were evaluated and diagnosed by an experienced neurologist with a specialisation in memory disorders. Assessment was based on the diagnostic guidelines of the National Institutes on Aging-Alzheimer's Association [188] [189]. In these guidelines, MCI is the term used to describe presymptomatic AD. MCI was identified by: assessment of change in cognition in comparison to a person's previous cognitive ability; impairment in one or more of memory, executive function, attention, language and episodic memory; problems with performing everyday tasks such as paying bills and not demented (i.e. cognitive changes significantly mild that they have no impaired social or occupational functioning) [189]. AD was identified in individuals in a similar fashion and were differentiated from MCI by an expert clinician on whether symptoms have significant interference with their ability to function at work and in everyday activities [188]. Participants did not undergo PET or CSF assessment to determine pathology or biomarker status. CH individuals were recruited from either the spouses of the patient volunteers or from the Joseph and Kathleen Bryan Alzheimer's Disease Research Centre that includes a database of cognitively normal individuals based on extensive testing including MoCA [32]. Exclusion criteria included non-AD dementia, diabetes, uncontrolled hypertension, demyelinating disorders, glaucoma, AMD and other vitreoretinal pathology that could interfere with the retinal image analysis. MMSE [190] was used to evaluate cognitive function on the same day of image acquisition. An Optos California UWF-SLO was used to image the retinal vasculature of 169 participants. Both left and right eyes were imaged where possible (212 CH, 62 AD, 64 MCI images). An additional 9

CH (3 left, 6 right), 4 MCI (3 left, 1 right) and 8 AD (2 left, 6 right) single eyes were obtained where images from both eyes was not feasible (e.g. challenges during image acquisition).

120 of these images were rejected (102 CH, 16 MCI, 2 AD) due to malfunctioning of the device thought to be an error in the detector. The detector is an avalanche photodiode (APD) which is responsible for converting light to electricity. There are two APD's in the device, one for the green laser to convert to the green channel in the image and one for the red laser to convert to the red channel in the image. Error in APD functioning would cause noise in the conversion of either channels. In this study, the green APD was the most likely cause of the malfunction as the noise was in the green channel. Figure 5.1 shows images with varying levels of detector error that occurred in the dataset. The result was a noisy green channel that obscured retinal landmarks including the vasculature and so measurements were not possible to obtain in these images.

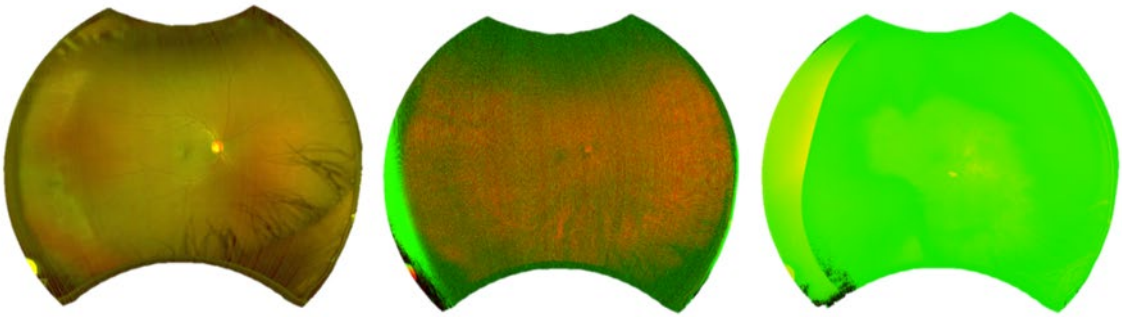

Figure 5.1 Example images displaying varying severity of detector error from low to high (left to right).

16 images were rejected due to obstruction of the OD or fovea (6 CH, 10 AD) which occurs when eyelashes, eye lids, retinopathy or media opacity are present in the image obscuring these landmarks (Figure 5.2), which are needed for the vessel segmentation process. In addition, media opacity reduces the definition of the vasculature and vessel segmentation would be unsuccessful for these images.

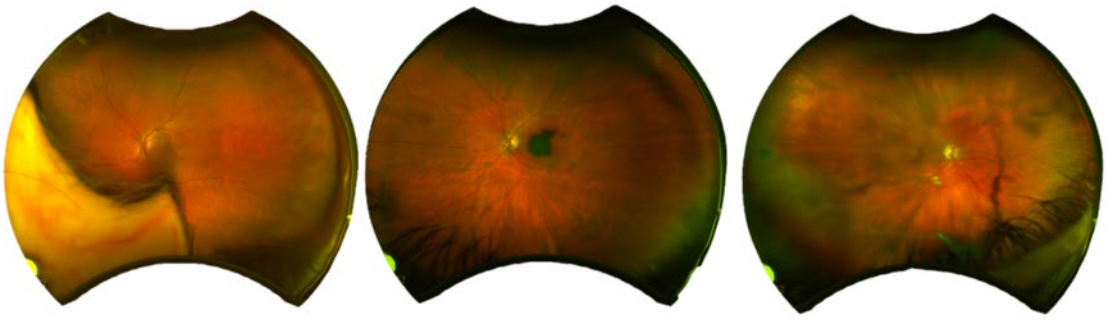

Figure 5.2 Examples of images with obstruction of the OD and fovea by eyelashes and eyelids (left), retinopathy (middle) and media opacity (right). Images such as these were rejected because vessel segmentation and the subsequent measurements would be unsuccessful in the analysis pipeline.

Images were exported from the device in non-compressed TIFF format to preserve detail. Images were sent for stereographic projection (see Chapter 2) and returned with the projected image along with a .txt file with the projection error recorded. This automatic process was made available by Optos. If the projection error was >1 the image was rejected as the projection process that allows pixel measurements of the vasculature to be converted to mm equivalents but would not be precise from such inaccurate projections. Projected images with a large error usually occur due to

incorrect automatic fovea detection in the Optos process arising either from retinopathy or from poor image quality. Manual identification of the fovea can be used to reduce the projection error. This was not performed in this dataset because the incorrectly projected images had no green channel, possibly arising from the error in the detector. 16 images were rejected due to unacceptable projection errors (6 Control, 8 MCI). Figure 5.3 shows an example of a successful (left) and unsuccessful projection (right). Following data cleaning 107 CH, 42, MCI and 60 AD images were sent for segmentation of the vasculature.

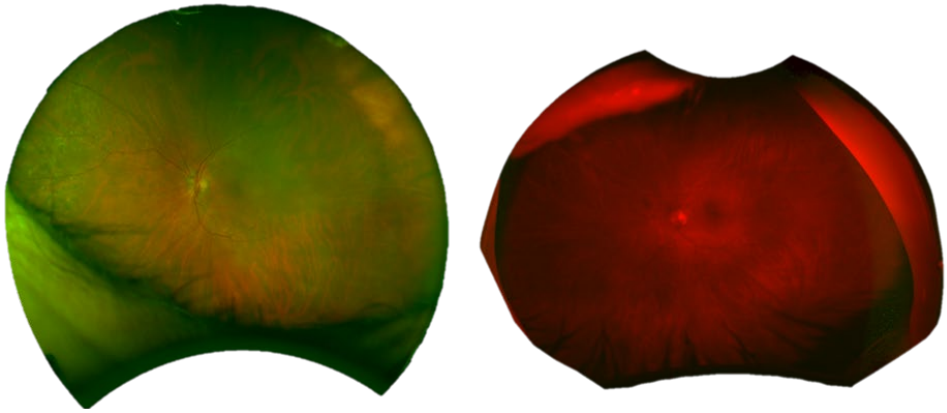

Figure 5.3 Shows a successful stereographically projected image (left) and an unsuccessful one (right), likely due to the absence of the green channel caused by a device malfunction.

5.2.2 Vessel segmentation and correction

The vessel analysis procedure involved: (1) definition of a ROI, (2) automatic segmentation of the retinal vasculature [191], (3) manual correction of the segmentation, (4) labelling of arteriolar and venular vessels [192] and (5) extraction of vasculature measurements pertaining to the vasculature [191][117].

To obtain an optimal segmentation of the vasculature, a ROI was manually defined first in order to select an area free from eyelashes, eyelids and artefacts (Figure 5.4 A-B). This produced a mask that maximises the usable FOV (typically larger than is available from a conventional fundus camera) whilst minimising any features that could lead to incorrect segmentation. In some occasions, due to suboptimal participant positioning during image acquisition, bright reflections occur in the perimacular zone (see Chapter 4 for the definition of this zone) (Figure 5.4 A). Such artefacts may be incorrectly segmented as vessel but can be removed in a manual correction stage. Within the ROI, a binary map of the vascular network was obtained from pixel-wise classification (assigns probabilities to pixels) by a previously developed fully connected neural network with a reported AUC 0.97 [117] (Figure 5.4 C). Some smaller vessels fail to segment, a common limitation of retinal vessel segmentation techniques, or parts of the vasculature were obscured by eyelashes that could not be avoided in the ROI selection stage and so were not segmented correctly. The segmentation may also contain inaccuracies due to complicated vessel junctions and crossings as well as intertwining arteriolar and venular paths. To simplify the vascular network, a skeletonised version of the binary vessel map was

created. This was used to represent the network as a graph structure where the edges are vessel paths and the nodes represent either the start/end of a vessel or a vessel bifurcation (Figure 5.4 D)[117].

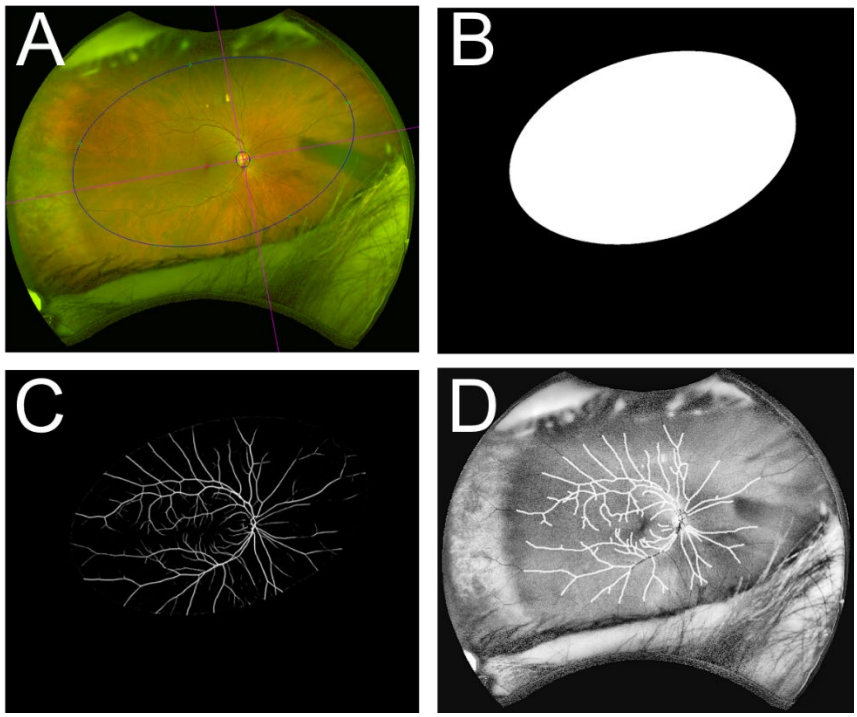

Figure 5.4 Output of the preparation and segmentation stages to vessel analysis in UWF [117]. **(A)** The OD and fovea are manually located as landmarks (red cross). A ROI is defined by either placing a resizable ellipse onto the image or by selecting 5 points around the ROI to create a tilted ROI. It can be seen in this image that a dark artefact exists in the right portion of the image and could not be avoided. **(B)** The resulting ROI binary mask. **(C)** The automatic segmentation within the ROI performed by the neural network. **(D)** The skeleton of the segmentation is displayed over the image for a manual correction stage that aims to obtain the most accurate

representation of the vascular network as possible by allowing a user to make a limited set of corrections.

To ensure that the vascular network is as accurate as possible, I completed a manual correction step. A vessel tree is represented as one vessel path that extends from the OD to the periphery. The aim was to create complete and fluent vessel trees to allow a comprehensive set of valid measurements to be extracted. The type of corrections that were enacted are (A) completing trees, (B) identifying vessel crossings, (C) removing false positives such as those that arise from retinopathy or artefact, and (D) untangling complicated network patterns.

The following descriptions refer to Figure 5.5**.**

Correction of type A were needed when either a reduction in image quality (i.e. a reduction in contrast) occurs along a vessel that leads to suboptimal segmentation, or from artefact that prevents accurate segmentation and breaks the vessel path (Figure 5.5 (*left)* 4, 6 and 9). These can be time-consuming corrections to make as UWF imaging captures a large amount of the vasculature in the retina. If these corrections were not performed, it would lead to disconnected vessel trees that are not representative of the true vascular network. The vessel segmentation is piecewise linear, whereby each segment of the skeleton is a straight line joined to another along the vessel path. This means that any small gaps between segments can be approximated by joining with a straight line. Corrections of type A were performed

by selecting the two edges that are to be joined and fitting a straight line. However, this was only performed when the distance between the two edges was small enough that a straight line represents the true path of the vessels. This means that larger gaps between segments that may have bends in the vessels were not corrected in this way because a straight line would not be representative of the vessel path. The judgement of performing such corrections (or not) was by a subjective decision made by me. Figure 5.5 (*left*) examples *4*, *6* and *9* shows these corrections at various distances. In each of these cases, a straight line was deemed representative of the network and the manual correction results in a fully connected tree as shown in Figure 5.5 (*right*).

Correction of type B were to ensure that when different vessels cross one another they are connected to the correct vessel tree. Figure 5.5 (*left*) examples *2, 5* and *9* shows overlapping vessels that have been connected and then coloured according to the tree they belong to. This again was an important but time-consuming correction to make because if a branch from, for example, the pink tree at example *2* was connected to the blue tree, the vessel path would join at an invalid (i.e. anatomically incorrect) right angle and the path would incorrectly follow back on itself. Vessel width measurements made on an incorrect path such as this would result in an erroneous increase in vessel thickness and could be misinterpreted as vessels thickening when this was not the real case.

Corrections of type C involved the removal of false positives that occurred as a result of artefacts being present in the image (Figure 5.5 (*left*) *1*), or from retinopathy such

as drusen and exudates or from deep lying choroidal vessels (Figure 5.6). Corrections of type D were performed by removing vessels that were too short to obtain a width measurement from (Figure 5.5 (*left) 10* and *8*) or could not be correctly joined to a tree (Figure 5.5 (*left) 3* and *7*). This occurred in images with complex networks and the correction avoided any errors in the vessel tree. For example, Figure 5.5(l*eft*) example *7* was removed because this segment was a branch from the blue tree, but the segment end is from the main path of the pink tree. Such instances were deleted in order to create a vessel network as representative of the true vasculature as possible.

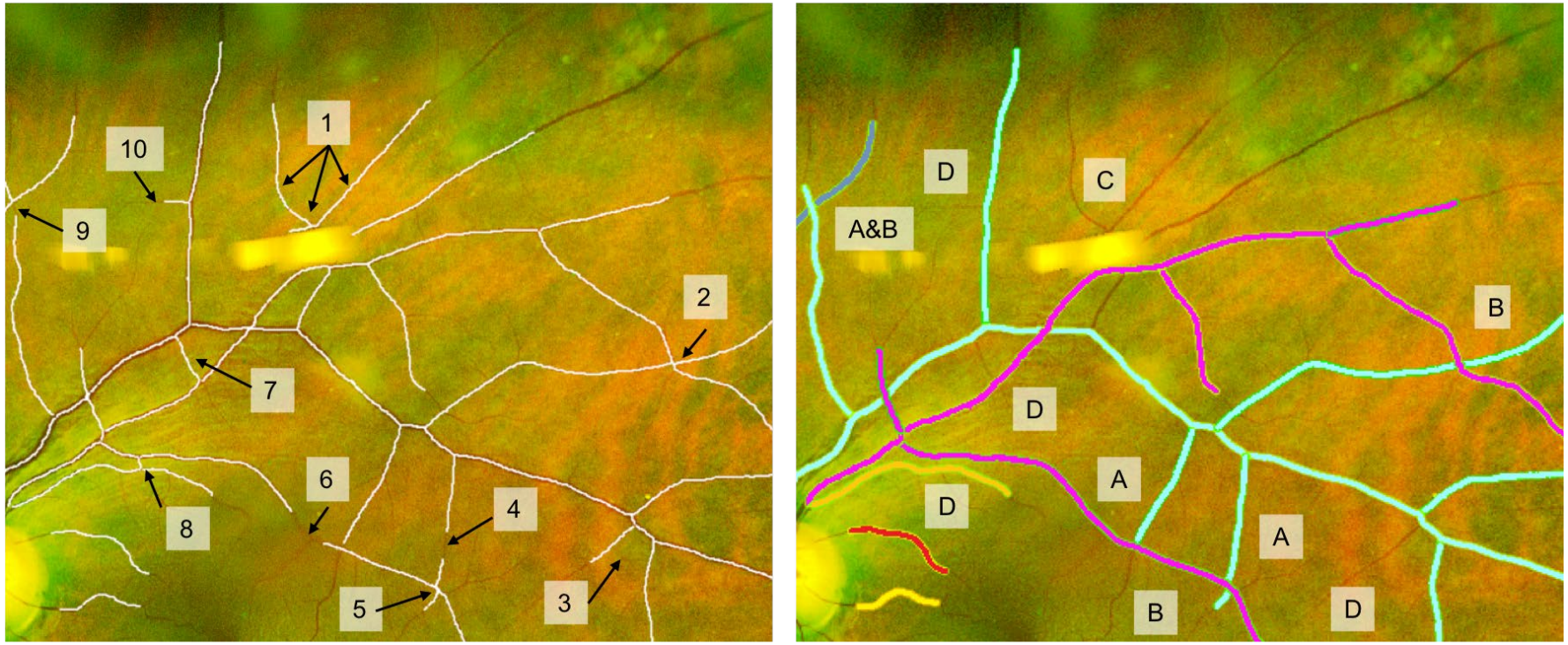

Figure 5.5 Overview of the manual corrections made in a portion of a segmented UWF image (*left*). The corrected network (*right*) where vessel trees are represented as different colours. We can see that by correcting the skeleton we have fluent trees extending from the OD to the periphery.

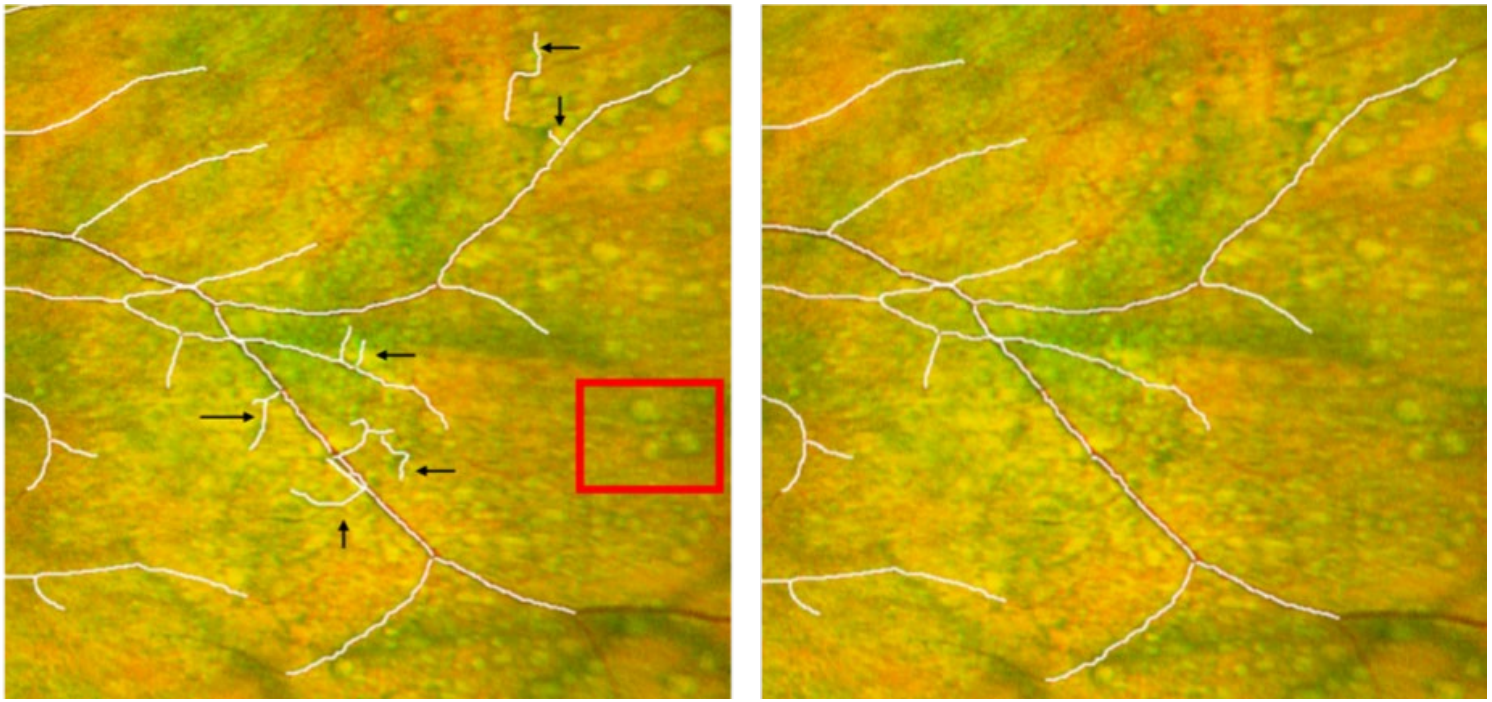

Figure 5.6 Example of false positives (*left arrows*) due to peripheral drusen appearing near vessels, similar to drusen highlighted in the *red box*. Left shows the correction of type C that involves deletion of false positives.

Following manual corrections, I labelled the vessel trees as arteriolar (A) or venular (V). AV labels were defined by visual inspection. Venules appear darker than arteries due to deoxygenated blood [192]. In addition, venular and arteriolar paths almost always cross one another (with arterioles on top) and not over themselves, and vessels usually alternate between venular and arteriolar paths around the OD. Following these rules, an accurate classification (previously reported to have an accuracy of 0.883 [192]) can be made in order to label each vessel tree as arteriolar or venular (Figure 5.7).

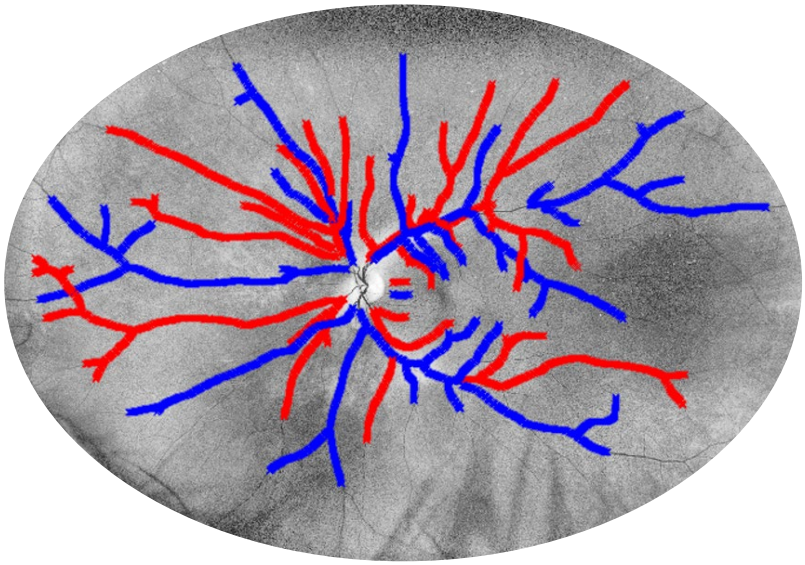

Figure 5.7 A UWF image after complete manual correction process with all vessel trees separated and highlighted as A (red) or V (blue).

5.2.3 Retinal vessel parameters

Table 5.1 shows a summary description of the retinal vessel parameters that were extracted from the corrected vessel trees

Parameter	Description	Retinal Zone
CRVE **CRAE** **AVR**	Summary width measures obtained by selecting 12 largest (6 arterioles and 6 venules) vessels and computing according to [193] AVR = CRAE/CRVE	Zone B Zone B
WGa **WGv**	Gradients of vessel widths of main arterial or venular paths estimated for each of the quadrants superotemporal (ST), inferior temporal (IT), superonasal (SN) and inferior nasal (IN) [91]	Zone B to periphery
FDa **FDv**	Global branching complexity of arteriolar and venular networks calculated from the corrected vessel skeleton. FD is calculated by a multifractal method that summarises how thoroughly the branching pattern fills 2-dimensional space. The higher the fractal dimension the more complex or space-filling the pattern[194][195].	ROI
Torta **Tortv**	Sum of squared curvatures divided by vessel path. Curvatures estimated by [196]. The global tortuosity values are calculated from each path by combining and calculating an average weighted by the length of the path (i.e. short paths will have less weight in the calculated average). The smaller the tortuosity the straighter the path.	Zone B to C

Table 5.1 Descriptions of the vessel parameters measured for each UWF image and the relevant zone or region of the retina that the measurement is derived.

An image is divided into the 4 quadrants (SN, IN, ST, IT) with the quadrants mirrored in the left and right eyes. Figure 5.8 (pink axes lines) illustrate the image quadrants which were centred and aligned according to the OD centre, the OD radius and the fovea centre (as observed by an axial tilt of the quadrants). The image is then divided into concentric zones - Zone A (0.5 to 1.0 OD diameters from the OD centre), Zone B (1.0 to 1.5 OD diameters from the OD centre), Zone C (1.0 and 2.5 OD diameters from the OD centre) and the periphery, which is any area outside of Zone C.

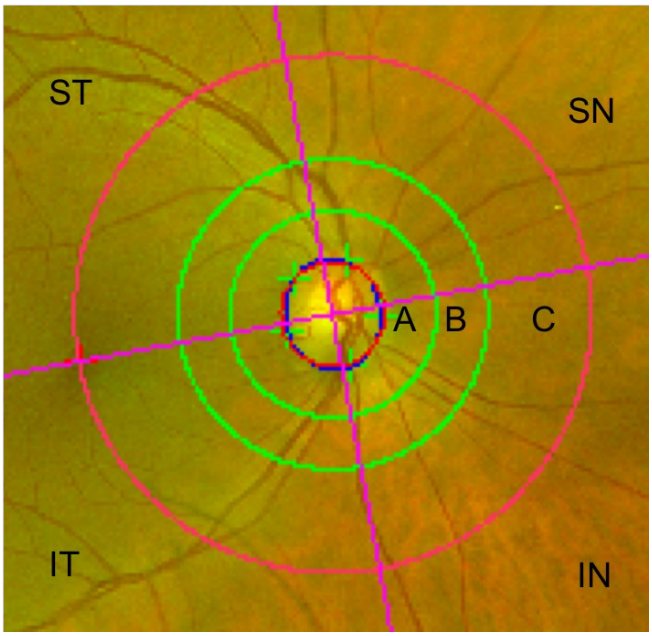

Figure 5.8 The image zones. Peripheral zones are any area outside of Zone C.

Within each quadrant, the longest arteriolar and venular path is selected to measure the WG (See Figure 5.9). The intuition behind this choice was to investigate the longest vessels that are carrying blood in and out of the retina in each quadrant. Once the arteriole and venule pairs are selected, the vessel widths are measured along

their paths where the WG is the gradient of the robust fit regression line that fits these data points [91]. WGa and WGv was calculated within each quadrant.

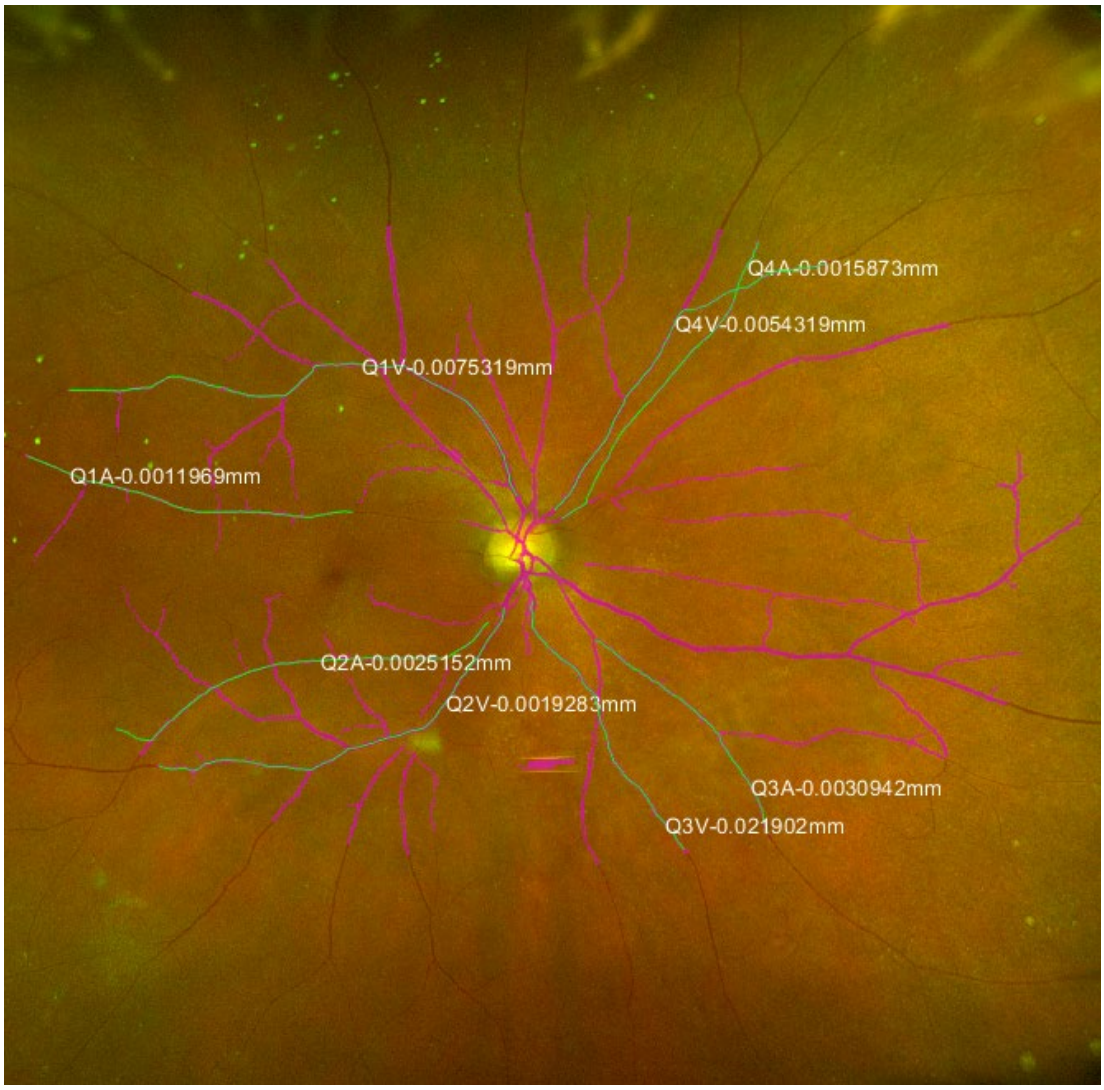

Figure 5.9 Example of the longest arteriolar and venular vessel paths (green) in each quadrant (Q) from which the vessel width is then measured (in mm). The labels show SN (Q1), SN (Q4), IN (Q3) and IT (Q2) along with the computed width gradient. The raw segmentation of the vessels is overlaid in pink.

Tortuosity is a unitless measure [197] and describes how an object curves and twists. Curvature is estimated by measuring the ratio of the length of a curve between two points on the vessel path to the distance between the two points. A larger curvature would indicate that the vessel bends more within the two points. The global tortuosity values were then calculated using the sum of the squared curvatures divided by the length of the path. Tortuosity was measured for arteriole and venule pairs for each vessel path using [196]. Fractal dimension (FD) is a unitless number used to quantify the complexity of a pattern and was first introduce in ophthalmology by Family et al [198]. FD is a method to measure an object that is made of repetitive patterns across different scales (i.e. self-similar). The retinal vasculature can be considered a self-similar pattern as it follows a tree structure where branches repeat and get smaller towards the periphery. FD is calculated by measuring the change in detail of the vessel tree (i.e. the white pixels) in the space it is contained (i.e. the black pixels) at different scales, which is achieved by overlaying boxes of decreasing size onto the vessel skeleton and counting the presence of white pixels [198].The FD was calculated for the arteriolar (FDa) and venular (FDv) binary maps using the methods from [194][195].

As FD measures were calculated from the user-defined ROI, the total area covered by the ROI was also determined (in millimetres, *Areamm2*) for subsequent analysis. Figure 5.10 shows a visualisation of the complete set of ROI's sampled for every image in the dataset. Overall, the FOV in a UWF image is larger than that of a fundus camera. The largest areas sampled were in the ST and IT zones. This is due to less

artefact from lashes and the device face plate being present in these regions compared to the IN and SN zones.

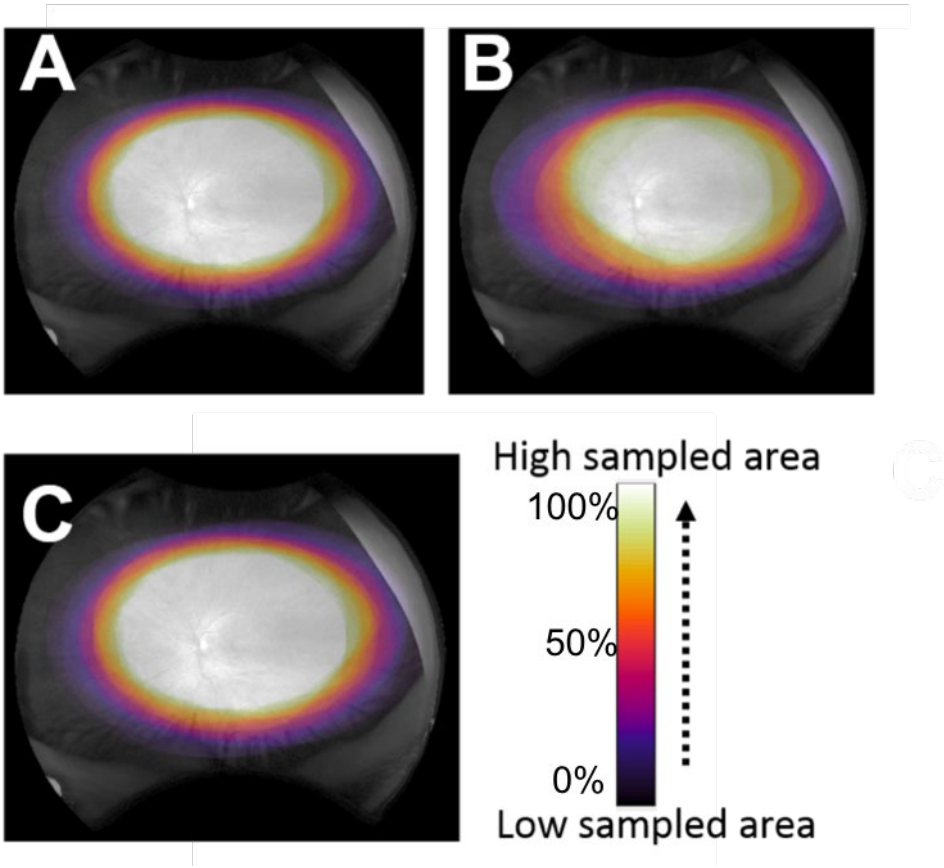

Figure 5.10 Visualising all the ROI's sampled over every image and in the whole dataset overlaid on a representative image. The whiter the area the more times this region was sampled. There is a skew to temporal quadrants as less artefact from eyelashes and eyelids was present in these regions.

5.2.4 Data cleaning and validation

Prior to statistical modelling it was essential to identify any outliers in the measured parameters and assess whether they were valid or erroneous measurements. The intra-observer agreement was also considered in order to determine the repeatability of the vessel measurement process performed by myself (which contained points at which decisions are made about enacting manual corrections).

WG parameters that could sometimes not be calculated due to obstruction of a vessel by an eyelash or eyelid. In the dataset, 58 CH, 26 MCI and 29 AD images had to be removed due to missing WG values. All vessel measures were assessed using boxplots to observe their distributions and identify potential outliers (see Figure 5.11). Any measure that was beyond the 95% confidence interval of the set of measures was visually inspected by looking at the originating image and the corresponding vessel segmentation. The distribution of the TortA parameter is shown where it can be seen that the y-axis ranges by several orders of magnitude. TortA and TortV were log transformed to normalise the values (Figure 5.11B). Figure 5.11C shows an example of a TortA outlier from the AD group with the arteriolar paths overlaid onto the image. There were some extremely tortuous arteriolar vessel paths but there was also an erroneous segmentation (highlighted by the white arrow) that led to an invalid measurement. This data point was removed from subsequent analysis where all other outliers were visually inspected and deemed not to contain segmentation errors and were included in the subsequent analysis.

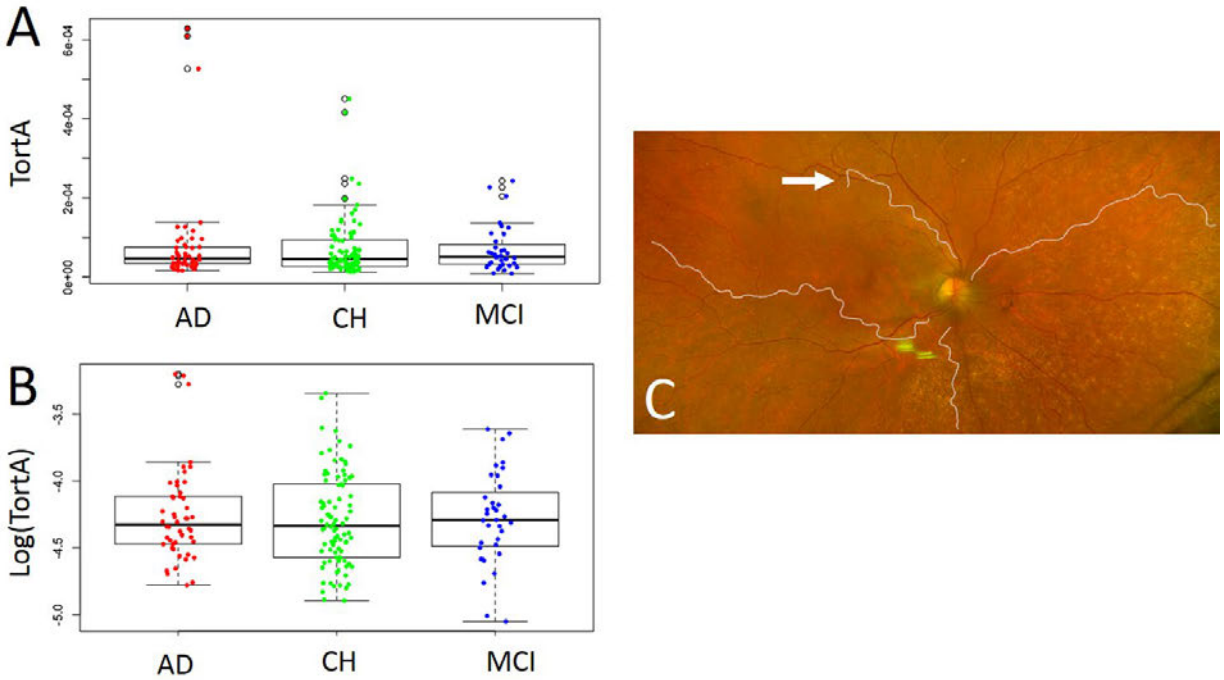

Figure 5.11. Example of a boxplots used to assess the data for outliers. **A)** A boxplot of the TortA vessel parameter. The y-axis is small and could be rescaled using a log transformation. **B)** Boxplot of the log transformed TortA. **C)** Example of a TortA outlier from the AD group with the arteriolar paths used to calculate TortA overlaid in white. This individual has extremely tortuous arteriolar paths giving a large tortuosity value, but there was an error in the segmentation as the path flows into the neighbouring venule (white arrow)

Figure 5.12 shows another example of a typical graphical representation of the vessel data used to investigate the sources of the outliers. The output of the vessel segmentation was overlaid onto the image where vessels that are close to one another have resulted in an erroneous classification of pixels as vessel. This results in a thicker and inaccurate width measure. 16 UWF images were removed from the analysis based on the aforementioned visual assessment (8 CH, 2 MCI, 6 AD).

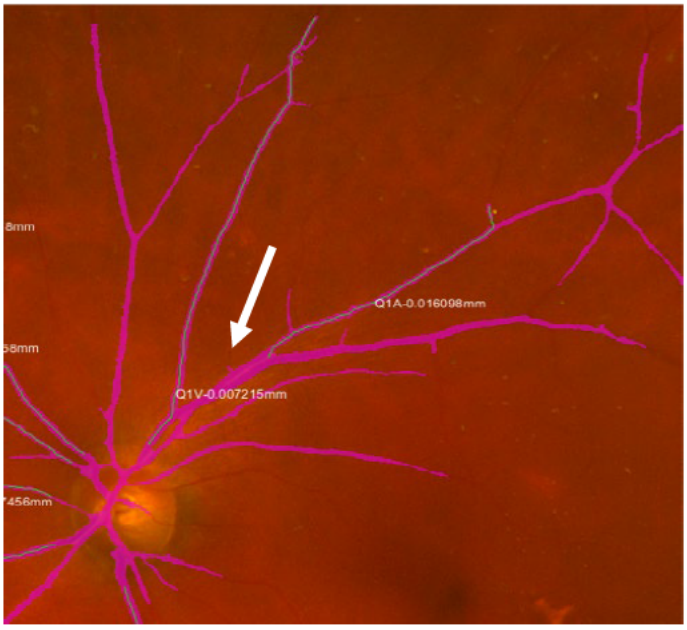

Figure 5.12 The 'raw' or uncorrected segmentation (pink) with the WG paths (blue) overlaid onto a UWF image. The outlier here was the WGaST (Q1A) path. The arteriolar width has been overestimated due to two vessels that are close together being merged by the segmentation process leading to an invalid thicker vessel-like structure (*arrow*). There was a slight media opacity to the overall image which could explain the reduced quality of the vessel segmentation.

To investigate the repeatability of the vessel measures, 12 participants with left and right eye images (i.e. 22 UWF images) were randomly selected from the dataset (4 CH, 4 MCI and 4 AD) for repeat ROI selection, manual correction and vessel measurements stages. Effect size (ξ) was calculated using a heteroskedastic one-way ANOVA for trimmed means[199] . This was chosen because it is a robust method to explore data without making any assumptions for equal variance or normal distribution and removes any possible outliers introduced in repeats [200]. The analysis was conducted in R (version 3.4.4) using the R package *ggstatsplot* [201]. An effect size value was scored as small if $0.10 < \xi < 0.30$, medium if $0.30 > \xi < 0.50$, and large if $\xi \geq 0.50$ [202]. Table 5.2 shows the results of comparing of the initial and repeated measure separated by left and right eye. It was observed that there was large effect size for area, FDa and FDv. In this thesis, a p-value is considered statistically significant if $p<0.05$ while small p-values are reported as '<0.0001'. There was a 40-70mm^2 (6- 10% of the total area that can be captured in a UWF image) mean difference in area sampled, which was less in the initial analysis than the repeats and had a non-significant ξ in the left eyes ($p = 0.12$) but a significant ξ in left eye ($p = 0.01$). This suggests that there was a significant variation in area as defined by me as the operator in left eyes. FDa was consistently larger in the left eyes of repeat measures than in the initial analysis and smaller in the right eyes of repeat measures, which was non-significant ($p > 0.05$). There was a mean difference of 0.20 in FDv of left and right eyes, which was significant ($p < 0.001$). This indicates that there was a significant increase in FDv in the repeat measures compared to the initial measures.

Given that significant large effect sizes were observed in both FDv and area measures of the ROI and the FD parameters are calculated from the entire ROI, two versions of the dataset will be analysed. The first is the initial data (referred to as the I-data) and a dataset with the original measures replaced with their repeats (referred to as the R-data). This was to investigate whether the differences between the initial and repeat measures effects the output of subsequent statistical analysis.

	Mean difference		ξ (95% CI)		p-value	
Vessel Parameter	Left eye	Right eye	Left eye	Right eye	Left eye	Right eye
Areamm2 (mm^2)	-70	-40	0.51 [-0.16, 0.77]	0.38 [-0.13, 0.76]	**0.02**	0.12
OD radius (mm^2)	-0.48	-0.25	0.04 [-0.06, 0.11]	0.09 [-0.14, 0.25]	0.50	0.30
OD fovea (mm^2)	-0.19	-1.56	0.02 [-0.17, 0.09]	0.04 [-0.14, 0.14]	0.85	0.53
CRAE (mm^2)	0.002	-0.002	0.04 [-0.25, 0.16]	0.13 [-0.09, 0.35]	0.80	0.42
CRVE (mm^2)	0.004	0.005	0.17 [-0.24, 0.44]	0.24 [-0.05, 0.49]	0.45	0.05
AVR (mm^2)	-0.01	-0.04	0.08 [-0.24, 0.25]	0.25 [-0.06, 0.53]	0.70	0.08
WGa ST	-0.001	-0.001	0.25 [-0.29, 0.59]	0.30 [-0.10, 0.67]	0.41	0.30
WGa IT	0.003	0.003	0.40 [-0.06, 0.84]	0.10 [-0.44, 0.37]	0.20	0.50
WGa IN	-0.004	0.001	0.44 [-0.04, 0.73]	0.14 [-0.15, 0.37]	0.24	0.40
WGa SN	-0.001	0.001	0.33 [-0.18, 0.79]	0.22 [-0.18, 0.53]	0.30	0.30
WGv ST	-0.0001	0.0001	0.05 [-0.48, 0.24]	0.16 [-0.27, 0.44]	0.82	0.53
WGv IT	0.0001	0.005	0.04 [-0.44, 0.19]	0.47 [-0.03, 0.86]	0.90	0.23
WGv IN	-0.002	0.001	0.26 [-0.14, 0.59]	0.20 [-0.17, 0.49]	0.30	0.50
WGv SN	0.001	-0.0003	0.03 [-0.41, 0.15]	0.00 [-0.30, 0.07]	0.80	0.97
Log(TortA)	-0.05	0.07	0.01 [-0.29, 0.08]	0.17 [-0.15, 0.42]	0.96	0.25
Log(TortV)	0.02	0.09	0.10 [-0.31, 0.30]	0.20 [-0.10, 0.47]	0.70	0.22
FDa	-0.0006	0.04	0.06 [-0.29, 0.17]	0.52 [-0.06, 1.08]	0.80	0.12
FDv	-0.20	-0.20	0.84 [-0.70, 0.88]	0.84 [-0.62, 0.89]	**<0.001**	**<0.001**

Table 5.2 Mean differences between initial and repeat measures of both left and right eyes with the calculated effect size (ξ) and 95% confidence intervals.

5.2.5 Cleaned dataset

Table 5.3 gives details of the participant demographics for the dataset following outlier removal. The final dataset contains 24 CH (41 eyes), 8 MCI (14 eyes) and 14 AD (25 eyes) individuals. Comparison of the means was obtained using 1-way ANOVA (when normally distributed) and Kruskalwallis (when non-normally distributed). The MCI group contained more individuals with hypertension than CH and AD. Whether an individual involved in the study was taking antihypertensives was not known. There was a statistically significant difference in age ($p= 0.01$) and in MMSE score ($p<0.05$) between groups. Post hoc Tukey Kramer test (for ANOVA) and Dunn & Sidak (for Kruskalwallis) was used to identify where this difference lies. For age there was a significant difference between CH and AD ($p<0.05$) and not CH vs MCI or MCI vs AD. There was a significant difference in MMSE between all groups ($p<0.05$).

Feature	CH	MCI	AD	P value for Difference [Test]
Number of participants (Female/Male)	24 (20/4)	8(3/5)	14(10/4)	-
(No. right) (No. left)	(17) (24)	(8)(6)	(14) (11)	-
Mean age (SD)	65 (6)	65(6)	72 (10)	0.01 †
Hypertension (%)	5	60	18	-
Mean MMSE score (SD)	29 (1)	26(1)	20 (5)	< 0.0001[1]

Table 5.3. Participant demographics of the cleaned dataset. If the values were normally distributed a 1-way ANOVA[1] was used to compare means, otherwise if not normally distributed Kruskalwallis[†]. Post hoc Tukey Kramer for ANOVA and Dunn and Sidak for Kruskalwallis was used to identify where differences lie. For age there was a significant difference between Controls and AD ($p<0.05$). There was a significant difference between all groups for MMSE scores ($p<0.05$). SD = standard deviation.

5.2.6 Statistical modelling

The aim of the analysis was to determine whether the vessel parameters differ between CH, MCI and AD groups. The methods of analysis that were considered will be discussed to highlight the reasoning for the final choice in the model used. In the

cleaned dataset the number of MCI participants was smaller than for the CH and AD groups and would not be enough to compare across 3 groups (i.e. using a test such as ANOVA) (see 5.2.5). For this reason, the analysis was conducted to investigate CH vs MCI and CH vs AD.

The dataset consisted of repeated measures, where two eyes of an individual were included (as well as single eye images for some participants). Repeated measures tend to be correlated with one another, especially when measured over short time intervals. For example, if we are interested in how height and weight changes over a 5-year period in 100 individuals, with measurements repeated twice a year, we know that there would be an underlying correlation between the repeated measures - i.e. two observations of an individual are more likely to be similar in value than two observations from two separate individuals. Similarly, we would not expect the two eyes in an individual to be independent as the underlying anatomy that connects the brain to the two eyes would be similar within an individual. For this reason, correlation between the measurements of two eyes of an individual should be considered during the analysis if possible.

Standard statistical techniques such as Wilcoxon rank sum and logistic regression (considered in this chapter, as they do not assume a normal distribution) assume an independence between measures. Wilcoxon rank sum is a nonparametric analysis that does not require normal distribution. Logistic regression is a method to model the group as the dependent variable (e.g. CH, MCI or AD) to the independent variable

(e.g. CRAE, CRVE) with use of multivariate analysis to control for potential confounders (e.g. age, ROI area). However, logistic regression does not permit a within-subject correlation structure. Table 5.4 summarises the models that were considered, the assumptions and the output.

Model	Assumptions	Reasoning	Output
Wilcoxon rank sum [203]	• Non-parametric (non-normal distribution) two-sample test. • Independence between samples.	Evaluate the influence of adjusting for the correlation between eyes	Estimates a p-value that describes whether the distribution between the groups have shifted to the left or right of one another.
Univariate/Multivariate logistic regression [204]	• Non-linear relationship between dependent and independent variables • Heteroskedasticity (unequal variance) • Non-normal distribution	Evaluate the influence of adjusting for the within-subject correlation and confounding variables (e.g. Age).	Log odds ratio. If the p-value of the odds ratio is significant (<0.05) the odds ratio can be interpreted as the odds that an outcome (i.e. CH) occurs in the presence of the independent variable (e.g. CRAE), compared to the odds of the outcome (i.e. MCI) occurring in the absence of the independent variable (e.g. CRAE). An odds ratio < 1 indicates a negative relationship and a positive odds ratio indicates a positive relationship.
Univariate/Multivariate GEE [205]	• An extension of linear regression. • No distribution assumed • Within-subject correlation structure defined	A univariate model aimed to establish the relationship between the dependent variable (e.g. CRAE) and the independent variable (e.g. CH). Multivariate model used to determine how holding the	β-coefficient is the estimate for one unit increase/decrease in the dependent variable (i.e. CRAE) for which there is an increase/decrease in the independent variable by the β-coefficient value in all subjects with adjustment for the correlation structure. A positive β-

• Heteroskedasticity (unequal variance)	other independent variables constant influences the output.	coefficient indicates an increase and a negative β-coefficient indicates a decrease. The p-value indicates whether this increase/decrease is statistically significant.

Table 5.4. Summarising the models that were considered for the analysis of vessel parameters between two groups.

Generalised estimation equations (GEE) is a method by which repeated measurements (covariates such as, for example, CRAE) from a group of subjects (e.g. CH vs MCI) are used to model an expected response for an individual based on the covariate [206]. GEE is an extension of linear regression. In linear regression a line is fit by iteratively rotating a line through the data and measuring the squared distance (residual) of the data point to the fitted line and fitting the sum of the squared residuals (so called ordinary least squares [207]). The linear regression equation m for the *j*th vessel parameter in the *i*th individual is given by,

$$m_{ij} = a + \beta s_{ij} + e_{ij} \qquad \textbf{(5.1)}$$

OR

$$E(m_{ij}) = a + \beta s_{ij}; \text{ error = Normal.} \qquad \textbf{(5.2)}$$

Where a is the intercept at s in the independent variable (e.g. CRAE), β is the gradient and e the error term uncorrelated with s. The error term assumes a normal distribution and equal variance and is the standard error of the estimates. $E(m_{ij})$ denotes the *j*th vessel parameter in the *i*th individual and is said to contain an *identity*

link because it is a direct linear prediction of the dependent variable [206]. Figure 5.13 shows what these parameters are in the model in a visual form.

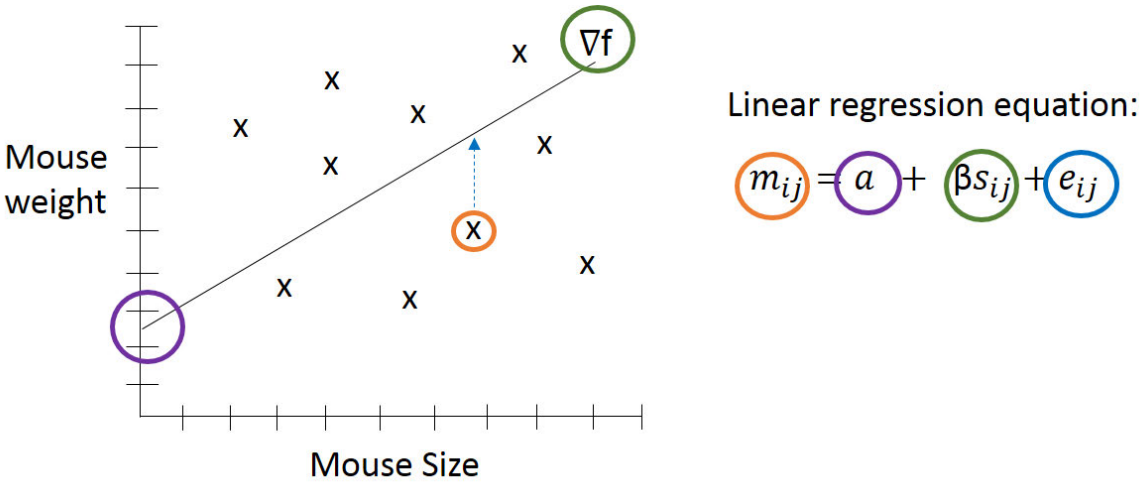

Figure 5.13 Illustrating the parameters in a linear regression equation. The hypothetical research question here is whether weight predicts the size. The equation for the regression coefficient (orange) is given by the intercept of the regression line a (purple) at the independent variable weight (green; s) with a gradient (∇f) given by β (green). The error (e) is the distance of the predicted value to the regression line.

A GEE model extends the linear model in two ways. First, the correlation outcomes within an individual are taken into account in the regression formulae that generates the regression coefficients and their standard errors. Secondly, the GEE model does not assume a normal distribution of the standard errors and therefore permits robust

estimates for the standard errors and their coefficients [205]. GEEs thus provide a method by which standard errors and their coefficients give consistent inferences even if the strength of the correlation within observations of an individual varies [205]. A GEE model is given by Equation 5.4 and the within-subject correlation is modelled separately. GEE modelling can be thought of in five steps:

1) Fit a standard regression model (Equation 5.1) where a and β are the regression coefficients.

2) Take the residual of the *j*th observation in the *i*th individual in the regression model, which is given by,

$$r_{ij} = m_{ij}\,(a + \beta s_{ij}) \quad . \qquad \textbf{(5.3)}$$

3) Use the residual r to form a matrix which is used to characterise the correlations between observations of an individual. This correlation structure is defined *a priori*. Equation 5.5 shows an exchangeable correlation structure and discussed more below).

4) Refit the regression model as a weighted regression model by modifying the formulae that incorporates the correlation matrix ($Corr$) given by,

$$m_{ij} = a + \beta s_{ij} + Corr + e_{ij} \quad . \qquad \textbf{(5.4)}$$

5) Generate new values for the regression coefficients (a and β) and calculate new residuals (r_{ij}).

6) Iterate through steps 2, 3, 4 and 5 until the estimates have stabilised and the model converges (i.e. the error is minimised between the residual and the regression line and the optimum model has been fitted).

GEEs permit a within subject correlation structure to be defined *a priori* that makes some assumptions. Although multiple correlation structures exist such as independence (no correlation), autoregressive (within-subject correlation decreases) or unstructured (no specification), in this thesis an *exchangeable* correlation structure was chosen (Equation 5.5). Also known as, *compound symmetry*, an exchangeable matrix assumes that every individual has an equal correlation between

his or her two eyes. This is characterised by the intraclass correlation coefficient (ICC) [208],

$$Corr = \begin{array}{c} \\ r_{ij} \\ r_{ij+1} \end{array} \begin{array}{c} \begin{array}{cc} r_{ij} & r_{ij+1} \end{array} \\ \begin{pmatrix} - & p \\ p & - \end{pmatrix} \end{array}$$

(5.5)

An exchangeable matrix was used to estimate the correlation parameter to quantify the within-subject correlation. In this matrix we have the residual (r) for the *i*th individual in for both eyes (*j, j+1)*. The correlation between repeated measures was calculated orthogonally using the ICC and is denoted as *p* and a dash (-) denotes which comparisons have not had applied – i.e. the *p* of the same eye of an individual.

Comparison of the outputs informs me whether the significant p-value appearing to show evidence of a significant difference is an over estimation due to the fact that both the Wilcoxon rank sum test and ordinary least squares regression model ignore the inherent correlation between measurements - i.e. a Type I error (false negative) has occurred [209]. The use of a multivariate GEE model also permits the comparison of the p-value signal when controlling for confounding variables. Linear regression

cannot not be performed because the data in this study does not meet the model assumptions of normal distribution.

It is against this backdrop that the group analysis was evaluated by comparing CH vs MCI, CH vs AD using Wilcoxon rank sum and univariate/multivariate GEE models to evaluate across all models how the signal from the p-value changed and evaluate possible Type I or Type II errors [210]. Table 5.5 in section 5.3.1 shows how the p-value changed between each test. Given the variety of the outputs across models it appears that the correlation between eyes while controlling for other parameters would be the most reliable and robust result. For this reasoning, the multivariate GEE model is presented.

Multivariate GEE analysis was performed whereby all independent variables were entered into the GEE formula and only included in the final multivariate model if statistically significant to the model. An independent variable is statistically significant if there is an association between the dependent variable and the independent variable of interest. If statistically significant, the variable is held constant so that the effects of the independent variable of interest to the dependent variable can be observed in the absence of this other association. Wilcoxon rank sum and GEE modelling was performed in R (version 3.4.3 [211]) using *geepack* R package [212].

Each GEE model was implemented using the following code:

```
geemodel = geeglm (Vessel parameter ~ Group, family = "gaussian",
corstr = "Exchangeable")
```

where the 'corstr' parameter is the assumed covariance matrix and the 'family' parameter assumes that the data follows Gaussian distributed (rather than a Poisson distribution which is typically applied in count data). The output of the model is the β-coefficient that is interpreted in a population averaged manner because it is calculated directly from the residuals after adjusting for all other covariates in a model. To illustrate, in this dataset I have calculated the regression coefficients from between the two eyes of an individual that can therefore be thought of as an '*average*' from that individual. When comparing between the groups, the comparison being made was essentially the comparison between the estimated average of all individuals. In this case, a positive or negative β-coefficient indicated the average increase or decrease of the population and the p-value and 95% confidence interval helped me to evaluate the statistical significance between of the association.

5.3 Results

5.3.1 Model comparison

Table 5.5 summarises where significant differences were observed for the Wilcoxon rank sum (WRS), univariate GEE (UGEE) and multivariate GEE (MGEE) models when comparing CH with MCI and CH with AD (using the I-data). There was a significant difference in TortV ($p = 0.01$) when comparing CH and AD using WRS but this was non-significant in the UGEE and MGEE models (corrected for area and age). This suggested that there is a correlation of TortV measures between eyes that was accounted for in the GEE models and therefore was a false positive in the WRS model. There was no evidence of a significant difference in the WRS model for FDv but a significant difference was observed in the UGEE model ($p = 0.01$) and the MGEE model ($p = 0.01$) following correction for area. There was a significant difference of WGaST ($p = 0.04$), WGaIN ($p = 0.02$) and WGvIT ($p=0.04$) when comparing CH with AD using WRS. Given that the signal for TortV, WGaST, WGaIN and WGvIT was not present when accounting for within subject variation, it suggested that the results would be more reliable when analysed using GEE's. WGaIN and WGvIT signal was not present following correction for age which suggested that a multivariate GEE would be a more robust and consistent method to analyse the dataset.

	CH vs MCI			CH vs AD		
	WRS	UGEE	MGEE	WRS	UGEE	MGEE
CRAE	0.40	0.50	0.50	0.90	0.90	0.90
CRVE	0.40	0.20	0.30	0.90	0.60	0.60
AVR	0.90	0.90	0.90	0.80	0.70	0.70
WGa ST	0.20	0.30	0.50	**0.04**	0.30	0.30
WGa IT	0.90	0.50	0.50	0.30	0.30	0.20
WGa IN	0.40	**0.02**	0.40	**0.02**	**0.02**	**0.02**
WGa SN	0.80	0.90	0.90	0.30	0.40	0.30
WGv ST	0.80	0.20	0.20	0.60	0.20	0.06
WGv IT	0.06	**0.05**	0.06	**0.04**	0.20	0.20
WGv IN	0.70	0.70	0.70	0.90	0.90	0.90
WGv SN	0.80	0.70	0.70	0.40	0.60	0.20
TortA	0.60	0.40	0.40	0.30	0.70	0.10
TortV	**0.01**	0.10	0.10	0.30	0.10	0.10
FDa	0.40	0.80	0.80	0.10	0.20	0.75
FDv	0.45	**0.01**	**0.01**	0.60	0.30	0.20

Table 5.5 P-values of statistical testing for each vessel parameter when comparing groups using the different models to evaluate possible Type I or Type II errors.

The purpose of this analysis was to determine if accounting for within subject correlations was required to further analyse the data. If the signal did not change between models, it would suggest that within-subject variation is not essential for the analysis of the I-data and the R-data. The use of univariate and multivariate models indicated that correcting for age or area is required to give consistent results – as shown by the change in signal when comparing CH with MCI for the WGaIN and WGaIT following correction for age.

5.3.2 Vessel width calibres

Figure 5.14 shows boxplots of the CRAE and CRVE parameters measured for the three participant groups. Table 5.6 and Table 5.7 display the β-coefficients for the GEE univariate models with the I-data and R-data respectively. When comparing CH to MCI there was a decrease in CRAE and CRVE observed in the I-data and R-data but with no evidence of a statistically significant difference. When comparing CH to AD there was an increase in CRAE in the I-data and a decrease in CRAE in the R-data, but again no evidence of a significant difference. There was a decrease in CRVE in the I-data and an increase in CRVE in the R-data that was non-significant. Outliers were observed in I-data and R-data for CRAE and CRVE and these were visually inspected to determine whether they might be invalid measurements. Figure 5.14 shows an example of this visual inspection in Zone B where the CRVE calibres are derived and displays the outlier (Figure 5.14A; CRVE = 0.16mm^2) next to an image with an average

CRVE (Figure 5.14B; 0.20mm^2). In the outlier skeleton there is a smaller superonasal vessel branch that has been incorrectly segmented leading to a small CRVE.

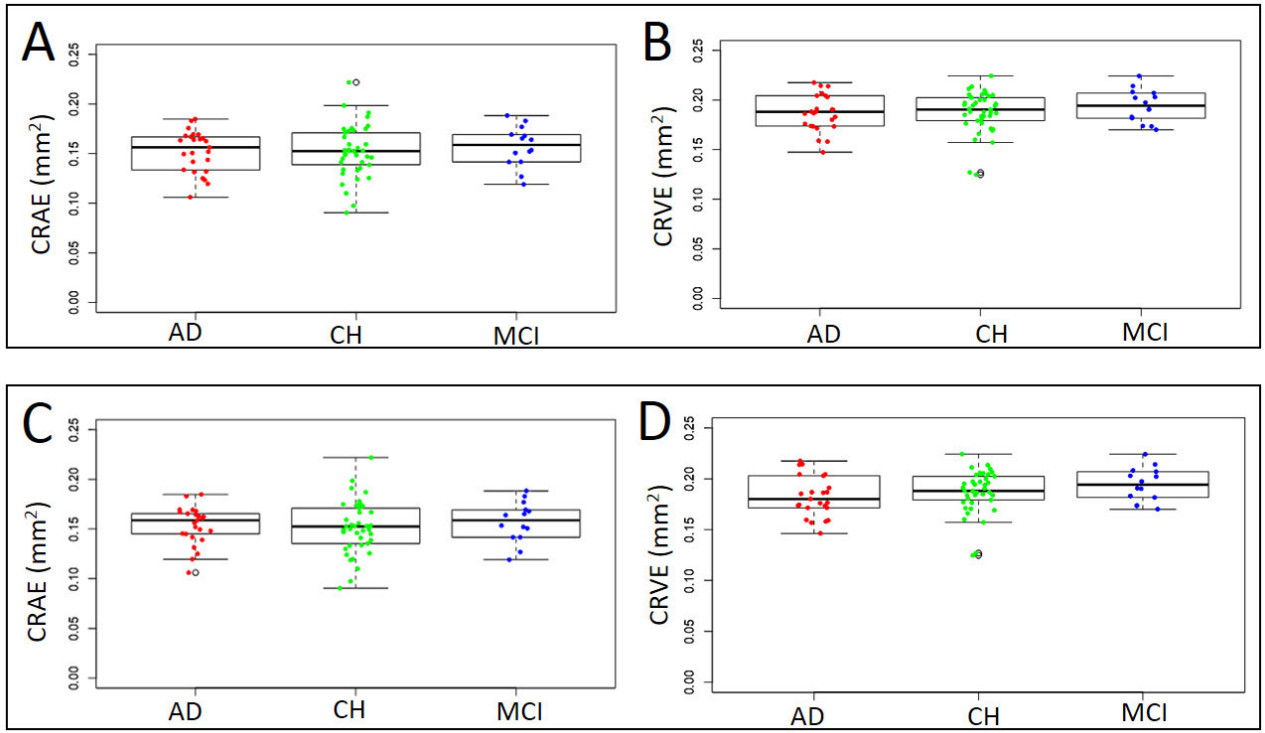

Figure 5.14 Boxplots of CRAE and CRVE measures in each group. **A-B)** boxplots of CRAE and CRVE measures from the I-data. **C-D)** boxplots of CRAE and CRVE measures from the R-data.

				β Coefficient (95% Confidence Intervals) p-value	
UWF vessel parameter	CH (Mean ± SD)	MCI (Mean ± SD)	AD (Mean ± SD)	CH vs MCI	CH vs AD
CRAE (mm^2)	0.2±0.02	0.2±0.02	0.2±0.02	-0.01 (-0.02 to 0.01) p = 0.50	0.001 (-0.01 to 0.01) p = 0.88
CRVE (mm^2)	0.2±0.01	0.2±0.01	0.2±0.03	-0.01 (0.02 to 0.01) p = 0.25	-0.003 (-0.02 to 0.01) p = 0.60
AVR (mm^2)	0.8±0.14	0.8±0.13	0.8±0.13	0.01 (-0.1 to 0.1) p = 0.88	0.01 (-0.07 to 0.10) p = 0.72

Table 5.6 Means and standard deviations (SD) of CRAE, CRVE and AVR from the I-data. β-coefficients with the 95% confidence intervals and p-values for differences between groups.

				β Coefficient (95% Confidence Intervals) p-value	
UWF vessel parameter	CH (Mean ± SD)	MCI (Mean ± SD)	AD (Mean ± SD)	CH vs MCI	CH vs AD
CRAE (mm^2)	0.2±0.03	0.2±0.02	0.2±0.02	-0.01 (-0.02 to 0.01) p = 0.50	-0.001 (-0.01 to 0.01) p = 0.89
CRVE (mm^2)	0.2±0.02	0.2±0.02	0.2±0.03	-0.01 (-0.02 to 0.01) p = 0.24	0.0001 (-0.01 to 0.01) p = 0.99
AVR (mm^2)	0.8±0.2	0.8±0.15	0.8±0.2	0.01 (-0.1 to 0.1) p = 0.89	-0.01 (-0.1 to 0.1) p = 0.78

Table 5.7. Means and standard deviations (SD) of CRAE, CRVE and AVR from the I-data. β-coefficients with the 95% confidence intervals and p-values for differences between groups.

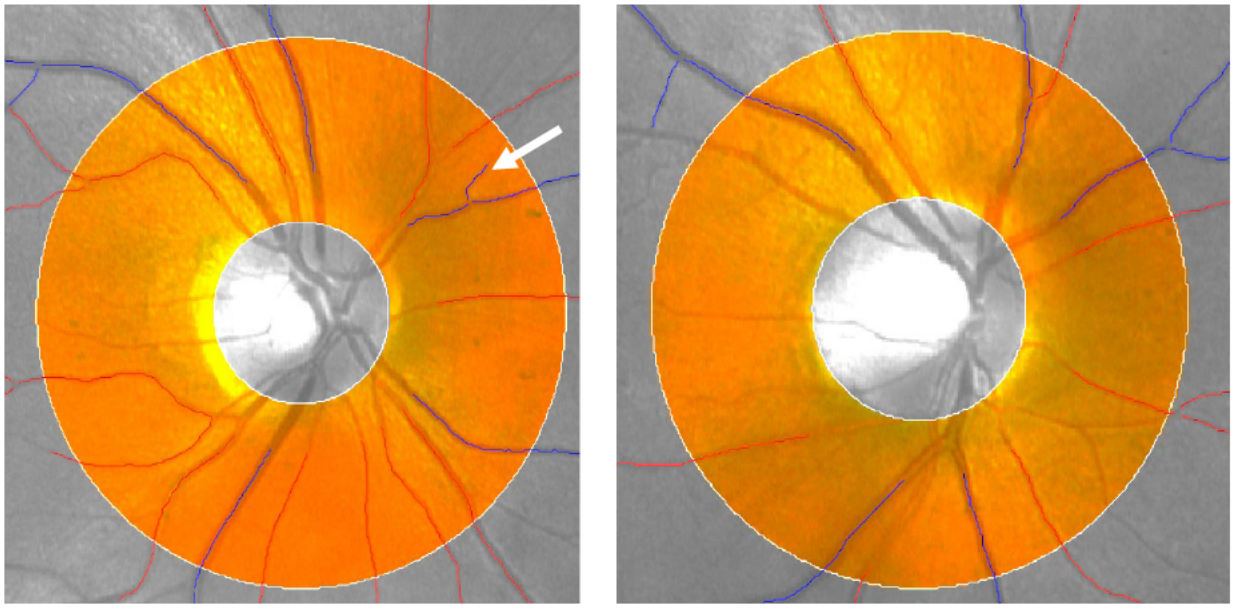

Figure 5.14 Visualisation of an outlier observed in the dataset. **A)** Zone B of a CRVE outlier (from the CH group) that was observed in the I-data and the R-data. There is a small venule branch (*arrow*) that has been incorrectly segmented and measured contributing to the small CRVE (0.16mm). **B)** Zone B of a CH individual with an average CRVE (0.2mm).

5.3.3 Vessel width gradients

Table 5.8 and Table 5.9 give the GEE analysis with the I-data and with the R-data, respectively. There was no evidence of a significant difference in WGaST, WGaIT, WGaSN, WGvST, WGvIT and WGvIN when comparing CH with MCI and CH with AD. There was a significant difference in WGaIN when comparing CH with AD in both the I-data ($p = 0.02$) and the R-data ($p = 0.01$). Figure 5.15 shows the distribution of

WGaIN of the I-data and the R-data in boxplot form. Figure 5.16A shows an individual with AD and a mean WGaIN (-0.006) and an individual who is CH and who has a mean WGaIN (-0.005). From the boxplot it can be seen that the difference in WGaIN was from the CH group given by the larger mean WGaIN (-0.006) compared to the AD group (-0.005). The representative examples in Figure 5.16 show the vessel path from which the width gradient was calculated (cyan) displayed over the image and the widths measured along the vessel path. The AD individual appears to have a thinner path in the inferior nasal quadrant compared to the individual who is CH.

				β Coefficient (95% Confidence Intervals) p-value	
UWF vessel parameter	CH (Mean ± SD)	MCI (Mean ± SD)	AD (Mean ± SD)	CH vs MCI	CH vs AD
WGa ST	-0.005±0.004	-0.005±0.002	-0.006±0.003	-0.006 (-0.005 to 0.001) p = 0.94	0.001 (-0.001 to 0.002) p = 0.60
WGa IT	-0.005±0.004	-0.007±0.01	-0.005±0.004	0.002 (-0.004 to 0.007) p = 0.60	0.0003 (-0.001 to 0.002) p = 0.71
WGa IN	-0.005±0.003	-0.006±0.004	-0.006±0.004	0.001 (-0.001 to 0.004) p = 0.32	0.002 (-0.001 to 0.003) **p = 0.01**
WGa SN	-0.003±0.002	-0.003±0.002	-0.003±0.003	-0.002 (0.001 to 0.001) p = 0.50	-0.001 (-0.002 to 0.001) p = 0.37
WGv ST	-0.006±0.003	-0.006±0.001	-0.006±0.002	-0.0005 (-0.001 to 0.0005) p = 0.34	-0.0002 (-0.001 to 0.001) p = 0.78
WGv IT	-0.007±0.004	-0.006±0.008	-0.006±0.003	-0.0004 (-0.005 to 0.004) p = 0.86	-0.001 (-0.003 to 0.003) p = 0.13
WGv IN	-0.007±0.004	-0.007±0.003	-0.006±0.003	-0.0003 (-0.003 to 0.002) p = 0.77	-0.001 (-0.002 to 0.001) p = 0.31

Table 5.8. Means and standard deviations (SD) of WG's in each region from the I-data. β-coefficients with the 95% confidence intervals and p-values for differences between groups with corrections for age[†] when statistically significant to the model, otherwise no corrections.

				β Coefficient (95% Confidence Intervals) p-value	
UWF vessel parameter	CH (Mean ± SD)	MCI (Mean ± SD)	AD (Mean ± SD)	CH vs MCI	CH vs AD
WGa ST	-0.005±0.003	-0.006±0.003	-0.006±0.002	0.0007 (-0.001 to 0.003) p = 0.48†	0.001 (-0.001 to 0.002) p = 0.33
WGa IT	-0.004±0.003	-0.004±0.002	-0.005±0.003	0.001 (-0.001 to 0.002) p = 0.50	0.001 (-0.001 to 0.003) p = 0.16
WGa IN	-0.005±0.003	-0.007±0.004	-0.007±0.004	0.002 (-0.001 to 0.003) p = 0.40†	0.003 (0.0005 to 0.005) **p = 0.02**
WGa SN	-0.003±0.002	-0.003±0.002	-0.004±0.002	0.00005 (-0.002 to 0.002) p = 0.99	0.0005 (-0.001 to 0.002) p = 0.39
WGv ST	-0.006±0.003	-0.006±0.001	-0.006±0.002	-0.001 (-0.002 to 0.0004) p = 0.19	-0.009 (-0.002 to 0.00006) p = 0.07†
WGv IT	-0.006±0.003	-0.005±0.003	-0.005±0.001	-0.001 (-0.003 to 0.0002) p = 0.05	-0.001 (-0.002 to 0.001) p = 0.22
WGv IN	-0.007±0.004	-0.007±0.003	-0.006±0.002	0.001 (-0.002 to 0.003) p = 0.71	0.00009(-0.002 to 0.002) p = 0.92

Table 5.9. Means and standard deviations (SD) of WG's in each region from the R-data. β-coefficients with the 95% confidence intervals and p-values for differences between groups with corrections for age† when statistically significant to the model, otherwise no corrections.

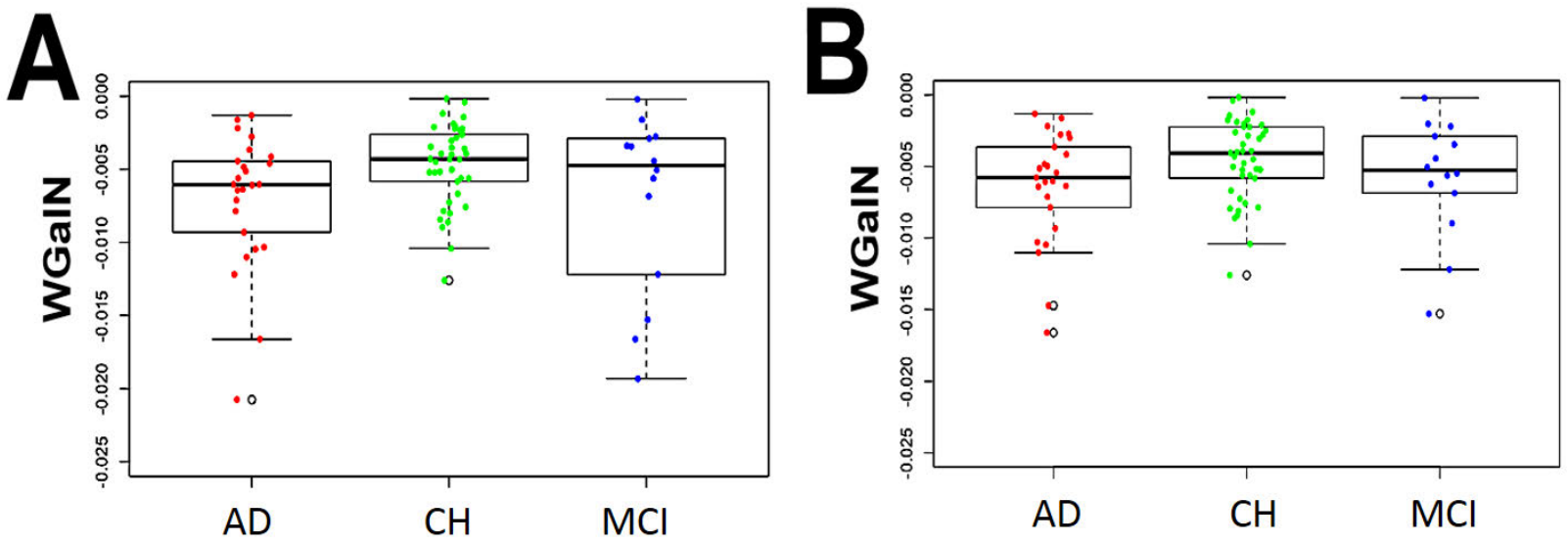

Figure 5.15 Boxplot of WGaIN for the I-data **(A)** and the R-data **(B)**. The mean WGaIN for CH was the same in the I-data and R-data (-0.005±0.003). There was an increase of 0.001 in the mean WGaIN of MCI and AD in the R-data compared to the I-data. This was

reflected as a 0.001 increase in the β-coefficient and was statistically significant in the R-data and I-data suggesting that the result was repeatable.

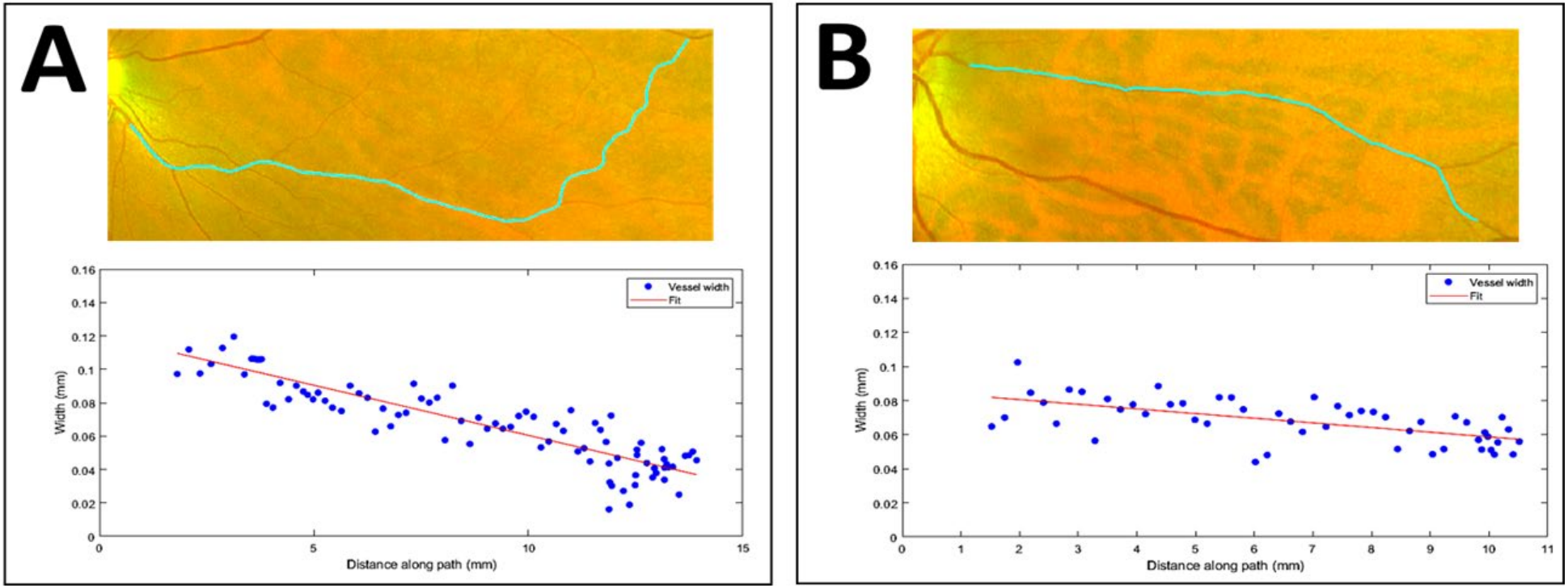

Figure 5.16 A and **B** shows representative examples of an AD participant and CH participant, respectively. The main vessel arteriolar path from which the WGaIN is calculated is overlaid in cyan on to the image. The individual widths measured along the path are displayed underneath where a straight line is fitted to the data points. The AD individual appears to have thinner widths measured along the path compared to the CH individual resulting in a higher negative width gradient.

5.3.4 Vessel tortuosity and fractal dimensions

Table 5.10 and Table 5.11 display the GEE analysis with the I-data and R-data, respectively. When comparing CH with MCI there appeared to be a decrease in TortA and TortV in the univariate GEE models using both the I-data and the R-data, though there was no evidence of a significant difference. Figure 5.16 shows the FDa and FDv distribution derived from the I-data and the FDa and FDv from the R-data. Comparison of CH with MCI showed an increase in FDa but no evidence of this being a significant difference (following correction for area). Comparing CH with AD showed an increase in FDa ($p = 0.20$) in the I-data, following and prior to, correcting for area and age where there was no evidence of a significant difference. In the R-data when comparing CH to AD showed an increase in FDa where a significant difference prior to and following correction for area was observed ($p = 0.01$). Comparing CH with MCI there was a significant increase in FDv ($p = 0.01$) prior to and following a correction for area in the I-data and an increase in FDv in the R-data, though not with a level considered significant ($p = 0.71$). In the R-data, comparing CH with AD seemed to indicate a decrease in FDv, but this was without achieving a level of significance. From distribution of FDa it appears as if the R-data contains more outliers as shown in the boxplot.

				β Coefficient (95% Confidence Intervals) p-value	
UWF vessel parameter	CH (Mean ± SD)	MCI (Mean ± SD)	AD (Mean ± SD	CH vs MCI	CH vs AD
Area (mm^2)	417±110	463±81	470±71	-	-
Log(TortA)	-4.3±0.4	-4.3±0.2	-4.2±0.4	1.7 (-2.3 to 5.7) p = 0.41	-1.6 (-3.7 to 0.46) p = 0.13
Log(TortV)	-4.3±0.2	-4.2±0.3	-4.3±0.3	-2.5(-5.8 to 0.80) p = 0.14	-9.2 (-4.7 to 72.2) p = 0.06†
FDaEntire	1.41±0.04	1.41±0.03	1.40±0.05	0.001(-0.001 to 0.03) p = 0.30$^{\wedge}$	0.02 (-0.01 to 0.04) p =0.20$^{\dagger\wedge}$
FDvEntire	1.25±0.04	1.28±0.03	1.27±0.06	0.002 (-0.004 to 0.005) **p = 0.01**$^{\wedge}$	0.02 (0.04 to 0.01) p = 0.20$^{\dagger\wedge}$

Table 5.10 Means and standard deviations (SD) of FD and Tort parameters from the I-data. β-coefficients with the 95% confidence intervals and p-values for differences between groups with corrections for corrections for age† and area$^{\wedge}$ (mm^2) when statistically significant to the model, otherwise no corrections.

				β Coefficient (95% Confidence Intervals) p-value	
UWF vessel parameter	CH (Mean ± SD)	MCI (Mean ± SD)	AD (Mean ± SD)	CH vs MCI	CH vs AD
Area (mm^2)	430±113	476±77	489±86	-	-
Log(TortA)	-4.3±0.4	-4.3±0.2	-4.2±0.4	-0.02 (-0.2 to 0.2) p = 0.86	-0.09 (0.3 to 0.1) p = 0.44
Log(TortV)	-4.3±0.2	-4.2±0.3	-4.3±0.3	-0.10 (-0.3 to 0.1) p = 0.39	-0.01 (-0.1 to 0.1) p = 0.90
FDaEntire	1.43±0.05	1.43±0.04	1.40±0.09	-0.002(-0.002 to 0.02) p = 0.90^	0.01 (0.01 to 0.1) **p = 0.01**^
FDvEntire	1.28±0.08	1.36±0.10	1.31±0.10	-0.07(-0.2 to 0.01) p = 0.07	-0.003 (-0.05 to 0.05) p = 0.90^

Table 5.11 Means and standard deviations (SD) of FD and Tort parameters from the R-data. β-coefficients with the 95% confidence intervals and p-values for differences between groups with corrections for corrections for age† and area^ (mm^2) when statistically significant to the model, otherwise no corrections.

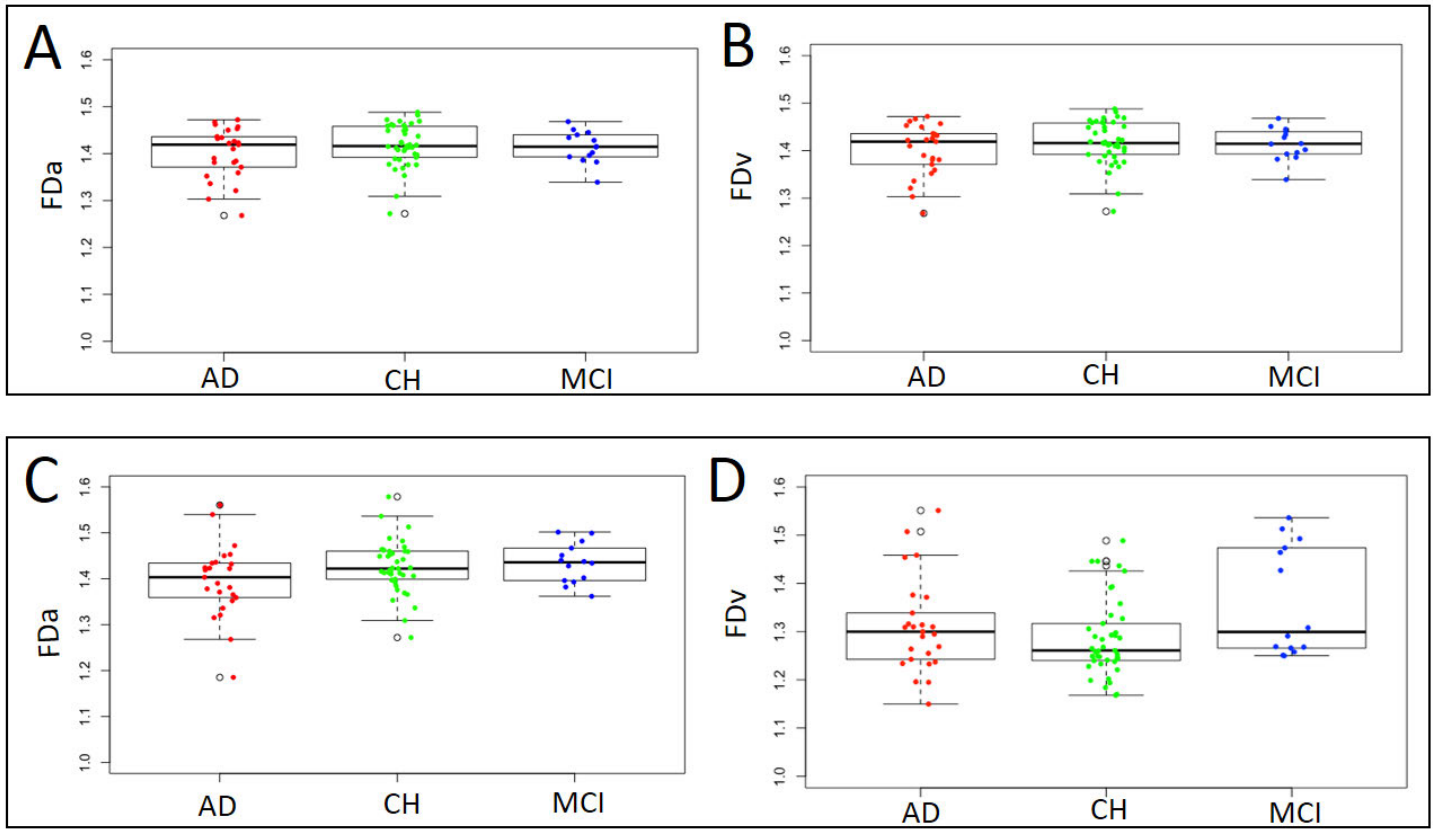

Figure 5.17 Boxplots of FDa and FDv for the O-data (A & B) and the R-data (C & D)

5.4 Discussion

Multivariate GEE modelling was concluded to be a reliable and consistent method for analysing datasets that contain within subject correlations. This was shown by the variation in signal when comparing to other non-parametric modelling techniques, which suggested that these methods may be prone to Type I errors when analysing retinal datasets. In [210], the authors showed that ignoring the correlation between eyes of an individual can lead to an inflated p-value when using paired tests. This was observed in this Chapter and showed that the result for my venular tortuosity parameter (TortV) was a false positive when using Wilcoxon rank sum and univariate logistic regression to compare between groups. Changes in tortuosity have previously been reported in the literature when analysing fundus images of individuals with AD, though results have been inconsistent. In [48], the authors reported an increase in arteriolar and venular tortuosity and used multivariate logistic regression in their dataset that contained both eyes of an individual. In [112], the authors reported a significant decrease in venular tortuosity but no change in arteriolar tortuosity and used a χ^2 test to compare a dataset containing both eyes of individuals with AD and healthy controls. In [213], authors used logistic regression to analyse fundus images of CH and AD individuals using right eyes only. The authors reported a significant decrease in arteriolar tortuosity of AD individuals following adjustment for

confounders (e.g. age, hypertension). Although the authors used only the right eyes it would be interesting to see whether this significance persists in the left eyes. If the significance did not persist when analysing eyes separately it would suggest that there is variation between eyes that needs to be included in the model to give a more accurate representation of the data and to determine any differences between AD and CH groups. Given the inconsistencies in the literature, more data is needed to understand the role of tortuosity as a potential biomarker for pathological cognitive decline but statistical techniques such as multivariate GEE would allow for a reliable and consistent analysis.

In my dataset, there was a significant increase in the complexity of the arteriolar network (FDa) observed when comparing CH with AD in the repeat data, but this was not recurrent in the initial data. Similarly, a significant increase in the complexity of the venular network (FDv) was observed when comparing CH with MCI in the initial data, but the finding was not observed in the repeat data. The likely source of this disparity is from the variation in the area which FD was calculated, which has three possible effects on the measure. In general, more area was included in the repeat data than the initial data (see section 5.3.4). An increase in area increases the amount of the vasculature that is used for the fractal dimension calculation and could result in a larger value. Figure 5.18 shows an example of a CH individual who showed an increase in FDa on repeat analysis. The initial ROI (white) and the repeat ROI (green) are overlaid onto the UWF image. The initial segmentation of the arteriolar and

venular paths are shown in red and blue, respectively. The extension of this segmentation in the repeat analysis is also displayed with arteriolar and venular paths are shown in cyan and magenta respectively. Either the additional paths can occur from the ROI positioning capturing more of the vasculature or from differences when performing the manual correction of stages (e.g. a path was removed in the initial analysis but not in the repeat). In this example, the ROI area has increased from 396mm^2 to 643mm^2, which has led to an increase in FDa from 1.38 to 1.46 and an increase in FDv from 1.25 to 1.45.

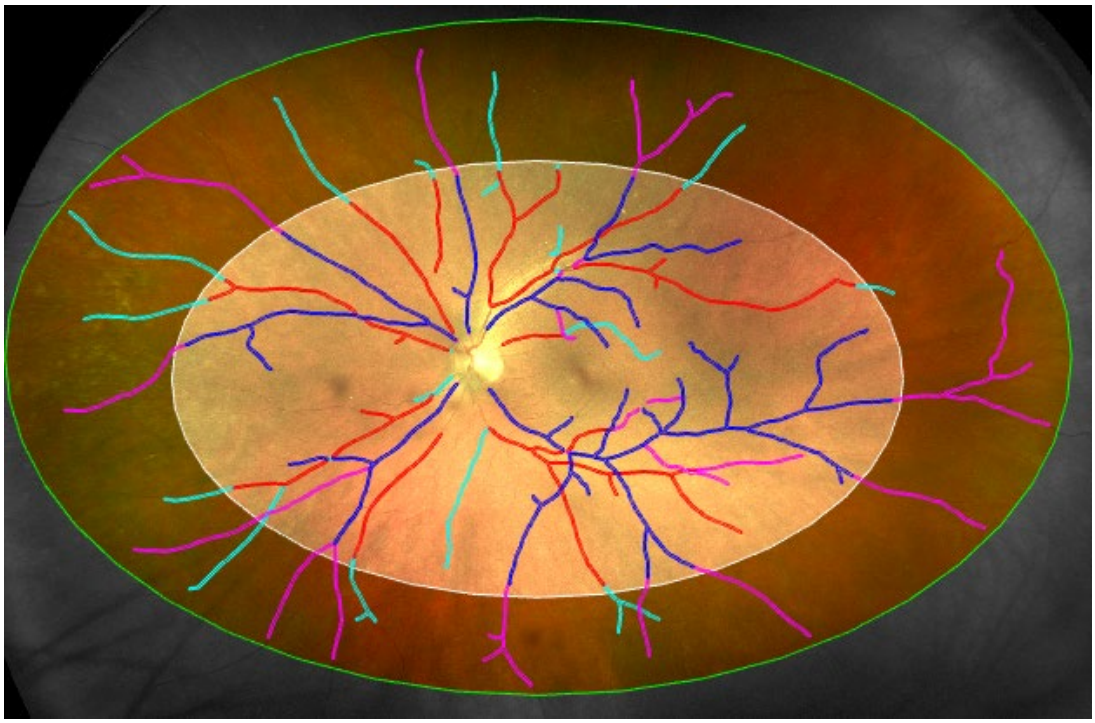

Figure 5.18 Image illustrating the increase in FDa and FDv that can occur due to a larger ROI in the repeat analysis (green circle) compared to the initial ROI (white circle). The initial arteriolar and venular paths are shown in red and blue respectively. The extension of the arteriolar and venular paths in the repeat analysis are shown in cyan and magenta, respectively.

It is also possible for an increase in ROI area to lead to a decreased fractal dimension. This is because the fractal dimension calculation takes into account the empty black space (i.e. non-vessel pixels) as well as the object (i.e. vessel pixels) to quantify the complexity of the structure (i.e. the vessel pattern). In individuals with a sparser vascular pattern, an increased ROI would result in a smaller fractal dimension. Figure

5.19 shows an example of an individual who had a low FDa and FDv measures in both the repeat and the initial analysis. The initial ROI (white) and the repeat ROI (green) has increased by 68.2mm^2, leading to a decrease in FDa and FDv by 0.12 even though the arteriolar and venular paths are the same in both analysis.

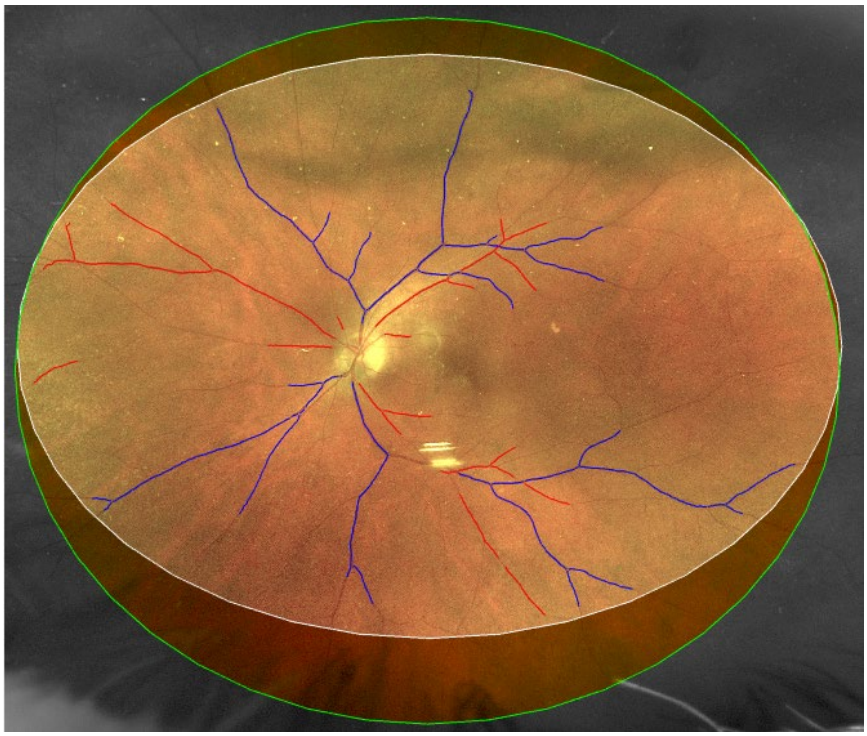

Figure 5.19 Image illustrating how an individual with a sparse vasculature network can show a decrease in FDv and FDa due to more empty space being introduced by a larger ROI in the repeat (green circle) compared to the initial ROI (white circle) even though the segmentation remained the same. Arteriolar network is labelled red and the venular network is labelled blue.

A different segmentation on the repeat analysis can lead to a different FDa or FDv value despite the size of the ROI. Figure 5.20 shows a CH individual with the initially segmented vasculature and the initial ROI (white) and the repeat ROI (green) overlaid onto the image. In this individual, the original ROI is 60mm^2 larger than the repeat ROI. The arteriolar (red) and venular (blue) trees from the initial segmentation are overlaid with the extension of this segmentation in the repeat (arteriolar – cyan; venular - magenta) analysis is also shown. In this individual, the repeat segmentation was different due to variation in manual correction stage and a smaller ROI than the initial ROI capturing less empty space. This resulted in an increase FDa from 1.47 to 1.48 and an increase FDv from 1.30 to 1.49.

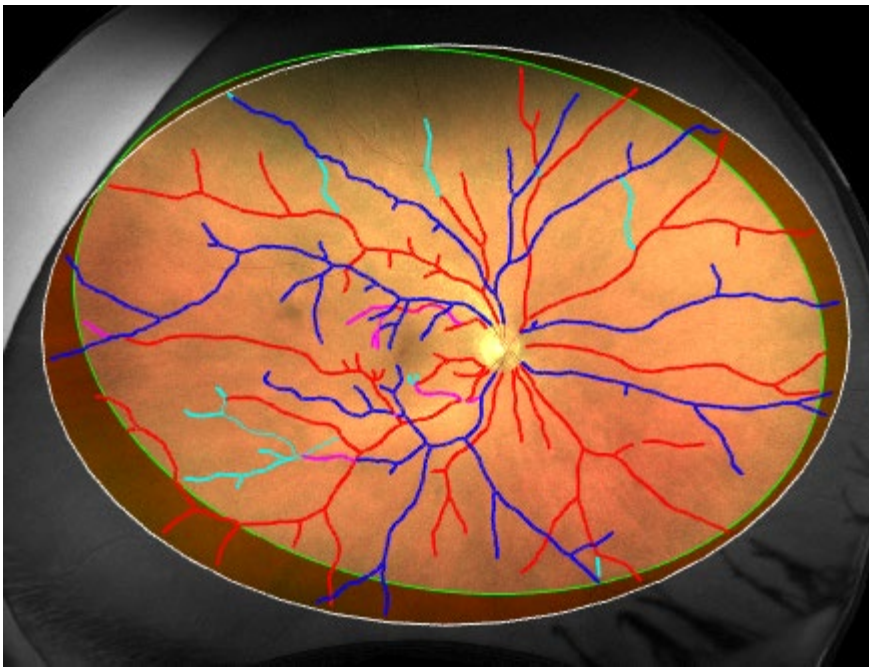

Figure 5.20 An image of an individual that showed an increase in FDa and FDv in the repeat analysis despite the larger ROI in the initial analysis (white circle) compared to the repeat analysis (green). The initial arteriolar (red) and venular (blue) segmentation original is overlaid with the extra (cyan) and venular (magenta) paths found in the repeat. The segmentation in the repeat analysis included more smaller branching vessels contributing to the increase in FDa and FDv values.

It can be concluded that in the repeat data more of the UWF image was sampled and this led to variations in the fractal dimension results. In previous work, a significant decrease in FDa was observed in AD individuals compared to controls, i.e. sparser arteriolar vascular network [91]. The sampling area affected the FD value even following correction for area in the model. The reliability of a fractal dimension

measure has previously been shown to be influenced by ROI placement [214]. The effects of the ROI placement on FD values could be overcome in future analysis by defining a consistent zone for extracting the FD measure throughout the dataset or to develop a new metric of branching complexity of the retinal vasculature in UWF images (i.e. standardisation of FD calculations) (see Chapter 7). My results show that researchers working in this field should take great care when interpreting FD measures and that reporting a repeat analysis would be important to demonstrate that the analysis is consistent and the results are reliable.

There was a consistent significant difference in WGaIN when comparing CH with AD that was present in the both the initial and repeat data. Previously, an increase in WGvSN was observed in UWF images of individuals with AD [91] but this was not observed in this dataset. In their study, the authors also described an increase in peripheral drusen that was located in the superonasal quadrant. They hypothesised that the increased WGvSN in AD resulted in venular thinning and suboptimal clearance of the peripheral retina leading to drusen deposition. This study involved 59 AD (mean age 79 years, SD 8.4 years) and 48 healthy controls (mean age 71 years, SD 10.4 years) and therefore had more AD participants than in my study while all their participants were older and better age-matched than the participants in the dataset analysed in this chapter. These factors could explain why I was not able to replicate previous findings. More data would be required to understand whether the

WGaIN increase, observed in this dataset, has a correlation to drusen deposition (see Chapter 6).

All participants in this study were over the age of 50 and changes to the retinal structure and function that occur in natural ageing of the eye would likely be present. Hence, subtle biomarkers specific to diseases such as AD might be difficult to discern from natural ageing.

5.5 Conclusions

GEE's are a consistent and robust method to provide valid inference of retinal datasets that contain both eyes of an individual. Given the inconsistent results of fractals that were observed, it is difficult to conclude whether fractal dimension might associate with AD. This chapter has highlighted the need for the standardisation of extracting fractal dimension measures, particularly when performed on a UWF image that has a larger FOV than a traditional fundus camera. The retinal vascular data presented here suggests that there may be subtle changes occurring in the retina in pathological cognitive decline pertaining to AD. It is believed that changes occurring in the brain connected to AD occur much earlier than the outward presentation of symptoms of significant cognitive decline. Thus, retinal vascular changed mirror brain vascular changed could also occur much earlier in life. Longitudinal images of mid-life individuals would therefore be necessary to reveal these.

It is believed that changes occurring in the brain connected to AD occur much earlier than the presentation of significant cognitive decline. Thus, retinal vascular changes that mirror brain vascular changes could also occur much earlier in life. Longitudinal retinal images of mid-life individuals would shed light on the nature of these changes.

Chapter 6 Exploratory drusen analysis

6.1 Introduction

This chapter outlines the exploratory analysis of the dataset presented in Chapter 5 using the automatic system described in Chapter 4**.** Quantitative measures performed on UWF images could be used to study cross-sectional differences between patients or longitudinal changes in individuals to facilitate investigations into drusen formation and progression as a biomarker of neurodegeneration. However, because the detector had a low sensitivity we should bear in mind that there will be a large proportion of drusen that are not detected giving an under estimation of the true amount present in an image. In this chapter I investigate differences between individuals deemed cognitively healthy and those with MCI and AD.

6.2 Materials and methods

93 CH, 41 MCI and 50 AD images were analysed for drusen. This dataset comprised images prior to data cleaning (Section 5.2.4) of invalid measurements of vessel parameters. One image from a CH individual was excluded due to the presence of artefact (see Figure 6.2). Drusen were detected using the automatic system described in Chapter 4 whereby a Manchester grid is overlaid over the input image and each

cell is classified as either *positive* for drusen or *negative* for drusen. Counts of grid cells deemed *positive* were obtained for each of the zones previously described in Section 4.3.4 as well as an overall total drusen load for every image. In Section 4.3.2 the ROC curve showed the performance of the classifier at various thresholds. At a threshold of 0.5 the curve was closest to the top left corner of the ROC plot and represents the optimum threshold whereby the true positive rate is maximised and false positives are minimised. A cell was considered *positive* if the output probability value was more than 0.5.

Figure 6.1 shows the distribution of the total counts per image visualised as a dot plot. Given the low sensitivity of the drusen detector the images with high counts were visually inspected to assess whether these featured cells that were false positives or true positives.

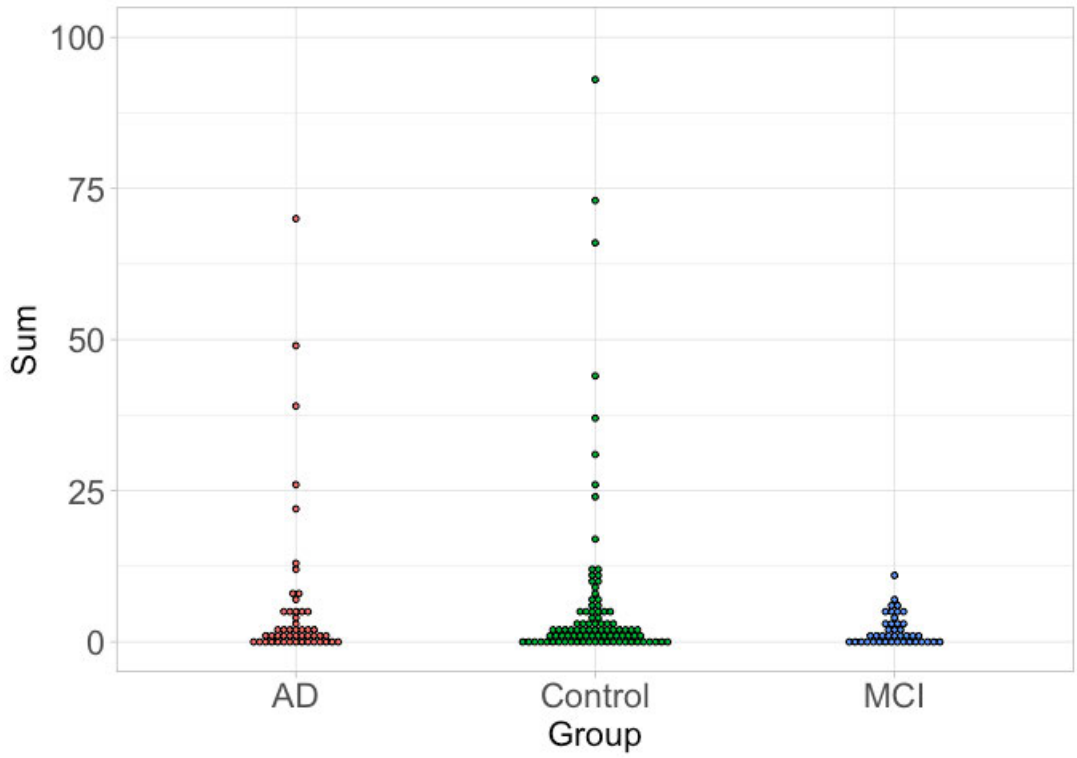

Figure 6.1 Dot plot of the total count of Manchester grid cells classified as *positive* for drusen. There were many images with high counts of cells detected as *positive* for drusen in the AD and CH group.

Figure 6.2A shows the output of the drusen detector visualised as a heatmap in an individual who had a high drusen count. Closer inspection of the image in Figure 6.2B shows that there were large amounts of bright material that is most likely vitreous material, which confounds the automatic system. As this image contained an extremely large number of false positives, the image was excluded from further analysis. This individual's fellow eye was visually inspected to assess whether the similar bright material was also present (see Figure 6.2C). Regions of high probability

were visually inspected (see Figure 6.2D) and found to contain clusters of hard and soft drusen rather than the previously observed bright vitreous material, and so the image was included in further analysis.

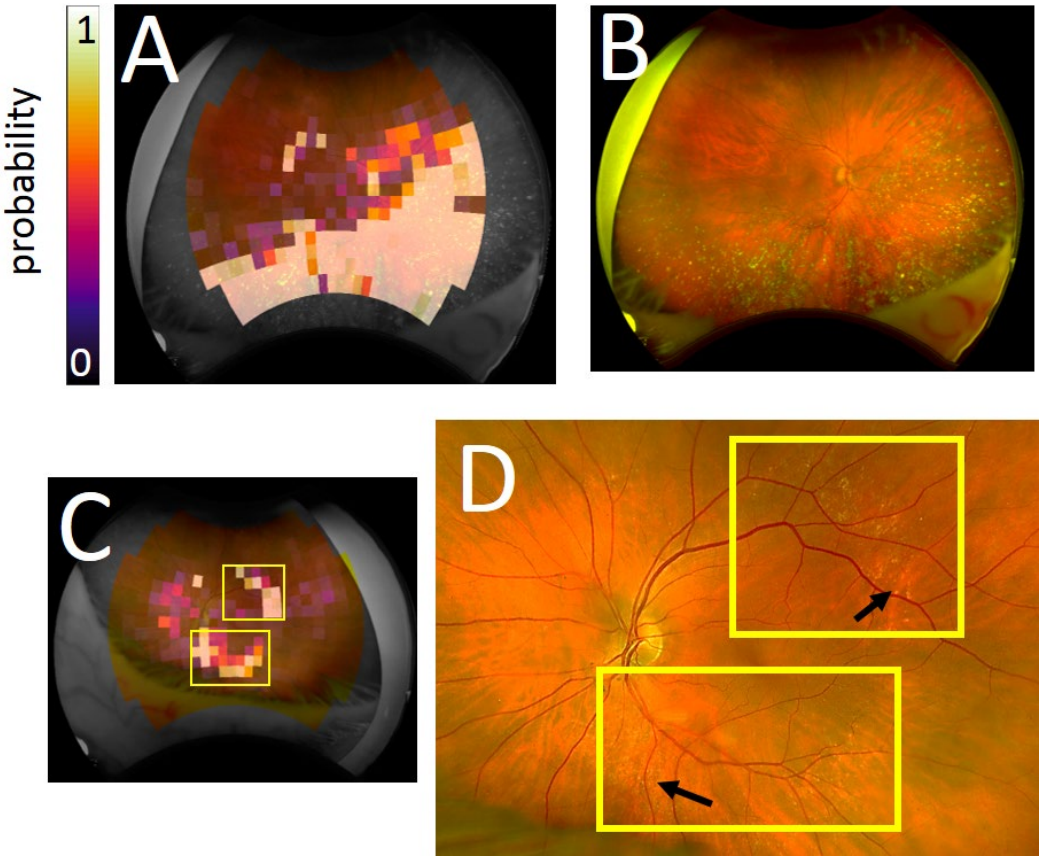

Figure 6.2 A CH individual with large amounts of vitreous material in their right eye but not the left. **A)** The output from the system was a high probability that these regions contained drusen (*white*). **B)** On closer inspection vitreous material was observed in the inferior region of the image and this has a similar appearance to drusen is present in the inferior region of the image. **C)** For the image of the fellow eye the output of the drusen detection was inspected for artefact (*yellow boxes*). **D)** The regions of high probability were visually assessed and deemed to contain clusters of hard and soft drusen (*black arrows*) and so were considered true positives.

On visual inspection of 16 images suspected of containing large numbers of false positives, they were actually from individuals who had large amounts of drusen. Figure 6.3 shows an example of this along with the image obtained from the fellow eye for comparison. There was a region of high probability near to the OD. Given that this was a source of false positives (see Section 4.3.2) the area was visually inspected and deemed to contain drusen (white arrows) but also false negatives (purple arrow). All 16 images appeared to contain low numbers of false positives and were all included in the subsequent analysis.

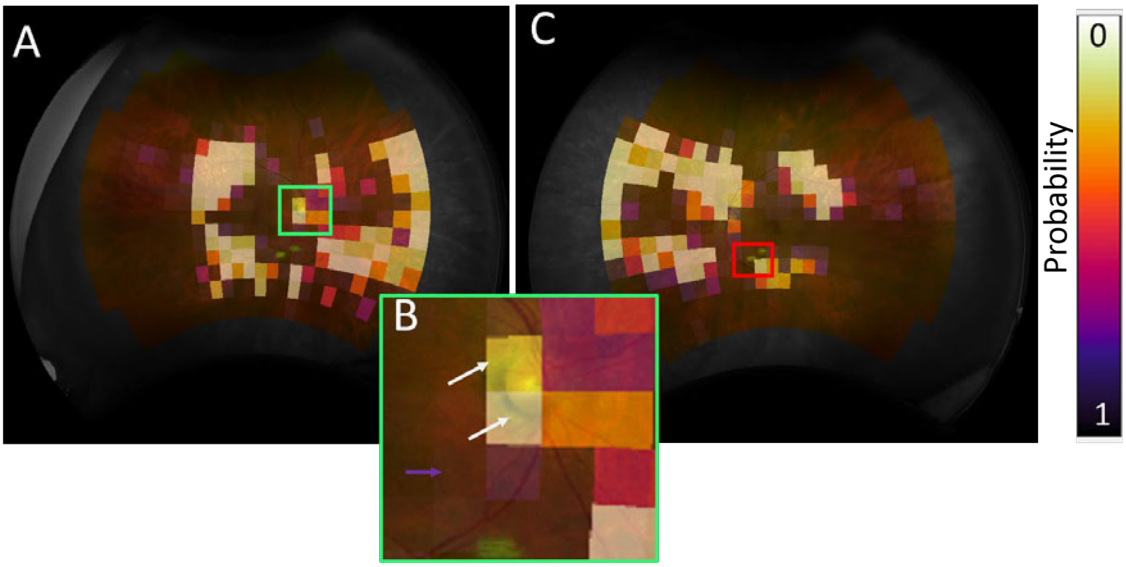

Figure 6.3 A) the right eye of the individual where a region near the OD was assessed by the automatic system to contain a high probability for drusen. **(B)** This region was visually inspected and shows true positives near the OD (white arrows) and false negatives (purple arrow). **(C)** This individual has a high abundance of drusen in the fellow eye.

Table 6.1 shows the participant demographics following removal of the one outlier (shown in Figure 6.2. There was no evidence of a significant difference between ages for the 3 groups but there was a significant difference in Mini Mental State Exam (MMSE) scores (as expected).

Feature	CH	MCI	AD	P value for Difference [Test]
Number of participants (Female/Male)	92(76/16)	41(31/10)	50(34/16)	-
(No. OS) (No. OD)	(49) (44)	(19)(22)	(23) (27)	-
Mean age (SD)	64(9)	70(8)	71 (9)	0.05[†]
MMSE	29 (1)	26(1)	20 (5)	< 0.0001[1]

Table 6.1. Participant demographics. If values were normally distributed, a 1-way ANOVA[1] was used to compare means, otherwise Kruskalwallis[†] was used. There was no evidence of significant difference in age. There was a significant difference between all groups for MMSE scores as identified using post-hoc Dunn and Sidak test for ANOVA[1]

Previous large population based surveys involving fundus photography and drusen grading have reported an association between increasing drusen and age [215][216][217]. To assess whether a similar association might be present in the dataset described above, Spearman correlation (R) test was used. To investigate the differences of total drusen counts between groups multivariate GEE modelling was used (see Section 5.2.6). The overall count as opposed to individual zones was used

for GEE analysis because some images could have no drusen detected in any one of the zones. This would result in many 0 values which can be avoided, wherever possible, by using the total count detected in the image.

6.3 Results

6.3.1 Drusen detection

Figure 6.4 shows as a percentage the total number of cells detected as positive for drusen per zone and split by group. There were a total of 635, 112 and 384 cells detected as drusen in the CH, MCI and AD groups, respectively. MCI had a higher number of cells detected as positive for drusen in the central zone (35 cells, 31%) than CH (19 cells, 3%) and AD (82 cells, 21%). CH had a higher number of drusen positive cells in the perimacular zone (289 cells, 46%) than MCI (35 cells, 31%) and AD (82 cells, 21%). AD had a higher number of cells detected as positive for drusen in the peripheral zone (220 cells, 58%) than CH (327 cells, 51%) and MCI (42 cells, 38%).

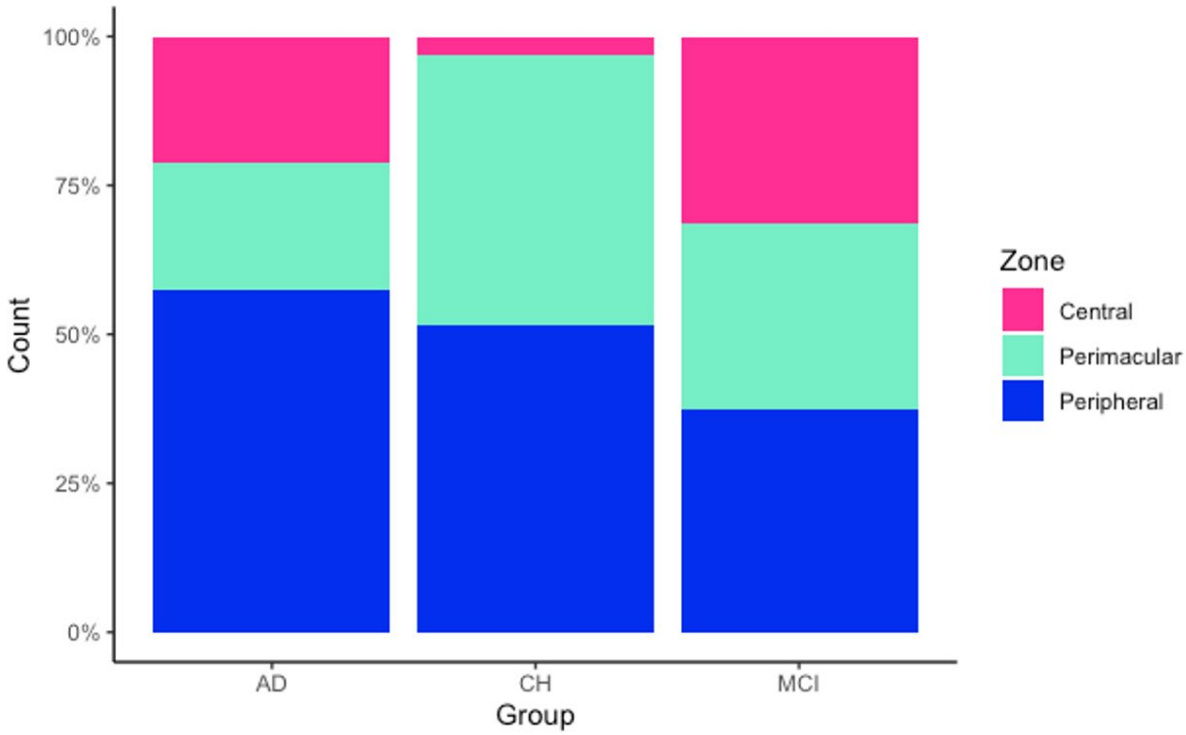

Figure 6.4 shows as a percentage the total number of cells detected as positive for drusen per zone and split by group.

6.3.2 Drusen and age

Figure 6.5 explores the relationship between age and the total number drusen positive cells in the whole dataset. I observed a statistically significant (weak) positive correlation ($R = 0.30$, $p < 0.01$). By group, there was a statistically significant (moderate) positive correlation between drusen count and age in CH ($R = 0.40$, $p < 0.01$) and a non-significant (weak) positive correlation between drusen count and age

in MCI (R = 0.30, p = 0.06) and in AD (R = 0.30, p = 0.06). The drusen detector and our dataset replicated the previously reported association of increasing drusen abundance with age, even though the network sensitivity was low and the counts were most likely under estimated.

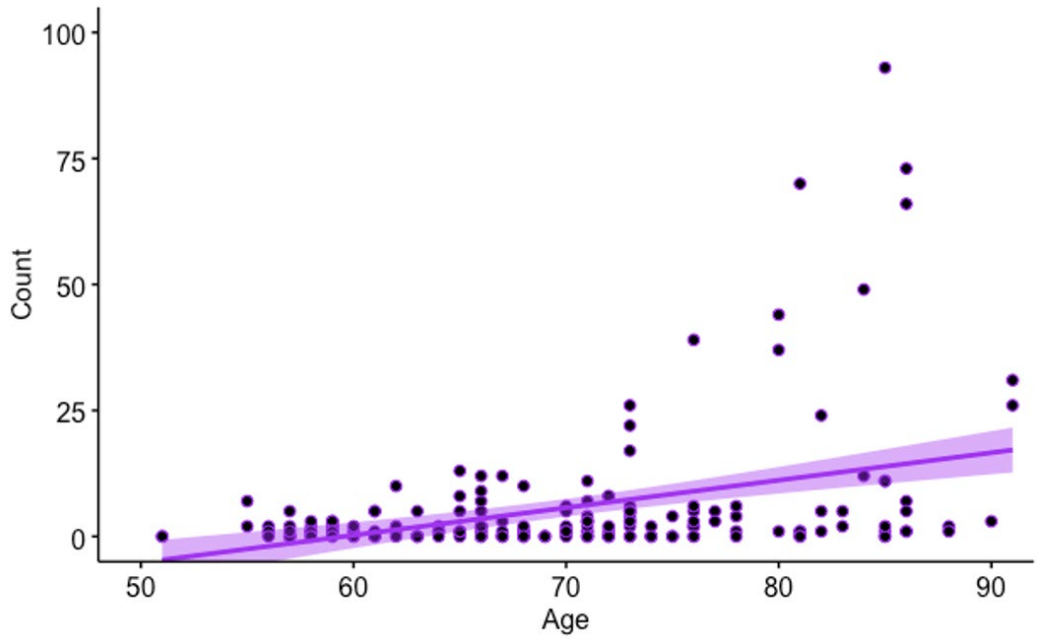

Figure 6.5 A plot of age against total drusen count for each individual, which shows a linear relationship. The purple line indicates the correlation of the two parameters and the 95% confidence intervals. There was a weak positive correlation between age and drusen counts (R = 0.30, p < 0.01). It can be seen that individuals with higher numbers of counts (>30) were older in age (> 70 years old).

6.3.3 Multivariate analysis

Table 6.2 shows the β-coefficients for the multivariate GEE analysis (corrected for age). When comparing CH to MCI there was statistically significant difference in the number of drusen positive cells (p = 0.02). When comparing CH to AD there appeared to be a difference but there was no evidence of this being statistically significant (p = 0.30). When comparing MCI to AD there was statistically significant difference in drusen counts (p = 0.003).

	β Coefficient (95% Confidence Intervals) p-value		
	CH vs MCI	CH vs AD	MCI vs AD
Total Count	7.0 (2.0 to 10.0) **p = 0.02**	3.0 (2.0 to 8.0) p = 0.30	-1.3 (0 to 2.0) **p = 0.003**

Table 6.2. Multivariate GEE β-coefficients with the 95% confidence intervals and p-values for differences between total drusen cell counts and groups, corrected for age.

6.3.4 Drusen distribution

In order to explore how drusen were distributed, across the retina for each group, an average probability was created and visualised using a heatmap, which colour codes the output of the detector (see Figure 6.5). Figure 6.5 A (left eye) and B (right eye)

show this for the CH group where it can be seen that there was a large proportion of high probability in the inferior nasal region of the perimacular and peripheral zone. Figure 6.5 C (left eye) and D (right eye) show the average probability in the MCI group where there were smaller regions of high probability cells than observed in the CH group. These look to be distributed around the perimacular and peripheral zones. Figure 6.5 E (left eye) and F (right eye) show the average probability for the AD group where there was a large high probability region in the inferonasal region (peripheral zone) of the left and right eyes that is similar in pattern to the CH group. This analysis suggests that there were less drusen in the perimacular and peripheral zones in MCI compared to CH and AD.

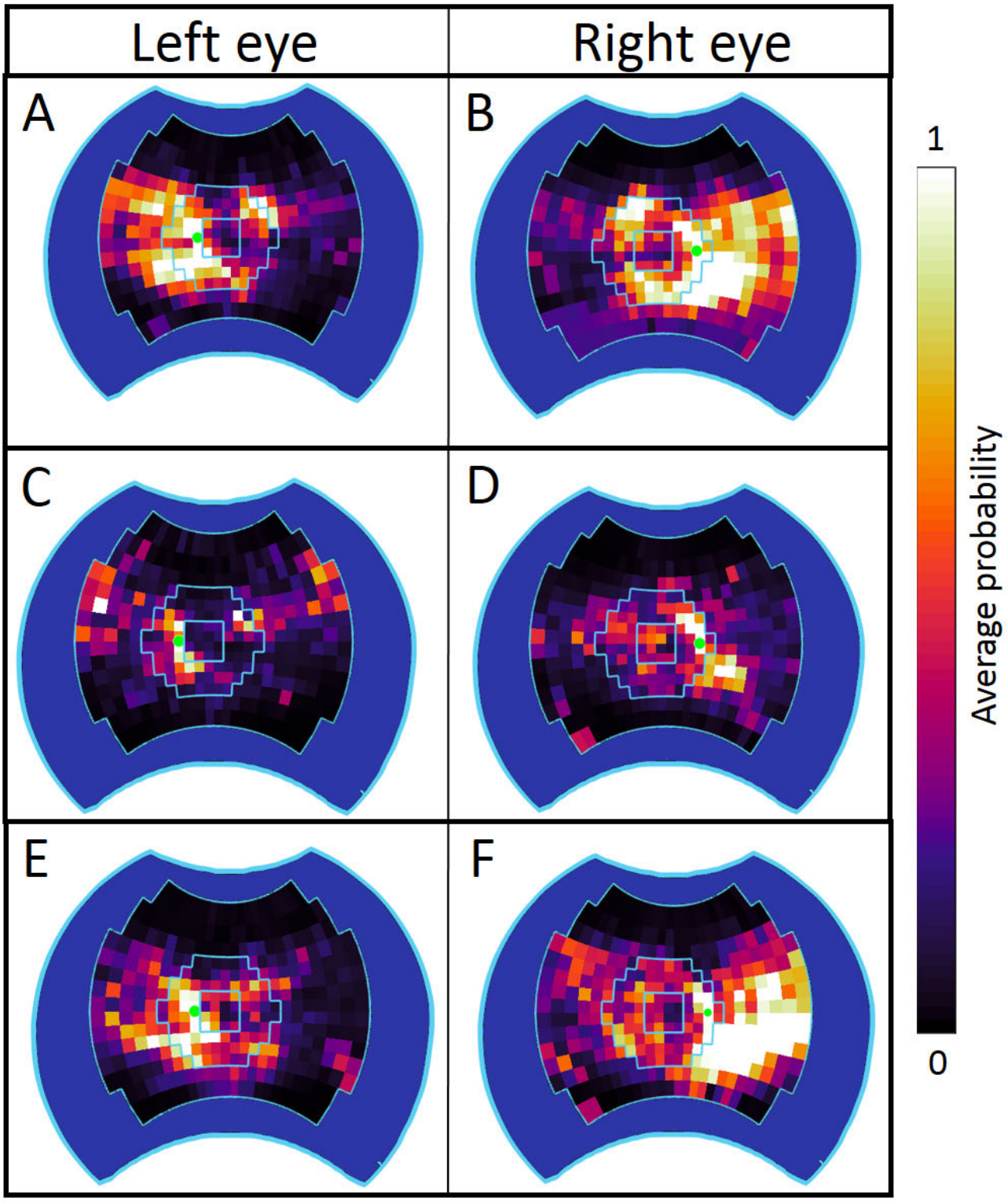
Left eye
Right eye
A
B
C
D
E
F
1
Average probability
0

Figure 6.5 Visualising the drusen cell probabilities, where higher average probabilities correspond to a whiter colour on the colour scale and lower probabilities correspond to the darker end. **(A)** and **(B)** correspond to the CH group, **(C)** and **(D)** to the MCI group, **(E)** and **(F)** to the AD group. It can be seen that there are different patterns in drusen distribution. The CH and AD groups both contained large regions of high drusen probability in the inferornasal regions whereas drusen probabilities look more dispersed in MCI eyes. The green marker indicates the OD. The blue lines delineate the zones as defined in Chapter 4.

6.4 Discussion

The aim of this chapter was to use the automatic drusen detection system to investigate associations between drusen deposition and neurodegeneration in the context of AD. Drusen deposition was associated with age in my dataset ($R = 0.40$, $p < 0.05$), which agrees with previous reports [215][216][217]. Though it should be remembered that my system has low sensitivity and so there could have been false negatives that would give an underestimation of the true drusen load in an image. It would be interesting to observe the heatmaps in longitudinal images to monitor drusen progression in people split by zones to see whether this differs between the groups. However, improvements to the system such as more training data would be required to increase the networks sensitivity (see Section 4.4). Following such

improvements, the analysis in this chapter could be conducted to highlight whether the trends observed in this thesis persists.

There was a higher abundance of drusen positive cells in the inferonasal region of the peripheral retina in CH (51%) and AD (58%) individuals and less was observed in MCI (38%). Csincsik et al [91] proposed that the thinning of the vasculature towards the periphery would decrease blood flow and could lead to decreased oxygenation, nutrient availability and suboptimal clearing of the peripheral retina. Previous studies demonstrating retinal changes in MCI and AD have shown that blood flow abnormalities precede neurodegeneration [218]. In Chapter 5 when comparing CH with AD there was significant difference between the inferonasal arteriolar width gradient. This could suggest that the high abundance of drusen detected in this region may be related to a thinner inferonasal arterioles leading to suboptimal clearance and drusen deposition.

In Chapter 5, there was no evidence of a significant difference between CH and MCI inferonasal arteriolar width gradient, but 60% of MCI individuals were hypertensive (Section 5.3.4). MCI had significantly different drusen counts when comparing to CH ($p = 0.02$) and AD ($p = 0.003$). It could be that drusen were not observed in this region for MCI because individuals were taking antihypertensives that may have improved blood flow and thus minimised drusen deposition. The data on hypertensive medication was not available for this thesis but would be important for future studies investigating vascular measures and drusen load.

There was a statistically significant difference between the number of drusen positive cells when comparing CH with AD ($p = 0.02$) and CH with MCI ($p = 0.003$). High probabilities were located in the perimacular and peripheral zones. A metanalysis of RNFL thickness measured using OCT in AD and MCI compared to healthy controls found that there was an association between RNFL thinning of the superior and inferior quadrants [219]. Additionally, there is a higher concentration of ganglion cells in the nasal quadrant than the temporal quadrant of the perimacular zone [75]. RGC degeneration has been associated with AD [101][220] as well as amyloid deposition in the retina [52][93]. This could explain the higher rates of drusen detection in this region as a higher concentration of degenerating cells may lead to increased drusen accumulation. More data and improvements to the drusen detector (as outlined above) would be required to determine whether this pattern persists.

Csincsik et al [91] found that there was a higher abundance of peripheral hard drusen in AD. This raises the question as to whether drusen type helps to distinguish accumulation associated with age or with ganglion cell degeneration and amyloid deposition that may precede cognitive decline. Post mortem histology analysis of drusen deposits composition may shed light onto the differences between age-related drusen and drusen associated with cognitive decline. It would be interesting to curate datasets featuring UWF images and corresponding post mortem histology to assess for correlations between the appearance of drusen *in vivo* and drusen

molecular composition. This may aide in the identification of imaging characteristics that define drusen containing amyloid and further identify AD retinal biomarkers.

6.5 Conclusions

This chapter explored drusen load in CH, MCI and AD individuals using an automatic deep learning-based system. Although the sensitivity of the system was low and the drusen counts were most likely underestimated, the output of the system was able to replicate a previously reported association of drusen and age. There was significantly less drusen detected in MCI compared to AD and CH. AD had a higher proportion of drusen counts in the periphery compared to MCI and CH that replicates previous work using manual annotations of drusen in UWF images. Improvements to the system, such as more training data, would be required to assess whether these trends persist in larger datasets.

Chapter 7 Conclusions and Future Work

7.1 Conclusions

Inspired by previous work in fundus photography, a novel computerised technique to automatically detect drusen in UWF images was developed in this thesis. Retinal vascular measurements and drusen counts were obtained and assessed in a patient cohort that featured AD and MCI as well as CH individuals - to identify potential biomarkers and retinal manifestations of disease.

The following were the major aims and advancements achieved by this thesis:

1) Aim 1 was to investigate methods for automatically detecting and analysing drusen in UWF images. This was achieved by investigating previous work in detecting drusen in fundus camera images (primarily in the field of AMD). This highlighted the direction of research moving towards deep learning and transfer learning, which subsequently inspired my development of drusen detection for UWF images
2) Aim 2 was to cultivate an automatic drusen detection algorithm. This was achieved through development of a supervised deep convolutional neural network to classify patches within a UWF image as *positive* or *negative* for drusen. Performances of 0.55-0.59, 0.62- 0.65 and 0.65-0.66 were achieved in

the central, perimacular and peripheral regions of the retina, respectively. This provided a novel method to identify regions that had a high probability of containing drusen. This has utility for quantifying drusen (by counting the number of *positive* patches) and for monitoring drusen progression (visualising the output via a heatmap). The system had a low sensitivity (17.7-21.8%) which means there would be a high number of false negatives, therefore utilising the system for quantifying drusen in a new dataset would result in an underestimation of the true amount of drusen present in the images. The nature of many of these false negatives were instances of small discrete drusen that might only be 1 pixel in size or dust artefact on the mirror of the SLO. In both cases, a human grader might have more success where the system failed. However, there was only a moderate agreement between two observers grading images ($\kappa = 0.53$, DSC = 0.38), which highlights the difficult nature of manually annotating drusen in UWF images. This also created a challenge to validating the system in the presence of observer disagreement. These limitations could be overcome by training the system on more data and additional observers to create a more consistent reference standard.

3) Aim 3 was to measure and analyse the retinal vasculature in AD to investigate retinal manifestations and discover potential biomarkers. Multivariate GEE analysis was used to compare between groups of CH, MCI (considered in this thesis as prodromal AD) and AD, whilst accounting for correlations between

eyes. I found a significant difference between arteriolar inferonasal width gradients of CH and AD. There were inconstancies between fractal dimension measurements in the repeat analysis, where this thesis identified limitations in marking non-standardised ROI in the UWF image set. My work did highlight a robust statistical analysis method for UWF retinal vasculature measurements as well as issues to overcome in future studies.

4) Aim 4 was to apply the drusen detector to the dataset consisting of CH, MCI and AD in order to quantify drusen in these images and identify differences between the groups. This was conducted as an exploratory analysis due to the limitations of the proposed system, i.e. low sensitivity. The exploratory analysis produced promising results as the output of the system was successful in replicating a known association between drusen and age as well as uncovering novel differences between groups. Using GEE, there was a significant difference in drusen count when comparing MCI to CH ($p = 0.02$) and MCI to AD ($p = 0.03$), but no evidence of a significant difference when comparing AD to CH ($p = 0.30$). Visual inspection of the output of the automatic system via a heatmap showed that there were higher probabilities of drusen in the inferonasal region of CH and AD, which corresponded with the difference observed in width gradients.

7.2 Future work

The human observers used a Manchester grid to grade cells for drusen, and the proposed system was validated in a way to correspond to this grading protocol. This limited the FOV due to the curvature of the grid towards the periphery that could not be resized for input into the neural network, and subsequently led to grid cells which could not be classified for drusen. Additionally, drusen grading is highly subjective as there was not a clear enough protocol for the human observers to follow, particularly when encountering a Manchester grid cell with drusen at the edge of the cell boundary. One observer may grade for drusen and the other may not. This created a level of disagreement between the observers and led to challenges when validating the neural networks performance. Future reference standards should therefore include a grid that extends to the periphery perhaps by using a system that does not account for the retinal curvature in this region (i.e. a square cell) and precise grading protocols as to how to grade drusen at the edge of a cell. This would improve the quality of the ground truth which is necessary for both training and validating an automatic system.

Neural network performance could be improved with more training data. This data could consist of patches graded for different subtypes of drusen and dust (to train the network to identify false positives). As drusen size is used in many AMD studies and may be important for AD studies, this could be obtained by segmenting drusen

in the UWF image. Semantic segmentation, whereby each pixel is given a class, could be used provided that the training data contains large amounts of hand drawn segmentations by multiple observers. Individual hand drawn boundaries of drusen would be challenging in UWF images because of the resolution of the imaging technique as drusen might be only a few pixels in size. Obtaining large amounts of pixel level annotations would be challenging, expensive and time consuming where multiple observers would be needed to obtain a reliable reference standard. Subsequently, ensembles of neural networks, whereby stacks of networks that perform different tasks in order to create a combined output, could be developed. For example, one network could classify the image as containing drusen or not, where images high risk for drusen are passed to a second network. The second network could classify patches for drusen where patches with a high probability for drusen are passed to a third network. The third network could perform semantic segmentation to delineate drusen boundaries for drusen quantification. This coupled with OCT images that could be used to segment drusen volumes, may provide further insight into the type of drusen present. Such multimodal imaging ensemble learning methods would require large amounts of data, meticulous hand drawn segmentations and gradings from many human observers and will be challenging to obtain.

There was inconsistency between repeat measures of fractal dimensions that resulted from the varying nature of the ROI used in performing the calculation. To overcome this, a predefined ROI could be used on each image, for example a 'Zone D' that is an annulus that extends further from Zone C. However, in a UWF image obtaining a circular area is challenging because of the eyelashes and eyelids that sometimes feature in the image. An ellipse shape is often more appropriate. This ellipse could be obtained by taking the average manually drawn ROI over the whole dataset (that is positioned according to the OD and fovea) so that the fractal dimension is calculated from the same region in each image. The challenge with such an approach is that not all of the vasculature might be measured in some images. Another way to address the challenge might be to develop a new metric of branching complexity. The data structure of the vascular segmentation is a graph, whereby the nodes are the branching points, the edges are the vessels to the node and there are multiple vessel graphs in one image that corresponds to the arteriolar or venular components. This data structure is synonymous to a computer program where the nodes can be a conditional statement, the edges are the routes from the statement and there can be a number of graphs that corresponds to the number of subroutines or classes. Software developers use this to measure the complexity of the graph (i.e. the complexity of the software), called cyclometric complexity [221]. This returns a value between 1 and 10, where 10 is most complex. Potentially the cyclometric complexity could be used to give a quantitative measure of the complexity of the

vasculature. The challenge here is that the performance of the segmentation would have an influence on the number of edges and nodes and therefore impact the value of the complexity – i.e. a low cyclometric complexity could be returned because not all of the vasculature could be segmented in a poor quality of the image. However, the question would be whether this method is more repeatable than a fractal dimension. Comparisons of fractal dimension and cyclometric complexity measures would be required to validate a new metric of vascular branching complexity.

Bibliography

[1] A. Alzheimer, R. A. Stelzmann, H. N. Schnitzlein, and F. R. Murtagh, "An English translation of Alzheimer's 1907 paper, 'Uber eine eigenartige Erkankung der Hirnrinde'.," *Clin. Anat.*, vol. 8, no. 6, pp. 429–431, 1995.

[2] H. J. Moller and M. B. Graeber, "The case described by Alois Alzheimer in 1911. Historical and conceptual perspectives based on the clinical record and neurohistological sections.," *Eur. Arch. Psychiatry Clin. Neurosci.*, vol. 248, no. 3, pp. 111–122, 1998.

[3] K. Beyreuther and C. L. Masters, "Amyloid precursor protein (APP) and beta A4 amyloid in the etiology of Alzheimer's disease: precursor-product relationships in the derangement of neuronal function.," *Brain Pathol.*, vol. 1, no. 4, pp. 241–251, Jul. 1991.

[4] J. Hardy and D. Allsop, "Amyloid deposition as the central event in the aetiology of Alzheimer's disease.," *Trends Pharmacol. Sci.*, vol. 12, no. 10, pp. 383–388, Oct. 1991.

[5] M. Jin, N. Shepardson, T. Yang, G. Chen, D. Walsh, and D. J. Selkoe, "Soluble amyloid beta-protein dimers isolated from Alzheimer cortex directly induce Tau hyperphosphorylation and neuritic degeneration.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 14, pp. 5819–5824, Apr. 2011.

[6] A. Hofman *et al.*, "Atherosclerosis, apolipoprotein E, and prevalence of

dementia and Alzheimer's disease in the Rotterdam Study," *Lancet*, vol. 349, no. 9046, pp. 151–154, 1997.

[7] T. J. MacGillivray, E. Trucco, J. R. Cameron, and B. Dhillon, "Retinal imaging as a source of biomarkers for diagnosis, characterisation and prognosis of chronic illness or long-term conditions," *Br J Radiol*, 2014.

[8] J. Savige, S. Ratnaike, and D. Colville, "Retinal Abnormalities Characteristic of Inherited Renal Disease," *J. Am. Soc. Nephrol.*, vol. 22, no. 8, pp. 1403 LP – 1415, Aug. 2011.

[9] F. Forrester, John, McMenamin, Paul, Pearlman, Eric, Andrew, Dick, Roberts, *The eye. Basic Sciences in Practice*, 4th ed. 2006.

[10] A. A. of Opthalmology, "Anatomy of the eye: image," 2017. [Online]. Available: https://www.aao.org/eye-health/anatomy/fovea. [Accessed: 25-Feb-2019].

[11] R. Brookmeyer, E. Johnson, K. Ziegler-Graham, and M. Arrighi, "Forecasting the global burden of Alzheimers disease," *Alzheimers Dement.*, vol. 3, no. 3, pp. 186–191, 2007.

[12] F. Lewis, S. K. Shauffer, J. Sussex, P. O'Neill, and L. Cockcroft, "The Trajectory of Dementian in the UK - Making a Difference," Office of Health Economics, 2014.

[13] J. C. Morris, "Early-stage and preclinical Alzheimer disease.," *Alzheimer Dis. Assoc. Disord.*, vol. 19, no. 3, pp. 163–165, 2005.

[14] B. Dubois *et al.*, "Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS–ADRDA criteria," *Lancet Neurol.*, vol. 6, no. 8, pp. 734–746, 2007.

[15] C. R. Jack Jr. *et al.*, "NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease," *Alzheimer's Dement. J. Alzheimer's Assoc.*, vol. 14, no. 4, pp. 535–562, Apr. 2018.

[16] R. J. O'Brien and P. C. Wong, "Amyloid precursor protein processing and Alzheimer's disease," *Annu. Rev. Neurosci.*, vol. 34, pp. 185–204, 2011.

[17] U. C. Muller and H. Zheng, "Physiological functions of APP family proteins.," *Cold Spring Harb. Perspect. Med.*, vol. 2, no. 2, p. a006288, Feb. 2012.

[18] S. S. Sisodia, "Beta-amyloid precursor protein cleavage by a membrane-

bound protease.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 89, no. 13, pp. 6075–6079, Jul. 1992.

[19] R. Vassar *et al.*, "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE.," *Science*, vol. 286, no. 5440, pp. 735–741, Oct. 1999.

[20] E. H. Corder *et al.*, "Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families.," *Science*, vol. 261, no. 5123, pp. 921–923, Aug. 1993.

[21] L. Chávez-Gutiérrez *et al.*, "The mechanism of γ-Secretase dysfunction in familial Alzheimer disease," *EMBO J.*, vol. 31, no. 10, pp. 2261–2274, May 2012.

[22] D. Campion *et al.*, "Early-onset autosomal dominant Alzheimer disease: prevalence, genetic heterogeneity, and mutation spectrum.," *Am. J. Hum. Genet.*, vol. 65, no. 3, pp. 664–670, Sep. 1999.

[23] G. M. Shankar *et al.*, "Amyloid-β protein dimers isolated directly from Alzheimer's brains impair synaptic plasticity and memory," *Nat. Med.*, vol. 14, p. 837, Jun. 2008.

[24] J. T. Jarrett, E. P. Berger, and P. T. Lansbury, "The carboxy terminus of the .beta. amyloid protein is critical for the seeding of amyloid formation: Implications for the pathogenesis of Alzheimer's disease," *Biochemistry*, vol. 32, no. 18, pp. 4693–4697, May 1993.

[25] R. A. Sperling *et al.*, "Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's Dement. J. Alzheimer's Assoc.*, vol. 7, no. 3, pp. 280–292, May 2011.

[26] M. Goedert, M. G. Spillantini, M. C. Potier, J. Ulrich, and R. A. Crowther, "Cloning and sequencing of the cDNA encoding an isoform of microtubule-associated protein tau containing four tandem repeats: differential expression of tau protein mRNAs in human brain," *EMBO J.*, vol. 8, no. 2, pp. 393–399, Feb. 1989.

[27] I. Grundke-Iqbal, K. Iqbal, Y. C. Tung, M. Quinlan, H. M. Wisniewski, and L. I. Binder, "Abnormal phosphorylation of the microtubule-associated protein tau (tau) in Alzheimer cytoskeletal pathology.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 83, no. 13, pp. 4913–4917, Jul. 1986.

[28] M. Bobinski *et al.*, “The histological validation of post mortem magnetic resonance imaging-determined hippocampal volume in Alzheimer’s disease.,” *Neuroscience*, vol. 95, no. 3, pp. 721–725, 2000.

[29] K. A. Jobst *et al.*, “Rapidly progressing atrophy of medial temporal lobe in Alzheimer’s disease.,” *Lancet (London, England)*, vol. 343, no. 8901, pp. 829–830, Apr. 1994.

[30] N. Mattsson, U. Andreasson, H. Zetterberg, and K. Blennow, “Association of Plasma Neurofilament Light With Neurodegeneration in Patients With Alzheimer Disease.,” *JAMA Neurol.*, vol. 74, no. 5, pp. 557–566, May 2017.

[31] M. F. Folstein, S. E. Folstein, and P. R. McHugh, “‘Mini-mental state’. A practical method for grading the cognitive state of patients for the clinician.,” *J. Psychiatr. Res.*, vol. 12, no. 3, pp. 189–198, Nov. 1975.

[32] Z. S. Nasreddine *et al.*, “The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment.,” *J. Am. Geriatr. Soc.*, vol. 53, no. 4, pp. 695–699, Apr. 2005.

[33] B. Schmand, C. Jonker, C. Hooijer, and J. Lindeboom, “Subjective memory complaints may announce dementia,” *Neurology*, vol. 46, no. 1, pp. 121 LP – 125, Jan. 1996.

[34] M. S. Mega, J. L. Cummings, T. Fiorello, and J. Gornbein, “The spectrum of behavioral changes in Alzheimer’s disease,” *Neurology*, vol. 46, no. 1, pp. 130 LP – 135, Jan. 1996.

[35] R. J. Bateman *et al.*, “Clinical and Biomarker Changes in Dominantly Inherited Alzheimer’s Disease,” *N. Engl. J. Med.*, vol. 367, no. 9, pp. 795–804, Jul. 2012.

[36] J. C. Morris *et al.*, “Mild Cognitive Impairment Represents Early-Stage Alzheimer Disease,” *JAMA Neurol.*, vol. 58, no. 3, pp. 397–405, Mar. 2001.

[37] J. C. Morris, “Revised criteria for mild cognitive impairment may compromise the diagnosis of Alzheimer disease dementia,” *Arch. Neurol.*, vol. 69, no. 6, pp. 700–708, Jun. 2012.

[38] M. J. de Leon *et al.*, “Prediction of cognitive decline in normal elderly subjects with 2-fluoro-2-deoxy-glucose/positron-emission tomography (FDG/PET),” *Proc. Natl. Acad. Sci.*, vol. 98, no. 19, pp. 10966 LP – 10971, Sep. 2001.

[39] B. A. Gordon *et al.*, “Spatial patterns of neuroimaging biomarker change in individuals from families with autosomal dominant Alzheimer’s disease: a

longitudinal study," *Lancet Neurol.*, vol. 17, no. 3, pp. 241–250, Mar. 2018.

[40] C. R. Jack Jr *et al.*, "Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade," *Lancet. Neurol.*, vol. 9, no. 1, pp. 119–128, Jan. 2010.

[41] R. S. Doody *et al.*, "Phase 3 trials of solanezumab for mild-to-moderate Alzheimer's disease.," *N. Engl. J. Med.*, vol. 370, no. 4, pp. 311–321, Jan. 2014.

[42] S. Salloway *et al.*, "Two phase 3 trials of bapineuzumab in mild-to-moderate Alzheimer's disease.," *N. Engl. J. Med.*, vol. 370, no. 4, pp. 322–333, Jan. 2014.

[43] E. R. Siemers *et al.*, "Phase 3 solanezumab trials: Secondary outcomes in mild Alzheimer's disease patients.," *Alzheimers. Dement.*, vol. 12, no. 2, pp. 110–120, Feb. 2016.

[44] F. Chiti and C. M. Dobson, "Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade," *Annu. Rev. Biochem.*, vol. 86, no. 1, pp. 27–68, Jun. 2017.

[45] A. Grimaldi *et al.*, "Inflammation, neurodegeneration and protein aggregation in the retina as ocular biomarkers for Alzheimer's disease in the 3xTg-AD mouse model," *Cell Death Dis.*, vol. 9, no. 6, p. 685, 2018.

[46] P. J. Morin *et al.*, "Amyloid precursor protein is synthesized by retinal ganglion cells, rapidly transported to the optic nerve plasma membrane and nerve terminals, and metabolized.," *J. Neurochem.*, vol. 61, no. 2, pp. 464–473, Aug. 1993.

[47] T. Y. Wong *et al.*, "Cerebral white matter lesions, retinopathy, and incident clinical stroke.," *JAMA*, vol. 288, no. 1, pp. 67–74, Jul. 2002.

[48] N. Cheung *et al.*, "Retinal microvascular abnormalities and subclinical magnetic resonance imaging brain infarct: a prospective study.," *Brain*, vol. 133, no. Pt 7, pp. 1987–1993, Jul. 2010.

[49] N. K. Archibald, M. P. Clarke, U. P. Mosimann, and D. J. Burn, "The retina in Parkinson's disease.," *Brain*, vol. 132, no. Pt 5, pp. 1128–1145, May 2009.

[50] M. M. Moschos *et al.*, "Morphologic changes and functional retinal impairment in patients with Parkinson disease without visual loss.," *Eur. J. Ophthalmol.*, vol. 21, no. 1, pp. 24–29, 2011.

[51] M. Soderstrom, “Optic neuritis and multiple sclerosis.,” *Acta Ophthalmol. Scand.*, vol. 79, no. 3, pp. 223–227, Jun. 2001.

[52] M. Koronyo-Hamaoui, Y. Koronyo, and A. V Liubimov, “Identification of amyloid plaques in retinas for Alzheimer’s patients and noninvasive in vivo optical imaging of retinal plaques in a mouse model,” *Neuroimage*, vol. 54, pp. 204–217, 2011.

[53] L. Gasparini *et al.*, “Tau inclusions in retinal ganglion cells of human P301S tau transgenic mice: effects on axonal viability.,” *Neurobiol. Aging*, vol. 32, no. 3, pp. 419–433, Mar. 2011.

[54] K. U. Loffler, D. P. Edward, and M. O. Tso, “Immunoreactivity against tau, amyloid precursor protein, and beta-amyloid in the human retina,” *Invest Opthalmol Vis Sci*, vol. 36, no. 1, pp. 24–31, 1995.

[55] T. Dentchev, A. H. Milam, V. M. Lee, J. Q. Trojanowski, and J. L. Dunaief, “Amyloid-beta is found in drusen from some age-related macular degeneration retinas, but not in drusen from normal retinas.,” *Mol Vis*, vol. 14, no. 9, pp. 184–190, 2003.

[56] T. Yoshida, K. Ohno-Matsui, and S. J. Ichinose, “The potential role of amyloid beta in the pathogenesis of age-related macular degeneration,” *Clin. Investig. (Lond).*, vol. 115, no. 10, pp. 2763–2800, 2005.

[57] S. W. Park *et al.*, “Intracellular amyloid beta alters the tight junction of retinal pigment epithelium in 5XFAD mice.,” *Neurobiol. Aging*, vol. 35, no. 9, pp. 2013–2020, Sep. 2014.

[58] V. Dinet *et al.*, “Distinct effects of inflammation on gliosis, osmohomeostasis, and vascular integrity during amyloid beta-induced retinal degeneration.,” *Aging Cell*, vol. 11, no. 4, pp. 683–693, Aug. 2012.

[59] W.-L. Ho, Y. Leung, A. W.-T. Tsang, K.-F. So, K. Chiu, and R. C.-C. Chang, “Review: tauopathy in the retina and optic nerve: does it shadow pathological changes in the brain?,” *Mol. Vis.*, vol. 18, pp. 2700–2710, 2012.

[60] T. R. 3rd Hedges, R. Perez Galves, D. Speigelman, N. R. Barbas, E. Peli, and C. J. Yardley, “Retinal nerve fiber layer abnormalities in Alzheimer’s disease.,” *Acta Ophthalmol. Scand.*, vol. 74, no. 3, pp. 271–275, Jun. 1996.

[61] C. Paquet, M. Boissonnot, F. Roger, P. Dighiero, R. Gil, and J. Hugon, “Abnormal retinal thickness in patients with mild cognitive impairment and Alzheimer’s disease.,” *Neurosci. Lett.*, vol. 420, no. 2, pp. 97–99, Jun. 2007.

[62] M. Parnell, L. Guo, M. Abdi, and M. F. Cordeiro, “Ocular manifestations of Alzheimer’s disease in animal models.,” *Int. J. Alzheimers. Dis.*, vol. 2012, p. 786494, 2012.

[63] E. S. Garcia-Martin *et al.*, “Macular thickness as a potential biomarker of mild Alzheimer’s disease.,” *Ophthalmology*, vol. 121, no. 5, pp. 1149-1151.e3, May 2014.

[64] D. R. Hinton, A. A. Sadun, J. C. Blanks, and C. A. Miller, “Optic-nerve degeneration in Alzheimer’s disease,” *N Engl J Med*, vol. 315, no. 8, pp. 485–487, 1986.

[65] Y. Koronyo, B. C. Salumbides, K. L. Black, and M. Koronyo-Hamaoui, “Alzheimer’s disease in the retina: imaging retinal abeta plaques for early diagnosis and therapy assessment.,” *Neurodegener. Dis.*, vol. 10, no. 1–4, pp. 285–293, 2012.

[66] S. Selvam, T. Kumar, and M. Fruttiger, “Retinal vasculature development in health and disease,” *Prog. Retin. Eye Res.*, vol. 63, pp. 1–19, 2018.

[67] N. Patton *et al.*, “Retinal image analysis: Concepts, applications and potential,” *Prog. Retin. Eye Res.*, vol. 25, no. 1, pp. 99–127, 2006.

[68] B. V Zlokovic, “Neurovascular pathways to neurodegeneration in Alzheimer’s disease and other disorders.,” *Nat. Rev. Neurosci.*, vol. 12, no. 12, pp. 723–738, Nov. 2011.

[69] S. M. Golzan *et al.*, “Retinal vascular and structural changes are associated with amyloid burden in the elderly: ophthalmic biomarkers of preclinical Alzheimer’s disease,” *Alzheimers. Res. Ther.*, vol. 9, no. 1, p. 13, Mar. 2017.

[70] A. Sommer, N. R. Miller, I. Pollack, and E. Maumenee, “The Nerve Fobre Layer in the diagnosis of Glaucoma,” *Clin. Sci.*, vol. 95, pp. 2149–2156, 1977.

[71] B. P. Buckingham *et al.*, “Progressive ganglion cell degeneration precedes neuronal loss in a mouse model of glaucoma.,” *J. Neurosci.*, vol. 28, no. 11, pp. 2735–2744, Mar. 2008.

[72] P. Mahendradas, A. Khanna, A. Kawali, and R. Shetty, “Quantification of inflammation in inflammatory eye diseases,” *Internet J. Rheumatol. Clin. Immunol.*, vol. 2(S1), pp. 1–10, Dec. 2014.

[73] A. Csutak *et al.*, “Agreement between image grading of conventional (45°) and ultra wide-angle (200°) digital images in the macula in the Reykjavik eye

study," *Eye (Lond).*, vol. 24, pp. 1568–1575, Oct. 2010.

[74] M. Sagong, J. van Hemert, L. C. Olmos de Koo, C. Barnett, and S. R. Sadda, "Assessment of Accuracy and Precision of Quantification of Ultra-Widefield Images," *Ophthalmology*, vol. 122, no. 4, pp. 864–866, Apr. 2015.

[75] N. Quinn *et al.*, "The clinical relevance of visualising the peripheral retina.," *Prog. Retin. Eye Res.*, vol. 68, pp. 83–109, Jan. 2019.

[76] J. Peng, Q. Zhang, H.-Y. Jin, W.-Y. Lu, and P.-Q. Zhao, "Ultra-wide field imaging system and traditional retinal examinations for screening fundus changes after cataract surgery.," *Int. J. Ophthalmol.*, vol. 9, no. 9, pp. 1299–1303, 2016.

[77] K. Ghasemi Falavarjani, I. Tsui, and S. R. Sadda, "Ultra-wide-field imaging in diabetic retinopathy.," *Vision Res.*, vol. 139, pp. 187–190, Oct. 2017.

[78] RNIB, "Key information and statistics on sight loss in the UK," *Key Information and Statistics*, 2011. .

[79] W. L. Wong, S. Xinyi, X. Li, and C. M. Cheung, "Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis," *Lancet Glob. Heal.*, vol. 2, no. 2, pp. 106–116, 2014.

[80] R. Klein, B. E. Klein, M. D. Knudston, and S. M. Meuer, "Fifteen-year cumulative incidene of age-related macular degeneration: the Beaver Dam Eye Study.," *Opthalmology*, vol. 114, no. 2, pp. 253–262, 2007.

[81] J. H. Lee, T. Lee, S. C. Lee, and C. S. Lee, "Disappearance of soft drusen and subsequent development of choroidal neovascularization following macular hole surgery: a case report," *BMC Ophthalmol.*, vol. 15, p. 43, Apr. 2015.

[82] N. M. Bressler *et al.*, "Five-year incidence and disappearance of drusen and retinal pigment epithelial abnormalities. Waterman study.," *Arch. Ophthalmol. (Chicago, Ill. 1960)*, vol. 113, no. 3, pp. 301–308, Mar. 1995.

[83] M. Sebag, E. Peli, and M. Lahav, "Image analysis of changes in drusen area," *Acta Ophthalmol.*, vol. 69, no. 5, pp. 603–610, Oct. 1991.

[84] T. A. Foundation, "Retinal OCT Fluid Challenge (RETOUCH)." .

[85] R. Klein, M. D. Davis, and Y. L. Magli, "The Wisconson age-related maculopathy grading system," *Opthalmology*, vol. 98, no. 7, pp. 1128–1134,

1991.

[86] A.-R. E. D. S. (AREDS) R. Group, “A randomized , placebo-controlled, clinical trial of high-dose supplementation and vitamins C and E, beta carotene, and zinc for age-related macular degeneration and vision loss: AREDS report no 8,” *Arch Opthalmology*, vol. 119, pp. 1417–1436, 2001.

[87] R. F. Spaide and C. A. Curcio, “Drusen characterization with multimodal imaging.,” *Retina*, vol. 30, no. 9, pp. 1441–1454, Oct. 2010.

[88] C.-M. Li, M. E. Clark, M. Rudolf, and C. A. Curcio, “Distribution and composition of esterified and unesterified cholesterol in extra-macular drusen.,” *Exp. Eye Res.*, vol. 85, no. 2, pp. 192–201, Aug. 2007.

[89] M. Rudolf, M. E. Clark, M. F. Chimento, C.-M. Li, N. E. Medeiros, and C. A. Curcio, “Prevalence and morphology of druse types in the macula and periphery of eyes with age-related maculopathy,” *Invest. Ophthalmol. Vis. Sci.*, vol. 49, no. 3, pp. 1200–1209, Mar. 2008.

[90] K. Ukalovic *et al.*, “Drusen in the Peripheral Retina of the Alzheimer’s Eye.,” *Curr. Alzheimer Res.*, vol. 15, no. 8, pp. 743–750, 2018.

[91] L. Csincsik, T. J. MacGillivray, E. Flynn, and G. Papanastasiou, “Peripheral Retinal Imaging Biomarkers for Alzheimer’s Disease: A Pilot Study.,” *Opathalmic Res*, vol. 59, no. 4, pp. 182–192, 2018.

[92] K. Kaarniranta, A. Salminen, A. Haapasalo, H. Soininen, and M. Hiltunen, “Age-Related Macular Degeneration (AMD): Alzheimer’s Disease in the Eye?,” *J. Alzheimers. Dis.*, vol. 24, pp. 615–631, Feb. 2011.

[93] K. Ohno-Matsui, “Parallel findings in age-related macular degeneration and Alzheimer’s disease.,” *Prog. Retin. Eye Res.*, vol. 30, no. 4, pp. 217–238, Jul. 2011.

[94] D. H. Anderson, K. C. Talaga, A. J. Rivest, E. Barron, G. S. Hageman, and L. V Johnson, “Characterization of beta amyloid assemblies in drusen: the deposits associated with aging and age-related macular degeneration.,” *Exp. Eye Res.*, vol. 78, no. 2, pp. 243–256, Feb. 2004.

[95] T. C. Saido and N. Iwata, “Metabolism of amyloid beta peptide and pathogenesis of Alzheimer’s disease. Towards presymptomatic diagnosis, prevention and therapy.,” *Neurosci. Res.*, vol. 54, no. 4, pp. 235–253, Apr. 2006.

[96] Y. Koronyo *et al.*, “Retinal amyloid pathology and proof-of-concept imaging trial in Alzheimer’s disease,” *JCI insight*, vol. 2, no. 16, p. e93621, Aug. 2017.

[97] E. Trucco, A. Ruggeri, T. Karnowski, and L. Giancardo, “Validating Retinal Fundus Image Analysis Algorithms: Issues and a Proposal,” *IOVS*, vol. 54, no. 5, pp. 3546–3559, 2013.

[98] G. Coppola *et al.*, “Optical Coherence Tomography in Alzheimer’s Disease: A Meta-Analysis.,” *PLoS One*, vol. 10, no. 8, p. e0134750, 2015.

[99] K. L. Thomson, M. J. Yeo, B. Waddell, and J. R. Cameron, “A systematic review and meta-analysis of retinal nerve fiber layer change in dementia, using optical coherence tomography,” *Alzheimers Dement (Amst)*, vol. 1, no. 2, pp. 136–143, 2015.

[100] M. Gharbiya *et al.*, “Choroidal thinning as a new finding in Alzheimer’s disease: evidence from enhanced depth imaging spectral domain optical coherence tomography.,” *J. Alzheimers. Dis.*, vol. 40, no. 4, pp. 907–917, 2014.

[101] C. Y. Cheung *et al.*, “Retinal ganglion cell analysis using high-definition optical coherence tomography in patients with mild cognitive impairment and Alzheimer’s disease.,” *J. Alzheimers. Dis.*, vol. 45, no. 1, pp. 45–56, 2015.

[102] U. Greferath, R. H. Guymer, K. A. Vessey, K. Brassington, and E. L. Fletcher, “Correlation of Histologic Features with In Vivo Imaging of Reticular Pseudodrusen,” *Ophthalmology*, vol. 123, no. 6, pp. 1320–1331, 2016.

[103] A. A. Khanifar, A. F. Koreishi, J. A. Izatt, and C. A. Toth, “Drusen Ultrastructure Imaging with Spectral Domain Optical Coherence Tomography in Age-related Macular Degeneration,” *Ophthalmology*, vol. 115, no. 11, pp. 1883-1890.e1, 2008.

[104] M. Bulut *et al.*, “Evaluation of optical coherence tomography angiographic findings in Alzheimer’s type dementia.,” *Br. J. Ophthalmol.*, vol. 102, no. 2, pp. 233–237, Feb. 2018.

[105] H. Jiang *et al.*, “Altered Macular Microvasculature in Mild Cognitive Impairment and Alzheimer Disease.,” *J. Neuroophthalmol.*, vol. 38, no. 3, pp. 292–298, Sep. 2018.

[106] S. H. Sarks, J. J. Arnold, M. C. Killingsworth, and J. P. Sarks, “Early drusen formation in the normal and aging eye and their relation to age related maculopathy: a clinicopathological study.,” *Br. J. Ophthalmol.*, vol. 83, no. 3,

pp. 358–368, Mar. 1999.

[107] D. Ardeljan and C.-C. Chan, “Aging is not a disease: distinguishing age-related macular degeneration from aging,” *Prog. Retin. Eye Res.*, vol. 37, pp. 68–89, Nov. 2013.

[108] J. Ding *et al.*, “Retinal vascular caliber and the development of hypertension: a meta-analysis of individual participant data.,” *J. Hypertens.*, vol. 32, no. 2, pp. 207–215, Feb. 2014.

[109] T. Y. Wong *et al.*, “Retinal arteriolar narrowing and risk of diabetes mellitus in middle-aged persons.,” *JAMA*, vol. 287, no. 19, pp. 2528–2533, May 2002.

[110] T. Y. Wong *et al.*, “Retinal microvascular abnormalities and incident stroke: the Atherosclerosis Risk in Communities Study.,” *Lancet (London, England)*, vol. 358, no. 9288, pp. 1134–1140, Oct. 2001.

[111] B. Aliahmad, D. K. Kumar, M. G. Sarossy, and R. Jain, “Relationship between diabetes and grayscale fractal dimensions of retinal vasculature in the Indian population,” *BMC Ophthalmol.*, vol. 14, no. 1, p. 152, 2014.

[112] S. Frost *et al.*, “Retinal vascular biomarkers for early detection and monitoring of Alzheimer’s disease,” *Transl. Psychiatry*, vol. 3, p. e233, Feb. 2013.

[113] G. Liew *et al.*, “Retinal microvascular signs and cognitive impairment.,” *J. Am. Geriatr. Soc.*, vol. 57, no. 10, pp. 1892–1896, Oct. 2009.

[114] C. Y.-L. Cheung *et al.*, “Retinal vascular tortuosity, blood pressure, and cardiovascular risk factors.,” *Ophthalmology*, vol. 118, no. 5, pp. 812–818, May 2011.

[115] E. Trucco *et al.*, *Novel VAMPIRE algorithms for quantitative analysis of the retinal vasculature*. 2013.

[116] S. McGrory *et al.*, “Towards Standardization of Quantitative Retinal Vascular Parameters: Comparison of SIVA and VAMPIRE Measurements in the Lothian Birth Cohort 1936,” *Transl. Vis. Sci. Technol.*, vol. 7, no. 2, p. 12, Mar. 2018.

[117] E. Pellegrini, G. Robertson, L. Ballerini, T. Pearson, T. J. MacGillivray, and E. Trucco, “Morphometric Retinal Measurements With Ultra-Wide-Field-Of-View Imaging,” *Conf. Comput. Math. Biomed. Eng.*, 2015.

[118] E. Pead *et al.*, “Automated detection of age-related macular degeneration in

color fundus photography: A systematic review.," *Surv. Ophthalmol.*, vol. 64, no. 4, pp. 498–511, Feb. 2019.

[119] C. Duanggate and B. Uyyanonvara, "A Review of Automatic drusen detection and segmentation from retinal images," *3rd Int. Symp. Biomed. Eng. (ISBME 2008)*, pp. 222–225, 2008.

[120] A. C. Bird *et al.*, "An international classification and grading system for age-related maculopathy and age-related macular degeneration. The International ARM Epidemiological Study Group.," *Surv. Ophthalmol.*, vol. 39, no. 5, pp. 367–374, 1995.

[121] C. A. Curcio, N. E. Medeiros, and C. L. Millican, "The Alabama Age-Related Macular Degeneration Grading System for donor eyes.," *Invest. Ophthalmol. Vis. Sci.*, vol. 39, no. 7, pp. 1085–1096, Jun. 1998.

[122] P. Checco and F. Corinto, "CNN-based algorithm for drusen identification," *2006 IEEE Int. Symp. Circuits Syst.*, p. 4, 2006.

[123] K. Rapantzikos and M. Zervakis, "Nonlinear enhancement and segmentation algorithm for the detection of age-related macular degeneration (AMD) in human eye's retina," *Proc. 2001 Int. Conf. Image Process. (Cat. No.01CH37205)*, vol. 3, pp. 1055–1058, 2001.

[124] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Trans. Syst. Man. Cybern.*, vol. 9, no. 1, pp. 62–66, 1979.

[125] S. S. Parvathi and N. Devi, "Automatic Drusen Detection from Colour Retinal Images," *Int. Conf. Comput. Intell. Multimed. Appl. (ICCIMA 2007)*, vol. 2, pp. 377–381, 2007.

[126] Xianghua Xie, "A Review of Recent Advances in Surface Defect Detection using Texture analysis Techniques," *ELCVIA Electron. Lett. Comput. Vis. Image Anal.*, vol. 7, no. 3, pp. 1–22, 2008.

[127] M. S. Nixon and A. S. Aguado, "Chapter 8 - Introduction to texture description, segmentation, and classification," pp. 399–434, 2012.

[128] P. J. Burt, "Fast filter transform for image processing," *Comput. Graph. Image Process.*, vol. 16, no. 1, pp. 20–51, 1981.

[129] L. Florack, B. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever, "Cartesian differential invariants in scale-space," *J. Math. Imaging Vis.*, vol. 3, pp. 327–348, Nov. 1993.

[130] Z. Ben Sbeh, L. D. Cohen, G. Mimoun, and G. Coscas, “A new approach of geodesic reconstruction for drusen segmentation in eye fundus images,” *IEEE Trans. Med. Imaging*, vol. 20, no. 12, pp. 1321–1333, 2001.

[131] G. Wang *et al.*, “Morphological background detection and illumination normalization of text image with poor lighting,” *PLoS One*, vol. 9, no. 11, pp. e110991–e110991, Nov. 2014.

[132] Y. Zheng, “ARIA.” The Foundation for the Prevention of Blindness, 2006.

[133] A. Hoover, “STructured Analysys of the Retina.” 2000.

[134] AREDS, “The Age-Related Eye Disease Study Severity Scale for Age-related Macular Degeneration,” *Arch Opthalmology*, vol. 123, no. 11, pp. 1484–1498, 2006.

[135] S. Cohen, “Retina Gallery,” *~full sized retina images.* [Online]. Available: available: http://retinagallery.com/index.php.

[136] E. S. Brandi C, Breinlich V, Stark KJ, “Features of Age-Related Macular Degeneration in the General Adults and Their Dependency on Age, Sex, and Smoking: Results from the German KORA Study,” *PLoS One*, no. 11, p. e0167181, 2016.

[137] S. M. Pizer *et al.*, “Adaptive histogram equalization and its variations,” *Comput. Vision, Graph. Image Process.*, vol. 39, no. 3, pp. 355–368, 1987.

[138] M. Hijazi, F. Coenen, and Y. Zheng, “Retinal Image Classification using Histogram Based Approach,” in *Neural Networks (IJCNN), The 2010 International Joint Conference on*, 2010.

[139] M. Mookiah, U. Acharya, J. Koh, and C. K. Chua, “Decision support system for age-related macular dgeneration using discrete wavelet transform,” *Med Biol Eng Comput*, vol. 52, pp. 781–796, 2014.

[140] Y. Zheng, M. Hijazi, and F. Coenen, “Automated ‘Disease/No Disease’ Grading of Age-Related Macular Degeneration by an Image Mining Approach,” *Investig. Opthalmology Vis. Sci.*, pp. 8310–8318, 2012.

[141] U. Acharya, Y. Hagiwara, J. Koh, and J. Salatha, “Automated Screening Tool for Dry Age-Related Macular Degeneration (ARMD) Using Pyramid of Histogram of Orientated Gradients (PHOG) and Nonlinear Features,” *Comput. Sci.*, vol. 20, pp. 41–51, 2017.

[142] S. Kankanaballi, P. Burlina, Y. Wolfson, and D. Freund, “Automated Classification of Severity of Age-Related Macular Degeneration from fundus photographs,” *Investig. Opthalmology Vis. Sci.*, vol. 54, no. 3, pp. 1789–1796, 2013.

[143] T. Phan, L. Seoud, H. Chakor, and F. Cheriet, “Automatic Screening and Grading of Age-Related Macular Degeneration from Texture Analysis of Fundus Images,” *J. Opthalmology*, vol. 8, pp. 1–11, 2016.

[144] M. Grinsven, Y. Lechanteur, J. van de Ven, and B. Ginneken, “Automatic Drusen Quantification and Risk Assessment of Age-Related Macular Degeneration on Color Fundus Images,” *Retina*, vol. 54, pp. 3019–3027, 2013.

[145] P. Burlina, D. Freund, B. Dupas, and N. Bressler, “Automatic Screening of Age-Related Macular Degeneration and Retinal Abnormalities,” in *33rd Annual International Conference of the IEE EMBS*, 2011, pp. 3962–3966.

[146] A. García-Floriano, Á. Ferreira-Santiago, O. Camacho-Nieto, and C. Yáñez-Márquez, “A machine learning approach to medical image classification: Detecting age-related macular degeneration in fundus images,” *Comput. Electr. Eng.*, vol. 75, pp. 218–229, 2019.

[147] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The WEKA Data Mining Software: An Update,” *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, Nov. 2009.

[148] M. R. K. Mookiah, U. Acharya, J. Koh, and V. Chandran, “Automated diagnosis of Age-related Macular Degeneration using greyscale features from digital fundus images,” *Comput. Biol. Med.*, vol. 53, pp. 55–64, 2014.

[149] S. Bonnel, S. Mohand-Said, and J.-A. Sahel, “The aging of the retina.,” *Exp. Gerontol.*, vol. 38, no. 8, pp. 825–831, Aug. 2003.

[150] H. Bay, T. Tuytelaars, and L. Van Gool, “SURF: Speeded Up Robust Features BT - Computer Vision – ECCV 2006,” 2006, pp. 404–417.

[151] G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.

[152] D. Shen, G. Wu, and H.-I. Suk, “Deep Learning in Medical Image Analysis,” *Annu. Rev. Biomed. Eng.*, vol. 19, no. 1, pp. 221–248, Jun. 2017.

[153] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, “Deep learning for healthcare: review, opportunities and challenges,” *Brief. Bioinform.*, vol. 19,

no. 6, pp. 1236–1246, Nov. 2018.

[154] D. H. Hubel and T. N. Wiesel, “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex,” *J. Physiol.*, vol. 160, no. 1, pp. 106–154, Jan. 1962.

[155] F. Rosenblatt, “The Perceptron: A Probabilistic Model for Information Storage and Organization in The Brain,” *Psychol. Rev.*, pp. 65–386, 1958.

[156] G. Cybenko, “Approximation by superpositions of a sigmoidal function,” *Math. Control. Signals Syst.*, vol. 2, no. 4, pp. 303–314, 1989.

[157] Y. Glorot, Xavier, Bordes, Antoine, and Bengio, “Deep sparse rectifier networks,” *Proc. 14th Int. Conf. Artif. Intell. Stat.*, pp. 315–323, 2011.

[158] B. Rohrer, “How do Convolutional Neural Networks work?,” *End to End Machine Learning*, 2018. .

[159] P. Y. L. Boureau, Y-Lan Jean, “A Theoretical Analysis of Feature Pooling in Visual Recognition,” in *Proceedings of the 27th International Conference on Machine Learning*, 2012.

[160] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, “Robust Object Recognition with Cortex-Like Mechanisms,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 3, pp. 411–426, Mar. 2007.

[161] A. Krizhevsky, I. Sutskever, and G. E. Hinton, *ImageNet Classification with Deep Convolutional Neural Networks*, vol. 25. 2012.

[162] et al. Y. Le Cun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel, “Handwritten digit recognition with a back-propagation network,” in . *In Advances in neural information processing systems*, 1990.

[163] A. Wald, “Statistical decision functions.,” *Stat. Decis. Funct.*, vol. 20, no. 2, pp. 165–205, 1949.

[164] H. Masnadi-shirazi and N. Vasconcelos, “On the Design of Loss Functions for Classification: theory, robustness to outliers, and SavageBoost,” *Adv. Neural Inf. Process. Syst. 21*, pp. 1049–1056, 2009.

[165] J. Kiefer and J. Wolfowitz, “tochastic Estimation of the Maximum of a Regression Function,” *Ann. Math. Stat.*, vol. 23, no. 3, pp. 462–466, 1952.

[166] O. Russakovsky *et al.*, “ImageNet Large Scale Visual Recognition Challenge,” *Int. J. Comput. Vis.*, vol. 115, pp. 211–252, 2015.

[167] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv*, no. 1409.1556, 2014.

[168] C. Szegedy, V. Vanhouke, S. Ioffe, J. Shlens, and Z. Wonja, “Rethinking the inception architecture for computer vision,” *arXiv*, no. 1512.00567, 2015.

[169] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *arXiv*, no. 1512.03385, 2015.

[170] C. Szegedy *et al.*, “Going deeper with convolutions,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1–9, 2015.

[171] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation.,” *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017.

[172] R. S. Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, “Dropout: a simple way to prevent neural networks from overfitting,” *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[173] J. H. Tan *et al.*, “Age-related Macular Degeneration detection using deep convolutional neural network,” *Futur. Gener. Comput. Syst.*, vol. 87, pp. 127–135, May 2018.

[174] P. Burlina, K. Pacheco, N. Joshi, D. Freund, and N. Bressler, “Comparing humans and deep learning performance for grading AMD:A study in using universal deep features and transfer learning,” *Comput. Biol. Med.*, vol. 82, pp. 80–86, 2017.

[175] F. Grassmann *et al.*, “A Deep Learning Algorithm for Prediction of Age-Related Eye Disease Study Severity Scale for Age-Related Macular Degeneration from Color Fundus Photography.,” *Ophthalmology*, vol. 125, no. 9, pp. 1410–1420, Sep. 2018.

[176] Q. Ji, W. He, J. Huang, and Y. Sun, “Efficient deep learning-based automated pathology identification in retinal optical coherence tomography images,” *Algorithms*, vol. 11, p. 88, 2018.

[177] T. K. Yoo, J. Y. Choi, J. G. Seo, B. Ramasubramanian, S. Selvaperumal, and D. W. Kim, “The possibility of the combination of OCT and fundus images for improving the diagnostic accuracy of deep learning for age-related macular degeneration: a preliminary experiment,” *Med. Biol. Eng. Comput.*, vol. 57, no. 3, pp. 677–687, 2019.

[178] P. Burlina, D. E. Freund, N. Joshi, and Y. Wolfson, “Detection of age-related macular degeneration via deep learning,” *Biomed. Imaging*, pp. 184–188, 2016.

[179] P. Sermanet, D. Eigen, X. Zhang, and M. Mathieu, “OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks,” *Int. Conf. Learn. Represent. (Banff).*, 2014.

[180] T. Chen *et al.*, “MXNet: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems,” Dec. 2015.

[181] K. He, X. Zhang, S. Ren, and J. Sun, *Identity Mappings in Deep Residual Networks*, vol. 9908. 2016.

[182] Y. Peng *et al.*, “DeepSeeNet: A Deep Learning Model for Automated Classification of Patient-based Age-related Macular Degeneration Severity from Color Fundus Photographs,” *Ophthalmology*, vol. 126, no. 4, pp. 565–575, 2019.

[183] A. Sala-Puigdollers *et al.*, “New software to assess retinal non-perfusion on Optomap® Wide-Field Fundus Fluorescein Angiography in Diabetic Macular Oedema,” *Invest. Ophthalmol. Vis. Sci.*, vol. 54, no. 15, 2013.

[184] P. Oellers *et al.*, “Novel grid combined with peripheral distortion correction for ultra-widefield image grading of age-related macular degeneration,” *Clin. Ophthalmol.*, vol. 11, pp. 1967–1974, Nov. 2017.

[185] Q. Xu, Y. Li, Y. Cheng, and Y. Qu, “Assessment of the effect of age on macular layer thickness in a healthy Chinese cohort using spectral-domain optical coherence tomography,” *BMC Ophthalmol.*, vol. 18, no. 1, p. 169, Dec. 2018.

[186] J. Cohen, “A Coefficient of Agreement for Nominal Scales,” *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, Apr. 1960.

[187] “The age-related eye disease study system for classifying age-related macular degeneration from stereoscopic color fundus photographs: the age-related eye disease study report number 6,” *Am. J. Ophthalmol.*, vol. 132, no. 5, pp. 668–681, 2001.

[188] G. M. McKhann *et al.*, “The diagnosis of dementia due to Alzheimer’s disease: Recommendations from the National Institute on Aging-Alzheimer’s Association workgroups on diagnostic guidelines for Alzheimer’s disease,” *Alzheimer’s Dement.*, vol. 7, no. 3, pp. 263–269, 2011.

[189] M. S. Albert *et al.*, “The diagnosis of mild cognitive impairment due to Alzheimer’s disease: Recommendations from the National Institute on Aging-Alzheimer’s Association workgroups on diagnostic guidelines for Alzheimer’s disease,” *Alzheimer’s Dement.*, vol. 7, no. 3, pp. 270–279, 2011.

[190] I. Arevalo-Rodriguez *et al.*, “Mini-Mental State Examination (MMSE) for the detection of Alzheimer’s disease and other dementias in people with mild cognitive impairment (MCI).,” *Cochrane database Syst. Rev.*, no. 3, p. CD010783, Mar. 2015.

[191] E. Pellegrini *et al.*, “Blood vessel segmentation and width estimation in ultra-wide field scanning laser ophthalmoscopy,” *Biomed. Opt. Express*, vol. 5, no. 12, pp. 4329–4337, Nov. 2014.

[192] E. Pellegrini, G. Robertson, T. MacGillivray, J. van Hemert, G. Houston, and E. Trucco, “A Graph Cut Approach to Artery/Vein Classification in Ultra-Widefield Scanning Laser Ophthalmoscopy,” *IEEE Trans. Med. Imaging*, vol. 37, no. 2, pp. 516–526, 2018.

[193] M. D. Knudtson, K. E. Lee, L. D. Hubbard, T. Y. Wong, R. Klein, and B. E. K. Klein, “Revised formulas for summarizing retinal vessel diameters.,” *Curr. Eye Res.*, vol. 27, no. 3, pp. 143–149, Sep. 2003.

[194] T. Stosic and B. D. Stosic, “Multifractal analysis of human retinal vessels,” *IEEE Trans. Med. Imaging*, vol. 25, no. 8, pp. 1101–1107, 2006.

[195] T. J. MacGillivray and N. Patton, “A reliability study of fractal analysis of the skeletonised vascular network using the ‘box-counting’ technique,” in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, 2006, pp. 4445–4448.

[196] R. Annunziata, A. Kheirkhah, S. Aggarwal, C. Bernardo, P. Hamrah, and E. Trucco, “Tortuosity classification of corneal nerves images using a multiple-scale-multiple-window approach,” in *Proceedings of the Opthalmic Medical Image Analysis (OMIS) First International Workshop, MICCAI*, 2014, pp. 1–113.

[197] A. A. Kalitzeos, G. Y. H. Lip, and R. Heitmar, “Retinal vessel tortuosity measures and their applications.,” *Exp. Eye Res.*, vol. 106, pp. 40–46, Jan. 2013.

[198] F. Family, B. R. Masters, and D. E. Platt, “Fractal pattern formation in human retinal vessels,” *Phys. D Nonlinear Phenom.*, vol. 38, no. 1, pp. 98–103, 1989.

[199] R. R. Wilcox and T. S. Tian, “Measuring effect size: a robust heteroscedastic approach for two or more groups,” *J. Appl. Stat.*, vol. 38, no. 7, pp. 1359–1368, Jul. 2011.

[200] E. Kulinskaya and M. B. Dollinger, “Robust weighted one-way ANOVA: Improved approximation and efficiency,” *J. Stat. Plan. Inference*, vol. 137, no. 2, pp. 462–472, 2007.

[201] I. Patil and C. Powell, “ggstatsplot: ‘ggplot2’ Based Plots with Statistical Details,” 2018.

[202] P. Mair and R. Wilcox, “Robust statistical methods in R using the WRS2 package.,” *Behav. Res. Methods*, May 2019.

[203] F. Wilcoxon, “Individual comparisons by ranking methods,” *Biometrics*, vol. 1, pp. 80–83, 1945.

[204] P. McCullagh and N. John, “Generalized linear models,” *CRC Press*, vol. 37, 1989.

[205] J. A. Hanley, A. Negassa, M. D. deB. Edwardes, and J. E. Forrester, “Statistical Analysis of Correlated Data Using Generalized Estimating Equations: An Orientation,” *Am. J. Epidemiol.*, vol. 157, no. 4, pp. 364–375, Feb. 2003.

[206] K.-Y. Liang and S. L. Zeger, “Longitudinal data analysis using generalized linear models,” *Biometrika*, vol. 73, no. 1, pp. 13–22, Apr. 1986.

[207] F. Galton, “Regression towards mediocrity in hereditary stature,” *J. Anthropol. Inst.*, vol. 15, pp. 246–63, 1886.

[208] T. K. Koo and M. Y. Li, “A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research,” *J. Chiropr. Med.*, vol. 15, no. 2, pp. 155–163, Jun. 2016.

[209] P. D. Bridge and S. S. Sawilowsky, “Increasing Physicians’ Awareness of the Impact of Statistics on Research Outcomes: Comparative Power of the t-test and Wilcoxon Rank-Sum Test in Small Samples Applied Research,” *J. Clin. Epidemiol.*, vol. 52, no. 3, pp. 229–235, 1999.

[210] G.-S. Ying, M. G. Maguire, R. Glynn, and B. Rosner, “Tutorial on Biostatistics: Statistical Analysis for Correlated Binary Eye Data,” *Ophthalmic Epidemiol.*, vol. 25, no. 1, pp. 1–12, Feb. 2018.

[211] R. C. Team, “R: A language and environment for statistical computing. R

Foundation for Statistical Computing." Vienna, Austria., 2013.

[212] H. Jsgaard, S. Halekoh, and J. Yan, "The R Package geepack for Generalized Estimating Equations," *J. Stat. Softw.*, vol. 2, no. 15, pp. 1–11, 2006.

[213] M. A. Williams *et al.*, "Retinal microvascular network attenuation in Alzheimer's disease," *Alzheimer's Dement. Diagnosis, Assess. Dis. Monit.*, vol. 1, no. 2, pp. 229–235, 2015.

[214] F. Huang *et al.*, "Reliability of Using Retinal Vascular Fractal Dimension as a Biomarker in the Diabetic Retinopathy Detection.," *J. Ophthalmol.*, vol. 2016, p. 6259047, 2016.

[215] R. Klein *et al.*, "Fifteen-year cumulative incidene of age-related macular degeneration: the Beaver Dam Eye Study.," *Opthalmology*, vol. 114, no. 2, pp. 253–262, Feb. 2007.

[216] C. C. Klaver, A. Ott, and A. Hofman, "Is age-related maculopathy associated with Alzheimer's disease? The Rotterdam study.," *Am. J. Epidemology*, vol. 120, no. 9, pp. 963–968, 1999.

[217] N. Joachim, P. Mitchell, G. Burlutsky, A. Kifley, and J. J. Wang, "The incidence and Progression of Age-Related Macular Degeneration over 15 Years: The Blue Mountains Eye Study," *Opthalmology*, vol. 1229, no. 12, pp. 2482–2489, 2015.

[218] G. T. Feke, B. T. Hyman, R. A. Stern, and L. R. Pasquale, "Retinal blood flow in mild cognitive impairment and Alzheimer's disease," *Alzheimers Dement.*, vol. 1, no. 2, pp. 144–151, 2015.

[219] M. Wang, Y. Zhu, Z. Shi, C. Li, and Y. Shen, "Meta-analysis of the relationship of peripheral retinal nerve fiber layer thickness to Alzheimer's disease and mild cognitive impairment," *Shanghai Arch. psychiatry*, vol. 27, no. 5, pp. 263–279, Oct. 2015.

[220] J. C. Blanks, D. R. Hinton, A. A. Sadun, and C. A. Miller, "Retinal ganglion cell degeneration in Alzheimer's disease.," *Brain Res.*, vol. 501, no. 2, pp. 364–372, Nov. 1989.

[221] J. D. Summers and J. J. Shah, "Developing Measures of Complexity for Engineering Design," pp. 381–392, Jun. 2008.