



# THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e. g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.



THE UNIVERSITY of EDINBURGH  
Edinburgh Medical School

---

**Molecular, Genetic and  
Population Health Sciences**

---

# Choroidal Image Analysis for OCT Image Sequences with Applications in Systemic Health

Jamie Burke

Doctorate of Philosophy in Precision Medicine

---

The University of Edinburgh - 2024



# Acknowledgements

First and foremost, I must extend my thanks to my supervisory team: Dr. Stuart King, Dr. Ian J.C. MacCormick, and Prof. Kenny Baillie. Your insights across both technical and clinical domains have been invaluable to the work presented for this degree and in this thesis. In particular, I want to thank Stuart for his unwavering technical and personal support throughout the COVID-19 lockdown(s), without which this doctorate degree would have certainly lacked direction. I also wish to express special gratitude to Ian for his seemingly infinite patience with my naive questions, manuscript reviews, expert ophthalmology input, and for helping to frame much of the research toward a clinical audience, always offering a long-term, forward-thinking perspective.

The first half of my PhD was completed in the School of Mathematics, and I'd like to thank the *Usual Suspects*, a group of excellent individuals with whom I had the pleasure of spending much of my first research period. You always managed to brighten my day, even in the deep and dark halls of JCMB.

Breaking up my PhD studies were two valuable internships hosted by Public Health Scotland and NHS England. Although unrelated to my PhD research, these were valuable experiences outside academia that helped broaden my understanding of both healthcare and data science. I'd like to acknowledge Theresa Ryan and Dr. Dan Schofield for their supervision at these institutions, both of whom made my time there worthwhile and enjoyable. I also want to give an honourable mention to Dr. Jim Rafferty, who helped supervise my work with NHS England, and who, alongside Dan, is fully convinced that everything, everywhere is a graph model, in some strange and peculiar way.

Upon returning to my PhD studies, I was fortunate enough to be invited to the Robert O Curle (ROC) Ophthalmology Suite at the Institute for Regeneration and Repair, thanks to the invitation of Dr. Tom MacGillivray. Consequently, I'd like to acknowledge him as an effective fourth supervisor, who helped me navigate and publish in the field of retinal imaging. It was at the ROC lab where I felt most at home, surrounded by like-minded computer vision researchers in ophthalmology, and where I navigated successful and collaborative projects. In

particular, I'd like to acknowledge Adam Threlfall, Borja Marin, Darwon Rashid, Fabian Yii, Miracle Ozzoude, Sam Gibbon, and Ylenia Giarratano for their collaboration, company and banter.

I'd also like to acknowledge Justin Engelmann for the fantastic work we've done together. Without your wizard-like skills in deep learning, this thesis would only be half of what it is today. It's been a pleasure working with someone as passionate, hard-working, and successful as you. I also want to give an honourable mention to Charlene Hamid. Without your expertise in OCT imaging and acquisition, much of the data in this thesis wouldn't have been available, and my time acquiring and extracting data was made significantly easier with your help.

It would be remiss not to mention the wonderful people in my life outside academia who have kept me afloat over the last five years. To my good friends Thomas Rodger and Conan Phimister, your company at the flat was incredible and always appreciated. Thank you for listening to me ramble on about my work and for letting me comfort-feed you some very (deliciously) gluttonous pesto pasta. To my family — Katherine, Jim, Ollie, Rachel, Nathan, and Katharine — thank you for putting up with my spells of absence and general grumpiness toward life and research over the past five years. I promise to get a real job now. You are all excellent individuals, and I'm very lucky to call you my family. I must also give a special acknowledgement to my brother Ollie, who always believed in me and encouraged me along this academic path. I therefore extend my endless gratitude (and occasional frustration) to you, good sir.

Finally, there is one very important individual I couldn't imagine not mentioning — my partner, Chloe Campbell. You have been an exceptional constant in my life these past five years and have somehow managed to put up with all my idiosyncrasies. I am forever grateful for your unwavering encouragement and emotional support, even though you still don't quite know what the choroid is. You, and the gang, have played a crucial role in keeping me motivated and goal-oriented, and you've been an indispensable distraction, allowing me to live my happiest life outside of work. You are simply fantastic, and I couldn't have asked for a better collaborator to navigate life with.

# Declaration

*I declare that this thesis has been composed by myself and that the work has not been submitted for any other degree or professional qualification. I confirm that the work submitted is my own, except where work which has formed part of jointly-authored publications has been included. My contribution and those of the other authors to this work is indicated below. I confirm that appropriate credit has been given within this thesis where reference has been made to the work of others.*

I outline below whether the content of any research chapter has been published, and highlight the author of declaration in bold type, with lead authors underlined and supervisors in italic type.

- I was the lead researcher for the work presented in chapter 2. The methodology was conceived by myself and Stuart King, and I wrote the software and carried out the method's evaluation. This work was previously published in [1, 2]:
  - “Edge Tracing using Gaussian Process Regression” by (in publication authorship order): **Jamie Burke** and *Stuart King*;
  - “Evaluation of an Automated Choroid Segmentation Algorithm in a Longitudinal Kidney Donor and Recipient Cohort” by (in publication authorship order): **Jamie Burke**, Dan Pugh, Tariq Farrah, Charlene Hamid, Emily Godden, Thomas J. MacGillivray, Neeraj Dhaun, *J. Kenneth Baillie, Stuart King and Ian J.C. MacCormick*.
- I was the lead researcher for the work presented in chapter 3. The methodology was conceived by myself and Stuart King, and I wrote the software and carried out the method's evaluation and reproducibility. This work has not been published.
- There is no lead author for the research presented in chapter 4. The methodology was conceived by myself and Justin Engelmann, under shared co-authorship. I helped collect the data, generate the ground-truth

labels for modelling and carried out the method's reproducibility. Justin Engelmann trained the deep learning model, and we both co-wrote the software and performed the method's evaluation. This work was previously published in [3]:

- “An Open-Source Deep Learning Algorithm for Efficient and Fully Automatic Analysis of the Choroid in Optical Coherence Tomography” by (in publication authorship order): **Jamie Burke**, Justin Engelmann, Charlene Hamid, Megan Reid-Schachter, Tom Pearson, Dan Pugh, Neeraj Dhaun, Amos Storkey, *Stuart King*, Thomas J. MacGillivray, Miguel O. Bernabeu and *Ian J.C. MacCormick*.
- There is no lead author for the research presented in chapter 5. The methodology was conceived by myself and Justin Engelmann, under shared co-authorship. I helped collect the data, generate the ground-truth labels for modelling and carried out the method's reproducibility. Justin Engelmann trained the deep learning model, and we both co-wrote the software and performed the method's evaluation. This work was previously published in [4]:
  - “Choroidalalyzer: An Open-Source, End-to-End Pipeline for Choroidal Analysis in Optical Coherence Tomography” by (in publication authorship order): Justin Engelmann, **Jamie Burke**, Charlene Hamid, Megan Reid-Schachter, Dan Pugh, Neeraj Dhaun, Diana Moukaddem, Lyle Gray, Niall Strang, Paul McGraw, Amos Storkey, Paul J. Steptoe, *Stuart King*, Thomas J. MacGillivray, Miguel O. Bernabeu and *Ian J.C. MacCormick*.
- I was the lead researcher who carried out the reproducibility analysis of the methods presented in chapters 4 and 5. The methodology was conceived by myself. I helped collect the data, wrote the software and carried out the reproducibility analysis. This work has been submitted and is in peer-review as [5]:
  - “OCTolyzer: Fully automatic analysis toolkit for segmentation and feature extracting in optical coherence tomography (OCT) data” by (in publication authorship order): **Jamie Burke**, Justin Engelmann, Samuel Gibbon, Charlene Hamid, Niall Strang, Neeraj Dhaun, Thomas J. MacGillivray, *Stuart King* and *Ian J.C. MacCormick*.
- I was the lead researcher who carried out the data analysis in section 6.1. I conceived the idea with help from senior researchers Neeraj Dhaun, Stuart King, J. Kenneth Baillie, Ian J.C. MacCormick and Thomas J.

## Declaration

---

MacGillivray. I carried out the image analysis, data analysis and the manuscript write-up. This work was previously published in [2]:

- “Evaluation of an Automated Choroid Segmentation Algorithm in a Longitudinal Kidney Donor and Recipient Cohort” by (in publication authorship order): **Jamie Burke**, Dan Pugh, Tariq Farrah, Charlene Hamid, Emily Godden, Thomas J. MacGillivray, Neeraj Dhaun, *J. Kenneth Baillie, Stuart King and Ian J.C. MacCormick*.
- There is no lead author for the research carried out in section 6.2. Samuel Gibbon and I conceived the idea, under shared co-authorship. I carried out the image analysis, and Samuel Gibbon and I both carried out the data analysis and the manuscript write-up. This work has been submitted and in peer-review as [6]:
  - “Association between choroidal microvasculature in the eye and Alzheimer’s disease risk in cognitively healthy midlife adults” by (in publication authorship order): **Jamie Burke**, Samuel Gibbon, Audrey Low, Charlene Hamid, Megan Reid-Schachter, Graciela Muniz-Terrera, Craig W. Ritchie, Baljean Dhillon, John T O’Brien, *Stuart King, Ian J.C. MacCormick, Thomas J. MacGillivray*.
- There is no lead author for the research carried out in section 6.3. J. Kenneth Baillie conceived the idea, with help from Ian J.C. MacCormick, David Griffiths and George Cooper on carrying out project administration and ethical approval. I carried out the data collection and the image analysis, and George Cooper and I both carried out the data analysis and manuscript write-up. A manuscript has been prepared but this work has yet to be submitted (as of December 2024):
  - “Direct Retinal Imaging for Shock Resuscitation in Critical Ill Adults II (D-RISC II)” by (in publication authorship order): George Cooper, **Jamie Burke**, Charlene Hamid, Emily Godden, Neeraj Dhaun, *Stuart King, Thomas J. MacGillivray, J. Kenneth Baillie, David M. Griffith and Ian J.C. MacCormick*.

Using TexCount, there are  $\approx 74,000$  words in this thesis. Finally, the use of “we”, “our” and “the author” in the main text is purely stylistic.

# Abstract

The retina is a light-sensitive tissue at the back of the eye and is responsible for vision. Light-sensitive photoreceptor cells in the outer retina detect light and, through a series of neuronal and vascular layers, process it into signals for the brain. The photoreceptors are perfused and maintained indirectly by the choroid and choriocapillaris, a highly vascularised layer posterior to the retina. The choroid is an extension of the central nervous system and has parallels with the renal cortex, but choroidal blood flow is four-fold higher per unit mass than the kidney and ten-fold higher than the brain. Thus, there has been growing interest in the structure and function of the choroidal circulation reflecting physiological status of systemic disease in the kidney and brain. The choroid can be imaged using optical coherence tomography (OCT), a non-invasive imaging technique which uses interferometry to capture three-dimensional, cross-sectional visualisations of ocular tissue at micron resolution. Advancements in OCT technology now permit deeper penetration and improved visualisation of the choroidal vessels. However, conventional methods of characterising and quantifying this vascular space have not kept pace with the improvements in OCT technology which visualise it, resulting in non-standardised manual or semi-automatic approaches as commonplace methods for choroidal measurement. The ability to measure anatomy consistently at micron-scale both intra- and inter-patient is paramount to capturing the inherent biological change or signal being studied — a signal which can be corrupted when exposed to human subjectivity.

In this thesis, I develop and evaluate several novel methods to analyse the choroid in OCT image sequences, with each successive method markedly improving on its predecessors. In the first instance, I develop two semi-automatic

approaches for choroid region (Gaussian Process Edge Tracing, *GPET*) and vessel (Multi-scale Median Cut Quantisation, *MMCQ*) analysis, which improve on manual approaches but ultimately are biased by the end-user's biological interpretation and technical experience. As a first step to fully automatic choroid segmentation, I develop *DeepGPET* as a deep learning-based method for choroid region segmentation, which significantly improves on semi-automatic approaches in terms of time, reproducibility, and end-user accessibility. However, *DeepGPET* lacks choroidal vessel quantification and still requires manual input for generating standardised, choroid-derived measurements. Improving on this, I developed *Choroidalizer*, a fully automatic, deep learning-based, end-to-end pipeline which fully characterises the choroidal space and vessels, and automatically generates clinically meaningful and reproducible choroid-derived metrics. I provide rigorous evaluation of these four approaches, and consider their use-case and potential clinical value in three distinct applications into systemic health: *OCTANE*: evaluating longitudinal choroidal change and its association with renal function in transplant recipients and donors; *PREVENT*: investigating associations between the choroid and risk factors for developing later-life Alzheimer's disease in a mid-life cohort; *D-RISCii*: assessing choroidal variation and feasibility of OCT imaging in critical care. This thesis has contributed several new approaches to the research community which are all open-source and freely available, enabling consistent and reproducible measurement of the choroid. This thesis also highlights the potential role the choroid may play in reflecting pathophysiology in the kidney, brain and wider systemic health from iatrogenic shock, thus helping accelerate the nascent field of choroidal analysis in OCT image sequences.

**Keywords** – choroid – optical coherence tomography – image analysis – deep learning – systemic health

**Lay summary** The retina is an essential part of the eye which is responsible for vision through detecting light and converting it into signals for the brain. The light-sensitive photoreceptors in the retina play a crucial role in this process and are supported and maintained by a dense network of blood vessels

in the choroid, a predominantly vascular layer located behind the retina. The choroid is part of the central nervous system, shares similarities with organs like the kidney, and has a much higher blood flow rate per unit mass compared to the kidney or brain, in part to support the energy demands of the photoreceptor cells. Consequently, there is growing interest from the research community for studying the choroid in the context of systemic diseases that affect these organs. The choroid can be visualised using optical coherence tomography (OCT), a non-invasive imaging technique that provides detailed, three-dimensional visualisations of the back of the eye at micro-metre resolution. Recent advancements in OCT technology have significantly improved visualisation of the choroidal vessels, thus permitting a unique assessment into the structure and function of this microvascular network. However, traditional, manual or semi-automatic approaches to measure the choroid have not kept pace with the improving OCT technology that visualise it, leading to inconsistent results which have the potential to be impacted by human subjectivity.

In this thesis, I develop and evaluate new methods for analysing the choroid in OCT image sequences. I first create two semi-automatic approaches to analyse the choroidal space (Gaussian Process Edge Tracing, *GPET*) and vessels (Multi-scale Median Cut Quantisation, *MMCQ*), which improve on manual methods but still depend on user expertise. I then develop a fully automatic deep learning method, *DeepGPET*, for isolating the choroidal space in OCT images. This method is faster and more consistent, though still requires manual input to measure the choroid. As a final improvement for analysing the choroid in OCT image sequences, I develop *Choroidalizer*, a fully automatic, deep learning-based toolkit that comprehensively quantifies the choroid and its blood vessels. *Choroidalizer* is a significant improvement on previous approaches and provides reliable and clinically meaningful measurements of the choroid without human bias. I thoroughly test these new approaches to choroidal image analysis, and explore their potential for characterising the choroid in the context of systemic disease using three distinct applications: *OCTANE*: evaluating longitudinal choroidal change and its association with renal function in transplant recipients and donors; *PREVENT*: investigating associations between the choroid and risk factors for developing later-life Alzheimer's dis-



## Abstract

---

ease in a mid-life cohort; D-RISCii: assessing choroidal variation and feasibility of OCT imaging in critical care. This thesis contributes several new, freely available tools for quicker, accurate, and more consistent measurement of the choroid compared with previous manual or semi-automatic approaches. This thesis also emphasises the potential role of the choroid in reflecting overall systemic health, ultimately helping accelerate the nascent field of choroidal analysis in OCT image sequences.

# Contents

|  |               |
|--|---------------|
| <b>Acknowledgements</b>                      | <b>ii</b>     |
| <b>Declaration</b>                           | <b>iv</b>     |
| <b>Abstract</b>                              | <b>vii</b>    |
| <b>Acronyms</b>                              | <b>xviii</b>  |
| <b>List of Figures</b>                       | <b>xxi</b>    |
| <b>List of Tables</b>                        | <b>xxviii</b> |
| <b>List of Algorithms</b>                    | <b>xxxii</b>  |
| <b>1 Introduction</b>                        | <b>1</b>      |
| 1.1 Guiding principles . . . . .             | 3             |
| 1.2 The structure of this thesis . . . . .   | 4             |
| 1.3 PhD outputs . . . . .                    | 6             |
| 1.3.1 Publication portfolio . . . . .        | 6             |
| 1.3.2 Technical portfolio . . . . .          | 9             |
| 1.4 The choroidal microvasculature . . . . . | 10            |
| 1.4.1 Systemic disease . . . . .             | 13            |
| 1.5 Optical coherence tomography . . . . .   | 15            |
| 1.5.1 Time-domain OCT . . . . .              | 17            |
| 1.5.2 Spectral-domain OCT . . . . .          | 18            |
| 1.5.3 Enhanced depth imaging OCT . . . . .   | 20            |

## Contents

---

|          |   |           |
|----------|---|-----------|
| 1.5.4    | Swept-source OCT . . . . .  | 23        |
| 1.6      | Segmenting the choroidal microvasculature . . . . .   | 26        |
| 1.6.1    | Region segmentation . . . . .   | 28        |
| 1.6.1.1  | Manual measurement . . . . .  | 28        |
| 1.6.1.2  | Semi- and Fully-automatic . . . . .   | 30        |
| 1.6.2    | Vessel segmentation . . . . .   | 35        |
| 1.6.3    | Synthesising region and vessel segmentation . . . . .   | 37        |
| 1.6.4    | Current challenges . . . . .  | 39        |
| 1.6.4.1  | Vessel segmentation . . . . .   | 39        |
| 1.6.4.2  | Bias, reproducibility and accessibility . . . . .   | 41        |
| 1.6.5    | Summary . . . . .   | 42        |
| 1.7      | Measuring the choroidal microvasculature . . . . .  | 43        |
| 1.7.1    | OCT B-scan . . . . .  | 46        |
| 1.7.1.1  | Pixel length-scales . . . . .   | 48        |
| 1.7.1.2  | Different viewing modes . . . . .   | 49        |
| 1.7.1.3  | Regions of interest . . . . .   | 50        |
| 1.7.2    | OCT volume scan . . . . .   | 55        |
| 1.7.3    | OCT peripapillary scan . . . . .  | 58        |
| 1.7.4    | Comparing measurements . . . . .  | 59        |
| 1.7.4.1  | Choroid-derived metrics . . . . .   | 60        |
| 1.7.4.2  | Segmentation metrics . . . . .  | 63        |
| 1.8      | Executive summary . . . . .   | 65        |
| <b>2</b> | <b>GPET: Semi-automatic choroid region segmentation using Gaussian process regression</b> . . . . . | <b>67</b> |
| 2.1      | Introduction . . . . .  | 67        |
| 2.1.1    | Gaussian process regression . . . . .   | 69        |
| 2.1.1.1  | Prior distributions and covariance functions . . . . .  | 71        |
| 2.1.1.2  | Posterior distributions . . . . .   | 75        |
| 2.2      | GPET's Methodology . . . . .  | 78        |
| 2.2.1    | Pre-processing . . . . .  | 78        |
| 2.2.1.1  | Gradient-based edge detection . . . . .   | 80        |

---

|          |   |            |
|----------|---|------------|
| 2.2.2    | Initialisation . . . . .  | 84         |
| 2.2.2.1  | Prior distribution . . . . .  | 84         |
| 2.2.2.2  | Posterior distribution . . . . .  | 85         |
| 2.2.3    | Curve scoring . . . . .   | 87         |
| 2.2.4    | Pixel scoring . . . . .   | 88         |
| 2.2.5    | Accept-discard scheme . . . . .   | 90         |
| 2.2.6    | Refitting . . . . .   | 92         |
| 2.2.7    | Convergence . . . . .   | 93         |
| 2.2.8    | Pipeline . . . . .  | 94         |
| 2.3      | GPET's evaluation . . . . .   | 96         |
| 2.3.1    | Data and methods . . . . .  | 96         |
| 2.3.1.1  | Study population . . . . .  | 96         |
| 2.3.1.2  | Image acquisition . . . . .   | 98         |
| 2.3.1.3  | Manual measurement of choroid thickness . . . . .   | 99         |
| 2.3.1.4  | Automatic measurement of choroid thickness . . . . .  | 100        |
| 2.3.1.5  | Statistical analysis . . . . .  | 101        |
| 2.3.2    | Results . . . . .   | 102        |
| 2.4      | Discussion . . . . .  | 106        |
| 2.4.1    | Evaluation . . . . .  | 108        |
| 2.4.2    | Limitations and future work . . . . .   | 110        |
| 2.4.3    | Outputs . . . . .   | 112        |
| 2.4.4    | Executive summary . . . . .   | 113        |
| <b>3</b> | <b>MMCQ: Semi-automatic choroid vessel segmentation using multi-scale median cut quantisation</b> . . . . . | <b>114</b> |
| 3.1      | Introduction . . . . .  | 114        |
| 3.1.1    | Niblack autolocal thresholding . . . . .  | 119        |
| 3.1.2    | Contrast limited adaptive histogram equalisation . . . . .  | 124        |
| 3.1.3    | Median cut image quantisation . . . . .   | 126        |
| 3.1.4    | Vessel shadow compensation . . . . .  | 128        |
| 3.2      | MMCQ's methodology . . . . .  | 129        |
| 3.2.1    | Pre-processing . . . . .  | 131        |

|          |  |            |
|----------|--|------------|
| 3.2.2    | Patch quantisation and enhancement . . . . .                               | 132        |
| 3.2.2.1  | Patch size . . . . .   | 132        |
| 3.2.2.2  | Quantisation . . . . .   | 133        |
| 3.2.2.3  | Enhancement . . . . .  | 135        |
| 3.2.3    | Merging . . . . .  | 140        |
| 3.2.3.1  | Patch merging . . . . .  | 140        |
| 3.2.3.2  | Scale merging . . . . .  | 142        |
| 3.2.4    | Vessel segmentation . . . . .  | 142        |
| 3.2.5    | Pipeline . . . . .   | 143        |
| 3.3      | MMCQ's evaluation . . . . .  | 146        |
| 3.3.1    | Study population . . . . .   | 146        |
| 3.3.2    | Statistical analysis . . . . .   | 147        |
| 3.3.3    | Results . . . . .  | 148        |
| 3.4      | MMCQ's Reproducibility . . . . .   | 155        |
| 3.4.1    | Study population . . . . .   | 155        |
| 3.4.2    | Statistical analysis . . . . .   | 156        |
| 3.4.3    | Results . . . . .  | 158        |
| 3.4.3.1  | Population-level . . . . .   | 158        |
| 3.4.3.2  | Eye-level . . . . .  | 160        |
| 3.5      | Discussion . . . . .   | 164        |
| 3.5.1    | Evaluation . . . . .   | 164        |
| 3.5.2    | Reproducibility . . . . .  | 167        |
| 3.5.3    | Limitations and future work . . . . .                                      | 168        |
| 3.5.4    | Outputs . . . . .  | 170        |
| 3.5.5    | Executive summary . . . . .  | 171        |
| <b>4</b> | <b>DeepGPET: Efficient and fully-automatic choroid region segmentation</b> | <b>172</b> |
| 4.1      | Introduction . . . . .   | 172        |
| 4.1.1    | Deep learning for image segmentation . . . . .                             | 174        |
| 4.2      | Study populations . . . . .  | 177        |
| 4.2.1    | Model training & evaluation . . . . .                                      | 177        |

|          |   |            |
|----------|---|------------|
| 4.2.1.1  | Sample derivation . . . . .   | 178        |
| 4.2.2    | Reproducibility . . . . .   | 180        |
| 4.3      | DeepGPET's deep learning model . . . . .  | 181        |
| 4.4      | DeepGPET's evaluation . . . . .   | 183        |
| 4.4.1    | Statistical analysis . . . . .  | 183        |
| 4.4.2    | Results . . . . .   | 185        |
| 4.4.2.1  | Quantitative evaluation . . . . .   | 185        |
| 4.4.2.2  | Qualitative evaluation . . . . .  | 188        |
| 4.4.2.3  | Potential generalisability . . . . .  | 189        |
| 4.5      | DeepGPET's reproducibility . . . . .  | 192        |
| 4.5.1    | Statistical analysis . . . . .  | 192        |
| 4.5.2    | Results . . . . .   | 193        |
| 4.5.2.1  | Population-level . . . . .  | 193        |
| 4.5.2.2  | Eye-level . . . . .   | 195        |
| 4.6      | Discussion . . . . .  | 197        |
| 4.6.1    | Evaluation . . . . .  | 197        |
| 4.6.2    | Reproducibility . . . . .   | 200        |
| 4.6.3    | Limitations and Future work . . . . .   | 201        |
| 4.6.4    | Outputs . . . . .   | 205        |
| 4.6.5    | Executive summary . . . . .   | 205        |
| <b>5</b> | <b>Choroidalizer: End-to-end choroidal analysis in optical coherence tomography</b> | <b>207</b> |
| 5.1      | Introduction . . . . .  | 207        |
| 5.1.1    | The UNet Deep Learning Model . . . . .  | 210        |
| 5.2      | Study populations . . . . .   | 212        |
| 5.2.1    | Model training and evaluation . . . . .   | 212        |
| 5.2.1.1  | Ground truth labels . . . . .   | 214        |
| 5.2.1.2  | Sample derivation . . . . .   | 216        |
| 5.2.2    | Reproducibility . . . . .   | 220        |
| 5.3      | Choroidalizer's deep learning model . . . . .                                       | 221        |
| 5.4      | Choroidalizer's evaluation . . . . .  | 224        |

|          |  |            |
|----------|--|------------|
| 5.4.1    | Statistical analysis   | 224        |
| 5.4.1.1  | Quantitative evaluation  | 224        |
| 5.4.1.2  | Qualitative evaluation   | 225        |
| 5.4.2    | Results  | 226        |
| 5.4.2.1  | Performance on internal and external test sets   | 226        |
| 5.4.2.2  | Comparison with manual segmentations   | 231        |
| 5.4.2.3  | Detailed error analysis  | 233        |
| 5.4.2.4  | Potential generalisability   | 235        |
| 5.5      | Choroidalizer's Reproducibility  | 237        |
| 5.5.1    | Statistical analysis   | 237        |
| 5.5.2    | Results  | 239        |
| 5.5.2.1  | Population-level   | 239        |
| 5.5.2.2  | Eye-level  | 241        |
| 5.6      | Discussion   | 243        |
| 5.6.1    | Evaluation   | 243        |
| 5.6.2    | Reproducibility  | 248        |
| 5.6.3    | Limitations and future work  | 249        |
| 5.6.4    | Outputs  | 251        |
| 5.6.5    | Executive summary  | 251        |
| <b>6</b> | <b>Applications of choroidal image analysis in OCT image sequences</b>                           | <b>253</b> |
| 6.1      | The OCTANE Study: Choriorretinal changes in a longitudinal cohort of kidney donor and recipients | 253        |
| 6.1.1    | Introduction   | 253        |
| 6.1.2    | Data   | 255        |
| 6.1.2.1  | Study population   | 255        |
| 6.1.2.2  | Image collection and analysis  | 255        |
| 6.1.2.3  | Statistical analysis   | 256        |
| 6.1.3    | Results  | 257        |
| 6.1.3.1  | Participants   | 257        |
| 6.1.3.2  | Choroidal variation  | 259        |
| 6.1.4    | Discussion   | 262        |

|         |  |     |
|---------|--|-----|
| 6.1.4.1 | Outputs . . . . .  | 266 |
| 6.1.4.2 | Executive summary . . . . .  | 267 |
| 6.2     | The PREVENT Dementia Study: Association between choroidal microvasculature and Alzheimer’s disease risk . . . . .          | 268 |
| 6.2.1   | Introduction . . . . .   | 268 |
| 6.2.2   | Data . . . . .   | 269 |
| 6.2.2.1 | Study population and image capture . . . . .   | 269 |
| 6.2.2.2 | Image analysis . . . . .   | 270 |
| 6.2.2.3 | Statistical analysis . . . . .   | 270 |
| 6.2.3   | Results . . . . .  | 272 |
| 6.2.3.1 | Participants . . . . .   | 272 |
| 6.2.3.2 | Choroidal variation . . . . .  | 275 |
| 6.2.4   | Discussion . . . . .   | 278 |
| 6.2.4.1 | Outputs . . . . .  | 285 |
| 6.2.4.2 | Executive summary . . . . .  | 285 |
| 6.3     | D-RISC II: A phase two evaluation of retinal choroid biomarkers for shock resuscitation in critically ill adults . . . . . | 286 |
| 6.3.1   | Introduction . . . . .   | 286 |
| 6.3.2   | Data . . . . .   | 288 |
| 6.3.2.1 | Study population . . . . .   | 288 |
| 6.3.2.2 | Image acquisition . . . . .  | 288 |
| 6.3.2.3 | Image analysis . . . . .   | 291 |
| 6.3.2.4 | Statistical analysis . . . . .   | 291 |
| 6.3.3   | Results . . . . .  | 293 |
| 6.3.3.1 | Participants . . . . .   | 293 |
| 6.3.3.2 | Feasibility . . . . .  | 293 |
| 6.3.3.3 | Choroidal variation . . . . .  | 297 |
| 6.3.4   | Discussion . . . . .   | 301 |
| 6.3.4.1 | Outputs . . . . .  | 305 |
| 6.3.4.2 | Executive summary . . . . .  | 305 |



|          |  |            |
|----------|--|------------|
| <b>7</b> | <b>Discussion</b>  | <b>307</b> |
| 7.1      | Core contributions to the field . . . . .                            | 307        |
| 7.2      | The technical work accomplished in this thesis . . . . .             | 309        |
| 7.2.1    | Outline of thesis progression . . . . .                              | 310        |
| 7.2.2    | Choroidalalyzer’s key contributions . . . . .                        | 312        |
| 7.2.3    | From imaging device to choroid measurement . . . . .                 | 314        |
| 7.2.4    | Closing the gap between research and clinical practice . . . . .     | 316        |
| 7.3      | Current limitations . . . . .  | 319        |
| 7.4      | Future work . . . . .  | 321        |
| 7.4.1    | Large-scale analysis, external validation and benchmarking . . . . . | 321        |
| 7.4.2    | Towards choroidal vessel standardisation . . . . .                   | 323        |
| 7.4.3    | Context-dependent image quality . . . . .                            | 323        |
| 7.4.4    | Choroid depth inference from en face imaging . . . . .               | 324        |
| 7.5      | Concluding remarks . . . . .   | 325        |
|          | <b>Bibliography</b>  | <b>326</b> |
|          | <b>Appendices</b>  | <b>376</b> |
| <b>A</b> | <b>Protocol Documentation</b>  | <b>378</b> |
| A.1      | OCT B-scan definitions . . . . .                                     | 378        |
| A.2      | Equipment check . . . . .  | 381        |
| A.3      | OCT B-scan quality guide . . . . .                                   | 382        |
| A.3.1    | Choroidal space OCT image quality . . . . .                          | 382        |
| A.3.2    | Choroidal vasculature OCT image quality . . . . .                    | 383        |
| A.4      | OCT B-scan vessel guide . . . . .                                    | 384        |
| A.5      | OCT B-scan manual segmentation protocol . . . . .                    | 386        |
| A.5.1    | ITK-SNAP setup . . . . .   | 386        |
| A.5.2    | Load in image, label definition, main tools . . . . .                | 386        |
| A.5.3    | Segmentation . . . . .   | 387        |
| A.5.3.1  | Region segmentation . . . . .  | 387        |
| A.5.3.2  | Vessel segmentation . . . . .  | 390        |
| A.5.4    | Saving segmentations . . . . .                                       | 391        |

# Acronyms

**AMD** Age-Related Macular Degeneration.

**APOE4** Apolipoprotein E4.

**ART** Automatic Real Time.

**AUC** Area Under the Receiver Operating Characteristic Curve.

**CFP** Colour Fundus Photography.

**CI** Confidence Interval.

**CKD** Chronic Kidney Disease.

**CLAHE** Contrast Limited Adaptive Histogram Equalisation.

**CNN** Convolutional Neural Network.

**CVI** Choroid Vascular Index.

**D-RISC** Direct Retinal Imaging for Shock Resuscitation in Critically Ill Adults.

**DVCKD** Diurnal Variation for Chronic Kidney Disease.

**EDI** Enhanced Depth Imaging.

**EDI-OCT** Enhanced Depth Imaging Optical Coherence Tomography.

**EOI** Edge Of Interest.

**ETDRS** Early Treatment Diabetic Retinopathy Study.

**FH** Family History.

**FOV** Field of View.

- GCU** Glasgow Caledonian University.
- GPET** Gaussian Process Edge Tracing.
- GPU** Graphics Processing Unit.
- HDU** High Dependency Unit.
- HRA** Heidelberg Retina Angiography.
- ITU** Intensive Therapy Unit.
- MAE** Mean Absolute Error.
- MD** Mean Difference.
- MMCQ** Multi-scale Median Cut Quantisation.
- OCT** Optical Coherence Tomography.
- OCT-A** Optical Coherence Tomography - Angiography.
- OCTANE** Optical Coherence Tomography And NEphropathy.
- OIT** Oblique Illumination Test.
- OR** Odds Ratio.
- PCA** Posterior Ciliary Arteries.
- PPD** Posterior Predictive Distribution.
- RBF** Radial Basis Function.
- ROI** Region Of Interest.
- RPE** Retinal Pigment Epithelium.
- SD** Standard Deviation.
- SD-OCT** Spectral Domain Optical Coherence Tomography.
- SFCT** Subfoveal Choroid Thickness.
- SLO** Scanning Laser Ophthalmoscopy.
- SNR** Signal-to-Noise Ratio.

**SOTA** State-Of-The-Art.

**SS-OCT** Swept Source Optical Coherence Tomography.

**TD-OCT** Time Domain Optical Coherence Tomography.

# List of Figures

|      |  |    |
|------|--|----|
| 1.1  | Schematic of work conducted during research period. . . . .  | 5  |
| 1.2  | Exemplar, annotated SLO and OCT scan. . . . .  | 16 |
| 1.3  | Simple schematic of an OCT imaging system. . . . .   | 18 |
| 1.4  | Comparison of TD-OCT and conventional SD-OCT. . . . .  | 19 |
| 1.5  | Comparison of conventional SD-OCT and EDI-OCT. . . . .   | 21 |
| 1.6  | Comparison of B-scans with different ART values with EDI mode toggled on and off. . . . .                        | 22 |
| 1.7  | SNR sensitivity roll-off curves for Spectralis SD-OCT and hypothetical SS-OCT. . . . .                           | 25 |
| 1.8  | Comparison of SS-OCT and EDI-OCT from different OCT devices. . . . .   | 25 |
| 1.9  | Choroid segmentation tasks and measurements of interest. . . . .   | 27 |
| 1.10 | Exemplar, annotated OCT B-scan. . . . .  | 44 |
| 1.11 | Different definitions of the Choroid-Sclera boundary. . . . .  | 45 |
| 1.12 | Demonstration of ROI constructing and thickness measurement. . . . .   | 47 |
| 1.13 | Relationship between scan focus and lateral pixel length-scale on the Heidelberg Engineering OCT system. . . . . | 49 |
| 1.14 | Viewing modes on the Heidelberg Engineering OCT system. . . . .  | 50 |
| 1.15 | ROIs aligned to the image- and choroid-axis for OCT B-scans with skew and/or curvature. . . . .                  | 51 |
| 1.16 | Different examples of choroid skew and curvature for varying refractive errors. . . . .                          | 55 |
| 1.17 | Procedure to generate spatial en face choroid maps from OCT. . . . .   | 56 |
| 1.18 | Diagram of generating spatial choroid measurements across the macula from OCT volume scans. . . . .              | 57 |

---

|      |  |     |
|------|--|-----|
| 1.19 | Diagram of generating spatial choroid measurements in OCT peripapillary B-scans. . . . .               | 59  |
| 2.1  | Example Gaussian process prior distributions with different covariance functions. . . . .              | 74  |
| 2.2  | Example Gaussian process posterior distributions with different covariance functions. . . . .          | 77  |
| 2.3  | Pre-processing for GPET's image analysis pipeline. . . . .   | 79  |
| 2.4  | Denoising and contrast enhancement of RPE-Choroid and Choroid-Sclera boundaries. . . . .               | 81  |
| 2.5  | Gradient-based edge maps for RPE-Choroid and Choroid-Sclera boundaries. . . . .                        | 82  |
| 2.6  | Enhanced, gradient-based edge maps for RPE-Choroid and Choroid-Sclera boundaries. . . . .              | 83  |
| 2.7  | Gaussian process prior distributions for the RPE-Choroid and Choroid-Sclera boundaries. . . . .        | 85  |
| 2.8  | Initial Gaussian process PPDs for the RPE-Choroid and Choroid-Sclera boundaries. . . . .               | 86  |
| 2.9  | Frequency density from posterior curves of the PPD during Choroid-Sclera boundary fitting. . . . .     | 89  |
| 2.10 | Influence of the observation noise during GPET's Choroid-Sclera boundary fitting. . . . .              | 93  |
| 2.11 | Influence of increasing the observation set on the PPD during Choroid-Sclera boundary fitting. . . . . | 94  |
| 2.12 | Convergence of GPET's image analysis pipeline. . . . .   | 95  |
| 2.13 | Schematic of GPET's image analysis pipeline. . . . .   | 95  |
| 2.14 | Sample derivation flowchart of the renal cohort used to validate GPET. . . . .                         | 98  |
| 2.15 | Demonstration of manual and automated methods to measure the choroid. . . . .                          | 100 |
| 2.16 | GPET's population-level agreement against manual choroid thickness. . . . .                            | 103 |
| 2.17 | GPET's cohort- and location-level agreement with manual choroid thickness. . . . .                     | 104 |
| 2.18 | Examples of major disagreements between manual and automatic choroid thickness. . . . .                | 106 |

|      |   |     |
|------|---|-----|
| 2.19 | Measurement error in manual choroid thickness. . . . .  | 107 |
| 3.1  | Relationship between total choroid area and vessel area. . . . .  | 115 |
| 3.2  | Vessel segmentation in an OCT B-scan. . . . .   | 116 |
| 3.3  | Poor agreement between manual vessel segmentation labels. . . . .   | 118 |
| 3.4  | Grid-searching optimal thresholding parameter configurations for Niblack's method against manual labelling. . . . . | 121 |
| 3.5  | Error sensitivity of Niblack's thresholding parameters against manual labelling. . . . .                            | 122 |
| 3.6  | Qualitative comparison of the Niblack method against manual labelling. . . . .                                      | 123 |
| 3.7  | Demonstration of CLAHE on an OCT B-scan. . . . .  | 125 |
| 3.8  | Demonstration of CLAHE at the patch-level during pre-processing. . . . .  | 126 |
| 3.9  | Demonstration of median cut quantisation on an OCT B-scan. . . . .  | 127 |
| 3.10 | Correspondence between en face retinal vessels and A-scan shadowing on an OCT B-scan. . . . .                       | 129 |
| 3.11 | Superficial retinal vessel shadow compensation scheme. . . . .  | 130 |
| 3.12 | Influence of vessel shadow compensation on downstream vessel segmentation performance. . . . .                      | 130 |
| 3.13 | Initial pre-processing of an OCT B-scan before MMCQ's pipeline. . . . .   | 132 |
| 3.14 | Pre-processing at the patch-level before applying MMCQ's pipeline. . . . .  | 134 |
| 3.15 | Median cut quantisation at the patch-level. . . . .   | 135 |
| 3.16 | Patch-level enhancement scheme. . . . .   | 136 |
| 3.17 | Custom Sigmoid function to weight enhancement strength. . . . .   | 137 |
| 3.18 | Schematic of quantisation and enhancement at the patch-level. . . . .   | 139 |
| 3.19 | Patch-level result after MMCQ's enhancement. . . . .  | 139 |
| 3.20 | MMCQ's enhancement at $40 \times 40$ scale. . . . .   | 140 |
| 3.21 | MMCQ's enhancement at choroid-dependent scales. . . . .   | 141 |
| 3.22 | MMCQ's enhancement compared to simple pre-processing. . . . .   | 141 |
| 3.23 | MMCQ's multi-depth segmentation scheme. . . . .   | 143 |
| 3.24 | Schematic diagram of MMCQ's image analysis pipeline. . . . .  | 144 |
| 3.25 | MMCQ's image analysis pipeline applied to several exemplar OCT B-scans. . . . .                                     | 144 |

---

|      |  |     |
|------|--|-----|
| 3.26 | Correlation in choroid vessel area between MMCQ, Niblack and the manual graders. . . . .         | 149 |
| 3.27 | Agreement in choroid vessel area between MMCQ, Niblack and the manual graders. . . . .           | 150 |
| 3.28 | Qualitative comparison between MMCQ, Niblack and manual graders. . . . .                         | 151 |
| 3.29 | Correlation in CVI between MMCQ, Niblack and the manual graders.                                 | 153 |
| 3.30 | Agreement in CVI between MMCQ, Niblack and the manual graders.                                   | 153 |
| 3.31 | OCT B-scans from the excluded eyes with poor vessel visualisation. . . . .                       | 157 |
| 3.32 | Population-level agreement in repeated SD-OCT volume data for MMCQ and Niblack. . . . .          | 160 |
| 3.33 | Population-level agreement in repeated SS-OCT B-scan data for MMCQ and Niblack. . . . .          | 161 |
| 3.34 | Eye-level agreement in repeated SD-OCT volume data for MMCQ and Niblack. . . . .                 | 162 |
| 3.35 | Eye-level agreement in repeated SS-OCT B-scan data for MMCQ and Niblack. . . . .                 | 162 |
|      |  |     |
| 4.1  | Comparison between GPET's complexity and DeepGPET's simplicity. . . . .                          | 173 |
| 4.2  | The concept of fine-tuning in deep learning for semantic image segmentation. . . . .             | 176 |
| 4.3  | Sample derivation for DeepGPET's modelling dataset. . . . .                                      | 178 |
| 4.4  | DeepGPET's agreement with GPET's ground truth labels. . . . .                                    | 186 |
| 4.5  | DeepGPET's error distribution in thickness and area against GPET's ground truth labels. . . . .  | 187 |
| 4.6  | Adjudication examples between DeepGPET and GPET. . . . .   | 189 |
| 4.7  | DeepGPET's potential generalisability to peripapillary OCT B-scans. . . . .                      | 190 |
| 4.8  | DeepGPET's qualitative performance on a small set of SS-OCT B-scans. . . . .                     | 191 |
| 4.9  | DeepGPET's population-level reproducibility performance on SD-OCT volume scans. . . . .          | 194 |
| 4.10 | DeepGPET's population-level reproducibility performance on SD-OCT peripapillary B-scans. . . . . | 195 |



## List of Figures

---

|      |   |     |
|------|---|-----|
| 4.11 | DeepGPET's eye-level reproducibility performance. . . . .   | 196 |
| 4.12 | High-level, qualitative comparison between DeepGPET and GPET.   | 200 |
| 5.1  | Choroidalyzer's proposed benefits over current state-of-the-art. . .  | 209 |
| 5.2  | Demonstrative UNet architecture for choroid segmentation. . . .   | 211 |
| 5.3  | Improved MMCQ vessel segmentation by varying brightness/-<br>contrast and taking majority vote. . . . .                   | 214 |
| 5.4  | Choroidalyzer's ground truth labels. . . . .  | 216 |
| 5.5  | Sample derivation flowchart for Choroidalyzer's modelling dataset.  | 217 |
| 5.6  | Choroidalyzer's UNet deep learning architecture. . . . .  | 222 |
| 5.7  | Choroidalyzer's segmentation performance compared to Deep-<br>GPET. . . . .   | 229 |
| 5.8  | Choroidalyzer's agreement in choroid-derived metrics in the in-<br>ternal test set. . . . .                               | 229 |
| 5.9  | Choroidalyzer's agreement in choroid-derived metrics in the ex-<br>ternal test set. . . . .                               | 230 |
| 5.10 | Choroidalyzer's fovea detection error analysis results. . . . .   | 230 |
| 5.11 | Choroidalyzer's fovea detection error distributions. . . . .  | 234 |
| 5.12 | Choroidalyzer's ability to locate the foveal pixel coordinate. . . .  | 236 |
| 5.13 | Choroidalyzer's ability to segment challenging cases of poor Choroid-<br>Sclera boundary visibility. . . . .              | 237 |
| 5.14 | Choroidalyzer's tunable region threshold enables different Choroid-<br>Sclera boundary definitions to be applied. . . . . | 238 |
| 5.15 | Choroidalyzer's population-level reproducibility agreement for de-<br>rived metrics. . . . .                              | 240 |
| 5.16 | Choroidalyzer's eye-level reproducibility agreement for derived<br>metrics. . . . .                                       | 242 |
| 5.17 | Examples of Choroidalyzer applied to OCT data from different<br>devices. . . . .  | 245 |
| 6.1  | Sample derivation flowchart for the renal cohort. . . . .   | 259 |
| 6.2  | Longitudinal choroid variation in the renal cohort. . . . .   | 261 |
| 6.3  | Derivation of choroidal measurements in PREVENT cohort. . . .   | 271 |
| 6.4  | Sample derivation flowchart for the PREVENT cohort. . . . .   | 273 |

---

|      |   |     |
|------|---|-----|
| 6.5  | Examples demonstrating poor choroid visualisation using conventional SD-OCT. . . . .  | 274 |
| 6.6  | Examples demonstrating the added benefit of EDI-OCT over conventional SD-OCT. . . . .   | 274 |
| 6.7  | Distribution histograms of choroidal measurements in the PREVENT cohort. . . . .  | 277 |
| 6.8  | Distribution of choroidal measurements for APOE4/FH carriers and non-carriers in the PREVENT cohort. . . . .                            | 278 |
| 6.9  | Distribution of choroidal measurements against age, sex and blood pressure in the PREVENT cohort. . . . .                               | 279 |
| 6.10 | Distribution of choroidal measurements between ordinal Dementia risk groups in the PREVENT cohort. . . . .                              | 280 |
| 6.11 | OCT FLEX Module in ITU and demonstrative acquisition. . . . .   | 290 |
| 6.12 | Description of choroidal measurements collected for the D-RISC-ii study. . . . .  | 292 |
| 6.13 | Sample derivation flowchart for the D-RISC-ii study. . . . .  | 296 |
| 6.14 | Patient-level OCT feasibility data for the D-RISC-ii study. . . . .   | 297 |
| 6.15 | Representative suprachoroidal space in patients in the D-RISC-ii study. . . . .   | 298 |
| 6.16 | Pairwise correlation analysis between choroidal measurements and clinical variables in the D-RISC-ii study. . . . .                     | 299 |
| 6.17 | Potential non-linear and monotonic association between choroidal measurements and fluid measurements in the D-RISC-ii study. . . . .    | 300 |
| 6.18 | Potential linear association between choroidal measurements and various clinical variables measurements in the D-RISC-ii study. . . . . | 301 |
| 6.19 | Observed longitudinal change in choroidal measurements and fluid in the D-RISC-ii study. . . . .  | 302 |
| 7.1  | Schematic of work conducted during research period. . . . .   | 308 |
| 7.2  | OCTolyzer's toolkit. . . . .  | 315 |
| S1   | Annotated OCT B-scan. . . . .   | 379 |
| S2   | Collection of annotated OCT B-scans outlining region and vessel landmarks. . . . .  | 380 |
| S3   | Example set-up for manual adjudication/segmentation. . . . .  | 381 |

## List of Figures

---

|    |   |     |
|----|---|-----|
| S4 | Exemplar OCT B-scans with varying image quality in choroidal space. . . . .       | 382 |
| S5 | Exemplar OCT B-scans with varying image quality in choroidal vasculature. . . . . | 384 |
| S6 | Example of over- and under-segmentation of choroidal vasculature. . . . .         | 385 |
| S7 | Screenshot of ITK-Snap for manual choroid segmentation. . . . .                   | 388 |
| S8 | Demonstration of manual choroid region segmentation. . . . .                      | 389 |
| S9 | Demonstration of manual choroid vessel segmentation. . . . .                      | 391 |

# List of Tables

|     |  |     |
|-----|--|-----|
| 1.1 | Previous semi- and fully-automatic approaches for choroid region segmentation in OCT. . . . .            | 34  |
| 1.2 | Previous semi- and fully-automatic approaches for choroid vessel segmentation in OCT. . . . .            | 36  |
| 1.3 | Previous semi- and fully-automatic approaches for choroid region and vessel segmentation in OCT. . . . . | 38  |
| 1.4 | Choroid measurement error between different ROI definitions. . . . .                                     | 52  |
| 1.5 | Description of choroid-derived and segmentation metrics used in this thesis. . . . .                     | 61  |
| 2.1 | Demographics of renal cohort used to validate GPET. . . . .  | 97  |
| 2.2 | Longitudinal sample information on renal cohort used to validate GPET. . . . .                           | 99  |
| 2.3 | GPET's population-level agreement against manual choroid thickness. . . . .                              | 102 |
| 2.4 | GPET's cohort- and location-level agreement with manual choroid thickness. . . . .                       | 103 |
| 2.5 | Blinded adjudication of major disagreements between manual and automatic choroid thickness. . . . .      | 105 |
| 3.1 | Demographic information on cohort used to evaluate MMCQ. . . . .   | 146 |
| 3.2 | Segmentation performance of MMCQ and Niblack against manual labels. . . . .                              | 148 |
| 3.3 | Execution times of MMCQ, Niblack and the manual graders. . . . .   | 149 |
| 3.4 | Demographics of samples used to assess the reproducibility of MMCQ and Niblack. . . . .                  | 156 |

## List of Tables

---

|      |   |     |
|------|---|-----|
| 3.5  | Reproducibility performance at the population-level for MMCQ and Niblack. . . . .                             | 159 |
| 4.1  | Cohort demographics of DeepGPET's modelling dataset. . . . .  | 179 |
| 4.2  | Demographics of the train, validation and test split for DeepGPET's modelling dataset. . . . .                | 179 |
| 4.3  | Cohort demographics for DeepGPET's reproducibility analysis. . . . .  | 181 |
| 4.4  | DeepGPET's modelling performance on GPET's ground truth labels. . . . .                                       | 185 |
| 4.5  | Blinded adjudication between DeepGPET and GPET for different cases of error. . . . .                          | 188 |
| 4.6  | DeepGPET's population-level reproducibility performance. . . . .  | 193 |
| 5.1  | Choroidalyzer's modelling dataset used for training and evaluation. . . . .                                   | 218 |
| 5.2  | Choroidalyzer's training, validation and testing dataset split. . . . .                                       | 219 |
| 5.3  | Choroidalyzer's reproducibility dataset. . . . .  | 220 |
| 5.4  | Choroidalyzer's segmentation performance across internal and external test sets. . . . .                      | 227 |
| 5.5  | Choroidalyzer's derived-measurement performance across internal and external test sets. . . . .               | 227 |
| 5.6  | Choroidalyzer's segmentation performance metrics for peripapillary B-scans. . . . .                           | 227 |
| 5.7  | Choroidalyzer's segmentation performance metrics against manual graders and state-of-the-art methods. . . . . | 231 |
| 5.8  | Choroidalyzer's derived-metric performance against manual graders and state-of-the-art methods. . . . .       | 232 |
| 5.9  | Choroidalyzer's execution time compared with manual graders and state-of-the-art methods. . . . .             | 232 |
| 5.10 | Choroidalyzer's manual adjudication against state-of-the-art methods. . . . .                                 | 233 |
| 5.11 | Choroidalyzer's population-level reproducibility agreement for derived metrics. . . . .                       | 239 |
| 6.1  | Baseline demographics and clinical variables for the renal cohort. . . . .                                    | 258 |
| 6.2  | Baseline demographics and clinical variables for the renal cohort. . . . .                                    | 258 |

---

|      |   |     |
|------|---|-----|
| 6.3  | Choroidal variation pre-transplant and one year post-transplant in renal cohort. . . . .                            | 260 |
| 6.4  | Association between choroidal measures and clinical variables in the renal cohort. . . . .                          | 260 |
| 6.5  | Linear mixed-effects modelling between the choroid and clinical variables in the renal cohort. . . . .              | 262 |
| 6.6  | Population demographics and clinical variables in the entire PREVENT cohort. . . . .                                | 275 |
| 6.7  | Population demographics and clinical variables in the PREVENT cohort eligible for analysis. . . . .                 | 276 |
| 6.8  | Distribution of choroidal measurements between ordinal Dementia risk groups in the PREVENT cohort. . . . .          | 277 |
| 6.9  | Ordinal logistic odds ratios from regressing choroidal measurements on Dementia risk in the PREVENT cohort. . . . . | 281 |
| 6.10 | Clinical variables collected for the D-RISC-ii study. . . . .   | 289 |
| 6.11 | Patient summary statistics for the D-RISC-ii study. . . . .   | 294 |
| 6.12 | Patient-level feasibility data for the D-RISC-ii study. . . . .   | 295 |

# List of Algorithms

- 1 Edge Tracing using Gaussian Process Regression . . . . . 96
- 2 Multi-scale Median Cut Quantisation . . . . . 145

# Chapter 1

## Introduction

The choroid is a crucial vascular layer at the back of the eye which sits posterior to the retina, effectively maintaining the visual system through nourishment of the photoreceptors [7]. The retina has often been the primary anatomical layer of interest in ophthalmology, while the choroid has seen neglect by virtue of the challenges faced by historical imaging methods. Previous methods used to image the retina relying on light reflection (laser doppler flowmetry), sound (ultrasonography) or fluorescence (fluorescein angiography) are hindered from imaging the choroid at high resolution by the pigment in the outer retinal pigment epithelium (RPE) and choroid [8]. Additional to the light absorption and back-scattering of the anterior retina, some of these approaches remain a challenge to collect repeated measurements of the choroid [8].

Optical coherence tomography (OCT) [9] has provided non-invasive access to cross-sectional retinochoroidal structures at the posterior pole since 1991. However, traditional time-domain (TD-OCT) and conventional spectral domain OCT (SD-OCT) are both affected by melanin and the scattering properties of blood vessels in the retina and choroid [10]. Improved visualisation of the choroid was introduced in the late 2000s with enhanced depth imaging OCT (EDI-OCT) [11] and later in 2012 with swept-source OCT (SS-OCT) which both permitted deeper penetration of ocular tissue over their traditional counterparts.

This enhanced appearance of the choroid has allowed many to study the



deeper vascular tissue of the eye in both retinal and systemic disease, conducting biomarker investigation with representative measurements like choroidal thickness [10]. However, approaches to measuring choroidal thickness vary significantly, both in terms of methodology [12, 13, 14, 15, 16, 17] and in terms of protocol [18, 19, 20, 21].

This has led to a deficit in the research community in accurately and reliably measuring the choroid between and within individuals [22, 23]. The ability to measure anatomy consistently at micron-scale is paramount to capturing the inherent biological change or signal being studied [24, 25] — a signal which can be corrupted by human subjectivity.

While the advent of Fourier domain EDI-OCT and SS-OCT has enabled improved visualisation of the choroid, current approaches to measure it have not kept pace with these advancements. This has led to the research community continuing to adopt traditional, semi-automatic image processing methods or manual grading in their studies [12, 26, 27, 28]. This is by virtue of their accessibility through mediums such as Fiji (ImageJ) [14, 29, 30] or availability on open-source platforms like GitHub [12, 26].

However, current methods are inadequate because of human subjectivity and improper usage and reporting of approaches to measuring the choroid. This has led to a concern around their overall reproducibility and accuracy. Additionally, more accurate and domain-specific approaches based on deep learning are often not openly available to the research community, leading to institutions using their own bespoke methods. Together, these contribute to a lack of standardisation within choroidal image analysis in OCT image sequences.

This absence of accurate and reliable methods for measuring the choroid drives a real need for more accurate approaches to choroid analysis which are open-source, easily accessible and rigorously validated. This would ultimately equip researchers with the ability to extract reproducible and clinically meaningful measurements of the choroid, thus improving standardisation in the field of choroidal image analysis in OCT image sequences.

In this thesis, we develop and evaluate several new approaches to measure the choroid from OCT image sequences. We first consider the applicability of tradi-

tional image processing techniques, which begins to address problems around reproducibility and human subjectivity. We then consider the utility of more advanced and domain-specific computer vision algorithms based on deep learning. In the latter, these approaches permit fully automatic, end-to-end choroidal image analysis in OCT image sequences, completely removing manual subjectivity from the measurement procedure. We show their ability to improve standardisation of choroidal image analysis by assessing their measurement error in detail.

Additionally, we release these tools to the wider research community as open-source to facilitate end-user accessibility. These tools are designed to not require proprietary software to use, nor require any domain-specific knowledge in image processing or computer science. We anticipate the accessibility, accuracy and reliability of these tools to help improve the standardisation of choroidal measurements in OCT image sequences. In addition to the development of these novel algorithms, we demonstrate their use-case on three novel applications of the choroid in systemic disease to show their wider applicability to the nascent field of choroidal image analysis in OCT image sequences.

## 1.1 Guiding principles

This thesis hopes to advance the field of choroidal image analysis in OCT through following a set of guiding principles:

1. **Measurement protocol:** Follow strict measurement protocol which accounts for pixel resolution, skew and curvature of intra-ocular structures, enabling improved measurement of the choroid in OCT image sequences between and within patients.
2. **Reproducibility:** Assess each method's measurement variability to help future end-users differentiate true biological change from measurement error in their studies. This will promote their reliability to produce clinically meaningful measurements of the choroid in OCT image sequences.

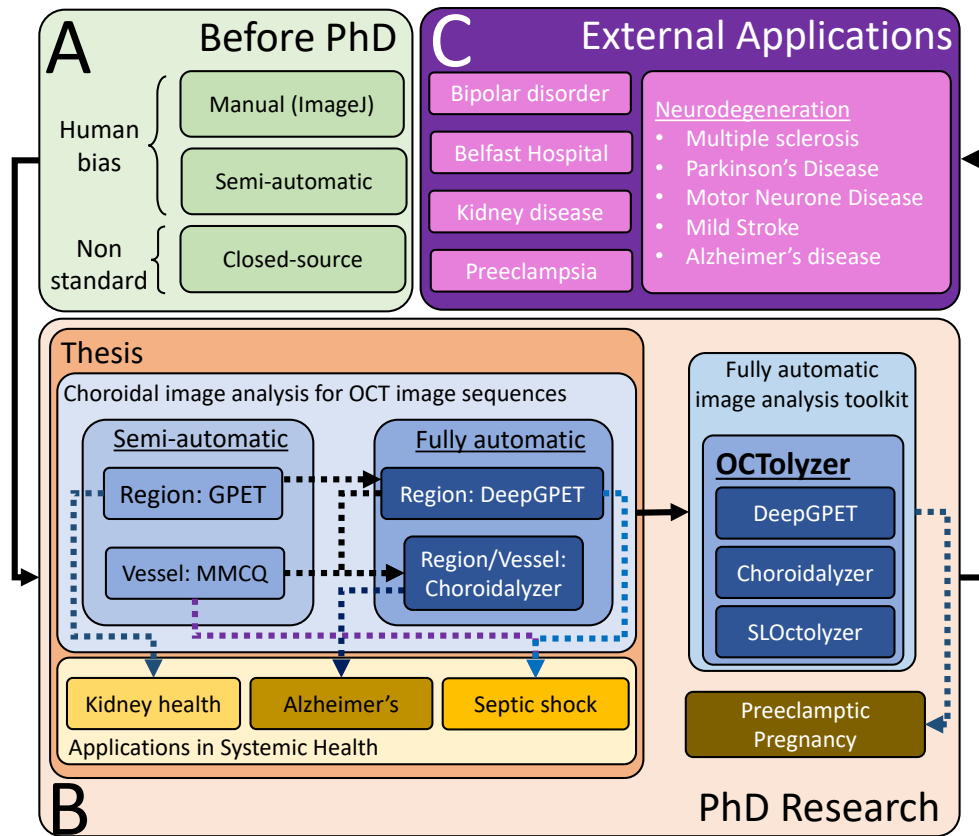
3. **Rigorous validation:** Conduct thorough validation of the new approaches developed in this thesis, both quantitatively and qualitatively, to transparently identify their strengths and limitations. This will communicate each method's applicability to choroidal image analysis in OCT image sequences.
4. **End-user accessibility:** Release these tools as open-source to ensure their accessibility to the broader research community. This will facilitate the adoption of these methods across varying levels of expertise, encouraging wider usage and standardisation of choroidal measurements in OCT image sequences.
5. **Application to systemic disease:** Demonstrate the potential of choroidal analysis in reflecting microvascular function, injury and response to systemic disease. The thesis will explore the emerging relevance of the choroid in this field and hope to encourage further research in this nascent area.

This thesis ultimately aims to contribute to the field of choroidal image analysis in OCT image sequences and its potential role in systemic disease. Through providing the research community with accessible and rigorously validated tools for extracting reproducible and clinically meaningful measurements of the choroid, the author hopes to help improve the standardisation of these measurements in the research community.

Finally, the author hopes this thesis will contribute to the nascent field of 'oculomics' by describing the potential for the eye to inform about systemic health. This will be accomplished through applying these novel methods in the context of renal transplant in chronic kidney disease, Alzheimer's disease risk and fluid resuscitation in multi-organ failure.

## 1.2 The structure of this thesis

This thesis is composed of an introductory chapter, five research chapters and a concluding discussions chapter. Chapter 1 introduces the choroidal



**Figure 1.1:** Schematic of the PhD research journey. (A) Current approaches to choroidal image analysis. (B) Work conducted during research period, with boxes and arrows identifying analysis methods from applications and their relationships. (C) Past, current and planned future applications of the methods developed from this thesis.

microvasculature, its anatomical composition and relationship with systemic disease. We also discuss how the choroid is imaged with OCT and how it is measured, with a brief overview on the current approaches used.

Research chapters 2 – 5 discuss, define and rigorously evaluate the accuracy and reproducibility of the automatic segmentation tools developed over the course of the research period. These research chapters introduce and evaluate Gaussian Process Edge Tracing (GPET, chapter 2) for semi-automatic choroid region segmentation; Multi-scale Median Cut Quantisation (MMCQ, chapter 3) for semi-automatic choroid vessel segmentation; DeepGPET (chapter 4), for fully automatic, deep learning based choroid region segmentation;

and Choroidalizer (chapter 5), for end-to-end, deep learning based choroidal region and vessel segmentation, alongside automatic fovea detection and choroid measurement.

The final research chapter, chapter 6, presents three applications of choroidal image analysis in OCT image sequences in systemic disease using the aforementioned tools. These include investigating choroidal changes in a longitudinal kidney donor and recipient cohort, estimating associations between choroidal microvasculature and Alzheimer's disease risk, and evaluating the feasibility of OCT imaging and choroidal variation for shock resuscitation in critically ill adults.

Finally, chapter 7 will discuss the key takeaways from this thesis and the core contributions this thesis provides to the research community. Moreover, this chapter will outline the current limitations and challenges faced by the work in this thesis. Lastly, this chapter will outline avenues of further work for choroidal image analysis in OCT image sequences. Figure 1.1 visualises the work contained in this thesis and over the PhD research period.

## 1.3 PhD outputs

### 1.3.1 Publication portfolio

Below lists publication outputs which were published and peer-reviewed by December 2024. Author of thesis in bold type, with lead authorship in underline type.

- **Burke, Jamie**, and Stuart King. “*Edge tracing using Gaussian process regression.*” IEEE Transactions on Image Processing 31 (2021): 138-148.
- **Burke, Jamie**, Dan Pugh, Tariq Farrah, Charlene Hamid, Emily Godden, Thomas J. MacGillivray, Neeraj Dhaun, J. Kenneth Baillie, Stuart King,

and Ian J.C. MacCormick. “*Evaluation of an automated choroid segmentation algorithm in a longitudinal kidney donor and recipient cohort.*” *Translational Vision Science & Technology* 12, no. 11 (2023): 19-19.

- **Burke, Jamie**, Neeraj Dhaun, Baljean Dhillon, Kyle J. Wilson, Nicholas A.V. Beare, and Ian J.C. MacCormick. “*The retinal contribution to the kidney–brain axis in severe malaria.*” *Trends in parasitology* 39, no. 6 (2023): 410-411.
- **Burke, Jamie**, Justin Engelmann, Charlene Hamid, Megan Reid-Schachter, Tom Pearson, Dan Pugh, Neeraj Dhaun, Amos Storkey, Stuart King, Thomas J. MacGillivray, Miguel O. Bernabeu, and Ian J.C. MacCormick. “*An Open-Source Deep Learning Algorithm for Efficient and Fully Automatic Analysis of the Choroid in Optical Coherence Tomography.*” *Translational Vision Science & Technology* 12, no. 11 (2023): 27-27.
- Engelmann, Justin, **Jamie Burke**, Charlene Hamid, Megan Reid-Schachter, Dan Pugh, Neeraj Dhaun, Diana Moukaddem, Lyle Gray, Paul McGraw, Amos Storkey, Niall Strang, Paul J. Steptoe, Stuart King, Thomas J. MacGillivray, Miguel O. Bernabeu, and Ian J.C. MacCormick. “*Choroidalyzer: An open-source, end-to-end pipeline for choroidal analysis in optical coherence tomography.*” *Investigative Ophthalmology & Visual Science* 65, no. 6 (2024): 6-6. (Mathematical and Computational Ophthalmology Special Issue).
- **Burke, Jamie**, Samuel Gibbon, Justin Engelmann, Adam Threlfall, Ylenia Giarratano, Charlene Hamid, Stuart King, Ian J.C. MacCormick, and Thomas J. MacGillivray. “*SLOctolyzer: Fully automatic analysis toolkit for segmentation and feature extracting in scanning laser ophthalmoscopy images.*” *Translational Vision Science & Technology* 13, no. 11 (2024): 7-7.

Below lists publication outputs which are either in peer-review, or prepared but currently not submitted by December 2024. Author of thesis in bold type, with shared lead authorship in underline type.

- **Burke, Jamie**, Ashley Akbari, Rowena Bailey, Kevin Fasusi, Ronan A.

Lyons, Jonathan Pearson, James Rafferty, and Daniel Schofield. “*Representing multimorbid disease progressions using directed hypergraphs.*” medRxiv (2023): 2023-08. Submitted and in peer-review at IEEE Journal of Biomedical and Health Informatics.

- **Burke, Jamie**, Justin Engelmann, Charlene Hamid, Diana Moukaddem, Dan Pugh, Neeraj Dhaun, Amos Storkey, Niall Strang, Stuart King, Thomas J. MacGillivray, Miguel O. Bernabeu, and Ian J.C. MacCormick. “*Domain-specific augmentations with resolution agnostic self-attention mechanism improves choroid segmentation in optical coherence tomography images.*” arXiv preprint arXiv:2405.14453 (2024). Manuscript prepared but currently not submitted as of December 2024.
- **Burke, Jamie**, Justin Engelmann, Charlene Hamid, Diana Moukaddem, Dan Pugh, Niall Strang, Neeraj Dhaun, Thomas J. MacGillivray, Stuart King, Ian J.C. MacCormick “*OCTolyzer: Fully automatic toolkit for segmentation and feature extraction in optical coherence tomography and scanning laser ophthalmoscopy data.*” arXiv preprint arXiv:2407.14128 (2024). Submitted and in peer-review at Elsevier Medical Image Analysis.
- **Burke, Jamie**, Samuel Gibbon, Audrey Low, Charlene Hamid, Megan Reid-Schachter, Graciela Muniz-Terrera, Craig W. Ritchie, Baljean Dhillon, John T. O’Brien, Stuart King, Ian J.C. MacCormick and Thomas J. MacGillivray “*Association between choroidal microvasculature in the eye and Alzheimer’s disease risk in cognitively healthy midlife adults.*” medRxiv 2024.08.27.24312649 (2024). Submitted and in peer-review at Alzheimer’s and Dementia: Diagnosis, Assessment and Disease Monitoring.
- George Cooper, **Burke, Jamie**, Charlene Hamid, Emily Godden, Neeraj Dhaun, Stuart King, Tom MacGillivray, J. Kenneth Baillie, David M. Griffith and Ian J.C. MacCormick. “*D-RISC II: Direct Retinal Imaging for Shock Resuscitation in Critical Ill Adults II.*” Manuscript prepared but currently not submitted as of December 2024.

### 1.3.2 Technical portfolio

- Primary owner / maintainer:
  - Semi-automatic choroid region segmentation using Gaussian process regression,  
GPET: [https://github.com/jaburkel66/gaussian\\_process\\_edge\\_trace](https://github.com/jaburkel66/gaussian_process_edge_trace)
  - Semi-automatic choroid vessel segmentation using multi-scale median cut quantisation,  
MMCQ: <https://github.com/jaburkel66/mmcq>
  - Fully automatic choroid region segmentation in optical coherence tomography data,  
DeepGPET: <https://github.com/jaburkel66/deepgppe>
  - Fully automatic pipeline for segmentation and feature extraction in scanning laser ophthalmoscopy images,  
SLOctolyzer: <https://github.com/jaburkel66/SLOctolyzer>
  - Fully automatic pipeline for segmentation and feature extraction in optical coherence tomography and scanning laser ophthalmoscopy data,  
OCTolyzer: <https://github.com/jaburkel66/OCTolyzer>
- Contributed significantly to:
  - Fully automatic, end-to-end choroid region and vessel analysis in optical coherence tomography data,  
Choroidalalyzer: <https://github.com/justinengelmann/choroidalalyzer>
  - Directed hypergraphs for longitudinal multimorbidity data,  
Hypergraphs-mm: <https://github.com/nhsx/hypergraph-mm>



## 1.4 The choroidal microvasculature

The choroid is a highly vascularised structure of the eye located between the retina and the sclera [7], and plays a crucial role in the physiology of the eye. It is one of the most vascularised tissues in the body [31], citing the highest blood flow per unit tissue weight over any other [7].

The uveal circulation branches from the ophthalmic artery and enters the eye external to the optic nerve, receiving 95% of the blood flow directed to the eye [8], splitting into anterior and posterior segments. The anterior segment reaches the iris and ciliary body, while the posterior segment, or posterior uveal tract, supplies the choroid and accounts for more than 70% of the entire uveal blood flow [32].

Embryologically, the eye first forms as an outpocketing of the forebrain, called the optic vesicle [33]. This undergoes invagination to form the retina and the RPE, each of which have their own dedicated circulations [8], the latter being supplied by the choroid eventually. The choroid begins to form around six weeks of gestation as the choriocapillaris [8], located posterior to the RPE. By eight weeks, the medial and lateral short posterior ciliary arteries (PCAs), which stem from the posterior uveal tract [34], enter the choroid and branch into smaller arterioles toward the choriocapillaris [8]. The choriocapillaris are drained exclusively by vortex veins, in the absence of venous equivalents to the posterior ciliary arteries [35].

Anatomically, the choroid is anchored posterior to the retina facilitating sub-retinal fluid exchange between the RPE and choriocapillaris and is connected laterally to the optic nerve. The choroid is attached anteriorly to the sclera by loose connective tissue which creates a potential space, the suprachoroid [8]. This potential space is traversed by vessels and nerves, and ranges in thickness from 10 to 35 microns [36]. However, this is often not visible using current imaging methods, with a thin suprachoroid reported to be visible in up to 44% of healthy people (74 eyes) aged 55 – 85 in EDI-OCT [18] and a similar proportion (from 200 eyes) in SS-OCT [37]. The inconsistency in the visibility of the suprachoroid in OCT when measuring the choroid can lead to poor agreement [18, 37], and so for the purposes of standardisation we consider

the suprachoroid to be a separate entity to the choroidal vascular layer.

The choroid, bound between the suprachoroid and Bruch's membrane, is structurally organised into three vascular layers commonly based on vessel size: the outer layer, known as Haller's layer, consists of large blood vessels which are sourced from the posterior uveal tract. These larger vessels branch into smaller vessels in the middle segment, known as Sattler's layer which then form capillary networks at Bruch's membrane, known as the choriocapillaris [8].

The choriocapillaris is a dense, anastomosed network of fenestrated vessels organised in a honeycomb-like, lobular pattern [38], forming a thin sheet at Bruch's membrane and is critical for the exchange of nutrients and waste with the RPE and outer retinal layers. The choriocapillaris layer has the highest capillary density in the body [39] and is most dense and thickest under the fovea, with an approximate depth of  $10\mu\text{m}$  [7]. Individual capillaries are densely packed together with broad and flat lumens of approximately  $20 - 50\mu\text{m}$  in diameter [38]. This means the choroidal capillaries have low vascular resistance, thus facilitating high choroidal blood flow [40], to enable sufficient exchange with the RPE [10].

The choroid is the major source of oxygen to the photoreceptors, acting as the only source for the fovea — where inner retinal vessels are absent [7]. The nature of choroidal blood flow is exceptional, with a blood flow rate approximately 20 to 30 times greater than that of the retina [41]. This immense flow rate is more than sufficient — but not necessary — to meet the energy demands of the photoreceptors, which are heavily reliant on the choroid for oxygen and nutrient supply at all times [31]. Any significant reduction in choroidal blood flow can lead to rapid photoreceptor degeneration, contributing to conditions like Age-Related Macular Degeneration (AMD) [7].

Moreover, The structural and functional integrity of the retinal and choroidal vasculatures is essential for normal retinal function [42], and deficits in choroidal perfusion cannot be maintained by the retina and vice versa [42]. Unlike the highly permeable and freely anastomosing choriocapillaris, the choroid has a segmental blood distribution network [43], delivering oxygen to localised, avas-

cular zones of the retina [44]. This leaves strips known as watershed zones between the various PCAs, the short PCAs, the choroidal arteries, the arterioles, and the vortex veins [45]. This can lead to localised areas of filling defects and lower vascular perfusion which is thought to contribute to AMD or choroidal neovascularisation [43].

Relative to the choroidal circulation, the retinal circulation is characterised by low flow and high oxygen extraction (approximate arteriovenous difference in partial pressure) of approximately 40% [41]. Conversely, choroidal blood flow has markedly higher flow rate with relatively low oxygen extraction of 3 – 4% [40]. The high blood flow is thus more than sufficient but not necessary to nourish the high metabolic demands of the photoreceptors [40], which have the highest rate of oxygen use per unit tissue of any tissue in the body [46]. This suggests that alongside support for the photoreceptor metabolic activity and continuous outer segment renewal [31], the excessive flow characteristics of the choroidal vasculature have other functions [40], including potentially thermoregulation to cool the retina [32] and compensation for lens focusing during accommodation [7]. However, some believe the high blood flow is necessary to promote diffusion of oxygen and nutrients through the outer blood-retina barrier [8, 31].

The retinal vasculature is thought to have efficient autoregulatory functions [42], an intrinsic property of many vascular tissues in the body which maintains perfusion pressure (constant blood flow) despite changes in systemic blood pressure and/or intracranial or intraocular pressure [47]. For the retina, blood flow is maintained across fluctuations in ocular perfusion pressure (the difference between mean arterial pressure and intra-ocular pressure) through various myogenic and metabolic functions [42]. This compensation constitutes to changes in retinal vessel calibre (vasodilation/vasoconstriction), which maintains the oxygen partial pressure at a low and normal physiological level [47].

Interestingly, choroidal blood flow is not entirely autoregulated using traditional myogenic responses like the retina [7]. Instead, blood pressure management of choroidal vessels and response to retinal activity in the choriocapillaris are directly and indirectly under control by intrinsic choroidal neurons, respectively [31]. The choroid is well endowed with several choroidal nerve fibers (neurons),

mostly situated at the central retina [7]. These neurons innervate (regulate) choroidal blood flow through vasodilatory (from parasympathetic and sensory nerve fibers) and vasoconstrictive influence (from sympathetic nerve fibers) [31]. Although, compensation for changes in ocular perfusion pressure is considered to be influenced by local, autoregulatory myogenic mechanisms [8].

### 1.4.1 Systemic disease

Examination of the choroid has become increasingly popular with the technological advancements in non-invasive OCT imaging devices. Conventional SD-OCT led to the primary focus of OCT research on the cross-sectional retinal layers, with little attention paid to the choroid due to limited visibility. However, the advent of EDI-OCT [11] and SS-OCT [48] have significantly improved imaging of choroidal structures. Improved, non-invasive imaging of the choroid has in turn contributed to the nascent field of 'oculomics' [49], where the role of the choroid as a potential biomarker in numerous systemic diseases has been investigated [50].

The choroid's poor autoregulation results in choroidal fluctuations corresponding to changes in intra-ocular pressure, which itself fluctuates in response to systemic blood pressure due to diurnal variation [51]. The extent of diurnal variation in the choroid has been investigated previously [52, 53, 54, 55, 56, 57], reporting choroidal fluctuations of approximately 30  $\mu\text{m}$  over the course of the day, with the majority of research studies reporting larger choroid thickness measurements in the morning, compared to the afternoon and evening [51].

Additionally, the choroid's poor autoregulation makes it more susceptible to systemic changes in blood pressure. This has helped motivate investigations into conditions affecting systemic blood pressure, such as hypertension [58] and chronic kidney disease [59]. Additionally, the eye and the kidney show a striking resemblance in anatomy, physiology and response to disease [60], and in particular the choroidal microcirculation seemingly reflects the renal microcirculation [61]. The endothelial cells of the choriocapillaris and renal cortex have similarly sized fenestrated vessel walls allowing subretinal fluid exchange in the choroid and blood filtration in the kidneys [61]. Moreover, the circulatory

systems in the choroid and renal cortex have similar proportions of blood flow in contrast to their retinal and renal medullary counterparts [7], with choroidal blood flow four-fold higher per unit mass than the kidney [61].

Consequently, there has been interest and recent evidence of the choroid reflecting renal function. Choroidal changes have been linked to chronic kidney disease (CKD), with patients often exhibiting choroidal thinning correlating with severity of kidney dysfunction [16, 62]. This relationship highlights the potential of choroidal imaging as a non-invasive method to monitor microvascular injury in CKD, as shown by Balmforth, et al. [59], who suggested that an overactive sympathetic nervous system and systemic endothelial dysfunction were potential causes to choroidal thinning in pre-dialysis CKD. Additionally, in patients undergoing haemodialysis, the choroid can show significant fluctuations in thickness [63, 64], possibly due to changes in fluid balance [65], offering insights into the impact of dialysis on systemic and ocular health.

Emerging evidence suggests potential links between the eye and neurodegeneration, such as in Alzheimer's disease [66], Parkinson's disease [67], multiple sclerosis [68] and cerebral small vessel disease [69]. For instance, Alzheimer's disease patients or those with mild cognitive impairment often exhibit thinning of the choroid [14, 70], which could be linked to hypoperfusion and atrophy. However, the research community acknowledges the need for more methods to measure the choroid [23, 68], as paradoxical results have also been found in similar research contexts [71, 72]. Although, the choroid's response in the early prodromal stages of Alzheimer's disease has not yet been elucidated, and this will be discussed in section 6.2.

Relative to invasive imaging of the cerebrovascular network [73], which is limited by resolution [74], non-invasive OCT imaging of the intra-ocular structures enables consistent and reproducible sampling of a microvascular site which is part of the central nervous system and may mirror cerebral hypoperfusion and microvascular dysfunction [75]. This has contributed to retinal biomarker investigation in systemic inflammatory conditions like septic shock [75, 76], which affects microvascular flow and permeability. Recent pre-clinical evidence [77] in a septic rat model reported reduced choroidal blood flow during shock, and other related studies have demonstrated feasibility of retinal imaging and mea-

surement in OCT and OCT-Angiography (OCT-A) in a critical care environment [78, 79].

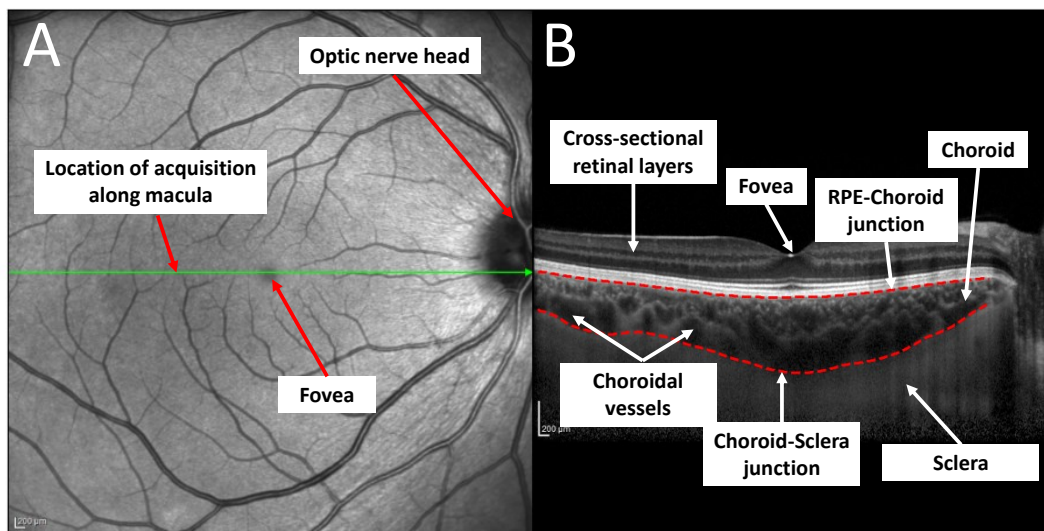
The choroid's utility in systemic disease will continue to grow with advancements in the technology and accessibility of OCT [80, 81]. Ultimately, with sufficient imaging technology and standardised analysis tools, the choroid has the potential to provide a valuable window into systemic microvascular health and the health of vital organs like the kidney and the brain.

### 1.5 Optical coherence tomography

Previous methods for imaging/measuring the choroid, which sits posterior to the hyperreflective RPE, have used reflectance, sound and fluorescence. Light reflectance based on laser Doppler flowmetry [82] suffers from its limited representation, through only measuring localised choroidal blood flow at the choriocapillaris, typically underneath the fovea [8], which is not representative across the posterior pole [44, 83]. Ultrasonography use sound waves and vibrations to detect ocular tissue based on sound reflectance [84], but typically have very low resolution and poor localisation which can result in less precise choroidal measurements [8].

Indocyanine green angiography is more accurate for en face choroidal perfusion measurements in the early angiographic sequence (dye fill time) [85], fluorescing over reasonably long wavelengths which can penetrate the choroid (790 – 805 nm) compared to fluorescein angiography [8]. However, this method is limited by its invasive methodology [10], light absorption and scattering by the pigment in the RPE and choroid, and blood flow in the choroid [8]. Other methods have been considered such as ultrasonic colour Doppler imaging (combining ultrasonography with Doppler flowmetry) [86] and laser interferometry (measuring perfusion based on the distance between the cornea and retina during the cardiac cycle) [87]. However, these methods also face the aforementioned challenges around localisation and reproducibility [10].

It has been established that optical imaging techniques help improve the resolution of imaging the choroid [8], and optical coherence tomography, or OCT,



**Figure 1.2:** Typical horizontal-line, OCT scan from a Heidelberg Engineering OCT imaging device. En face localiser SLO image (A) and corresponding OCT B-scan (B), with annotations overlaid of major anatomical landmarks.

has become the gold standard for clinical evaluation and imaging of the chorioretinal structures [10, 88]. OCT was first introduced in 1991 [9] and uses low coherence interferometry to construct a three-dimensional map of the intra-ocular structures at the posterior pole.

The general principle is that a tightly focused light beam is scanned across the posterior pole and measurements on back-scattered light which reflects from ocular tissue at various depths can then be processed. Interferometry leverages the phase difference between different light beams to infer reflectivity of the ocular tissue, otherwise known as *interference*. This is done first in the axial direction (A-scan) at a single lateral position, and iterated over in the transverse direction to create a two-dimensional cross-sectional image of the chorioretinal structures (B-scan). Doing this across the macula along the frontal plane (the final perpendicular axis to the transverse and axial direction) will enable imaging of three-dimensional intra-ocular structures [88].

Figure 1.2 shows a standard OCT B-scan capture in panel (B), with the corresponding *localiser* en face image in panel (A). Landmarks are annotated on both images, and in particular the green horizontal line in panel (A) cor-



responds to the cross-sectional location of the B-scan on the macula. The localiser image is an en face scan of the retinal vasculature which can vary depending on the OCT system. For most datasets used in this thesis, we adopt the Heidelberg Engineering OCT system which leverages confocal imaging methods to capture an infra-red scanning laser ophthalmoscopy (SLO) image. The primary purpose of this image is to fixate the OCT beam at the correct location in the posterior pole (the green horizontal line in figure 1.2(A)).

### 1.5.1 Time-domain OCT

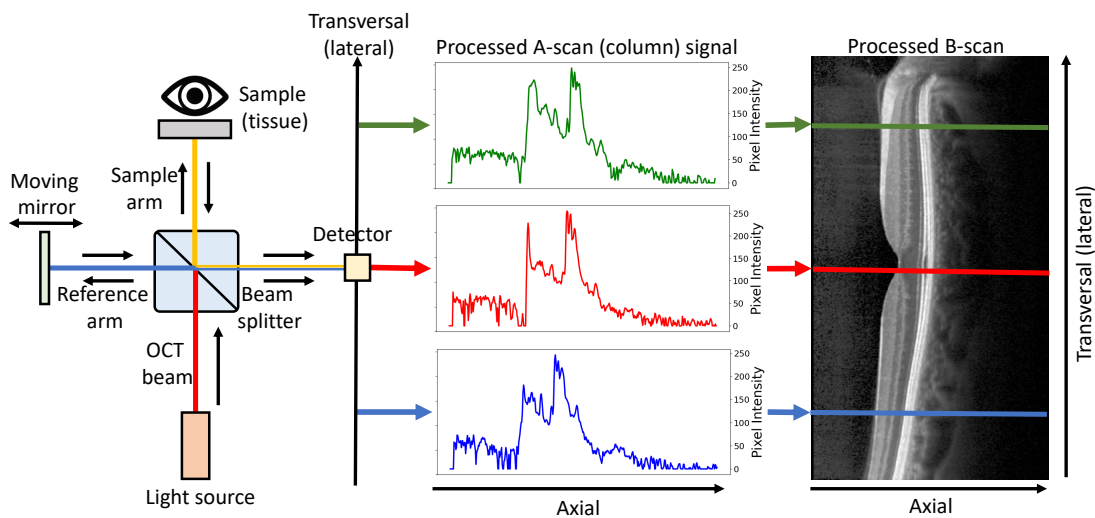
Figure 1.3 shows a schematic of the conceptually simplest approach to OCT, time-domain OCT (TD-OCT). TD-OCT works using low coherence light, i.e. light waves which have a short coherence length and can only interfere with each other over very short distances [89]. This interference is leveraged in OCT to create precise depth measurement with high axial resolution.

In the OCT setup in figure 1.3, a single light source (superluminescent diode) emits a beam of broadband light with a short coherence length which is split into identical beams along two separate arms using a beam splitter. One beam penetrates the ocular tissue along the sample arm and back-scatters features of depth and reflectivity to the beam splitter, while the other beam moves along the reference arm and reflects off a mirror which is moving in the axial direction.

Because we use low coherence light, each position of the moving mirror generates a reference beam which interferes with one point in the sample beam to produce a signal. This location along the sample beam is equal to the optical length of the reference arm and corresponds to an axial depth in the ocular tissue. It is this optical path length and resulting interference between the reference and sample beams which creates the axial position and reflectance (pixel intensity) for a single pixel in an OCT B-scan [90].

Varying the reference arm in length, i.e. the axial position of the moving mirror, has the effect of generating a reflectivity depth profile for a specific transverse (lateral) position. This is known as an A-scan, and represents a whole column (in the axial direction) of an OCT B-scan. By scanning a reflectiv-





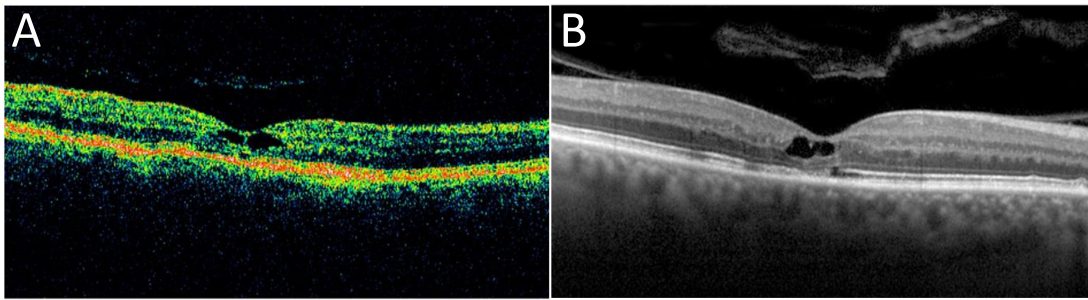
**Figure 1.3:** Schematic of how TD-OCT works to generate an OCT B-scan. Note that this B-scan is from an SD-OCT device and is purely demonstrative of how the system works in general.

ity depth profile (A-scan) along the transverse axis for the same sample, a two-dimensional OCT B-scan is generated by stitching the corresponding A-scans together. Finally, by repeating this process along the sagittal direction (vertically), three-dimensional OCT volume data (sequential and parallel OCT B-scans) is collected of the intra-ocular structures at the posterior pole.

While TD-OCT is conceptually simple and easy to manufacture [8], it has some limitations. To create a good quality image requires significant sampling time as the OCT beam goes through the full depth during the sampling of each axial location per lateral position. However, this is limited due to safety standards, thus lowering the overall signal-to-noise ratio (SNR) [8]. The slower acquisition speed (400 A-scans/second) [19], light absorption from the RPE and limited axial resolution ( $\approx 10 \mu\text{m}$ ) [91] prevent sufficient choroid visualisation.

### 1.5.2 Spectral-domain OCT

A significant advance in OCT technology was removing the need for a moving reference arm and being able to collect reflectivity depth profiles (A-scans) in a single pass by working in the Fourier domain. This is called spectral do-



**Figure 1.4:** Comparison of TD-OCT and SD-OCT for a single OCT B-scan of the same eye and approximately the same anatomical location. Reprinted with permission from Christopher Mody [92].

main OCT (SD-OCT) [93]. Low coherence light is split along the reference and sample arms, but the reference arm is fixed in SD-OCT image acquisition (the mirror in figure 1.3 does not move anymore). For each A-scan, the reference and sample beams are combined at the beam-splitter to create an interference pattern, as is typical in TD-OCT. However, in SD-OCT this pattern is passed through a diffraction grating which disperses the signal into individual frequency components [94] (the detector in figure 1.3 is now a spectrometer with a diffraction grating). While TD-OCT encodes the interference pattern (and depth of ocular tissue) as a function of the reference arm distance, SD-OCT encodes the interference pattern (and corresponding depth) as a function of frequency (or wavelength).

This spectral data encodes depth and reflectivity information of the ocular tissue. In particular, each frequency (or wavelength) in the spectra corresponds to a particular depth of the ocular tissue back-scattered from the sample beam, and is associated with a specific strength (or power). A charge-coupled device (spectrometer) decodes this spectral data using a Fourier transform, resolving and reconstructing a depth reflectivity profile, which is the A-scan. Once A-scans have been collected along the transversal axis at multiple lateral positions, the B-scan can be generated to create a two-dimensional OCT B-scan. Figure 1.4 shows the difference in image quality and signal between the same eye using TD-OCT in panel (A) and using SD-OCT in panel (B) at the same anatomical location.

Advantageously, SD-OCT generates an A-scan in a single pass, rather than

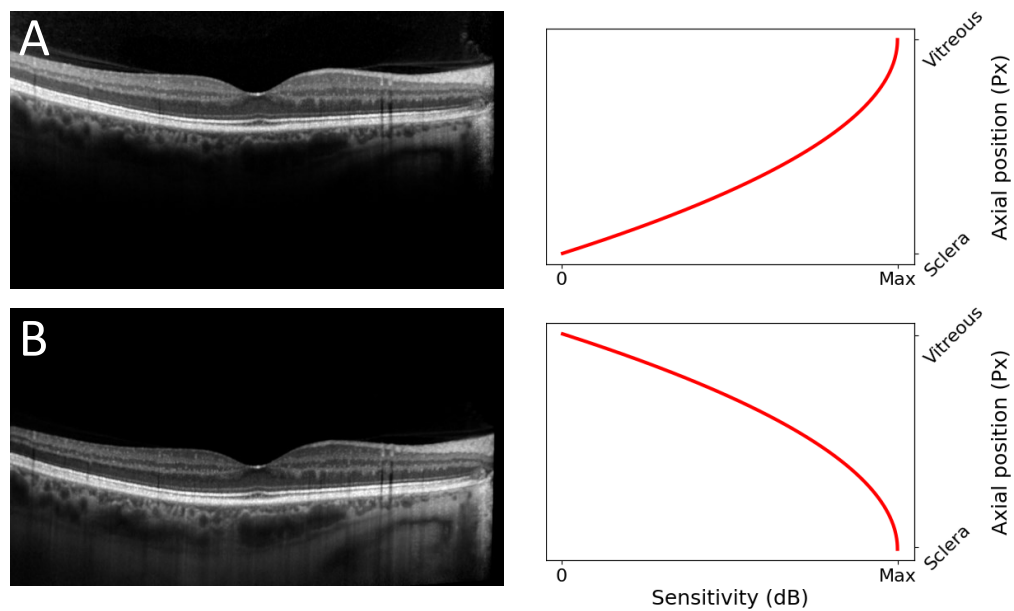
requiring multiple sampling points like in TD-OCT. Therefore, SD-OCT is much more efficient and has a significantly higher scanning rate between 20,000 – 52,000 A-scans/second with a higher axial resolution of around 5 – 7  $\mu\text{m}$  [95]. However, while SD-OCT utilise infra-red wavelengths up to 850nm for greater penetration of the retinal and subretinal space [8], light scattering from the RPE and the vascular nature of the choroid prevent penetration to the Choroid-Sclera boundary often [11]. Thus, conventional SD-OCT struggles to visualise the choroid and Choroid-Sclera boundary well, as shown in figure 1.4(B).

Additionally, SD-OCT suffers from the sensitivity ‘roll-off’ [10], where deeper ocular structures tend to have poorer SNR. This sensitivity roll-off is in part due to the sampling of the spectrometer once it detects the interferometric signal, after it’s been passed through the grating [11]. The sampling is non-linear in the axial direction (as each A-scan forms), resulting in lower resolution (and signal) with greater depth. Additionally, increasing depth is associated with increasing frequency in the spectral data, and the limitations of the spectrometer prevents sampling of the highest frequency responses, resulting in a loss of measured signal intensity with greater depth [11].

### 1.5.3 Enhanced depth imaging OCT

The axial position (ocular tissue depth) which has the highest sensitivity (or signal) is called the ‘zero-delay line’, and corresponds to the ocular tissue depth which equals the length of the reference arm [8]. The sensitivity, and thus signal, decreases as a function of distance away from this depth, as these represent higher frequencies which are harder to sample. In conventional SD-OCT the zero-delay line is at the vitreoretinal interface, which is sensible since the vitreous is almost transparent and requires high sensitivity for good imaging [8]. The trade-off is that this produces poor quality choroid visualisation, as exemplified in figure 1.4(B).

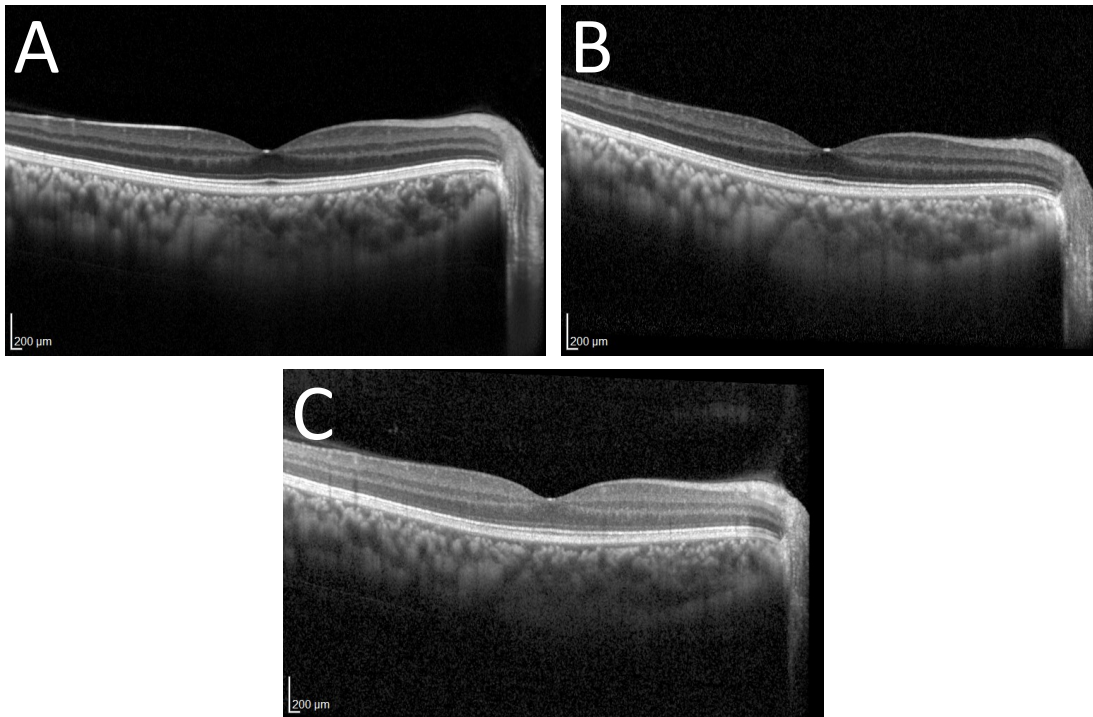
To mitigate for the sensitivity roll-off at the depth of the choroid, SD-OCT can leverage the conjugate property of the Fourier transform and alter the position of this zero-delay line. These adaptations to conventional SD-OCT give rise to enhanced depth imaging (EDI-OCT) [11].



**Figure 1.5:** Comparison of an SD-OCT B-scan without EDI mode (A) and with EDI mode (B). Beside each B-scan plots the approximate signal sensitivity as a function of axial distance (from sclera to vitreous).

A consequence of taking the Fourier transform of the interferometric signal is that it gives rise to the real and conjugate A-scan, which is referred to as the real and inverted A-scan [11] (or B-scan after stitching the A-scans along the transverse axis). In practice, the real B-scan is used with peak sensitivity at the vitreoretinal interface, which corresponds to a sensitivity profile which decreases with greater depth. If the camera head is positioned nearer to the eye, this shifts the zero-delay line toward the inner sclera. Here, the inverted B-scan has an inverted sensitivity profile to that of the real B-scan, such that signal strength *increases* with greater depth (toward the zero-delay line, i.e. inner sclera) [8].

This has the effect of enhancing deeper structures such as the choroid, as shown in figure 1.14. In panel (A), conventional SD-OCT captured a B-scan (using the ‘real’ A-scan signals) whose sensitivity profile is shown to the right, and decreases non-linearly with increasing depth through the eye. This, as well as the light scattering properties of the choroid, prevent sufficient visualisation. In panel (B) we show an EDI-OCT B-scan. For this acquisition, the camera



**Figure 1.6:** Different OCT B-scans of the same eye using the Heidelberg Engineering Spectralis OCT1 module. (A) EDI mode with an ART of 100. (B) EDI mode with an ART of 25. Conventional SD-OCT, known as non-EDI OCT in this thesis, with an ART of 9.

head is moved closer, and the 'inverted' B-scan is used as the output, which shows an inverted sensitivity profile to the right.

Heidelberg Engineering imaging devices now have EDI-OCT as an explicit option during image capture, by selecting the 'EDI' button. In addition to EDI as an option, Heidelberg Engineering also have active eye tracking (TruTrack, Heidelberg Engineering, Heidelberg, Germany) of the pupil during image capture [11]. This can allow for multiple (up to 100) B-scans to be captured during the same acquisition and averaged over to reduce speckle noise, which is called Automatic Real Time (ART) averaging. Figure 1.6 shows an OCT B-scan of the same eye with decreasing image quality, where EDI mode with an ART of 100 provides the most enhanced visualisation of the choroidal vessels and Choroid-Sclera boundary.

### 1.5.4 Swept-source OCT

A further advancement in Fourier domain OCT was the introduction of swept-source OCT (SS-OCT), which utilise longer wavelengths for greater depth penetration. Additionally, through the use of a wavelength-swept laser and photodetector hardware [48], SS-OCT is able to mitigate for the sensitivity roll-off from which SD-OCT suffers [10]. In figure 1.3, the OCT beam source emits a sweeping light source across a range of wavelengths and the detector is now a simple photodiode like in TD-OCT, rather than a broadband light source and spectrometer as in SD-OCT.

In SS-OCT image capture, a tunable laser is swept over the interference signal between the reference and sample arms across a range of wavelengths and in an ordinal fashion. Through the use of a simple photodiode, the laser's output has a much higher throughput, processing A-scan reflectivity depths much more efficiently [8] and without the need for a grating or spectrometer [96]. This resulted in improved acquisition speed of up to 4x when it was introduced in 2012, relative to commercial SD-OCT systems at the time [97].

In today's commercial market, Topcon's commercial SS-OCT DRI Triton reports a superior A-scan rate of 100,000 A-scans/second [98] over Heidelberg Engineering's Spectralis OCT2 module's 85,000 A-scans/second [99]. Nevertheless, some commercial SD-OCT systems are still rivalling SS-OCT systems, with Optopol's commercially available NX Revo-130 reporting scan rates of 130,000 A-scans/second [100]. However, it is thought that SS-OCT systems have potential for enabling OCT imaging speeds beyond 400,000 A-scans/second and even reaching speeds up to 1,000,000 A-scans/second [90].

While acquisition rate is important for overcoming barriers to patient focus and concentration, especially in the context of disease, it is not the only metric on which to evaluate OCT systems. In the particular application of choroidal image analysis, deep, subretinal tissue penetration and signal strength are key. Interestingly, there exists a direct trade-off between acquisition speed and signal strength [101]. Signal strength is positively associated with *integration time* [97], the length of time it takes to capture an A-scan, and is thus inversely associated with acquisition rate. However, modern systems can counteract

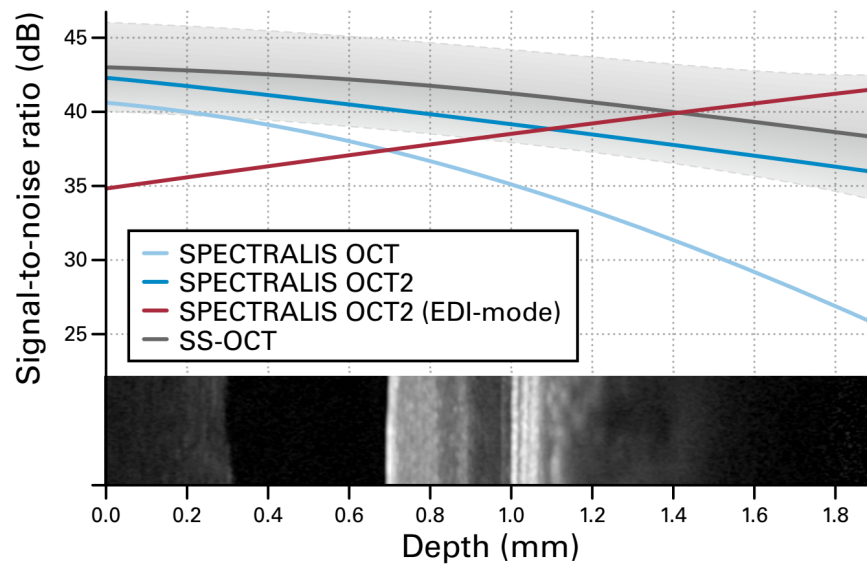


higher acquisition rate (and thus low signal strength) using post-acquisition image processing methods such as B-scan averaging (ART) [11].

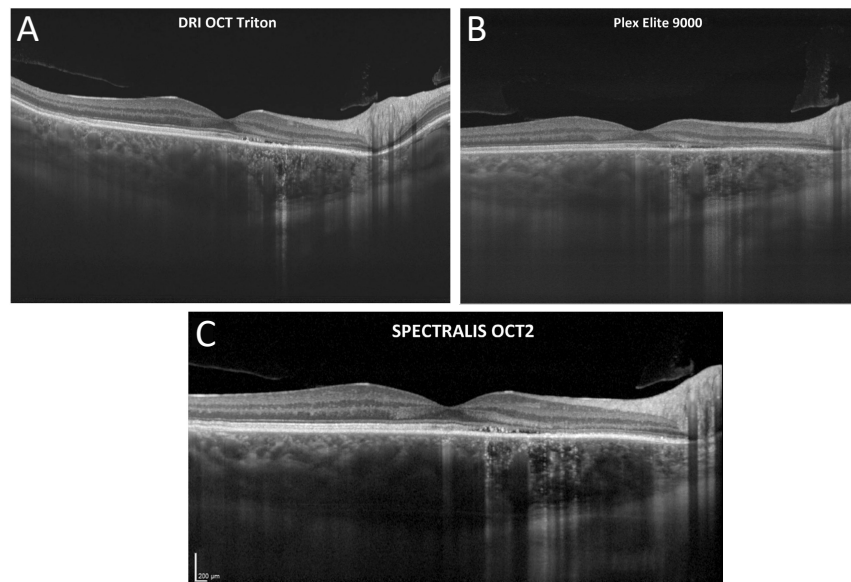
SS-OCT can provide greater depth into the ocular tissue due to the longer coherence length of the wavelength-swept laser, relative to the shorter coherence length of the broadband light source of SD-OCT systems. This has a direct impact on the sensitivity roll-off, which SD-OCT suffers from more so than SS-OCT [8]. Additionally, the use of a wavelength-swept laser enables a broader range of wavelengths up to 1050nm (with an approximate bandwidth of 100nm [96]) which allow for both the vitreous and choroid to be imaged simultaneously, rather than having to prioritise one as in SD-OCT and EDI-OCT [8]. Ultimately, this allows for greater axial depth penetration in scattering tissues relative to shorter wavelength systems such as SD-OCT [97].

However, the longer coherence length of the swept-source laser negatively impacts axial resolution, which is partially compensated for by the higher bandwidth of 100nm (compared to SD-OCT systems which usually have a bandwidth of 60nm around the central 850nm wavelength). Nevertheless, this leads to commercial SD-OCT systems having better optical in-depth resolution (7  $\mu\text{m}$  for Heidelberg Engineering's Spectralis standard OCT module [102] compared with 8  $\mu\text{m}$  for Topcon's DRI Triton [98]). Additionally, although longer wavelengths penetrate farther, they also suffer from greater water absorption which can lead to light attenuation [97], as the beam has to penetrate through the vitreous which is primarily water [8]. This ultimately has an impact on the signal strength of SS-OCT systems.

Signal strength is often measured as the sensitivity of the OCT system or SNR. Teussink, et al. [97] state that "SNR is determined by the signal of a specific sample, how much light reaches the sample, how much light is reflected, and how much this light is attenuated before reaching the detector". In their work, they estimate that longer wavelengths of 1050nm (with a bandwidth of 100nm) only lead to a factor increase in SNR of approximately 1.7, relative to an 850nm wavelength SD-OCT system (with a bandwidth of 60nm). This was observed even with a hypothetically higher input power (at the maximum safety limits) of the wavelength-swept laser source (meaning that more light would be reaching the sample) [97].



**Figure 1.7:** SNR (sensitivity) roll-off curves (against ocular tissue depth) for SD-OCT Spectralis OCT, Spectralis OCT2 and Spectralis OCT2 with EDI-mode, compared to a hypothetical SS-OCT device. Reprinted from Teussink, et al. [97] with permission from Heidelberg Engineering GmbH.



**Figure 1.8:** Comparison of visibility of the Choroid-Sclera boundary in a patient with central serous chorioretinopathy, acquired using the Topcon DRI Triton (A), Zeis Plex Elite 9000 (B) and Heidelberg Engineering OCT2 Module with EDI mode (C). Note that panels (B) and (C) use an ART of 100. Reprinted from Teussink, et al. [97] with permission from Heidelberg Engineering GmbH.



For choroidal image analysis in OCT image sequences, maximal SNR at greater depth is optimal. As the sensitivity roll-off is improved in SS-OCT capture relative to conventional SD-OCT, this system is more optimal for choroidal visualisation. However, for spectral domain EDI-OCT, this sensitivity roll-off curve is inverted (figure 1.14). Thus, regardless of higher wavelength and lower sensitivity roll-off, EDI-OCT still obtains higher SNR values for deeper ocular structures [97], such as the larger choroidal vessels nearer the Choroid-Sclera boundary. Figure 1.7 shows the sensitivity roll-off curves for the Heidelberg Engineering Spectralis OCT1, OCT2 and EDI-OCT2 systems, and comparison to a hypothetical SS-OCT device.

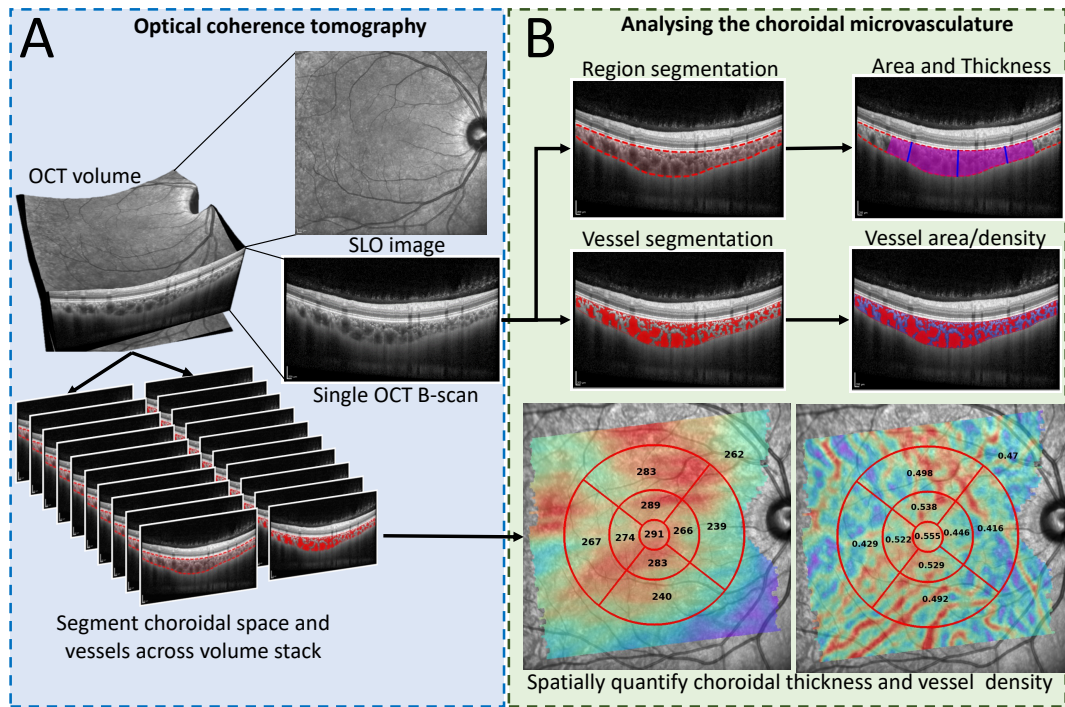
Figure 1.8 compares the visibility of the Choroid-Sclera boundary for a patient with an enlarged choroidal space due to pachychoroid disease (central serous chorioretinopathy). The same location was scanned using Topcon's SS-OCT DRI Triton, Zeiss' SS-OCT Plex Elite 9000 and Heidelberg Engineering's SD-OCT Spectralis OCT2 module with EDI mode activated. Note that for Zeiss and Heidelberg Engineering, these B-scans were captured with an ART of 100. While the Choroid-Sclera boundary is visible across all three devices, the SD-OCT capture in panel (C) has greater contrast of the choroidal space.

## 1.6 Segmenting the choroidal microvasculature

Figure 1.9 shows a diagram with the output from a typical OCT volume scan acquisition in panel (A), showing the OCT volume block with the sequential B-scans and corresponding localiser SLO image. In panel (B), we show the segmentation tasks necessary for choroidal measurement, region and vessel segmentation. We also show the typical measurements of interest that are commonly sought after in choroidal image analysis including, for a single B-scan, thickness and area of the choroidal space and choroidal vessel density. This can also be extrapolated for an OCT volume scan to allow full choroidal mapping across the macula, measuring local choroid thickness and vessel density at different spatial locations of the macula.

Measurement of the choroid in OCT images require segmentation of the choroidal

## 1.6. Segmenting the choroidal microvasculature



**Figure 1.9:** EDI-OCT volume scan acquisition (A), with segmentation tasks and measurements sought for choroidal image analysis (B).

space and vessels, which is generally a harder problem than retinal layer segmentation. This is primarily due to the hyperreflective Bruch's complex which can influence the OCT beam's optical signal as it penetrates deep to the choroid. This can result in poor Choroid-Sclera boundary visualisation. While EDI-OCT and SS-OCT improve choroid visualisation, natural speckle noise still plays a significant role in poor Choroid-Sclera boundary visualisation [97]. Furthermore, the Choroid-Sclera boundary is typically identified from locating the posterior-most vessels appearing as dark blobs suspended in brighter-appearing connective tissue [103, 104]. However, this is not always the case and the homogeneous appearance of the sclera and choroidal interstitial space can often obscure the true boundary location.

The choroidal vessels are highly heterogeneous in terms of size, shape, and edge contrast which make them hard to discern exactly. The choroid is also often plagued by low SNR which corresponds to high speckle and low contrast between the interstitial and vascular compartments. A final challenge is the

oblique presentation of the vessels from cross-sectional visualisations. The choroidal vascular network is a dense mesh of blood vessels which makes it challenging to visualise individual vessels at perfect, circular cross-sections. Instead, oblique viewpoints of the choroidal vessels are seen on each OCT B-scan. Additionally, certain A-scans of the OCT B-scan are corrupted by shadows cast by superficial retinal vessels which branch out of the optic nerve head lying along the inner retinal layers. During acquisition, the OCT beam may penetrate through the retinal vasculature lying along the inner surface of the retina and sitting perpendicular to the incident laser light. This darkens the appearance of the interstitial and vascular compartments, which can often confuse automatic segmentation approaches [105].

These reasons all contribute to the challenging tasks of choroid segmentation. While manual segmentation of the choroidal space and vessels is often done using Fiji (ImageJ) [29] or on the imaging device, this is a complex and time-consuming task, and prohibitively so for choroid vessel segmentation. Moreover, manual choroid segmentation suffers from high intra- and inter-rater variability [106, 107, 108]. Ultimately, manual segmentation is subject to the expertise of the grader, which may vary depending on experience as different graders may have different definitions of the Choroid-Sclera boundary — especially if the suprachoroidal space is present [18]. The subjective nature of manual segmentation prevents standardisation of choroidal measurements in the research community.

### 1.6.1 Region segmentation

#### 1.6.1.1 Manual measurement

The lack of open-source and easily accessible computational approaches to characterise the choroid in OCT often lead to the research community leveraging manual methods [13, 18, 59, 106], and this has also continued as a standard approach even in recent years [14, 16, 21, 64, 72]. Due to the time it would take a manual grader to segment the choroid manually across the macula, representative thickness measurements are taken, usually underneath the

## 1.6. Segmenting the choroidal microvasculature

---

fovea, i.e. subfoveal choroid thickness (SFCT). However, these often suffer from wide variability among graders [106, 107, 108] — especially when detecting the Choroid-Sclera boundary — due to poor visualisation [22] and also inconsistent definition [37].

Rahman, et al. [106] showed that intra- and inter-observer reproducibility had an average difference of 23  $\mu\text{m}$  and 32  $\mu\text{m}$ , respectively, while Shao, et al. [107] showed good agreement between graders of around 3  $\mu\text{m}$  on average. Boonarpa, et al. [22] showed that the repeatability of manual choroid thickness measurements vary according to image quality, ranging from a mean difference between raters of 4  $\mu\text{m}$  and 27  $\mu\text{m}$  across the spectrum of Choroid-Sclera boundary visualisation. Sim, et al. [108] found that inter-rater variability in thickness from two graders improved from 6 – 40  $\mu\text{m}$  on average to 4 – 21  $\mu\text{m}$  after introducing a strict manual measurement protocol for individual Sattler, Haller and choroid layer segmentation.

Yiu, et al. [18] found that varying the definition of the Choroid-Sclera boundary to include/exclude the suprachoroidal space lead to error in manual SFCT ranging from 17  $\mu\text{m}$  to 70  $\mu\text{m}$ . Chandrasekera, et al. [37] showed that agreement between automatic and manual detection of the Choroid-Sclera boundary worsened when the suprachoroidal space was visible (increase of mean difference in thickness of 17  $\mu\text{m}$  to 23  $\mu\text{m}$ ). Conversely, Vuong, et al. [109] found that only 15% of 91 eyes had a visible suprachoroidal space in a population with average age 59.5, and that it's appearance did not impact manual measurement error, which ranged between 1 – 37  $\mu\text{m}$  on average when considering different Choroid-Sclera boundary definitions.

Given that effect size can vary across disease, with some being as low as 20 – 30  $\mu\text{m}$  change (such as when tracking myopia progression [24, 110]), the wide variability in reproducibility from manual grading from poor measurement protocol and human bias [22, 23] poses a problem for standardised choroidal image analysis in OCT image sequences.

Additionally, measuring choroidal thickness on OCT has also been poorly standardised in the research community, as the choroid can appear skewed on an OCT B-scan and some institutions account for this skew [19, 22, 59, 111],

while others do not [13, 18, 21, 64, 106, 112]. We will show in section 1.7.1.3 that accounting for this choroidal curvature can have an inordinate impact on measuring choroidal thickness because of the pixel-to-micron scaling imposed by the OCT B-scan. We will also show in chapter 2 that manual graders are a poor reference standard for measuring choroid thickness consistently within patients when accounting for this choroidal curvature.

### 1.6.1.2 Semi- and Fully-automatic

The inconsistencies introduced from manual grading has been a driving force for more objective approaches based on computer science and mathematics, with traditional image processing methods initially playing a significant role. These methods often required manual intervention or pre-defined parameters, making them semi-automatic and less efficient compared to more modern, fully automatic methods which use deep learning. Table 1.1 gives a brief overview of the range of semi-automatic and fully-automatic approaches developed for choroid region segmentation over the last 12 years. Performance metrics have explicitly not been included due to lack of publicly available datasets to standardise and benchmark performance with.

Briefly, many semi-automatic approaches utilised graph-based methods such as Dijkstra's algorithm [113] for individual edge tracing of the RPE-Choroid and Choroid-Sclera boundaries. These methods represent the image as a graph, where each pixel is a node, and the segmentation task is modelled as a graph partitioning problem [12, 114, 115, 116, 117, 118, 119, 120, 121, 122]. Other traditional, semi-automatic approaches leveraged statistical models based on level sets and different enhancement strategies to simplify the edge tracing problem [123, 124, 125, 126, 127, 128]. While these semi-automatic approaches improve on just manual labelling alone, they typically rely on handcrafted features and pre-defined criteria which lack generalisation. Additionally, many of these approaches take a local perspective on optimisation or require many tuning parameters and thus domain-specific knowledge to operate successfully. Moreover, these methods are multi-stage and heavily rely on many pre-processing steps which, if sub-optimal, can affect the output segmentation.

| Study                        | Method         | Data                           | Description   | Advantages  | Limitations   | Reproducibility | Open-source |
|------------------------------|----------------|--------------------------------|---|---|---|-----------------|-------------|
| Kajic, et al. [117]          | Semi-automatic | N=12 (871), HD-OCT**, Manual   | Two-stage statistical model based on texture, shape and Dijkstra's algorithm [113].   | Reliable performance on low SNR B-scans.                    | Many tuning parameters requiring domain-knowledge.                            | No              | No          |
| Torzicky, et al. [123]       | Semi-automatic | N=5 (5), PS-OCT+, Manual       | threshold technique based on the degree of polarisation uniformity image.   | Leverages phase retardation to identify choroid.            | Limited evaluation data, not generalisable outside PS-OCT.                    | Yes             | No          |
| Duan, et al. [124]           | Semi-automatic | N=6 (384), PS-OCT+, Manual     | Gradient and regression lines on phase retardation tomography from SS-OCT   | Leverages phase retardation to identify choroid.            | Jagged boundary, requires error correction, not generalisable outside PS-OCT. | Yes             | No          |
| Alonso-Caneiro, et al. [118] | Semi-automatic | N=119 (1173), EDI-OCT, Manual* | Two-stage use of Dijkstra's algorithm [113] after several brightness-adjusted enhancement with gradient-based edge weights. | Large evaluation set.                                       | Extensive and interdependent pre-processing, slow (45 seconds/B-scan).        | No              | No          |
| Tian, et al. [148]           | Semi-automatic | N=45 (45), EDI-OCT, Manual     | Gradient-based Dijkstra's algorithm [113], posed as dynamic programming.  | Fast (1.25 seconds/B-scan).                                 | Non-smooth Choroid-Boundary segmentations from local optimisation.            | No              | No          |
| Lu, et al. [114]             | Semi-automatic | N=30 (30), EDI-OCT, Manual     | Two-stage contour model and gradient-based graph search, posed as dynamic programming.                                      | Easy-to-use.  | Human-interactive, small evaluation set.                                      | No              | No          |
| Hu, et al. [121]             | Semi-automatic | N=30 (<1100), SD-OCT, Manual*  | 3D graph-search method for OCT volumes.   | Allows whole-volume segmentation.                           | Evaluated on poor quality data, tuned parameters specific to pathology.       | No              | No          |
| Danesh, et al. [122]         | Semi-automatic | N=10 (100), EDI-OCT, Manual    | Two-stage tracing using Gaussian mixture models with wavelet features and graph-cut.  | Multiresolution pipeline obtaining coarse and fine details. | Requires manual labelling for training for inference, slow (13 – 21s/B-scan). | No              | No          |
| Vupparaboina, et al. [127]   | Semi-automatic | N=5 (485), EDI-OCT, Manual     | Tracing using tensor voting, Hessian analysis and structural similarity index.  | Promotes smoothness and connectivity.                       | Computationally expensive and thus slow.                                      | Yes             | No          |

|                          |                |                              |  |  |   |     |                  |
|--------------------------|----------------|------------------------------|--|--|---|-----|------------------|
| Chen, et al. [125]       | Semi-automatic | N=66 (212), HD-OCT**, Manual | Graph min-cut-max-flow and the energy minimisation.  | Suitable for HD-OCT segmentation.  | Sensitive pre-processing, threshold-based upper bound detection not robust, slow (9s/B-scan). | Yes | No               |
| Masood, et al. [115]     | Semi-automatic | N=21 (525), EDI-OCT, Manual  | Graph-based energy minimisation with normalised cuts.  | More precise than manual methods, better than conventional image processing alone. | Requires manual parameter tuning, slow (40s/B-scan).  | No  | No               |
| Salafani, et al. [116]   | Semi-automatic | N=11 (32), EDI-OCT, Manual   | Neutrosophic space Dijkstra's algorithm [113] extension for boundary detection.  | Handles uncertainty well to improve over standard graph methods.                   | Sensitise pre-processing and post-manual correction.  | No  | No               |
| Chen, et al. [120]       | Semi-automatic | N=42 (978), EDI-OCT, Manual* | 3D choroid segmentation on SD-OCT using EDI-OCT and registration with graph methods.                                       | Enables low-quality SD-OCT segmentation, slow (4.5min/volume).                     | Restricted to volume segmentation, dependent on high quality EDI-OCT.                         | No  | No               |
| Wang, et al. [126]       | Semi-automatic | N=30 (7680), SS-OCT, Manual  | Combined Markov Random Fields (MRF) with level sets.   | MRF and Regularisation to promote smoothness and noise robustness.                 | Multiple parameter tuning, questionable ground-truth comparison.                              | No  | No               |
| Mazzaferrri, et al. [12] | Semi-automatic | N=30 (30), EDI-OCT, Manual*  | Dijkstra's algorithm [113] with several pre-processing steps, including B-scan flattening, Gabor transform edge detection. | Encoded uncertainty in edge-map detection.   | Sensitive and multi-stage pre-processing, requires domain-specific knowledge to use.          | No  | Yes <sup>†</sup> |
| George, et al. [128]     | Semi-automatic | N=19 (589), EDI-OCT, Manual  | Two-stage contour evolution model RPE-Choroid and Choroid-Sclera boundaries  | General purpose for choroid and cornea segmentation.                               | Outputs bumpy contours, sensitive to posterior choroidal vessels.                             | No  | No               |



|                         |               |                               |   |  |  |     |    |
|-------------------------|---------------|-------------------------------|---|--|--|-----|----|
| Chen, et al. [130]      | Deep learning | N=31 (62), EDI-OCT, Manual    | SegNet [149] with graph-based post-processing.  | Local graph-method, global CNN feature learning. Evaluated on pathology.   | Dependent on graph-based post-processing, jagged boundaries.                             | No  | No |
| Sui, et al. [132]       | Deep learning | N=73 (912), EDI-OCT, Manual   | Learned graph-edge weights from CNN with Dijkstra's algorithm [113].  | Smooth segmentations from global/local features from CNN/graph theory. Leverages off-line augmentation to increase dataset | Data leakage in evaluation.  | No  | No |
| Al-Bander, et al. [133] | Deep learning | N=169 (169), EDI-OCT, Manual  | Patch-based CNN based on super-pixel segmentation.  | Leverages global CNN representation learning.  | Unsmooth boundary segmentations from super-pixel CNN classification.                     | No  | No |
| Masood, et al. [134]    | Deep learning | N=21 (525), EDI-OCT, Manual   | Patch-based CNN classifier into "choroid", "not-choroid".   | Removes handcrafted features from pipeline.  | Semi-automatic RPE-Choroid tracing, not pixel-level segmentation.                        | No  | No |
| He, et al. [135]        | Deep learning | N=108 (146), EDI-OCT, Manual  | Patch-based CNN classifier into "choroid boundary", "not-choroid boundary".                                   | Regularisation ensures smoothness and removes handcrafted features from pipeline.  | Semi-automatic RPE-Choroid tracing and RANSAC [150] for patch-stitching.                 | No  | No |
| Hsia, et al. [137]      | Deep learning | N=30 (750), EDI-OCT, Manual   | Mask R-CNN for feature extraction, region proposal and boundary detection.                                    | Leverages fine-tuning and ensemble modelling on small dataset.   | Slow execution ( $\approx 6s/B$ -scan on GPU), required post-processing to remove dents. | No  | No |
| Chen, et al. [136]      | Deep learning | N=123 (3075), EDI-OCT, Manual | Combines Mask R-CNN to extract square choroid ROI and use fully convolutional network [151] for segmentation. | Leverages fine-tuning.   | Over complex, required post-processing to remove dents.                                  | No  | No |
| Tsuji, et al. [131]     | Deep learning | N=34 (290), SS-OCT, Manual    | SegNet [149] for pixel-level segmentation with class weight balancing.  | No requirement of pre-processing, less sensitive to noise than graph methods   | Tunable, post-processing to smooth Choroid-Sclera boundary.                              | Yes | No |



|                       |               |                                 |  |   |   |    |    |
|-----------------------|---------------|---------------------------------|--|---|---|----|----|
| Mao, et al. [142]     | Deep learning | N=20 (5120), SS-OCT, Manual     | Modified UNet [138] with learnable skip connections and global pooling.                        | Leverages augmentation on-the-fly, outperforms standard UNet. | Post-processing necessary in non-signal regions, data leakage potential, no details on evaluation data. | No | No |
| Devalla, et al. [143] | Deep learning | N=100 (100), EDI-OCT, Manual    | UNet [138] with dilated convolutional blocks for multi-layer retinal and choroid segmentation. | Comprehensive intra-ocular segmentation.                      | Evaluation focus not on the choroid.  | No | No |
| Chai, et al. [144]    | Deep learning | Unknown (790), SS-OCT, Manual   | Adversarial learning to minimise domain-shift discrepancy from Topcon to Nidek.                | No requirement of ground truths in target domain (Nidek).     | Low-quality perceptual loss, No detail on OCT mode of devices, small evaluation.                        | No | No |
| Zhang, et al. [146]   | Deep learning | N=20 (1280), SS-OCT, Manual     | Bio-Net: Several UNet's [138], regularised to match choroid thickness.                         | Leveraged to find shadow-free en face choroid visualisations. | High complexity from multiple models.   | No | No |
| Cahyo, et al. [145]   | Deep learning | N=99 (42496), SS-OCT, Manual    | SA-Net: UNet [138] with spatial feature reconstruction using consecutive B-scans.              | Large dataset, leverages features from adjacent B-scans.      | High complexity from multiple models, unfeasible on edge-devices.                                       | No | No |
| Xu, et al. [15]       | Deep learning | N=35 (1400), SS-OCT, Manual     | Enhanced UNet [138] by adding dense connection per block [140].                                | Dense feature extraction, evaluated on high myopia.           | Poor segmentation on raw B-scans, sensitive pre-processing required.                                    | No | No |
| Wang, et al. [147]    | Deep learning | Unknown (1062), EDI-OCT, Manual | PGKD-Net: Transformer-assisted cascade network   | Evaluated on several tasks, and high myopia.                  | Requires prior mask to reproduce, high complexity implies scalable only with GPU resources.             | No | No |

**Table 1.1:** Semi-automatic and fully-automatic approaches to choroid region segmentation over the last 12 years. The data column is structured to describe the number of eyes (total B-scans) used for training and/or evaluation, the type of OCT data and whether model development/training used manual labels. \* : Manual labelling by two or more graders; \*\* : High-penetration OCT; + : Polarization Sensitive OCT; †: Requires proprietary software MatLab.

As datasets grew and computational resources became more available, fully automatic approaches based on convolutional neural networks (CNNs) become preferable over semi-automatic ones and could provide faster and more generalisable inference [129]. The earlier approaches were a combination of feature extracting CNNs which took a patch-based binary classification approach to region segmentation, utilising semi-automatic algorithms for pre/post-processing and model fitting [130, 131, 132, 133, 134, 135, 136, 137]. More recently, there has been a push to develop fully automatic methods using fully convolutional networks (FCNs), such as the popularised UNet deep learning architecture [138] with residual [139] or dense learning [140]. This permits pixel-level segmentation and Kugelman, et al. [141] found that this type of encoder–decoder deep learning architecture outperformed patch-based methods. Thus, the UNet model has become the reference standard for deep learning approaches toward choroid region segmentation [15, 142, 143, 144, 145, 146, 147].

### 1.6.2 Vessel segmentation

While manual segmentation of the choroidal region requires identification of a single landmark in the OCT B-scan, segmenting the choroidal vessels require multiple, interdependent and heterogeneous blobs to be identified which is prohibitively time consuming and subject to human bias. Thus, approaches toward choroid vessel segmentation leveraged semi-automatic and fully-automatic approaches from the get-go. Table 1.2 gives a brief overview of the range of semi-automatic and fully-automatic approaches developed for choroid vessel segmentation over the last 12 years. Performance metrics have explicitly not been included due to lack of publicly available datasets to standardise and benchmark performance with.

Some of the earliest methods were oriented around binarisation and thresholding, whereby choroidal pixels are classified based on pixel intensity values into choroid vessel and surrounding interstitial tissue (choroid stroma) [104, 152]. Other methods were more domain-specific to address the heterogeneity of the choroidal vessel size and contrast [153, 154, 155, 156] but were heavily inter-

| Study                      | Method         | Data                          | Description  | Advantages  | Limitations  | Repro-ducibil-ity | Open-source      |
|----------------------------|----------------|-------------------------------|--|---|--|-------------------|------------------|
| Zhang, et al. [153]        | Semi-automatic | N=43, SD-OCT, None            | Multiscale Hessian matrix analysis.  | Addresses size-based heterogeneity of vessels.  | Assumed circularity of vessel, large data loss (N=19) from poor choroid visualisation.   | Yes               | No               |
| Kajic, et al. [154, 155]   | Semi-automatic | N=12 (871), EDI-OCT, Manual   | 3D multi-scale edge filtering.   | Addresses size-based heterogeneity of vessels. Pathological evaluation.                       | Vessel estimation assumed circularity.   | No                | No               |
| Mahajan, et al. [156]      | Semi-automatic | N=12 (12), SD-OCT, Manual     | Multiple global thresholding with Frangi's filter [157].                     | Domain-specific method to address size/contrast heterogeneity of vessels. Simple application. | Inter-dependent pipeline on global thresholding, limited evaluation.   | No                | No               |
| Branchini, et al. [104]    | Semi-automatic | N=42 (42), SD-OCT, None       | Otsu's global thresholding after manually selecting choroidal area.          |   | Sensitive to noise, limited to large vessels, requires manual selection.   | No                | No               |
| Sonoda, et al. [27]        | Semi-automatic | N=20 (20), EDI-OCT, Manual    | Niblack's local thresholding after manual selection of vessel brightness.    | Open-source protocol, commonly used, assessed in pathology.                                   | Did not specify Niblack parameters, requires manual brightness adjustment, limited to high quality choroids and strict grading protocol, slow ( $\approx$ 5 min/B-scan). | Yes               | Yes**            |
| Vupparaboina, et al. [152] | Semi-automatic | N=2 (388), SD-OCT, None       | Multi-stage enhancement with histogram equalisation and global thresholding. | Less human involvement than Sonoda, et al. [27], less noise sensitive.                        | Limited and only qualitative evaluation, global thresholding not adept to multi-scale vessels.   | No                | No               |
| Agrawal, et al. [158]      | Semi-automatic | N=345 (345), EDI-OCT, Manual* | Modified Niblack's method with morphology.                                   | More automated than Sonoda, et al. [27] Commonly used across studies.                         | Lack of standardisation of Niblack's parameters across studies, slow ( $\approx$ 1 min/B-scan).  | Yes               | Yes <sup>†</sup> |
| Liu, et al. [159]          | Deep learning  | N=10 (40), SS-OCT, Manual*    | UNet based architecture with a large ResNet encoder.                         | No pre-processing required, assessed inter-rater variability.                                 | High computational demand, manual vessel labelling.  | No                | No               |

**Table 1.2:** Semi-automatic and fully-automatic approaches to choroid vessel segmentation over the last 12 years. The data column is structured to describe the number of eyes (total B-scans) used for training and/or evaluation, the type of OCT data and whether model development/training used manual labels. \*: Manual labelling by 2 or more graders; \*\*: Measurement protocol provided in [27]; <sup>†</sup>: Website available but membership, permission and co-author requirements [160].

dependent on a priori information such vessel shape.

Local thresholding methods were preferred over global ones as an attempt to address the heterogeneity in size of the choroidal vasculature. Namely, the Niblack autolocal thresholding algorithm [161] became commonly used in the research community by virtue of two research groups [27, 158] releasing measurement protocols for Fiji (ImageJ) [29] or membership-only image analysis platforms [160] which use modified versions of Niblack. These semi-automatic approaches suffer from manual intervention as part of the protocol, thus making large-scale volumetric analysis unfeasible. Therefore, some attempts have been made to replace semi-automatic algorithms using deep learning [159], with UNet architectures having superior performance over SegNet [149] and DeepLab [162] models, according to ground-truth labels generated by Niblack's method, as reported by Muller, et al. [163].

### 1.6.3 Synthesising region and vessel segmentation

All of the previous approaches have considered the tasks of region and vessel segmentations as separate ones. However, it is advantageous to consider the inter-dependence of the choroidal space and vessels during the modelling process, and recent advancements have focused on synergistically segmenting both the choroidal region and its vasculature to provide more comprehensive analyses of the choroid. Table 1.3 gives a brief overview of the range of semi-automatic and fully-automatic approaches developed for choroid region and vessel segmentation. Performance metrics have explicitly not been included due to lack of publicly available datasets to standardise and benchmark performance with.

Traditional methods have been proposed for choroid layer and vessel segmentation [119, 164], but were complex, multi-stage and heavily parameter dependent. Attempts have been made at deep learning-based segmentation of Haller's and Sattler's layers [165], despite the oblique presentation of choroidal vessels on OCT and no reliable ground truth or standardised protocol to differentiate the two vascular layers [8]. Other deep learning methods segmented

| Study                 | Method         | Data                            | Description   | Advantages  | Limitations   | Repro-ducibil-ity | Open-source |
|-----------------------|----------------|---------------------------------|---|---|---|-------------------|-------------|
| Srinath, et al. [164] | Semi-automatic | Unknown, OCT, None              | EDI- Hessians analysis for boundary with structural similarity index, level sets for vessels.                           | Domain-specific algorithm based on heterogeneity of intensities in choroid.               | Sensitive pre-processing, jagged segmentation boundaries, fails to segment small vessels.   | No                | No          |
| Hussain, et al. [119] | Semi-automatic | N=10 (190), OCT, Manual*        | EDI- Global Otsu thresholding for vessels, leverages vessel features to obtain edge weights for Dijkstra's algorithm.   | Novel depth-intensity normalisation for choroid enhancement.                              | Requires several tuning parameters, domain-specific knowledge, noise-sensitive.             | No                | No          |
| Li, et al. [165]      | Deep learning  | N=92 (92), OCT, Manual*         | EDI- Two separate UNet's [138] for vessels and (Haller-Sattler, Choroid-Sclera) multi-layer segmentation.               | Comprehensive choroid segmentation.   | Challenging ground truth validation.  | No                | No          |
| Zheng et al. [166]    | Deep learning  | N=23 (667), SS-OCT, Auto/Manual | SS- UNet [138] with residual connections for boundary, Niblack method for vessels.                                      | Improved feature learning and map reconstruction.   | Semi-automatic Niblack vessel segmentation, unnecessary loss of resolution, healthy cohort. | Yes               | No          |
| Khaing, et al. [168]  | Deep learning  | N=160 (160), SD-OCT, Manual     | SD- ChoroidNET: Two dilated UNet [138] models for sequential choroid layer and vessel segmentation.                     | Utilises inter-dependency of choroidal space and vessels, tested on pathology.            | Small training set, manual vessel labelling.  | No                | No          |
| Zhu, et al. [169]     | Deep learning  | N=92 (2852), EDI-OCT, Manual*   | EDI- CUNet: UNet [138] style with single feature extraction encoder, two task-specific decoders for region and vessels. | Adaptive multi-task loss as regularisation, end-to-end, evaluated on myopia.              | Potential data leakage. Manual vessel annotation.   | No                | No          |
| Wang, et al. [170]    | Deep learning  | N=65 (396), SS-OCT, Manual*     | SS- CVI-Net: UNet [138] and pyramid pooling layers for multi-scale feature learning.                                    | Addresses heterogeneity of choroidal vessel size.   | Small dataset, extensive pre-processing required, manual vessel labels.                     | No                | No          |
| Xuan, et al. [17]     | Deep learning  | N=162 (2162), SS-OCT, Manual    | SS- Medical transformer [171] for region, UNet [138], Niblack for vessels.  | Permits standardised choroid measurements, attempted automation of Niblack.               | Small reproducibility dataset (20 B-scans), Use of semi-automatic Niblack method.           | Yes               | Yes**       |
| Arian, et al. [167]   | Deep learning  | N=156 (537), EDI-OCT, Manual    | EDI- Modified-loss UNet [138] for retinal pathology.  | Specified Niblack training parameters.  | Semi-automatic Niblack method. Limited to fovea-centred evaluation.                         | No                | Yes†        |
| Wen, et al. [172]     | Deep learning  | N=10 (120), SS-OCT, Manual      | SS- TACLNet: Transformer-assisted UNet [138].   | Multi-scale feature learning, generalisable to other tasks, outperforms previous methods. | Small dataset, manual vessel labelling  | No                | No          |

**Table 1.3:** Approaches to choroid region and vessel segmentation. The data column is structured to describe the number of eyes (total B-scans) used for training and/or evaluation, the type of OCT data and whether model development/training used manual labels. \* : Manual labelling by two or more graders. \*\* : Website inactive as of December 2024; † : Codebase not plug-and-play, only fully-automatic region segmentation, released dataset.

the region and vessels but still considered their problem formulations as separate entities [17, 166, 167], applying semi-automatic Niblack's method for vessel segmentation. Although, Xuan, et al. [17] did acknowledge the need for standardisation of choroidal measurements and included additional OCT B-scan foveal pit detection in their model DCAP.

Finally, more comprehensive approaches combined the tasks of region and vessel segmentation into single deep learning-based models, with the creation of models based on UNet architectures [138] like ChoroidNet [168] CUNet [169] and CVI-Net [170]. A final advancement was the development of transformer-assisted TACLNet, proposed by Wen, et al. [172] which showed superiority performance across CUNet [169], ChoroidNet [168], CVI-Net [170] and the regular UNet [138].

### 1.6.4 Current challenges

#### 1.6.4.1 Vessel segmentation

Regarding choroid vessel segmentation, the research community largely adopt the semi-automatic vessel segmentation protocols outlined by Sonoda, et al. [27] and Agrawal, et al. [158] which both use the semi-automatic Niblack thresholding algorithm. However, despite the widespread use of Niblack's algorithm for choroid vessel segmentation, there is concern surrounding its improper usage. A crucial element missing from the established software and measurement protocols [27, 160] is how to use the Niblack method and, specifically, how to tune the internal parameters of the local thresholding tool. The platforms available to conduct image analysis, such as Fiji (ImageJ) [29], have further compounded this issue, as they often default to specific parameter settings that may not be reported by researchers. To the best of our knowledge, only two studies have reported specific Niblack parameters [163, 167]. This lack of transparency makes it difficult to compare results across different studies or to replicate findings, thus hampering efforts to standardise the reporting of choroidal measurements.

There have also been several other reports hindering the standardisation of

Niblack's autolocal thresholding algorithm [163, 173, 174, 175, 176]. Wei, et al. [173] found a reported difference of around 3% in choroid vessel density (and an intra-class correlation of 0.35) measured following the protocols outlined by Sonoda, et al. [27] and Agrawal, et al. [158]. Elbay, et al. [174] observed the modified Niblack method proposed by Agrawal, et al. [158] to be sensitive to the type of scaling used to view the OCT B-scan (see section 1.7.1.2), reporting errors of up to 4% in choroid vessel density. Muller, et al. [163] found that tuning of only one of the internal Niblack parameters could result in differences in choroid vessel density between 2% and 20%. Rosa, et al. [175] found a significant difference of around 7.5% in choroid vessel density when using the Niblack method proposed by Agrawal, et al. [158] and altering the image brightness of the OCT B-scan. Finally, Ma, et al. [176] reported that the mean difference between SD-OCT and SS-OCT data in three dimensional vessel density was 13% using the protocol outlined by Sonoda, et al. [27].

These are crucial studies which accompany the concern around the Niblack method's potential lack of reproducibility and measurement error. Agrawal, et al. [28] published a major review on choroid vessel density in retinal and systemic disease reporting average changes between healthy and diseased eyes between 2% and 6%. This suggests that improper usage of these current methods and their lack of standardisation could result in measurement error exceeding observed effect sizes in disease, which could lead to paradoxical results inherently skewed by measurement variability. This was posited by Rosa, et al. [175] who, in their discussion, found contradictory results among the literature when measuring the choroid vessel density in Vogt-Koyanagi-Harada disease [177, 178, 179, 180] and Non-Arteritic Anterior Ischemic Optic Neuropathy [181, 182], using the methods outlined by Sonoda, et al. [27] and Agrawal, et al. [158].

Conversely to Ma, et al. [176], Agrawal, et al. [183] reported a mean difference in choroid vessel density of only 3% between two-dimensional SD-OCT and SS-OCT data. While not perfectly comparable to three-dimensional vessel density, recent evidence suggest that choroid vessel density does not change across different macular scanning locations [184, 185, 186] (but does between macular and peripheral locations [187]). Thus, it is possible that the contradic-

tory findings between Ma, et al. [176] and Agrawal, et al. [183] may be linked to improper use of Niblack’s method, such that it may be sufficient to measure the choroidal vessels only with suitable domain knowledge and expertise of the algorithm. Given that Agrawal, et al. [183] developed and released the modified Niblack method, it’s likely they know how best to use it. Unfortunately, there lacks a standardised protocol for using and specifying internal parameters for semi-automatic methods like Niblack which is so heavily adopted in the community [28]. It would be pertinent for one to exist to educate researchers on the pitfalls of its parameter specification, similar to Chu, et al. [188] who do so for using the Phansalkar thresholding algorithm for the choriocapillaris in OCT-A.

### 1.6.4.2 Bias, reproducibility and accessibility

A problem in the literature is the generation of high quality ground truth data for deep learning model development, particularly for choroid vessel segmentation. The majority of the sophisticated deep learning-based are developed and evaluated against manual ground truth labels which are subject to human bias (tables 1.1 – 1.3, Data column). Few consider the value of developing models based on ground truth labels generated using semi-automatic or fully automatic approaches, which explicitly encode a degree of determinism and consistency into the ground truth generation process. Maloca, et al. [189] argues that such approaches to generating ground truth labels are preferable as they reduce subjectivity and thus bias and inconsistency, especially for the choroid.

Furthermore, there is a distinct lack of reporting the reproducibility of automatic methods in medical imaging research and deep learning-based segmentation (tables 1.1 – 1.3, Reproducibility column, only 9 (18%) out of all 49 reviewed papers conducted reproducibility analysis). For those studies which do report reproducibility, they typically adopt population-level, dimensionless metrics like the intra-class correlation or coefficient of variation/repeatability. Even unit-dependent metrics like average signed differences can be subject to cancellation effects and asymmetric error distributions. Without an *interpretable* and



*contextual* assessment of reproducibility, there is no way to differentiate measurement variability from the true biological effect being measured. This is particularly pertinent for interpreting study findings where effect sizes are likely to be small, such as in choroid vessel density between healthy and disease eyes (2% – 6%) [28] or choroid thickness in myopia progression (20 – 30  $\mu\text{m}$ ) [24, 110].

There are currently no accessible, open-source algorithms for large-scale, fully-automatic choroid region and vessel segmentation (tables 1.1 – 1.3, Open-source column, only 5 (10%) out of all 49 reviewed papers released their code-base). All available algorithms fall into one of three categories: First, manual or semi-automatic methods [26, 27, 29, 158, 190] that require human supervision (introducing subjectivity) and/or membership-access only. Second, fully-automatic deep learning-based methods that are not openly accessible, either only providing the code but not the trained model necessary to use the method [141, 147] or not providing any access at the time of writing [17] (December 2024). Third, releasing the model as fully-automatic but comprising of many steps, requiring underlying domain-specific knowledge or a license for proprietary software [12, 167].

### 1.6.5 Summary

The field of choroid region and vessel segmentation has seen significant evolution. Beginning from traditional image processing like graph-based [12, 117, 119] and thresholding methods [27, 28, 158], which suffer from domain-specific expertise, free-parameter tuning, inter-dependent pipelines and limitations on generalisability to complex edge-cases, and ending at advanced, fully-automatic deep learning-based methods, which offered superior accuracy and large-scale automation [129], marking a significant advancement in the field of ophthalmic image analysis. While initially these methods treated the two tasks separately [15, 146, 147, 159], researchers leveraged deep learning methods for synergistic segmentation of the region and vessels [17, 168, 169, 170, 172]. However, these methods are still limited. Their ground truth label generation is still subject to human subjectivity when produced manually, and there is a lack

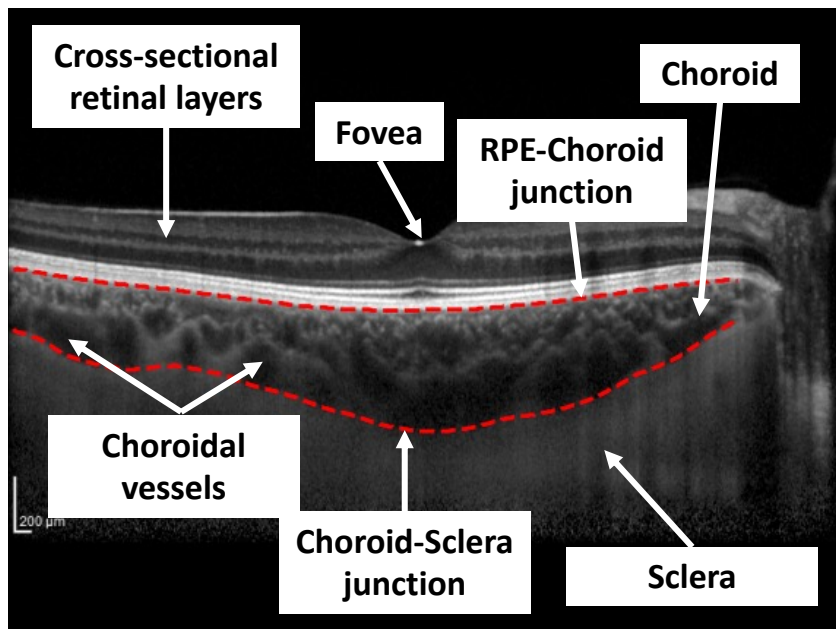
of interpretable reproducibility analysis to help the research community differentiate measurement variability from true biological change. Additionally, there is a significant absence of open-source, deep learning methods available to the wider research community. This has resulted in much of the research community still dependent on manual or semi-automatic methods which are made available by image analysis platforms like Fiji (ImageJ) [29]. This can have unfortunate consequences regarding the standardisation of choroidal measurements in OCT image sequences, particularly if different definitions of the Choroid-Sclera boundary are assumed [18, 22, 23, 37], or improper usage of Niblack's method continues [163, 173, 174, 175, 176].

## 1.7 Measuring the choroidal microvasculature

Figure 1.10 shows an exemplar EDI-OCT B-scan with the RPE-Choroid and Choroid-Sclera boundaries delineating the choroidal space, with important biological landmarks commonly captured in the B-scan also visually defined. Conceptually, there is debate around the exact location of the Choroid-Sclera boundary, and figure 1.11 demonstrates this. Depending on the appearance of the choroid (and the visibility of the suprachoroidal layer) in an OCT B-scan, this may either be [18]:

1. the hyperreflective line at the posterior most location of the largest choroidal vessels in Haller's layer, i.e. inner choroid junction (figure 1.11, blue arrow);
2. the hyporefective line separating the posterior of Haller's layer from the suprachoroid, i.e. inner suprachoroid junction (figure 1.11, yellow arrow);
3. the hyperreflective line separating the suprachoroidal layer from the sclera, i.e. inner scleral junction (figure 1.11, green arrow).

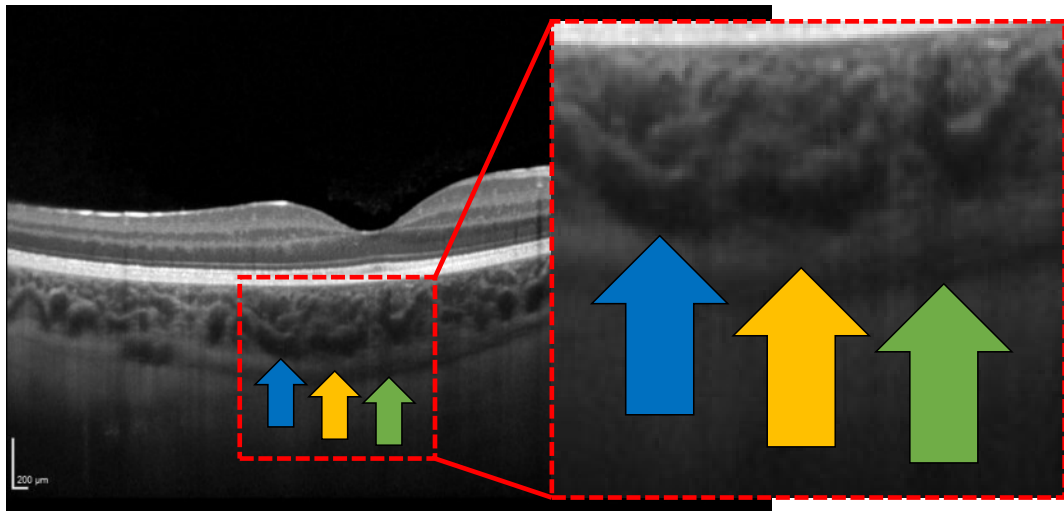
The disparity in these definitions can have an inordinate impact on choroidal measurements — Yiu, et al. [18] found that error in manual SFCT can vary from  $17\ \mu\text{m}$  to  $70\ \mu\text{m}$  when different definitions of the Choroid-Sclera boundary were used, especially when the suprachoroidal layer was observed.



**Figure 1.10:** EDI-OCT B-scan showing the choroid beneath the retina, posterior to the hyperreflective outer retinal layers and fixed to the RPE. Red-dashed lines mark the RPE-Choroid and Choroid-Sclera boundaries, with annotated landmarks. Due to the limited resolution of current OCT devices (about  $3.87 \mu\text{m}$ -per-pixel for Heidelberg Engineering Spectralis OCT1 devices), the chorio-capillaris layer is barely visible. The choroidal space appears as bright regions (stroma) with darker areas representing vasculature, a pattern empirically observed [103, 104] and accepted in the research community [28].

The original authors who proposed EDI-OCT [11] define the choroid to include the suprachoroidal layer [8], and thus define the Choroid-Sclera boundary at the point of the inner sclera [191], i.e. definition (3) (green arrow in figure 1.11). However, even that definition is still debated and can be hard to discern in practice, as the suprachoroidal space is not always present due to OCT device limitations, age, health and other factors [18]. Inconsistency arises when certain choroids exhibit a suprachoroidal space and others do not, under the assumption that this space has no relationship to the research objective [18, 37].

Chandrasekera, et al. [37] showed that only 38% of 200 eyes showed a visible suprachoroid in SS-OCT data, and Yiu, et al. [18] showed a similar proportion (45%) in 74 eyes using EDI-OCT data, noting that visibility of the suprachoroid



**Figure 1.11:** OCT B-scan with visible suprachoroidal space, with coloured arrows corresponding to the three distinct definitions of the Choroid-Sclera boundary.

was higher in older eyes. Additionally, both studies found that choroidal thickness was significantly different when considering the different definitions of the choroid outlined above. Moreover, Chandrasekera, et al. [37] additionally found that agreement between semi-automatic and manual detection of the Choroid-Sclera boundary worsened when the suprachoroidal space was visible (increased mean difference in thickness of  $17 \mu\text{m}$  to  $23 \mu\text{m}$ ), reporting that automatic methods would prefer definition (1) while manual measurement would opt for definition (3).

Thus, to remain objective in our analyses throughout this thesis — given the inconsistent visibility of the suprachoroidal layer, and the fact that it is an avascular potential space [192] unlike the highly vascularised choroidal layers [7] — we opted to follow definition (1) (blue arrow in figure 1.11).

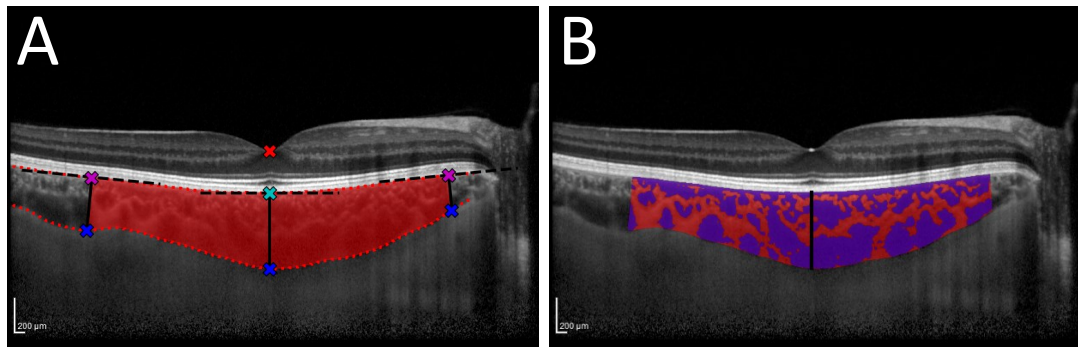
Therefore, the choroidal space in an OCT B-scan for this thesis is defined as the space posterior to the boundary delineating the RPE layer and Bruch's membrane complex (RPE-Choroid) and superior to the boundary delineating the sclera or suprachoroid, if visible, from the posterior most point of Haller's layer (Choroid-Sclera).

### 1.7.1 OCT B-scan

Measuring the choroid in the macula is usually done with reference to a particular region of interest (ROI). This is to ensure comparisons between and within patients can be done accurately and reliably. For the work in this thesis, we consider measuring the macular choroid in a fovea-centred ROI, with some micron distance either side of the fovea. The specific location of the fovea is its centre, at the foveal pit. The foveal pit in an OCT B-scan is defined in this thesis as a single pixel coordinate at the centre of the foveola centralis. The foveola is a circular zone of approximately 350 micron width which is avascular, creating a depression at the centre of the macula which pushes the inner retinal layers laterally [193]. The single pixel coordinate is one which appears most brightly illuminated on the B-scan in this depression, often aligning with a ridge formed at the photoreceptor layer.

Figure 1.12 demonstrates the construction of a two-dimensional ROI in an OCT B-scan (A), with measurements visualised in (B). In panel (A), we show the landmarks used to define a two-dimensional ROI (red), given a 3000  $\mu\text{m}$  distance either side of the fovea and describe its construction below:

1. Given the location of the fovea (red cross in 1.12(A)), find the nearest location along the RPE-Choroid boundary, which defines the centre of the ROI (cyan cross in 1.12(A)).
2. Traversing the RPE-Choroid boundary some micron distance (3000  $\mu\text{m}$  in this example) either side of this central pixel defines the endpoints of the ROI (purple crosses in 1.12(A)) along this boundary.
3. Tangential lines at each RPE-Choroid reference point are drawn (black dashed lines in 1.12(A)), using an offset of 15 pixels in each direction to the reference point to estimate tangential gradient around reference points.
4. Tangential lines are rotated 90 degrees (black solid lines in 1.12(A)) and extended until they intersect the Choroid-Sclera boundary, detecting the reference points along this boundary (blue crosses in 1.12(A)).



**Figure 1.12:** (A) OCT B-scan with measurement landmarks overlaid to describe construction of the fovea-centred ROI. (B) OCT B-scan with prescribed ROI in red, SFCT in black, and vessels in purple.

The ROI is defined as the choroidal space (red shaded region in 1.12(A)) which is bound between these solid black lines which extend *locally perpendicular* from the RPE-Choroid boundary. Importantly, the prescribed, fovea-centred ROI accounts for any choroidal curvature from poor acquisition, or degree of myopia, and ensures a standardised measurement protocol for measuring the choroid.

In figure 1.12(B), we show the 3000  $\mu\text{m}$ , fovea-centred ROI in red, with segmented vessels in this ROI in blue, and a black solid line representing the one-dimensional, subfoveal choroid thickness (SFCT) for this B-scan. Within this ROI, there are four metrics of interest to measure the choroid: SFCT, choroid area, choroid vascular area and choroid vascular index (CVI). SFCT measures the thickness of the choroid directly underneath the fovea, and locally perpendicular to the RPE-Choroid boundary (black solid line in figure 1.12(B)). Using the pixel length-scales in the axial and lateral directions, the pixel length of this black line is converted into microns. It is most common to measure the choroid through a one-dimensional measurement like SFCT [10]. However, because of the segmental nature of the choroidal vascular bed [44], this is misguided [83]. Thus, it is common to take several thickness measurements across the ROI and average them [23].

Instead of one-dimensional thickness, two-dimensional measurements like area are more representative and robust measurements of the choroid [23]. Choroid

area counts the number of pixels in the total ROI (red and purple pixels in figure 1.12(B)), and this count is converted into  $\text{mm}^2$  by a conversion factor equal to the total  $\text{mm}^2$  area of a single pixel, dictated by the axial and lateral pixel length-scales. Choroid vessel area is computed much the same but only for the vessel pixels (purple pixels in 1.12(B)). CVI is a dimensionless ratio of choroid vascular area to total choroid area within the prescribed ROI (ratio of purple pixels to red and purple pixels in figure 1.12(B)).

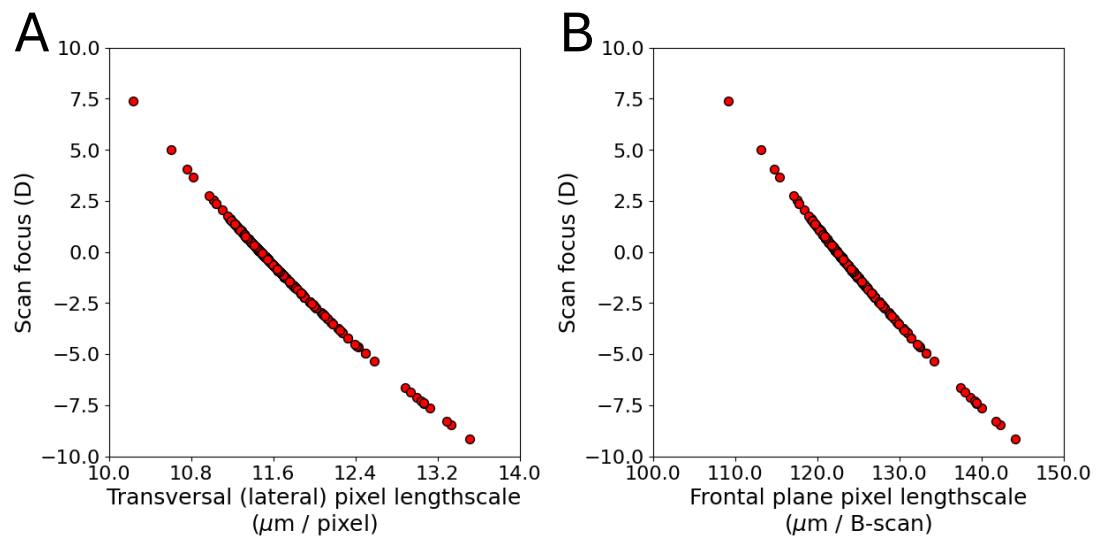
### 1.7.1.1 Pixel length-scales

The axial and lateral pixel length-scales in an OCT B-scan are typically not equal, which is known as an *anisotropic* scaling. For OCT imaging devices, the ratio of axial to lateral pixel length-scale is approximately 1:3. Heidelberg Engineering's Spectralis OCT1 module has a fixed axial pixel length-scale of  $3.87 \mu\text{m}$ -per-pixel and, for a perfectly emmetropic model eye, a lateral pixel length-scale of  $11.47 \mu\text{m}$ -per-pixel for pixel resolution  $768 \times 768$ . Topcon's DRI Triton Plus has a fixed axial pixel length-scale of  $2.61 \mu\text{m}$ -per-pixel and a fixed lateral pixel length-scale of  $8.79 \mu\text{m}$ -per-pixel for pixel resolution  $992 \times 1024$ .

For some OCT imaging devices (Heidelberg Engineering's OCT Spectralis, for example), the lateral pixel length-scale has the ability to change across different examinations depending on the corneal surface properties and refraction of the patient's eye, as well as the operators use of the machine. While the former can be approximately encoded into the device, this is not always known and thus a phantom eye model is used instead. In the latter, during acquisition the operator can adjust various settings to compensate for the individual refractive surfaces in each patient's eye [102].

An important tuning parameter during acquisition for the technician is the scan focus knob on the imaging device. This compensates for the refractive error of the patient's eye, and thus acts as an approximation for the patient's refractive error (measured in dioptres, D). While suitable focus on the chorioretinal structures is possible with the scan focus knob set to '0' — perfectly emmetropic — it can be adjusted freely and is particularly useful in cases of high myopia or hy-





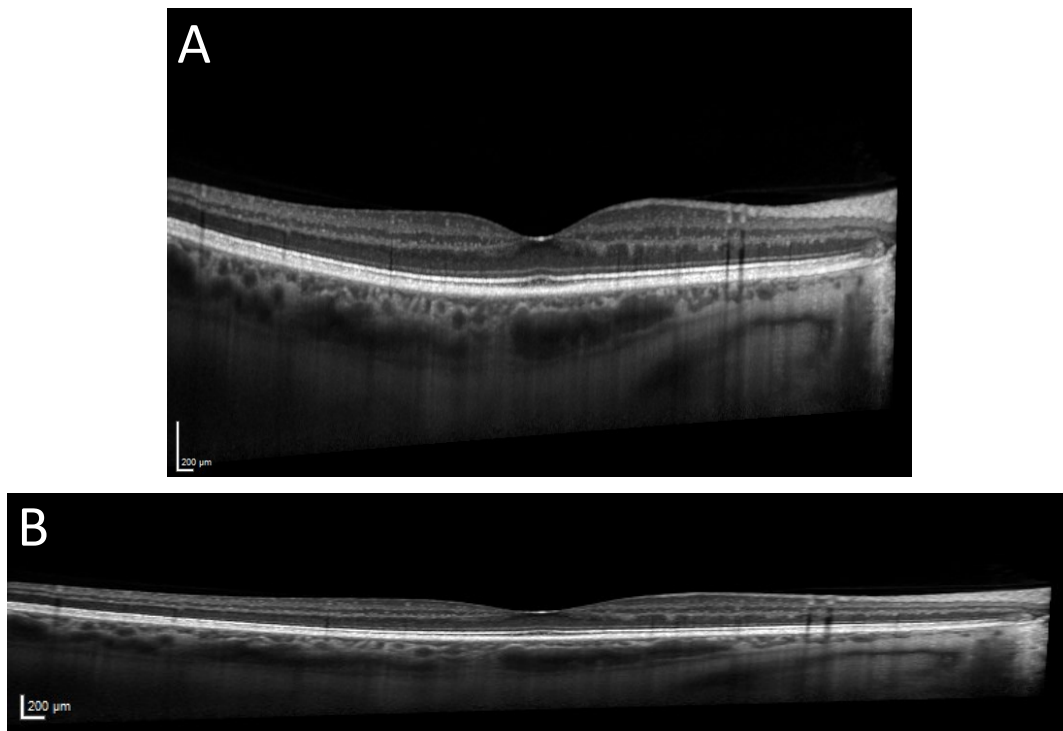
**Figure 1.13:** Relationship between the scan focus value and the lateral (A) and frontal (B) pixel length-scales for Heidelberg Engineering OCT systems.

peropia [102]. Adjusting the focus alters the lateral pixel length-scale in an OCT B-scan, as well as the distance between sequential B-scans in an OCT volume scan (the frontal plane length-scale). This relationship can be observed in figure 1.13, which is approximately linear in the emmetropic range, between -1D and 1D. This is a crucial detail for producing standardised choroidal measurements, and can in particular have significant consequences for measurements which are computed using different ROI definitions.

### 1.7.1.2 Different viewing modes

A way to circumvent anisotropic scaling is to present the chorioretinal structures in a B-scan whose axial and lateral pixel length-scales are equal. Default scaling used by most in the research community presents the OCT B-scan in ‘1:1 pixel mode’ (figure 1.14(A)), allowing chorioretinal structures to be visualised more easily for the end-user by appearing more bloated in the axial direction which aids the interpretation and readability of the B-scan. Another option available to present the chorioretinal structures is ‘1:1 micron mode’ (figure 1.14(B)) which, although squashes the chorioretinal structures axially, is advised by device manufacturers to help mitigate for potential manual mea-





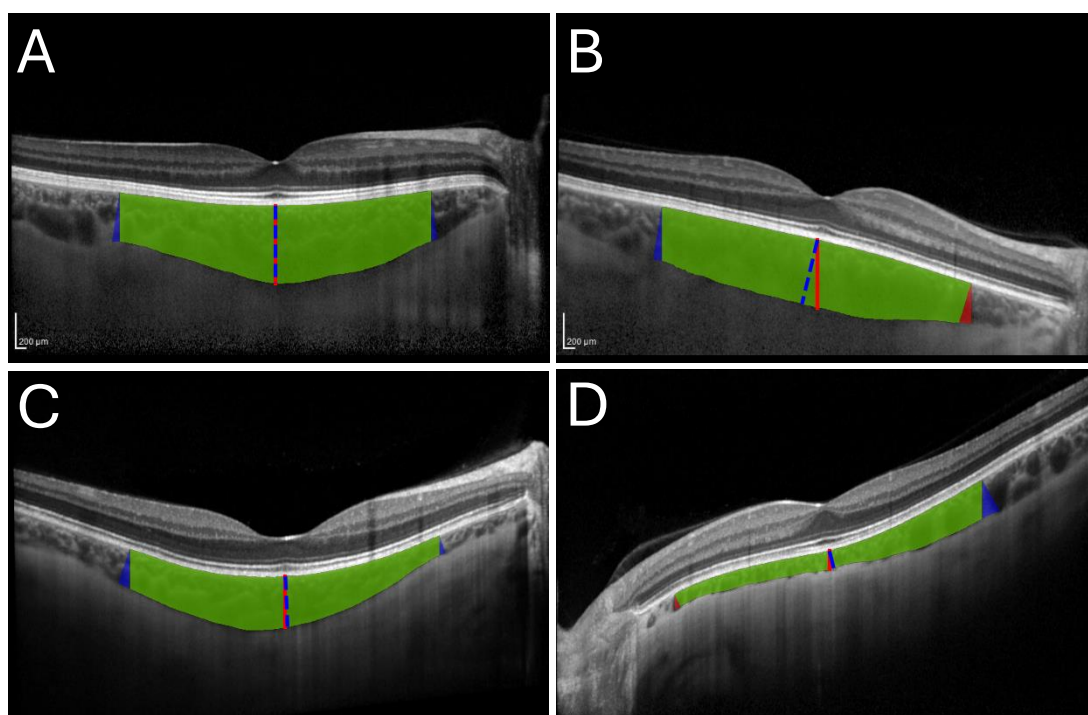
**Figure 1.14:** An EDI-OCT B-scan presented in 1:1 pixel mode (A) and 1:1 micron mode (B).

surement error [194].

Indeed, Lawali, et al. [195] found that different scale types produced significantly different subfoveal retinal thickness measurements when measured manually. However, with the introduction of automatic methods which can segment and subsequently measure the chorioretinal anatomy in a programmatic way, presenting the B-scan using the default '1:1 pixel mode' is sufficient, and is used for the remainder of this thesis. However, it is for images with anisotropic scaling that the anatomic curvature is compensated for at the point of measurement.

### 1.7.1.3 Regions of interest

Historically, retinochoroidal metrics are measured according to the image-axis [13, 18, 21, 64, 106, 112]. That is, the ROI and subsequent measurements are defined and computed per A-scan (column-wise) which most imaging de-



**Figure 1.15:** OCT B-scans with SFCT and ROI overlaid for a flat (A), skewed (B), curved (C) and skewed-and-curved (D) choroids. Lines drawn in solid red and dashed blue for image- and choroid-aligned SFCT. Green pixels represent overlap between the two image- and choroid-aligned ROIs, with red and blue pixels representing regions exclusive to the image- and choroid-aligned ROIs.

vices also adopt. However, this assumes there is little to no curvature in the retinochoroidal tissue in the B-scan. The shape of the patient's eye (degree of myopia) and quality of image capture both contribute to the appearance of the retina and choroid on the OCT B-scan. This is why in this thesis we ensure the ROI, and subsequent measurements, are choroid-aligned, i.e. locally perpendicular to the RPE-Choroid boundary. This ROI is aligned with the predominant axis which the choroid's RPE-Choroid boundary lies along, which can deviate from the lateral image-axis. This deviation is known as *skew* or *curvature*, defined as consistent or inconsistent deviation in the angle of elevation of the choroid from the lateral image-axis, respectively.

Figure 1.15 shows OCT B-scans whose intra-ocular structures exhibit different types of skew or curvature. The choroid in panel (A) appears predominantly

| Choroid curvature | Method of measurement | Choroid measurement    |       |            |                                     |       |            |                        |       |            |         |       |          |
|-------------------|-----------------------|------------------------|-------|------------|-------------------------------------|-------|------------|------------------------|-------|------------|---------|-------|----------|
|                   |                       | SFCT [ $\mu\text{m}$ ] |       |            | Average thickness [ $\mu\text{m}$ ] |       |            | Area [ $\text{mm}^2$ ] |       |            | CVI     |       |          |
|                   |                       | Choroid                | Image | $\Delta_r$ | Choroid                             | Image | $\Delta_r$ | Choroid                | Image | $\Delta_r$ | Choroid | Image | $\Delta$ |
| Flat              | Hyperope              | 424                    | 423   | 0%         | 335                                 | 332   | 1%         | 1.915                  | 1.902 | 1%         | 0.483   | 0.482 | 0.001    |
|                   | Emmetrope             | 424                    | 423   | 0%         | 342                                 | 337   | 1%         | 2.071                  | 2.054 | 1%         | 0.484   | 0.482 | 0.002    |
|                   | Myope                 | 424                    | 423   | 0%         | 361                                 | 351   | 3%         | 2.173                  | 2.138 | 2%         | 0.49    | 0.496 | 0.006    |
| Skew              | Hyperope              | 409                    | 369   | 11%        | 347                                 | 305   | 14%        | 1.847                  | 1.838 | 0%         | 0.543   | 0.544 | 0.001    |
|                   | Emmetrope             | 424                    | 369   | 15%        | 374                                 | 315   | 19%        | 1.914                  | 1.902 | 1%         | 0.554   | 0.555 | 0.001    |
|                   | Myope                 | 448                    | 369   | 21%        | 413                                 | 327   | 26%        | 1.981                  | 1.982 | 0%         | 0.573   | 0.581 | 0.008    |
| Curved            | Hyperope              | 277                    | 278   | 0%         | 222                                 | 204   | 9%         | 1.259                  | 1.246 | 1%         | 0.535   | 0.537 | 0.002    |
|                   | Emmetrope             | 278                    | 278   | 0%         | 239                                 | 213   | 12%        | 1.325                  | 1.305 | 2%         | 0.54    | 0.546 | 0.006    |
|                   | Myope                 | 278                    | 278   | 0%         | 261                                 | 225   | 16%        | 1.406                  | 1.375 | 2%         | 0.566   | 0.567 | 0.001    |
| Skew & Curved     | Hyperope              | 111                    | 100   | 11%        | 168                                 | 127   | 32%        | 0.722                  | 0.695 | 4%         | 0.345   | 0.329 | 0.016    |
|                   | Emmetrope             | 116                    | 100   | 16%        | 179                                 | 127   | 41%        | 0.792                  | 0.762 | 4%         | 0.379   | 0.363 | 0.016    |
|                   | Myope                 | 123                    | 100   | 23%        | 189                                 | 123   | 54%        | 0.776                  | 0.745 | 4%         | 0.349   | 0.338 | 0.009    |

**Table 1.4:** Choroidal measurements for the flat, skewed, curved and skewed-and-curved choroids from figure 1.15, while simulating different refractive errors. Highly myopic: -9.17D, emmetropic: 0D, highly hyperopic: 7.37D.  $\Delta_r$  is absolute, relative difference measured as  $|c-i|/i$  where  $c$  and  $i$  are the choroid- and image-aligned values.  $\Delta$  is absolute difference.

flat, while the skewed choroid in panel (B) shows consistent deviation away from the horizontal image-axis. The choroid in panel (C) is curved and thus shows inconsistent deviation from the horizontal image-axis. In panel (D) we see both skew and curvature.

In figure 1.15, we overlay the image-aligned and choroid-aligned ROIs and SFCT. While the majority of the ROIs overlap (green), there are distinct regions which are mutually exclusive between the ROI definitions which, depending on the extent of skew and/or curvature, can have an impact on downstream choroidal measurements. For example, the skew exhibited by the choroids in panels (B, D) underneath and fovea result in deviation between the lines drawn measuring SFCT when considering different ROIs (red solid line compared to blue dashed-line). This is in contrast to the choroids in panels (A, C), which are locally flat underneath the fovea, resulting in identical lines drawn to measure SFCT regardless of the ROI definition.

The extent of deviation in choroidal measurements was investigated for these four examples by considering a fovea-centred ROI covering a 6000  $\mu\text{m}$  distance and measuring SFCT, average thickness across the ROI, total choroid

area and CVI. For these four eyes, we consider three cases where the patient's eye is highly myopic, emmetropic and highly hyperopic, simulating different lateral pixel length-scales (for a Heidelberg Engineering OCT imaging device). We define emmetropic as when the scan focus is 0, which corresponds to a lateral scaling of  $11.47 \mu\text{m-per-pixel}$ . We define highly myopic and hyperopic as the extremal observations from figure 1.13, i.e. refractive errors (approximated by the scan focus knob in dioptres, D)  $-9.17\text{D}$  and  $7.37\text{D}$ , respectively. These correspond to lateral pixel length-scales of  $13.51 \mu\text{m-per-pixel}$  and  $10.24 \mu\text{m-per-pixel}$ , respectively.

Table 1.4 shows the extent of deviation between the image- and choroid-aligned choroidal measurements. As expected, measurements were stable for the flat choroid, but significant deviations in SFCT and average thickness was observed across the board for the remaining choroids (the curved choroid did not deviate in SFCT because it is locally flat under the fovea as in figure 1.15(C)). Interestingly, deviation in all choroidal measurements (but primarily in SFCT and average thickness) increased as the eye became more myopic. This is due to the anisotropic scaling which becomes more disproportionate with increasing degree of myopia (figure 1.13). This suggests not only the importance of ensuring a consistent ROI definition, but also that these pixel length-scales are taken into account when reporting physical measurements of populations which could have a wide variety of refractive errors.

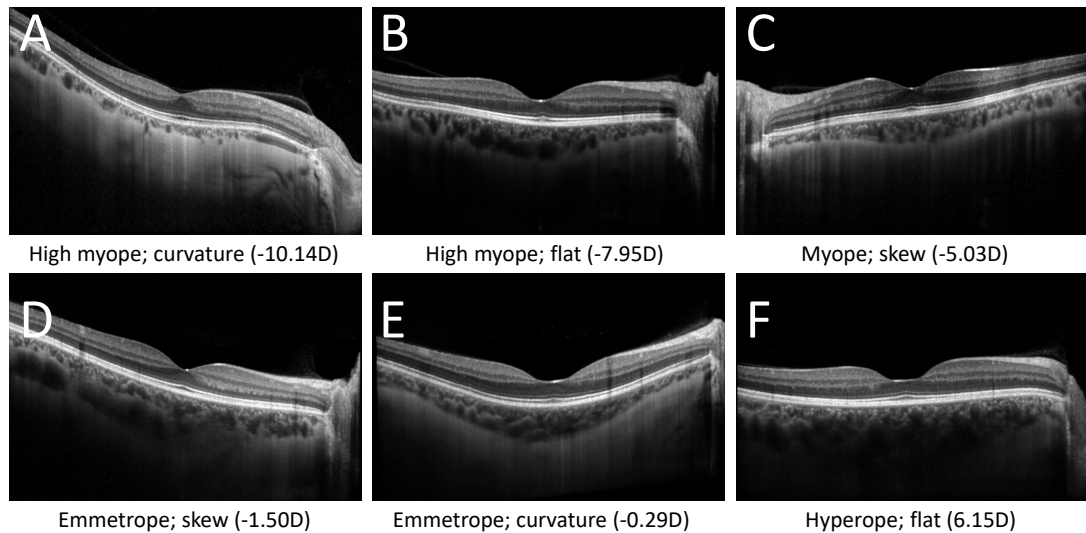
For the skewed choroids, which were not aligned with the image axis underneath the fovea, the average SFCT deviation was 16% ( $37 \mu\text{m}$  in absolute value), and 31% ( $58 \mu\text{m}$  in absolute value) for average thickness across the ROI. Across all choroids, there was a deviation in average thickness of 19% ( $37 \mu\text{m}$  in absolute value) and the largest deviation observed was the skewed choroid for a hypothetically myopic eye, which saw a deviation of 23% and 54% ( $79 \mu\text{m}$  and  $86 \mu\text{m}$  in absolute value) for SFCT and average thickness, respectively.

Two-dimensional measurements like area and CVI appear more robust to these different ROI definitions, with an average absolute difference in CVI of 0.006 (0.6%) across all choroids. For area, choroidal curvature did cause some minor deviation of 2% ( $0.026 \text{ mm}^2$  in absolute value) for the curved choroids.

While the deviations in two dimensional measurements are unlikely to be clinically significant (Agrawal, et al. [28] reported differences in CVI of between 2% – 6% between healthy and diseased eyes), the reported deviations in thickness are. For example, reported effect sizes in choroidal thickness can be as low as 20 – 30  $\mu\text{m}$  when considering myopia progression [110, 24], while changes in SFCT is reported to decrease between 5% – 12% during haemodialysis, as reported by Farrah, et al. [61], and choroidal thinning by approximately 25% in end-stage CKD, as reported by Balmforth, et al. [59].

SFCT is a one-dimensional straight line distance measure which is highly susceptible to changes in anisotropic pixel length-scales. As will be discussed in chapter 2, this anisotropic scaling and difficulty to follow standardised measurement protocol can have an inordinate impact on comparing measurements between and within patients. Two-dimensional measurements showed less bias because of a potential cancellation effect between the different ROIs. For example, in figure 1.15(B), the red and blue shaded regions form two distinct triangles either side of the overlapped portion of each ROI (shaded green region), and thus contribute similar amounts to their respective image- and choroid-aligned area value, respectively. However, in cases where there is choroidal curvature, this cancellation effect is lost and can cause measurement error like the choroid in figure 1.15(C) where area deviated most (table 1.4).

Figure 1.16 shows six OCT B-scans from eyes with varying refractive error — measured using the Oculus Myopia Master (Oculus, Wetzlar, Germany) — visualising the different kinds of curvature and skew that can occur across the spectrum of refractive error. Skewedness of the intra-ocular structures is primarily due to poor acquisition, and curvature from high myopia [196]. However, modern day OCT image capture has added functionality to circumvent choroidal curvature and high myopia [194]. This can be seen in panel (B), whose eye had a refractive error of approximately 8D, but showed no signs of curvature. This is in contrast to the degree of curvature and skew observed in panel (A) for the eye whose refractive error was -10D. The intra-ocular structures are skewed in panels (C) and (D) from poor acquisition, while in panel (E) we see significant curvature for an emmetropic eye (-0.29D). Finally, in panel (F) we see that for this hyperopic eye (6.15D), the choroid is large and the



**Figure 1.16:** A selection of OCT B-scans which show varying curvature and skew, with their refractive error annotated.

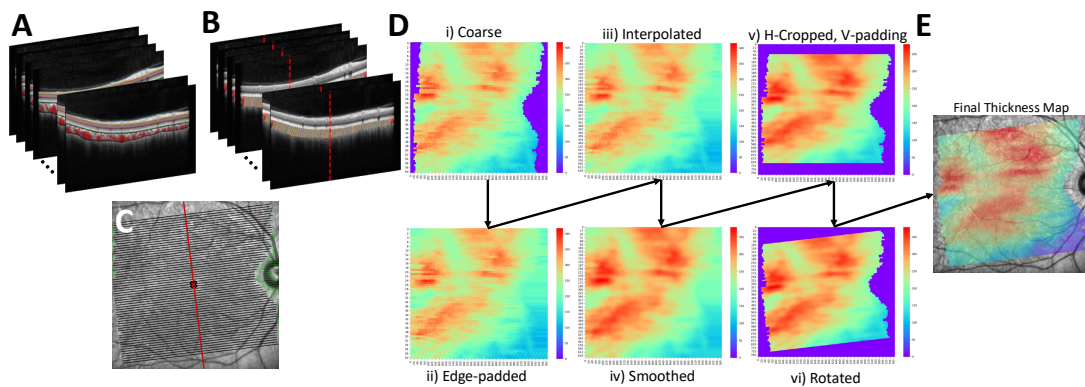
intra-ocular structures are flattened due to the size of the eye. These examples suggest that the extent of curvature and skewedness can occur across the range of refractive errors.

### 1.7.2 OCT volume scan

The introduction of automatic segmentation methods has permitted near real-time segmentation of the choroid across macular OCT volume scans. Measuring the macular choroid is a more representative quantification than subfoveal choroid thickness [23], allowing choroid volume and thickness to be estimated at different locations in the macula.

An OCT volume is made up of several, parallel and linear OCT B-scans spaced evenly across the macula. Given an OCT volume whose choroid has been fully segmented, measurements across the macular choroid can be estimated by interpolating choroid-derived thickness and volume (and vessel density) across the B-scans. This constitutes generating a two-dimensional thickness map which encodes the depth of the choroid at each location on the macula and en face localiser.





**Figure 1.17:** Procedure to generate a choroid thickness map. (A) OCT volume scan with segmentations overlaid. (B) Valid thicknesses measured for the choroid across the volume scan. (C) Thickness array alignment using fovea. (D) Step-wise process to generate thickness map. (E) Final thickness (heat)map overlaid onto en face localiser, with redder colours implying higher depth/density.

The procedure generating the thickness map is shown in figure 1.17(D) and described below. the procedure for measuring choroid vessel maps is analogous.

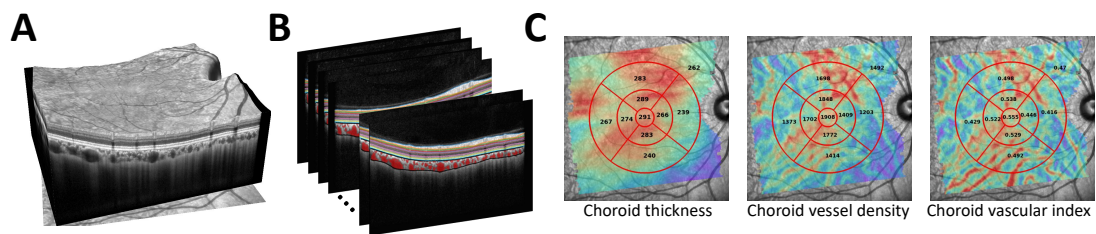
1. For every B-scan, thickness measurements are taken at every available A-scan, measuring the micron distance between the RPE-Choroid and Choroid-Sclera boundary, locally perpendicular to the RPE-Choroid boundary (figure 1.17(B));
2. Each B-scans' thickness array is aligned (figure 1.17(C)) with the lateral position of the fovea detected on the en face localiser (red dotted and solid lines in figures 1.17(B) and 1.17(C), respectively). Thickness array alignment per B-scan is required since the anatomical layer segmentations (black diagonal lines in figure 1.17(C)) do not necessarily cover the full lateral width of the acquisition region of interest (green diagonal lines in figure 1.17(C));
3. The aligned thickness arrays are stacked along the frontal plane to create a coarse, two-dimensional map of thickness values (1.17(Di));
4. This is padded horizontally by duplicating the edge values to prevent edge artefacts from interpolation (1.17(Dii));

## 1.7. Measuring the choroidal microvasculature

5. This is then interpolated to the same pixel resolution as the SLO localiser using bi-linear interpolation (1.17(Diii));
6. Since the raw thickness arrays only sample the depth of the choroid at discrete locations, a Gaussian filter whose standard deviation equals the frontal-plane pixel resolution — the pixel distance between sequential B-scans — is applied for smoothing (1.17(Div));
7. The smoothed map is vertically padded to centre the map onto the fovea of the en face localiser SLO, and the padded edges from step (2) are removed (1.17(Dv));
8. Finally, this map is rotated to the angle of elevation of the acquisition ROI (1.17(Dvi)), as shown by the green and black diagonal lines in figure 1.17(C).

The thickness map (in microns) can then be interpolated to estimate choroidal volume (in  $\text{mm}^3$ ), or can be averaged for more localised measurements. Choroid vessels maps of interest include vessel density in  $\text{micron}^2$ , volume in  $\text{mm}^3$  and CVI (dimensionless).

A standard macular grid was developed over 30 years ago from a landmark, multi-centre clinical trial the Early Treatment Diabetic Retinopathy Study (ETDRS) [197]. The ETDRS grid centres three concentric circles with 1mm, 3mm and 6mm diameters at the foveal pit as seen on the en face localiser, and further splits the 3mm and 6mm circles into four, equally sized quadrants. The



**Figure 1.18:** Diagram of how measurements are computed for a macular OCT volume scan. (A) 3D render of an OCT volume scan. (B) Sequential B-scans with segmentations overlaid. (C) Choroid thickness, vessel density and CVI (heat)maps, with average measurements along the ETDRS grid overlaid (redder colours imply higher depth/density).



quadrants are aligned with the angle of elevation of the OCT volume's acquisition ROI. Average thickness, vessel density, CVI and interpolated volumes are measured in these 9 sub-fields. Figure 1.18 shows representative choroid thickness, vessel density and CVI maps across the macula, generated from an OCT volume scan, superimposed onto the en face localiser with measurements in each sub-field of the ETDRS grid.

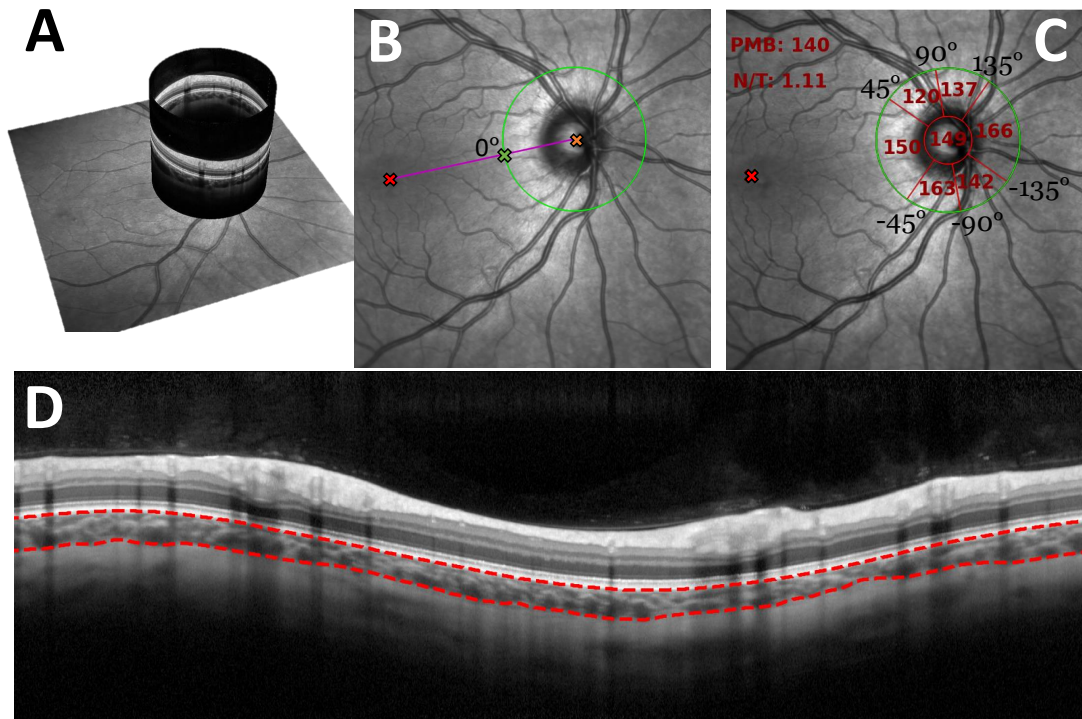
### 1.7.3 OCT peripapillary scan

A peripapillary B-scan is a circular scan centred at the optic disc and is continuous along the frontal plane (the lateral ends of the B-scan are continuous). Figure 1.19 shows the procedure to obtain average choroid thickness in each of the peripapillary sub-fields, often used for assessing glaucoma [198]. These sub-fields measure chorioretinal thickness spatially around the optic disc, as the circular B-scan traverses the disc's circumference at an offset. Panel (A) shows the en face localiser with the circular B-scan overlaid.

Upon choroid layer segmentation of the choroid in a peripapillary OCT B-scan (figure 1.19(D)), we first generate a thickness array. This measures a micron distance for every A-scan (column). We do not measure locally perpendicular to the RPE-Choroid boundary because the sinuous nature of the peripapillary B-scan is a result of the circular-shaped acquisition line, and not natural curvature from the anatomy. This is in contrast to the linear, macular line-scans previously discussed.

To create the peripapillary grid, we first define the centre (0 degrees) of the temporal peripapillary sub-field (figure 1.19(B)). This is done by cross-referencing the point along the circular acquisition region (on the en face localiser) which is co-linear with the optic disc centre and fovea centre. This position has a corresponding lateral position along the circular B-scan. If the en face localiser is unavailable, then the centre is taken as the lateral centre of the B-scan. However, without the en face localiser, the peripapillary grid will likely be off-centre which could have significant clinical consequences [199].

Using this temporal centre, the thickness array can then be divided into tempo-



**Figure 1.19:** Diagram demonstrating how peripapillary thickness measurements for an OCT peripapillary B-scan are made. (A) Circular OCT B-scan overlaid onto the en face localiser. (B) Detection of the centre of the temporal peripapillary grid. (C) Peripapillary grid definition on the en face localiser, with average choroid thickness values overlaid per sub-field. (D) Circular OCT B-scan with choroid layer segmentation overlaid.

ral (-45 – 45 degrees), supero-temporal (45 – 90 degrees), supero-nasal (90 – 135 degrees), nasal (135 – -135 degrees), infero-nasal (-135 – -90 degrees) and infero-temporal (-90 – -45 degrees) sub-fields. Average thickness values are taken per sub-field, as well as the nasal-to-temporal (N/T) ratio and papillo-macular bundle (PMB) (-30 – 30 degrees), with a final average over the whole B-scan. Figure 1.19(C) shows these values overlaid onto the peripapillary grid, with the N/T and PMB values shown in the top-left.

#### 1.7.4 Comparing measurements

Table 1.5 defines and describes the metrics used throughout this thesis to compare agreement between two sets of measurements, either choroid-derived

measurements or segmentation maps outputted from a segmentation model. This section serves as a brief description of these metrics for the reader. For choroid-derived metrics, consider sets  $x$  and  $y$  to be of length  $n$ . For segmentation metrics, we define the  $2 \times 2$  confusion matrix  $C$  which compares the agreement between two binary segmentation maps as

$$C = \begin{pmatrix} TP & FP \\ FN & TN \end{pmatrix}. \quad (1.1)$$

Upon comparing a predicted, binary segmentation map  $S_p$  to the true segmentation map  $S_t$ , we define  $TP$  and  $TN$  as the number of true positives and negatives, i.e. the sum of all positive (1) and negative (0) matches between maps  $S_p$  and  $S_t$ , respectively.  $FP$  is the number of false positives, i.e. where  $S_p$  is equal to 1 but  $S_t$  is equal to 0, and vice versa for  $FN$ , the number of false negatives.

#### 1.7.4.1 Choroid-derived metrics

For comparing choroid-derived measurements, we use Pearson  $r$  and Spearman  $s$  correlation coefficients, mean difference, mean absolute error, and element-wise measurement noise  $\lambda$ . Pearson  $r$  measures the linear relationship between measurements in  $x$  and  $y$ , while Spearman  $s$  measures their monotonic relationship. In table 1.5,  $\text{Cov}(x, y)$  is the covariance between measurement sets  $x$  and  $y$ ,  $\sigma(x)$  measures the standard deviation of the values in  $x$  and  $d_i$  measures the difference in rank between  $x_i$  and  $y_i$ , after ordering sets  $x$  and  $y$ . Pearson's  $r$  assumes a linear relationship between  $x$  and  $y$ , while Spearman does not. On the assumption that  $x$  and  $y$  are measuring the same thing, then there should by definition hold a linear relationship. Anything other than a perfect linear relationship would imply disagreement between the approaches which computed the measurements in  $x$  and  $y$ . In the case where a perfect linear relationship exists, Pearson and Spearman correlation coefficients are equal.

While  $r$  and  $s$  provides a dimensionless representation of agreement between  $x$  and  $y$ , this sometimes lacks clear interpretation on how close  $x$  and  $y$  match

## 1.7. Measuring the choroidal microvasculature

each other. Mean difference (MD) and mean absolute error (MAE) are more interpretable measures of agreement between  $x$  and  $y$ . This is because they are measured in the same units as the original data. However, MAE is a more robust metric to assess agreement than MD. This is because errors are averaged using their absolute value, so they will contribute positively toward the error regardless of their sign (positive/negative). Conversely, MD has the potential to cancel out error if their magnitude is equal but with opposite signs, and this cancellation effect also has the effect of shrinking the systematic bias estimate of asymmetric error distributions. Thus, MAE provides a more accurate and reliable reflection of the magnitude of error compared to MD, making it a more informative metric.

| Name   | Range    | Formula   | Optimum | Description                               |
|--|----------|---|---------|---|
| <u>Choroid-derived metrics</u>                 |          |   |         |   |
| Pearson ( $r$ )                                | [0, 1]   | $\text{Cov}(x,y)/\sigma(x)\sigma(y)$  | High    | Linear correlation                        |
| Spearman ( $s$ )                               | [0, 1]   | $1 - 6/n(n^2-1) \sum_{i=1}^n d_i^2$   | High    | Monotonic correlation                     |
| Mean Difference (MD)                           | Any      | $1/n \sum_{i=1}^n (x_i - y_i)$  | Low     | Prediction error                          |
| Mean Absolute Error (MAE)                      | $\geq 0$ | $1/n \sum_{i=1}^n  x_i - y_i $  | Low     | Unsigned prediction error                 |
| Element-wise measurement noise ( $\lambda_i$ ) | $\geq 0$ | $\sigma([x_i, y_i])/\sigma(x)$  | Low     | Proportional measurement variability      |
| <u>Segmentation metrics</u>                    |          |   |         |   |
| Precision                                      | [0, 1]   | $\text{TP}/(\text{TP}+\text{FP})$   | High    | Positive predictive value                 |
| Recall   | [0, 1]   | $\text{TP}/(\text{TP}+\text{FN})$   | High    | True positive rate (sensitivity)          |
| Dice coefficient                               | [0, 1]   | $2(\text{Precision} \times \text{Recall})/(\text{Precision} + \text{Recall})$ | High    | Harmonic mean of precision and recall     |
| Area-under-curve (AUC)                         | [0, 1]   | NA  | High    | True positive to false positive trade-off |

**Table 1.5:** Definition and brief description of metrics used throughout this thesis for choroid-derived measurements and for segmentation maps.

The aforementioned metrics are a population-level description of agreement between  $x$  and  $y$ . However, it is of particular interest to measure an individual-level description of agreement, which is what  $\lambda$  achieves, introduced by Engelmann, et al. [200] in his study on assessing the repeatability of retinal fractal dimension in oculomics. Say that each element of  $x$  and  $y$  represents a measurement of an individual eye. Then

$$\lambda = \frac{\text{SD of within-eye measurements}}{\text{SD of between-eye measurements}}. \quad (1.2)$$

Specifically, from the equation in table 1.5, for the  $i^{\text{th}}$  eye the measurement noise is

$$\lambda_i = \frac{\sigma([x_i, y_i])}{\sigma(x)} \quad (1.3)$$

$$= \frac{1/\sqrt{2}|x_i - y_i|}{\sigma(x)}, \quad (1.4)$$

where the numerator,  $\sigma([x_i, y_i])$ , is the within-eye standard deviation between values  $x_i$  and  $y_i$ , which is simply the range between  $x_i$  and  $y_i$ . The denominator is the standard deviation across all eye measurements in  $x$ , arbitrarily chosen between  $x$  and  $y$ . This ratio expresses the size of the error for the  $i^{\text{th}}$  eye in units of the overall population variability, i.e. how many times the size of the error for the  $i^{\text{th}}$  eye is contained within the overall population variability. This is particularly helpful to assess the effect size of the measurement error at the subject level.

Lower values of  $\lambda_i$  are desirable, indicating better reproducibility and less noise in individual measurements.  $\lambda_i=0$  is the optimal value, representing perfect precision for the  $i^{\text{th}}$  eye's measurements in  $x$  and  $y$ . Values where  $\lambda_i < 1$  would mean that the within-eye variability is less than the between-eye variability, indicating greater measurement reliability and consistency at the eye-level. A smaller  $\lambda_i$  suggests the measurement process between  $x$  and  $y$  is stable such that any observed variability is more attributable to differences between eyes rather than the measurement error within the  $i^{\text{th}}$  eye. Computing  $\lambda_i$  for all  $i = 1, \dots, n$  outputs a distribution of eye-level measurement error, which can be represented with high-level summary statistics.

Note that this metric is particularly helpful for differentiating between measurement error and the overall measured effect in a study population. Say  $x$  and  $y$  represent the same measurements produced by the same method, thus they are *repeated* measurements. For a given population (with repeated data to generate  $x$  and  $y$ ), the average measurement noise  $\bar{\lambda} = 1/n \sum_i^n \lambda_i$  represents the average individual-level error relative to the population variability. Using the same population for a particular research objective, such as measuring the effect of a drug or treatment, if the measured effect relative to the population variability is greater than the estimated  $\bar{\lambda}$ , you can be confident that measurement error does not fully explain the observed signal, and thus what remains is true biological change. This kind of interpretation is crucial for ensuring confidence in an observed effect in research studies.

A final consideration is when there are repeated measurements per eye to take into account (as will be the case in this thesis). During each examination, multiple measurements  $x_i^{(j)}, y_i^{(j)}$  for  $j = 1, \dots, m$  may be taken for each eye  $i = 1, \dots, n$ . For example, an OCT volume scan may be measured in the 9 sub-fields of the ETDRS grid [197] ( $m = 9$ ). In this case, we would average over the  $m$  measurements to create a representative value for the  $i^{\text{th}}$  eye for both sets of measurements,  $\bar{x}_i, \bar{y}_i$ , where

$$\bar{x}_i = 1/m \sum_j^m x_i^{(j)}, \quad (1.5)$$

for  $i = 1, \dots, n$  and similarly for  $\bar{y}_i$ . Then we arrive at our setup from before, measuring the eye-level measurement noise for all  $n$  eyes using a single comparative pair of values in  $\bar{x}$  and  $\bar{y}$ .

### 1.7.4.2 Segmentation metrics

Standard segmentation metrics of interest to compare a predicted segmentation map and ground truth segmentation map are *precision*, *recall*, the *Dice* similarity coefficient and the area under the receiver operating characteristic (AUCROC) curve (area-under-curve, *AUC*).

Given a predicted binary segmentation map, compared to the true segmenta-

tion mask, precision measures the positive predictive value of your segmentation task. Thus, precision measures the proportion of correctly labelled pixels against all pixels positively labelled in the predicted map. For choroidal space/vessel segmentation, precision measures the extent of over-segmentation in the model. Recall is the sensitivity of the segmentation model and is the true positive rate, measuring how accurate the model is at predicting positive instances overall. For choroidal space/vessel segmentation, recall measures the extent of under-segmentation in the model [201].

In machine learning, there is an inverse relationship between precision and recall, commonly referred to as the precision-recall trade-off. Increasing precision typically reduces recall and vice versa, because optimising for one often comes at the expense of another. High precision is achieved by making more conservative positive predictions, which reduces the  $FP$  count but may increase the  $FN$  count, lowering recall. Conversely, taking a liberal approach to positive predictions (decreasing  $FN$ ) increases recall, but may lower precision (increasing  $FP$ ) [202].

Instead of reporting precision and recall separately, the Dice similarity coefficient combines precision and recall to give an overall representation of predictive performance for a segmentation model. In particular, the Dice coefficient will measure the extent of over- and under-segmentation of a predicted binary segmentation map relative to its ground-truth segmentation map. It is referred to as the *harmonic* mean between precision and recall, as it assumes an equal weighting between precision and recall [203].

So far, we have considered segmentation metrics for a predicted binary segmentation map. However, segmentation models often output a probabilistic segmentation map, where elements take on values in the range  $[0, 1]$ . This map represents the degree of confidence the model has in predicting a positive label for each pixel. The map can be binarised through selecting a threshold value  $T$  between  $[0,1]$ , where precision, recall and Dice coefficient can then be computed. Often we wish to measure the overall model performance using these probability values, rather than the thresholded binary map, and the AUC does exactly this.

AUC measures the trade-off between the true positive rate (recall) and false positive rate ( $FP/FP+TN$ ) across different thresholds  $T$  in  $[0, 1]$ . For choroidal space/vessel segmentation, AUC provides a single value representing how well the segmentation model's raw predictions are at differentiating between the choroid/vessel and the rest of the OCT B-scan. Note that all aforementioned segmentation metrics here are bound between 0 and 1, with higher values indicating better performance.

## 1.8 Executive summary

In this chapter, we introduced the choroid, a crucial vascular layer of the eye which has historically been neglected due to poor imaging quality. However, recent advancements in OCT, particularly with EDI-OCT and SS-OCT, have improved the ability to visualise and measure this microvascular bed. This has contributed to new investigations into its role in systemic disease given emerging evidence of its ability to reflect microvascular dysfunction in hypertensive conditions, renal function, neurodegeneration and systemic inflammation.

We also observed that the choroid's complex vascular structure presents significant challenges in measurement due to imaging issues like speckle noise, low signal-to-noise ratio, retinal vessel shadowing and ambiguous vessel wall definition. Current approaches to measure the choroid often rely on manual measurement or improper usage of open-source semi-automatic tools, which introduces variability across and within patients, limiting reproducibility and accuracy. Other, more accurate methods remain closed-source, preventing standardisation of choroidal measurements. This emphasises the need for open-source, reproducible methods to enable consistent and accurate measurement of the choroid.

We also introduced choroidal measurements of interest and found that measurement error can arise from differences in definitions of the Choroid-Sclera boundary, measurement protocol, or image characteristics such as skew and curvature. Therefore, a strict protocol accounting for factors like the presentation of intra-ocular structures on the B-scan is essential for consistent and stan-



standardised measurements of the choroid. This is especially pertinent when pixel length-scales are subject to change across individuals. Otherwise, this has the potential to exacerbate potential measurement error in one-dimensional metrics such as choroid thickness.

## Chapter 2

# GPET: Semi-automatic choroid region segmentation using Gaussian process regression

### 2.1 Introduction

In this chapter, we have a particular interest in segmenting the choroidal space from an OCT B-scan. Here, we introduce a method which takes a traditional approach to semantic image segmentation using an edge-based model based on Bayesian machine learning. That is, we formulate the problem of segmenting the choroidal space in an OCT B-scan as identifying the upper and lower boundaries delineating the choroid, i.e. the RPE-Choroid and Choroid-Sclera boundaries. Advantageously, this method leverages uncertainty quantification and non-parametric learning using a Bayesian modelling approach known as Gaussian process regression.

Semantic image segmentation is an important tool for deriving information from an image, and in medical imaging plays a crucial role in many applications, facilitating higher-level image analysis such as feature measurement [204] or disease detection [205]. The goal of any image segmentation algorithm is to partition an image into meaningful sub-regions. Edge-based segmentation algorithms do this by identifying boundaries, or edges between those regions.

Human visual systems are able to interpret and recognise scenes and objects just based on the basic structure of their outlines [206], making edge-based approaches useful in mimicking some of the mechanisms of our own vision.

The most popular frameworks to employ semantic segmentation in medical imaging utilise deep learning and CNNs. However, this can be a challenge when high quality annotated datasets are limited, which was the case at the time of developing this algorithm. Nevertheless, even if data was indeed available then the ground labels generated for model development would have been done manually. This has the potential to inject human bias into the model development and segmentation procedure, particularly for manual detection of the Choroid-Sclera boundary [189].

Two common types of models suited for tracing individual edge structures in an image like an OCT B-scan are those which consider the pixels in an image as nodes of a graph, and those which adapt a contour to an edge structure using splines or level sets. Note that both of these methods have been used previously for semi-automatic choroid region segmentation (table 1.1). However, and in particular to the Choroid-Sclera boundary, these approaches are typically unable to capture uncertainty or incorporate prior knowledge into the segmentation procedure.

The use of Bayesian frameworks in computer vision has been popularised due to their non-parametric learning and uncertainty quantification. Some image processing applications involve denoising [207], object categorisation [208], detection of pneumonia in chest X-rays [209] or super-resolution of images [210]. However, there has not been an extensive amount of work on applying Gaussian processes to image segmentation. Freytag, et al. [211] performed semantic segmentation by using histogram intersection kernels for fast and exact Gaussian process classification. Tu and Zhu [212] proposed a Bayesian statistical framework to image segmentation by learning object appearance models for different region types, placing prior distributions over region size, region number and boundary smoothness. Simek and Barnard [213] used a two-dimensional Gaussian process regression model to segment Arabidopsis leaves by modelling the blade and petiole as two random functions, joining them at their boundaries using a smoothing constraint.

The method introduced in this chapter, Gaussian Process Edge Tracing (GPET), provides a tunable, robust and probabilistic approach to edge-based image segmentation of the choroid in OCT B-scans. GPET models an edge of interest (EOI) in an image as a Gaussian process in a regression setting. Using a suitable estimate of the EOI using gradient-based pre-processing steps and tunable initialisation, the model iteratively searches for edge pixels which it uses to update its belief on the location of the true EOI. Moreover, the internal parameters of the model allows the user to tune the fitting procedure to the properties of the EOI.

### 2.1.1 Gaussian process regression

Gaussian processes are a Bayesian machine learning technique used to perform regression on a set of observations. Bayesian methods are able to treat the learning process incrementally by updating their beliefs in regions near these observations once they have been incorporated into the model. While providing precision in regions near observations, Gaussian processes also provide a way of capturing the uncertainty in regions farther away from these observations, where the model has little information to learn from. The modelling of this uncertainty induces a distribution of functions that plausibly fits the set of observations.

For Gaussian processes to be used in a regression setting, we typically have some noisy observations  $y$  which correspond to some underlying function  $f$  which we want to model. A Gaussian process regressor models this underlying function  $f$  as a Gaussian process and assumes the following relationship

$$y(x) = f(x) + \epsilon, \quad (2.1)$$

where  $\epsilon$  represents the underlying noise — assumed to be Gaussian — perturbing the true function values  $f(x)$  (which we *don't* have access to during modelling) to the noisy observations  $y$  (which we *do* have access to during modelling). Formally, the identically, independently distributed Gaussian noise  $\epsilon \sim \mathcal{N}(0, \sigma_y^2)$  is characterised by the observation noise variance  $\sigma_y^2$  which gov-

erns the uncertainty that the model assumes in the set of observations  $y(x)$ .

We can think of any continuous, real-valued function  $f$  as being an infinitely long list, extending to the far ends of the number line (of real values), i.e. a list  $\{f(x) : x \in (-\infty, \infty)\}$  defined at input values  $x$  within the open interval  $(-\infty, \infty)$ . However, to prevent dealing with the mathematics of infinity, we can consider the function  $f$  evaluated at inputs  $x$  which comes from a large, finite discretisation of this infinite input space,  $X$ , yielding a very long list of function values  $\{f(x) : x \in X\}$ .

Although, while we still can't deal mathematically with the extortionate size of this list, we can however place a multivariate Gaussian distribution on them. This distribution exerts a probabilistic belief on what these function values look like and how pairwise function values relate (or co-vary) with each other. This induces a *Gaussian process distribution* which is analogous to a stochastic process — a generalisation of a joint probability distribution of random variables to a probabilistic distribution of functions [214]. Simply put, a Gaussian process is a multivariate Gaussian distribution over function values in the very long list  $\{f(x) : x \in X\}$ .

By modelling  $f$  as a multivariate Gaussian distribution, the function values  $f(x)$  become random values, and thus  $f$  becomes a random function which can be fully specified by a mean function  $m(x)$  and covariance function  $k(x, x')$ , defined by

$$\mu(x) = \mathbb{E}[f(x)]; \tag{2.2}$$

$$k(x, x') = \mathbb{E}[(f(x) - \mu(x))(f(x') - \mu(x')))], \tag{2.3}$$

where  $x, x'$  are inputs,  $\mu(x)$  is the expected (average) value of the random function  $f$  at input  $x$ , and covariance  $k(x, x')$  measures the correlation value between the random function values at inputs  $x$  and  $x'$ .

With these ingredients, the problem formulation can be formalised as a random function  $f(x)$  modelled by a Gaussian process such that

$$f(x) \sim \mathcal{GP}(\mu(x), k(x, x')). \tag{2.4}$$

In relation to segmenting the RPE-Choroid (upper) and Choroid-Sclera (lower) boundaries in an OCT B-scan, we consider the input values  $x$  to be the column indexes (the horizontal axis) of the B-scan,  $f$  to be the true row locations of the upper (or lower) boundary at inputs  $x$ , and  $y$  to be some subset of row locations which we use to model the Gaussian process as closely to  $f$  as possible. Thus, since we are working with one-dimensional column and row indexes, we can assume for the following high-level description that inputs  $x, x'$  and outputs  $f(x), y(x)$  are all one-dimensional values. Consider chapter 2.2 of Rasmussen and William's book for more generalisable details [214].

### 2.1.1.1 Prior distributions and covariance functions

Using Bayesian statistics to fit functions requires constructing a *prior distribution* for the function values. This prior distribution encodes our initial belief of the properties exhibited by the function  $f$  that we wish to fit to the data  $y$ . The choice of prior reduces to choosing an initial mean function  $\mu(x)$  and covariance function  $k(x, x')$ . However, we can assume without any loss of generality that  $\mu(x) = 0$  for all inputs  $x$ , because it is the covariance function which has the largest impact on model specification and fitting.

The covariance function is used to build the covariance matrix between random function values and typically depends on the input observations of these function values. The choice of covariance function determines the properties of the random functions drawn from the Gaussian process distribution. Considering that this approach is being used to model boundaries separating biological tissue, these are typically continuous (or approximately smooth). This is an essential property we want to encode in our covariance function.

Therefore, because we want to model continuous functions, function values of nearby inputs should be of similar value and hence have large correlation. Thus, a covariance function is chosen which measures the similarity between function values based on the distance between their input locations. These are known as isotropic stationary covariance functions, which are translation invariant and only depend on the distance between inputs  $x$  and  $x'$ .

The stationary kernels which are of interest in this chapter are the squared exponential and Matern covariance functions. The squared exponential, also known as the radial basis covariance function (RBF), produce smooth and continuous functions. For input locations  $x, x'$ , the squared exponential kernel is defined as

$$k(x, x') = \sigma_f^2 \exp\left(-\frac{1}{2\sigma_l^2}r^2\right), \quad (2.5)$$

where  $r = |x - x'|$ , i.e. the absolute difference between inputs  $x$  and  $x'$ . Importantly, this function value is largest when  $x$  is closest to  $x'$ , i.e. when the input locations of the function values  $f(x)$  and  $f(x')$  are closest to each other. There are two parameters  $\sigma_f$  and  $\sigma_l$  which act as the typical deviation (amplitude) from the mean function and sinuosity (curvature) of the random functions drawn from the Gaussian process, respectively. Formally,  $\sigma_f$  and  $\sigma_l$  are the function variance and length-scale hyperparameters of the covariance function, respectively.  $\sigma_f$  dictates the typical amplitude of the prior functions, where approximately 66% of the values making up each random function lie within  $\pm\sigma_f$ .  $\sigma_l$  is the length-scale which roughly describes the typical distance between turning points of the random function.

RBFs produce very smooth curves when sampling from the Gaussian process (because this function is infinitely differentiable), but its strong assumption on smoothness and continuity is one of its pitfalls when modelling real-life physical processes. For example, the Choroid-Sclera boundary is not always perfectly smooth, especially in the presence of retinal pathology. A collection of more flexible covariance functions is the Matern class, of which for our purposes we consider only two types, defined as

$$k_{\nu=3/2}(x, x') = \sigma_f^2 \left(1 + \frac{\sqrt{3}r}{\sigma_l}\right) \exp\left(-\frac{\sqrt{3}r}{\sigma_l}\right) \quad (2.6)$$

$$k_{\nu=5/2}(x, x') = \sigma_f^2 \left(1 + \frac{\sqrt{3}r}{\sigma_l} + \frac{5r^2}{3\sigma_l^2}\right) \exp\left(-\frac{\sqrt{5}r}{\sigma_l}\right). \quad (2.7)$$

Here,  $\nu$  is a positive parameter that dictates the differentiability, and therefore continuity, of the random functions. Rasmussen and Williams consider  $\nu = 3/2$  and  $\nu = 5/2$  to be particularly useful for modelling machine learning problems [214]. The generalised formula for the Matern class, as well as many other

examples of covariance functions, can be found in chapter 4.2 of Rasmussen and Williams's book [214].

The use of  $r$  in these formulas means that Gaussian processes which use these covariance functions will generate function values with higher covariance when their input locations lie close to each other. This property (and the differentiability of the covariance functions) defines how smooth the random functions drawn from the Gaussian process are.

Once a covariance function has been chosen, assuming a trivial mean function, then the initial prior distribution of the Gaussian process is fully defined and we can draw *realisations* from it, which are just random samples drawn from a multivariate Gaussian distribution. However, we interpret these random samples as random functions. Define the prior distribution of the Gaussian process,  $f(x)$ , such that

$$f(x) \sim \mathcal{GP}(0, k(x, x')). \quad (2.8)$$

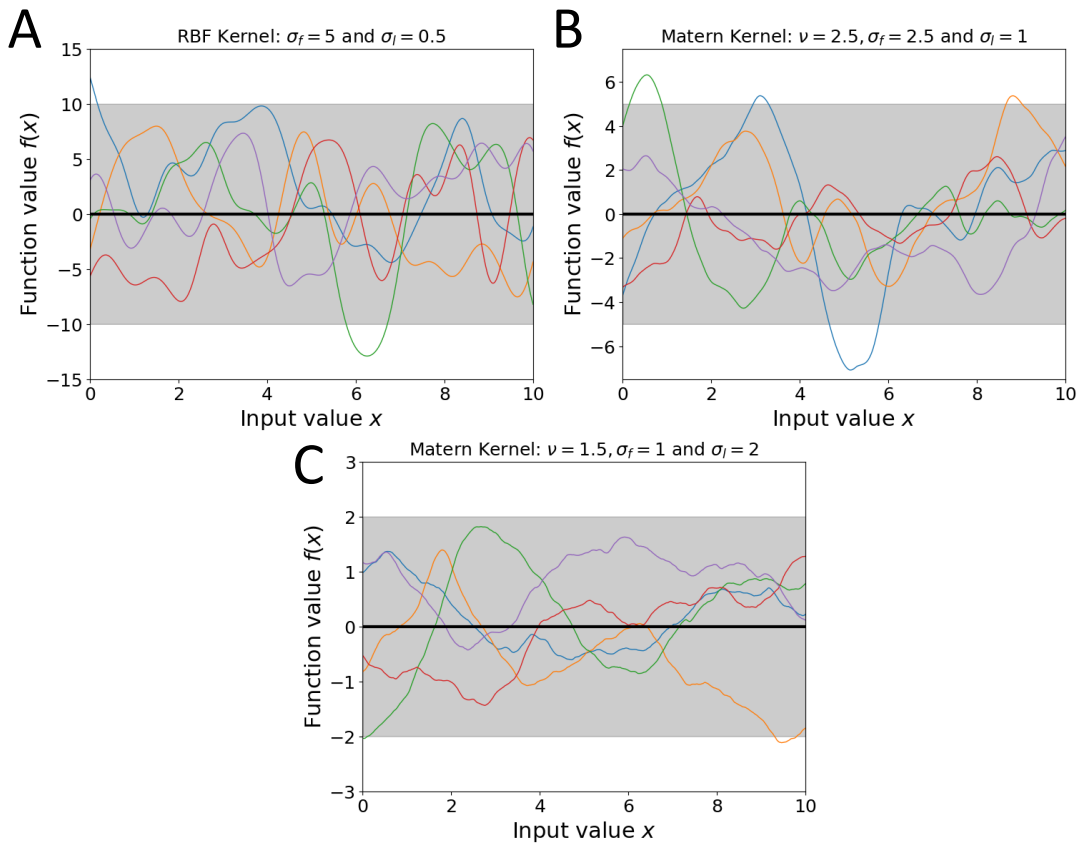
Samples can then be drawn from this prior distribution by defining input points  $X_* = \{x_i : i = 1, \dots, n\}$  and drawing a random list of function values  $\{f_*(x_i) : i = 1, \dots, n\}$ , or random function  $f_*$ , from a multivariate Gaussian distribution with an  $n \times n$  covariance matrix  $K(X_*, X_*)$  such that

$$f_* \sim \mathcal{N}(0, K(X_*, X_*)), \quad (2.9)$$

where  $K(X_*, X_*)$  is a square  $n \times n$  matrix whose  $(i, j)^{\text{th}}$  element is the covariance  $k(x_i, x_j)$  between function values  $f(x_i)$  and  $f(x_j)$ , dependent on inputs  $x_i, x_j$ .

Figure 2.1 shows 3 example Gaussian processes using the RBF (A) and Matern (B – C) covariance functions, with varying settings of  $(\sigma_f, \sigma_l)$  (and  $\nu$  for the Matern covariance function). Here, we've defined  $X_*$  to be the interval  $[0, 10]$  and drawn five random functions from their respective prior distributions. The shaded regions represent the point-wise mean-function value for each input location  $\pm 2$  standard deviations, corresponding to each prior distributions' 95%





**Figure 2.1:** Five random functions drawn from the prior distribution of three different Gaussian processes, each with a different kernel and hyperparameters.

credible region<sup>1</sup> — note that the mean function (solid black line) is 0 by design as  $\mu(x) = 0$  for all  $x$ . In these examples, the 95% credible region of each Gaussian process is a rectangular area with its length spanning the input range since there are no observations fitted to the model (as is the case for a prior distribution). Note that different settings of  $(\sigma_f, \sigma_l)$  (and  $\nu$ ) yield different curvatures and amplitudes of the random functions drawn from each Gaussian process, while the different covariance function changes the smoothness of the random functions.

<sup>1</sup>This is the preferred terminology for a Bayesian model's 95% confidence interval.

### 2.1.1.2 Posterior distributions

In the aim of model-fitting its now in our interest to update our Gaussian process by incorporating a set of observations  $\{(x_i, f(x_i)) : i = 1, \dots, n\}$  representing the true function  $f$ . However, the observations are typically corrupted by an element of noise in real-world problems. Denote these noisy observations as  $\mathcal{D} = \{(x_i, y(x_i)) : i = 1, \dots, n\}$  where  $y(x_i) \sim \mathcal{N}(f(x_i), \sigma_y^2)$ , where we assume  $\sigma_y^2$  is fixed and known and that  $y(x_i)$  deviates around the underlying true function value  $f(x_i)$  for all inputs  $x_i$ . By integrating these noisy observations into our Gaussian process, we update our probabilistic belief about the true underlying function which induces a *posterior distribution*.

This posterior distribution allows us to make predictions, or sample random functions as we did in the previous section, using the properties imposed by the covariance function *and* after taking into account the noisy function values,  $y$ , we've observed. In order to do this, we first form the joint distribution between  $y$  and the function values we want to know (or predict/sample at),  $f_*$ .

Define vector  $\mathbf{y} = \{y(x_i) : i = 1, \dots, n\}$  as the list of observed function values at locations  $X = \{x_i : i = 1, \dots, n\}$ . Similarly define  $\mathbf{f}_* = \{f_*(x_j) : j = 1, \dots, n_*\}$  as the function values we want to predict at locations  $X_* = \{x_j : j = 1, \dots, n_*\}$ . This joint distribution is defined as

$$p(\mathbf{y}, \mathbf{f}_*) = \mathcal{N}\left(\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix}; \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} K(X, X) + \sigma_y^2 \mathbb{I}_n & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix}\right), \quad (2.10)$$

where  $\mathbb{I}_n$  is the  $n \times n$  identity matrix. The covariance matrix of this joint distribution is a block matrix comprising of 4 matrices computed using the chosen covariance function and the input locations of the observations,  $X$  and input locations we want to predict/sample at,  $X_*$ .

The posterior distribution of the Gaussian process is then induced by conditioning on the known, but noisy, observations. We restrict the samples we can draw from the joint distribution to only describe functions which accommodate these noisy observations. The posterior distribution of this Gaussian process,

after observing our noisy observations, is defined as

$$p(\mathbf{f}_* : \mathbf{y}) = \mathcal{N}(\mathbf{f}_*; \mathbb{E}[\mathbf{f}_*], \text{Cov}[\mathbf{f}_*]); \quad (2.11)$$

$$\mathbb{E}[\mathbf{f}_*] = K(X_*, X)[K(X, X) + \sigma_y^2 \mathbb{I}_n]^{-1} \mathbf{y}; \quad (2.12)$$

$$\text{Cov}[\mathbf{f}_*] = K(X_*, X_*) - K(X_*, X)[K(X, X) + \sigma_y^2 \mathbb{I}_n]^{-1} K(X, X_*). \quad (2.13)$$

The additive diagonal matrix  $\sigma_y^2 \mathbb{I}_n$  is used to perturb the posterior functions away from the observations (and also for numerical stability during sampling). This conditioning updates the mean function and covariance matrix shown by equations (2.12) and (2.13), respectively. This new distribution extracts an update on our belief on the mean and covariance of the unobserved function values  $\mathbf{f}_*$  of the Gaussian process. The mathematical details of constructing this posterior distribution is shown in appendix A.2 of Rasmussen and William’s book [214]. The second term in  $\text{Cov}[\mathbf{f}_*]$  implies that with the more data observed, the more certain we are of predictions — especially for function values whose input locations are near the observations’ input locations.

Typically in Bayesian machine learning, the hyperparameters of the covariance function  $\sigma_f, \sigma_l$  and observation noise  $\sigma_y$  are also updated to accommodate the new observations in the posterior distribution<sup>2</sup>. Hyperparameter optimisation collates all kernel hyperparameters and observation noise into  $\theta$ , and these are optimised through maximising the log marginal likelihood of the observation set,

$$\theta = \underset{\theta}{\text{argmin}} \left\{ \log p(\mathbf{y} | X, \theta) \right\}; \quad (2.14)$$

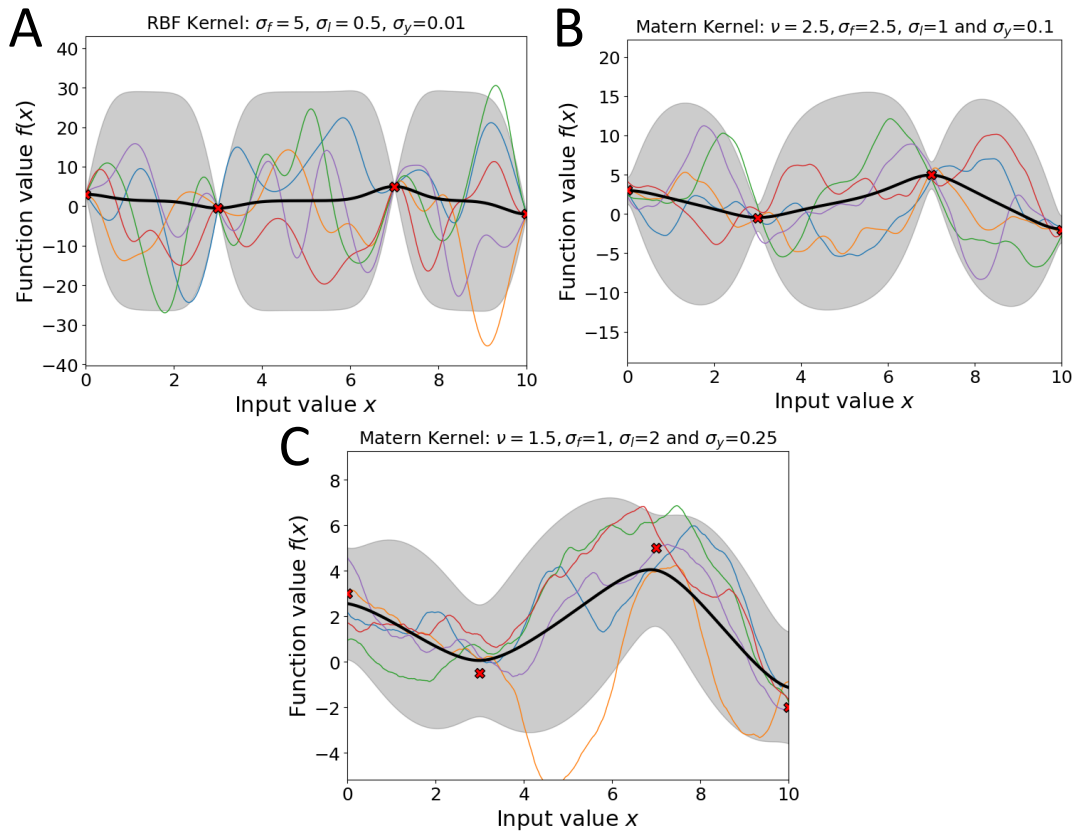
$$\log p(\mathbf{y} | X, \theta) = -\frac{1}{2} \mathbf{y}^\top A^{-1} \mathbf{y} - \frac{1}{2} \log |A| - \frac{n}{2} \log 2\pi, \quad (2.15)$$

with  $n = |\mathcal{D}|$ , i.e. the sample size, and  $A = K(X, X) + \sigma_y^2 \mathbb{I}$ . For the equations above and for the remainder of this chapter, we implicitly assume conditioning on  $\theta$  and test input locations  $X_*$  for readability.

Equipped with this posterior distribution we can now draw functions  $f_*$  from it in a similar fashion to the prior distribution. Therefore, we draw a list of func-

---

<sup>2</sup>However, this is not always necessary and they can remain fixed, if desired.



**Figure 2.2:** Five random functions drawn from the posterior distribution of three different Gaussian processes, each with a different covariance function, hyperparameters, and different observation noise  $\sigma_y$ .

tion values  $\{f_*(x_i) : i = 1, \dots, n\}$  for a random function  $f_*$  from the posterior distribution, which is simply a multivariate Gaussian distribution such that

$$f_* \sim \mathcal{N}(\mathbb{E}[f_*], \text{Cov}[f_*]), \quad (2.16)$$

where  $\mathbb{E}[f_*]$  and  $\text{Cov}[f_*]$  are given by equations (2.12) and (2.13). Figure 2.2 shows the Gaussian processes from figure 2.1, whose prior distributions have now incorporated new observations (red crosses) to generate posterior distributions. Note that hyperparameter optimisation has not happened here because the edge-fitting procedure to be described below does not leverage hyperparameter optimisation until convergence, contrary to typical application of Bayesian machine learning methods.

For each Gaussian process, a different level of noise  $\sigma_y^2$  has been chosen to demonstrate how the value of  $\sigma_y$  influences the posterior distribution, and the randomly sampled functions, as they near the observations. The change in mean function and 95% credible regions represent how each of the Gaussian process models have changed their belief in what the underlying function looks like after fitting the data. Note how the credible regions reduce in area near observations, and especially so when  $\sigma_y$  is very small. It's important to note that the choice of covariance function, and its hyperparameters ( $\sigma_f, \sigma_l$ ), must be chosen according to the data being fitted. Furthermore, we must also have an idea of how much noise,  $\sigma_y$ , exists in the observations so that the model doesn't become too confident (figure 2.1(A)), or too uncertain (figure 2.1(C)) in its predictions around these observations.

## 2.2 GPET's Methodology

This methodology works as an individual edge tracer. As such, the methodology is applied to the RPE-Choroid and Choroid-Sclera boundaries in series. Given an OCT B-scan with  $M$  rows and  $N$  columns, we consider each boundary as being modelled by an underlying true function  $f$ , whose input values  $X = \{i : i = 1, \dots, N\}$  are the column index, and function values  $\{f(x_i) : x_i \in X, i = 1, \dots, N\}$  are row indexes, i.e.  $f(x_i) \in [1, M]$ .

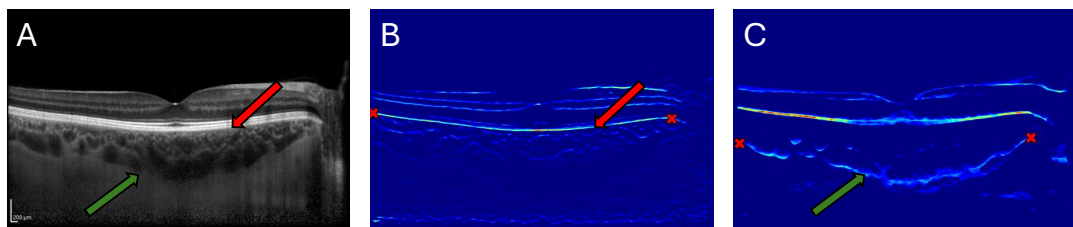
### 2.2.1 Pre-processing

In order to perform edge-tracing for an image with GPET, we must first generate an edge map for each EOI. An edge map gives an approximation to where the EOI lies in the image, but not an indication on the specific coordinates — this latter task is the edge tracing which GPET performs. Edge maps are common to aid edge-based segmentation methods. Graph-based and level set-methods use approximate edge maps to define a cost function to optimise over [113, 215], and GPET also leverages the edge map in a similar way.

The following steps are used to pre-process an OCT B-scan to obtain approximate edge maps for the RPE-Choroid and Choroid-Sclera boundaries:

1. Denoising: a  $5 \times 5$  median smoothing filter to reduce speckle noise inherent in an OCT B-scan. This slides a  $5 \times 5$  array (or window) across the image and replaces the pixel intensity which the window is centred on with the median value of the pixels in the window. Median smoothing is preferable because the median value is more robust to outliers in localised regions of the image relative to the mean value.
2. Contrast enhancement: application of contrast limited adaptive histogram equalisation (CLAHE). We break the image into  $8 \times 8$  tiles, and apply histogram equalisation to each tile using a histogram clipping limit of 5. See section 3.1.2 for more details on this method.
3. Gradient-based edge detection: a common way to obtain edge maps in an image is to apply a discrete derivative filter across the image, which have been well studied [216, 217] and include filters such as Sobel's [218], Robert's [219], Prewitt's [220] and the Canny edge detector [221]. See below for more details.
4. Edge map enhancement: once edge maps have been obtained using discrete derivative filters, these are post-processed to enhance each edge of interest by using minimum filters and morphological operations, as is discussed below in more detail.

Once the edge maps (also referred to as image gradients) have been gen-



**Figure 2.3:** (A) OCT B-scan. (B – C) Gradient-based edge maps for the RPE-Choroid boundary (red arrows) and Choroid-Sclera boundary (green arrows). Red crosses overlaid represent the edge endpoints selected manually.

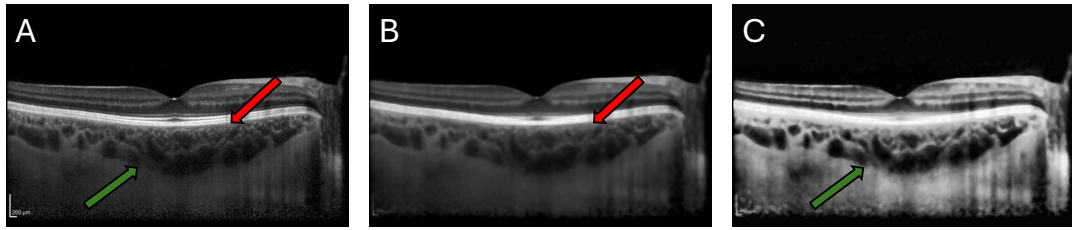
erated, an additional pre-processing steps requires the end-user to manually detect the endpoints of the EOI on the image gradient for both boundaries. This is done through a simple, high-level graphical user interface (GUI) developed using Python, where the image gradient appears on the end-user's home screen and they are prompted to select the endpoints using their track-pad or mouse. Note that this step could be automated to prevent manual intervention, and has indeed been done for a similar application related to a graph-based approach to retinal layer segmentation [222]. To enable endpoint detection for retinal layer segmentation, Chiu, et al. [222] leveraged the enhanced contrast between the innermost retinal layer (inner limiting membrane) relative to the vitreous body. While the RPE-Choroid is typically well illuminated for this automated procedure, this is not the case for the Choroid-Sclera boundary.

Figure 2.3 shows an OCT B-scan in panel (A) with the approximate edge maps (image gradient) for the RPE-Choroid and Choroid-Sclera boundaries of the choroid in panels (B) and (C), respectively. Red crosses show the edge endpoints manually selected for each EOI. It is the image gradient and edge endpoints which initialise the algorithm.

### 2.2.1.1 Gradient-based edge detection

Many early approaches to image segmentation used edge detection methods based on zero-crossings, discrete derivative operators or histogram thresholding [220, 223, 219, 216]. A variety of these early approaches to image segmentation can be found in Rosenfeld's work [217]. The problem of edge detection has been well studied and with methods such as Canny [221], it can be straightforward to obtain some approximation of the edges in the image.

However, there is a distinct difference between identifying the edges in an image and linking them to form a meaningful contour, with direct knowledge of the pixel coordinates. Szeliski [224] notes that this latter task is far more useful than the former for strict boundary delineation and downstream feature measurement, but is more challenging because of its inter-dependency on the quality of the resulting edge map.



**Figure 2.4:** (A) OCT B-scan. (B) Overly smoothed B-scan after median filtering. (C) Application of CLAHE with high clipping limit for severe enhancement.

Thus, it's important to extract a good quality edge map. In the first instance, we pre-process the OCT B-scan to denoise and enhance the boundaries of interest. Figure 2.4 shows the result of smoothing with a large window ( $5 \times 5$ ) in panel (B) which improves the contrast between the RPE and Choroid junction (red arrow). In panel (C), this shows the application of CLAHE with a reasonably large clip limit (5), which improves the contrast between the choroid's posterior vessels and sclera, subsequently improving the visible junction separating the posterior choroid from the sclera (green arrow). Thus, the RPE-Choroid and Choroid-Sclera edge-maps are detected from the denoised and the enhanced B-scans, respectively.

Discrete derivative filters are small arrays which encode a set of values which, when *convolved* with the input image, are able to quantify the strength of intensity transition over local regions of the image, such as the Sobel's [218], Robert's [219] or Prewitt's [220] filter. Given that the choroid boundaries are predominantly horizontal, our filters must be encoded to look for predominantly horizontal intensity transitions. Moreover, the RPE-Choroid boundary transitions from the bright, hyperreflective RPE to the darker choriocapillaris, while the Choroid-Sclera boundary transitions from the dark, posterior choroidal vessels in Haller's layer to the brighter, homogeneous sclera. Thus, the transitions are the opposite of one another. This is another important detail to encode into the filters.

A  $5 \times 5$  extension to the discrete derivative filters proposed by Sobel [218] for detecting horizontal edges with different transition patterns, i.e. for light-to-dark RPE-Choroid and dark-to-bright Choroid-Sclera boundaries are shown below



as arrays  $f_{b2d}$  and  $f_{d2b}$ , respectively.

$$f_{b2d} = \begin{pmatrix} -1 & -1 & -2 & -1 & -1 \\ -1 & -2 & -3 & -2 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 2 & 1 & 1 \\ 1 & 2 & 3 & 2 & 1 \end{pmatrix} \quad \text{and} \quad f_{d2b} = -f_{d2b}. \quad (2.17)$$

The larger filter compared to traditional  $3 \times 3$  filters promotes smoothness and reduces noise, while the increasing integer values toward the filter's centre ensures that pixels local to the centre contribute more, preventing nearby artefacts from negatively influencing the gradient approximation.

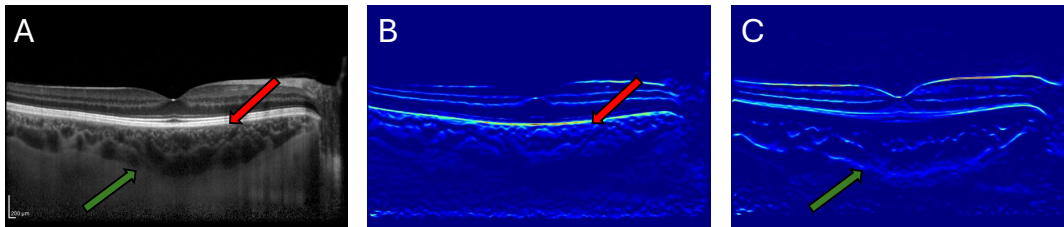
We obtain the image gradients by convolving these filters with the original image  $I$ , which are defined for the RPE-Choroid and Choroid-Sclera edge maps as

$$\hat{G}_{\text{upper}} = I * f_{b2d} \quad (2.18)$$

$$\hat{G}_{\text{lower}} = I * f_{d2b}. \quad (2.19)$$

Figure 2.5 show the approximate edge maps for the RPE-Choroid and Choroid-Sclera boundaries in panels (B) and (C) from an OCT B-scan.

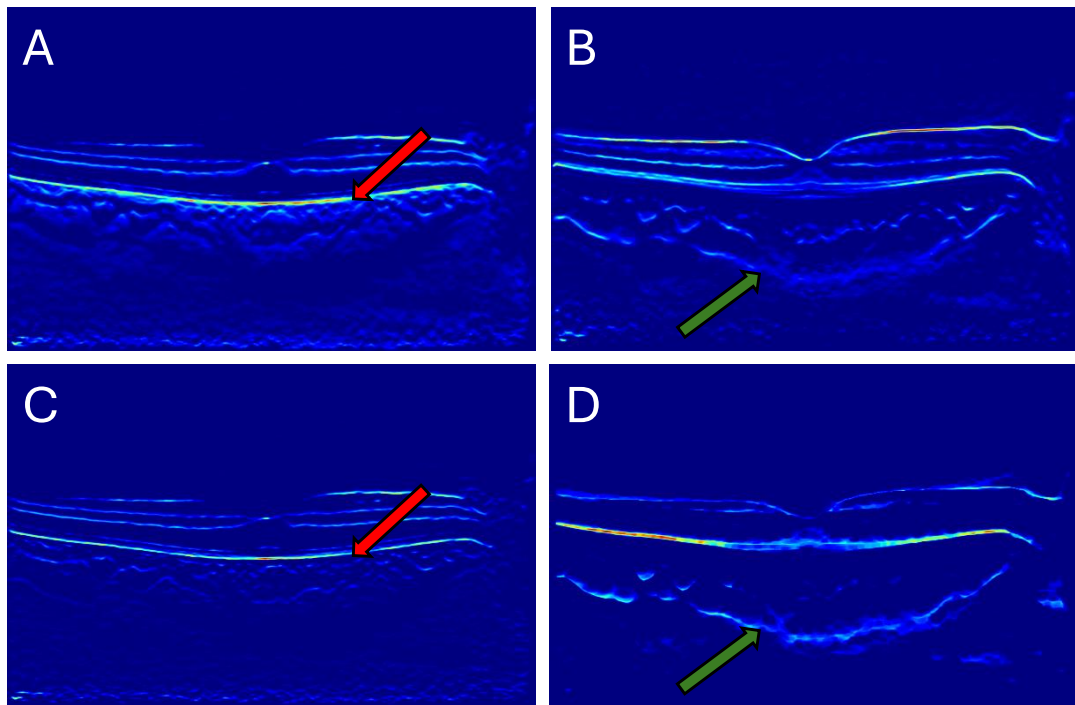
A final step is the enhancement of these edge maps. Note that in figures 2.5(B) and 2.5(C), while there is an obvious signal denoting the approximate location of the RPE-Choroid and Choroid-Sclera boundaries, there is also signal com-



**Figure 2.5:** (A) OCT B-scan. (B – C) Initial, gradient-based edge maps for the RPE-Choroid boundary (red arrows) and Choroid-Sclera boundary (green arrows).

ing from where individual choroidal vessels meet interstitial space. Moreover, the edge map in panel (C) shows the Choroid-Sclera boundary as quite a fuzzy object.

To improve the RPE-Choroid edge map, we can actually apply the dark-to-bright filter because of the thickness of the approximated edge seen in figure 2.5(B). To improve the Choroid-Sclera edge map, we can apply *morphological opening* using a  $5 \times 15$  window, which removes small, bright and predominantly horizontal objects, and *morphological erosion* using a  $5 \times 5$  window which removes small objects. These steps are important for removing the signal from individual choroidal vessels and tightening up the Choroid-Sclera boundary signal in the edge map. Thus, the final gradient-based edge maps  $G_{\text{upper}}$  and  $G_{\text{lower}}$  for the RPE-Choroid and Choroid-Sclera boundaries, respec-



**Figure 2.6:** (A–B) Naive edge maps  $\hat{G}_{\text{upper}}$  and  $\hat{G}_{\text{lower}}$ . (C–D) Enhanced edge maps  $G_{\text{upper}}$  and  $G_{\text{lower}}$ .

tively, are defined as

$$G_{\text{upper}} = (I * f_{\text{b2d}}) * f_{\text{d2b}} \quad (2.20)$$

$$G_{\text{lower}} = \text{erode}_{5 \times 5} \left( \text{open}_{5 \times 15} (I * f_{\text{d2b}}) \right). \quad (2.21)$$

Figure 2.6 shows the original, naive edge maps in the top row (panels A–B), with the enhanced edge maps shown in the bottom row (panels C–D).

## 2.2.2 Initialisation

For tracing each boundary of the choroid in series, GPET requires initialisation. To initialise the model for an EOI, we assume that its edge endpoints and image gradient  $G$  are known. Let the set of EOI endpoints be defined as  $\mathcal{D}^{(0)} = \{X^{(0)}, \mathbf{y}^{(0)}\}$ . We assume for simplicity that the EOI endpoints span the width of the image.

### 2.2.2.1 Prior distribution

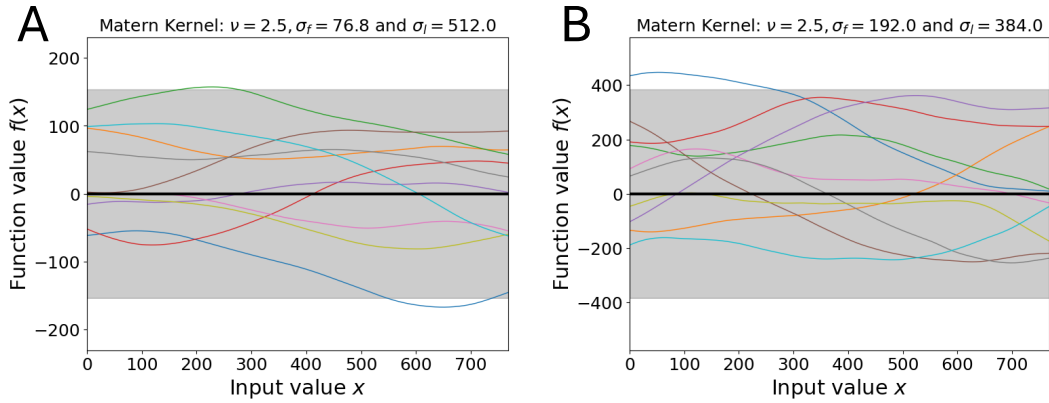
At the beginning of the procedure, following a similar formulation to equation (2.4), our prior distribution comes from a zero mean Gaussian process with covariance function  $k$  (with specified hyperparameters  $\theta = [\sigma_f, \sigma_l]$ ) such that

$$f^{(0)}(x) \sim \mathcal{GP}(0, k(x, x'; \theta)). \quad (2.22)$$

This induces a prior distribution where a random function  $f^{(0)}$  can be sampled using

$$f^{(0)} \sim \mathcal{N}(0, K_{**}), \quad (2.23)$$

where  $K_{**} = K(X_*, X_*)$  and  $X_* = \{i : i = 1, \dots, N\}$ . The choice of covariance function for the RPE-Choroid and Choroid-Sclera boundary is typically dependent on the choroid. However, the RPE-Choroid often appears as smooth and very slightly parabolic in shape, with little curvature. The Choroid-Sclera boundary is typically less smooth with more curvature and turning points. Thus,



**Figure 2.7:** 10 random functions drawn from the initial prior distribution of Gaussian process regressors for RPE-Choroid (A) and Choroid-Sclera (B) boundary detection. The mean function (black solid line) and 95% credible intervals (grey shaded region) are also shown.

the default covariance functions for the upper and lower choroid boundaries — for an OCT B-scan with pixel resolution  $768 \times 768$  (height  $\times$  width) — are defined below as

$$k_{\text{upper}} = \text{Matern}(\sigma_f = M/10, \sigma_l = 2N/3; \nu = 2.5) \quad (2.24)$$

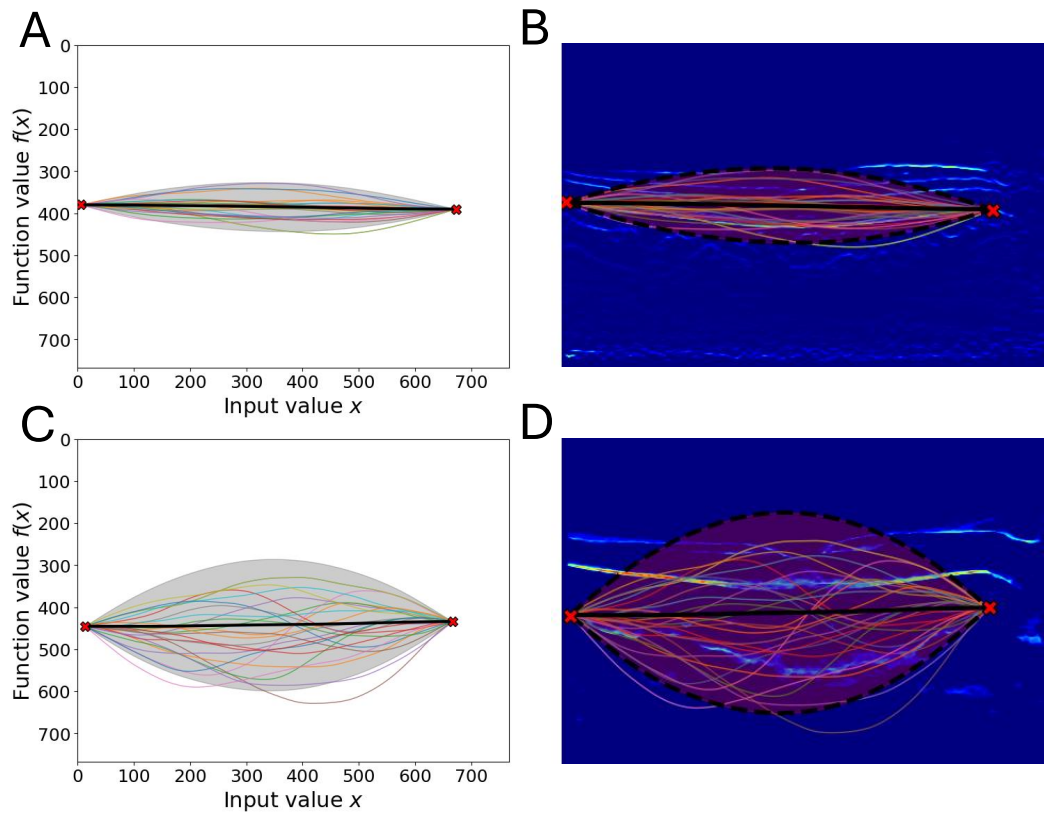
$$k_{\text{lower}} = \text{Matern}(\sigma_f = M/4, \sigma_l = N/2; \nu = 2.5) \quad (2.25)$$

Figure 2.7 shows the initial prior distributions for the Gaussian process regressors modelling the RPE-Choroid and Choroid-Sclera boundaries in panels (A) and (B), respectively.

### 2.2.2.2 Posterior distribution

To induce the initial posterior (predictive) distribution (PPD), we condition the Gaussian process on the edge endpoints  $\mathcal{D}^{(0)}$ . These user-specified pixel coordinates act as domain-specific knowledge and are assumed to be true realisations of the EOI. This step tells the algorithm which edge the end-user wishes to trace and fitting  $\mathcal{D}^{(0)}$  induces the initial PPD,  $\mathbf{f}^{(1)}|\mathcal{D}^{(0)}$ , from which we can draw samples evaluated at the image column indexes in  $X_*$ .

The objective is to now predict the missing pixels between the edge endpoints.



**Figure 2.8:** 25 random functions sampled from the initial PPD of the Gaussian process regressors modelling the RPE-Choroid (A–B) and Choroid-Sclera boundaries (C–D). Panels (B) and (D) plot the random functions (multi-coloured), mean function (solid black) and 95% credible interval (red shaded, with black dashed outlines) onto the respective edge maps.

The predictive variance (the 95% credible intervals shown in figure 2.8) of the PPD increases in regions far away from the edge endpoints. This allows the random functions we draw from the PPD to evolve freely across the image — according to the properties imposed by the selected covariance function — as shown in figure 2.8(B, D).

In fact, the signal variance  $\sigma_f^2$  dictates the size of this predictive variance. With enough distance from the edge endpoints, the initial PPD will have a 95% credible interval of  $\bar{y}^{(0)} \pm 2\sigma_f$ , where  $\bar{y}^{(0)}$  is the mean of the edge endpoint pixel row indexes. This predictive variance is something we take advantage of throughout the fitting procedure so that the random functions can explore around the

EOI and propose possible locations for it.

We will now refer to random functions as “posterior curves”, and use the terminology “predictive variance” to mean the 95% credible interval of the PPD. The remainder of this section describes the methodology using an arbitrary iteration  $n \geq 0$ .

### 2.2.3 Curve scoring

$L$  posterior curves  $f_l^{(n)}$ ,  $l = 1, \dots, L$  are drawn at random from the PPD for the current iteration which are defined at the column indexes in  $X_*$ . These curves are a representative sample of the model's belief in where the EOI lies, given the (current) observation set defined from the last iteration  $\mathcal{D}^{(n-1)}$ .

$L$  is set at 500 so that there are enough posterior curves to represent the PPD to sufficiently explore the image gradient  $G$ . Lower values of  $L$  will not explore the space well enough leading to an under-specified model. Additionally,  $L$  should not be arbitrarily large due to the computational overhead of curve sampling.

These curves are scored using the pre-computed image gradient  $G$  and the highest scoring posterior curves are used to decide on the next set of pixels that will be added to the model's observation set. The score for each posterior curve computes the cumulative value of the gradient response, per unit length, along each posterior curve. For a given posterior curve  $f_l^{(n)}$  the score  $I(f_l^{(n)})$  is

$$I(f_l^{(n)}) = \int_{f_l^{(n)}} G(\mathbf{r}(s)) ds / \int_{f_l^{(n)}} ds. \quad (2.26)$$

This score is analogous to computing the area under the posterior curve along the interpolated image gradient surface, divided by the posterior curve's arc length.

We use Simpson's and finite difference rules to estimate the line integral and arc length. The advantage of scaling a curve's score by their arc length is to penalise them for making longer excursions than necessary, as these are unlikely to happen for continuous edges such as the upper and lower choroid

boundaries. Moreover, this can also be particularly helpful in the presence of noise in the image gradient.

We invert this gradient score to create a cost function such that

$$C(f_l^{(n)}) = \left(I(f_l^{(n)})\right)^{-1} \quad (2.27)$$

This is so that higher scoring curves are minimally costly and the minimisation of this cost becomes part of the optimisation scheme to model the edge.

## 2.2.4 Pixel scoring

the 50 most optimal posterior curves according to the cost function defined above are selected to determine the next observation set, with all other curves discarded. These 50 curves were defined as 10% of  $L=500$ . Selecting fewer than 10% risks under-specification, as the model is forced to rely on a small subset of posterior curves to choose pixels with. Moreover, too large a proportion may introduce unwanted noise into the fitting procedure from low-scoring posterior curves.

These optimal curves provide a best guess at where the current PPD believes the EOI lies and their coordinates together map out which regions have been visited the most. Given these curves lie in the vicinity of the EOI, we can obtain an empirical approximation for where the PPD believes the EOI is by generating a two dimensional density function which measures the location frequency of the most optimal posterior curves. Think of this as a two dimensional (continuous) histogram the size of the original image, keeping track of where and how many times a pair of optimal posterior curves intersected in the image.

This frequency density function is defined by

$$\phi^{(n)}(z) = \frac{1}{\sum_{l,i} w(z_{li})} \sum_{l,i} w(z_{li}) K_{\mathbb{I}}(z - z_{li}), \quad (2.28)$$

where  $z, z_{li}$  are two dimensional coordinates such that  $z = [x, y]$  with  $x \in [1, N]$  and  $y \in [1, M]$ .  $z_{li}$  corresponds to the  $i^{\text{th}}$  coordinate of the  $l^{\text{th}}$  optimal posterior curve,  $f_l^{(n)}$ .

The weighting for coordinate  $z_{li}$  scores the point according to how optimal the curve  $f_l^{(n)}$  is relative to all other optimal curves, using  $I(f_l^{(n)})$ . This weighting will result in a higher density for regions where most optimal curves intersected in the image. The weight function  $w$  is defined by

$$w(z_{li}) = \frac{I(f_l^{(n)})}{\sum_{l=1}^{L/10} I(f_l^{(n)})}, \quad (2.29)$$

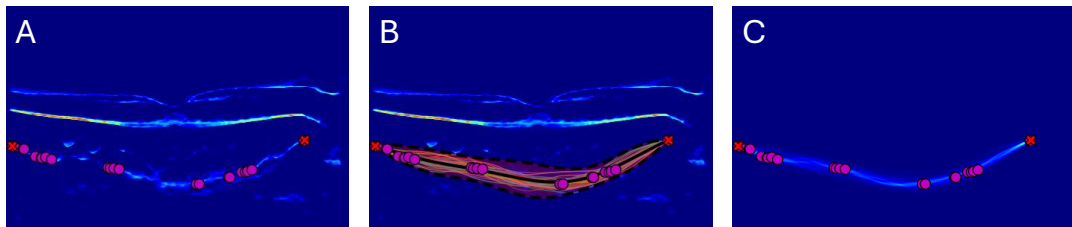
for all coordinates  $i$  for the  $l^{\text{th}}$  optimal curve  $f_l^{(n)}$ .

$K_{\mathbb{I}}(z - z_{li})$  is a symmetric, multivariate kernel (covariance function) of the density  $\phi^{(n)}$ . Here, we use the isotropic, two-dimensional Gaussian kernel with the identity matrix as its length-scale,

$$K_{\mathbb{I}}(z - z_{li}) = (2\pi)^{-1} \exp\left(-\frac{1}{2}(z - z_{li})^{\top}(z - z_{li})\right). \quad (2.30)$$

The isotropic and unitary length-scale imposes no dominant orientation in the Gaussian kernel which is centred at each real-valued point  $z_{li}$ . Moreover, the unitary property ensures each point's kernel contributes the majority of its density to only the 4 nearest pixel coordinates in the discretised space.

Figure 2.9 illustrates the fitting procedure for the Choroid-Sclera boundary whose initial PPD was shown in figure 2.5(D). In panel (A) we show the image gradient with the current observation set  $\mathcal{D}^{(n-1)}$ . This observation set has



**Figure 2.9:** Frequency density and PPD during an arbitrary iteration of the fitting procedure for the Choroid-Sclera boundary. (A) Image gradient. (B) PPD with optimal posterior curves overlaid. (C) Weighted frequency density function. Current observation set  $\mathcal{D}^{(n-1)}$  overlaid as purple dots, with edge endpoints as red crosses.



updated the PPD around the EOI as shown in panel (B), whose 95% credible interval has shrunk and is centred around the EOI. Note the central column of the image has the largest predictive variance (uncertainty) because there are no observations present in that region. The most optimal posterior curves, overlaid in panel (B) are able to collectively map out the approximate location of the EOI, as shown by the weighted kernel density in panel (C).

Once  $\phi^{(n)}$  is defined, individual pixel coordinates,  $\mathbf{p} = [x, y]$ , can be scored by combining their density,  $\phi^{(n)}(\mathbf{p})$  and gradient response,  $G(\mathbf{p})$ . This score,  $s(\mathbf{p}) \in [0, 1]$ , is defined by

$$s(\mathbf{p}) = \frac{1}{3} \left( \phi^{(n)}(\mathbf{p}) \cdot G(\mathbf{p}) + \phi^{(n)}(\mathbf{p}) + G(\mathbf{p}) \right). \quad (2.31)$$

This scoring function gives equal weighting to the local information supplied by the image gradient and the global perspective provided by the PPD.

## 2.2.5 Accept-discard scheme

The primary advantage of using a score function which includes the iteration-dependent frequency density  $\phi^{(n)}$  is that densities are updated in each consecutive iteration using the new PPD. This means that pixels fitted during previous iterations are re-scored using an updated density allowing a more accurate representation of how probable those pixels are of belonging to the EOI, given the new set of optimal posterior curves. This allows the possibility of older, incorrectly fitted pixels to be discarded from the model if their score drops sufficiently in later iterations.

An adaptive, user-specified value  $T \in [0, 1]$  is chosen to threshold these pixels such that

$$P_T^{(n)} = \left\{ \mathbf{p} \mid s(\mathbf{p}) \geq T \right\}, \quad (2.32)$$

where  $T$  is initially set at 1 for every iteration (see below).

The thresholded pixels are binned into  $\lceil N/\Delta x \rceil$  linearly spaced subsets  $S_i$ ,  $i = 1, \dots, \lceil N/\Delta x \rceil$ , according to their horizontal column index, such that

$$S_i = \left\{ \mathbf{p} = [x, y] \mid \mathbf{p} \in P_T^{(n)}, x \in [(i-1)\Delta x, i\Delta x] \right\}_{i=1}^{\lceil N/\Delta x \rceil}, \quad (2.33)$$

where  $\Delta x$  is the length of each sub-interval. That is,  $S_i$  contains all thresholded pixel coordinates whose column index is in the interval  $[(i-1)\Delta x, i\Delta x]$ .  $\Delta x$  defines how many pixels are used to reconstruct the EOI and the choice of  $\Delta x$  depends on the EOI's complexity and number of columns it spans in the image. However, For practical purposes,  $\Delta x = 10$  can reliably reconstruct the upper and lower choroid boundaries well. Lowering this value may be required when the EOI is less smooth.

As a continuous function is fitted, we enforce a property called injectivity. This is a property where two different row indexes uniquely corresponds to two different column indexes, such that for some function  $f$ ,  $f(x) \neq f(y)$  implies that  $x \neq y$ , i.e. the function  $f$  does not have any regions which are entirely vertical, or loops back on itself. Injectivity is enforced by using non-max suppression, selecting the highest scoring pixel per sub-interval  $S_i$ , yielding

$$P_A^{(n)} = \left\{ \operatorname{argmax}_p \{s(\mathbf{p})\} \mid \mathbf{p} \in S_i \right\}_{i=1}^{\lceil N/\Delta x \rceil}. \quad (2.34)$$

Coordinates are binned because it is unnecessary to fit the model with points which are very close in the input space, unless the underlying edge is very complex, in which case we reduce the value of  $\Delta x$ .

The observation set for the next iteration  $\mathcal{D}^{(n)}$  is formed by re-scoring the previous iteration's observation set according to the new set of optimal posterior curves,  $\mathcal{D}^{(n-1)}$ , and combining this with  $P_A^{(n)}$  through binning (equation (2.33)) and non-max suppression (equation (2.34)) yielding

$$\mathcal{D}^{(n)} = \operatorname{suppression} \left( \operatorname{binning} \left( P_A^{(n)} \cup \mathcal{D}^{(n-1)} \right) \right). \quad (2.35)$$

$P_A^{(n)}$  is combined with  $\mathcal{D}^{(n-1)}$  because  $\phi^{(n)}$  changes with each iteration, meaning that the highest scoring pixel coordinate in each set  $S_i$  may change dependant on the score it obtains compared to other possible pixels in its neighbourhood.

The accept-discard procedure in this section is repeated if  $|\mathcal{D}^{(n)}| \leq |\mathcal{D}^{(n-1)}|$  by reducing  $T$  by 1%. Thus, setting this adaptive threshold at 1 at every iteration ensures that only the best scoring pixels are selected since  $s(\mathbf{p}) \leq 1$  for all pixel coordinates  $\mathbf{p}$ . This adaptive property ensures that  $|\mathcal{D}^{(n)}| > |\mathcal{D}^{(n-1)}|$ , and that the fitting procedure is guaranteed to converge in a finite number of iterations, which is exactly  $\lceil N/\Delta x \rceil$ .

## 2.2.6 Refitting

The observations in  $\mathcal{D}^{(n)}$  represent the most probable pixel coordinates which form part of the EOI for the next iteration. An improved PPD can be formed using this new set of pixel coordinates, updating the model's belief in where the EOI lies, such that

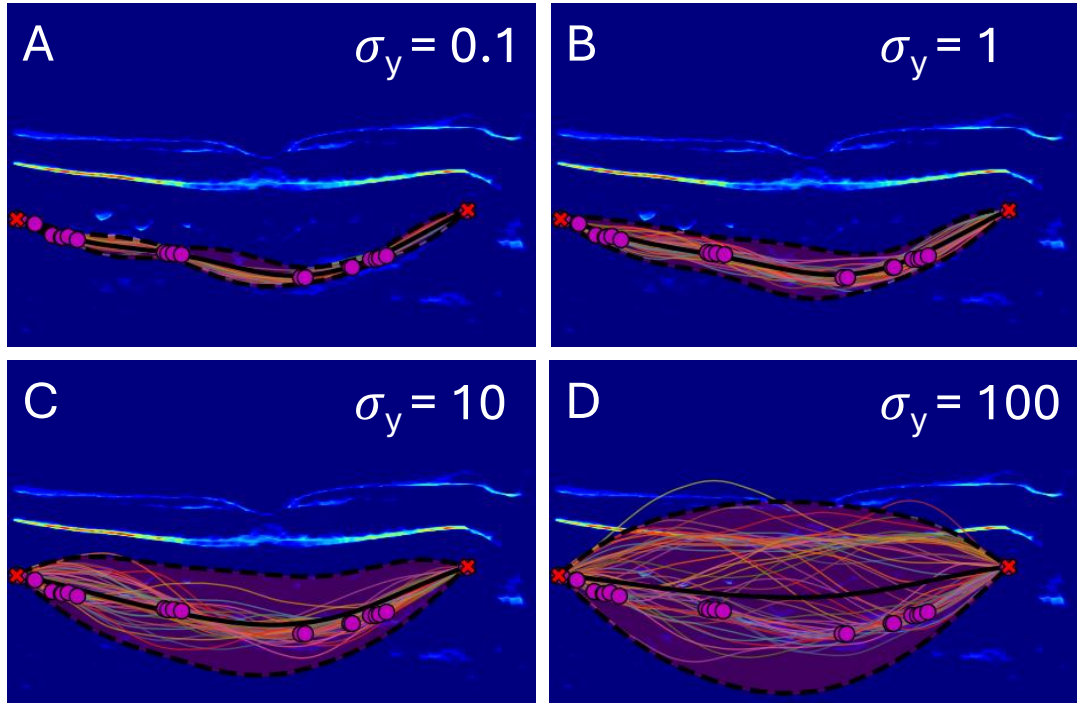
$$\mathbf{f}^{(n+1)} \mid \mathcal{D}^{(n)} \sim \mathcal{N}(\boldsymbol{\mu}^{(n+1)}, \boldsymbol{\sigma}^{(n+1)}); \quad (2.36)$$

$$\boldsymbol{\mu}^{(n+1)} = K_{*n} [K_{nn} + \sigma_y^2 \mathbb{I}]^{-1} \mathbf{y}^{(n)}; \quad (2.37)$$

$$\boldsymbol{\sigma}^{(n+1)} = K_{**} - K_{*n} [K_{nn} + \sigma_y^2 \mathbb{I}]^{-1} K_{n*}, \quad (2.38)$$

where  $K_{nn} = K(X^{(n)}, X^{(n)})$  and similarly for  $K_{n*}$  and  $K_{**}$ . The posterior curves  $\mathbf{f}^{(n+1)}$  we can sample should now be more accurate in mapping regions of the EOI than in previous iterations.

Observation noise  $\sigma_y$  is user-specified and encodes the confidence in the models' observation set  $\mathcal{D}^{(n)}$ . Specifically,  $\sigma_y$  is used to perturb the posterior curves from passing through the observation points exactly, allowing posterior curves to continue searching for pixels in the image that are in the vicinity of previously fitted pixels. This permits the possibility of accepting better edge pixels in each sub-interval in future iterations. The lower the value of  $\sigma_y$  implies a lower predictive variance in regions near observations in the PPD. Too large a value implies the model lacks confidence during pixel selection which can reduce accuracy. Figure 2.10 shows the effect of  $\sigma_y$  on the same observation set used in figure 2.9. Here, a value of 1 balances the degree of uncertainty away from  $\mathcal{D}^{(n)}$  with the confidence near  $\mathcal{D}^{(n)}$  most optimally.

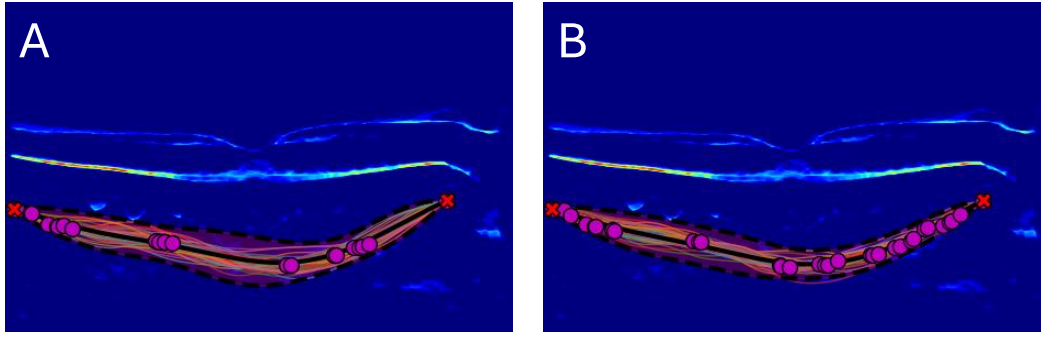


**Figure 2.10:** Influence of the observation noise during GPET's Choroid-Sclera boundary fitting. PPDs (with the 50 optimal posterior curves overlaid with mean function and 95% credible interval) for the same observation set as in figure 2.9, with different observation noise  $\sigma_y = 0.1$  (A),  $\sigma_y = 1$  (B),  $\sigma_y = 10$  (C),  $\sigma_y = 100$  (D).

Figure 2.11 shows the PPD with the observation set  $\mathcal{D}^{(n-1)}$  used throughout this section in panel (A), with the new observation set  $\mathcal{D}^{(n)}$  in panel (B). Clearly, there is smaller predictive variance as the number of observations grow, and the PPD has converged closer to the Choroid-Sclera boundary.

### 2.2.7 Convergence

Convergence is when the final observation set  $\mathcal{D}^{(n)}$  has enough pixels to reconstruct the EOI. Specifically, the algorithm reaches termination once all sub-intervals have selected a high-scoring pixel, i.e.  $|\mathcal{D}^{(n)}| = \lceil N/\Delta x \rceil$ . The model is refitted once more using  $\mathcal{D}^{(n)}$ , this time optimising  $\theta$  by maximising the log marginal likelihood according to equation (2.14), yielding  $\hat{\theta}$  and  $\hat{\sigma}_y^2$ . This max-



**Figure 2.11:** Influence of increasing the observation set on the PPD during Choroid-Sclera boundary fitting. (A) Observation set  $\mathcal{D}^{(n-1)}$  and corresponding PPD shown from figure 2.9(B). (B) Observation set  $\mathcal{D}^{(n)}$  and corresponding PPD after re-fitting.

imisation scheme chooses kernel hyperparameters and noise variance which make  $\mathcal{D}^{(n)}$  seem probable.

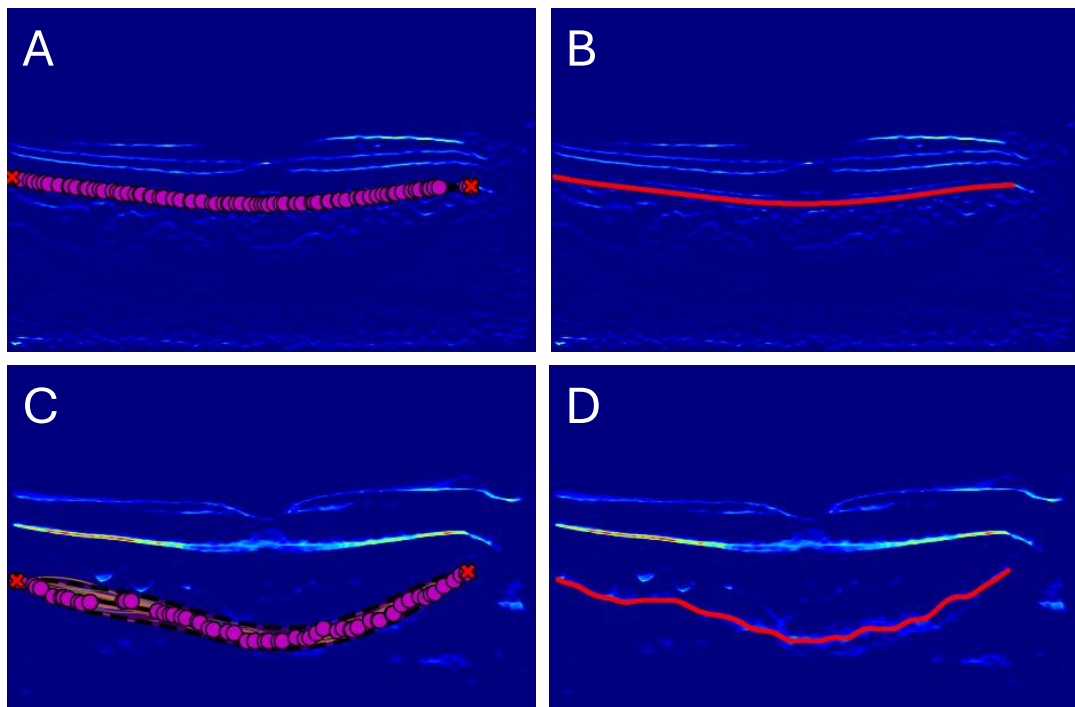
This optimisation drives  $\sigma_y^2$  toward 0 because there are enough edge pixels in  $\mathcal{D}^{(n)}$  to interpolate through the observations almost exactly. Note via equation (2.12) that the smaller  $\sigma_y^2$  is to 0, the closer the posterior predictive mean is to all observations. It is this optimisation scheme which converges the final PPD to the EOI, using the optimised posterior predictive mean,

$$\hat{\mathbf{f}}^{(n+1)} \mid \mathcal{D}^{(n)}, \hat{\boldsymbol{\theta}} = K_{*n} [K_{nn} + \hat{\sigma}_y^2 \mathbb{I}]^{-1} \mathbf{y}^{(n)}, \quad (2.39)$$

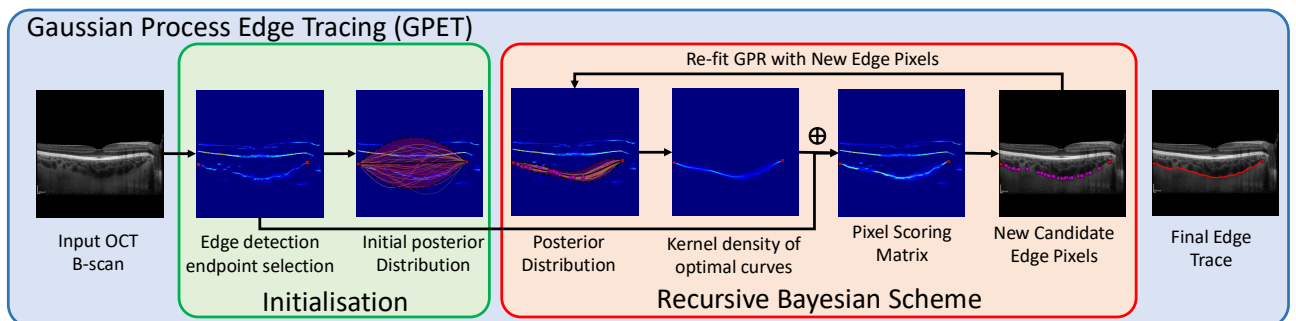
as the proposed edge. This, as well as a 95% credible band, evaluated at  $X_*$ , is outputted to the end-user. Figure 2.12 shows the final observation set and PPD overlaid onto the RPE-Choroid and Choroid-Sclera image gradients in panels (A) and (C), with the final edge trace (the optimised posterior predictive mean) shown in panels (B) and (D), respectively.

## 2.2.8 Pipeline

Figure 2.13 presents a schematic diagram of GPET's image analysis pipeline. Additionally, Algorithm 1 outlines the pseudocode of the proposed edge tracing algorithm.



**Figure 2.12:** (A, C) Image gradients with final PPD and observation set overlaid for RPE-Choroid (A) and Choroid-Sclera boundary (C). Final edge trace overlaid onto image gradient for RPE-Choroid (B) and Choroid-Sclera boundary (D).



**Figure 2.13:** Schematic diagram of GPET segmenting the Choroid-Sclera boundary of the choroid on an OCT B-scan.

**Algorithm 1** Edge Tracing using Gaussian Process Regression

---

**Inputs:** Image gradient  $G$  and edge endpoints  $\mathcal{D}^{(0)}$ . Covariance function  $k$  and its hyperparameters  $\theta$ , including observation noise  $\sigma_y$ .

**Output:** Optimal posterior predictive mean with a 95% credible band, evaluated at  $X_*$ .

- 1: Induce prior predictive distribution,  $f^{(0)}(x) \sim \mathcal{GP}(0, k(x, x'; \theta))$ ;
  - 2: Fit edge endpoints in  $\mathcal{D}^{(0)}$  to the Gaussian process and induce the initial PPD,  $\mathbf{f}^{(1)}|\mathcal{D}^{(0)}$ , corrupting edge endpoints with observation noise  $\sigma_y^2$ ;
  - 3: **while**  $|\mathcal{D}^{(n)}| \neq \lceil N/\Delta x \rceil$  **do**
  - 4:     Draw  $L=500$  posterior curves from the current PPD;
  - 5:     Compute cost of each posterior curve and select the most optimal 10% (50) of curves;
  - 6:     Estimate frequency density function  $\phi^{(n)}$  using optimal posterior curves, weighted by their curve score;
  - 7:     Score and threshold pixel coordinates with  $T$  initialised as 1, yielding  $P_T^{(n)}$ ;
  - 8:     Bin and use non-max suppression to obtain  $P_A^{(n)}$ ;
  - 9:     Combine  $\mathcal{D}^{(n-1)}$  and  $P_A^{(n)}$  to obtain  $\mathcal{D}^{(n)}$  through binning and non-max suppression;
  - 10:     While  $|\mathcal{D}^{(n)}| \leq |\mathcal{D}^{(n-1)}|$ , reduce  $T$  by 1% and repeat steps 7–9;
  - 11:     Retrain the model to induce an updated PPD,  $\mathbf{f}^{(n+1)}|\mathcal{D}^{(n)}$ , corrupting  $\mathcal{D}^{(n)}$  with observation noise  $\sigma_y$ .
  - 12: **end while**
  - 13: Optimise hyperparameters  $\theta$  and retrain the Gaussian process using final set of observations  $\mathcal{D}^{(n)}$  with optimised hyperparameters  $\hat{\theta}$ ;
  - 14: **return** optimised posterior predictive mean,  $\hat{\mathbf{f}}^{(n+1)}|\mathcal{D}^{(n)}$ ,  $\hat{\theta}$ , and 95% credible band evaluated at  $X_*$ .
- 

## 2.3 GPET’s evaluation

We evaluate GPET by comparing its performance against manual grading for measuring choroid thickness. This is accomplished using a longitudinal cohort of renal transplant and donor patients.

### 2.3.1 Data and methods

#### 2.3.1.1 Study population

We prospectively analysed a longitudinal cohort of healthy kidney donors and patients with end-stage kidney disease (CKD) undergoing living donor kidney transplantation (NCT0213274) [225]. Data collection and subsequent analyses were conducted after ethical approval from the South East Scotland Research Ethics Committee, in accordance with the principles of the Declaration

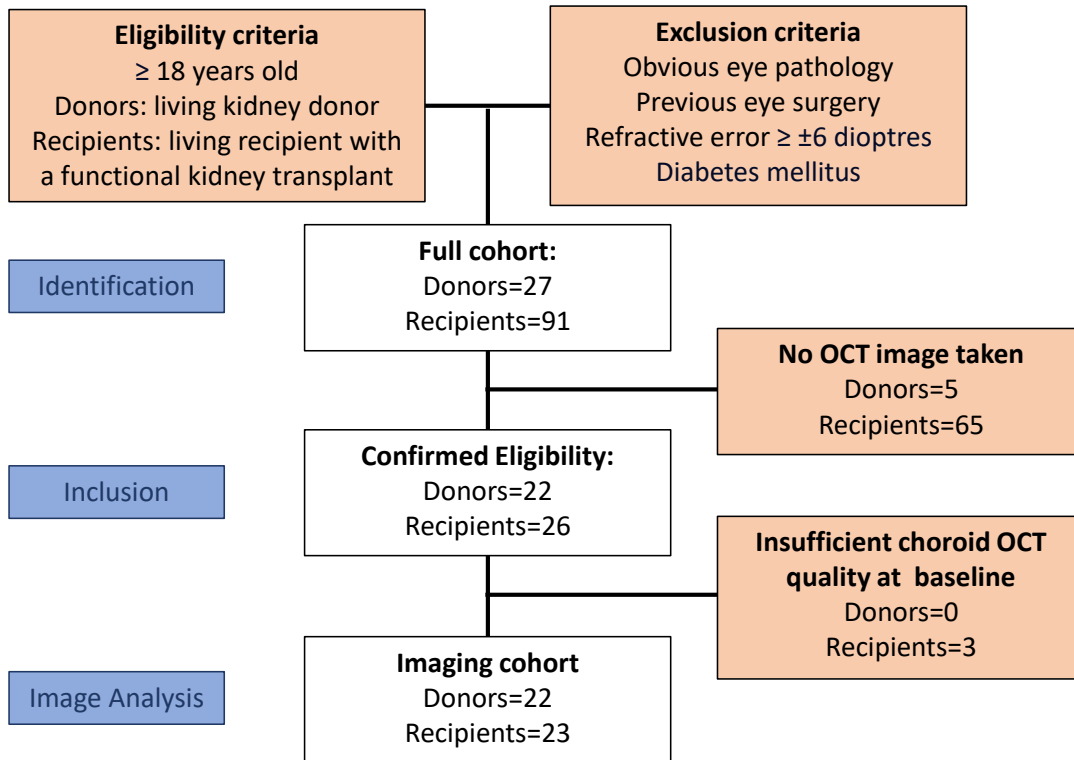
of Helsinki and all participants gave informed consent to recruitment. Eligibility criteria for recruitment were (1) donors must be living and healthy throughout the period of analysis, (2) recipients with end-stage CKD must have had a functional kidney transplant and (3) must be aged 18 or over. Exclusion criteria were (1) any ocular pathology pre-transplant, (2) any previous eye surgery, (3) a refractive error exceeding  $\pm 6$  dioptres or (4) a diagnosis of diabetes mellitus. We judged image quality by the OSCAR-IB criteria [226] and excluded images with B-scan signal quality  $\leq 15$ , indistinguishable Choroid-Sclera boundaries due to speckle noise, or partial image cropping of choroid. Figure 2.14 presents a sample derivation flowchart on how the population was selected for the wider research study, and then used for GPET's evaluation.

At the time of data extraction, 22 donors and 23 recipients were eligible for analysis, after excluding a total of 65 recipients and 5 donors. The majority of exclusions were because there were no eye scans performed for these individuals. This was because, once recruited, they were scheduled for transplantation and then immediately returned to their referring institution after transplant away from Edinburgh, thus making it impractical to retain in the study and re-scan. Each participant was assessed on the day of transplant and then attempts were made at 7 other time periods over the following year (8 time points in total). We used the entire analysis cohort for performance evaluation of the automatic approach against manual choroid thickness, resulting in 480 choroid thickness measurements for comparison from 160 OCT B-scans in 45 eyes. Table 2.1 shows some population statistics for the analysis cohort at baseline, with table 2.2 showing sample size and time-point statistics across all other longitudinal time-points sampled.

|            | Cohort      |             |
|------------|-------------|-------------|
|            | Donors      | Recipients  |
| Sample, N  | 22          | 23          |
| Age, years | $50 \pm 11$ | $47 \pm 12$ |
| Sex (male) | 9 (41)      | 15 (65)     |

**Table 2.1:** Basic demographic information for the donors and recipients used to evaluate GPET. Where appropriate, values are shown as mean  $\pm$  SD.





**Figure 2.14:** Population derivation for CKD sample used to assess GPET’s performance against manual grading.

### 2.3.1.2 Image acquisition

We imaged each participant’s right eye using the Heidelberg Engineering SD-OCT Spectral OCT1 Standard Module. Patients were examined between 9am and 5pm and were endeavoured to be followed up at around the same time of day at each time point so as to prevent any diurnal effects on downstream statistical analyses. The average (and standard deviation) daytime which scans were taken at each time point are listed in table 2.2. Follow-up was attempted at 7 different time points, but circumstances related to the availability of the imaging device, technician and participant resulted in a number of follow-up imaging not taking place for many participants.

During each patient’s examination, a single horizontal-line, EDI-OCT B-scan centred at the foveal pit was taken. Each B-scan covered a 30° (approximately 9 mm laterally) region and was extracted as a 768 × 768 (pixel height × width)

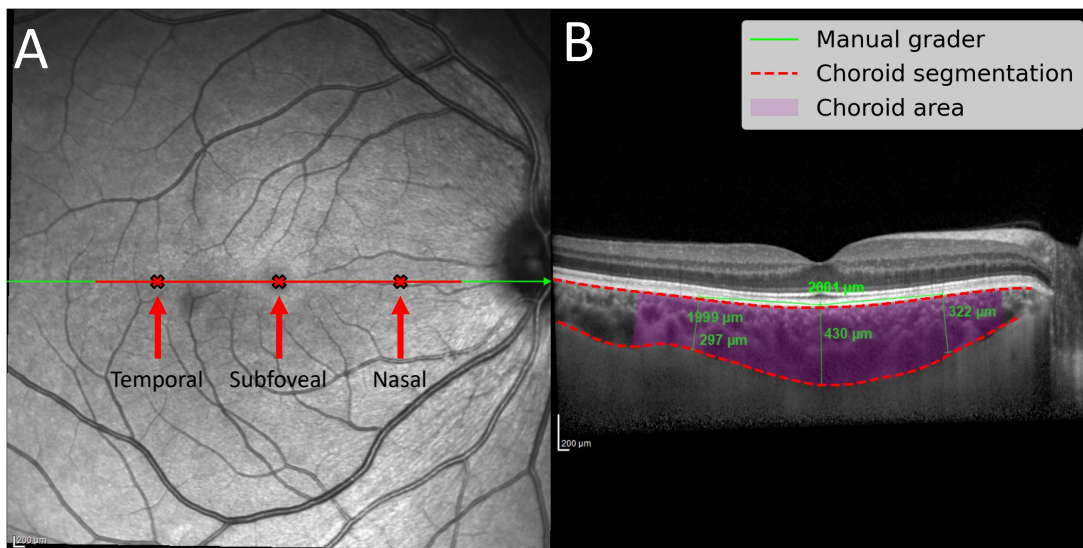
| Timepoint | Cohort        |                  |                            |               |                  |                            |
|-----------|---------------|------------------|----------------------------|---------------|------------------|----------------------------|
|           | Donors        |                  |                            | Recipients    |                  |                            |
|           | Sample, n (%) | Daytime (Hr:Min) | Time from baseline (weeks) | Sample, n (%) | Daytime (Hr:Min) | Time from baseline (weeks) |
| Baseline  | 22 (100)      | 13:50 ± 1:41     | 0                          | 23 (100)      | 13:59 ± 2:19     | 0                          |
| 1 Week    | 11 (50)       | 13:48 ± 1:20     | 1 ± 0                      | 13 (61)       | 12:44 ± 1:42     | 1 ± 0                      |
| 2 Weeks   | 2 (9)         | 12:04 ± 1:34     | 2 ± 0                      | 5 (48)        | 11:45 ± 1:50     | 2 ± 0                      |
| 4 Weeks   | 8 (36)        | 12:48 ± 3:47     | 4 ± 1                      | 5 (48)        | 12:23 ± 2:01     | 4 ± 1                      |
| 8 Weeks   | 10 (45)       | 13:43 ± 2:12     | 8 ± 2                      | 10 (43)       | 11:31 ± 1:13     | 8 ± 2                      |
| 16 Weeks  | 8 (36)        | 13:54 ± 2:00     | 16 ± 4                     | 6 (26)        | 12:29 ± 1:55     | 17 ± 4                     |
| 28 Weeks  | 10 (45)       | 13:30 ± 2:03     | 29 ± 3                     | 8 (35)        | 11:47 ± 1:39     | 29 ± 4                     |
| 52 Weeks  | 11 (50)       | 13:48 ± 2:07     | 53 ± 6                     | 8 (35)        | 12:25 ± 1:31     | 52 ± 4                     |

**Table 2.2:** Information on sample size, time of day and approximate time from baseline for donor and recipient samples. Values are shown as mean ± SD and those in parentheses are percentages of the total baseline sample size.

high-resolution image for downstream image processing. Each B-scan was captured using active eye tracking with an ART of 100. Figure 2.15 shows an example EDI-OCT B-scan with results from manual and automatic assessment.

### 2.3.1.3 Manual measurement of choroid thickness

Manual grading for measuring choroid thickness was performed by a single, trained operator across all image sessions for each participant. Manual choroid thickness was measured using the calliper tool in the Heidelberg Eye Explorer (HEYEX) software (version 1.10.4.0, Heidelberg Engineering, Heidelberg, Germany). The calliper tool was used to mark the centre of the fovea at the level of the photoreceptor outer segment and mark 2000 microns temporal and nasal to this point, parallel to the RPE-Choroid boundary. Choroid thickness was defined as the straight line distance between the RPE-Choroid and Choroid-Sclera boundaries, measured locally *perpendicular* to the RPE-Choroid boundary. Choroid thickness was measured at 3 distinct locations across the macula, subfoveal and 2000 microns nasal and temporal to the fovea.



**Figure 2.15:** (A) En face SLO image with the location of the B-scan in green and markers for measuring thickness (crosses) and area (line) in red. (B) B-scan showing chorioretinal structures with manual grading and automatic segmentation overlaid.

Figure 2.15 shows three choroid thickness markings measured by the manual grader in green and the choroid boundaries delineated using GPET. Note that the reference lines used by the manual grader are also visible, measuring approximately 2000 microns temporal and nasal to the foveal pit along the RPE-Choroid boundary.

#### 2.3.1.4 Automatic measurement of choroid thickness

GPET was used to segment the RPE-Choroid and Choroid-Sclera boundaries in every OCT B-scan following the pipeline outlined in figure 2.13.

Automated choroid thickness was measured at the same subfoveal, temporal and nasal locations as manual grading but using two separate approaches, yielding *perpendicular* measurements and *parallel* measurements. *Perpendicular* choroid thickness was measured locally *perpendicular* to the RPE-Choroid boundary, while *parallel* choroid thickness measured parallel to the line drawn by the manual grader. These lines were overlaid on each B-scan

using the calliper tool (figure 2.15, green), allowing an automated procedure to detect each lines' endpoints and measure thickness between the GPET-detected boundaries, as well as measure the difference in angle between perpendicular and parallel lines drawn by GPET. These parallel measurements mimic any potential angular error made by the manual grader, which we know can have an inordinate impact on choroid thickness measurements (section 1.7.1.3).

### 2.3.1.5 Statistical analysis

To determine any systematic or proportional differences between the two approaches, we measured agreement between approaches by comparing (parallel) automatic measurements with manual ones using correlation plots. We also investigated residuals, calculated by subtracting manual measurements from automatic ones (automatic - manual), using a Bland-Altman plot [227]. Mean absolute error (MAE) and Pearson correlation coefficient were also computed between both approaches, stratified by cohort and macular location. To compare the consistency of manual measurements across longitudinal data, we investigated the angular deviation made by the manual grader. This was done by computing angles between the manual calliper measurements made from manual grading against the perpendicular automatic choroid thickness measurements (whose perpendicularity was determined programmatically and thus acts as ground truth). Assessment of the choroid and quantitative analyses comparing manual and automatic measurements were done in Python (version 3.11.3).

Rahman, et al. [106] found an unsigned residual of 32  $\mu\text{m}$  as a threshold suggested to exceed inter-observer variability between manual graders of OCT B-scans. Therefore, we defined major discrepancies between automatic and manual choroid thickness to be any residual greater than 32  $\mu\text{m}$  in absolute value. These were selected for external adjudication by a clinical ophthalmologist (Supervisor Dr. Ian J.C. MacCormick). For each discrepancy, the adjudicator (I.M.) was shown two choroid thickness measurements overlaid on the corresponding OCT B-scan using equipment described in appendix A.2, and

was blinded to which approach was used in each case. He was asked to rate each measurement in terms of its accuracy in measuring thickness from the upper to lower boundaries. He was also asked to rate the overall quality of the OCT B-scan regarding the visibility of the Choroid-Sclera boundary. Quality and ratings were graded using a 3-ordinal scale: “bad”, “okay” and “good”. The adjudicator followed a protocol set out in appendix A.3 for measuring image quality, while the scales for rating the thickness measurements were based on the definitions set out in appendix A.1. Finally, I.M. was asked to select the approach which was preferred — options for both or neither approaches were also available.

### 2.3.2 Results

Performance results between automatic and manual choroid thickness measurements are shown in table 2.3. There was good agreement between manual and automatic measures, with an average residual of  $1.8 \pm 22.0\mu\text{m}$  and MAE of  $14.1 \pm 16.9\mu\text{m}$  across all 480 choroid thickness measurements. The largest residuals were at the subfoveal macular location. In table 2.4 we show the estimated slope and intercept coefficients, as well as the Pearson correlation coefficient, which show excellent agreement after stratifying across location and cohorts.

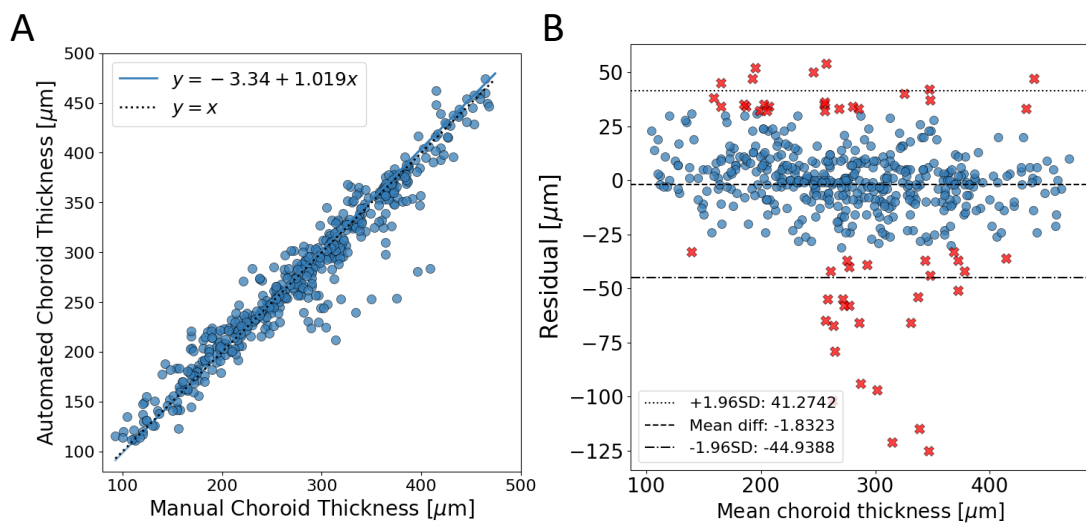
Figure 2.16 shows the correlation and Bland-Altman plots comparing GPET’s measurements against manual ones. Across the whole sample of 480 mea-

|            |           | GPET ( $\mu\text{m}$ ) | Manual ( $\mu\text{m}$ ) | Residual ( $\mu\text{m}$ ) | MAE ( $\mu\text{m}$ ) |
|------------|-----------|------------------------|--------------------------|----------------------------|-----------------------|
|            | All       | $278 \pm 83$           | $277 \pm 79$             | $1.8 \pm 22.0$             | $14.1 \pm 16.9$       |
| Donors     | Temporal  | $290 \pm 65$           | $285 \pm 57$             | $5.1 \pm 18.4$             | $13.8 \pm 13.2$       |
|            | Subfoveal | $287 \pm 77$           | $293 \pm 71$             | $-5.9 \pm 17.4$            | $13.4 \pm 12.5$       |
|            | Nasal     | $215 \pm 71$           | $213 \pm 65$             | $2.1 \pm 18.8$             | $12.9 \pm 13.8$       |
|            | Temporal  | $304 \pm 61$           | $302 \pm 58$             | $2.9 \pm 24.3$             | $13.5 \pm 20.4$       |
| Recipients | Subfoveal | $338 \pm 79$           | $336 \pm 77$             | $3.0 \pm 28.1$             | $17.3 \pm 22.3$       |
|            | Nasal     | $235 \pm 78$           | $233 \pm 72$             | $3.8 \pm 21.0$             | $13.8 \pm 16.2$       |

**Table 2.3:** Performance evaluation between manual and automatic measurements, stratified by cohort and macular location. Where appropriate, values are shown as mean  $\pm$  SD.

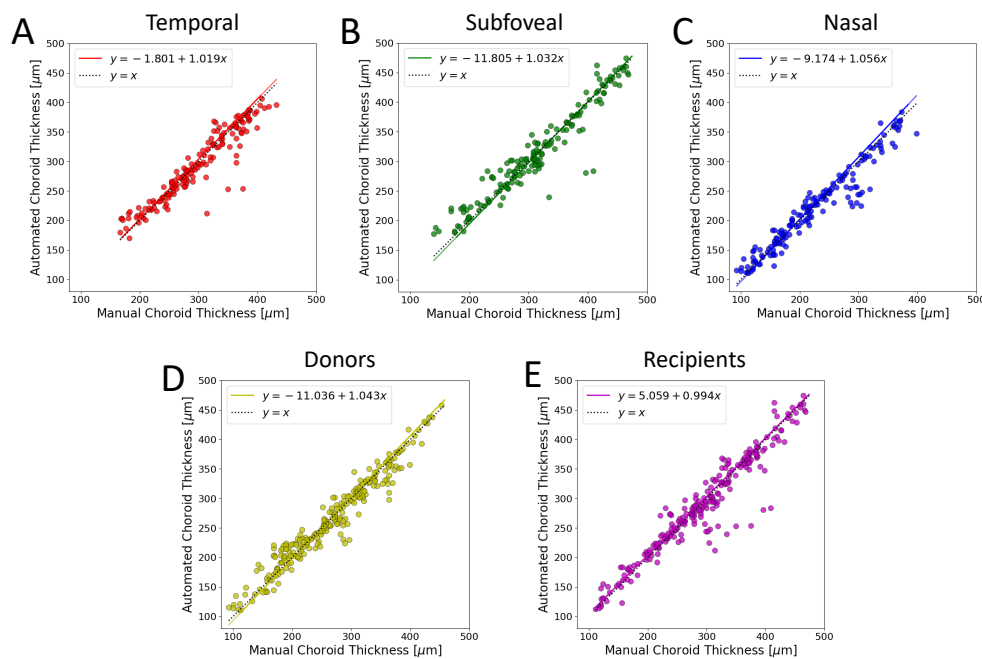
| Metric   |            | Slope [95% CI]    | Intercept [95% CI] ( $\mu\text{m}$ ) | Pearson correlation |
|----------|------------|-------------------|--------------------------------------|---------------------|
| Location | All        | 1.02 [0.99, 1.04] | -3.34 [-10.77, 3.50]                 | 0.97                |
|          | Temporal   | 1.02 [0.96, 1.07] | -1.80 [-17.60, 16.60]                | 0.94                |
|          | Subfoveal  | 1.03 [0.98, 1.08] | -11.81 [-27.97, 3.36]                | 0.96                |
|          | Nasal      | 1.06 [1.01, 1.10] | -9.17 [-19.83, 0.83]                 | 0.97                |
| Cohort   | Donors     | 1.04 [1.01, 1.07] | -11.04 [-20.42, -3.00]               | 0.97                |
|          | Recipients | 0.99 [0.96, 1.03] | 5.06 [-6.53, 16.56]                  | 0.96                |

**Table 2.4:** Correlation linear fits between manual and automatic measurements. Where appropriate, values are shown as mean  $\pm$  SD. All Pearson correlations were statistically significant with  $p < 0.0001$ .



**Figure 2.16:** GPET's population-level agreement against manual choroid thickness. (A) Correlation plot of manual and (parallel) automatic measurements. (B) Bland-Altman plot showing residual values, with 52 major discrepancies highlighted as red crosses whose residual are greater than  $32\mu\text{m}$  in absolute value.

measurements compared, the linear fit estimated a slope value of 1.02 (95% confidence interval of 0.99 to 1.04) and an intercept of  $-3.34\ \mu\text{m}$  (95% confidence interval of  $-10.77\ \mu\text{m}$  to  $3.50\ \mu\text{m}$ ) (figure 2.16(A)). These 95% confidence intervals include 1 and 0, respectively, suggesting that there is no real systematic bias or proportional bias between manual and automatic methods to measure choroid thickness (according to the automated, parallel measurements) [228]. This was also the case when stratified across macular locations and cohorts,



**Figure 2.17:** Correlation plots comparing manual and automatic choroid thickness measurements, stratified by macular location and cohort. Macular locations shown in (A – C), with (A) 2000 microns temporal to fovea, (B) subfoveal and (C) 2000 microns nasal to fovea. Cohorts shown in (D – E), with (D) donors and (E) recipients.

with exception for donors and nasal measurements, which showed slight proportional and systematic bias, as shown in table 2.4 with supporting correlation plots shown in figure 2.17.

Bland-Altman analysis on figure 2.16(B) computed an average residual of  $-1.83 \mu\text{m}$  with limits of agreement of  $-44.94 \mu\text{m}$  to  $41.27 \mu\text{m}$ . Approximately 11% of all 480 choroid thickness measurements (52 / 480) exceeded an unsigned residual of  $32 \mu\text{m}$  [106].

Table 2.5 presents the results from the external adjudication of all 52 major discrepancies. Overall, the automatic choroid thickness measurements scored higher in terms of preference, where 79% (41/52) of the automatic thicknesses were preferred compared to 58% (30/52) for manual thicknesses. Moreover, the automatic approach scored higher in terms of quality, with only 1 measurement judged as “bad”, versus 10 measurements judged “bad” for the manual approach. In addition to manual measurements voted “bad” because they



| C-S Junction<br>Visibility | Score     |        |
|----------------------------|-----------|--------|
|                            | Automated | Manual |
| Good ( $N = 9$ )           | 8         | 2      |
| Okay ( $N = 15$ )          | 9         | 13     |
| Bad ( $N = 28$ )           | 24        | 15     |
| Total ( $N = 52$ )         | 41        | 30     |

| Method    | Discrepancy Measurement     |
|-----------|-----------------------------|
| Automated | Good: 32, Okay: 19, Bad: 1  |
| Manual    | Good: 24, Okay: 18, Bad: 10 |

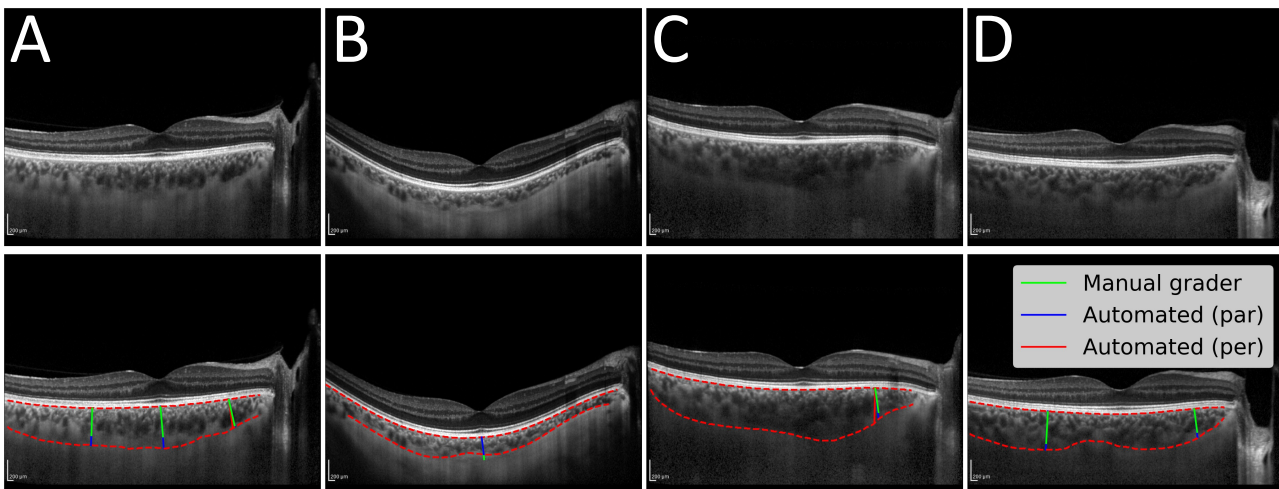
**Table 2.5:** (Top) Numerical score of both approaches from masked adjudication of 52 major discrepancies, stratified by the visibility of the Choroid-Sclera junction. (Bottom) Qualitative preference scores for both approaches.

failed to identify the Choroid-Sclera junction, the adjudicator noted that the manual grader did not follow the protocol consistently and failed to measure locally perpendicular to the RPE-Choroid boundary.

Figure 2.18 shows four major discrepancies from four different OCT B-scans. Red, blue and green lines represent perpendicular, parallel and manual measurements, respectively. The lines in red are shown in panels (A,C) to demonstrate the observed angular deviation between perpendicular automatic and manual measurements. Figure 2.18(A) shows temporal and subfoveal major discrepancies from a donor, residuals  $97 \mu\text{m}$  and  $114 \mu\text{m}$ , respectively. These discrepancies were the largest, with adjudicator I.M. preferring GPET's thickness measurements. Additionally, figure 2.18(A) shows manual angular deviation from perpendicular at the nasal macular location ( $5.4^\circ$ ), resulting in perpendicular and parallel residuals of  $22 \mu\text{m}$  and  $2 \mu\text{m}$ , respectively.

In figure 2.18(B), we observe a subfoveal major discrepancy in a recipient with residual  $-47 \mu\text{m}$ , of which adjudicator I.M. preferred the manual grader. In figure 2.18(C), we observe angular deviation at the nasal macular location ( $8.1^\circ$ ), resulting in perpendicular and parallel residuals of  $106 \mu\text{m}$  and  $79 \mu\text{m}$ , respectively. Finally, figure 2.18(D) shows temporal and nasal major discrepancy from a donor, residuals  $66 \mu\text{m}$  and  $65 \mu\text{m}$ , respectively. The adjudicator preferred GPET's thickness measurements for the choroids in panels (C–D), but noted significant error in the manual grader failing to measure perpendicularly





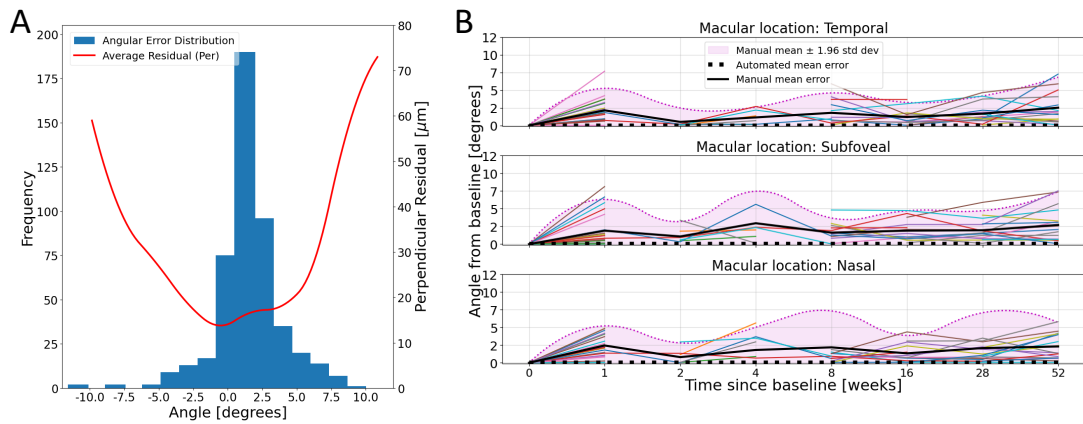
**Figure 2.18:** A selection of major discrepancies between automatic and manual choroid thickness measurements.

in panel (C).

Interestingly, we found that deviation in the angle of manual measurement from perpendicular led to a disproportionate error between manual and perpendicular automatic choroid thickness measurements. This is illustrated in figure 2.19(A) which shows the distribution (mean  $0.5^\circ$  and standard deviation  $2.5^\circ$ ) of angular errors made from manual grading. Absolute differences between manual and perpendicular automatic measurements grew quadratically as the angular error deviated from 0 (red curve). We also found large deviation in manual grading when tracking the longitudinal, within-patient angular difference from baseline, as shown in figure 2.19(B). Distributions at each time point were estimated through empirical mean and standard deviation values. Note that the mean fluctuation in manual grading was always above 0 in comparison to the mean fluctuation from GPET which was constant at 0 (by design).

## 2.4 Discussion

Low signal quality, vessel shadowing, lack of signal penetration and the potential for extravascular fluid to obscure and occlude the choroid all contribute to



**Figure 2.19:** (A) Distribution of angular errors from manual choroid thickness measurements (blue). Average unsigned difference between manual and automated perpendicular thickness measurements (red). (B) Unsigned, longitudinal within-patient angular deviations, relative to baseline measurements. Coloured lines represent distinct individuals using manual measurements.

the challenging task of tracing the Choroid-Sclera boundary in OCT B-scans. However, GPET has the ability to overcome many of these problems due its tunable hyperparameters and model uncertainty quantification, the latter of which many other semi-automatic methods based on graph theory and level sets cannot take into account (table 1.1).

The algorithm can successfully trace discontinuous edges through its ability to interpolate between areas of high image gradient response. The hyperparameters of the selected covariance function governs the properties of the types of curves which are used to model the choroid boundaries, and can be chosen to reflect the underlying properties of the RPE-Choroid and Choroid-Sclera boundaries. Combining this with being able to quantify uncertainty in regions of discontinuity, the model’s curve sampling scheme can search for edge pixels which conform to the properties of these interpolates, particularly around areas of known edge pixels. Therefore, it is well adept to accurately quantifying and interpolating occluded regions caused by vessel shadowing or poor reflectivity.

A huge advantage to GPET over larger models which require expertly annotated datasets, such as deep learning-based approaches, is that it requires no prior training. GPET learns each boundary on-the-fly, enabling each fitting pro-

cedure to be tuned to the underlying properties of both choroidal boundaries. Moreover, the methodology provides an analytical and functional representation of the edge whose fitting procedure is both interpretable and explainable. These attributes are often lacking in many deep learning models employed in medical image segmentation.

The GPET algorithm is also flexible in terms of input and usage. While the fitting procedure requires initial guidance from the edge endpoints inputted manually, GPET can also be further initialised with intermediary pixels between the two endpoint edge pixels in cases where there is poor visualisation along the Choroid-Sclera boundary. Selecting the edge endpoints and an additional pixel to guide tracing around the central column of the edge would likely add a negligible amount of time for the end-user, but potentially lead to improved segmentation performance. Finally, while this was not the focus of this chapter, GPET can also be used for processing an OCT volume scan in a two-step procedure. Once the first B-scan in the stack is fully traced, a subset of the edge pixels in the output can be used as part of the initialisation stage for the next B-scan. Thus, by recursively propagating the output trace of one B-scan to be used as the initialisation for consecutive B-scans, an entire volume scan can be traced in a single sweep. However, the accuracy of such volumetric segmentations has not been fully investigated.

### **2.4.1 Evaluation**

We found that GPET could replace human measures of the choroid. We observed no systematic or proportional differences when comparing manual to automatic measurements, when measuring parallel to the grader. Advantageously, automatic choroid thickness eliminated angular error and therefore had greater precision when assessing within-patient longitudinal data. The elimination of inconsistent measurement angle was achieved by GPET, since the automatic method is always able to measure choroid thickness locally perpendicular at the same location and across all time points for each patient. This is an important attribute of measurement protocols so that choroidal curvature or skew can be accounted for. In contrast, error in the angle of manual

measurement had a disproportionate impact when comparing these to the perpendicular automated thickness measurements.

Measurement error is a major problem in research studies, and especially in longitudinal studies where repeated measurements are made from highly sensitive retinal imaging modalities. The RPE-Choroid boundary is rarely parallel to the imaging axis across the width of the B-scan, making it a challenge for humans to measure perpendicular angles by eye across repeated measures, which is a major source of error. The extent of angular error from manual grading across individual patients (figure 2.19(B)) necessarily adds noise into the signal being measured, and is likely to have a significant negative impact for predicting clinical outcomes.

What exacerbates this potential measurement error is the extent that choroid thickness can change as the angle deviates from perpendicular, as exemplified by the quadratic relationship between GPET measurements made perpendicularly against the manual ones. The performance evaluation of the automatic method would have been unjustly skewed had only perpendicular measurements been used for the analyses seen in figure 2.16(A). This is because the default presentation for OCT B-scans have different axial and lateral pixel length-scales, which can result in the slightest change in marking a lateral position during manual grading to misrepresent the true choroid thickness. While this can be accounted for by viewing the B-scan in '1:1 micron mode', this squashes the choroid in the axial direction, which can make manual measurements even more prone to error [195] (section 1.7.1.2).

Rahman, et al. [106] reported an average inter-rater agreement of 32  $\mu\text{m}$  between manual raters for measuring choroid thickness, which sets a high threshold to exceed to ensure that the signal measured manually isn't driven by human error, but true biological change. These observations from manual grading ultimately highlight a potential issue of reporting manual measurements within choroid image analysis, and make the case for releasing methods which can provide consistent and clinically meaningful measurements of the choroid, particularly in cases where effect sizes are likely to be small, such as in tracking myopia progression, which has seen effect sizes approximately in the range of 20 $\mu\text{m}$  – 30 $\mu\text{m}$  [24, 110].

Much of the discrepancy between manual and automatic measurements arose from choroidal features that are inherently difficult to measure. For example, the Choroid-Sclera boundaries of larger choroids (figure 2.18(A,C,D)) are prone to poor quality image acquisition due to higher incidence of poor signal penetration. However, the external adjudicator preferred the automatic measurements over the manual grader for the major discrepancies in panels (A, C, D). This was because of the propensity for the manual grader to fail to measure perpendicularly, as this inconsistency becomes more obvious (and larger in error) with larger choroids. However, the manual grader was preferred for the discrepancy shown in panel (B) because GPET failed to identify the Choroid-Sclera boundary underneath the fovea because it was obscured as the posterior of Haller's layer was low contrast.

The choroid in figure 2.18(A) represents a source of major disagreement between the manual grader and the automatic approach. The manual grader defined the Choroid-Sclera boundary as the junction below the most visible posterior vessels (green), while the automatic approach has successfully identified the boundary as a smooth and continuous contour tracking posterior vessels with much lower visibility (red), of which the external adjudicator agreed with.

The choroids in figure 2.18 represent just over half (52%) of all 52 major discrepancies lying outside the threshold suggested to exceed inter-observer variability between graders of EDI-OCT images [106]. That is, most outliers came from the same four choroids at different time points. Moreover, of the 28 thickness measurements from choroids described as having bad visibility (table 2.5), 20 of these came from the larger choroids (figure 2.18(A,C,D)). Thus, it is encouraging to see that most areas of poor disagreement come from large choroids with poor visibility due to degrading optical signal.

## 2.4.2 Limitations and future work

GPET provides a more consistent approach to measuring choroid thickness than manual grading, and permits measurement of more robust metrics such as choroid area and volume (across the macula). These are two and three

dimensional measurements which are able to average out measurement error posed by one-dimensional point source metrics like thickness. Moreover, as discussed in section 1.7.1.3, manual thickness is more sensitive to changes in scaling and the presentation of the OCT B-scan [195]. Thus, we expect it to be a suitable replacement for manual methods for choroid image analysis in OCT image sequences. GPET has also been released as open-source on GitHub to enable the research community to promote the transition from manual annotation to automatic measurement.

However, while the fitting procedure of GPET is fully automatic, the entire image analysis pipeline is not. The pipeline involves a series of pre-processing steps to obtain the desired choroid region segmentation. This includes denoising, contrast enhancement and edge map detection. Thus, the success rate of the fitting procedure is completely dependent on the quality of the resulting edge maps, which can have a significant effect on the final edge trace and thus downstream choroidal measurements.

Failure to trace the boundaries correctly in a one-shot fashion will require further manual intervention to tune the hyperparameters (of the covariance function), select intermediary pixels between the edge endpoints, or refine the pre-processing steps to enhance the image gradient. All these additional steps not only adds precious time to the end-user, but also requires domain knowledge of the algorithm's steps, such as image processing techniques for obtaining edge maps, and underlying mathematical theory of Gaussian process regression for hyperparameter tuning. The domain-specific knowledge necessary for applying GPET ultimately prevents this method's accessibility to those who are likely to apply segmentation methods most to choroidal image analysis in OCT image sequences, i.e. imaging technicians who may have little experience in mathematics, programming or image processing.

Automated tools to assist prognosis and treatment will strengthen future studies, enabling the delivery of robust, reproducible and responsible research in areas which have previously been at risk of human error. However, the aforementioned limitations of which many semi-automatic approaches suffer contribute to the barrier between research software and their routine use in the broader research community (i.e. clinicians and technicians), regardless of

their ability to produce more consistent measurement of the choroid compared to manual grading.

While GPET provides a progressive step forward in obtaining the goal of producing standardised choroid measurement, this method leaves a lot of space for developing better methods. These methods would be ones which are not multi-stage, inter-dependent pipelines or require manual intervention/initialisation, and whose accessibility does not require pre-requisite knowledge of theory in image analysis, mathematics or computer science. In an ideal world, these methods would be fully automatic and require very little effort to run regardless of the end-user's experience, and would be deterministic in nature such that their approach to choroidal image analysis in OCT image sequences would be robust and reproducible.

### 2.4.3 Outputs

In this chapter, there have been two publication outputs, both of which has been published and peer-reviewed (by December 2024). The author's name is in bold type.

- **Burke, Jamie**, and Stuart King. “*Edge tracing using Gaussian process regression.*” IEEE Transactions on Image Processing 31 (2021): 138-148.
- **Burke, Jamie**, Dan Pugh, Tariq Farrah, Charlene Hamid, Emily Godden, Thomas J. MacGillivray, Neeraj Dhaun, J. Kenneth Baillie, Stuart King, and Ian J.C. MacCormick. “*Evaluation of an automated choroid segmentation algorithm in a longitudinal kidney donor and recipient cohort.*” Translational Vision Science & Technology 12, no. 11 (2023): 19-19.

The software associated with this method, GPET, has been published as open-source and is freely available on GitHub [here](#).

### 2.4.4 Executive summary

In this chapter, we introduced a semi-automatic, edge-based approach to choroid region segmentation using Gaussian process regression, Gaussian Process Edge Tracing (GPET). Advantageously, GPET is able to consider local and global optimisation through use of a pre-processed edge map and a Bayesian recursive scheme using Gaussian process regression. This is particularly relevant for detecting the Choroid-Sclera boundary, which suffers from optical signal degradation and choroid stroma obstructions. GPET removes the need for intensive, manual segmentation of the choroid and reduces the degree of human subjectivity injected into the measurement procedure. We found that GPET agreed strongly with manual measurement of choroid thickness and, importantly, was more consistent at measuring thickness in the presence of choroid skew and curvature. This is crucial for producing standardised measurements of the choroid in OCT image sequences.

However, GPET is a multi-stage pipeline, requiring manual initialisation and crucial knowledge of the underlying theory of Gaussian processes to work effectively. These limitations significantly impact GPET's accessibility to the wider research community and limits its ability to produce reproducible measurements given the, albeit minimal, human subjectivity injected into the segmentation procedure. Therefore, future work should consider a more fully automatic approach which does not require specialist training, pre-processing or manual intervention.

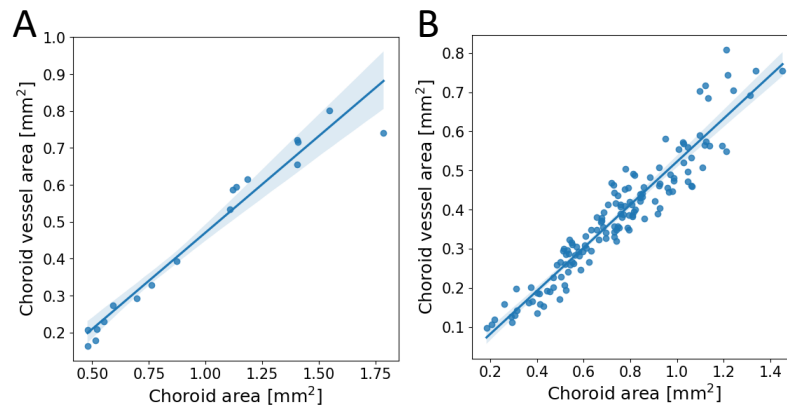


## Chapter 3

# MMCQ: Semi-automatic choroid vessel segmentation using multi-scale median cut quantisation

### 3.1 Introduction

In the last chapter we developed a semi-automatic procedure to segment the choroidal space, which enables measurement of choroid-derived thickness and area in a single OCT B-scan (and volume across the macula). However, the choroid is primarily a vascular layer, with the highest blood flow per unit tissue weight of any body organ [7]. Thus, enabling quantification of the size of choroidal vessels would be significantly more informative than the overall choroidal space. Understanding pathophysiology through vascular metrics, instead of regional metrics, may provide a more stable and representative way of studying pathophysiological changes in the presence and during treatment of systemic disease which affect the microvascular system [158]. Because of the historic challenges of measuring the choroidal vasculature from OCT reliably, choroidal thickness has been used instead as a surrogate.



**Figure 3.1:** Relationship between choroidal total and vessel area for two cohorts described in sections 6.2 (A) and 6.3 (B).

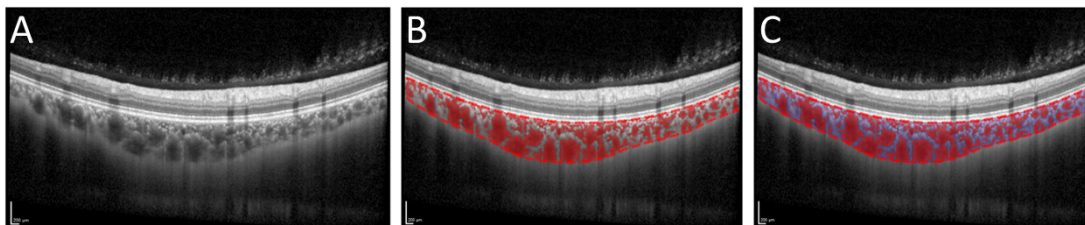
It may be that an increase in choroidal vessel density is directly proportional to changes in the choroidal space. In figure 3.1 we visualise the relationship between total choroid area against choroid vessel area for two populations discussed in more detail in sections 6.2 and 6.3. In each case, measurements were taken from a 6 mm, fovea-centred ROI from horizontal-line EDI-OCT B-scans and each scatter-point represents an individual eye. See section 1.7.1.3 for more details on how total area and vessel area were measured and sections 6.2 and 6.3 for information on the relevant populations. In both cases, vessel area and total choroid area are significantly correlated with each other (panel A, Pearson  $r=0.99$ ,  $P<0.0001$ ; panel B, Pearson  $r=0.95$ ,  $p<0.0001$ ). Thus, information on the choroidal space may reflect that of the choroidal vasculature, but measuring the choroidal vasculature directly is likely to be more representative of the choroid if done accurately.

A popular metric to quantify the choroidal vasculature in an eye is the choroid vascular index (CVI), which is a measure of an individual choroid's vessel density in an OCT B-scan. CVI describes the proportion of the choroid in an OCT B-scan made up of vessels, and is thus a dimensionless ratio, bound between 0 and 1. In an OCT B-scan, CVI is simply the number of pixels classified as choroid vessel divided by the number of pixels classified as choroid. The choroidal space is made up of interstitial fluid, or stroma, seen as brightly illuminated regions in the OCT B-scan, with interspersed, irregular areas of darker

intensity representing choroidal vasculature. This has been both empirically observed [103, 104] and widely accepted among the research community [28].

CVI has become particularly promising metric in choroidal image analysis and biomarker investigation because of its potential invariance to certain physiological factors which often affect choroid thickness like systemic blood pressure [158]. This is a direct consequence of its formula, as CVI is inversely proportional to choroid regional metrics, and its signal is primarily driven by choroidal vascular metrics. Thus, if the magnitude of change in choroidal thickness reflect a similar magnitude of change in choroidal vessel area, the change in CVI will be minimal. Therefore, CVI has been observed as a more robust index of choroidal variation compared with a one-dimensional, point-source measurement like subfoveal choroid thickness [83]. This relatively new choroidal measurement has been used among the research community to quantify the choroidal vasculature in ocular pathology — see Agrawal, et al. [28] for an extensive review on CVI in chorioretinal disease. Figure 3.2 shows an OCT B-scan with region and vessel segmentations overlaid.

Estimating CVI requires segmentation of the choroidal vasculature, which is a significantly more challenging task than choroid region segmentation. Segmenting the choroid region only requires a single shape of heterogeneous vascular structure to be identified, which sits below the hyperreflective RPE layer and above the sclera which appears relatively homogeneous in pixel intensity — two well-defined landmarks. Conversely, segmenting the choroidal vessels is a much more ambiguous task. In terms of size and shape, the



**Figure 3.2:** Vessel segmentation in an OCT B-scan. (A) OCT B-scan. (B) Choroid region demarcated with dashed red lines and vessel segmented as red blobs. (C) CVI is measured as the ratio of red pixels to the sum of red and shaded blue pixels.

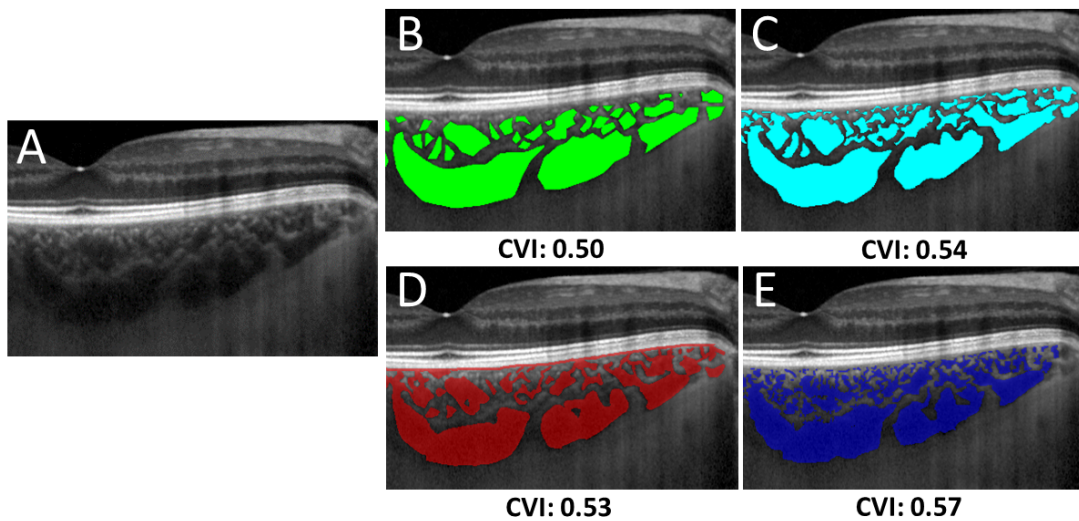
choroid contains a dense, heterogeneous population of vessels interspersed and suspended by connective tissue in its interstitial space. Cross-sectional image capture using OCT of this microvascular bed often means that individual choroidal vessels are captured at oblique cross-sections and thus do not have clear vessel wall boundaries. This is exacerbated by poor illumination, contrast and inherent speckle noise. This can make manual segmentation prohibitively time consuming, and has the potential to be impacted by human subjectivity.

Figure 3.3 highlights the potential variation in vascular metrics such as CVI when manual segmentation is used for choroid vessel segmentation. Three raters were asked to manually segment the same choroid (supervisor Dr. Ian J.C. MacCormick, author Jamie Burke and image technician Charlene Hamid) using ITK-Snap, which permits pixel-level annotation [229]. The segmentations and resulting CVI are shown in panels (B – E).

While the segmentations from panels (B, C) were done according to each graders' predisposed notion of the choroidal vasculature and without measurement protocol, the segmentations in panels (D, E) followed the measurement protocol outlined in appendix A.5 using definitions and equipment from appendices A.1 and A.2 (grader I.M. segmented the choroid twice, 2 years apart).

Note the significant variation in granularity from each segmentation, and resulting difference in CVI. Importantly, even following the same protocol (panels (D, E)) led to a difference in CVI of 4%, and the largest difference of 7% came from the same individual (panels (B, E)) after a time difference of 2 years before and after a well defined measurement protocol. These error are likely to be clinically significant, given Agrawal, et al.'s review [28] suggested effect sizes between 2% – 6% between healthy and diseased eyes.

While this only describes the potential error in manual vessel labelling for a single example, it alludes to a wider issue around manual labelling in and of itself. Thus, in lieu of manual annotation, local thresholding algorithms became commonplace for this segmentation task [27, 104, 158] (see section 1.6.2 for a brief literature review on the topic). Briefly, the most widely used approach is the Niblack algorithm [28, 160]. This adaptive thresholding algorithm is semi-automatic and requires parameter specification, and we have briefly described



**Figure 3.3:** Poor agreement between manual vessel segmentation labels. (A) OCT B-scan. (B – E) Manual choroid vessel segmentations from three raters, supervisor Ian J.C. MacCormick (B, E), image technician Charlene Hamid (C) and author Jamie Burke (D). The corresponding CVI is annotated, using the same choroid region segmentation to facilitate interpretation.

its potential problems (see section 1.6.4.1). Importantly, very few studies report the parameter configuration used for Niblack [163, 167]. Instead, many utilise the version supported by Fiji (ImageJ) [29] and either manually tune the internal algorithm parameters for every B-scan, or use the default setting. Moreover, ad-hoc pre-processing steps such as manual demarcation of the ROI [158] and manual measurement of the average pixel intensity of specific choroidal vessels [27] are typically used in research. These reasons alone can lead to under-specification of the algorithm, potentially poor quality segmentation and non-standardisation across studies and populations.

Therefore, we propose that the inconsistent and widespread usage of Niblack’s method has the potential to prevent the reporting of reproducible and standardised ocular measurements of the choroidal vasculature. We further propose that there is space for better methods which are able to target the core challenges of choroid vessel segmentation, which a fixed parameter setup such as Niblack, whose creation was designed around word document thresholding [161], is not able to address. Finally, Niblack is likely the most widely used because of its availability on Fiji (ImageJ) [29], or through membership-only

access [160], while other state-of-the-art methods based on deep learning are not openly available [169, 172]. Thus, there is an unmet need for more standardised approaches to choroid vessel segmentation which are domain-specific and freely available to the research community.

Because of the poor contrast and heterogeneity of the choroidal vasculature as seen on OCT, we hypothesised that a multi-scale procedure based on enhancement would address these two issues. Here, we propose a multi-scale approach to choroidal vessel segmentation, after sufficiently enhancing the choroidal vasculature through an extension to contrast limited adaptive histogram equalisation. This method performs image quantisation, enhancement and pixel clustering at several scales designed for the accurate segmentation of both small- and large-scale vasculature in the choroid seen in EDI-OCT B-scans. Importantly, we sought to develop a method which did not require a human element, and which was reproducible, high quality and aimed to release it as open-source.

### 3.1.1 Niblack autolocal thresholding

The Niblack algorithm [161] is a local thresholding technique which binarises individual pixels within an image based on the summary statistics of a patch centred on each pixel. For each pixel  $(x, y)$ , a threshold  $T(x, y)$  is estimated using

$$T(x, y) = \mu(x, y) + k \cdot \sigma(x, y), \quad (3.1)$$

where  $\mu$  and  $\sigma$  are the mean and standard deviation within a neighbourhood centred on pixel  $(x, y)$ . The size of this neighbourhood is defined by parameter  $w$ , which should be dependent on the size of the choroidal vessels being segmented.  $k$  is the standard deviation offset, which is a constant that determines the weighting of the standard deviation against the overall mean intensity of the window. Thus,  $k$  defines how sensitive the local threshold value is to local variance — if  $k$  is large and negative then the threshold will be very low and over-segmentation will occur, and vice versa for where  $k$  is large and positive. Muller, et al. [163] reported that this algorithm suffers when a neighbourhood

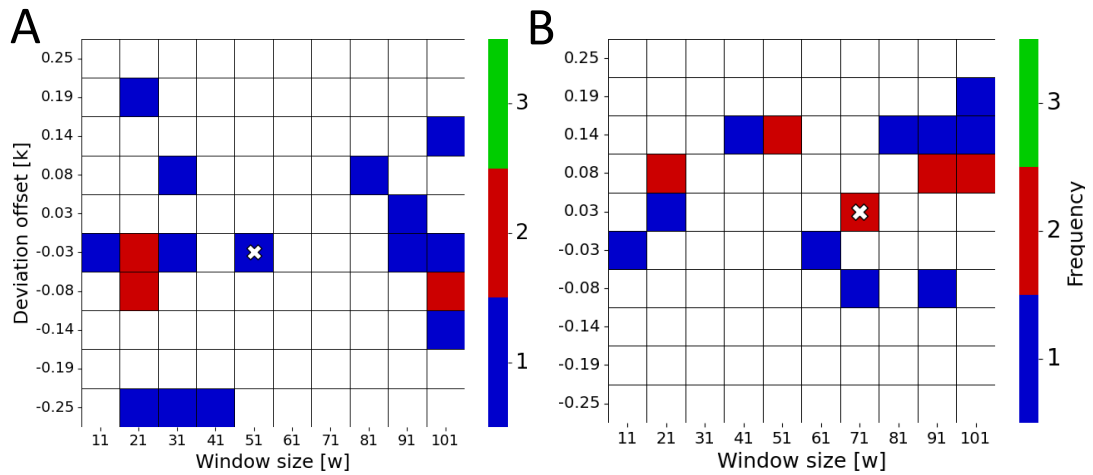
around a pixel is homogeneous in pixel intensity, which can result in noise amplification.

There are two widely used approaches to semi-automatic choroid vessel segmentation which use Niblack's local thresholding method, Sonoda's [27] and Agrawal's [28, 158, 160]. Sonoda, et al. [27] is dependent on the end-user using Fiji (Imagej) [29] and selecting the choroidal space, identifying the central location of three choroidal vessels (of size greater than  $100 \mu\text{m}$ ) and lowering the mean intensity of the B-scan to the minimum value of the three individual vessels' mean intensity. Only after this pre-processing is Niblack's method applied, with no details on parameter specification. This approach is described in detail in the supplementary material S1 of the original paper by Sonoda, et al. [27]. Agrawal's approach removed the need to select individual choroidal vessels, and used several morphological operations to enhance the vessel segmentation [160]. In both cases, manual identification of the ROI to generate measurement is required. To our knowledge, the most detailed description (which is publicly available) of their modified Niblack method — which is available under membership-only — can be found in the work by Betzler, et al. [160]. Importantly, in both cases [27, 28, 158, 160], there is no mention of the window size  $w$  or standard deviation offset  $k$ , which potentially means the default setting is used, or the choice is at the end-user's discretion.

The rule of thumb for selecting  $w$  is based on the size of the choroid and choroidal vessels of interest, while  $k$  is more abstract and therefore more difficult to define as a sensitivity measure. Muller, et al. [163] decided to fix  $k$  and allow manual annotators to tune  $w$ , as this was a more intuitive tunable parameter. During the manual annotation procedure, Muller, et al. [163] reported that varying  $w$  between 20 and 75 with  $k$  fixed as -0.05 resulted in a mean (standard deviation) difference in CVI of  $11.2\% \pm 4.1\%$  (range between 2% – 21%). To put this into context, Agrawal, et al.'s review [28] discussed multiple studies which reported differences in CVI values between healthy and diseased eyes in the range of 2% – 6%. Thus, it would appear that variations of Niblack's internal parameter could have a clinically significant impact to measurements like CVI.

To understand the importance of these two parameters in more detail, we in-





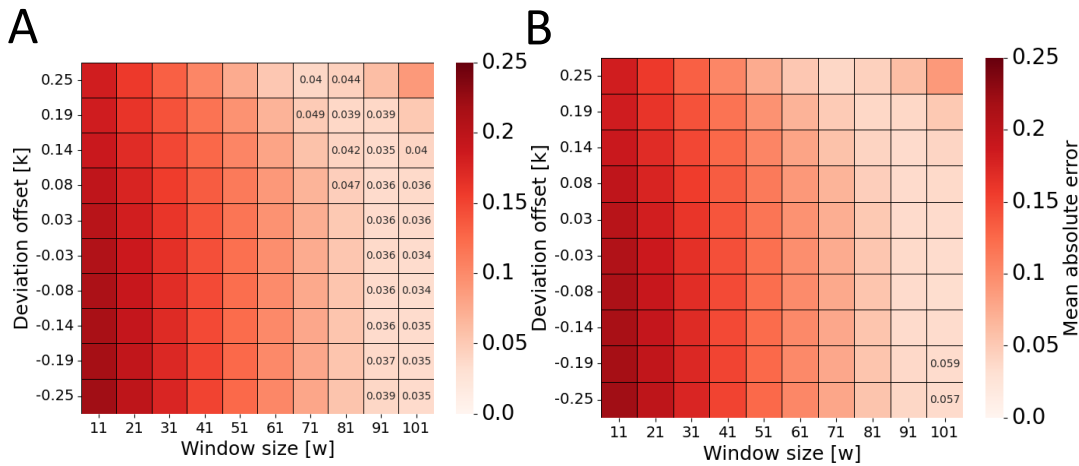
**Figure 3.4:** Frequency map after grid-searching over 100 possible pairs of Niblack parameters  $(w, k)$  and optimising for the best segmentation (using Dice) against raters M1 (A) and M2 (B). The white cross represents the optimal parameter configuration for the same B-scan which had the highest Dice similarity score between the manual segmentations performed by M1 and M2.

investigated the possible variation of CVI and segmentation performance for different paired parameter configurations  $(w, k)$  in a small dataset of 20 horizontal-line, EDI-OCT B-scans of 20 right eyes (10 healthy participants and 10 participants receiving renal transplantation — data sourced from the larger cohort described in section 2.3, table 2.1).

Choroidal vessels were manually segmented by two raters M1, supervisor Dr. Ian J.C. MacCormick; and M2, author Jamie Burke using ITK-Snap [229], which permits pixel-level annotation. Graders followed the measurement protocol outlined in appendix A.5 and were blinded to each others segmentations. Graders used the definitions and equipment outlined in appendices A.1 and A.2 to aid in region and vessel detection. Niblack autolocal thresholding was applied to each B-scan using 100 different parameter configurations (following Agrawal’s procedure [160]), with window size  $w$  varying between  $11 \times 11$  and  $101 \times 101$  over 10 equally spaced intervals, and similarly for  $k$  from -0.25 to 0.25.

An optimal parameter configuration was defined as the pair  $(w, k)$  which had the highest Dice score comparison between the resulting segmentation and



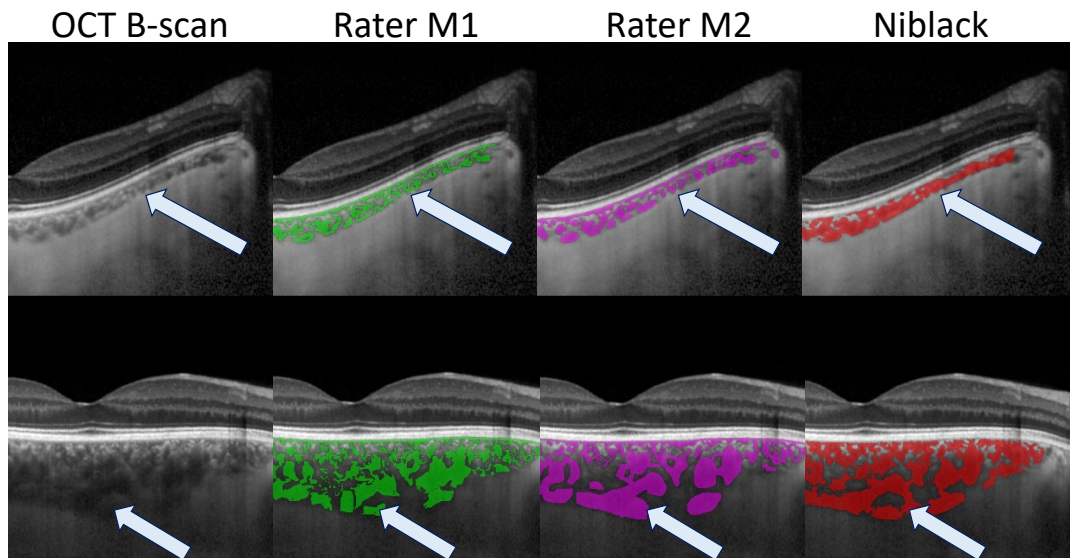


**Figure 3.5:** Heat map of mean absolute errors (MAEs) comparing CVI computed with Niblack, using 100 possible Niblack parameters  $(w, k)$ , against manual graders M1 (A) and M2 (B). Cells annotated are those whose MAE was below 0.06 (6%).

manual segmentation. As the parameter space of  $w$  and  $k$  are discretised, we can count the number of occurrences for each optimal parameter configuration. We also measured the CVI in a standard 6 mm fovea-centred, ROI using each paired parameter setting, and measured the mean absolute error (MAE) between the CVI outputted using Niblack and those outputted using the manual graders' segmentations. MAEs were averaged across the 20 B-scans, creating an overall averaged, MAE of CVI per parameter configuration.

Figure 3.4 shows the frequency count of optimal parameter settings for all 20 B-scans, when compared with manual segmentations from rater M1 (A) and rater M2 (B). While this is only a small dataset, what is clear is the extent of the variation of optimal parameter configurations across the B-scans, and across raters. In fact, for the choroid with the highest Dice similarity between raters (Dice score 0.83), the optimal parameter configuration for Niblack agreeing with M1 was  $(51, -0.03)$  and M2 was  $(71, 0.03)$  (white crosses). The extent of variation of optimal parameter configuration and the different pattern of occurrences both suggest that no one optimal configuration is applicable across all B-scans or applicable across different raters.

Figure 3.5 shows the heatmaps of averaged MAE between CVI measured us-



**Figure 3.6:** Choroid vessel segmentation for two choroids from raters M1, M2 and Niblack (with the same parameter configuration).

ing Niblack and raters M1 (A) and M2 (B). A similar pattern is observed across raters, which suggest that higher window sizes overall produce less error (when compared to manual segmentations), and that this is consistent across most standard deviation offsets. The minimum, averaged MAE for M1 and M2 were 0.034 (3.4%) and 0.057 (5.7%), respectively. Thus, while expert manual grading is not a gold standard, this does raise some concern over the accuracy of Niblack’s CVI, particularly given the small effect sizes observed in the existing literature [28]. Moreover, these minimum MAE values corresponded to different parameter configurations (Minimum MAE for M1,  $w = 101$ ,  $k \approx -0.055$ ; for M2,  $w = 101$ ,  $k = -0.25$ ).

Thus, no one parameter configuration is optimal across different choroids, and it should be tailored for the particular choroid. Figure 3.6 shows the segmentations by raters M1 (green) and M2 (pink) as well as Niblack (red) using the same parameter setting ( $w = 51$ ,  $k = -0.05$  [163]). Compared with the choroid in the bottom row, Niblack over-segments the choroid in the top row (arrows). Quantitatively, Niblack over-segmented both choroids according to the manual graders, producing an average MAE of 0.158 (15.8%) and 0.104 (10.4%) for the top and bottom choroids, respectively (averaged across raters).

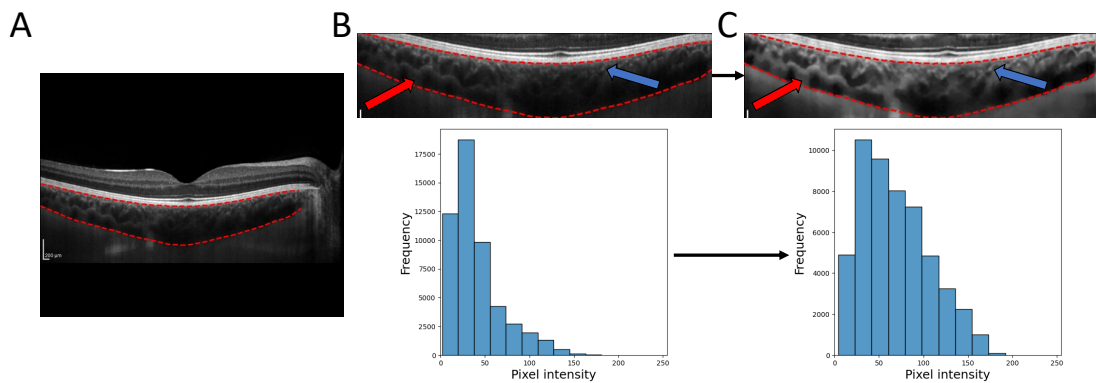
Therefore, we propose that Niblack’s autolocal threshold algorithm may be sufficient for individual choroid analysis, but likely requires significant tuning of parameters for optimal segmentation quality, thus rendering it unfeasible for large-scale choroidal image analysis. Moreover, we hypothesise that poor selection of parameters may result in over-segmentation of the interstitial space, particular in the anterior section of the choroid where smaller vessels and the choriocapillaris are situated.

### 3.1.2 Contrast limited adaptive histogram equalisation

Contrast limited adaptive histogram equalisation (CLAHE) [230] is an image enhancement method for increasing the contrast in an image. For the choroid, increased contrast can improve the definition of vessel walls, thus making the task of identifying vessel from interstitial space an easier one. Often the choroidal vessels have poor contrast with the interstitial space and significantly so for smaller vessels, and CLAHE is an approach to rectify this.

CLAHE is a local enhancement method, which splits the image into patches and adjusts the contrast of each patch locally. For each patch, it adjusts the contrast by spreading out the brightness levels evenly using its histogram, known as histogram *equalisation*. CLAHE is particularly helpful for ensuring that different parts of the image aren’t overly enhanced, as it is able to limit the redistribution of each patch’s brightness levels. Each patch is then combined using interpolation to preserve continuity between the newly enhanced patches. CLAHE has two parameters, one for deciding how many patches to generate (and consequently how local the enhancement will be) and another for deciding the strength of enhancement.

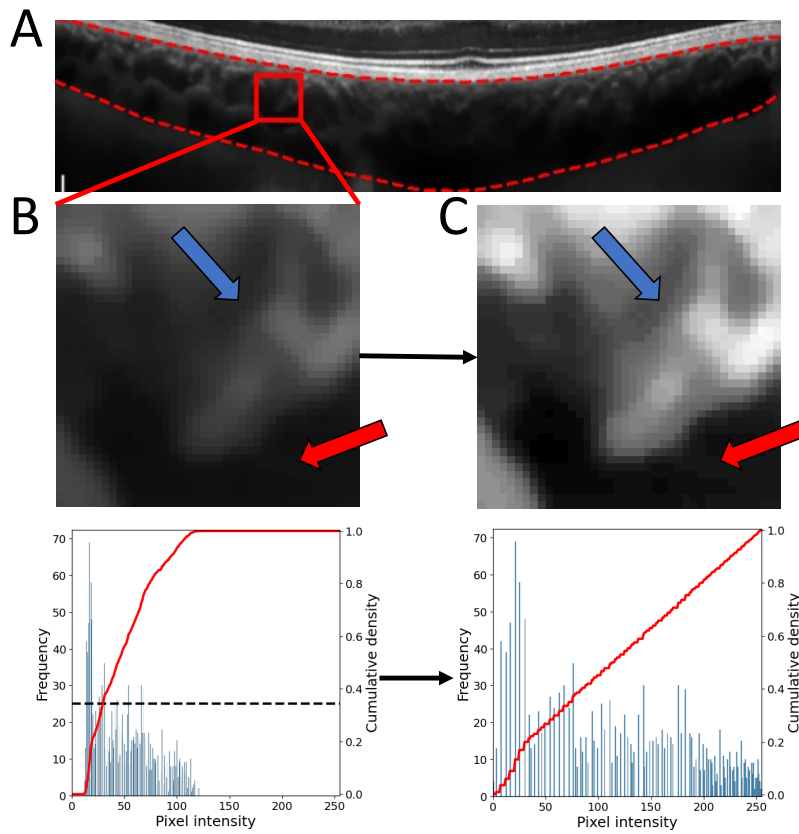
In figure 3.7 we show an OCT B-scan with a particularly low contrast choroid with poor vessel wall definition. After cropping out only the choroid and applying CLAHE by splitting the region into an  $8 \times 8$  grid of patches and limiting each patch’s histogram bins to 25 pixels, choroidal vessels suspended in connective tissue in the interstitial space have greater definition. However, note that the contrast of larger vessels is greater (red arrows) than for smaller smaller vessels (blue arrows).



**Figure 3.7:** Demonstration of CLAHE on an OCT B-scan. (A) OCT B-scan with poor contrast choroidal vessels, with choroid demarcated in red. Cropped choroid and histogram of choroid pixels in raw B-scan (B) and after application of CLAHE (C). Red arrows show good enhancement, blue arrows show poor enhancement.

A closer inspection for a small, square patch in figure 3.8 describes the CLAHE process in more detail. The patch's histogram undergoes a transformation which results in clipping the distribution of histogram bins (brightness levels) and re-distributing excess pixels uniformly across the rest of the histogram bins. Excess pixels are defined by those in each histogram bin which fill it beyond the *clip limit*, seen by the black dashed-line in panel (B). This process helps in preventing over-amplification of noise in relatively homogeneous areas (in pixel intensity) of the image.

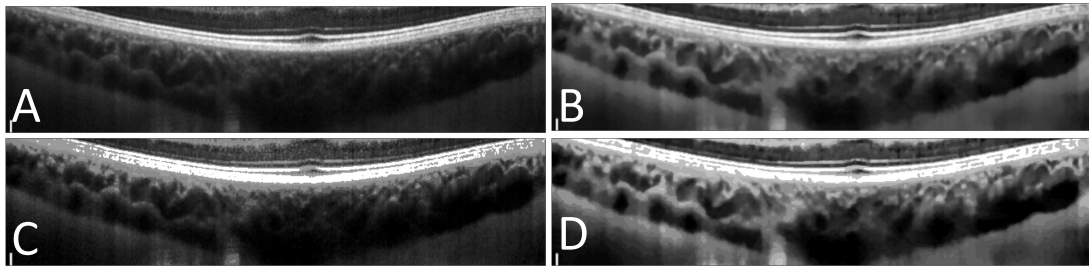
After uniform re-distribution of excess pixels, the resulting histogram is then stretched to cover the possible range of bins (typically 256 for an 8-bit image), which produces the adjusted patch you see in panel (C). From the perspective of cumulative distributions, CLAHE promotes linearity in the cumulative density function for each patch's new histogram, as evidenced by the attempted flattening of the red curve between panels (B) and (C). The linearity of this cumulative density function represents the distributing of pixels evenly across the range of brightness levels. Similar to the global view from figure 3.7, CLAHE is helpful in improving the contrast for larger vessels (red arrows) but not for smaller vessels (blue arrows), whose choroidal vessel walls still have relatively poor contrast relative to the choroidal interstitial space.



**Figure 3.8:** Demonstration of CLAHE at the patch-level during pre-processing. (A) Choroid cropped from OCT B-scan in figure 3.7 with boundaries annotated and a red square representing patch of interest in panels (B, C). Zoomed-in patch before (B) and after (C) contrast enhancement with CLAHE. Below each patch is the patch histograms shown as blue bars with their normalised cumulative distribution function shown in red. In (B), the black dashed line represents the histogram clipping limit. Red arrows show good enhancement, blue arrows show poor enhancement.

### 3.1.3 Median cut image quantisation

A problem with application of CLAHE is its inability to enhance smaller vessels. To overcome this limitation we propose that instead of enhancing the raw pixel intensities, we enhance a quantised equivalent. Quantisation is common in image compression and is simply a way of representing an image using fewer intensity values. Thus, instead of allowing all 256 brightness levels in an 8-bit image, we can reduce (or quantise) this to 16 levels (4-bit) instead. A popular



**Figure 3.9:** Demonstration of median cut quantisation on an OCT B-scan. (A) Raw OCT B-scan. (B) B-scan after median filter denoising and CLAHE contrast enhancement. Median cut quantisation of raw (C) and enhanced (D) B-scans into 20 quantisation levels/clusters.

approach to quantisation is the median cut algorithm [231], and for a grayscale image, this quantisation procedure is in a similar vein to pixel clustering but is far less computationally expensive.

Median cut quantisation clusters are computed by repeatedly splitting the image histogram at its median value into two groups, and splitting each resulting group by their own median value and so on recursively until a desired number of clusters are produced. The quantised image is the result of converting the original intensity of each pixel to the average value of all pixel intensities assigned the same cluster. This simplification of the original image's histogram is particularly useful for assigning pixels to categories of similar pixel value, and can be used at a local patch-based level, i.e. for small patches in the choroid of OCT B-scans.

The method developed in this chapter leverages median cut quantisation to enhance the choroid in OCT B-scans at a local level, and across multiple scales. However, a slight variation made to this algorithm is that the median value for each cluster group is assigned rather than the average value. This is because, relative to the mean, the median value is more robust to outliers and is a better measure of centrality for asymmetrical distributions, such as those found here in the pixel intensity histograms of the choroid in OCT B-scans.

Figure 3.9 shows the raw OCT B-scan and an enhanced B-scan in panels (A) and (B) with their quantised equivalents (using 20 quantisation levels/clusters) in panels (C) and (D). Note that the simpler representation is advantageous for

vessel classification, and that the enhanced OCT B-scan provides a suitable compression for differentiating large vessel from interstitial space, but not for smaller vessels.

### 3.1.4 Vessel shadow compensation

Choroidal vessel visualisation can be corrupted by superficial retinal vessel shadowing, where the OCT beam penetrates through the retinal vasculature lying along the inner surface of the retina and sitting perpendicular to the incident laser light. This consequently darkens the corresponding A-scans of the deeper outer retinal and choroidal structures behind it [232]. This can make choroidal stroma appear as vasculature due to the high contrast between A-scans adjacent to the shadowing. Figure 3.10 shows an OCT B-scan with its corresponding en face localiser SLO image with arrows on both scans showing the correspondence between superficial retinal vessels corrupting A-scans on the OCT B-scan through light attenuation.

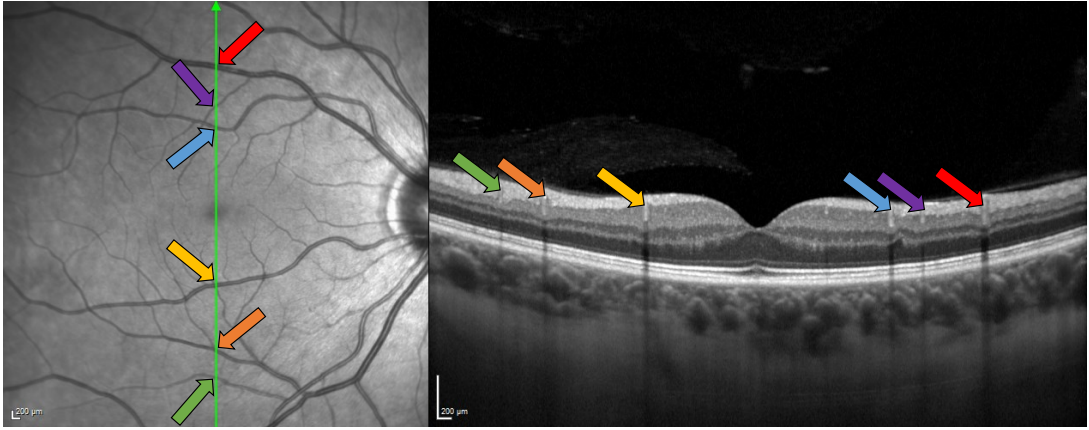
These shadows can often confuse segmentation algorithms into incorrectly believing stromal area is part of the choroid's vascular compartment. To prevent this over-segmentation, we apply a multiplicative column-wise compensation factor to automatically brighten shadow-corrupted A-scans [105] as a pre-processing step before the vessel segmentation procedure.

Column-wise compensation factors (which are multiplied to each A-scan (column) of the raw OCT B-scan) are computed as the element-wise ratio between each A-scans' averaged signal (axially), and the lateral moving average over these averaged A-scans. Note that the lateral moving average of A-scans smooths out any sudden dips in brightness from the shadow-corrupted A-scans. For an image array  $I$  with  $M$  rows indexed by  $y$  and  $N$  columns indexed as  $x$ , the pixel indexed by  $(x, y)$  in the shadow compensated image  $I_s$  is

$$I_s(x, y) = I(x, y) \cdot \frac{\mu_I(x)}{\text{MA}(\mu_I)(x)}, \quad (3.2)$$

where  $\mu_I(x)$  is the average pixel intensity of the A-scan indexed at column  $x$





**Figure 3.10:** Demonstration of superficial retinal vessels on the localiser SLO causing shadows on the OCT B-scan, corrupting the relevant A-scans of the choroid.

such that

$$\mu_I(x) = 1/M \sum_j^M I(x, j). \quad (3.3)$$

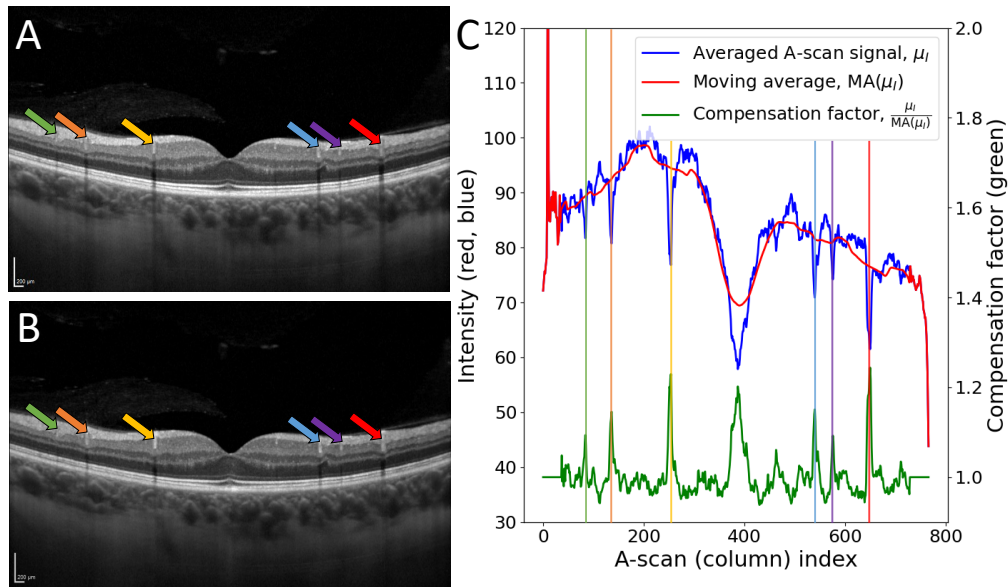
$\mu_I$  is thus a list of A-scan averages, indexed at all columns  $x = 1, \dots, N$  and  $MA(\mu_I)$  is the moving average of  $\mu_I$ , with  $MA(\mu_I)(x)$  the value of this moving average at column  $x$ .

Figure 3.11 shows the raw OCT B-scan from figure 3.10 in panel (A), with the shadow compensated result in panel (B) and the plots for the signals  $\mu_I$ ,  $MA(\mu_I)$  and the compensation factor signal  $\mu_I/MA(\mu_I)$  shown in blue, red and green, respectively. Note that the number of columns in this B-scan was 768. The arrows on the B-scans correspond to the shaded vertical lines overlaid onto panel (C). Figure 3.12 shows the impact of vessel shadow compensation on the resulting choroid vessel segmentation using the methodology introduced in this chapter.

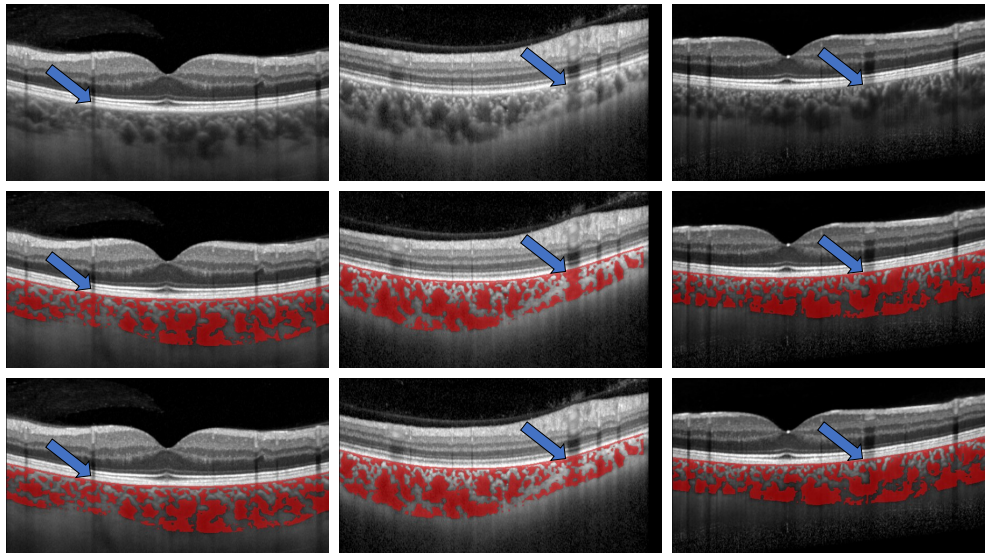
## 3.2 MMCQ's methodology

For choroid vessel segmentation, our task is to identify which pixels within the choroid belong to vasculature — typically seen as dark blobs [28, 103, 104] —





**Figure 3.11:** Demonstration of superficial retinal vessel shadow compensation. (A) Raw OCT B-scan. (B) Shadow compensated B-scan. (C) Graphs of  $\mu_I$ ,  $MA(\mu_I)$  and the compensation factor. Coloured, shaded vertical lines correspond to shadow locations (coloured arrows) in panels (A) and (B).



**Figure 3.12:** Choroid vessel segmentation before and after retinal vessel shadow compensation. (Top) raw OCT B-scans. (Middle) choroid vessel segmentation before shadow compensation. (Bottom) choroid vessel segmentation after shadow compensation. Arrows indicate locations of improved vessel segmentation.

and which do not. However, the problem is made difficult because of the poor visualisation of the vasculature within the choroid, particularly for small-scale vasculature, i.e. the choriocapillaris located at the anterior point of the choroid, and the smaller vessels which supply it. Larger vessels have greater contrast because there is more vessel lumen visible on the OCT device — contributing to a stronger and more consistent intensity transition between interstitial and intravascular compartments — while smaller vessels do not.

In this section we describe a methodology for semi-automatic choroid vessel segmentation known as multi-scale median cut quantisation (MMCQ). MMCQ takes a multi-scale approach to choroid vessel segmentation by enhancing choroidal vessels at different scales, thus addressing their heterogeneity in both size and shape. We also combine the effects of quantisation with histogram equalisation at the local level to address the poor contrast often exhibited by the choroidal vasculature.

### 3.2.1 Pre-processing

Before applying MMCQ, we apply some pre-processing to do some initial denoising and contrast enhancement. These steps are outlined below:

1. Shadow compensation: we compensate for A-scans which have been corrupted by superficial retinal vessel shadowing using the compensation scheme described in section 3.1.4 [105].
2. Cropping: we crop the OCT B-scan using a previously calculated segmentation mask or edge-based segmentation of the choroidal space. This not only decreases execution time, but also means that superfluous regions outside the choroid do not influence the segmentation procedure.
3. Denoising: we reduce the speckle noise by using a  $3 \times 3$  median filter, which applies the same procedure as described in section 2.2.1. The window size is smaller here so as to ensure smaller vessels are not overly smoothed.
4. Contrast enhancement: we apply the CLAHE algorithm by breaking the image into a set of  $8 \times 8$  tiles with a clipping limit of 2. The clip limit is

lower here than in section 2.2.1 so as to not over-enhance the smaller vessels, which would harm the enhancement procedure.

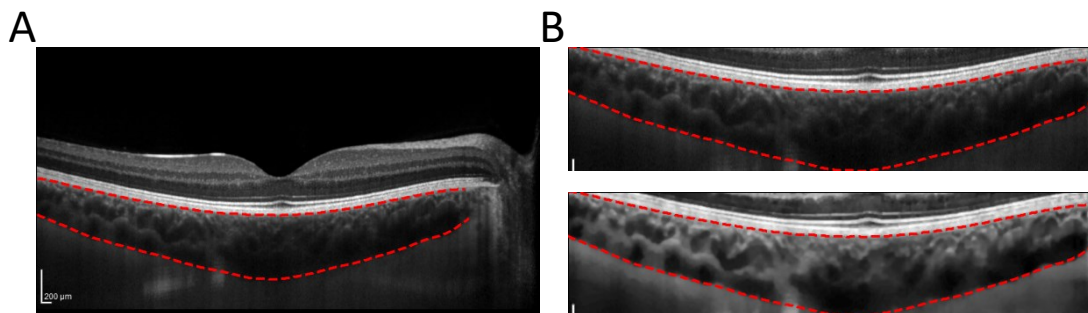
Figure 3.13 shows a horizontal-line EDI-OCT scan of the choroid with its region demarcated in panel (A), with the choroid region cropped out in the top row of panel (B) and the result of pre-processing in the bottom row of panel (B). In this example, the larger vessels have seen improved definition but less so for smaller vessels.

## 3.2.2 Patch quantisation and enhancement

### 3.2.2.1 Patch size

The multi-scale procedure applies quantisation and histogram equalisation at a local, patch-based level. We select three distinct scales to extract patches from the choroid, based on the (image-aligned) pixel thickness of the choroid in the OCT B-scan. The pixel thickness for an A-scan is defined here as the number of (row) pixels between the RPE-Choroid and Choroid-Sclera boundary for that particular lateral position (column). The pixel thickness  $T$  of the choroid is the average pixel thickness across all A-scans which have the choroid segmented.

Given some pixel thickness  $T$ , we apply the quantisation and enhancement scheme by decomposing the cropped choroid into square patches of size



**Figure 3.13:** Initial pre-processing of an OCT B-scan before MMCQ's pipeline. (A) OCT B-scan with choroid region segmented using red-dashed lines. (B) the OCT B-scan cropped to only show the choroid (top) and the result after pre-processing (bottom).

$\{T/10, T/5, T/2\}$ . The average choroid thickness in a healthy population is approximately 300 microns [233]. Thus, given that pixel length-scales for commercial OCT devices (Heidelberg Engineering and Topcon) range axially between approximately 3 and 4 microns-per-pixel, this corresponds to patch size scales for the average choroid of approximately  $\{10 \times 10, 20 \times 20, 50 \times 50\}$ . These approximate scales assume coverage of the range of different sizes of choroidal vessels found in Sattler's and Haller's layer and, importantly, scales with the size of the choroid. For the remainder of this section, we will consider a patch size  $s$ , i.e. the cropped choroid will be divided into mutually exclusive, non-overlapping patches of size  $s \times s$ .

### 3.2.2.2 Quantisation

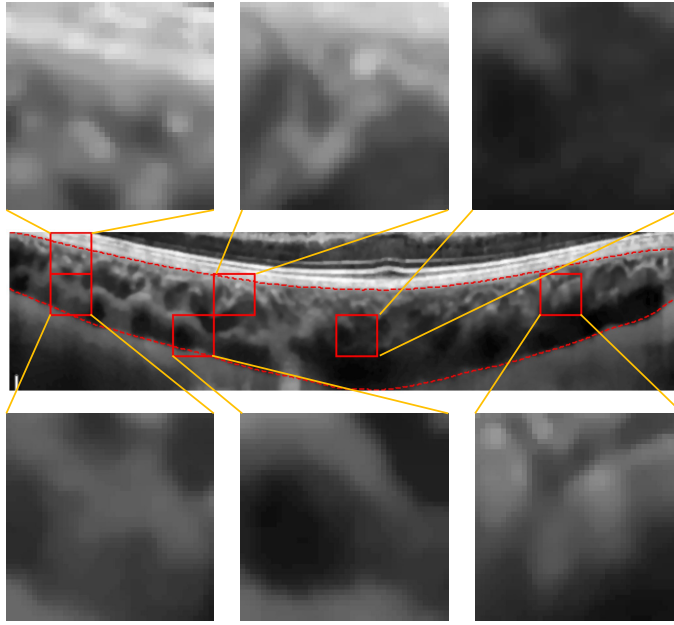
Figure 3.14 shows several patches extracted from the pre-processed choroid in figure 3.13(B) (bottom) which we will use as running examples for the rest of this section.

Each patch  $P$  is quantised using median cut quantisation,  $P_q$ , replacing the intensity value of every pixel assigned to the same cluster group with their median value. The number of clusters,  $K_P$  is patch-dependent and is based on the intensity dispersion of the patch histogram across the range of 256 brightness levels available for an 8-bit grayscale image. For a perfectly imaged choroid without any speckle noise, we would be able to set  $K = 2$  for all patches, i.e. one cluster for vessel pixels and another for stromal pixels. In reality, we must allow for poor contrast, low resolution and speckle noise to create a heterogeneous visualisation of the choroid in terms of pixel intensity. Thus, we allow  $K_P$  to vary between 2 and 5.

We first split the 256 brightness levels into a set of 5 equally spaced intervals,  $S$  defined by

$$S = \left\{ I_i = \left[ (i-1) \cdot 255/5, i \cdot 255/5 \right) \mid i = 1, \dots, 5 \right\} \quad (3.4)$$

$$= \{[0, 51), [51, 102), [102, 153), [153, 204), [204, 255]\}, \quad (3.5)$$



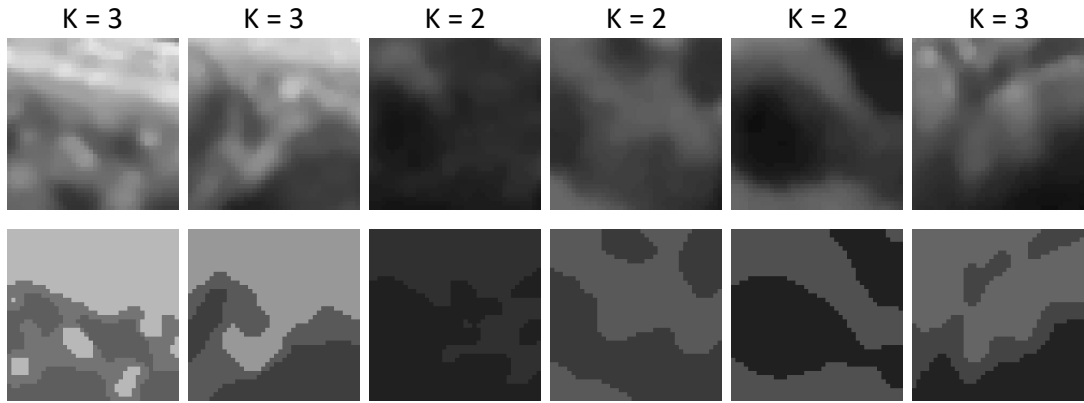
**Figure 3.14:** Six patches extracted from the pre-processed choroid of size  $40 \times 40$ .

where  $[x_1, x_2)$  is the half-open interval of integers (brightness levels) from integers  $x_1$  up to but excluding  $x_2$  (with exception of the last interval which is closed, i.e.  $[204, 255]$ ).  $K_P$  is then the number of intervals in  $S$  which the intensity range of patch  $P$  overlaps with, i.e.

$$K_P = \left| \left\{ I_i \mid I_i \in S, \exists j \in I_i, \hat{H}_P(j) > 0 \right\} \right|, \quad (3.6)$$

where  $\hat{H}_P$  represents a modified histogram of patch  $P$  such that  $\hat{H}_P(j)$  is the number of pixels in  $P$  which have intensity (brightness level)  $j$ , assuming that  $j$  exists within the 99<sup>th</sup> percentile range of patch  $P$ . This is to prevent outliers in the patch from influencing the estimate for  $K_P$ . The modified histogram  $\hat{H}_P$  is related to the raw histogram  $H_P$  such that for brightness levels  $j = 0, \dots, 255$

$$\hat{H}_P(j) = \begin{cases} 0 & j < \text{perc}(P, 0.005) \\ H_P(j) & \text{perc}(P, 0.005) \leq j \leq \text{perc}(P, 0.995) \\ 0 & j > \text{perc}(P, 0.995). \end{cases} \quad (3.7)$$



**Figure 3.15:** Six quantised patches using median cut clustering and replacing pixel intensities with median value for each cluster group.

Figure 3.15 show the six patches extracted from figure 3.14 in the top row, their quantised equivalents in the bottom row, with the number of quantisation levels estimated above as  $K$ .

### 3.2.2.3 Enhancement

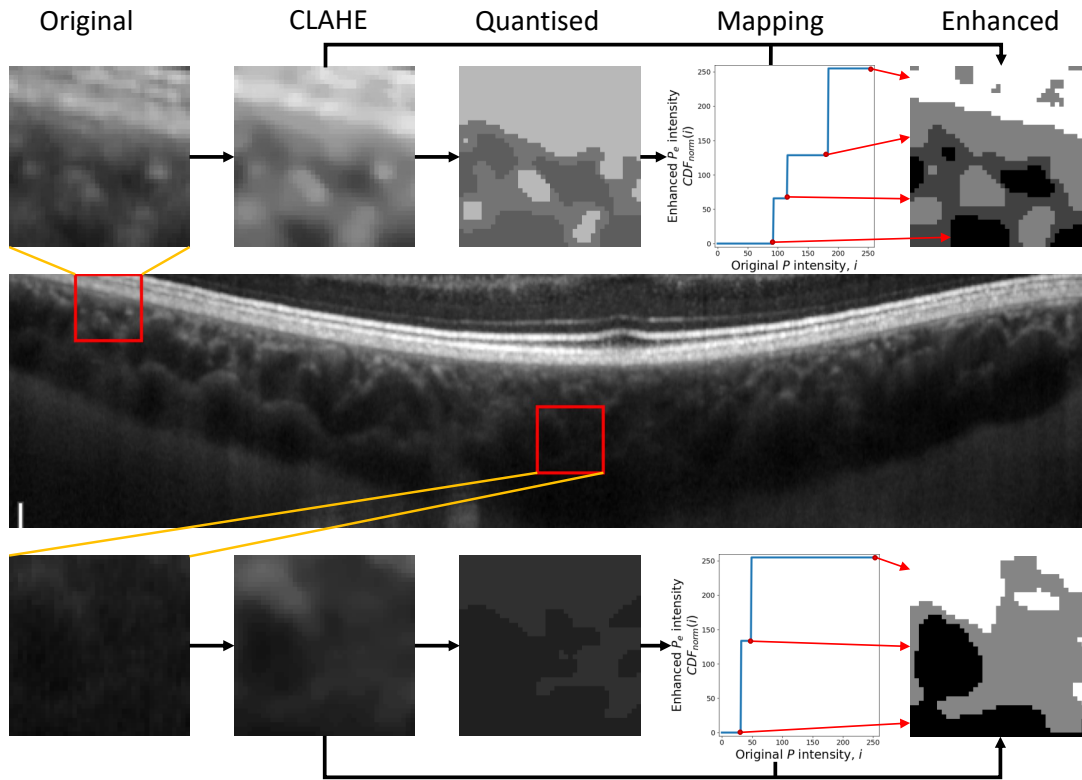
The enhancement procedure performs histogram equalisation on each quantised patch. We first construct the cumulative distribution function,  $CDF$  of the quantised patch  $P_q$ . This function is defined at brightness level  $i$  as

$$CDF(i) = \sum_{j=0}^i H_q(j), \quad (3.8)$$

where  $H_q(j)$  is the histogram of quantised patch  $P_q$  at brightness level  $j$ . This is normalised such that

$$CDF_{\text{norm}}(i) = \frac{CDF(i) - CDF(0)}{s^2 - CDF(0)} \cdot (L - 1). \quad (3.9)$$

Here, the  $CDF$  is normalised between  $[0, 1]$  and then scaled to the bit depth  $L$ , i.e. 8-bit depth ( $L = 256$ ). This provides a brightness level mapping for the original patch's pixels, spreading the distribution of brightness levels across the full range of available levels, i.e. between 0 and 255. Thus, the enhanced



**Figure 3.16:** The quantisation and enhancement procedure for two  $40 \times 40$  patches from the choroid in the raw OCT B-scan. Red arrows represent the pixel intensity on the new patch which the  $CDF_{\text{norm}}$  maps to from the original pixel intensities of patch  $P$ .

patch  $P_e$  is computed by mapping the pixel intensities from the original patch  $P$  using  $CDF_{\text{norm}}$  such that

$$P_e(x, y) = CDF_{\text{norm}}(P(x, y)), \quad (3.10)$$

where  $x, y$  are the column and row index locations of the patch. Since the quantisation is preserved by  $CDF_{\text{norm}}$  having been generated using the quantised patch  $P_q$ ,  $P_e$  is thus both quantised and enhanced. Figure 3.16 shows two patches from the original choroid and the evolution of their pre-processing, quantisation and enhancement.

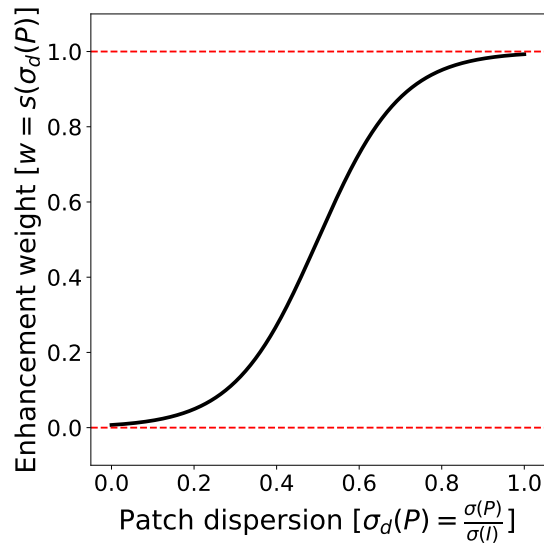
This enhancement is suitable for differentiating vessel from stroma for heterogeneous patches like the one in the top row of figure 3.16. However, the same



enhancement for relatively homogeneous patches, i.e. those which are primarily vessel or stroma, has the effect of over-enhancing small pixel intensity variations which introduces unwanted noise into the enhancement procedure. This is evident in the patch shown in the bottom row of figure 3.16. Thus, homogeneous patches should not be enhanced as much as heterogeneous patches. In CLAHE, preventing homogeneous areas from being over-enhanced is addressed through the *clip limit*. However, clipping the quantised patch's histogram will redistribute any excess pixels uniformly across the enhanced histogram, thus rendering the quantisation step meaningless. Moreover, selecting the optimal clipping limit is challenging for multiple patches at different sizes across different choroids.

Therefore, we apply a simple weighting scheme to control the strength of enhancement. We do this by measuring the proportion of the patch  $P$ 's intensity dispersion relative to the intensity dispersion of the cropped image (only showing the choroid),  $I$ . We characterise dispersion as the standard deviation over the region, so that patch  $P$ 's *relative* dispersion,  $\sigma_d(P)$  is

$$\sigma_d(P) = \frac{\sigma(P)}{\sigma(I)}, \quad (3.11)$$



**Figure 3.17:** Sigmoid function evolving over input range  $[0, 1]$  following equation (3.12).



where  $\sigma(\cdot)$  is just the standard deviation of the flattened array  $P$  or  $I$ .  $\sigma_d(P)$  is bound between 0 and 1, and to promote minimal enhancement for low values of  $\sigma_d(P)$  and maximum enhancement for high value of  $\sigma_d(P)$  we use the Sigmoid function to dictate the weight of enhancement strength. The Sigmoid function is ideal for computing weight  $w$  due to the sharp plateaus at 0 and 1 at either end of the input range. The enhancement weight for patch  $P$ ,  $w(P)$  is defined using the Sigmoid formula

$$w(P) = s(\sigma_d(P)) = (1 + e^{-2\pi(\sigma_d(P)-1/2)})^{-1}, \quad (3.12)$$

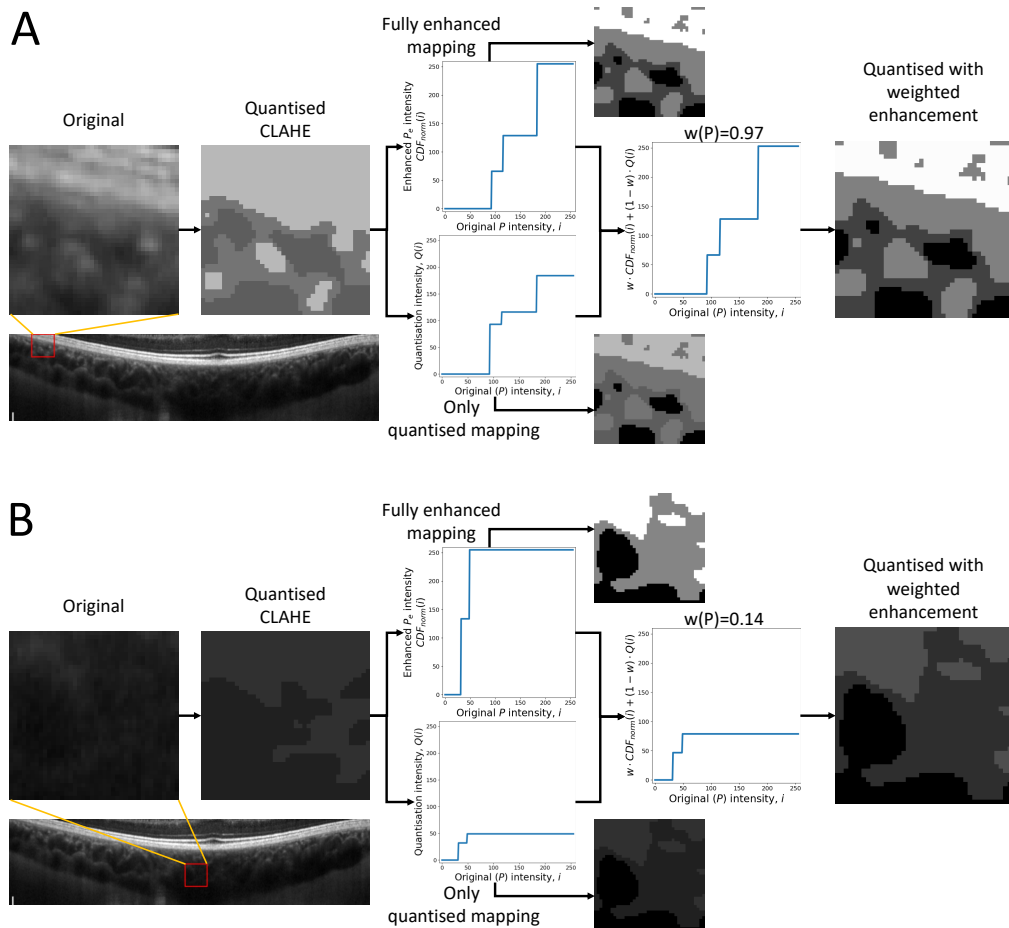
which yields an output range  $w(P)$  in the interval  $[0, 1]$ , given an input interval  $[0, 1]$  of which  $\sigma_d(P)$  of bound by. The Sigmoid function is plotted in figure 3.17. The enhancement of patch  $P$  is then scaled by  $w(P)$ , with the remaining  $(1 - w(P))$  factor made up by an intensity mapping equivalent to the quantised patch  $P_q$ , called  $Q$ .  $Q$  converts the pixel intensities in patch  $P$  to the brightness levels dictated by the quantised patch  $P_q$ . The piece-wise formula for mapping  $Q$  for a patch quantised into  $n$  levels  $\{q_1, \dots, q_n\}$  is

$$Q(i) = \begin{cases} 0 & i < q_1 \\ q_1 & q_1 \leq i < q_2 \\ \vdots & \\ q_n & i \geq q_n, \end{cases} \quad (3.13)$$

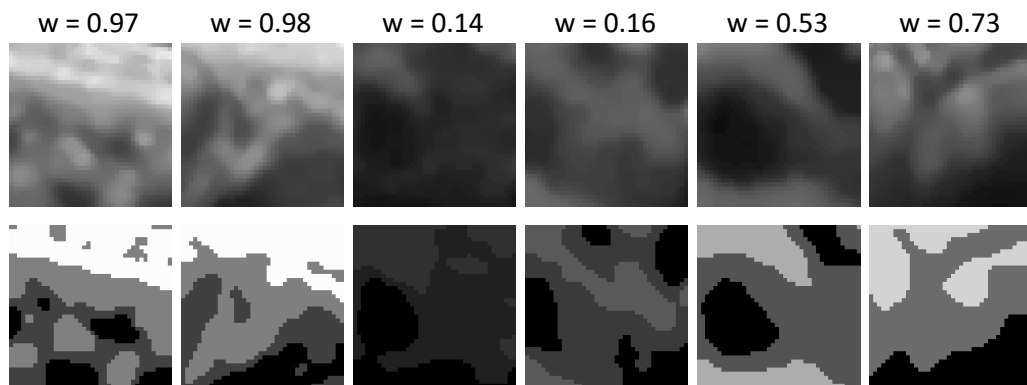
where  $i$  and  $q_j$  are both in the interval  $[0, 255]$  for all  $j = 1, \dots, n$ . This mapping is equivalent to zero enhancement and only quantisation. The enhanced patch  $P_e$  is thus a linear combination between  $Q$  and  $CDF_{\text{norm}}$  with  $w(P)$  as the multiplicative factor balancing the strength of enhancement, such that

$$P_e(x, y) = w(P) \cdot CDF_{\text{norm}}(P(x, y)) + (1 - w(P)) \cdot Q(P(x, y)). \quad (3.14)$$

The computation of  $P_e$  is shown for the two demonstrative patches from figure 3.16 in figure 3.18. In panel (A) we can see the heterogeneous patch is mostly fully enhanced as the weight is 0.97, while the homogeneous patch in panel



**Figure 3.18:** The weighted, quantisation and enhancement procedure for two  $40 \times 40$  patches from the choroid in the raw OCT B-scan.



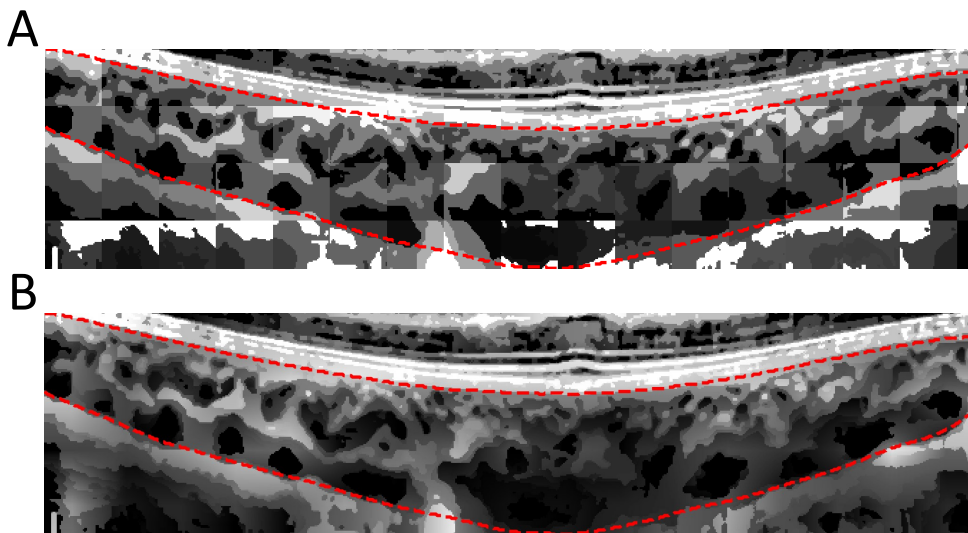
**Figure 3.19:** Enhanced patches extracted from the raw OCT B-scan for a single scale ( $40 \times 40$ ).

(B) is mostly quantised as the weighting is only 0.14. For completeness, figure 3.19 shows all six patches extracted from the raw OCT B-scan in the top row and their weighted, quantised and enhanced patches in the bottom row for a single scale ( $40 \times 40$ ).

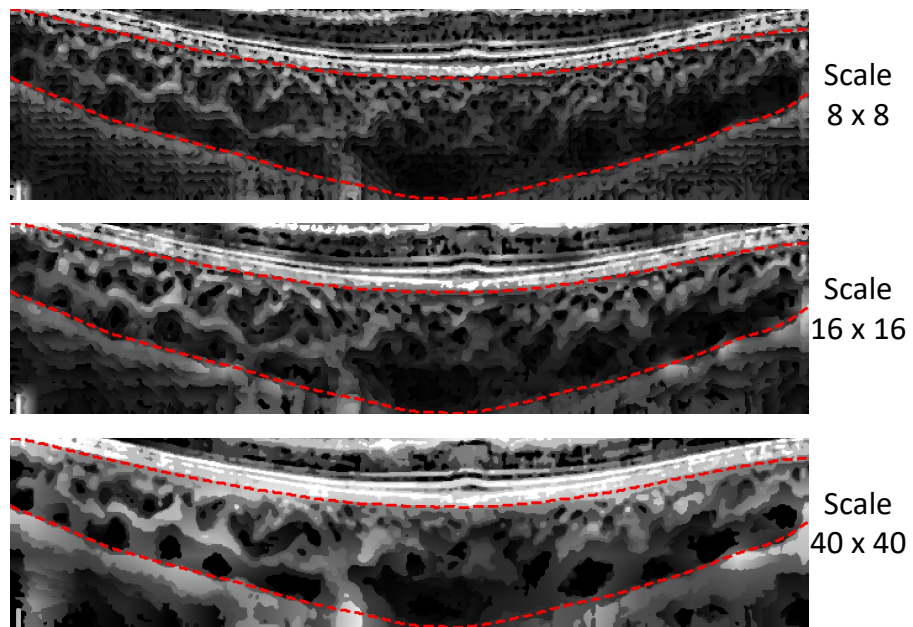
### 3.2.3 Merging

#### 3.2.3.1 Patch merging

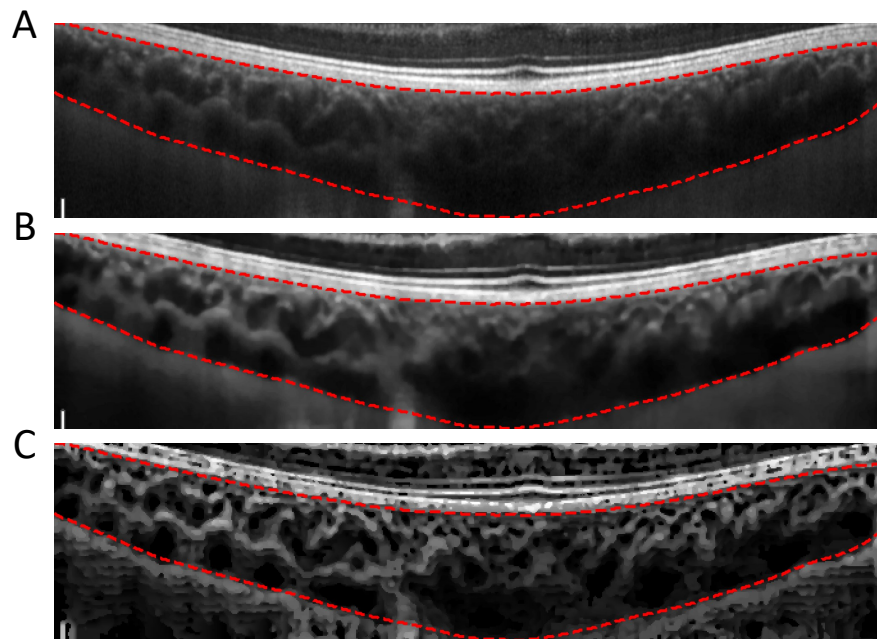
After all patches have been enhanced according to the procedure in the previous section, adjacent patches are blended together in both axial and lateral directions using bi-linear interpolation. Figure 3.20 shows all of the non-overlapping patches at scale  $40 \times 40$  stacked together in panel (A) with the final result after bi-linear interpolation in panel (B). Note that in panel (A) there are discontinuities at the patch edges.



**Figure 3.20:** (A) Enhanced patches stacked in order to recreate a discontinuously enhanced choroid. (B) Final enhanced choroid at scale  $40 \times 40$  after bi-linear interpolation.



**Figure 3.21:** Enhanced choroid at scales  $8 \times 8$ ,  $16 \times 16$  and  $40 \times 40$  for a choroid whose pixel thickness was 80 pixels.



**Figure 3.22:** (A) Original choroid from raw OCT B-scan. (B) Pre-processed choroid after pixel smoothing and simple CLAHE application. (C) Enhanced choroid using MMCQ.

### 3.2.3.2 Scale merging

After enhancement at all scales has been performed we are left with several quantised and enhanced choroids. At each scale, a particular scale of vasculature is prioritised in the enhancement and quantisation procedure. Figure 3.21 shows the enhanced choroids for each of the three scales,  $8 \times 8$ ,  $16 \times 16$  and  $40 \times 40$ . These scales were dictated by the pixel thickness of this particular choroid which was  $T = 80$  pixels, i.e. patch sizes were chosen as  $\{T/10, T/5, T/2\} = \{80/10, 80/5, 80/2\} = \{8, 16, 40\}$ .

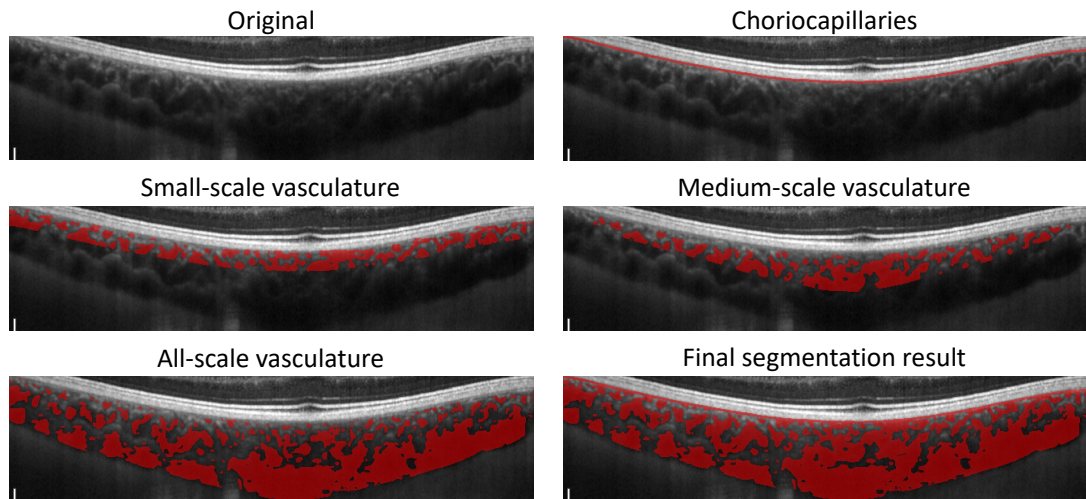
We synthesise the enhanced choroids by taking the element-wise minimum across the scale dimension. This ensures that we prioritise the darkest pixels systematically across all enhancements. Figure 3.22 shows the raw choroid, the pre-processed choroid and the MMCQ-enhanced choroid in series.

### 3.2.4 Vessel segmentation

We use a multi-depth approach to vessel segmentation on the MMCQ-enhanced choroid using the median cut algorithm. To ensure that small- and large-scale vessels are clustered together, we repeatedly apply the median cut algorithm at different depths of the choroid, starting from the choriocapillaris, then the small-scale choroidal vessels sitting posterior to these, then the smaller and medium-sized vessels and finally the whole choroid.

The choriocapillaris layer sits anterior-ward in the choroid, and is approximately 10 microns in thickness at the posterior pole [7]. However, the individual capillaries in this layer are too small to resolve given current OCT resolution limitations. Given how densely packed the choriocapillaries are [8], we assume in our vessel segmentation procedure that the first 10 microns of depth is vascular, and since the axial and lateral pixel length scales are known, this can be approximated well enough. For example, for Heidelberg Engineering OCT1 devices, the axial length-scale is 3.87 microns, and so we assign the first 3 rows posterior and parallel to the RPE-Choroid boundary as vessel.

For the remaining depths of the choroid, we apply median cut clustering to the grayscale values of only the choroid pixels (those between the segmented



**Figure 3.23:** The choroid segmented at different depths of the choroid. The segmentations are combined by taking their union (bottom right).

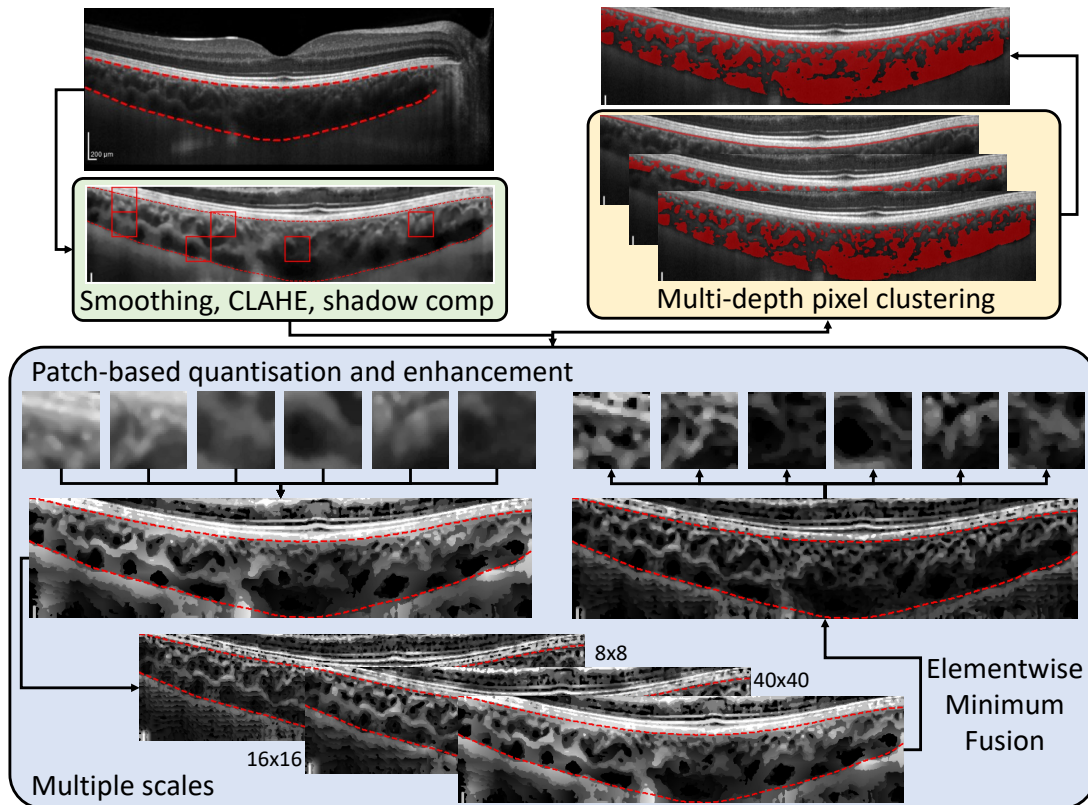
boundaries). Fortunately, due to the nature of the median cut and MMCQ algorithms, clusters can be ordered via their average intensity to work out which are vascular and which are not. The only set of parameters to define are the number of clusters  $K$  to generate and how many to assign as choroidal vasculature,  $k$ . While these parameters can be left free to the end-user, we have empirically found  $(K, k) = (20, 11)$  to be sensible given earlier parameter choices.

We use the average pixel thickness  $T$  to determine the different depth-levels, segmenting small- and medium-scale vessels from the choriocapillaris to  $T/3$  and  $3T/5$  posterior to the RPE-Choroid boundary, respectively. Finally, the whole choroid is segmented for large-scale vessels. Figure 3.23 shows the choroid used throughout this section, with segmentations for each of the depth-levels. In this case, the setting  $(K, k) = (20, 11)$  was used. The final segmentation result (bottom-right panel) is the element-wise union of all segmentations.

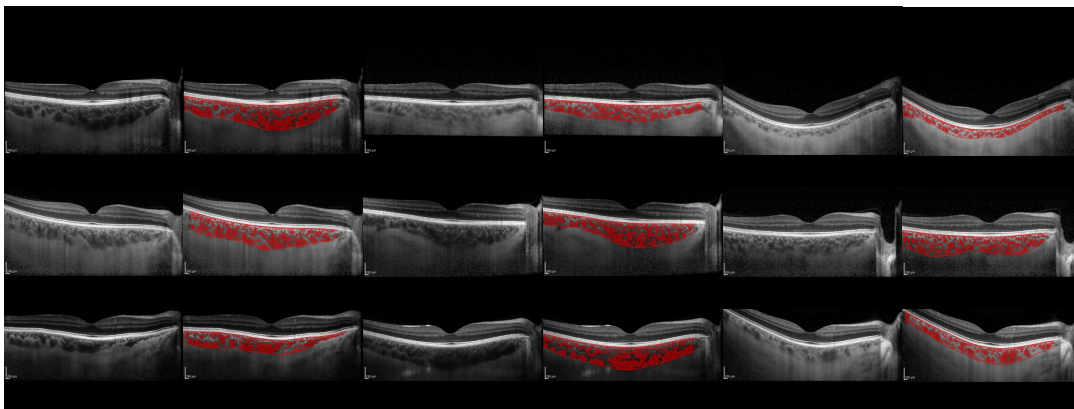
### 3.2.5 Pipeline

Figure 3.24 presents a schematic of MMCQ's pipeline with figure 3.25 showing the output segmentations on a selection of choroids using  $(K, k) = (20, 11)$ . Additionally, algorithm 2 outlines the pseudocode for the MMCQ algorithm.





**Figure 3.24:** Schematic diagram of MMCQ segmenting the choroidal vessels in the EDI-OCT B-scan used throughout this section.



**Figure 3.25:** Vessel segmentations using MMCQ for a variety of choroids large and small with varying contrast, vessel distribution and speckle noise.

**Algorithm 2** Multi-scale Median Cut Quantisation

**Inputs:** 8-bit OCT B-scan  $I$ , choroid region segmentation boundaries, pixel length-scales and cluster configuration  $(K, k)$ .

**Output:** Segmented image  $S$  of choroid vessels.

- 1: Crop image  $I$  to only choroid  $I_c$ , and pre-process  $I_c$  using vessel shadow compensation, a simple  $3 \times 3$  median filter and CLAHE application with a tile size of 8 and clip limit of 2;
- 2: Measure average choroid thickness  $T$  (in pixels) between upper and lower choroid boundaries;
- 3: Three square patch sizes defined according to  $\{T/10, T/5, T/2\}$ ;
- 4: **for** square patch size  $p_i \in \{T/10, T/5, T/2\}$  **do**
- 5:     Split choroid image  $I_c$  into non-overlapping patches of size  $p_i \times p_i$ ;
- 6:     **for** patch  $P$  **do**
- 7:         Number of quantisation levels  $K_P$  (clusters) via equation (3.6);
- 8:         Quantise patch  $P$  into  $K_P$  levels using median cut quantisation, called  $P_q$ , replacing cluster group pixel intensities with their collective median value;
- 9:         Compute the normalised cumulative distribution  $CDF_{\text{norm}}$  via equation (3.9);
- 10:         Enhancement weight  $w = w(P) = s(\sigma_d(P_q))$  given by equation (3.12) and  $\sigma_d(P_q)$  given by equation (3.11);
- 11:         Generate intensity mapping  $Q$  of the quantised patch  $P_q$  representing only quantisation and zero enhancement via equation (3.13);
- 12:         Compute enhanced patch  $P_e$  via equation (3.14);
- 13:     **end for**
- 14:     Enhanced patches are combined and smoothed using bi-linear interpolation to create the enhanced choroid at patch size  $p_i, I_{p_i}$ ;
- 15: **end for**
- 16: Final, enhanced choroid  $E(x, y) = \min\{I_{p_i}(x, y) : i = 0, 1, 2\}$ , i.e. fuse enhancements by taking element-wise minimum value across all  $I_{p_i}$  across all scale dimensions;
- 17: Thin,  $10\mu\text{m}$  layer sub-RPE-Choroid boundary segmented as choriocapillaris  $S_c$  using axial pixel length-scale;
- 18: **for** depth-level  $d \in \{T/3, 3T/5\}$  **do**
- 19:     Duplicate RPE-Choroid boundary and shift  $d$ -pixels vertically downward, cropping out all choroid pixels outside of region in between these two boundaries;
- 20:     Quantise choroid pixels within region of enhanced choroid  $E$ , using median cut quantisation into  $K$  levels, keeping  $k$  as vessel pixels,  $S_d$ ;
- 21: **end for**
- 22: Quantise whole enhanced choroid  $E$  using median cut quantisation into  $K$  levels, keeping  $k$  as vessel pixels  $S_w$ ;
- 23: Segmented choroid  $S = S_c + \sum_d S_d + S_w$ ;
- 24: **return**  $S$  after clipping values such that  $S$  only takes on values 0 and 1 to represent interstitial and vessel pixels.



### 3.3 MMCQ’s evaluation

We evaluate MMCQ by comparing its segmentation performance against two manual graders, and Agrawal’s modified Niblack method [28], as this is the most widely adopted approach for choroid vessel segmentation in the research community.

#### 3.3.1 Study population

We used a subsample of the dataset used to evaluate GPET’s choroid region segmentation described in section 2.3. Briefly, the cohort [225] was collected to investigate chorioretinal changes in response to kidney transplantation in healthy donors and recipients with end-stage CKD.

For this evaluation, as we required manual segmentation of the choroidal vessels, we selected a random subsample of 20 B-scans at the patient-level from 10 renal transplant recipients and 10 donors. As the temporal nature of this cohort was longitudinal, we only selected images at baseline for this evaluation to remain objective in our analysis. Thus, there is equal proportion of cases and controls in our evaluation (10 cases of end-stage CKD about to undergo kidney transplant and 10 controls about to undergo kidney donation).

The basic demographics and image characteristics for this subsample are

|                  | Cohort  |            | Total     |
|------------------|---------|------------|-----------|
|                  | Donors  | Recipients |           |
| Sample, N        | 10      | 10         | 20        |
| Right eye, N     | 10      | 10         | 20        |
| Age, years       | 49 ± 14 | 44 ± 12    | 47 ± 13   |
| Sex (male %)     | 5 (50)  | 7 (70)     | 12 (60)   |
| Pixel resolution |         |            | 768 × 768 |
| EDI mode         |         |            | Yes       |
| ART              |         |            | 100       |

**Table 3.1:** Population demographics and image characteristics of dataset used to evaluate MMCQ against manual grading and Agrawal’s Niblack method [158].

shown in table 3.1. Please see section 2.3.1.2 for a description of the image acquisition for this subsample.

### 3.3.2 Statistical analysis

Two manual graders segmented the choroidal vessels in each OCT B-scan, M1 and M2. M1 was a clinical ophthalmologist (Supervisor Dr. Ian J.C. McCormick) and M2 was the author of this thesis (Jamie Burke). Graders segmented the choroid region and vessels using ITK-Snap [229] and were blinded to each others segmentations, following the pre-specified measurement protocol outlined in appendix A.5 using the definitions and equipment outlined in appendices A.1 and A.2. We also applied MMCQ and Agrawal's modified Niblack method [28] to these 20 OCT B-scans for quantitative comparison with the manual annotations.

To keep the use of MMCQ and Niblack consistent, we fixed the free parameters of both algorithms. For MMCQ, we used the segmentation cluster configuration  $(K, k) = (20, 11)$  (see section 3.2.4) and for Niblack, we fixed the window size  $w$  to 51 and the standard deviation offset  $k$  to -0.05 — values proposed by Muller, et al. [163] as the mean values of  $w$  and  $k$  when the algorithm was used manually to create a ground truth label dataset for a separate deep learning model.

We used the Dice coefficient to measure overall segmentation agreement between the automatic methods (MMCQ/Niblack) and the manual graders (M1/M2), and also used Precision and Recall to quantify the positive predictive value and sensitivity of the two approaches for choroid vessel segmentation. These metrics will quantify miss-classification of interstitial and vessel pixels, respectively. Segmentation metrics were measured only for the region in the image which constituted the choroidal space, dictated by Choroidalizer, a method described later in chapter 5 for segmenting the choroidal space, vessel and detecting the fovea on OCT B-scans. The segmentation metrics used here have been previously defined in section 1.7.4.

We also compare MMCQ and Niblack to the manual grader on downstream

choroidal measurements including vessel area and CVI measured using a 6 mm, fovea-centred ROI. The fovea was detected also through Choridalyzer to remain objective in our analyses. These metrics have previously been defined in section 1.7. To compare derived metrics, we use mean absolute error (MAE) as well as Pearson and Spearman correlation coefficients previously defined in section 1.7.4.

Finally, we also compared the execution time of each approach for region and vessel segmentation, and report mean and standard deviation in seconds. MMCQ and Niblack were run on a laptop with a 4 year old Intel Core i5 (8<sup>th</sup> generation) CPU and 16 Gb of RAM. Manual graders timed themselves during manual vessel segmentation using ITK-Snap.

### 3.3.3 Results

Table 3.2 shows the main results of comparing all four methods with each other in a pairwise fashion. Overall, MMCQ had better agreement with both manual graders, both from the segmentation and choroid-derived evaluation metrics. For segmentation metrics, MMCQ had a Dice score with graders M1 and M2 of 0.7148 and 0.7350, respectively, while Niblack had 0.6838 and 0.7065, respectively. Interestingly, MMCQ achieved higher precision than Niblack (0.6921 vs. 0.6238 for comparisons with rater M1) with similarly comparative values for rater M2. Conversely, Niblack had higher recall (0.7486 vs. 0.7644 for compar-

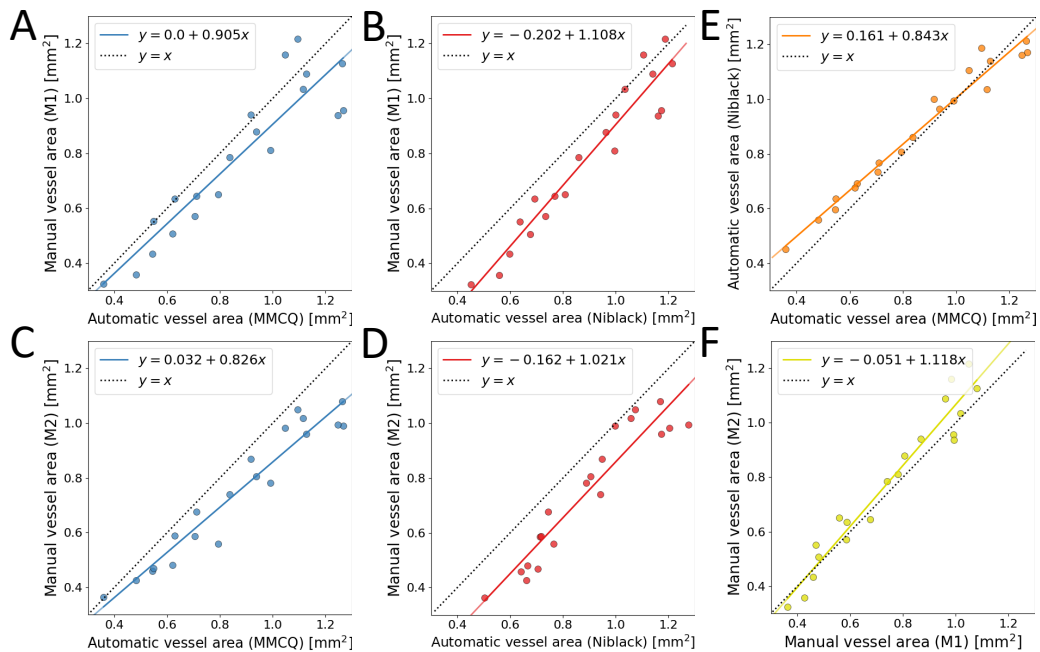
| Comparison       | Segmentation  |               |               | Vessel area     |                 |                        | Vascular Index      |                     |               |
|------------------|---------------|---------------|---------------|-----------------|-----------------|------------------------|---------------------|---------------------|---------------|
|                  | Dice          | Precision     | Recall        | Pearson         | Spearman        | MAE (mm <sup>2</sup> ) | Pearson             | Spearman            | MAE           |
| MMCQ vs. M1      | <b>0.7148</b> | <b>0.6921</b> | 0.7486        | <b>0.9622**</b> | <b>0.9564**</b> | <b>0.1080</b>          | 0.5503*             | <b>0.5242*</b>      | <b>0.0759</b> |
| Niblack vs. M1   | 0.6838        | 0.6238        | <b>0.7644</b> | 0.9187**        | 0.9113**        | 0.1153                 | <b>0.5958*</b>      | 0.5092*             | 0.1901        |
| MMCQ vs. M2      | <b>0.7350</b> | <b>0.6890</b> | 0.7911        | <b>0.9873**</b> | <b>0.9700**</b> | <b>0.1182</b>          | 0.5590*             | 0.5323*             | <b>0.0816</b> |
| Niblack vs. M2   | 0.7065        | 0.6262        | <b>0.8136</b> | 0.9643**        | 0.9474**        | 0.1435                 | <b>0.6068*</b>      | <b>0.5353*</b>      | 0.1464        |
| M1 vs. M2        | 0.7699        | 0.7937        | 0.7479        | 0.9718**        | 0.9504**        | 0.0637                 | 0.8074**            | 0.6857**            | 0.0618        |
| MMCQ vs. Niblack | 0.7366        | 0.7866        | 0.6959        | 0.9851**        | 0.9744**        | 0.0568                 | 0.1637 <sup>†</sup> | 0.0903 <sup>†</sup> | 0.0461        |

**Table 3.2:** Comparison of all four approaches in a pairwise fashion. Bold values represents the best method (of MMCQ/Niblack) for each comparison with the manual grader. \*\*:  $P < 0.0001$ , \*:  $P < 0.05$ , <sup>†</sup>:  $P > 0.05$ .

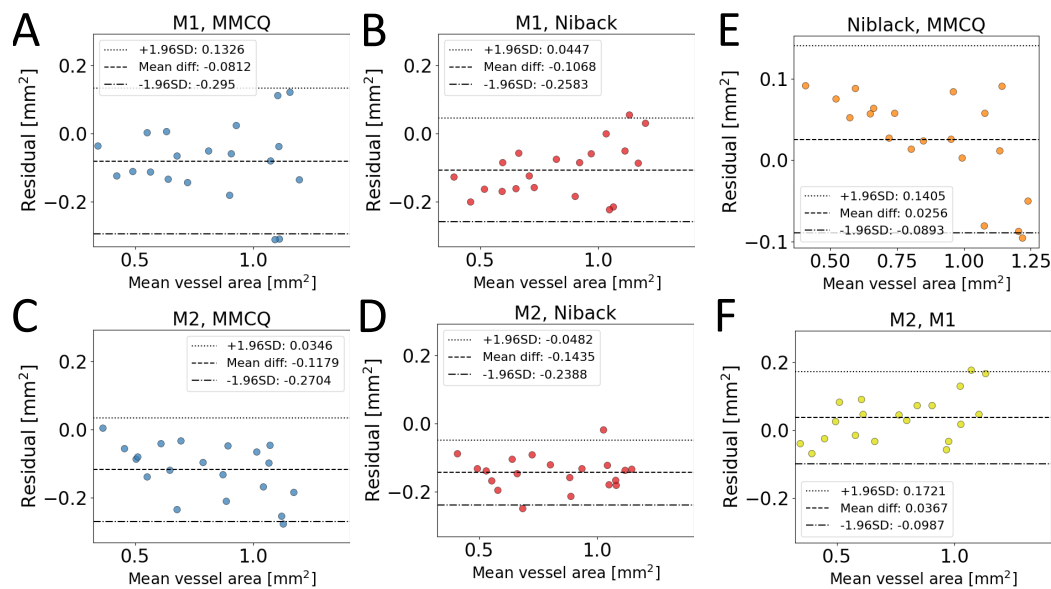
| Method  | Execution time (s) |
|---------|--------------------|
| M1      | 1506.000 ± 744.073 |
| M2      | 1176.700 ± 263.968 |
| Niblack | 0.370 ± 0.105      |
| MMCQ    | 2.148 ± 0.738      |

**Table 3.3:** Execution time (in seconds) for the manual graders and semi-automatic approaches for choroid vessel segmentation.

isons with rater M1) with similarly comparative values with rater M2. This would suggest that MMCQ-segmented choroids have higher positive predictive value, with fewer false positives than Niblack, at the cost of slight under-segmentation (more false negatives). Thus, according to the manual segmentations, Niblack over-segments the choroidal space, classifying more interstitial pixels as vessel compared with MMCQ. However, Niblack's cost of over-segmenting the choroidal space is somewhat balanced by its ability to have fewer false negatives, i.e. is able to correctly identify vessel pixels better than MMCQ.



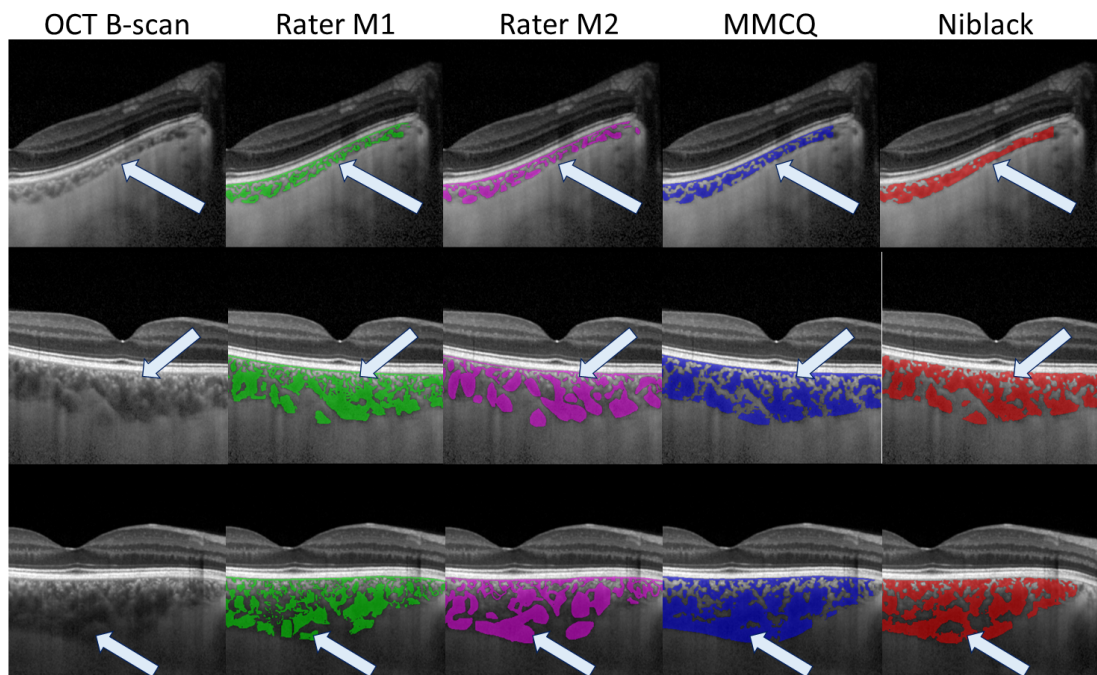
**Figure 3.26:** Scatter plots demonstrating correlation in choroid vessel area between different methods. (A) MMCQ vs. M1, (B) Niblack vs. M1, (C) MMCQ vs. M2, (D) Niblack vs. M2, (E) MMCQ vs. Niblack, (F) M1 vs. M2.



**Figure 3.27:** Bland-Altman plots of agreement in choroid vessel area between different methods. (A) MMCQ vs. M1, (B) Niblack vs. M1, (C) MMCQ vs. M2, (D) Niblack vs. M2, (E) MMCQ vs. Niblack, (F) M1 vs. M2.

There were strong and statistically significant correlations between both semi-automatic methods with the manual graders for choroid vessel area. MMCQ-derived vessel area had higher correlations and lower MAE than Niblack (Pearson correlation between MMCQ and raters M1 and M2 was 0.9622 and 0.9873, respectively, while for Niblack this was 0.9113 and 0.9474, respectively). These relationships can be visualised through scatter plots and Bland-Altman plots of choroid vessel area for each of the comparisons made, as shown in figures 3.26 and 3.27.

Both figures help describe why there are such large MAEs when comparing the semi-automatic methods with the manual graders (0.1080 and 0.1182 for MMCQ vs. M1 and M2, respectively). In all cases where we compare semi-automatic to manual, the semi-automatic method segments significantly more of the choroid than the manual grader. Interestingly, MMCQ and Niblack have opposite effects when compared to manual segmentations, i.e. MMCQ is closest to the manual graders for smaller choroids, and vice versa for Niblack. This may suggest, on the assumption of manual grading reflecting accurate



**Figure 3.28:** Qualitative comparison of the vessel segmentations for semi-automatic and manual approaches for the smallest and largest choroid in the dataset. Light blue arrows indicate sources of disagreement.

segmentation, that MMCQ could be better suited for smaller choroids than Niblack, and vice versa for larger choroids.

Figure 3.28 shows three OCT B-scans of which have been magnified to show only half of the choroid, with segmentations from all four methods (rater M1, green; rater M2, pink; MMCQ, blue; Niblack, red) overlaid in series. The choroids were selected based on their total choroid area, representing the smallest (top), median (middle) and largest (bottom) choroids in the dataset. Blue arrows indicate sources of disagreement between the methods.

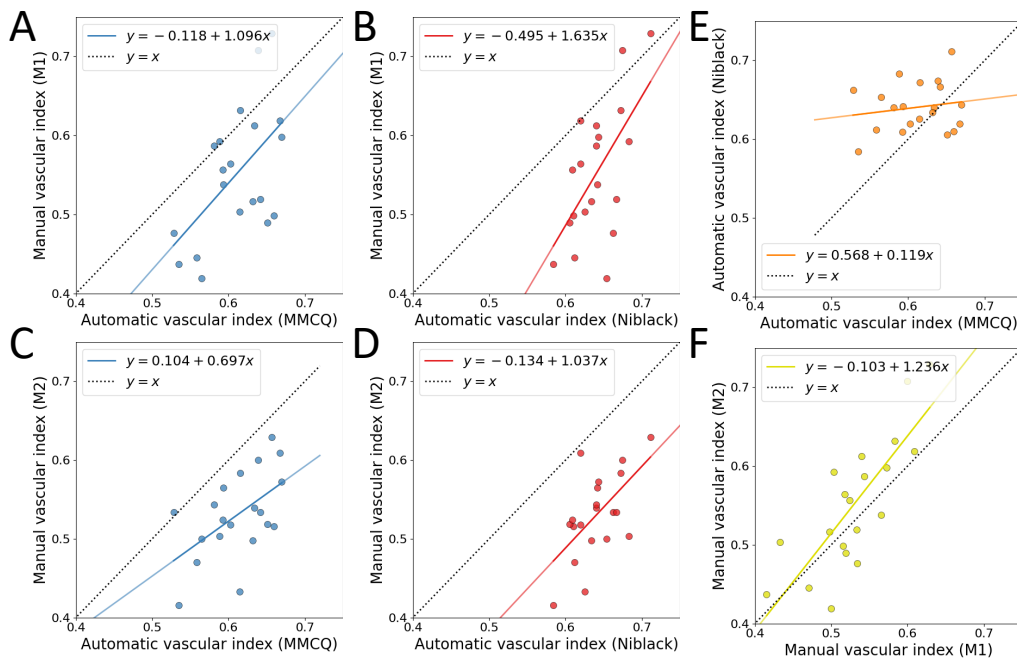
In the top row, we can see that Niblack over-segments the choroidal space significantly more than the other three methods, which are all able to identify the smaller vessels with precision. This translated into an overestimation of CVI of 0.66, compared with 0.47, 0.54 and 0.53 for raters M1, M2 and MMCQ, respectively. In the middle row, rater M1 segmented the choroidal vessels in more detail (CVI, 0.57) compared with rater M2 (CVI, 0.52) translating to a 0.05 discrepancy in CVI, while Niblack (CVI, 0.62) appeared to over-segment

the smaller vessels nearer the RPE-Choroid boundary, unlike MMCQ (CVI, 0.59). In the bottom row, MMCQ was subject to over-segmentation in this case because of the poor vessel boundary definition in the posterior segment of the choroidal space. This translated to MMCQ and Niblack both over-estimating the CVI as 0.63 and 0.61, respectively, with raters M1 and M2 estimating CVI as 0.50 and 0.51, respectively. Thus, we propose that MMCQ is more faithful to the manual raters for smaller choroids, and produces higher quality segmentations over the Niblack algorithm, on the assumption these manual graders represent accurate ground truths.

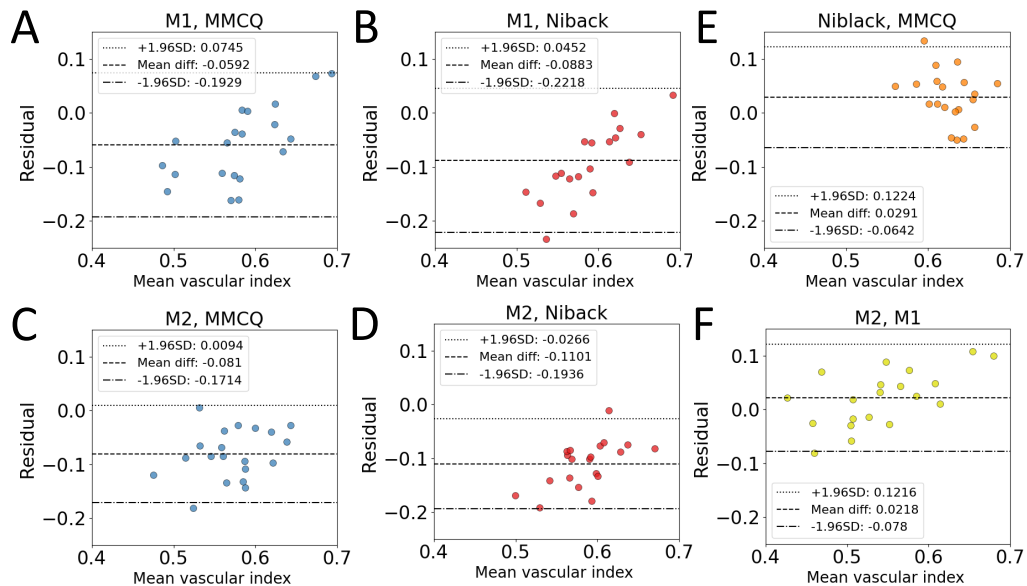
While the Bland-Altman plot in figure 3.27 demonstrate significant systematic bias between semi-automatic and manual methods, it is encouraging to observe from figure 3.26 strong linear and monotonic relationships between semi-automatic and manual methods in this small dataset. Based on this evaluation, MMCQ provides more accurate segmentations relative to Niblack, promoting vessel fidelity and positive predictive value over vessel pixel completeness at the cost of over-segmentation. However, the differences in these population-level performance metrics between MMCQ and Niblack are not large in absolute terms, and actually tend to average over specific, eye-level differences between the methods, such as Niblack's tendency to over-segment small-scale choroidal vasculature (figure 3.28).

For CVI, semi-automatic methods had poor agreement with the manual methods, both observed by Pearson/Spearman correlation (MMCQ vs M1, 0.5503/0.5242 and Niblack vs. M1, 0.5958/0.5092) which were weakly significant ( $P < 0.05$ ) and also exhibited large MAEs (MMCQ vs M1, 0.0759 and Niblack vs. M1, 0.1901). While MMCQ performed better than Niblack according to these manual labels, agreement was poor in absolute terms. The discrepancy between the correlations observed with choroid vessel area and CVI is an interesting contradiction. Figures 3.29 and 3.30 show scatter and Bland-Altman plots in a similar style to figure 3.26 but for CVI. Overall, there was poor agreement, with exception for the two manual graders, which reported reasonable agreement (Pearson/Spearman correlation, 0.7937/0.6857). However, the manual graders do not represent a gold standard approach, especially for larger choroids as shown in the figures 3.26(F) and 3.29(F).





**Figure 3.29:** Scatter plots demonstrating correlation in CVI between different methods. (A) MMCQ vs. M1, (B) Niblack vs. M1, (C) MMCQ vs. M2, (D) Niblack vs. M2, (E) MMCQ vs. Niblack, (F) M1 vs. M2.



**Figure 3.30:** Bland-Altman plots of agreement in CVI between different methods. (A) MMCQ vs. M1, (B) Niblack vs. M1, (C) MMCQ vs. M2, (D) Niblack vs. M2, (E) MMCQ vs. Niblack, (F) M1 vs. M2.



MMCQ and Niblack had very poor agreement for CVI (Pearson/Spearman correlation, 0.1637/0.0903, figure 3.29(E)). While at the population-level, the extent of disagreement in CVI appears random, the disagreement is actually ordered by the choroid vessel area (and thus the size of the choroid). The largest disagreement between semi-automatic methods in CVI (absolute difference of 0.13, 13%) corresponded to the choroid with the smallest vessel area, of which Niblack segmented a significant proportion more than MMCQ (figure 3.27(E)). This pattern was consistent such that smaller choroids corresponded to proportionately greater disagreement (relative to the size of choroid) in both CVI and vessel area between the two semi-automatic methods. However, unlike the linear relationship observed for choroid vessel area across the population (figure 3.26(E)), once an eye's choroid vessel area is normalised by its total choroid area (calculating CVI), this relationship disappears quite dramatically (figure 3.29(E)). This is likely because any error in vessel area in the numerator of CVI is further exacerbated by any error in the denominator, given the size of the choroid. Larger error in vessel area in smaller choroids introduces proportionately larger error in CVI. Thus, measurement error in vessel area is made significantly worse when translated into CVI for smaller choroids.

In terms of execution time, manual graders M1 and M2 took approximately 25 and 20 minutes to segment the choroidal vessels in an OCT B-scan, respectively, while Niblack and MMCQ took 0.37 and 2.14 seconds, respectively, with no human input. While Niblack has a faster execution time, by virtue of only working at a single scale and measuring summary statistics, both execution times are entirely reasonable for large-scale OCT image analysis. For a typical OCT volume scan with 61 successive B-scans, the manual graders would take approximately 23 hours on average to complete, compared with Niblack and MMCQ which would take 23 seconds and 131 seconds, respectively. This translates to significant time saving for the researcher, allowing them to focus on the science rather than the image analysis. Note that Agrawal's original method [158] took approximately 1 minute per B-scan due to additional manual identification of the ROI to measure the choroid. This execution time has been significantly reduced by virtue of automatically detecting the fovea to produce standardised, choroid-derived measurements.

## 3.4 MMCQ's Reproducibility

We investigate the reproducibility of MMCQ and Agrawal's Niblack method [28] using two datasets which contain repeated OCT samples.

### 3.4.1 Study population

Two samples were used in this section, the i-Test and Glasgow Caledonian University (GCU) Topcon datasets. The i-Test cohort [225] is being collected to investigate the maternal vascular response to pregnancy that may manifest within the eye. Specifically, the i-Test study is interested in the utility of ocular biomarkers being predictive of late-gestation pre-eclampsia. The GCU Topcon cohort [234] was collected to investigate diurnal variation of the choroid in eyes of varying refractive status, and had a preference for recruiting hyperopic eyes. The study population and imaging characteristics of the two samples used to assess the reproducibility of MMCQ (and Niblack) are shown in table 3.4. Briefly, there were 120 eyes from 60 participants in the i-Test sample and 33 eyes from 21 participants in the GCU Topcon sample with repeated data available at the point of analysis.

Acquisition for i-Test [225] used the SD-OCT Heidelberg OCT1 Spectralis Standard and FLEX modules (Heidelberg Engineering, Heidelberg, Germany). During acquisition two unregistered, macula-centred volume scans of both eyes were taken covering a  $30 \times 20$  degree (approximately  $9 \times 6.6$  mm, horizontal  $\times$  vertical) field of view (FOV). Volume scans were taken immediately (within 1 minute) after one another, with EDI mode toggled on and off. For each EDI-OCT volume, 31 parallel line-scans were collected, linearly spaced approximately 240 microns apart with an ART of 50. For each non-EDI OCT volume, 61 parallel line-scans were collected, linearly spaced approximately 120 microns apart with an ART of 12. OCT B-scans were of pixel resolution  $496 \times 768$  (pixel height  $\times$  width).

Acquisition for the GCU Topcon sample used the SS-OCT Topcon DRI Triton Plus device (Topcon, Tokyo, Japan). During acquisition, a twelve-line, fovea-centred radial scan was captured, starting from a horizontal-line B-scan and

## Chapter 3. MMCQ: Semi-automatic choroid vessel segmentation using multi-scale median cut quantisation

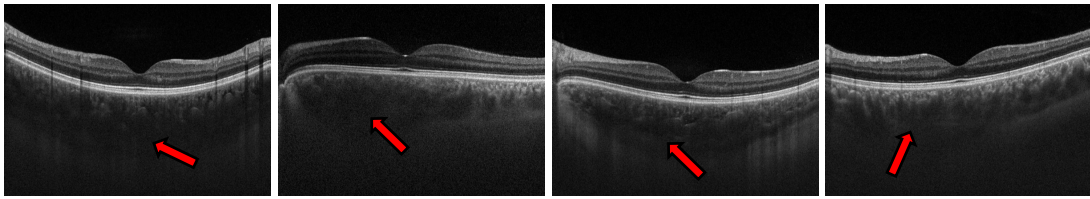
|                                | Study   |   |
|--------------------------------|---|---|
|                                | i-Test [225]  | GCU Topcon [234]                                |
| <b>Cohort demographics</b>     |   |   |
| Participants (Eyes)            | 60 (120)  | 21 (33)   |
| Right eyes (%)                 | 60 (50)   | 15 (45.5)                                       |
| Age (SD)                       | 34.7 (5.2)  | 23.9 (4.2)                                      |
| Sex, F (%)                     | 60 (100)  | 9 (43)  |
| Ethnicity                      | 52 White, 6 Asian, 2 Mixed                            | 12 White, 5 Asian, 2 Black, 2 Middle Eastern    |
| Gestation                      | 35.6 (3.4)  | NA  |
| Refractive status <sup>†</sup> | 3 hyperopes, 31 emmetropes, 26 myopes                 | 9 hyperopes, 8 emmetropes, 3 myopes, 1 missing  |
| Study purpose                  | Pre-eclamptic / normative pregnancy at late gestation | Diurnal variation                               |
| Control/Case                   | 45/15   | 21/0  |
| <b>Image characteristics</b>   |   |   |
| Device                         | Spectralis (Heidelberg)                               | DRI Triton Plus (Topcon)                        |
| OCT Type                       | Spectral-domain                                       | Swept-source                                    |
| Scan Pattern                   | Macular volume  | Macular radial                                  |
| Mode                           | HRA+OCT   | CFP+OCT   |
| Time of day (Interval)         | All in afternoon (1 minute)                           | 13 morning, 12 afternoon, 8 evening (5 minutes) |
| B-scans per eye                | 31 (EDI) / 61 (non-EDI)                               | 12  |
| ART                            | 50 (EDI) / 12 (non-EDI)                               | NA  |
| SNR                            | 35.61   | >88   |
| Pixel resolution               | 496 × 768   | 992 × 1024                                      |

**Table 3.4:** Population demographics and image characteristics of the datasets used to assess the reproducibility of MMCQ and Niblack. <sup>†</sup>: myopic/hyperopic status defined as  $< -1 / > 1$  dioptres. Dioptre measurements for i-Test sample was scan focus collected from the OCT imaging device metadata, and this was the spherical equivalent for the GCU Topcon sample.

rotating clockwise in 30 degree intervals. Each B-scan covered approximately a 9 mm FOV laterally and had a pixel resolution of  $992 \times 1024$  pixels. Any set of 12 B-scans with an average image quality score less than 88 — reported by the SNR index built-in to the Topcon scanning device — was excluded. Repeated acquisitions were taken at the same location during each visit within 5 minutes. Given repeated acquisition was not the primary aim of the GCU Topcon study, only 21 participants had repeated OCT data, with nine only in one eye, and twelve in both eyes.

### 3.4.2 Statistical analysis

For all repeated pairs, we segmented the choroidal vessels using both algorithms, after segmenting the choroidal space and detecting the fovea using



**Figure 3.31:** OCT B-scan slices from each eye excluded from the vessel segmentation reproducibility analysis due to poor vessel definition (arrows).

Choroidalyzer (see chapter 5, as this method is out of the scope of this chapter). Similar to section 3.3, we fixed the parameter configuration for MMCQ and Niblack such that MMCQ's vessel segmentation cluster configuration was  $(K, k) = (20, 11)$  and for Niblack we fixed the window size  $w$  to 51 and the standard deviation offset  $k$  to -0.05 [163].

For the i-Test sample, we measured the choroid vessel volume (in  $\text{mm}^3$ ) and CVI for the nine sub-fields in the ETDRS grid [197] for each EDI-OCT and non-EDI OCT volume pair. Thus, there were 1080 comparisons made per choroid metric for this sample ( $9 \text{ values} \times 120 \text{ eyes}$ ). Details on measuring the vessel and CVI maps can be found in section 1.7.2.

For each participant in the GCU Topcon sample, OCT B-scan capture was collected throughout the day, with some examinations containing repeated acquisitions. There were 79 instances of repeated OCT B-scan data overall, and approximately 25% of these included more than two acquisitions per examination. Additionally, each acquisition included 12 OCT B-scans per eye. To prevent any sampling bias and remain objective in our reproducibility analysis, we selected one repeated pair per eye to maximise equal sample size for morning, afternoon and evening acquisition where possible. This led to selecting 13 repeated pairs from the morning, 12 from the afternoon and 8 from the evening. Upon further inspection of each participant's choroid, four eyes from two participants were excluded because of poor vessel visualisation. An OCT B-scan slice from each eye can be seen in figure 3.31, with arrows indicating very poor vessel definition in the choroidal space. Thus, there were 348 comparisons made ( $29 \text{ eyes} \times 12 \text{ B-scans}$ ), using choroid vessel area (in  $\text{mm}^2$ ) and CVI measured in a 6 mm, fovea-centred ROI. Details on making these

choroid measurements can be found in section 1.7.1.

We measure the reproducibility of MMCQ and Niblack at the population-level, reporting population mean and standard deviation (SD), as well as mean absolute error (MAE) and Pearson and Spearman correlation coefficients. Additionally, we show correlation plots and use Bland-Altman [227] plots to assess the relationship between repeated measurements and the distribution of residuals. We also report the reproducibility of MMCQ and Niblack at the eye-level using the measurement noise parameter  $\lambda$  [200]. All aforementioned metrics have been defined and described earlier in section 1.7.4.

### 3.4.3 Results

#### 3.4.3.1 Population-level

Table 3.5 presents the population-level reproducibility results for MMCQ and Niblack. Both MMCQ and Niblack had strong levels of reproducibility for vessel metrics across the macula and at the fovea. For MAE in the i-Test SD-OCT sample, MMCQ had greater reproducibility than Niblack (ETDRS CVI for MMCQ vs. Niblack was 0.0112 vs. 0.0134; ETDRS vessel volume for MMCQ vs. Niblack was 0.0251 vs 0.0257 mm<sup>3</sup>), while the opposite was observed for the GCU Topcon SS-OCT dataset (fovea CVI for MMCQ vs. Niblack was 0.0113 vs. 0.0083; foveal vessel area for MMCQ vs. Niblack was 0.0400 vs 0.0363 mm<sup>2</sup>). This could be because, at the time of MMCQ's methodology development, there was only access to SD-OCT data, which is also why the evaluation in section 3.3 only featured SD-OCT data.

Encouragingly, both methods had strong linear and monotonic correlations between repeated samples (All Pearson and Spearman correlations were greater than 0.96 and 0.89 for MMCQ and Niblack, respectively). The MAE for ETDRS CVI of approximately 0.01 corresponds to a 1% error on average across individual sub-fields of the ETDRS grid. Additionally, a similarly low CVI MAE was found for fovea-centred SS-OCT B-scans in GCU Topcon sample, in which we report a 1% and 0.8% MAE for MMCQ and Niblack, respectively.

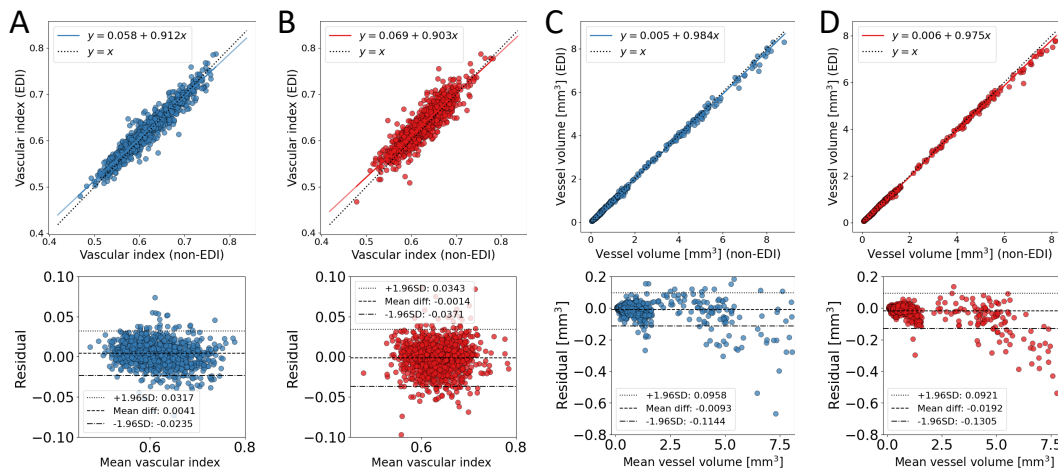
### 3.4. MMCQ's Reproducibility

| Dataset                             | Metric [unit]                       | Method  | Mean (SD)       | MAE           | Pearson       | Spearman      |
|-------------------------------------|-------------------------------------|---------|-----------------|---------------|---------------|---------------|
| i-Test<br>(SD-OCT<br>ETDRS)         | CVI                                 | MMCQ    | 0.6121 (0.0351) | <b>0.0112</b> | <b>0.9633</b> | <b>0.9600</b> |
|                                     |                                     | Niblack | 0.6335 (0.0333) | 0.0134        | 0.8993        | 0.9022        |
|                                     | vessel volume<br>[mm <sup>3</sup> ] | MMCQ    | 0.9424 (0.3233) | <b>0.0251</b> | 0.9994        | 0.9987        |
|                                     |                                     | Niblack | 0.9662 (0.2741) | 0.0257        | <b>0.9995</b> | <b>0.9990</b> |
| GCU<br>Topcon<br>(SS-OCT<br>Foveal) | CVI                                 | MMCQ    | 0.6332 (0.0332) | 0.0113        | <b>0.9266</b> | <b>0.9226</b> |
|                                     |                                     | Niblack | 0.5899 (0.0209) | <b>0.0083</b> | 0.9084        | 0.8949        |
|                                     | vessel area<br>[mm <sup>2</sup> ]   | MMCQ    | 1.1994 (0.3522) | 0.0400        | <b>0.9871</b> | <b>0.9881</b> |
|                                     |                                     | Niblack | 1.0994 (0.2522) | <b>0.0363</b> | 0.9796        | 0.9806        |

**Table 3.5:** Reproducibility performance for MMCQ and Niblack. All Pearson and Spearman correlations were statistically significant with P-values  $P < 0.0001$ . Bold values represent where MMCQ or Niblack had the better score, per metric.

To put these errors into context, Breher, et al. [235] tested the reproducibility of Sonoda's Niblack method [27] across different sub-fields of the ETDRS macular region and reported a mean difference ranging from 3.9% to 5.1%. Additionally, Agrawal's major review [28] on CVI as a biomarker in retinal pathology reported changes between healthy and diseased eyes between 2% and 6% [28]. Therefore, it is unlikely that our MAEs of approximately 1% will be clinically significant.

Figures 3.32 and 3.33 show the population-level reproducibility for MMCQ (blue) and Niblack (red) as correlation and Bland-Altman plots for the SD-OCT i-Test and SS-OCT GCU Topcon repeated samples, respectively. Across the board, SD-OCT vessel volume and SS-OCT vessel area had strong linear fits using both methods, with residuals centred around 0. The absolute value of residuals for vessel area increase with the magnitude of vessel pixels in the choroid, and this observation is exacerbated for vessel volume which have an obvious negative skew, especially for the Niblack method. This is perhaps unsurprising because one member of the i-Test volume pair was collected with EDI mode toggled off, which has poorer image quality than the corresponding EDI-OCT member of the pair. In these figures, we can also observe MMCQ's stronger performance on the SD-OCT dataset (figure 3.32) and, conversely, Niblack's superior performance on the SS-OCT dataset (figure 3.33).



**Figure 3.32:** Correlation and Bland-Altman plots for assessing the reproducibility of MMCQ and Niblack using repeated, SD-OCT volume data (i-Test). (A) MMCQ CVI, (B) Niblack CVI, (C) MMCQ vessel volume, (D) Niblack vessel volume.

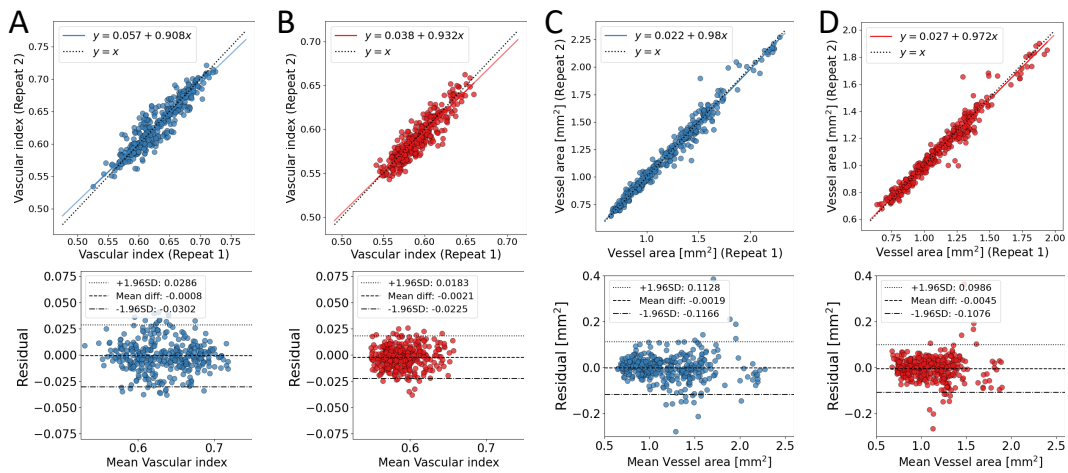
Finally, CVI had a reasonably strong linear fit for both methods across both datasets, but was weaker than vessel area and volume by virtue of errors in both numerator and denominator having an exacerbated effect on overall CVI error (as discussed in section 3.3.3). However, their residual distributions appeared to show no apparent trend across both datasets, and were centred around 0, albeit Niblack’s residual distributions were more spherical and tighter than MMCQ. Note also that CVI was significantly less variable than vessel volume (MMCQ standard deviation for SD-OCT (ETDRS) CVI and vessel volume was 0.0351 and 0.2741 mm<sup>3</sup>, respectively, and for SS-OCT (foveal) CVI and vessel area was 0.0332 and 0.3522 mm<sup>2</sup>, respectively).

### 3.4.3.2 Eye-level

Eye-level reproducibility helps quantify the effect size of your measurement error relative to the overall variability of the population. Effect size is critical in understanding the degree to which measurements correspond to real biological change compared to measurement error. Figures 3.34 and 3.35 show the eye-level reproducibility of MMCQ (blue) and Niblack (red) for the SD-OCT i-Test



### 3.4. MMCQ's Reproducibility



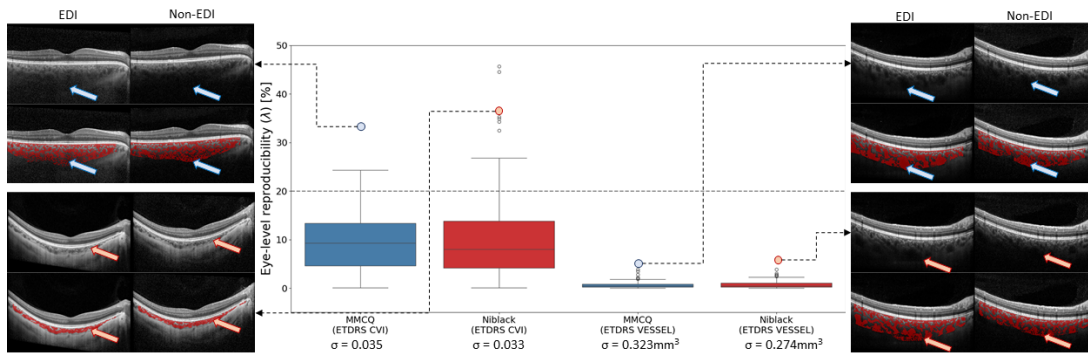
**Figure 3.33:** Correlation and Bland-Altman plots for assessing the reproducibility of MMCQ and Niblack using repeated, SS-OCT B-scan data (GCU Topcon). (A) MMCQ CVI, (B) Niblack CVI, (C) MMCQ vessel area, (D) Niblack vessel area.

and SS-OCT GCU Topcon samples, respectively. In each case, we show distributions of eye-level measurement noise  $\lambda$  [200]. A value of  $\lambda = 20\%$  has been overlaid as a black, horizontal dashed line to aid interpretation and readability, with representative B-scans of major outliers shown with their corresponding vessel segmentations overlaid in red, with arrows indicating the source of the error. The between-eye standard deviations are shown below each box-plot as  $\sigma$ .

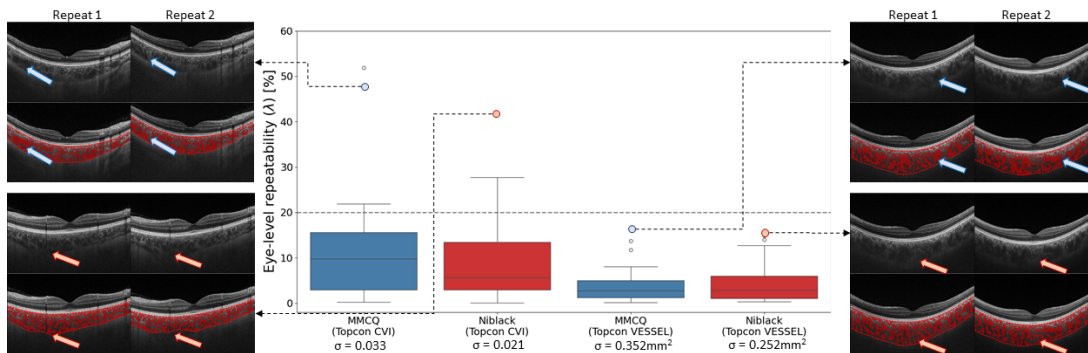
At the eye-level, vessel area and volume showed significantly lower measurement variability across the population compared with CVI for both MMCQ and Niblack methods. Specifically, 100% and 75% of the eyes in the i-Test and GCU Topcon samples were below about 7% of the overall population variability for vessel volume, respectively. Conversely, for CVI, 75% of eyes for both datasets were below 15% of the overall population variability. It's perhaps unsurprising that CVI is less reproducible than vessel area or volume because it is more prone to larger error from both the numerator (vessel area) and denominator (choroid area). CVI is dependent on both vessel and region segmentation. Thus, error in CVI incorporates measurement error from both vessel and total choroid area in the numerator and denominator of its formula, and the size of



### Chapter 3. MMCQ: Semi-automatic choroid vessel segmentation using multi-scale median cut quantisation



**Figure 3.34:** Reproducibility for MMCQ (blue) and Niblack (red) at the eye-level for the SD-OCT volume data (i-Test). (A) MMCQ CVI, (B) Niblack CVI, (C) MMCQ vessel volume, (D) Niblack vessel volume.



**Figure 3.35:** Eye-level for assessing the reproducibility of MMCQ and Niblack using repeated, SS-OCT B-scan data (GCU Topcon). (A) MMCQ CVI, (B) Niblack CVI, (C) MMCQ vessel area, (D) Niblack vessel area.

the choroid can also disproportionately inflate error.

Moreover, CVI had significantly smaller population variability than vessel area/volume (also observed by Agrawal, et al. [158]). This smaller variability exacerbates this potential for larger error even further. Additionally, smaller variability implies that effect sizes are likely to be small, which consequently means that reported differences between groups need to be interpreted with care. All of these points indicate that vessel area and volume are more reliable a measurement of the choroidal vasculature than CVI, at least according to MMCQ and Niblack.

Figures 3.34 and 3.35 also show the major outliers according to the eye-level

reproducibility analysis. For the i-Test sample, unsurprisingly, the greatest outliers from both methods for ETDRS CVI and vessel volume came from the same two eyes, which both had large choroids. For ETDRS CVI, even the EDI-OCT scan did not visualise the choroid particularly well (top-left OCT B-scans in figure 3.34), leading to unwanted noise in the output vessel segmentations which translated to high error. For ETDRS vessel volume (right-hand side OCT B-scans in figure 3.34), under-segmentation from both approaches for the non-EDI volume scan lead to large error in vessel volume. This highlights a significant limitation behind choroid vessel segmentation which is the pre-requisite segmentation of the choroidal space. Thus, the reproducibility of choroid-derived vessel metrics are dependent on the reproducibility of choroid-derived regional metrics [235].

Interestingly, one of the major outliers for Niblack's ETDRS CVI was an eye whose choroid was reasonably small and visualised well, as shown in the bottom-left set of OCT B-scans in figure 3.34. Here we observe major sources of over-segmentation in both sets of scans. This highlights Niblack's inability to segment smaller choroids well, which is likely a consequence of poorly selected window size  $w$ , which translated to a  $\lambda$  value of approximately 36% (of the population's variability). Conversely, MMCQ reported only a 3%  $\lambda$  value for this eye, suggesting MMCQ's precision for smaller choroids.

The major outliers for the SS-OCT GCU Topcon sample in figure 3.35 follow suit to the SD-OCT i-Test sample for foveal CVI, i.e. the same eye corresponded to the largest eye-level reproducibility error for both methods (bottom-left set of B-scans in figure 3.35). For this particular choroid, the region was not segmented correctly for the repeated scan, thus overestimating the CVI in both methods leading to large measurement error, relative to the population variability (MMCQ and Niblack  $\lambda$  were 52% and 42%, respectively). This further highlights the dependence that choroid-derived vessel metrics have on choroid-derived regional metrics [235].

Interestingly, the second largest outlier for foveal CVI for both methods (top-left set of B-scans in figure 3.35) came from a choroid with good visualisation of the Choroid-Sclera boundary and vessels. However, upon comparing the appearance of the vessels in the pair it was found that they were not registered

properly. Topcon imaging devices do not have the same advanced eye tracking system which Heidelberg Engineering imaging devices do. Thus, there is potential for repeated acquisitions to not be perfectly registered, which can lead to a very slightly different cross-sectional visualisation of the choroidal vasculature. While emerging research suggests that macular CVI is stable [184, 185, 186], in this analysis this actually produced large measurement error. Although, the measurement error was worse in MMCQ than it was for Niblack ( $\lambda$  value of 47% and 28%, respectively) suggesting that Niblack may be marginally more generalisable to macular scanning locations than MMCQ [184, 185, 186].

For choroid vessel area, the same eye produced the largest measurement error  $\lambda$  for both methods (top- and bottom-right side OCT B-scans in figure 3.35), which was 16% and 15% for MMCQ and Niblack, respectively. While the SS-OCT imaging properties enables sufficient Choroid-Sclera boundary visualisation, this comes at decreased SNR for deeper ocular structures (larger choroids) [97], which can result in poorly defined choroidal vessels with a lack of contrast and vessel boundary definition. While MMCQ was able to make a reasonable effort of segmenting whole vessels, Niblack struggled to segment the posterior segment of the choroid, leading to a noisy segmentation (red arrows). Moreover, the representative pair of OCT B-scans shown in the right-hand side of figure 3.35 do not appear to be perfectly registered (blue arrows) which also likely played a role in the reported measurement error.

## 3.5 Discussion

### 3.5.1 Evaluation

We showed that the proposed method, MMCQ, agreed better with the segmentations from two manual graders compared to Agrawal's Niblack method [28]. Our method achieved lower MAEs and stronger correlations for the majority of comparisons with manual labelling on choroid-derived measurements of vessel area and CVI. MMCQ also had greater segmentation agreement, albeit

with lower sensitivity than Niblack. However, while our comparison suggests MMCQ's superiority, these are according to manual vessels labels and the observed differences in these population-level performance metrics are small in absolute terms.

There does not yet exist a gold standard for choroid vessel segmentation, which makes quantitative evaluation of automatic methods for vessel segmentation a challenge (hence the lack thereof in table 1.2). To attempt evaluation rigour and facilitate segmentation consistency, our evaluation procedure used two expert analysts to conduct segmentation, both following the same pre-specified protocol which ensured a similar setup of screen size, image resolution and brightness, and application-dependent instructions on ITK-Snap [229]. This protocol ultimately helped facilitate reasonable agreement in segmentation and measurements between graders M1 and M2.

Relative to the manual segmentations in our evaluation, we showed that Agrawal's Niblack method [158] has an inherent problem of over-segmentation, particularly for small choroids and small-scale vessels. This is an observation which, to our knowledge, is only mentioned very briefly in the original study proposing Agrawal's method for choroid vessel segmentation using Niblack [158]. While a reasonable circumvention would likely be to tune the window size  $w$  and standard deviation offset  $k$ , this becomes unfeasible as datasets grow in size.

While MMCQ was superior to Niblack in this evaluation, the magnitude of agreement in CVI overall between the semi-automatic methods and the manual segmentations was generally poor and exhibited significant cases of systematic bias. Encouragingly, the semi-automatic methods did exhibit a linear and monotonic relationship with the manual methods for choroid vessel area, but this was not observed for CVI.

The poor agreement observed between the MMCQ and Niblack methods, and their tendency to agree for choroids with more vessel area (or thicker choroids) is an essential observational difference between the two approaches (figure 3.26(E)). MMCQ and Niblack represent two common ideologies which are akin to the precision-recall trade-off that exists within machine learning. Niblack prioritises recall over precision, which ensures all vessels are classified, at

the cost of miss-classifying some of the interstitial space as vessel. On the other hand, MMCQ prioritises precision (or positive predictive value), which promotes vessel pixel fidelity, i.e. at the cost of under-segmentation of the choroidal vessels, MMCQ minimises miss-classifying interstitial pixels as choroid vessel. This was exemplified by MMCQ scoring higher precision and Niblack scoring higher recall against the manual graders (table 3.2).

An important consequence of evaluating against manual vessel labels in this chapter is the assumption that the manual vessel labels are accurate and of a gold standard. While there was reasonable agreement between graders, it was by no means clinically insignificant [28]. This highlights the potential variation that exists within manual choroid vessel segmentation in OCT B-scans, given the potential for different definitions, experience and effort that two different human observers may apply to the same OCT B-scan. While our evaluation shows MMCQ as a more faithful methodology according to these manual graders, the lack of gold standard for this segmentation task, accompanied by human subjectivity and poor vessel visualisation, contribute to the overall problem of identifying a singularly superior approach to choroid vessel segmentation.

Agrawal's modified Niblack method [158] (and its older variant [27]) is clearly the most widespread approach for choroid vessel segmentation in OCT image analysis [28]. However, it remains subjective in nature due to its poor agreement [173] and its ambiguity around parameter selection (section 3.1.1). There should be more concern among the research community around this lack of parameter reporting and its subsequent improper use by the research community. To our best knowledge, only two studies report the values for  $w$  and  $k$  [167, 163] for ground truth label generation of deep learning-based algorithms, with other studies *presumably* resorting to default application from Fiji (ImageJ) or from Agrawal's implementation [28]. However, default application fixes  $w$  and  $k$  which is not optimal across choroids of different sizes as previously observed (figures 3.4, 3.5 and 3.6). This has the potential to impact standardisation and reproducibility of choroid-derived vascular metrics in ocular and systemic health.

Various image quality issues contribute to poor vessel visualisation, including

speckle noise, poor eye tracking, difficult patient fixation and poor acquisition from image technicians. There is an inherent assumption made by manual graders, and indeed the semi-automatic methods discussed in this chapter, which is that vessel segmentation in OCT images is seen as a binary classification problem. However, poor vessel visualisation tends to create an ambiguous choroidal space where smaller vessels, and large vessel boundaries may be misconstrued as interstitial fluid, or vice versa for interstitial space around vessels which are visualised obliquely at image capture. This all carries major challenges to produce an accurate binary classification of the choroidal vessels. Thus, we suggest that future work would make significant progress by providing a probabilistic solution to account for the uncertainty of low contrast vessel boundary regions.

### 3.5.2 Reproducibility

Both MMCQ and Niblack showed strong levels of reproducibility at the population-level, which was lower than previously reported effect sizes between healthy and diseased eyes (between 2% and 6%) [28], and previously reported studies of reproducibility [235] (3.9% to 5.1%). Thus, regardless of their ability to accurately segment the choroidal vasculature, their measurement error is lower than what current effect sizes in disease are. This suggests that these approaches should not significantly diminish the biological signal being measured, assuming they can do so accurately in the first place, as the measured effect should be larger than any measurement variability inherent in the method.

Moreover, at the eye-level, it's encouraging to see that this reproducibility analysis showed major outliers representing challenging cases of thick choroids with poor Choroid-Sclera boundary and/or vessel visualisation, or where OCT B-scans did not appear to be sufficiently registered in the case of the GCU Topcon sample. Heidelberg Engineering are able to leverage the confocal SLO image to enable real time eye tracking [97], unlike Topcon imaging devices like the DRI Triton Plus, which use CFP as the corresponding en face localiser during image acquisition.

Finally, we concluded at the eye-level that choroid-derived vessel area and vessel volume appeared to be significantly more reliable than CVI. This was because of the dependence CVI had on both the reproducibility of region and vessel segmentation which has the propensity to exacerbate measurement error unlike metrics like vessel area and volume. Additionally, the significantly lower population variability in CVI, whose standard deviation across both datasets was approximately 10% of the standard deviation reported for vessel area and vessel volume, also played a contributory role in exacerbating measurement variability.

Thus, it is imperative that results from CVI measurements are interpreted according to the effect size associated to real biological change, which is dependent on the method used to estimate it [235]. Fortunately, the reproducibility analysis conducted for MMCQ and Niblack help provide a interpretable metric for end-users to apply to their own studies to help differentiate measurement error from true biological change. Ultimately, the potential for higher measurement error in CVI suggests that predominantly vascular metrics like area and volume are more reliable and should be considered just as important to report in future studies alongside CVI.

### 3.5.3 Limitations and future work

While we propose MMCQ as a promising approach for choroidal vessel segmentation, it still faces several limitations that affect its current applicability. The very nature of the choroidal vasculature presents a challenge to segment accurately, given its complex and dense mesh network. This means that OCT B-scans only provide oblique cross-sectional visualisations of a microvascular bed which is highly heterogeneous in terms of size and shape. This makes it difficult to detect individual vessels and their morphology. As a result, while MMCQ can measure overall vessel area and density within the choroid, it falls short in offering more localised metrics, such as vessel calibre distributions or 3D rendering of the entire vascular network across the macula.

Another potential limitation of MMCQ is its inability to accurately quantify individual vascular layers within the choroid, particularly Sattler's and Haller's lay-



ers. However, while previous research has attempted this [155, 165], there is currently no agreed-upon and standardised protocol in the literature. Sim, et al. [108] did draft a protocol for multi-layer choroid segmentation, but only in a relatively small sample looking at diabetic eyes. There is also currently no reliable ground truth/gold standard data which exists to differentiate the choroidal vasculature into these individual vascular layers, or distinguish large vessels from medium-sized ones [8]. It's possible that histological samples of the choroid may help with developing an automatic procedure in the future, but validating any such approach would prove a significant challenge compared with only identifying vasculature across the choroidal space alone.

MMCQ is a pipeline with multiple stages, which introduces additional complications. For example, the pipeline's performance is highly interdependent, with each stage influencing the subsequent one. Sub-optimal performance during pre-processing or enhancement can negatively impact the subsequent vessel segmentation for example. Moreover, while domain-specific decisions such as scaling configurations (section 3.2.2.1) and patch-based quantisation level estimation (section 3.2.2.2) were made with experienced judgement in choroid image analysis, these were not rigorously validated across a diverse source of OCT choroid data (such as retinal pathology). Thus, it would be pertinent to conduct rigorous testing on each stage of the pipeline to assess its feasibility across different choroids in both health and retinal disease.

Additionally, the algorithm was primarily designed and tested on SD-OCT data from Heidelberg Engineering OCT devices due to a lack of access to SS-OCT data at the time of algorithm development. This likely contributed to the lower reproducibility observed in the GCU Topcon SS-OCT sample, relative to the i-Test SD-OCT sample. This further highlights the need for additional testing and validation of MMCQ across different OCT systems to ensure generalisability.

A key assumption in MMCQ is its focus on improving precision (positive predictive value) in choroidal vessel segmentation by avoiding over-segmentation of connective tissue within the choroid. While this approach is sensible given the ambiguous nature of the choroidal vessel walls appearing on OCT B-scans, it may result in under-segmentation (lower recall, or sensitivity). This would render vessel wall tissue potentially miss-classified, particularly for larger choroidal



vessels.

Another limitation is MMCQ's reliance on prior segmentation of the choroidal space — MMCQ requires another approach to quantify the choroidal space before MMCQ can be applied. Ideally, an end-to-end procedure able to simultaneously segment the choroidal space and vessels would be advantageous for choroidal image analysis in OCT image sequences. Moreover, MMCQ also requires an element of human subjectivity by requiring detection of the fovea on OCT B-scans. This is so that CVI and vessel area can be reliably measured in a standardised, fovea-centred ROI. These limitations add technical and clinical complexity, decreasing its end-user accessibility for automatic choroidal vasculature measurement.

Both MMCQ and Niblack operate under the assumption that the choroidal vasculature can be classified into binary “vessel”/“not-vessel” categories in OCT B-scans. However, this approach is misguided due to the inherent uncertainty and non-binary nature of the choroidal vessels on OCT B-scans, compounded by their limited resolution and oblique appearance. MMCQ begins to address this issue by allowing researchers to adjust the number of quantised levels  $k$  to assign as choroidal vasculature (out of a total of  $K$  clusters). However, a more robust solution would be to characterise the uncertainty of choroid pixels, and particularly vessel wall pixels.

Ultimately, the future direction for MMCQ lies in developing an approach that not only handles this uncertainty but also provides a fully automatic, end-to-end solution. Such a method would enable reproducible and accurate choroid-derived measurements directly from an OCT B-scan, through the automatic segmentation of both the choroidal space and its vessels.

### 3.5.4 Outputs

In this chapter, there were no publication outputs, but this method was used to generate ground truth labels for the method introduced in chapter 5, and is briefly described in the publication therein [4].

The software associated with this method, MMCQ, has been published as open-source and is freely available on GitHub [here](#).

### 3.5.5 Executive summary

In this chapter, we discussed how quantifying choroidal vasculature is likely more informative than the choroidal space alone, allowing measurement of vessel area and CVI. We found that the current reference standard approach, the Niblack method, has internal parameters which greatly influence the output segmentation and are not reported in the literature, which can lead to poor standardisation and reproducibility. We introduced a domain-specific, semi-automatic approach to choroid vessel segmentation, Multi-scale Median-cut Quantisation (MMCQ), to replace this current reference standard, and address the heterogeneity and contrast issues of the choroidal vasculature in OCT B-scans.

There is currently no gold standard to choroid vessel segmentation, thus making it a challenge to evaluate automatic methods accurately. According to manual ground truth labels from two experienced raters, MMCQ was more precise but Niblack was more sensitive. However, these differences were small, and inherently biased by comparing to segmentations flawed by human bias. MMCQ and Niblack together agreed reasonably well in vessel area and volume, but not for CVI. We found that both methods were reproducible at the population- and eye-level, with errors below reported effect sizes from the literature in retinal disease. However, measurement error in vessel segmentation in both methods was exacerbated more in CVI than in purely vascular metrics like vessel area and volume. This was due to error exacerbation in the formula for CVI. Thus, while it remains a popular metric in ocular and systemic health, CVI should be interpreted with care and reported alongside more reliable vascular metrics like vessel area and volume.

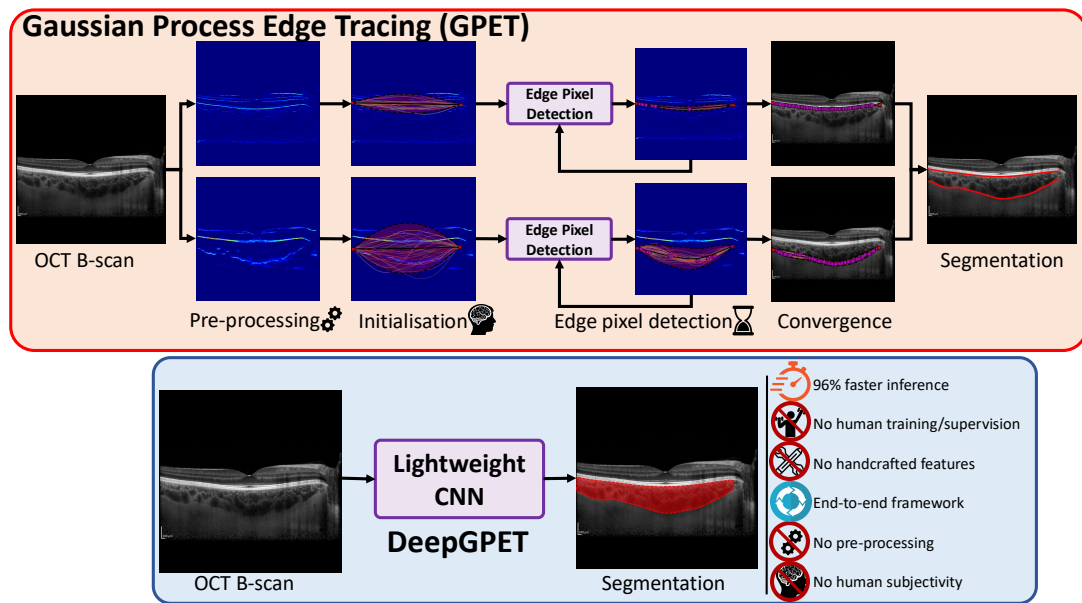
# Chapter 4

## DeepGPET: Efficient and fully-automatic choroid region segmentation

### 4.1 Introduction

Earlier in this thesis we introduced GPET, a semi-automatic approach to pixel-based edge tracing (edge-based image segmentation). We showed that GPET was suitable for choroid region segmentation through detection of the RPE-Choroid and Choroid-Sclera boundaries, and that this method was a significant improvement over manual measurement for downstream measurement of the choroid.

However, the semi-automatic nature of GPET requires time-consuming manual interventions by specifically trained personnel which introduces subjectivity. This limits GPET's potential for processing larger datasets or being used in research studies by those without experience in image processing or mathematics. Unfortunately, the training required for this approach extends not only from computer vision and image processing but also a sufficient theoretical background in Gaussian process regression. Additionally, execution time depends on image quality, with successful segmentation requiring manual identification of boundary endpoints for each boundary for a single B-scan, and potentially



**Figure 4.1:** Detailed schematic comparison between GPET and the proposed deep learning alternative, DeepGPET.

requires a range of pre-processing techniques to overcome noise to provide a target edge for GPET to trace in the first instance.

Thus, providing a fully automatic solution to choroid region segmentation in substitute of GPET which not only has quicker execution time but is end-to-end, i.e. without manual input or pre-processing, would be optimal. A solution based on deep learning methods has several advantages over semi-automatic solutions. Figure 4.1 presents a direct comparison between GPET’s semi-automatic solution (A) and the proposed deep learning alternative (B). The deep learning alternative removes the multi-stage procedure of semi-automatic GPET, requiring no pre-processing to prepare images for downstream segmentation, and removes any need for human intervention or initialisation at the point of image analysis.

A common way to replace an existing method with one based on deep learning is through *distillation*. Deep learning has been used to distil existing semi-automatic traditional image processing pipelines into fully-automatic methods. That is, the semi-automatic method has been used to generate ground truth labels which are used to train and evaluate the corresponding deep learning

model. Recently, Engelmann, et al. [236] used deep learning to emulate a previous semi-automatic approach [237] for retinal fractal dimension approximation, the aim to overcome image quality in CFP.

Following a similar suit, we expect the proposed deep learning-based solution to be more robust to poorer image quality than GPET because of the lack of necessary human input and subjectivity, and the substitution of handcrafted features for automatic, hierarchical feature learning. Additionally, a deep learning alternative will be able to take into account the entire B-scan for training and inference, segmenting the entire choroidal space in a single sweep, rather than segmenting each individual upper and lower boundary. The former only requires identification of a reasonably large area of the B-scan, rather than the latter which requires precise contour tracing and linkage of individual and adjacent pixels in series, thus treating the choroids' anterior and posterior surfaces effectively as independent entities.

Thus, we aimed to develop a fully-automatic method for end-to-end choroid region segmentation that can be easily used without specialised training. We do this through distilling GPET into a deep learning algorithm, DeepGPET, which can process images without human supervision in a fraction of the time. These attributes permit analysis of large-scale datasets and potential deployment into research practice without prerequisite knowledge in image processing or computer science. Moreover, DeepGPET will be released to address the distinct lack of open-source methods for automatic choroid region segmentation, and in order to replace pre-existing manual measurement of the choroid as the current reference standard.

### 4.1.1 Deep learning for image segmentation

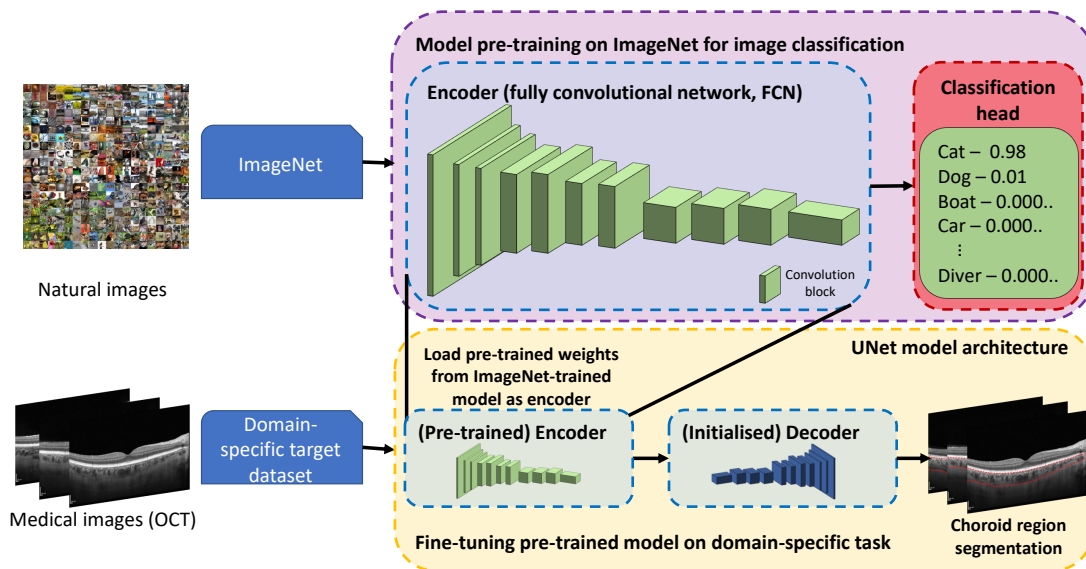
Deep learning has revolutionised the field of image segmentation, enabling more accurate and efficient delineation of objects within images than traditional methods [129]. Traditional image segmentation methods often rely on handcrafted features, which require domain expertise and are not always effective across different datasets or image modalities. In contrast, deep learning models can automatically extract relevant features from raw image data without the

need for manual intervention. This automation not only reduces the time and effort required for feature engineering but also enables the models to adapt to a wide range of segmentation tasks and domains. Deep learning's hierarchical model architectures (such as CNNs) can automatically learn hierarchical features (from low-level to high-level) from the input data. This capability allows them to capture complex patterns and details within images, leading to more robust performance compared to traditional imaging methods [129].

Among the numerous models developed for this purpose, the UNet [138] model stands out as one of the most influential and widely used architectures [240], originally introduced for biomedical image segmentation. The model's architecture is characterised by its U-shaped structure, comprising a contracting path (known as the encoder arm) to capture context and a symmetric expanding path (known as the decoder arm) that enables precise localisation [240]. This design allows UNet to excel in segmenting images with complex structures, even when training data is limited [138]. A more detailed description of the UNet deep learning architecture can be found in section 5.1.1.

In recent years, the practice of fine-tuning pre-trained models has been a standard approach to model training in the field of image segmentation [241]. Fine-tuning involves taking a model that has been pre-trained on a large dataset and adapting it to a specific task by training it on a smaller, domain-specific dataset. For fine-tuning semantic segmentation models, the encoder arm has typically been pre-trained on some large dataset like ImageNet [238], such that the weights already represent low- and high-level features of some natural image dataset, while the weights in the decoder arm are semi-randomly initialised — care is given to ensure random initialisation does not harm training [242]. Fine-tuning then constitutes training the decoder weights from scratch, while the encoder weights can either be trained simultaneously, or frozen altogether. In the latter, this is so that the learned representations from pre-training are not hindered by fine-tuning on an image domain entirely different to the larger dataset it was previously trained on. Although, the former allows the encoder to be tuned to the domain of interest.

Figure 4.2 shows the typical setup for fine-tuning a deep learning model pre-trained on a large and diverse dataset such as ImageNet [238]. Note that the



**Figure 4.2:** Diagram describing the process of fine-tuning a model for image segmentation using a pre-trained model from a large natural image dataset, such as ImageNet [238]. ImageNet demonstrative image courtesy of [239].

convolutional block shown in green is a combination of linear and non-linear mappings which enable hierarchical feature representation learning. This approach is particularly beneficial in scenarios where annotated data is scarce, as it allows leveraging the learned features from the pre-trained model, reducing the need for large amounts of labelled data. Fine-tuning can be used in combination with models like UNet [138], as the model’s architecture can be pre-trained on generic image datasets and later fine-tuned on domain-specific data, such as medical images for specific segmentation tasks, enhancing both the speed and accuracy of the training process. The combination of UNet’s robust architecture and the fine-tuning of pre-trained models offers a powerful strategy for image segmentation, particularly in medical imaging and other domains where data is limited. As deep learning continues to evolve, these methods are likely to remain at the forefront of image segmentation research, driving further advancements in accuracy, efficiency, and applicability across various fields.

## 4.2 Study populations

### 4.2.1 Model training & evaluation

We used 715 OCT B-scans belonging to 82 subjects from three studies of healthy and diseased individuals (unrelated to ocular pathology):

- **OCTANE** [225], a longitudinal cohort ( $N=47$ ) study investigating choroidal changes in renal transplant recipients and donors. Note that this is 2 participants and 6 B-scans more than the dataset in GPET's evaluation. This was because of additional data found available on the OCT imaging device during data extraction;
- **i-Test** [225], a cohort ( $N=5$ ) of pregnant women evaluating retinochoroidal changes in normative, pre-eclamptic or fetal growth restricted pregnancies;
- **Normative**, a cohort ( $N=30$ ) of healthy volunteers collected as a control group for a study investigating retinal biomarkers in multiple sclerosis [243].

All studies conformed with the Declaration of Helsinki and received relevant ethical approval and informed consent from all subjects.

Two Heidelberg Engineering SD-OCT Spectralis OCT1 modules were used for image acquisition: the Spectralis Standard and FLEX modules. Horizontal- and vertical-line scans with a  $30^\circ$  (approximately 9 mm laterally) fovea-centred FOV were captured using active eye tracking with an ART of 100. Posterior pole macular line scans covered a  $30 \times 20$  degree rectangular ROI using 31 consecutive line scans using an ART of 50 for the i-Test cohort, and 61 consecutive line scans using an ART of 9 for the Normative and OCTANE cohorts.

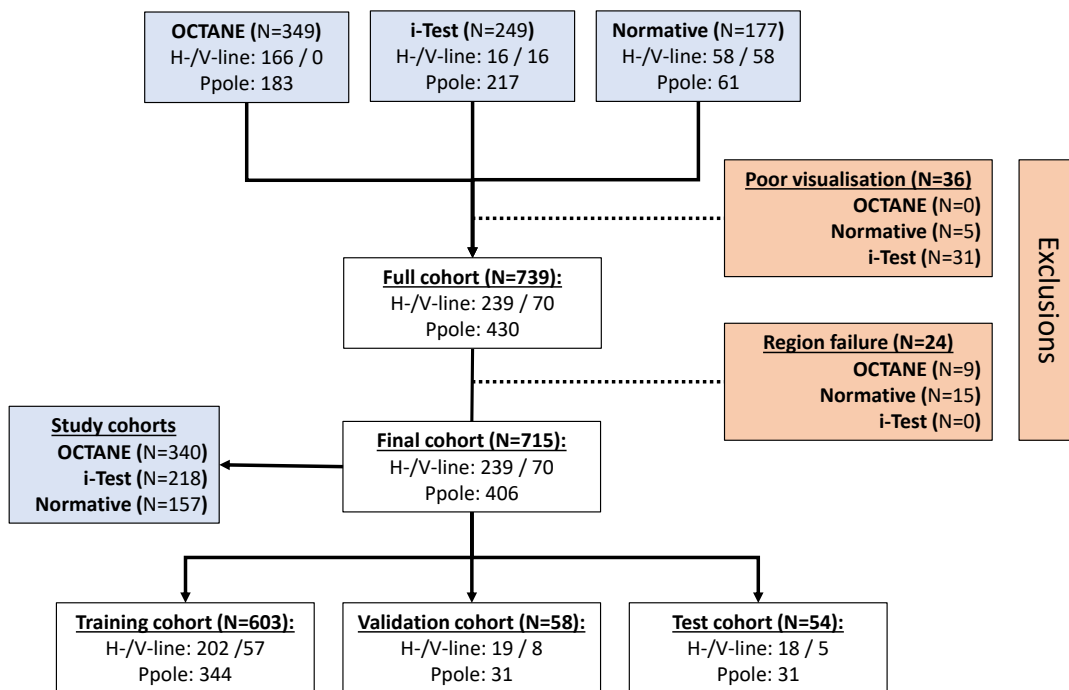
OCT B-scans were captured at a pixel resolution of  $768 \times 768$  pixels using the Standard Spectralis module and  $496 \times 768$  using the FLEX module which was then padded vertically to a common resolution of  $768 \times 768$ . To reduce processing, images were then cropped to remove unnecessary blank space



above and below where the choroid and retina appear on the OCT B-scan such that input images for training were size  $544 \times 768$ .

#### 4.2.1.1 Sample derivation

Figure 4.3 shows a sample derivation flowchart of how DeepGPET’s modelling dataset was constructed. At the time of DeepGPET’s development, there were a total of  $N=775$  OCT B-scans available from 82 participants. Upon visual inspection,  $N=36$  B-scans were removed from the dataset due to poor quality image acquisition, which resulted in an invisible Choroid-Sclera boundary. Of the remaining  $N=739$  OCT B-scans, ground-truth labelling for the choroidal space was provided by GPET (chapter 2), after manual selection of each boundary’s edge endpoints using a custom-built graphical user interface in Python (version 3.11). Upon manual adjudication of these ground-truth labels,  $N=24$  B-scans constituting posterior pole volume B-scans, were excluded. This was because



**Figure 4.3:** Sample derivation flowchart of how OCT data from three cohorts were combined to create DeepGPET’s modelling dataset.

## 4.2. Study populations

|                           | OCTANE      | i-Test     | Normative  | Total       |
|---------------------------|-------------|------------|------------|-------------|
| Subjects                  | 47          | 5          | 30         | 82          |
| Male/Female               | 24 / 23     | 0 / 5      | 20 / 10    | 44 / 38     |
| Right/Left eyes           | 47 / 0      | 5 / 5      | 29 / 29    | 81 / 34     |
| Age (mean (SD))           | 48.8 (12.9) | 34.4 (3.4) | 49.1 (7.0) | 48.0 (11.2) |
| Machine                   | Spectralis  | FLEX       | Spectralis | Both        |
| Horizontal/Vertical scans | 166 / 0     | 16 / 16    | 57 / 54    | 239 / 70    |
| Volume scans              | 174         | 186        | 46         | 406         |
| EDI / non-EDI             | 340 / 0     | 218 / 0    | 57 / 100   | 615 / 100   |
| Total B-scans             | 340         | 218        | 157        | 715         |

**Table 4.1:** Overview of population characteristics for DeepGPET’s training and evaluation dataset.

|                        | Training    | Validation | Testing    | Total       |
|------------------------|-------------|------------|------------|-------------|
| Subjects, N            | 66          | 9          | 7          | 82          |
| OCTANE cohort          | 38          | 5          | 4          | 47          |
| i-Test cohort          | 3           | 1          | 1          | 5           |
| Normative cohort       | 25          | 3          | 2          | 30          |
| Male/Female            | 32 / 34     | 7 / 2      | 5 / 2      | 44 / 38     |
| Age (mean (SD))        | 49.2 (11.4) | 40.2 (8.2) | 44.9 (8.2) | 48.0 (11.2) |
| Right/Left eyes        | 66 / 27     | 9 / 4      | 6 / 3      | 81 / 34     |
| Spectralis/FLEX Device | 63 / 3      | 8 / 1      | 6 / 1      | 77 / 5      |
| EDI/non-EDI            | 51 / 15     | 6 / 3      | 5 / 2      | 62 / 20     |
| B-scans                |             |            |            |             |
| Horizontal/Vertical    | 202 / 57    | 19 / 8     | 18 / 5     | 239 / 70    |
| Volume                 | 344         | 31         | 31         | 406         |
| Total                  | 603         | 58         | 54         | 715         |

**Table 4.2:** Overview of population and image characteristics of the training, validation and test sets for DeepGPET’s model training and evaluation.

GPET failed to segment the Choroid-Sclera boundary in regions of low SNR, as a result of low-ART volume B-scans.

Table 4.1 provides an overview of basic population characteristics and number of subjects/images of these studies. Notably, 14% of the OCT B-scans were non-EDI and thus presented more challenging cases with lower SNR.

We split the data into approximately an 85:8:7 ratio split between training (603 B-scans, 66 subjects), validation (58 B-scans, 9 subjects) and test sets (54 B-scans, 7 subjects). To avoid within-patient data leakage between sets, when splitting the data we did so at the patient-level, i.e. each subject’s OCT images

were present in only one set. Moreover, the sets were selected so that each set had proportionally equal amounts of scan types (EDI/non-EDI) to best represent image quality. Table 4.2 provides for an overview of basic population and imaging characteristics for each set.

## 4.2.2 Reproducibility

We also assess the reproducibility of DeepGPET on downstream choroidal measurements on two types of OCT data: macular volume scans and peripapillary B-scans. Two cohorts were used in this analysis, i-Test<sup>1</sup>[225] and Diurnal Variation for Chronic Kidney Disease (DVCKD) [225]. We used all available eyes which had repeated data available (at the point of analysis) from each cohort (120 eyes from 60 participants from the i-Test cohort and 22 eyes from 22 participants in the DVCKD cohort). Table 4.3 describes their population and imaging characteristics. Note that the i-Test sample of repeated data is the same dataset used to assess the reproducibility of MMCQ and Niblack in chapter 3, which was described in section 3.4.1.

In brief, the i-Test sample collected paired EDI and non-EDI macular SD-OCT volume scans (acquired within 1 minute of one another), both covering a  $30 \times 20$  degree (approximately  $9 \times 6.6$  mm) FOV. EDI volume scans contained 31 line-scans with an ART of 50 while non-EDI scans contained 61 line-scans with an ART of 9. Line-scans had a pixel resolution of  $496 \times 768$ .

DVCKD participants were recruited to assess diurnal variation of the choroid in healthy volunteers to support research in CKD [16]. During acquisition, the SD-OCT Heidelberg Spectralis Standard Module was used to collect horizontal-line OCT B-scans, OCT volume scans and OCT peripapillary scans of the right eye only. To assess DeepGPET's generalisability to other locations in the posterior pole, and to minimise the affect of diurnal variation on this reproducibility analysis, the macular OCT scans were not selected for analysis, and only the peripapillary B-scans were. Peripapillary B-scans collected for the DVCKD study were circular cross-sections of the retina and choroid and centred on the

---

<sup>1</sup>Note that this cohort was collected after developing DeepGPET, hence the increase in sample size from table 4.1.

### 4.3. DeepGPET's deep learning model

|                              | Study   |   |
|------------------------------|---|---|
|                              | i-Test [225]  | DVCKD [225, 16]                               |
| <b>Cohort demographics</b>   |   |   |
| Participants (Eyes)          | 60 (120)  | 22 (22)                                       |
| Right eyes (%)               | 60 (50)   | 22 (100)                                      |
| Age (SD)                     | 34.7 (5.2)  | 21.3 (2.2)                                    |
| Sex, F (%)                   | 60 (100)  | 10 (45.5)                                     |
| Ethnicity                    | 52 White, 6 Asian, 2 Mixed                            | Unknown                                       |
| Gestation                    | 35.6 (3.4)  | NA  |
| Study purpose                | Pre-eclamptic / normative pregnancy at late gestation | Diurnal variation                             |
| Control/Case                 | 45/15   | 22/0  |
| <b>Image characteristics</b> |   |   |
| Device                       | Spectralis (Heidelberg)                               | Spectralis (Heidelberg)                       |
| OCT Type                     | Spectral-domain                                       | Spectral-domain                               |
| Scan Pattern                 | Macular volume  | Peripapillary                                 |
| Mode                         | HRA+OCT   | HRA+OCT                                       |
| Time of day (Interval)       | All in afternoon (1 minute)                           | Each at 9am, 12:30pm, 4pm ( $\pm$ 40 minutes) |
| B-scans per eye              | 31 (EDI) / 61 (non-EDI)                               | 1   |
| ART                          | 50 (EDI) / 12 (non-EDI)                               | 100   |
| Pixel resolution             | 496 $\times$ 768                                      | 496 $\times$ 1536                             |

**Table 4.3:** Population demographics and image characteristics of the reproducibility cohorts to evaluate DeepGPET's reproducibility.

optic disc. B-scans were collected with EDI mode activated using an ART of 100 and had a pixel resolution of 496  $\times$  1536. The study collected scans at three time points during daylight, in the morning (09:12  $\pm$  12 minutes), early afternoon (12:36  $\pm$  7 minutes) and early evening (16:08  $\pm$  8 minutes).

## 4.3 DeepGPET's deep learning model

DeepGPET used a UNet deep learning architecture [138] and we employed fine-tuning. DeepGPET's encoder arm is a CNN architecture called MobileNetV3 [244] whose model weights were pre-trained using the large bank of natural images in ImageNet [238]. DeepGPET's decoder arm was symmetrical in terms of blocks and spatial resolution to the encoder arm, with default random ini-

tialisation of model weights [245]. The MobileNet CNN architectures were designed to be efficient, scalable and lightweight in terms of parameter (model weights) size and inference time [244], with the overall aim of these architectures being used for mobile applications.

DeepGPET was fine-tuned by retraining the encoder weights, as well as the decoder weights from scratch, for 60 epochs — an epoch representing one entire pass through the training dataset — using a batch size of 16. Model training leveraged the AdamW optimiser (using standard parameters  $\beta_1=0.9$ ,  $\beta_2=0.999$  and weight decay  $10^{-2}$ ) [246], and the learning rate was  $10^{-3}$ . The Adam optimiser [247] combines older concepts around momentum [248] and parameter-specific adapting learning rates [249] to improve convergence rates [250], and AdamW extends this further by using additional weight decay as implicit regularisation to help prevent overfitting [246].

Exponential moving average (EMA) is a simple yet effective approach to overcome occasionally poor quality ground truth labels, prevent noisy predictions during training and enhance transfer learning [251]. EMA of the model weights simply computes a moving average for each model parameter across epochs, placing higher emphasis to more recent epochs — this is in contrast to the linear moving average applied to retinal vessel shadow compensation in section 3.1.3. Thus, to enhance model performance and help overcome any poorly labelled ground-truths, we utilised EMA across model parameters which started after epoch 30.

We employ common data augmentation techniques such as brightness and contrast alterations ( $p = 0.5$ ), Gaussian blur ( $p = 0.25$ ) and scale, rotation and translation transformations ( $p = 0.25$ ). More realistic augmentations include horizontal flipping ( $p = 0.5$ ) to synthetically produce a right eye scan from a left eye scan and vice versa, and applying Gaussian noise followed by multiplicative noise ( $p = 0.5$ ) to corrupt images with custom speckle noise to simulate the natural noise degradation seen in OCT image data. Data augmentation is very useful for artificially obtaining a larger, and more diverse dataset. Data augmentation can be done off-line in advance, but is better applied during training where randomly augmented samples are generated anew within an epoch, and therefore are different per epoch. While this approach to generating ‘new’ data

is only artificial augmentations from pre-existing data, empirically this does improve the model's robustness and generalisation to augmented instances not readily available in the original dataset [252].

DeepGPET was developed, trained and evaluated using Python (3.11), PyTorch (2.0) and Segmentation Models PyTorch [245].

## 4.4 DeepGPET's evaluation

### 4.4.1 Statistical analysis

For evaluation, a threshold of  $p=0.5$  was used to convert the model's probabilistic prediction map to a binary segmentation mask.

We used the Dice similarity score and Area Under the Receiver Operating Characteristic Curve (AUC) for evaluating segmentation agreement, as well as the Pearson correlation  $r$  and mean absolute error (MAE) for segmentation-derived choroid thickness and area. These metrics have been defined previously in this thesis in section 1.7.4. We measure choroid thickness and area following the definitions outlined in section 1.7.1, take a representative thickness value as the average of three locations; subfoveal and 2000 microns temporal and nasal to the fovea, while area is measured in a 6 mm fovea-centred ROI.

To aid comparison of DeepGPET and GPET, we compare DeepGPET's agreement with GPET's ground-truth segmentations, and this is further compared against the *repeatability* of GPET. GPET was used to process all 54 test images 2 months after it was used to generate the initial set of segmentations used for DeepGPET's model evaluation. Both times, the same experienced analyst who developed the GPET method (author J.B.) segmented the choroidal space, which acts as an upper bound of performance of GPET which we compare DeepGPET with.

We define *repeatability* here as the ability to reproduce segmentations (and choroid-derived measurements) on the same input, i.e. when the same OCT

B-scan is processed by GPET twice. Conversely, *reproducibility* — used elsewhere in this thesis — is defined as the ability to reproduce segmentations (and choroid-derived measurements) from different inputs of the same object, i.e. when two OCT B-scans are taken 5 minutes apart for the same location of the same eye. Thus, the expected output should be the same between repeatability and reproducibility, but the inputs are different.

We also measured the execution time of both approaches on the test set and reported mean and standard deviation values in seconds. The execution time for GPET was measured on the re-segmentations, with the author timing himself from when each image was first loaded up on the screen, to when a successful choroid segmentation was obtained after application of the GPET image analysis pipeline (figure 2.13). Experiments were run on a standard, Windows laptop with a 3.5 year old Intel Core i5 (8<sup>th</sup> generation) CPU with 16Gb of RAM, but no graphics processing unit (GPU).

In addition to quantitative evaluations, we also compared segmentations by GPET and DeepGPET for 20 test set OCT B-scans qualitatively by having them rated by an experienced clinical ophthalmologist (supervisor Dr. Ian J.C. MacCormick, I.M.). We selected 7 examples with the highest disagreement in thickness and area, 7 examples with disagreement closest to the median, and 6 examples with the lowest disagreement. Thus, these 20 examples cover cases where both methods are very different, cases of typical disagreement, and cases where both methods are very similar. In each instance, I.M. was shown the segmentations of both methods overlaid on the OCT — blinded to which method produced which segmentation — and also provided with the raw, full-resolution B-scan. I.M. was then asked to rate each one along three dimensions: Quality of the upper boundary, quality of the lower boundary and overall smoothness using an ordinal scale: “Very bad”, “Bad”, “Okay”, “Good”, “Very good”. Quality was defined by how well the detected boundary traced along each true boundary, given the definitions set out in appendix A.1 using equipment specified in appendix A.2. Smoothness was defined by a lack of bumps or jagged edges in the predicted boundary segmentations for the RPE-Choroid and Choroid-Sclera boundaries.

## 4.4.2 Results

### 4.4.2.1 Quantitative evaluation

Table 4.4 shows the results for DeepGPET and a repeat GPET, compared to the initial GPET segmentation as ground truth.

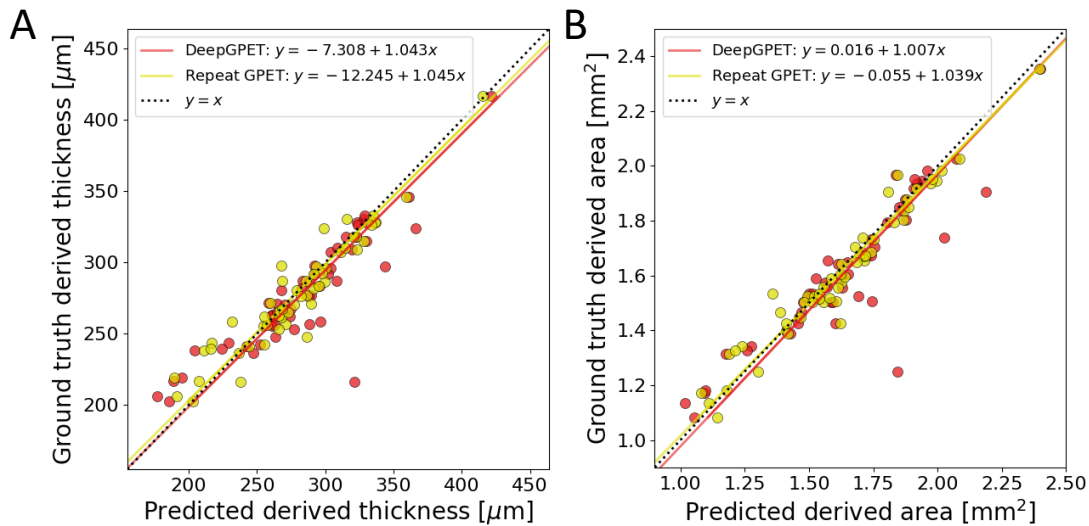
Both methods have excellent agreement with the original segmentations. DeepGPET’s agreement is comparable to the repeatability of GPET itself, with DeepGPET’s AUC being slightly higher (0.9994 vs 0.9812) and Dice coefficient slightly lower (0.9664 vs 0.9672). DeepGPET performing better in terms of AUC but worse in terms of Dice suggests that there is higher uncertainty (lower confidence) for pixels where it disagrees with GPET after thresholding, compared with pixels where it agrees with GPET. This in turn suggests that the raw (probabilistic) predictions made by DeepGPET are well-calibrated to the original GPET segmentations for each pixel. Thus, the raw DeepGPET probabilities provide a reasonable measure of uncertainty of the choroidal space.

DeepGPET processed the images in only 3.6% of the time that GPET needed. DeepGPET ran fully-automatic and successfully segmented all images in 1.25 seconds on average, whereas GPET took 34.50 seconds and required 1.27 manual interventions on average. Manual interventions were defined where the initial segmentation failed because of poor pre-processing or initialisation of GPET parameters (covariance function hyperparameters). Moreover, each additional manual intervention per B-scan corresponded to re-selecting the initial edge endpoints as well as intermediary pixels between the edge endpoint pixels to help guide tracing for the Choroid-Sclera boundary. In contrast, DeepGPET ran end-to-end without the need for pre-processing or manual input.

| Method      | AUC    | Dice   | Time<br>(s/img)   | Thickness   |                       | Area        |                       |
|-------------|--------|--------|-------------------|-------------|-----------------------|-------------|-----------------------|
|             |        |        |                   | Pearson $r$ | MAE [ $\mu\text{m}$ ] | Pearson $r$ | MAE [ $\text{mm}^2$ ] |
| DeepGPET    | 0.9994 | 0.9664 | 1.25 $\pm$ 0.10   | 0.8908      | 13.3086               | 0.9082      | 0.0699                |
| Repeat GPET | 0.9812 | 0.9672 | 34.49 $\pm$ 15.09 | 0.9527      | 10.4074               | 0.9726      | 0.0486                |

**Table 4.4:** Metrics for DeepGPET and repeated GPET using the initial GPET annotation as ground truth. Time given as mean  $\pm$  standard deviation.

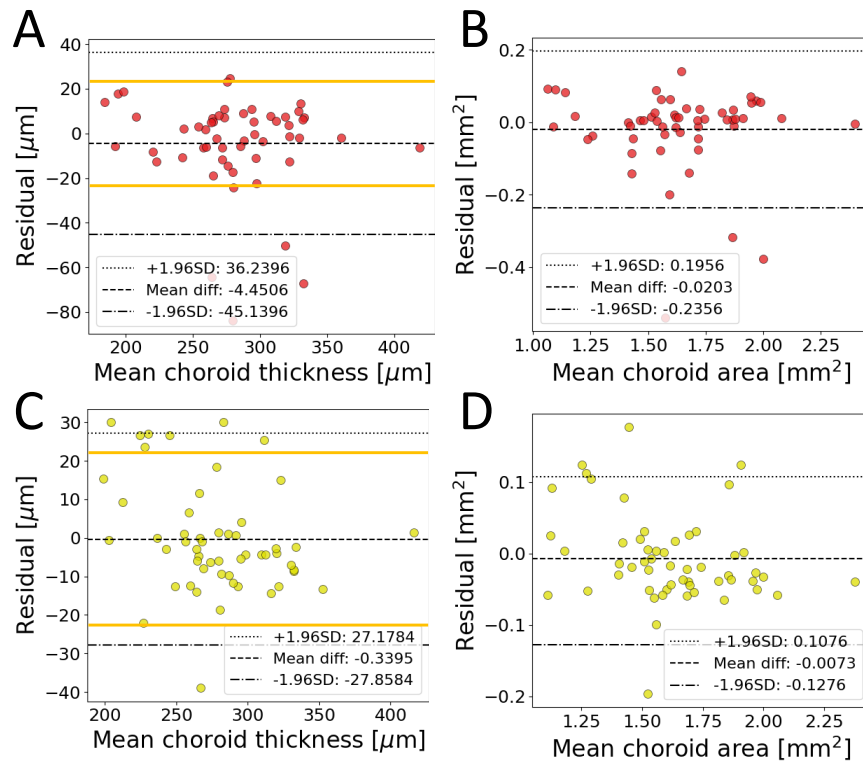




**Figure 4.4:** Correlation plots comparing derived measures of mean choroid thickness (A) and choroid area (B) using DeepGPET and the re-segmentations using GPET.

GPET showed very high repeatability for thickness (Pearson  $r=0.9527$ , MAE =  $10.4074 \mu\text{m}$ ) and area (Pearson  $r=0.9726$ , MAE =  $0.0486 \text{ mm}^2$ ). DeepGPET achieved slightly lower, yet also very high agreement for both thickness (Pearson  $r=0.8908$ , MAE =  $13.3086 \mu\text{m}$ ) and area (Pearson  $r=0.9082$ , MAE =  $0.0699 \text{ mm}^2$ ). Figure 4.4 shows correlation plots for thickness and area. DeepGPET's agreement with GPET does not quite reach the repeatability of GPET itself, when used by the same experienced analyst, but it is quite comparable and high in absolute terms.

Furthermore, the regression fits in both derived measures for DeepGPET are closer to the identity line than for the repeated GPET measurements. For choroid thickness, the linear fit estimated a slope value of 1.043 (95% confidence interval of 0.895 to 1.192) and intercept of  $-7.308 \mu\text{m}$  (95% confidence interval of  $-48.967 \mu\text{m}$  to  $34.350 \mu\text{m}$ ). For choroid area, the linear fit estimated a slope value of 1.01 (95% confidence interval of 0.878 to 1.137) and an intercept of  $0.016 \text{ mm}^2$  (95% confidence interval of  $-0.195 \text{ mm}^2$  to  $0.226 \text{ mm}^2$ ). All confidence intervals contained 1 and 0 for the slope and intercepts, respectively, suggesting no systematic bias or proportional difference between GPET and DeepGPET [228, 253].



**Figure 4.5:** Bland-Altman plots comparing agreement in thickness and area for DeepGPET (A – B) and GPET's repeatability (C – D). Orange lines in panels (A, C) at  $\pm 23 \mu\text{m}$  represent intra-rater repeatability threshold [106].

Figure 4.5 shows the residuals between DeepGPET and the re-segmentations of GPET against the original GPET ground truth labels from the held-out test set using Bland-Altman plots [227]. Rahman, et al. [106] found that intra-rater agreement and inter-rater agreement of manual subfoveal choroidal thickness measurements were  $23 \mu\text{m}$  and  $32 \mu\text{m}$ , respectively. For DeepGPET's residual error in choroid thickness, only 9.3% (5 / 54) were greater than  $23 \mu\text{m}$  in absolute value, while 15% (8 / 54) of the GPET re-segmentations exceeded this threshold (figure 4.5). For choroid area, the majority of residuals were centred around 0 (mean residual for DeepGPET:  $-0.02 \text{mm}^2$ , GPET re-segmentations:  $-0.0073 \text{mm}^2$ ), however, GPET had tighter distribution of residuals, and were better centred around 0 for both thickness and area (95% confidence intervals for thickness: DeepGPET,  $(-45.14, 36.24) \mu\text{m}$ ; GPET,  $(-27.17, 27.86) \mu\text{m}$ ; and area: DeepGPET,  $(-0.24, 0.20) \text{mm}^2$ ; GPET,  $(-0.13, 0.11) \text{mm}^2$ ).

#### 4.4.2.2 Qualitative evaluation

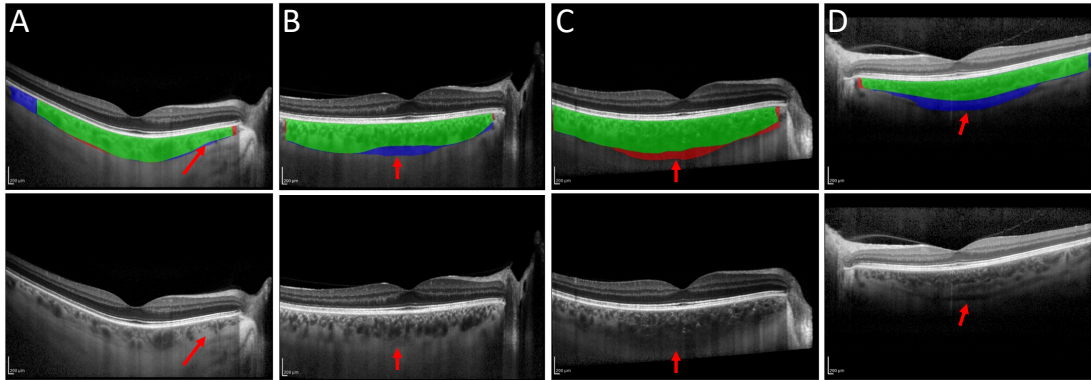
Table 4.5 shows the results of the manual adjudication between different cases comparing the segmentations outputted from GPET and DeepGPET. The upper boundary was rated as “Very good” for both methods in all 20 cases. However, for the lower boundary, DeepGPET was rated as “Bad” in 2 cases for the lower boundary and 1 case for smoothness. Otherwise, both methods performed very similarly.

Figure 4.6 shows some examples. In panel (A), DeepGPET segmented more of the temporal region than GPET did, providing a full width segmentation which was preferred by the adjudicator I.M. Additionally, both approaches were able to segment a smooth boundary, even in regions with stroma fluid obscuring the lower boundary (red arrow). In panel (B), the lower boundary for this choroid is very faint and is actually below the majority of the vessels sitting most posterior (red arrow). DeepGPET produced a smooth and concave boundary preferred by the adjudicator, while GPET fell victim to hugging the posterior most vessels in the subfoveal region. In panel (C), DeepGPET rejected the true boundary in the low contrast region (red arrow) and opted for a more well-defined one, while GPET segmented the more uncertain path. Since GPET permits human intervention, there is more opportunity to fine tune its parameters to fit what the analyst believes is the true boundary. Here, the adjudicator preferred GPET, while DeepGPET’s under-confidence led to under-segmentation and to a bad rating.

In panel (D), the Choroid-Sclera boundary is difficult to delineate due to a thick suprachoroidal space (red arrow) and thus a lack of lower boundary definition. In this case, DeepGPET segmented the choroidal space, as well as

| Method   | Upper boundary | Lower boundary                          | Smoothness                              |
|----------|----------------|---|---|
| DeepGPET | Very good: 20  | Very good: 4, Good: 10, Okay: 4, Bad: 2 | Very good: 5, Good: 12, Okay: 2, Bad: 1 |
| GPET     | Very good: 20  | Very good: 6, Good: 6, Okay: 8, Bad: 0  | Very good: 6, Good: 13, Okay: 1, Bad: 0 |

**Table 4.5:** Qualitative ratings of 20 test set segmentations along 3 key dimensions. The rater was blinded to the identity of the methods and their order was randomised for every example.

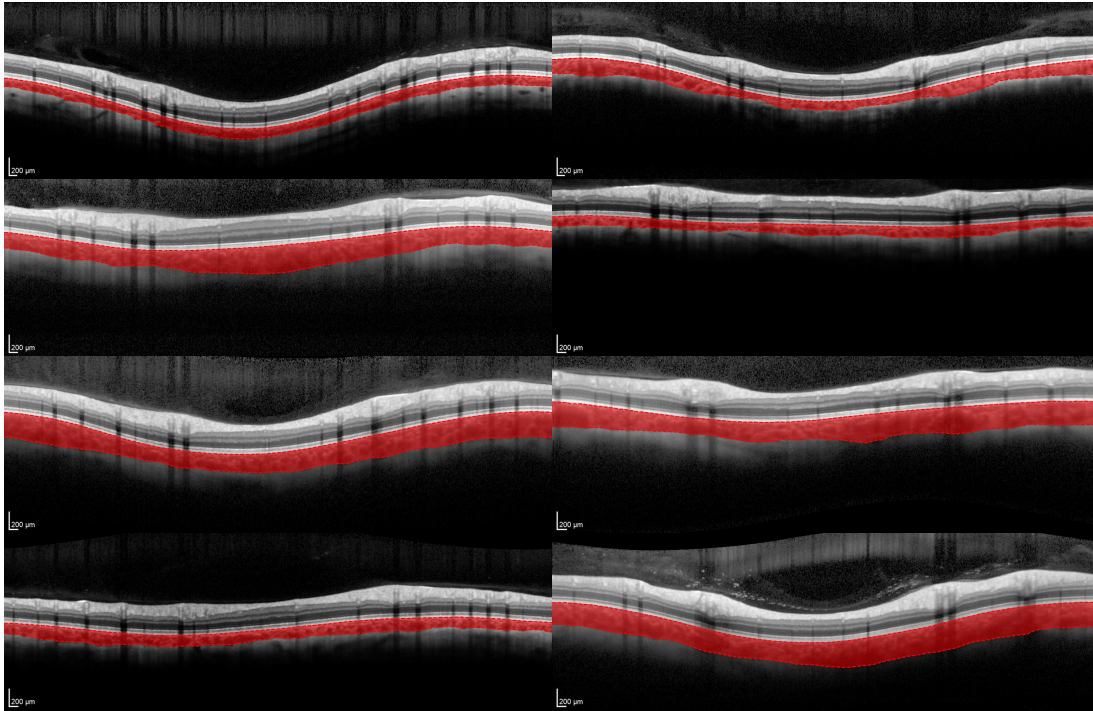


**Figure 4.6:** Four examples from the adjudication. The rater preferred DeepGPET for (A – B) and GPET for (C – D). Top row: green, segmented by both GPET and DeepGPET; red, GPET only; and blue, DeepGPET only. Bottom row: arrows indicate important choroidal features which can make segmentation challenging.

the suprachoroid which became more obvious near the centre of the B-scan. Thus, DeepGPET over-segmented the scan overall, as we exclude the suprachoroidal space from our definition of the choroid (section 1.7.1). While GPET excluded the suprachoroidal space, it actually segmented through posterior choroid vessels, thus under-segmenting. Nevertheless, the adjudicator preferred GPET in this case because we considered the suprachoroid and choroid itself as separate spaces (according to the definitions set out in appendix A.1). Note that the choroids in figure 4.6(B – D) are the choroids with the largest choroid thickness and area disagreement between DeepGPET and GPET as observed in figure 4.5.

#### 4.4.2.3 Potential generalisability

DeepGPET was only trained on macular SD-OCT B-scans using data from Heidelberg Engineering imaging devices. Thus, we do not know its performance on SS-OCT data, or at locations around the periphery of the posterior pole. While not explicitly part of the evaluation of DeepGPET, we conducted some initial experimentation on DeepGPET's potential to generalise to peripapillary SD-OCT and macular SS-OCT data. Note that while we highlight the following potential functionality of DeepGPET, these are not officially validated.

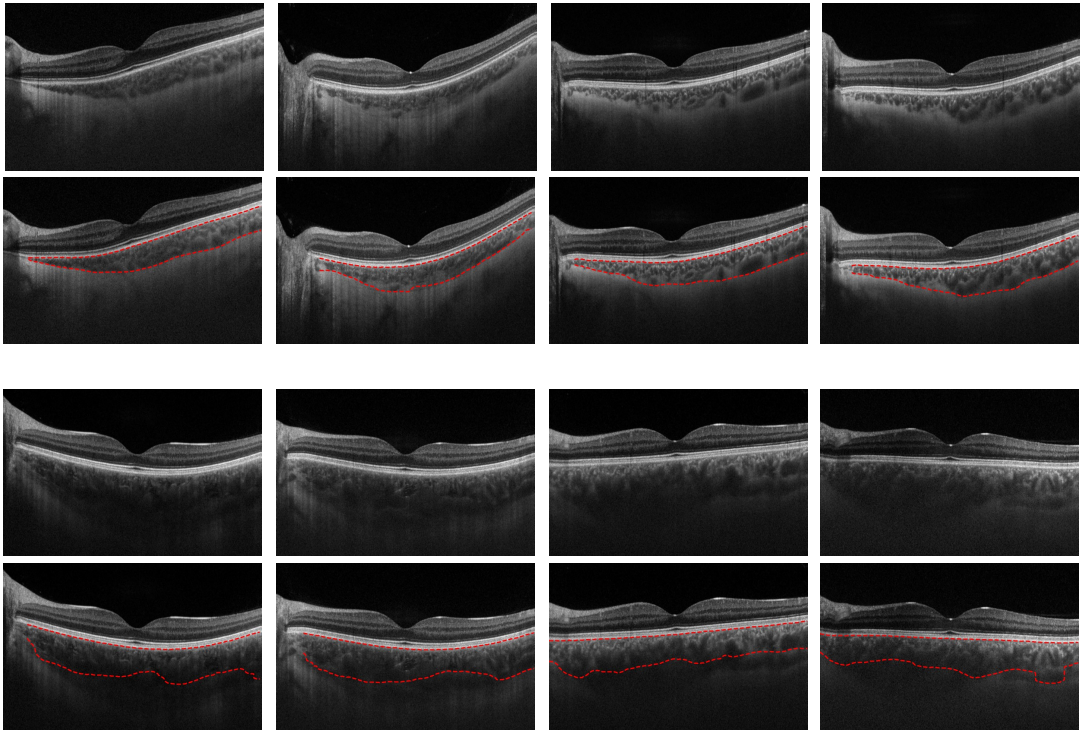


**Figure 4.7:** A selection of peripapillary choroids from 8 different eyes in the DVCKD cohort, with successful choroid region segmentation from DeepGPET.

Firstly, we considered DeepGPET’s potential generalisability to peripapillary OCT B-scans. These are circular-shaped B-scans centred on the optic nerve head and particularly useful for assessing retinal nerve fiber layer thickness. Figure 4.7 shows eight peripapillary choroids using the Heidelberg Engineering SD-OCT Spectralis OCT1 device. These B-scans are from participants in the DVCKD cohort, and are used to assess DeepGPET’s reproducibility. In this exploration, the peripapillary choroids were selected based on image quality, choroid size and extent of sinuosity of the circular peripapillary B-scan.

While DeepGPET was only trained on macular OCT B-scans, it appeared to generalise well to segmenting peripapillary choroids, but tended to not fully segment the whole width. Peripapillary B-scans from Heidelberg Engineering OCT imaging devices had a lateral pixel resolution of 1536 pixels, while the training data only had a lateral pixel resolution of 768 pixels. Ensuring the peripapillary choroid is segmented from edge-to-edge is particularly important since the peripapillary B-scan is circular and is thus continuous at either edge





**Figure 4.8:** A selection of small and large choroids from OCT B-scans collected using the Topcon SS-OCT DRI Triton plus (Topcon, Tokyo, Japan), with segmentations of the upper and lower boundaries as outputted by DeepGPET.

of the image. Therefore, a full segmentation is required to ensure consistent measuring of the different peripapillary sub-fields (section 1.7.3). However, a suitable remedy is to pad the B-scan horizontally by padding each edge with its opposite edge by a few hundred pixels, thus artificially elongating the B-scan. This is a valid approach because of the circular and continuous nature of the B-scan and is sufficient for DeepGPET to segment the entire peripapillary choroid, allowing average thickness in the peripapillary sub-fields to be measured.

Secondly, we wanted to experiment with how robust DeepGPET was to processing macular SS-OCT B-scans with choroids of various size. We selected 8 OCT B-scans from eight eyes from the GCU Topcon cohort used to assess MMCQ's reproducibility from chapter 3. Figure 4.8 shows a row of four small and large choroids, with DeepGPET's region segmentations overlaid in the

subsequent rows. While DeepGPET appeared to be robust to smaller choroids from these SS-OCT B-scans (top rows), for larger choroids whose Choroid-Sclera boundaries had poor visibility, DeepGPET's robustness failed (bottom rows). For these cases, DeepGPET is simply under-segmenting the choroid, and failing to detect the Choroid-Sclera boundary with any certainty.

## 4.5 DeepGPET's reproducibility

### 4.5.1 Statistical analysis

For the OCT volume data in the i-Test sample, we measured average choroid thickness in the 9 sub-fields in the ETDRS grid [197]. For each peripapillary OCT B-scan in the DVCKD sample, we measured average choroid thickness in the 6 peripapillary sub-fields, and the papillomacular bundle. The methods used to generate these sub-field measurements for volume and peripapillary data have been described previously in sections 1.7.2 and 1.7.3, respectively. For the DVCKD sample, there are three images per eye so we compare pairwise, chronological measurements (morning – afternoon, afternoon – evening). Thus, there were 1080 (120 eyes  $\times$  9 locations) and 308 (22 eyes  $\times$  7 locations  $\times$  2 pairwise comparisons) repeated samples to compare for the i-Test and DVCKD samples, respectively.

As we are comparing measurements across the day in the DVCKD sample, diurnal variation of the choroid is a potential confounding factor for this sample. Thus, we address the longitudinal evolution of the peripapillary choroid too.

We measure the reproducibility of DeepGPET at the population-level, reporting population mean and standard deviation (SD), as well as mean absolute error (MAE) and Pearson and Spearman coefficients. These metrics have been defined previously in section 1.7.4. Additionally, we show correlation plots and use Bland-Altman [227] plots to assess the relationship between repeated measurements and the distribution of residuals. We also report the reproducibility of DeepGPET at the eye-level using the measurement noise parameter  $\lambda$  [200] described in section 1.7.4.

## 4.5.2 Results

### 4.5.2.1 Population-level

Table 4.6 presents the reproducibility performance of DeepGPET at the population-level. Peripapillary choroids were thinnest in the DVCKD sample (mean  $\pm$  SD of  $169.9 \pm 54.1 \mu\text{m}$ ) compared to the i-Test sample (mean  $\pm$  SD of  $275.68 \pm 91.8 \mu\text{m}$ ). This makes sense given peripapillary choroids are thinner near the optic disc than in the central macula [254].

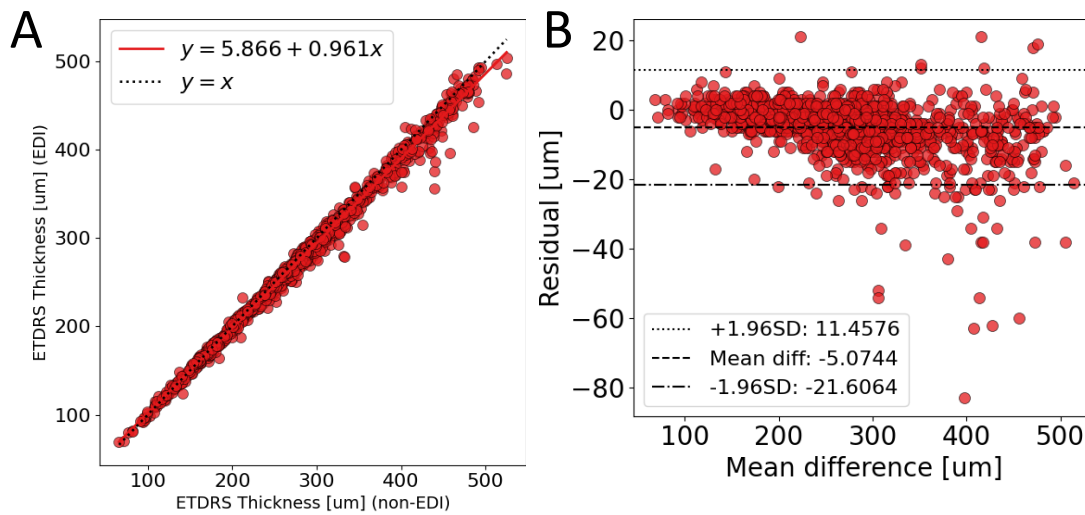
In figure 4.9 we show the correlation and Bland-Altman plots for the i-Test sample. DeepGPET showed very strong and statistically significant correlations between non-EDI and EDI paired ETDRS measurements (Pearson/Spearman for choroid thickness in i-Test: 0.9965/0.9968) and achieved an MAE of  $5.5 \mu\text{m}$ . Compared to manual thickness measurement, this MAE is well within the threshold expected to exceed manual inter- and intra-rater repeatability [106] ( $32$  and  $23 \mu\text{m}$ , respectively). Additionally, from a semi-automatic perspective, an MAE of  $5.5 \mu\text{m}$  was almost half of GPET's repeatability (table 4.4). Moreover, effect size is dependent on the context, and in systemic health this MAE is still below reported differences in choroid thickness which can be as low as  $20 - 30 \mu\text{m}$  such as in myopia progression [24, 110].

Generally, we see that for larger choroids there is larger error in choroid thickness. In the case for i-Test, this is entirely expected because the optical signal for non-EDI OCT acquisition degrades the deeper through the eye it penetrates, thus the visibility of the deeper choroidal vasculature decays rapidly also. This larger error translated into very slight under-estimation of choroid thickness for the non-EDI OCT volume scans of each pair, as seen in the

| Study                 | Mean (SD) [ $\mu\text{m}$ ] | MAE [ $\mu\text{m}$ ] | Pearson | Spearman |
|-----------------------|-----------------------------|-----------------------|---------|----------|
| i-Test (ETDRS)        | 275.68 (85.69)              | 5.5510                | 0.9984  | 0.9979   |
| DVCKD (Peripapillary) | 169.60 (54.07)              | 5.0320                | 0.9942  | 0.9940   |

**Table 4.6:** Reproducibility performance for DeepGPET using macular and peripapillary choroid thickness. All Pearson and Spearman correlations were statistically significant with P-values  $P < 0.0001$ .



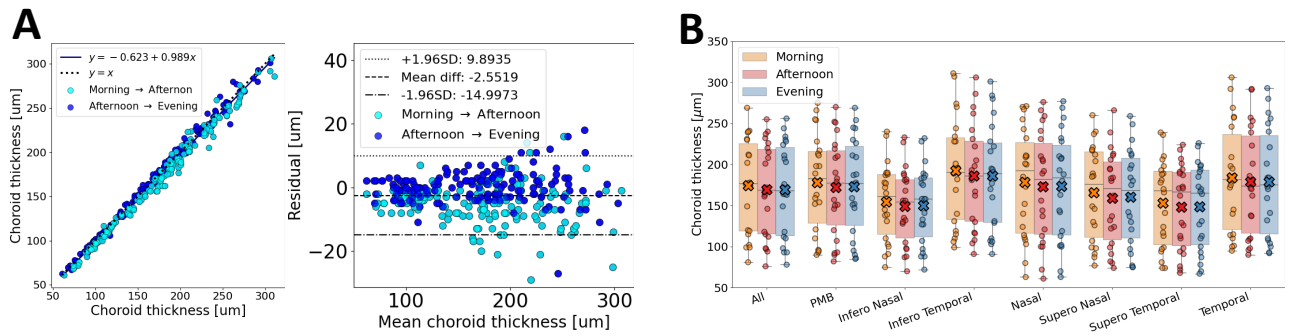


**Figure 4.9:** Correlation (A) and Bland-Altman (B) plots for assessing the reproducibility of DeepGPET on macular OCT volume data from the i-Test sample.

Bland-Altman plot in figure 4.9(B). However, this under-prediction with mean residual of  $-5.07 \mu\text{m}$  (95% confidence interval of  $-21.61 \mu\text{m}$  to  $11.46 \mu\text{m}$ ) is by no means large or clinically significant, highlighting the potential robustness to non-EDI scans of the choroid in macular OCT data.

Figure 4.10(A) show the correlation and Bland-Altman plots for comparing pairwise peripapillary sub-field thickness measurements for morning vs. afternoon and afternoon vs. evening (with scatter points coloured accordingly). The first striking observation is the degree of separation which measurements have between pairwise time-point comparisons. The residuals of measurements comparing afternoon and evening are centred around 0, while residuals comparing morning and afternoon appear to be systematically below the origin (and the identity line in the correlation plot).

Figure 4.10(B) shows the longitudinal evolution of peripapillary choroid thickness for each peripapillary sub-field between morning, afternoon and evening, with the mean value per time point and sub-field shown as coloured crosses. Across the board, morning measurements appear larger than afternoon and evening. This corresponds to the trend observed in measuring the same choroids manually in the macula [16] (shown in their Supplementary Figure 8C), as well as the natural diurnal variation of the choroid observed in other

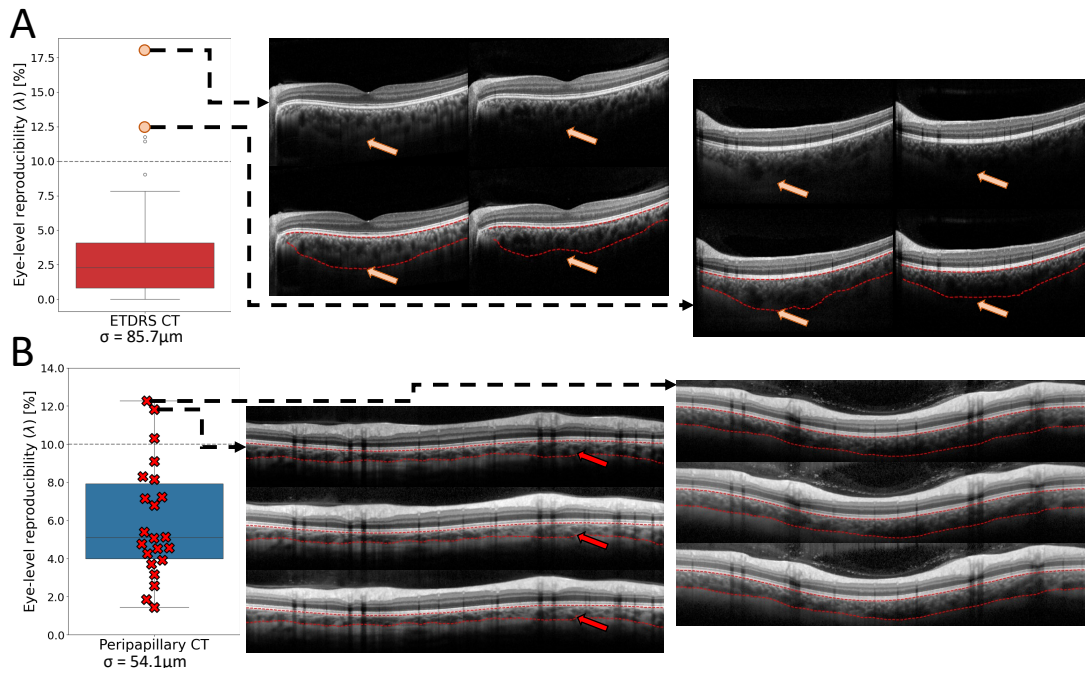


**Figure 4.10:** (A) Correlation and Bland-Altman plot for all chronologically paired (morning – afternoon, afternoon – evening) choroid thickness measurements in all peripapillary sub-fields. (B) Longitudinal evolution of average choroidal thickness in each peripapillary sub-field shown as box-plots, with mean values overlaid as crosses.

studies [52, 53, 54, 55, 56, 57]. These previous studies have reported choroidal fluctuations over the course of day (in the macula and peripapillary regions) with an amplitude of approximately  $30 \mu\text{m}$ , and the majority reporting larger choroidal variation in the morning, compared to the afternoon and evening. Thus, given that the largest residuals were those observed between morning and afternoon, and that we know the choroid thins between this time period, we propose that the diurnal variation of the choroid is the primary driver of the largest residuals in measuring peripapillary choroid thickness with DeepGPET, and not segmentation error.

#### 4.5.2.2 Eye-level

In figure 4.11, we present the eye-level reproducibility of DeepGPET for macular OCT volume data (A) and peripapillary OCT B-scan data (B). DeepGPET's eye-level measurement noise was very low compared to the population's variability, with the upper quartile of the box-plot distributions for ETDRS thickness sitting below 5% of the population's (i-Test) variability, and for peripapillary thickness this was below 10% of the population's (DVCKD) variability. Measurement error was higher for peripapillary thickness likely due to the lower population variability in the cohort relative to i-Test ( $\sigma=54.1 \mu\text{m}$  for DVCKD



**Figure 4.11:** Eye-level reproducibility of DeepGPET for ETDRS (A) and peripapillary (B) choroid thickness. We plot distributions of measurement noise  $\lambda$  [200], with a horizontal dashed-line marking 10% to aid readability and interpretation. Representative B-scans of major outliers are shown with segmentations overlaid in red and arrows indicating the source of the error. The between-eye standard deviations are shown below each box-plot.

sample which was 63% of the i-Test macular thickness population variability,  $\sigma=85.7 \mu\text{m}$ ). Nevertheless, for choroid region metrics segmentation error contributed to only 5–10% of the overall population variability in both cohorts.

In figure 4.11(A), we show two pairs of representative B-scans from the OCT volume scan data which were major outliers. The major outliers were primarily driven by large choroids whose non-EDI counterpart had poor Choroid-Sclera boundary visibility. This optical signal degradation had a significant impact on DeepGPET’s ability to detect this lower boundary, especially underneath the fovea. The arrows indicate the location of the major sources of error, which highlight the under-prediction of ETDRS thickness observed from the population-level reproducibility analysis. Nevertheless, DeepGPET is able to approximate the shape of the choroid well in regions of low visibility, albeit under-segmenting the choroid in regions of invisible Choroid-Sclera boundary.

In figure 4.11(B), we show the B-scans of the major outliers at each time point with the segmentations overlaid using DeepGPET. Any differences (red arrows) are minimal, and the segmentations do not look qualitatively different. These two cases were the major outliers due to the difference in choroid thickness observed from morning – afternoon (a negative differential of 15 and 14  $\mu\text{m}$  in average thickness, respectively). Comparing this to the difference observed between afternoon – evening, which was a positive differential of 1 and 2  $\mu\text{m}$ , we propose the major outliers are primarily driven by marginal diurnal variation of the choroid, and not segmentation error.

## 4.6 Discussion

### 4.6.1 Evaluation

DeepGPET is a fully-automatic and efficient method for choroid region segmentation. DeepGPET achieved excellent agreement with GPET on held-out data in terms of segmentation and choroid-derived measurements, approaching the repeatability of GPET itself, i.e. best case repeatability where the author (who developed GPET and arguably has the most domain knowledge around its usability) did both sets of annotation. Since the author generated both the original and repeated GPET segmentations for evaluating DeepGPET’s performance, there was no inter-rater subjectivity at play. Thus, the intra-rater agreement measured in this work is a best case scenario and forms an upper-bound for agreement with the original segmentations.

However, note that table 4.4 and figures 4.4 and 4.5 are comparing GPET’s *repeatability* with DeepGPET’s *generalisability*. Here, GPET is being measured against its own performance on the same 54 images it processed 2 months prior with manual input. Thus, GPET’s MAE of 10.41  $\mu\text{m}$  provides an upper bound for its segmentation performance when applied to the exact same input. Conversely, DeepGPET’s model evaluation shown in table 4.4 represents its generalisability to entirely new OCT data, which was slightly higher at 13.31  $\mu\text{m}$ . However, DeepGPET’s repeatability, i.e. application to the same image

twice is 100% accurate and will output the same prediction when given the same input. This is because hyperparameter-free deep learning models, post-training, like DeepGPET are *deterministic*.

In contrast, GPET has many tunable parameters and relies heavily on hand-crafted edge maps, manual input and image pre-processing. Holding all of GPET's free parameters and pre-processing steps fixed does make its output deterministic. However, the end-user is free to tune its parameters to fit what they believe is the true boundary — by selecting a different covariance function of the Gaussian process for example — necessarily injecting human subjectivity into the segmentation procedure. This subjectivity was ultimately quantified as the MAE of 10.41  $\mu\text{m}$  from GPET's repeatability analysis.

The improvement on segmentation inference time with DeepGPET over GPET results in massive time savings. A standard OCT volume scan collected on a Heidelberg Engineering imaging device usually consists of 61 B-scans. With GPET, processing such a volume for a single eye takes about 35 minutes during which the operator has to select initial pixels to guide tracing and adjust parameters if GPET initially failed (which may happen for about 25% of images given 1.27 manual interventions were required across 54 test images in section 4.4). In contrast, DeepGPET could do the same processing in about 76 seconds on the same hardware, during which no manual input is needed. DeepGPET could even be GPU-accelerated to cut the processing time by another order of magnitude.

The lack of manual interventions required by DeepGPET means that no subjectivity is introduced unlike GPET, particularly when used by different end-users. Importantly, DeepGPET does not require specialist training in image processing and can segment images fully-automatically in a fraction of the time, enabling analysis of large-scale datasets and greater potential for its applicability in future research studies. Thus, GPET can be replaced by a much quicker and more effective end-to-end segmentation pipeline which is significantly more accessible for those without a background in image processing (figure 4.12). We anticipate DeepGPET to facilitate efficient knowledge transfer of expertise in choroidal image analysis to the wider research community, ultimately helping improve the standardisation of segmentation-derived choroidal

measurements.

While the observed agreement was very high between DeepGPET and GPET, it was not perfect. However, even higher agreement with GPET would not necessarily produce a more generalisable method as GPET itself is not perfect, i.e. a perfect correspondence between DeepGPET and GPET would mean that DeepGPET learns the biases of GPET. For example, in figure 4.6(A) GPET fails to segment the entire width of the choroid, and in figure 4.6(B) under-segments the large choroid.

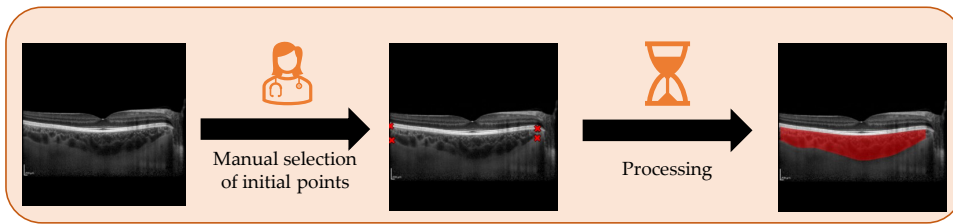
As described earlier (section 1.7), the definition of the Choroid-Sclera boundary can vary depending on the end-user and application [18]. This poses a huge challenge for manual segmentation of the choroid and standardisation of choroid-derived measurements [22, 23]. The variable definition, coupled with potentially low SNR, superficial retinal vessel shadowing, vessels passing from the sclera into the choroid and stroma fluid obscurations all contribute to the ambiguous nature of defining and detecting the Choroid-Sclera boundary [37].

However, for quantitative analysis of choroidal measures, the specific definition of the Choroid-Sclera boundary is secondary to applying the same, consistent definition across and within patients. Here, fully-automatic methods like DeepGPET provide a significant benefit by removing the subjectivity present in semi-automatic methods. Where semi-automatic methods require manual input, two analysts with different understandings of the Choroid-Sclera boundary could produce vastly different segmentations. With DeepGPET as a deterministic and automatic approach, the same image is always segmented in the same way, removing subjectivity.

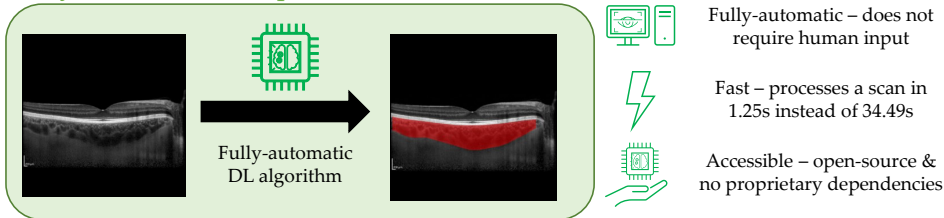
Finally, while DeepGPET was only trained on macular SD-OCT, we found the model to be robust to peripapillary SD-OCT data after suitable opposite-edge padding to ensure the whole width was segmented. This suggests DeepGPET's potential to generalise to different locations in the posterior pole, and the excellent reproducibility on the wider DVCKD cohort further provides evidence for this. Additionally, we observed that DeepGPET was able to successfully detect the Choroid-Sclera boundary in certain choroids of macular SS-OCT data which weren't too large. For large-choroid SS-OCT B-scans, DeepGPET



### Semi-Automatic GPET



### Fully-Automatic DeepGPET



**Figure 4.12:** Comparison between the semi-automatic GPET (top) and fully-automatic DeepGPET (bottom).

showed poor performance (figure 4.8). This is perhaps unsurprising given that DeepGPET was only trained on SD-OCT images. Thus, we recommend those using DeepGPET on SS-OCT images to review the segmentations afterwards as a sanity check.

## 4.6.2 Reproducibility

We found DeepGPET to be highly reproducible for measuring downstream choroidal measurements using macular SD-OCT volume and peripapillary B-scan data. DeepGPET reported MAEs well within thresholds expected to exceed manual, inter-rater agreement [106], GPET’s repeatability and small effect sizes observed from the literature [24, 110]. Moreover, DeepGPET achieved an eye-level measurement error which contributed to a very small proportion (5 – 10%) of the overall population variability in both cohorts. Advantageously, this eye-level reproducibility helps provides an upper bound on DeepGPET’s measurement variability which can be used to differentiate between measurement error and true biological change, a key step in the interpretation of choroidal measurements in research studies [235].

The reproducibility of DeepGPET for SD-OCT peripapillary B-scans were confounded by natural diurnal variation of the choroid. Analysis of residuals suggested that the major outliers were sourced primarily between morning – afternoon comparisons, rather than afternoon – evening comparisons, and major outliers at the eye-level showed no qualitative difference in segmentation error. Manual measurements of fovea-centred choroidal thickness showed a similar longitudinal trend [16] (supplementary figure 8C), as well as a wealth of literature suggesting choroidal thinning throughout daytime, and especially between the morning and afternoon [52, 53, 54, 55, 56, 57]. Thus, this suggests the error observed was primarily driven by diurnal variation and not segmentation error.

While the reproducibility was high, the values compared were averages across regions around the posterior pole, and so reproducibility for point-source measurements like subfoveal choroidal thickness has not been measured. Thus, there is potential for the values reported in each macular/peripapillary region to be averaging over local errors. However, the ETDRS and peripapillary subfields are commonly used by the research community and provide better ways to characterise the retina and choroid rather than extremely localised, point-source measurements. Therefore, assessing DeepGPET’s reproducibility according to these spatial regions is an entirely valid approach to investigate the method’s reproducibility.

However, we must make the reader aware of any potential data leakage in the reproducibility analysis. For the i-Test sample, approximately 7% (4 / 60) of women were part of the model training and validation for DeepGPET. However, only the EDI counterpart of their SD-OCT volume pair were used, and not their non-EDI volume scan. Thus, while the eye itself had been observed before by DeepGPET, the image artefacts introduced by deactivating EDI mode still posed a significant challenge to DeepGPET’s reproducibility as a method.

### 4.6.3 Limitations and Future work

There are some limitations associated with DeepGPET as a method. Firstly, DeepGPET was trained on OCT B-scans from examples related to systemic



health, and thus its generalisability to ocular health is unknown. It would be important to quantify the reliability of DeepGPET in extreme scenarios of pathology which affect the choroid, such as choroidal thinning from pathological myopia [255], thickening from central serous chorioretinopathy [256] or optical disturbance from photoreceptor degeneration in geographic atrophy [257]. Nonetheless, DeepGPET remains a highly reproducible approach for generating choroidal region metrics from OCT B-scans for the nascent field of ophthalmics [49], and we anticipate it will be a useful tool therein.

Additionally, by distilling GPET into DeepGPET, there is potential that it may have learned certain biases from GPET's semi-automatic procedure. For example, GPET does not always segment the whole width of the choroid, as it's dependent on the user's manual edge endpoint selection. Currently, area and thickness are only calculated for a specific fovea-centred ROI, so this is not a problem. However, segmenting the whole visible choroid would be desirable when generating maps across the macula from an OCT volume scan. Interestingly, DeepGPET is already capable of segmenting the whole choroid in most cases, i.e. figure 4.6(A), but as some of the training data did not cover the full width, DeepGPET sometimes emulates this, too.

Finally, and perhaps most importantly, DeepGPET is only focused on segmentation of the choroidal layer. While this is the most important and time-consuming step for measuring choroidal thickness, area and volume, the location of the fovea on OCT B-scans is required to define the fovea-centred ROI for choroid-derived measurements. This is critical for producing standardised measurements across and within patients. Given that identifying the fovea is less time-consuming or ambiguous than choroid segmentation, extending DeepGPET to output the fovea location is a natural next step.

Fortunately, DeepGPET's robustness to peripapillary B-scans permit fully automatic, segmentation-derived measurements of peripapillary choroids, as these B-scans do not require fovea detection for derived measurements. However, while this is the case for end-users who are reasonably familiar with the Python programming language, this may not be the case to all members of the research community. Therefore, it would be wise to wrap DeepGPET into a truly automatic 'raw data to measurement' toolkit which end-users can run simply

from the terminal on OCT data extracted from an imaging device. In future work, we plan to release such a toolkit, OCTolyzer, which would leverage DeepGPET's ability to provide reproducible and clinically meaningful measurements of choroids in peripapillary B-scans.

Another limitation of DeepGPET is that it also does not quantify the choroidal vasculature, which would arguably be a more representative measure of the choroid (figure 3.1). Segmenting the choroidal vessels is a very challenging task even for humans and is prohibitively time-consuming to do manually (table 3.3). Permitting additional functionality for DeepGPET to automatically segment the vasculature within the choroid as well would make DeepGPET a fast and efficient end-to-end framework capable of fully segmenting the choroidal space and vessels from a raw OCT B-scan, converting the B-scan into a set of standardised, clinically meaningful and reproducible measurements. Fortunately, the semi-automatic approach defined in chapter 3, MMCQ, could be used for ground truth label generation.

DeepGPET was the first foray into deep learning during this doctorate degree, and while deep learning in general remains as the current state-of-the-art for medical image segmentation [129, 240], there are a number of drawbacks. For example, traditional image processing methods like GPET and MMCQ did not require any prior training data, while model training in deep learning requires expertly annotated datasets which can be scarce in the medical imaging domain. Although techniques like transfer learning and data augmentation can mitigate this issue [241], the reliance on large datasets remains a fundamental limitation.

Furthermore, model training usually requires GPU resources which may not necessarily be at everyone's disposal, and while DeepGPET is able to be run on a standard laptop CPU reasonably fast (at 1.25 seconds/image), much larger models such as vision transformers [258] or parameter-intense architectures like DenseNet [140] require GPU-accelerated model training and inference to be run effectively, utilising large amounts of temporary memory (RAM). This complexity can make it difficult to run on edge compute devices, with limited access to high-performance compute resources to deploy deep learning-based segmentation models. Nevertheless, we anticipate that DeepGPET will

be able to cover a broad range of end-users with different levels of hardware.

Additionally, DeepGPET, unlike traditional image processing methods like GPET and MMCQ, is seen as a “black box” where the hierarchical nature of DeepGPET’s CNN makes it difficult for humans to interpret the decision making during inference. This is because each layer deeper into the model abstracts away the tangible relationship we end-users can attribute to the groups of pixels in the original input image. In critical applications like medical imaging, where understanding the reasoning behind a model’s decision is essential, this lack of transparency can be a significant drawback, especially when compared to traditional image processing methods which have a clear mechanistic procedure which is both interpretable and explainable, like GPET and MMCQ. However, segmentation models like DeepGPET are still more interpretable than deep learning-based disease detection models, since DeepGPET outputs a visual result which can be critiqued, unlike the output of, say, a continuous risk score for disease.

Finally, while fine-tuning from large natural image datasets such as ImageNet remains at the forefront of state-of-the-art deep learning model training [241], foundational models are making traction in the field of medical image segmentation [259, 260]. Pre-trained models from datasets like ImageNet use supervised learning to train models on millions of natural images for classification. This ultimately means the features learned by the model are attributed to natural images. Conversely, foundational models take a self-supervised approach to model training, learning feature representations without distinct labels and, compared to ImageNet-based model pre-training, are more task- and domain-specific to medical images [259] and indeed retinal images [260]. These models can utilise large label-free datasets and have previously been shown to outperform similar models following supervised learning [261].

Zhou, et al. [260] found that their foundational model, RETFound, already showed superior performance over traditional ImageNet-based supervised (and self-supervised) learning for the detection of ocular and systemic diseases using OCT images [260]. With freely available models pre-trained on retinal images [260], we suspect that foundational models will undoubtedly improve the performance of future OCT choroid segmentation models. Additionally, while

these models are traditionally thought to require significant compute resources to train, recent evidence suggests that this cost can be significantly reduced without harming performance and generalisation [262]. Thus, this approach may be worthwhile for OCT image segmentation problems.

#### 4.6.4 Outputs

In this chapter, there have been two publication outputs, one of which has been published and peer-reviewed (by December 2024) and another which is currently in peer-review but has been uploaded as a pre-print on Arxiv. The author's name is in bold type, with the names of any co-lead authors underlined.

- **Burke, Jamie**, Justin Engelmann, Charlene Hamid, Megan Reid-Schachter, Tom Pearson, Dan Pugh, Neeraj Dhaun et al. “*An Open-Source Deep Learning Algorithm for Efficient and Fully Automatic Analysis of the Choroid in Optical Coherence Tomography.*” *Translational Vision Science & Technology* 12, no. 11 (2023): 27-27.
- **Burke, Jamie**, Justin Engelmann, Charlene Hamid, Diana Moukaddem, Dan Pugh, Neeraj Dhaun, Niall Strang et al. “*OCTolyzer: Fully automatic analysis toolkit for segmentation and feature extracting in optical coherence tomography and scanning laser ophthalmoscopy data.*” arXiv preprint arXiv:2407.14128 (2024). Submitted and in peer-review at Elsevier's Medical Image Analysis

The software associated with this method, DeepGPET, has been published as open-source and is freely available on GitHub [here](#). Similarly, OCTolyzer is freely available on GitHub [here](#).

#### 4.6.5 Executive summary

In this chapter, we distilled the previous semi-automatic approach to choroid region segmentation, GPET, into a fully automatic deep learning-based method, DeepGPET. DeepGPET removes the need for manual segmentation of the

choroidal space, and prevents any human subjectivity from being injected into the segmentation procedure. DeepGPET performed well on unseen test data related to systemic health, and was preferred by a clinical ophthalmologist when compared with its predecessor, GPET, on various examples of disagreement. We also found that DeepGPET was reproducible across SD-OCT macular data and, interestingly, peripapillary data. This was observed both at the population- and eye-level, with clinically insignificant errors. Importantly, the reproducibility analysis conducted provides an interpretable measure of DeepGPET's measurement variability, enabling the research community to differentiate between measurement error and true biological change in future studies. While DeepGPET was a marked improvement over GPET, it was only trained on examples related to systemic disease and thus has not been evaluated against abnormal retinæ. Moreover, while initial experimentation showed DeepGPET's robustness to macular, SS-OCT with good choroid visualisation, this was not observed for more challenging cases. Importantly, DeepGPET was unable to automate computation of segmentation-derived measurements as the fovea required detection — typically manually — thus injecting human subjectivity into the measurement procedure. Future work would look to validate DeepGPET on examples related to retinal pathology, extend the model to be generalisable to SS-OCT data and automate fovea detection to permit end-to-end measurement of the choroidal space from raw OCT B-scans.

# Chapter 5

## Choroidalyzer: End-to-end choroidal analysis in optical coherence tomography

### 5.1 Introduction

In this work we have so far considered solutions for choroid region and vessel segmentation separately (choroid region: GPET, DeepGPET and choroid vessel: MMCQ). While these approaches are sufficient for their own use-case, they do not leverage information shared between both segmentation tasks, given that the choroidal vasculature is interspersed within the choroidal space. In an optimal setting, a unified approach to choroid analysis through solving both tasks simultaneously in an end-to-end framework would be desirable (and significantly more practical), thus utilising information from both the choroidal space and vasculature during model inference.

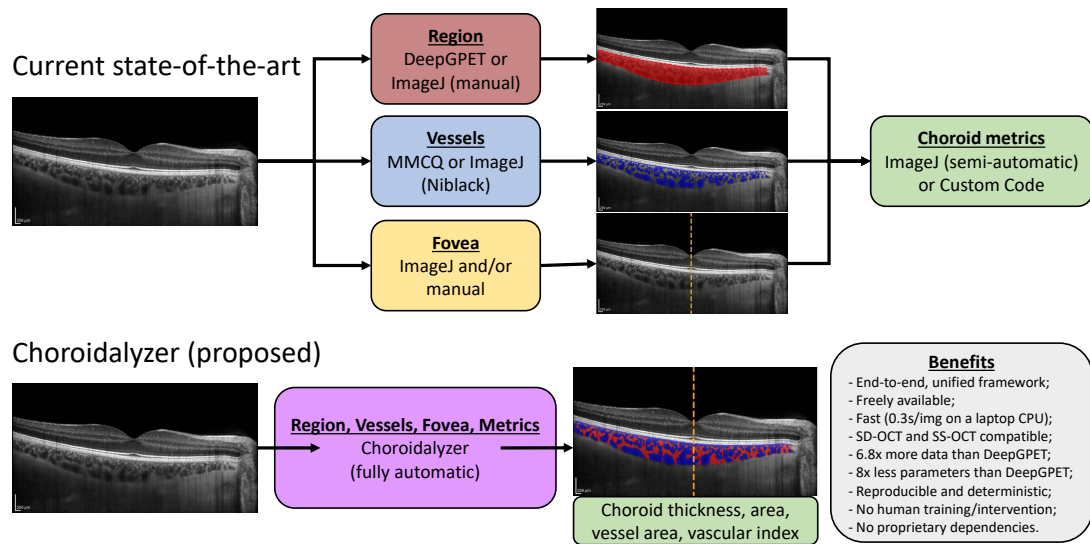
Furthermore, in addition to accurate and reliable segmentation of the choroidal region and vessels, there are two more necessary steps that are often overlooked, namely fovea detection and computation of choroidal metrics. Choroid-derived measurements require careful consideration of the ROI used for analysis, typically using the location of the fovea as a reference guide. This is to

ensure measurements are standardised for proper comparison across different studies, populations and imaging devices. Moreover, OCT B-scans are not necessarily perfectly centred, with the chorioretinal structure presented at a skew or appearing curved with respect to the image axis, which is often ignored while deriving measurements [13, 18, 21, 64, 106, 112]. We know that this can have an inordinate impact on downstream measurements given the 1:3 ratio between axial and lateral pixel length-scales in commercial OCT imaging devices (section 1.7). Thus, once the region, vessels, and fovea are extracted, choroidal metrics should be computed in a fovea-centred ROI which must account for key details like the skew or curvature of the chorioretinal structure and the pixel length-scales of each scan.

While DeepGPET and MMCQ are able to provide fast, accessible, reproducible and clinically meaningful solutions to choroid region and vessel segmentation, respectively, they ultimately only solve a fraction of the problem for choroidal analysis in OCT image sequences.

As discussed in section 1.6, the application of deep learning based algorithms to the problem of choroidal segmentation in OCT image sequences is not new (reviews: Eghtedar, et al. [263] and Selvam, et al. [264]). However, in the wider research community the steps required for measuring the choroid and locating the fovea are commonly performed using separate tools for region (table 1.1) and vessel segmentation (table 1.2), often using non-standardised manual approaches for fovea detection or ROI definition.

While recently there has been a push for synthesising the two segmentation tasks into a single model (table 1.3), there are still inherent issues in these methods. For example, most are either closed-source or are not packaged well for end-user accessibility to the research community [17, 167]. Only a small handful of these approaches report their reproducibility [17, 166], and some still consider each task as separate ones but package them together [17, 166, 167], depending on semi-automatic Niblack for vessel segmentation. In the case for using Niblack, only two studies reports the parameters they use [163, 167], making it impossible to reproduce the results of others in confidence. For those which do provide end-to-end region and vessel analysis using deep learning, their ground truths were generated manually [168, 169, 170, 172],



**Figure 5.1:** A comparison between the current state-of-the-art (top) — defined here as what’s most widespread and available to the research community — in choroidal image analysis in OCT image sequences (top), and the proposed method, Choroidalizer (bottom). To obtain choroidal metrics in a fovea-centred ROI, researchers currently need to combine many different tools. Choroidalizer unifies everything into an end-to-end pipeline that is very fast and convenient to use.

which can be subject to human bias [189]. Importantly, very few consider the importance of automatically measuring fovea-centred, choroid-derived measurements through detecting the fovea on the B-scan [17]. All of these aforementioned reasons contribute to the current lack of standardisation of choroid-derived measurements in OCT image sequences.

Thus, we address these current issues by proposing Choroidalizer, an end-to-end pipeline for choroidal analysis. Choroidalizer consists of a single deep learning model that simultaneously segments the choroidal region and vessels and detects the fovea location, combined with all the code needed to extract choroidal thickness, area, and CVI in a fovea-centred ROI. Figure 5.1 shows how Choroidalizer improves on the current state-of-the-art by providing a comprehensive solution for all elements of choroidal analysis. To our knowledge, Choroidalizer is the first open-source tool for fully automatic, end-to-end choroidal analysis which reports interpretable reproducibility metrics, and is developed using a large dataset from 6 datasets related to systemic health,



utilising both SD-OCT and SS-OCT data from two market-leading manufacturers for OCT imaging systems (Heidelberg Engineering, Topcon).

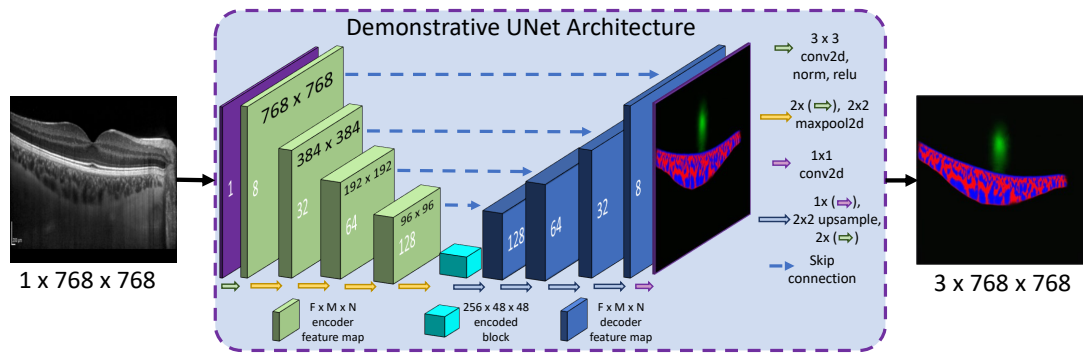
### 5.1.1 The UNet Deep Learning Model

We briefly discussed deep learning for image segmentation and the UNet architecture in section 4.1.1, focusing on the general advantages and disadvantages which deep learning methods provide for medical image segmentation over traditional image processing methods. DeepGPET was a pre-trained UNet architecture which was trained from scratch on ImageNet [238], while in this chapter, we develop a custom UNet deep learning architecture which Choroidalyzer leverages to segment the choroidal space, vessels and simultaneously detect the fovea on OCT B-scans.

The UNet architecture has become a cornerstone in deep learning for medical image segmentation due to several key factors that make it particularly well-suited for this domain. Consequently, the UNet architecture has been successfully applied across multiple medical imaging modalities, such as in MRI [265], CT [266], X-ray [267], ultrasound [268] and indeed OCT [269].

Medical image segmentation requires pixel-level precision to accurately delineate anatomical structures. The UNet deep learning architecture is a type of CNN for pixel-level semantic image segmentation, and excels in this regard due to its design, which captures both the global context and fine-grained details within images. The CNN architecture is constructed in a U-shape with two distinctive arms, the *encoder* and *decoder* arms. The encoder (contracting) arm takes the input image and sequentially applies blocks of convolutional filters, learning low- to high-level descriptor features at every progressive block, and simultaneously pooling these features into smaller spatial dimensions to collect higher-level, global contextual features. After encoding, what is left is a number of coarse grids (feature maps) representing abstract, space-invariant and global features of interest from the input — note that these are large in the number of feature map, but small in spatial dimension [138].

The decoder (expanding) arm uses upsampling (via transposed convolutions



**Figure 5.2:** Demonstrative UNet deep learning architecture which is used in this chapter, taking an input OCT B-scan of resolution  $1 \times 768 \times 768$  and outputting a probability segmentation map for three segmentation tasks, choroid region, vessel and fovea detection with resolution  $3 \times 768 \times 768$ .

or bi-linear interpolation) to increase the spatial dimensions while reducing the number of feature maps in a symmetrical and sequential manner to the encoder arm. This has the effect of decoding the global representations learned by the encoder into local ones, eventually outputting a pixel-level probability map of the model's belief at the original input pixel resolution. The symmetry is leveraged by combining (simply through addition) the upsampled features with corresponding high-resolution features from the encoder arm (through skip connections), enabling precise localisation and facilitating the output of high-fidelity segmentation maps. The additional U-shaped architecture is a significant improvement over traditional, fully convolutional networks which only have a contracting (encoder) path [138].

The inclusion of skip connections is one of UNet's most distinctive features. These connections directly link the encoder layers to the corresponding decoder layers, allowing the model to bypass certain levels of abstraction and preserve spatial information from earlier layers. In medical image segmentation fine details, such as the boundaries between tissues like the choroid extravascular and vascular compartments, need to be accurately segmented and these skip connections are vital in facilitating this. They help the model retain high-resolution features that might otherwise be lost, leading to more accurate and detailed segmentations. Thus, this architecture ensures that important spatial details are preserved, which is crucial for the high accuracy required in

medical applications [138].

Figure 5.2 shows a detailed and demonstrative diagram of the UNet deep learning architecture for the method proposed in this chapter. This UNet architecture presented here is a deep learning model with a depth of 4, i.e. has four distinct encoder (green) and decoder (blue) blocks. Note the symmetrical nature of the architecture, where each encoder block has a corresponding decoder block at the same spatial resolution, which allows skip connections (blue dashed lines). The black text denotes the spatial resolution of the encoder and decoder feature maps, which are scaled by  $1/2$  and 2 as you move between encoder blocks and decoder blocks, respectively. As the spatial resolution changes in one direction, the number of feature maps (white text) changes in the opposite direction, i.e. spatial maps become coarser and increase in quantity with depth into the encoder arm, as seen by the final encoded (cyan) block with 256 feature maps at a spatial resolution of  $48 \times 48$ .

In practice, scaling the feature maps by a higher magnitude would be more optimal in order to reduce memory and improve training/inference time. Additionally, having a higher depth (more convolutional blocks in the encoder and decoder arms) would likely lead to improved performance by having a very coarse, final encoded block to maximise spatial invariance and global representation learning [270]. This is entirely to do with the receptive field of each neuron in the network [270], which is not within the scope of this thesis.

## 5.2 Study populations

### 5.2.1 Model training and evaluation

The dataset used for Choroidalalyzer's training and evaluation contains 5,600 OCT B-scans of 233 participants from 6 cohorts of healthy and diseased individuals (unrelated to ocular pathology):

- **OCTANE** [225], a longitudinal cohort ( $N=46$ ) study investigating choroidal changes in renal transplant recipients and donors. Note that this cohort is

one participant fewer and 3 H-line scans greater than DeepGPET's OCTANE cohort from table 4.1. This is because of an administrative error during data curation and does not impact the rest of the analysis;

- **Diurnal Variation for Chronic Kidney Disease (DVCKD)** [225], a cohort ( $N=20$ ) of young participants collected as part of the OCTANE study (but mutually exclusive to the longitudinal data above), for assessing diurnal variation in the choroid of healthy volunteers to support research related to chronic kidney disease [16];
- **Normative**, a detailed OCT examination of the author of this thesis with informed consent ( $N=1$ );
- **i-Test** [225], a cohort ( $N=21$ ) of pregnant women evaluating retinochoroidal changes in normative, pre-eclamptic or fetal growth restricted pregnancies;
- **Prevent Dementia** [271], a mid-life cohort ( $N=121$ ), approximately half of whom have risk of developing later life Alzheimer's disease;
- **Glasgow Caledonian University Topcon (GCU Topcon)** [234], a cohort ( $N=24$ ) of healthy participants investigating diurnal variation of the choroid in individuals of varying refractive status.

All studies adhered to the Declaration of Helsinki and received relevant ethical approval, and informed consent from all subjects was obtained in all cases from the host institution.

Three OCT device types were used from two device manufacturers: the SD-OCT Spectralis Standard and FLEX Modules (both OCT1 systems from Heidelberg engineering, Heidelberg, Germany), and the SS-OCT DRI Triton Plus (Topcon, Tokyo, Japan).

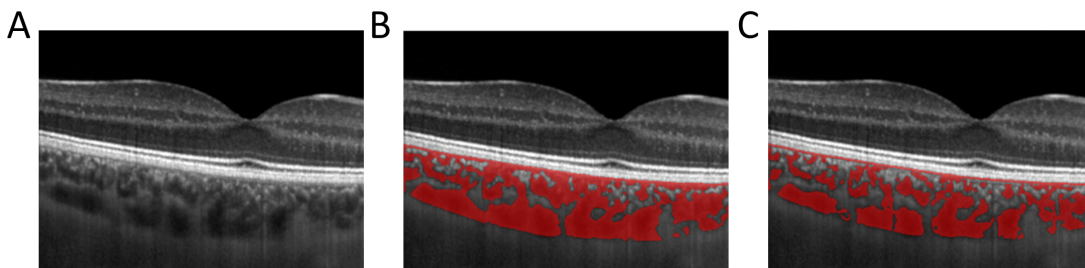
The majority of the Heidelberg data acquisition has been described previously in section 4.2.1. Briefly, fovea-centred horizontal- and vertical-line scans were collected with active eye tracking using an ART of 100, covering a 30-degree FOV. Posterior pole volume scans covered a  $30 \times 20$  degree rectangular ROI using 31 consecutive scans with an ART of 50 for the i-Test cohort and 61

line scans with an ART of 9 for the OCTANE and Normative cohorts. All Heidelberg data had signal quality  $\geq 15$  and was collected at a pixel resolution of  $768 \times 768$  pixels, except for i-Test which had a native resolution of  $496 \times 768$ , which was then padded vertically to  $768 \times 768$ . Additionally, we collected peripapillary OCT B-scans in the DVCKD cohort, whose acquisition has been previously described in section 4.2.2. Briefly, EDI-OCT circular B-scans centred at the optic nerve head were collected for each participant’s right eye at three times on the same day, using active eye tracking with an ART of 100.

For the Topcon device, during acquisition, a twelve-line, fovea-centred OCT radial scan was captured, starting from a horizontal-line B-scan and rotating clockwise in 30 degree intervals. Each B-scan covered approximately a 9 mm FOV laterally and had a resolution of  $992 \times 1024$  pixels which was cropped horizontally by 32 pixels and resized to the resolution of the Heidelberg scans of  $768 \times 768$ . Any set of 12 B-scans with an average image quality score less than 88 — reported by the SNR index built-in to the Topcon scanning device (Topcon, Tokyo, Japan) — was excluded.

### 5.2.1.1 Ground truth labels

The choroidal region in an OCT B-scan has been defined previously in section 1.7, as well as the suprachoroidal space. Because of the inconsistent appearance of the suprachoroidal space [18, 37], we consider this space not to be part of the choroid itself, which ensures that Choroidalyzer learns a con-



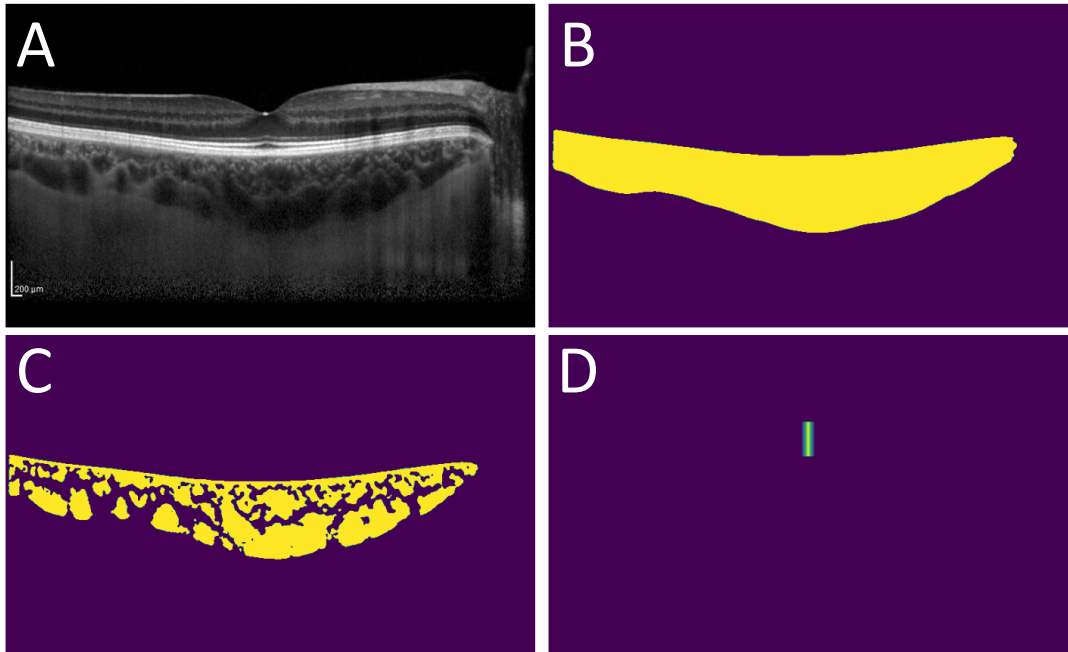
**Figure 5.3:** A comparison between applying MMCQ with and without the majority voting scheme. (A) OCT B-scan snippet, (B) default MMCQ application, (C) majority voting scheme MMCQ application.

sistent definition for the choroidal space. Ground truth labels for the choroidal space were generated using DeepGPET after thresholding the resulting binary segmentation maps with a value of 0.5. The choroidal vasculature in an OCT B-scan has already been defined in section 1.7, and the ground truth labels were generated using MMCQ.

Figure 5.3 shows a snippet from an OCT B-scan (A), with the original vessel segmentation from MMCQ (B) and after applying the majority voting scheme based on varying contrast and brightness adjustments (C). To improve the fidelity and robustness of our vessel segmentation labels, we randomly varied the brightness and contrast of each OCT B-scan before application of MMCQ. We used 5 linearly spaced gamma levels to fix the mean brightness of each image between 0.2 and 0.5 and simultaneously altered the contrast using 5 linearly spaced factors between 0.5 and 3. A 3:2 majority vote for vessel label classification was used across all 25 variants. Robustness is achieved by averaging out any erroneous over- or under-segmented pixels produced from each brightness and contrast combination (which produce a different set of image statistics and noise distribution).

The foveal pit has been defined previously in section 1.7. For a fovea-centred OCT B-scan, the lateral position of the foveal pit was defined as the horizontal (column) pixel index which aligned with the deepest point of the foveal pit depression [17]. For OCT B-scans centred at the fovea (i.e. horizontal, vertical and radial scans), the foveal column location was detected manually using a custom-built graphical user interface in Python (version 3.11). Those not centred at the fovea had their fovea label set as the origin  $(0, 0)$  ( $N=2178$ ). The fovea is only a single point which would be difficult for a segmentation model to learn, as predicting close to 0 for all pixels would yield virtually the same loss as a perfect prediction. Thus, we create a target 51 pixels high and 19 pixels wide centred at the fovea location. The exact fovea location is set to 1, the column of the fovea location to 0.95 for 25 pixels in both vertical directions, and adjacent columns to  $0.95-0.1d$ , where  $d$  is the column distance from the fovea, for  $d = 1, \dots, 9$ .

Figure 5.4 shows the region, vessel and fovea ground truth segmentation labels for an OCT B-scan.



**Figure 5.4:** Ground truth segmentation labels for an OCT B-scan (A), showing region label (B), vessel label (C) and fovea label (D).

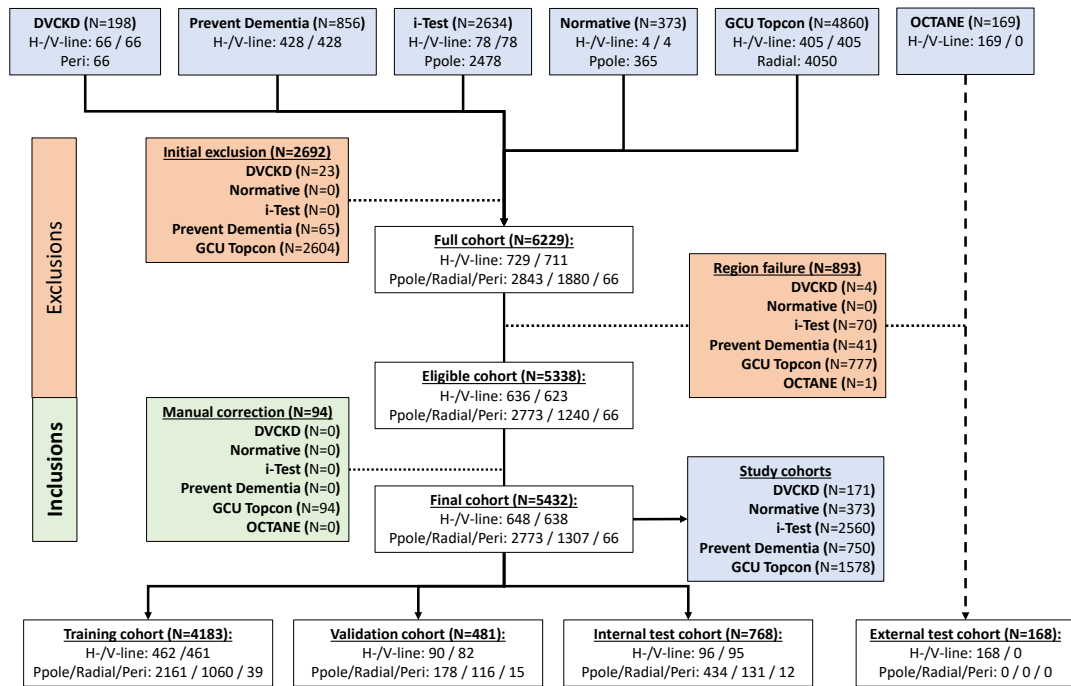
### 5.2.1.2 Sample derivation

Figure 5.5 illustrates the sample derivation for the dataset used to train and evaluate Choroidalalyzer. A total of 9090 OCT B-scans were available from all six cohorts. Before ground truth labels were generated, an initial exclusion criteria was set as 1) Choroid-Sclera boundary was invisible, 2) the choroid was clipped or cropped and 3) initial experimentation with DeepGPET in SS-OCT resulted in segmentation failure. During data extraction of the six cohorts, 23 and 65 B-scans from the DVCKD and Prevent Dementia cohorts, respectively, were excluded due to invisible Choroid-Sclera boundary and choroid clipping/cropping.

DeepGPET’s generalisability on SS-OCT data was tested on a small sample of SS-OCT data from the GCU Topcon cohort (see section 4.4.2.3). Through randomly sampling 6 OCT B-scans per eye, per participant, manual adjudication (from author J.B.) found that DeepGPET failed to segment the choroids from 13 out of 24 participants. Failure was defined as significant under-/over-



## 5.2. Study populations



**Figure 5.5:** Sample derivation flowchart of how OCT data from six cohorts were combined to create Choroidalalyzer's modelling dataset.

segmentation of the choroidal space (figure 4.8, bottom). All data from these participants were excluded from Choroidalalyzer's datasets ( $N=2604$ ). Thus, initial exclusion saw a total of  $N=2692$  B-scans removed, leaving  $N=6229$  B-scans to be processed by DeepGPET and MMCQ for choroidal space and vessel segmentation.

Ground truth labels for the choroidal space were reviewed by author J.B. and  $N=893$  B-scans were excluded because of segmentation failure from DeepGPET. The majority of these failures were from large choroids in the GCU Topcon cohort ( $N=777$ ). The remaining exclusions ( $N=116$ ) from the five other cohorts were also from large choroids with poor Choroid-Sclera boundary visibility. This left  $N=5506$  B-scans to train and evaluate Choroidalalyzer with.

DeepGPET's lack of generalisability to SS-OCT choroids of larger size resulted in an inordinately large exclusion rate of GCU Topcon data. Since the GCU Topcon cohort primarily targeted hyperopes during recruitment, many of the choroids were thus large and grainy in appearance at the point of the Choroid-



Sclera boundary due to natural optical signal degradation and higher speckle noise at depth compared to EDI-OCT data (figure 1.7), of which DeepGPET was primarily trained. Some examples of these large choroids can be found in the bottom row of figure 4.8 discussed in chapter 4 and also in figure 3.31 discussed in chapter 3. While disappointing to see such a large number of SS-OCT data excluded from Choroidalyzer’s model development, it also represents an obvious motivation for developing Choroidalyzer.

Therefore, in order for Choroidalyzer to generalise well to large SS-OCT choroids, author J.B. manually segmented 94 random OCT B-scans from the 13 participants initially excluded from the GCU Topcon cohort, and this was added to the overall dataset. Note that the initial expectation was to manually segment 100 B-scans, but a simple administrative error on behalf of the author lead to only 94 B-scans being segmented. Manual segmentation was performed semi-automatically using a custom-built graphical user interface in Python (version 3.11). For each of the RPE-Choroid and Choroid-Sclera boundaries, the end-user selected points along the boundary and these were interpolated using a univariate spline whose polynomial degree was 3.

Thus, Choroidalyzer’s dataset for training and evaluation consists of  $N=5600$  B-scans. Table 5.1 gives a detailed overview of the population and image characteristics of the final dataset used to develop Choroidalyzer, comprising 648 horizontal-line (H-line), 638 vertical-line (V-line), 2,773 posterior pole volume (Ppole), 1,307 radial and 66 peripapillary (Peri) scans. Notably, 15% of the

|                             | DVCKD [225] | Normative     | i-Test [225]   | Prevent Dementia [271] | GCU Topcon [234] | OCTANE [225] | Total               |
|-----------------------------|-------------|---------------|----------------|------------------------|------------------|--------------|---------------------|
| Device manufacturer         | Heidelberg  | Heidelberg    | Heidelberg     | Heidelberg             | Topcon           | Heidelberg   | All                 |
| Device type                 | Spectralis  | FLEX          | FLEX           | Spectralis             | DRI Triton Plus  | Spectralis   | All                 |
| Subjects, N                 | 20          | 1             | 21             | 121                    | 24               | 46           | 233                 |
| Control/Case                | 20 / 0      | 1 / 0         | 11 / 10        | 56 / 65                | 24 / 0           | 0 / 46       | 112 / 121           |
| Male/Female                 | 11 / 9      | 1 / 0         | 0 / 21         | 66 / 55                | 14 / 9*          | 24 / 22      | 116 / 116*          |
| Right/Left eyes             | 20 / 0      | 1 / 1         | 21 / 21        | 117 / 115              | 22 / 21          | 46 / 0       | 227 / 158           |
| Age (mean (SD))             | 21.4 (2.3)  | 23.0 (0.0)    | 32.8 (5.4)     | 50.8 (5.6)             | 21.8 (7.9)       | 42.9 (13.7)  | 47.5 (12.3)         |
| B-scans                     |             |               |                |                        |                  |              |                     |
| EDI / non-EDI / SS-OCT      | 171 / 0 / 0 | 247 / 126 / 0 | 2438 / 122 / 0 | 394 / 356 / 0          | 0 / 0 / 1,578    | 168 / 0 / 0  | 3,418 / 604 / 1,578 |
| Horizontal/Vertical         | 55 / 50     | 4 / 4         | 76 / 76        | 381 / 369              | 132 / 139        | 168 / 0      | 816 / 638           |
| Volume/Radial/Peripapillary | 0 / 0 / 66  | 365 / 0 / 0   | 2,408 / 0 / 0  | 0 / 0 / 0              | 0 / 1,307 / 0    | 0 / 0 / 0    | 2,773 / 1,307 / 66  |
| Total                       | 171         | 373           | 2,560          | 750                    | 1,578            | 168          | 5,600               |

**Table 5.1:** Overview of population characteristics. \*: One participant’s sex from the GCU Topcon cohort was not recorded.

## 5.2. Study populations

|                                   | Training            | Validation      | Testing         | External test | Total                 |
|-----------------------------------|---------------------|-----------------|-----------------|---------------|-----------------------|
| Subjects                          | 122                 | 28              | 37              | 46            | 233                   |
| Male/Female                       | 64 / 57*            | 12 / 16         | 16 / 21         | 24 / 22       | 116 / 116*            |
| Control/Case                      | 76 / 46             | 16 / 12         | 20 / 17         | 0 / 46        | 112 / 121             |
| Right/Left eyes                   | 117 / 107           | 27 / 23         | 37 / 28         | 46 / 0        | 227 / 158             |
| Standard/FLEX/DRI Triton Plus     | 88 / 14 / 20        | 24 / 2 / 2      | 29 / 6 / 2      | 46 / 0 / 0    | 187 / 22 / 24         |
| Heidelberg/Topcon                 | 102 / 20            | 26 / 2          | 35 / 2          | 46 / 0        | 209 / 24              |
| Age (mean (SD))                   | 40.7 (14.2)         | 42.5 (11.9)     | 44.5 (13.4)     | 47.5 (12.3)   | 42.9 (13.4)           |
| Cohort                            |                     |                 |                 |               |                       |
| OCTANE [225]                      | 0                   | 0               | 0               | 46            | 46                    |
| DVCKD [225]                       | 12                  | 4               | 4               | 0             | 20                    |
| Normative                         | 1                   | 0               | 0               | 0             | 1                     |
| i-Test [225]                      | 13                  | 2               | 6               | 0             | 21                    |
| Prevent Dementia [271]            | 76                  | 20              | 25              | 0             | 121                   |
| GCU Topcon [234]                  | 20                  | 2               | 2               | 0             | 24                    |
| B-scans                           |                     |                 |                 |               |                       |
| Spectralis/Flex/DRI Triton Plus   | 582 / 2,281 / 1,281 | 136 / 190 / 140 | 137 / 462 / 157 | 168 / 0 / 0   | 1,023 / 2,933 / 1,578 |
| Heidelberg/Topcon                 | 2,863 / 1,281       | 326 / 140       | 599 / 157       | 168 / 0       | 3,956 / 1,578         |
| Horizontal/Vertical scans         | 462 / 461           | 90 / 82         | 96 / 95         | 168 / 0       | 816 / 638             |
| Volume/Radial/Peripapillary scans | 2,161 / 1,060 / 39  | 178 / 116 / 15  | 434 / 131 / 12  | 0 / 0 / 0     | 2,773 / 1,307 / 66    |
| Total B-scans                     | 4,183               | 481             | 768             | 168           | 5,600                 |

**Table 5.2:** Overview of population and image characteristics of the internal training, validation and test sets, and also the external test set. \*: One participant's sex from the GCU Topcon cohort was not recorded.

Heidelberg SD-OCT data ( $N=4022$ ) was captured without EDI mode ( $N=604$ ).

Approximately 10% of the entire dataset was randomly sampled (495 B-scans) and stratified to represent different eyes and locations on the macula. This subsample was used to assess Choroidalyzer's fovea detection accuracy.

Five of the six cohorts were split into training (4,183 B-scans, 122 subjects), validation (481 B-scans, 28 subjects) and *internal* test sets (768 B-scans, 37 subjects) containing approximately 75, 10, and 15% of the B-scans. We split the data at the subject-level, such that no individual ended up in more than one set to prevent data leakage during training and evaluation. The remaining cohort, OCTANE, was entirely held-out as an *external* test set (168 B-scans, 46 subjects). Table 5.2 presents the population and image characteristics for each of the four sets and describes their population and image acquisition statistics.

External validation is helpful for testing the generalisability of a model using out-of-domain (out-of-distribution) data, i.e. OCT images which are from a source external to the internal set [272]. In this particular case, the external test set here indicates how well Choroidalyzer might generalise to an entirely different cohort and disease context (CKD).

## 5.2.2 Reproducibility

|                                | Study   |   |
|--------------------------------|---|---|
|                                | i-Test [225]  | GCU Topcon [234]                                |
| <b>Cohort demographics</b>     |   |   |
| Participants (Eyes)            | 60 (120)  | 21 (33)   |
| Right eyes (%)                 | 60 (50)   | 15 (45.5)                                       |
| Age (SD)                       | 34.7 (5.2)  | 23.9 (4.2)                                      |
| Sex, F (%)                     | 60 (100)  | 9 (43)  |
| Ethnicity                      | 52 White, 6 Asian, 2 Mixed                            | 12 White, 5 Asian, 2 Black, 2 Middle Eastern    |
| Gestation                      | 35.6 (3.4)  | NA  |
| Refractive status <sup>†</sup> | 3 hyperopes, 31 emmetropes, 26 myopes                 | 9 hyperopes, 8 emmetropes, 3 myopes, 1 missing  |
| Study purpose                  | Pre-eclamptic / normative pregnancy at late gestation | Diurnal variation                               |
| Control/Case                   | 45/15   | 21/0  |
| <b>Image characteristics</b>   |   |   |
| Device                         | Spectralis (Heidelberg)                               | DRI Triton Plus (Topcon)                        |
| OCT Type                       | Spectral-domain                                       | Swept-source                                    |
| Scan Pattern                   | Macular volume  | Macular radial                                  |
| Mode                           | HRA+OCT   | CFP+OCT   |
| Time of day (Interval)         | All in afternoon (1 minute)                           | 13 morning, 12 afternoon, 8 evening (5 minutes) |
| B-scans per eye                | 31 (EDI) / 61 (non-EDI)                               | 12  |
| ART                            | 50 (EDI) / 12 (non-EDI)                               | NA  |
| SNR                            | 35.61   | >88   |
| Pixel resolution               | 496 × 768   | 992 × 1024                                      |

**Table 5.3:** Population and image characteristics of the three samples used for assessing Choroidalyzer’s reproducibility. †: myopic/hyperopic status defined as  $< -1 / > 1$  dioptres. Dioptre measurements for i-Test sample was scan focus collected from the OCT imaging device metadata,, and this was the spherical equivalent for the GCU Topcon sample.

Two cohorts were used to assess Choroidalyzer’s reproducibility, the i-Test [225] and GCU Topcon [234] cohorts. Note that these were two cohorts used for training Choroidalyzer, and that the reproducibility analysis was performed after Choroidalyzer was constructed and during continued recruitment for the i-Test study. Thus, the reproducibility analysis contains more i-Test participants than the dataset used for training. We used all available eyes which had repeated data available (at the point of analysis) from each study (120 eyes from

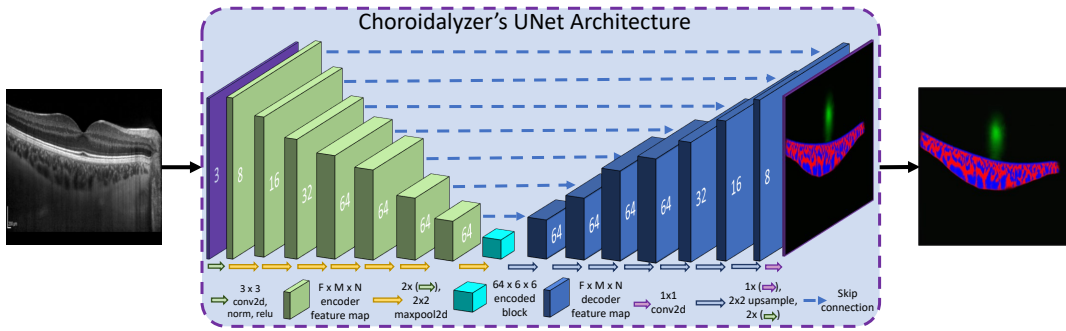
60 participants from the i-Test study and 33 eyes from 21 participants in the GCU Topcon study). Table 5.3 describes the population and image characteristics for these two samples.

The image acquisition for the i-Test reproducibility sample was previously described in section 4.2.2, which used an SD-OCT Spectralis OCT1 imaging device (Heidelberg Engineering, Heidelberg, Germany). Briefly, acquisition for an i-Test participant included a repeated OCT volume scan (collected within 1 minute) per eye, with EDI mode toggled on and off using active eye tracking with an ART of 50 and 9, respectively, collecting 31 and 61 parallel line-scans, respectively.

Acquisition for the GCU Topcon sample was described in section 5.2.1, which used an SS-OCT imaging device (Topcon, Tokyo, Japan). Briefly, acquisition for a GCU Topcon participant collected a twelve-line, fovea-centred OCT radial scan — with the native resolution kept fixed at  $992 \times 1024$ , unlike in Choroidalyzer’s modelling dataset. Repeated acquisitions were taken at approximately the same location during each visit and within 5 minutes of each other. Given repeated acquisition was not the primary aim of the GCU Topcon study (of which 24 participants were originally included in Choroidalyzer’s modelling dataset), only 21 participants had repeated OCT data, with nine only in one eye and twelve in both eyes.

## 5.3 Choroidalyzer’s deep learning model

Choroidalyzer segments the choroid region and vessels, and detects the fovea using a UNet deep learning model [138] with a depth of 7, utilising skip connections between symmetrical encoder (down) and decoder (up) blocks to aid information flow and enhance representation power. This depth allows Choroidalyzer to better consider the global context. The first 3 decoder blocks increase the internal channel dimension (number of feature maps) from 8 to 64, after which it is kept constant to reduce memory consumption and parameter count. The spatial dimensions of the feature maps are scaled by half at each layer which reduces the feature map of the deepest block to  $6 \times 6$  (for an input



**Figure 5.6:** Diagram of Choroidalalyzer’s UNet deep learning architecture. An input OCT B-scan is fed into Choroidalalyzer’s UNet architecture and outputs three probability segmentation maps for choroid region, vessel and fovea detection. The UNet architecture has 7 encoder (green) and 7 decoder (blue) blocks, with a  $6 \times 6$  encoded block (cyan) of dense features for scale invariance and global feature representation learning. Block transitions in the encoder and decoder are annotated below the diagram by green, yellow, blue and purple arrows. The white text overlaid on each block describes the number of feature maps per block, with the spatial dimension halving and doubling per block transition in the encoder and decoder blocks, respectively. conv2d, two dimensional convolution filter; norm, batch normalisation [273]; relu, non-linear rectified linear unit activation; maxpool2d, two-dimensional maximum pooling filter; upsample, bi-linear interpolation.

resolution of  $768 \times 768$ ) with 64 feature maps which enables spatial invariant feature extraction and global contextual learning.

Convolutions blocks (in both encoder and decoder arms) consist of two convolutional layers, each followed by batch normalisation [273] and non-linear rectified linear unit (ReLU) activation. Choroidalalyzer’s down-blocks (in the encoder arm) use maximum pooling at the end of each convolutional block to scale the spatial dimension of each feature map by  $1/2$ . Before every convolutional block, Choroidalalyzer’s up-blocks (in the decoder arm) use a  $1 \times 1$  convolution to reduce the channel dimension followed by bi-linear interpolation to increase the spatial dimension, which is more compute and memory efficient than the standard transposed convolutions, as there are no model parameters associated with this interpolation.

Figure 5.6 shows the UNet deep learning architecture designed for Choroidalalyzer. Given an input OCT B-scan of  $1 \times 768 \times 768$ , we expect Choroidalalyzer’s

inference procedure to output a probability segmentation map of resolution  $3 \times 768 \times 768$  — encoding each segmentation task’s probability map as a separate colour channel: choroid region in red, vessels in blue and fovea detection in green.

We trained Choroidalizer for 40 epochs after establishing convergence through negligible improvement on the validation set. Training was conducted using the AdamW optimizer [246] with a learning rate of  $5 \times 10^{-4}$  and weight decay of  $10^{-8}$  to minimise binary crossentropy loss, clamping the maximum gradient norm to 3 before each step to maintain numerical stability at back-propagation during training. We use automatic mixed precision to speed up training dramatically while reducing memory consumption by almost half. Forward pass and loss computation are done in bfloat16, a half-precision datatype optimised for machine learning.

During training, we apply the following data augmentations in random order per sample: Horizontal flip ( $p = 0.5$ ), changing the brightness and contrast independently (factors  $\sim U(0.5, 1.5)$ ,  $p = 0.95$ ), random rotation and shear (degrees  $\sim U(-25, 25)$  and  $\sim U(-15, 15)$  respectively,  $p = 1/3$ ), scaling the image (factor  $\sim U(0.8, 1.2)$ ,  $p = 1/3$ ), where  $U(a, b)$  denotes a uniform distribution between  $a$  and  $b$ , and  $p$  the probability of the transform being applied. For peripapillary B-scans which have a resolution of  $768 \times 1536$ , we place a window of size  $768 \times 768$  to crop the scan to, shifted horizontally at random multiples of 192 per example and epoch. For effective back-propagation of gradients during training, Choroidalizer’s fovea detection module, we employ one-sided label smoothing and set all other pixels outside the rectangular segmentation mask (figure 5.4(D)) to 0.01 instead of 0, stabilising training.

At inference, Choroidalizer outputs three probability maps, one each for the choroidal region, vessels and fovea. The choroidal region is thresholded at 0.5 to create a binary region mask. Note that this threshold can be fine-tuned by the end-user depending on the definition of the Choroid-Sclera boundary they prescribe to (see section 1.7 and figure 1.11). As detection of choroidal vessels is not a binary classification problem (as previously discussed in chapter 3), we leverage the output probabilities as a measure of uncertainty and leave the vessel mask as the raw probability map, after element-wise multi-

plication with the binarised region mask to remove any potential erroneously segmented areas outside the detected choroidal space. Finally, Choroidalyzer extracts the lateral position of the fovea by applying a 21-wide triangular filter to the column-wise sums of our model’s raw probability map, taking the column with the highest value. While we only evaluate the predicted lateral location of the fovea, Choroidalyzer can output the axial location too by passing 51-long triangular filters over the raw probability map and similarly taking the row with the highest value.

## 5.4 Choroidalyzer’s evaluation

### 5.4.1 Statistical analysis

#### 5.4.1.1 Quantitative evaluation

In this evaluation, we compare Choroidalyzer’s performance to the ground-truth labels generated using DeepGPET, MMCQ and manual labels for the choroidal space, vessels and fovea (section 5.2.1.1), respectively. To evaluate agreement in segmentations, we use the Dice coefficient and AUC, with the former metric used after applying a fixed threshold of 0.5 to binarise our model’s predictions. These metrics were defined previously in section 1.7.4.

Alongside standard segmentation metrics, we evaluate Choroidalyzer using choroid-derived metrics of thickness, area and CVI in a 6 mm, fovea-centred ROI, corresponding to the entire ETDRS macular grid [197]. The ROI for volume B-scans without the fovea present were centred at the middle column index of the image. The definition of these metrics and how they are computed can be found earlier in this thesis in section 1.7.1. As Choroidalyzer outputs raw probabilities for vessel segmentation predictions, rather than a binary mask, we propose a ‘soft’ CVI which takes the ratio between the predicted probabilities of the vessel map against the binarised region map. We evaluate the agreement of Choroidalyzer’s predicted metrics against ground truth-derived metrics using Pearson and Spearman correlations and mean absolute error (MAE), previously defined in section 1.7.4.



To measure agreement of fovea column predictions, we use MAE and median absolute error (Median AE). Additionally, we tested the effect of perturbing the fovea column laterally on segmentation-derived metrics. This was done by comparing fovea-centred metrics with metrics derived after the fovea column was randomly perturbed using a discretised uniform distribution  $\sim U(-6, 6)$  (excluding 0). 50 simulations were run on a random sample of 495 OCT B-scans, previously described in section [5.2.1.2](#).

### 5.4.1.2 Qualitative evaluation

In addition to quantitative evaluation against previously derived values from DeepGPET and MMCQ, we also used human graders to examine and characterise the behaviour of Choroidalyzer in cases of high error. Concretely, for each of the three tasks (region and vessel segmentation, fovea detection), we selected the 30 examples from both test sets where Choroidalyzer produced the highest errors. For redundant cases — such as adjacent, highly-similar slices from a volume scan — only one was retained. This left 28 cases for region, 29 for vessel, and 25 for fovea.

An adjudicating clinical ophthalmologist (Supervisor Dr. Ian J.C. MacCormick) was provided with the original image, Choroidalyzer’s prediction and the ground truth label while being masked to the identity of the methods. Images and labels were provided individually and as composites and the images were viewed according to the equipment listed in appendix [A.2](#). For each example, the adjudicator was asked which label they preferred. They also rated each label qualitatively on a 5-level ordinal scale (“Very bad”, “Bad”, “Okay”, “Good” and “Very good”) for region segmentation quality, as well as intravascular and interstitial vessel segmentation quality according to the definitions in appendix [A.1](#). The latter two were used to quantify any potential under-segmentation of vessels (false negative rate) and over-segmentation of the interstitial space (false positive rate), and guidance was provided using a pre-specified protocol in appendix [A.4](#).

Finally, we selected a random subsample of 20 B-scans at the subject-level from the external test set to be manually segmented by two graders, M1 and



M2. M1 was a clinical ophthalmologist (Supervisor Dr. Ian J.C. MacCormick) and M2 was the author of this thesis. Manual graders segmented the region and choroidal vessels using ITK-Snap [229], which permits pixel-level annotation, and were blinded to each others segmentations following a pre-specified protocol outlined in appendix A.5. The manual segmentations were compared to Choroidalyzer and to the current state-of-the-art, namely DeepGPET for region segmentation and Niblack for vessel segmentation. To keep Niblack's use consistent, we fixed the window size to 51 and the standard deviation offset to -0.05, as proposed by Muller, et al. [163]. We used Dice and AUC to measure segmentation agreement, as well MAE, and Pearson and Spearman correlation coefficients to measure agreement of choroidal metrics between the four methods for both region and vessel segmentation tasks. Evaluation metrics have been previously defined in section 1.7.4. We also compared the execution time of each approach for region and vessel segmentation, and report mean and standard deviation (SD) in seconds. Automatic methods were run on a laptop with a 4 year old Intel Core i5 (8<sup>th</sup> generation) CPU and 16 Gb of RAM, but no GPU.

The small sample size for these evaluations were chosen to strike a balance between the involvement of human experts to qualitatively assess Choroidalyzer's performance, and the time intensive nature of manual grading and segmentation of the choroid in OCT B-scans. Thus, this qualitative evaluation is able to make optimal use of a costly resource (human time). Moreover, the use of measurement and adjudication protocols promoted consistency between manual labelling, and ensured alignment between the definitions imposed by Choroidalyzer and applied by the adjudicator.

## 5.4.2 Results

### 5.4.2.1 Performance on internal and external test sets

Tables 5.4 and 5.5 shows the performance of Choroidalyzer on the internal and external test sets for the pixel-wise segmentations and choroid-derived measurements, respectively. Choroidalyzer achieved very good performance

#### 5.4. Choroidalalyzer's evaluation

| Set           | Region |        | Vessel |        | Fovea    |                |
|---------------|--------|--------|--------|--------|----------|----------------|
|               | AUC    | Dice   | AUC    | Dice   | MAE [px] | Median AE [px] |
| Internal test | 0.9998 | 0.9789 | 0.9982 | 0.8817 | 3.9      | 3.0            |
| External test | 0.9998 | 0.9749 | 0.9980 | 0.8703 | 3.4      | 3.0            |

**Table 5.4:** Segmentation metrics for Choroidalalyzer against ground-truth annotations from DeepGPET (region), MMCQ (vessel) and manual fovea detection for the internal and external test sets. Fovea error is in pixels (px).

| Set           | Thickness |          |                       | Area    |          |                       | CVI     |          |        |
|---------------|-----------|----------|-----------------------|---------|----------|-----------------------|---------|----------|--------|
|               | Pearson   | Spearman | MAE [ $\mu\text{m}$ ] | Pearson | Spearman | MAE [ $\text{mm}^2$ ] | Pearson | Spearman | MAE    |
| Internal test | 0.9754    | 0.9692   | 8.2252                | 0.9815  | 0.9786   | 0.0385                | 0.8285  | 0.8097   | 0.0206 |
| External test | 0.9831    | 0.9868   | 8.0888                | 0.9779  | 0.9848   | 0.0487                | 0.7948  | 0.7991   | 0.0306 |

**Table 5.5:** Agreement in choroid-derived metrics for Choroidalalyzer against ground-truth annotations from DeepGPET (thickness, area) and MMCQ (CVI) from the internal and external test sets. All correlation coefficients were statistically significant such that  $P < 0.0001$ .

in terms of AUC and Dice for region and vessels on both sets. Metrics for region are higher than for vessels, which is expected as vessel segmentation is much more difficult and ambiguous than region segmentation, and thus the ground truths (generated using MMCQ) are themselves not representative of a gold-standard. Performance was slightly higher for the internal test set than the external test set, which is expected, but only marginally so, indicating that our model potentially generalises well to this new cohort and potentially others related to systemic disease.

Table 5.6 shows the segmentation results when isolating the internal test set

|                     | Region |        | Vessel |        |
|---------------------|--------|--------|--------|--------|
|                     | AUC    | Dice   | AUC    | Dice   |
| Internal validation | 0.9983 | 0.9910 | 0.9514 | 0.7315 |
| Internal test       | 0.9996 | 0.9636 | 0.9925 | 0.7155 |

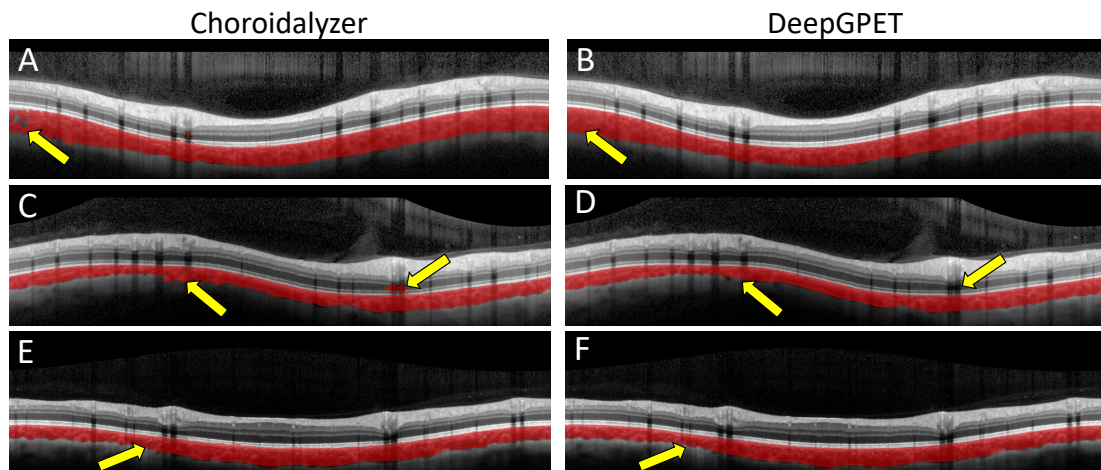
**Table 5.6:** Segmentation metrics for peripapillary OCT B-scans for Choroidalalyzer against ground-truth annotations from DeepGPET (region) and MMCQ (vessel) from the internal validation and test sets.

to only peripapillary B-scans. Our model achieved an AUC of 0.9996 (region) / 0.9925 (vessel) and Dice of 0.9636 (region) / 0.7155 (vessel). This is reasonable performance but lower than for other scans. Upon further investigation, we found that Choroidalyzer's performance on peripapillary B-scans was worse than DeepGPET's performance, even though DeepGPET was not trained on peripapillary choroids. Figure 5.7 shows a qualitative comparison of peripapillary choroids segmented by Choroidalyzer (left) and DeepGPET (right). Yellow arrows indicate area of disagreement, where Choroidalyzer either over- or under-segments, unlike DeepGPET.

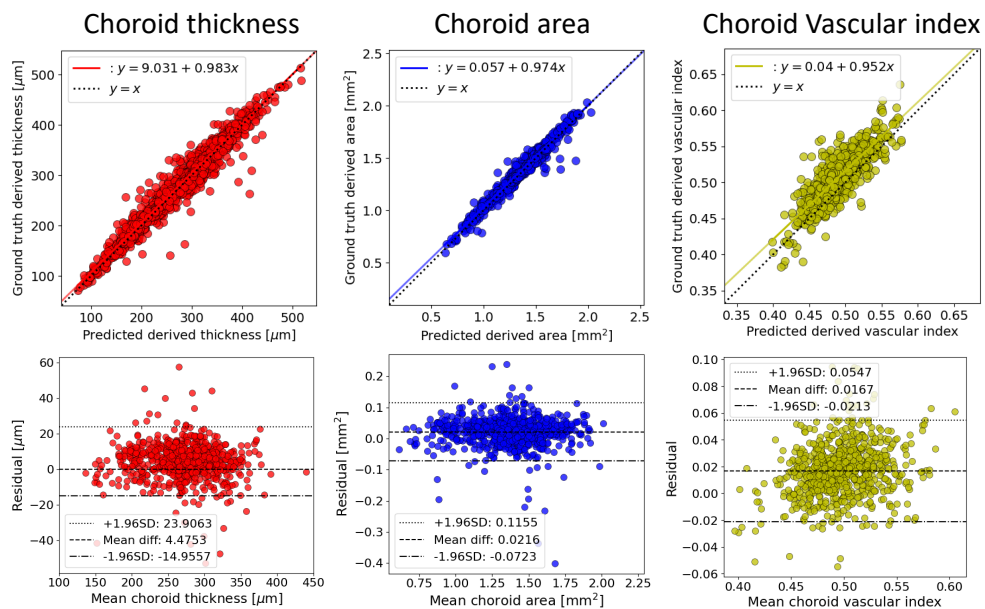
For the derived choroid metrics, Choroidalyzer shows excellent agreement with the ground truth labels derived from DeepGPET for thickness and area, with Pearson and Spearman correlations of 0.9692 or greater for both internal and external test sets. For CVI, performance is a bit lower, with correlations between 0.7948 and 0.8285. Although CVI depends on both region and vessel segmentation, the other metrics indicate that the error in CVI are driven primarily by differences in the vessel segmentation. Still, the observed correlations are high in absolute terms. Figures 5.8 and 5.9 shows correlation and Bland-Altman plots for the three derived metrics on internal and external test sets, respectively, which likewise indicate generally very good agreement.

For detection of the lateral position of the fovea, the model had an MAE of 3.9 pixels for the internal test set and 3.4 pixels for the external test set, with the median absolute error of 3 pixels for both. This is excellent performance, as an error of 3 pixels on a 768 pixel-wide image will not meaningfully change our ROI or choroid-derived metrics. Testing this formally on approximately 10% of the data by randomly perturbing the fovea column  $\pm 6$  pixels for 50 simulations yielded excellent Pearson correlation ( $r > 0.99$ ,  $P < 0.001$ ) across all derived choroid-derived metrics.

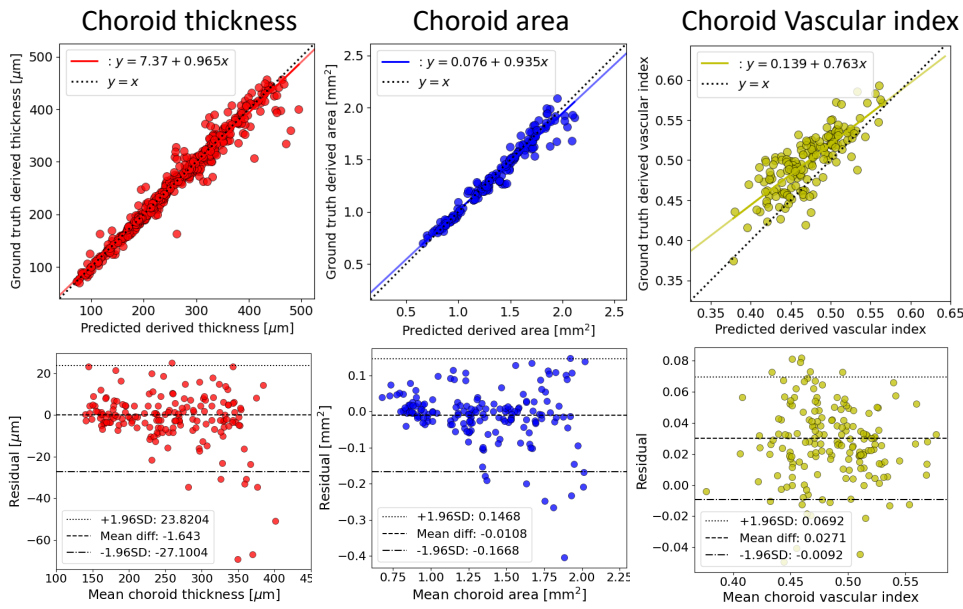
Figure 5.10 shows the distribution of correlations across all simulations for each metric in panel (A), with residual distributions for the poorest performing simulation per metric shown in panel (B). Even the poorest performing simulation generally showed strong agreement, with zero-centred residual distributions, and limits of agreement in the Bland-Altman plots well within acceptable bounds [106, 28, 235].



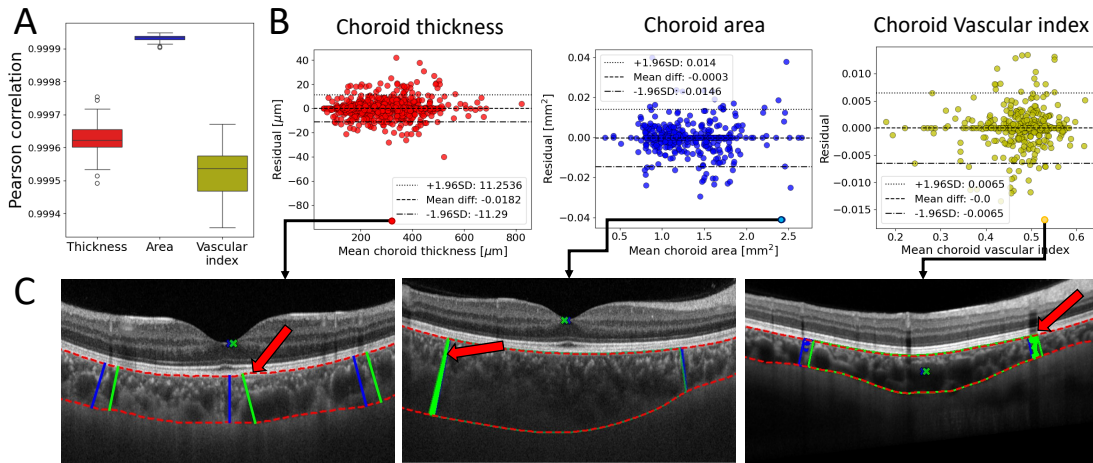
**Figure 5.7:** Peripapillary B-scans where DeepGPET (right) performed better than Choroidalzer (left). (A, C, E) Choroid region segmentations after application of Choroidalzer, and similarly for DeepGPET (B, D, F), Yellow arrows in each row represent sources of poor quality segmentation for Choroidalzer compared to DeepGPET.



**Figure 5.8:** Agreement in choroid-derived metrics against ground truth labels from DeepGPET (thickness, area) and MMCQ (CVI) for the internal test set. Top row are scatter plots with best regression fit and identity lines, bottom row are Bland-Altman plots.



**Figure 5.9:** Agreement in choroid-derived metrics against ground truth labels from DeepGPET (thickness, area) and MMCQ (CVI) for the external test set. Top row are scatter plots with best regression fit and identity lines, bottom row are Bland-Altman plots.



**Figure 5.10:** Results from comparing choroid-derived measurements after simulating 50 perturbations in the lateral position of the fovea using a uniform distribution  $\mathcal{U}(-6, 6)/\{0\}$ . (A) Distribution of Pearson correlation coefficients. (B) Bland-Altman plots for the worst performing simulation. (C) OCT B-scans with the highest error from the worst performing simulation. Original and perturbed fovea and measurements in blue and green, respectively. Red arrows to show the major source of the error.

Figure 5.10(C) shows the major outliers for each metric in the worst performing simulation. For choroid thickness, a kink in the upper boundary segmentation (red arrow) skewed the subfoveal thickness measurement, since thickness is measured *locally perpendicular* to the upper choroid boundary (see section 1.7.1). This was further exacerbated by unequal pixel length-scales, ultimately leading to a residual of 93 microns. For choroid area, this choroid was significantly thicker in the temporal region than it was in the nasal region. Thus, the fovea perturbation (6 pixels to the left) captured a significantly higher proportion of the choroid in its area calculation (red arrow), producing high error. Similarly, for CVI, the fovea perturbation (6 pixels to the right) captured a region of the choroid not only vascular but also artificially darkened by superficial retinal vessel shadowing (red arrow), compared to the original ROI which was not, causing high error. Nevertheless, on average the agreement is strong across the majority of cases. Thus, we propose that the fovea column quantitative error observed from Choroidalyzer does not significantly impact choroidal metrics.

#### 5.4.2.2 Comparison with manual segmentations

Tables 5.7 and 5.8 show the results using segmentation metrics and choroid-derived metrics when comparing Choroidalyzer and the current state-of-the-art (SOTA) (DeepGPET and Niblack for region and vessel analysis, respectively) to manual segmentations. Comparisons with Choroidalyzer/SOTA against individual manual graders were averaged for brevity.

| Comparison               | Region |        | Vessel |        |
|--------------------------|--------|--------|--------|--------|
|                          | AUC    | Dice   | AUC    | Dice   |
| M1 vs. M2                | 0.9639 | 0.9474 | 0.8891 | 0.7699 |
| Choroidalyzer vs. Manual | 0.9978 | 0.9375 | 0.9914 | 0.7669 |
| SOTA vs. Manual          | 0.9444 | 0.9333 | 0.9223 | 0.7742 |

**Table 5.7:** Segmentation metrics for the 20 images assessed manually and algorithmically from the external test set. M1 was grader supervisor Ian J.C. MacCormick and M2 was Jamie Burke (the author). SOTA: state-of-the-art.



## Chapter 5. Choroidalyzer: End-to-end choroidal analysis in optical coherence tomography

| Comparison               | Thickness |          |                       | Area    |          |                       | Vascular Index |          |        |
|--------------------------|-----------|----------|-----------------------|---------|----------|-----------------------|----------------|----------|--------|
|                          | Pearson   | Spearman | MAE ( $\mu\text{m}$ ) | Pearson | Spearman | MAE ( $\text{mm}^2$ ) | Pearson        | Spearman | MAE    |
| M1 vs. M2                | 0.9503    | 0.9521   | 17.8833               | 0.9516  | 0.9248   | 0.1096                | 0.8074         | 0.6857   | 0.0618 |
| Choroidalyzer vs. Manual | 0.9534    | 0.9663   | 20.9750               | 0.9554  | 0.9368   | 0.1150                | 0.6654         | 0.7383   | 0.0530 |
| SOTA vs. Manual          | 0.9676    | 0.9636   | 19.9250               | 0.9548  | 0.9233   | 0.1202                | 0.6907         | 0.6105   | 0.1682 |

**Table 5.8:** Derived metrics of thickness, area and CVI for the 20 images assessed manually and algorithmically from the external test set. M1 was supervisor Ian J.C. MacCormick and M2 was Jamie Burke (the author). SOTA: state-of-the-art.

| Method        | Region (s)           | Vessel (s)             | Total (s)              |
|---------------|----------------------|------------------------|------------------------|
| M1            | 78.400 $\pm$ 12.261  | 1506.000 $\pm$ 744.073 | 1584.400 $\pm$ 771.284 |
| M2            | 165.000 $\pm$ 23.889 | 1176.700 $\pm$ 263.968 | 1341.700 $\pm$ 265.800 |
| SOTA          | 0.751 $\pm$ 0.081    | 0.370 $\pm$ 0.105      | 1.121 $\pm$ 0.140      |
| Choroidalyzer | -                    | -                      | 0.299 $\pm$ 0.018      |

**Table 5.9:** Mean  $\pm$  SD execution time in seconds (s) of the four different approaches to region and vessel segmentation for the 20 images assessed manually and algorithmically from the external test set. SOTA: state-of-the-art.

Similar to the evaluation to ground truth labels, vessel Dice for Choroidalyzer (0.7410 vs M1 and 0.7927 vs M2; mean 0.7669) is worse than region Dice and even worse than the vessel Dice on both test sets. However, it is very similar to the manual grader’s (inter-grader) agreement of 0.7699. More generally, the inter-grader agreements for all other metrics are similar to Choroidalyzer’s agreement with the graders, with the notable exception of CVI. Here, Choroidalyzer’s MAE is better (0.0555 vs M1 and 0.0506 vs M2; mean 0.0530) than the inter-grader agreement (0.0618), as is the Spearman correlation but Pearson correlation is worse. Compared to the respective SOTA, Choroidalyzer has better agreement with the graders for most of the metrics, although methods are generally comparable.

Table 5.9 shows the execution per B-scan for all approaches. The manual graders on average needed more than 26 and 22 minutes (mean 24), with the vast majority of that time spent on the vessel segmentation. By contrast, the automatic methods on a standard laptop CPU needed about a second per B-scan and no human time at all. Thus, to get through a dataset of 100 scans, it would take manual graders about 40 hours of work, but with automatic meth-

ods it would be less than 2 minutes. With GPU-acceleration, Choroidalyzer and DeepGPET could achieve throughputs of dozens or hundreds of B-scans per second even on consumer-grade hardware. Comparing the automatic methods with each other, Choroidalyzer took 73% less time than DeepGPET and Niblack, while also detecting the fovea location. All three methods are very fast but for very large datasets or deployment on edge devices, Choroidalyzer’s efficiency is an additional advantage over existing automatic methods.

### 5.4.2.3 Detailed error analysis

Table 5.10 shows the results of manual inspection of scans where Choroidalyzer produced the highest error compared to the ground truth on the test sets. For region segmentation, Choroidalyzer was preferred in 8 cases, the ground truth in 5, and both methods were considered equally good in 15 cases. In terms of quality, Choroidalyzer was “Very bad” in only one case compared with 2 for the ground truth and “Very good” 3 times compared to none for the ground truth. For the vessels, Choroidalyzer was preferred in 13 cases, the ground truth in 4, and both were tied in 12 cases. Vessel segmentation is a harder task, with no methods achieving “Very good”. However, the intravascular scores for Choroidalyzer are substantially better, with no “Bad” or “Very bad” (vs. 3 and 2, respectively for ground truth) and far more “Good” (17 vs. 5), and the interstitial scores are similarly better. Finally, for the fovea, Choroidalyzer

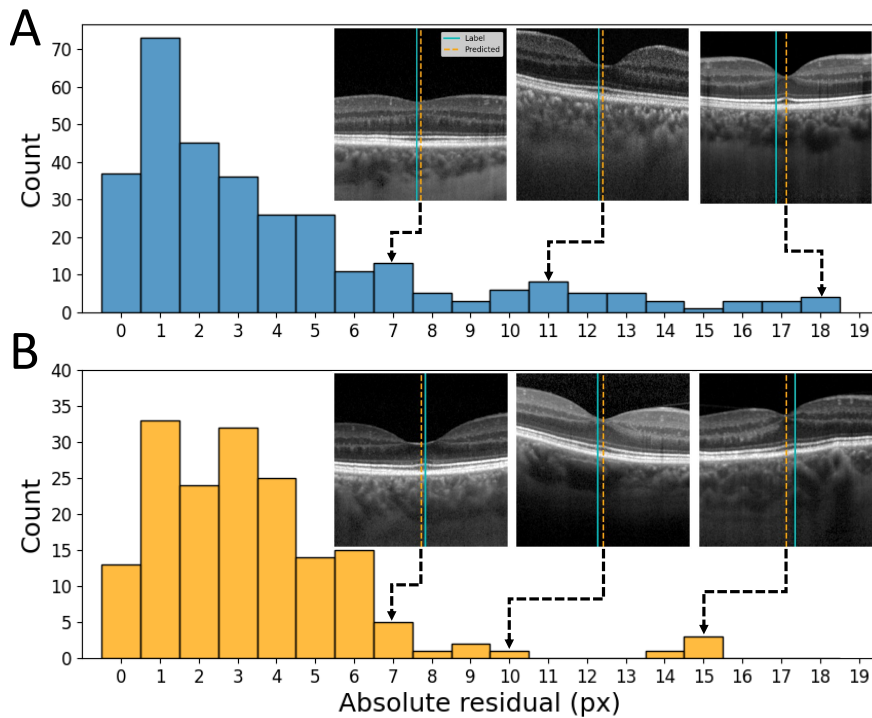
|        | Preferred Choroidalyzer | Preferred SOTA | Both equally good |
|--------|-------------------------|----------------|-------------------|
| Region | 8/28                    | 5/28           | 15/28             |
| Vessel | 13/29                   | 4/29           | 12/29             |
| Fovea  | 23/25                   | 1/25           | 1/25              |

| Method        | Region: quality                 | Vessel: intravascular            | Vessel: interstitial            |
|---------------|---------------------------------|----------------------------------|---------------------------------|
| Choroidalyzer | VG: 3, G: 14, O: 9, B: 1, VB: 1 | VG: 0, G: 17, O: 12, B: 0, VB: 0 | VG: 0, G: 20, O: 9, B: 0, VB: 0 |
| SOTA          | VG: 0, G: 17, O: 8, B: 1, VB: 2 | VG: 0, G: 5, O: 19, B: 3, VB: 2  | VG: 0, G: 17, O: 8, B: 2, VB: 2 |

**Table 5.10:** Preference and segmentation scores from expert adjudicator (I.M.) comparing the highest region segmentation, vessel segmentation and fovea column errors between Choroidalyzer and the ground truth labels. VG: very good; G: good; O: okay; B: bad; VB: very bad.





**Figure 5.11:** Histogram of absolute errors for fovea column detection for the internal (A) and external (B) test sets. Examples for different levels of error are shown with dotted lines indicating which part of the distribution they come from. In the examples, the teal line indicates the ground truth label, the dashed orange line the prediction. px: pixels.

was preferred 23/25 times and the ground truth only once, indicating that large fovea errors are almost exclusively due to mistakes in the manual ground truth labels.

Figure 5.11 shows the distributions of fovea errors for both test sets along with each example in both sets. For very large absolute error (>10 pixels), the ground truth labels are wrong and Choroidalalyzer correctly identifies the fovea location. For errors around 7 pixels, still twice the MAE, both methods are similar with either method sometimes being more correct. Further exploration revealed the majority of incorrectly labelled ground truths to be Topcon OCT B-scans, as each radial scan did not have the foveal pit assigned to the same lateral position in each B-scan, unlike a Heidelberg Engineering OCT volume scan. This was only identified during error analysis, which was after the initial

manual annotation which only identified the foveal pit for one representative B-scan per radial stack. Despite this oversight, Choroidalyzer learned to detect the fovea accurately.

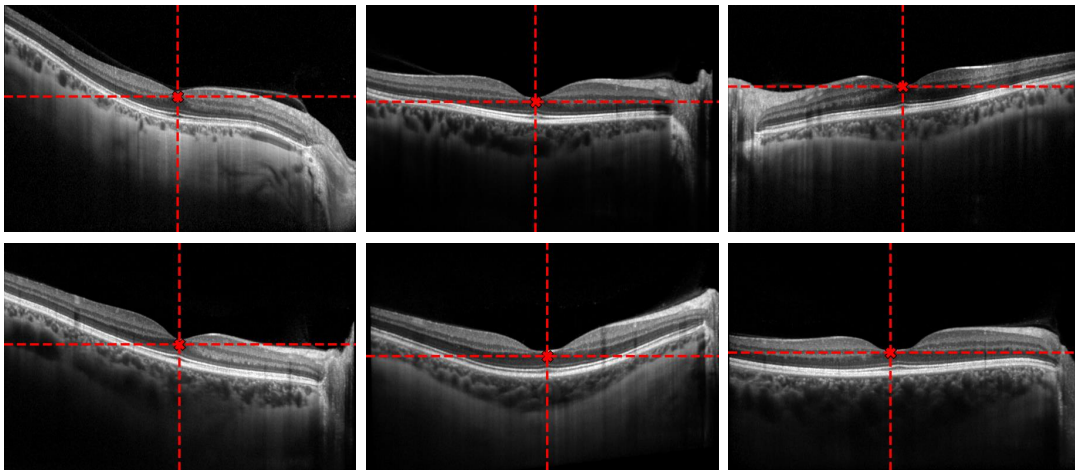
### 5.4.2.4 Potential generalisability

While not explicitly part of Choroidalyzer's evaluation, we conducted some initial experiments on Choroidalyzer's potential to generalise to different presentations of the choroid in OCT B-scans. Note that while we highlight the following potential functionality of Choroidalyzer, these are yet to be officially validated.

Firstly, our training procedure prioritised the detection of the lateral position over the axial position of the foveal pit. While this was a rational decision because the chorioretinal structures often appear predominantly aligned with the image-axis, it is often preferable to locate the axial position too. This is particularly important for choroids which exhibit skew (section 1.7.1.3). In figure 5.12, we show the axial and lateral positions of the fovea detected through Choroidalyzer for choroids with a range of curvature and skew (selected from figure 1.16 in section 1.7.1.3).

Note that across the board the general location of the fovea is correct, although the bottom-right prediction is off-centre. This particular choroid was from a highly hyperopic eye, which explains why the the intra-ocular structures are flattened and the foveola centralis is so large. Here, Choroidalyzer detected the fovea location as the point of highest illumination, but this was not centrally located, nor directly above the ridge which typically forms at the anterior point of outer retina. Nevertheless, even though Choroidalyzer's fovea detection map outputs a predominantly vertically-aligned probability map, our method of fovea pixel coordinate extraction (section 5.3) does reasonably well for these choroids which present skew and/or curvature.

Secondly, we tested Choroidalyzer's potential generalisability on a small sample of large and difficult choroids from the eyes of the GCU Topcon cohort which were discarded during data collection (section 5.2.1.2). In figure 5.13 we show three particularly challenging choroids which were part of the 94 images which were manually segmented and added to the dataset (in the internal test

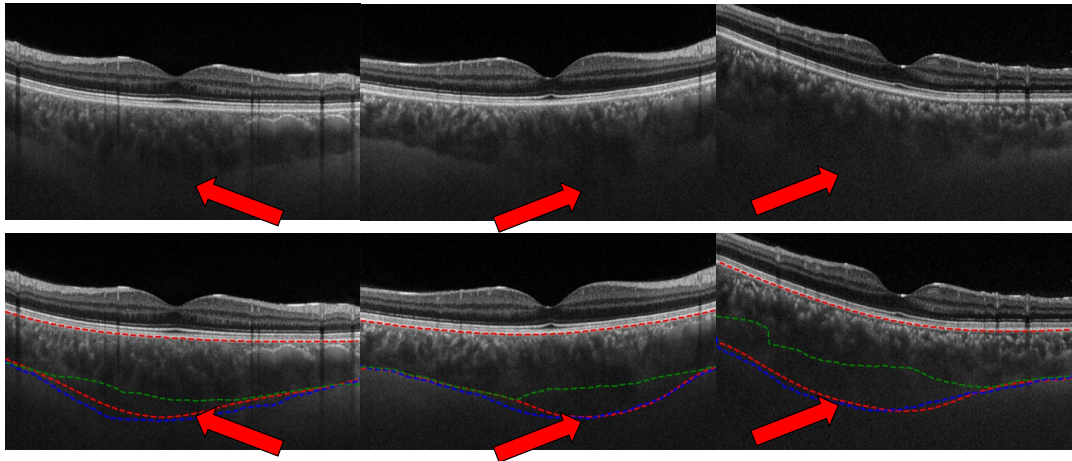


**Figure 5.12:** OCT B-scans from figure 1.16 with the detected axial and lateral foveal pit location overlaid as a red cross and the intersection of red vertical and horizontal lines.

set) to help improve Choroidalyzer’s generalisability to difficult choroids. In all cases, Choroidalyzer (red) lies closer to the manually detected Choroid-Sclera boundary (blue) than DeepGPET (green). This is an encouraging result, suggesting that Choroidalyzer is able to segment large choroids from SS-OCT data with poor Choroid-Sclera boundary visibility better than DeepGPET.

Finally, the majority of our dataset used to build Choroidalyzer were below the age of 50. Therefore, it’s very likely that much of the data used to train and evaluate Choroidalyzer did not have a visible suprachoroidal space [18]. Thus, we could not strictly evaluate Choroidalyzer’s ability to include/exclude the suprachoroid in it’s region detection. While we anticipate Choroidalyzer to exclude the suprachoroidal space from the choroid (following definition (1) from section 1.7) end-users may wish to follow a different definition of the Choroid-Sclera boundary which does include the suprachoroid.

Fortunately, Choroidalyzer’s region detection outputs a raw probability map which provides an analogue scale and can be tuned by the end-user according to the particular definition of the Choroid-Sclera boundary they choose for their study. In figure 5.14, the suprachoroid is visible for all four OCT B-scans. Setting a very high threshold ( $> 0.99$ ) allows Choroidalyzer to segment only as far as the largest, posterior-most vessels of the choroid (following definition (1)



**Figure 5.13:** Three example Topcon OCT B-scans comparing performance of DeepGPET (green) and Choroidalyzer (blue) with manual labelling (red). Red arrows indicate poor, but visible Choroid-Sclera boundary.

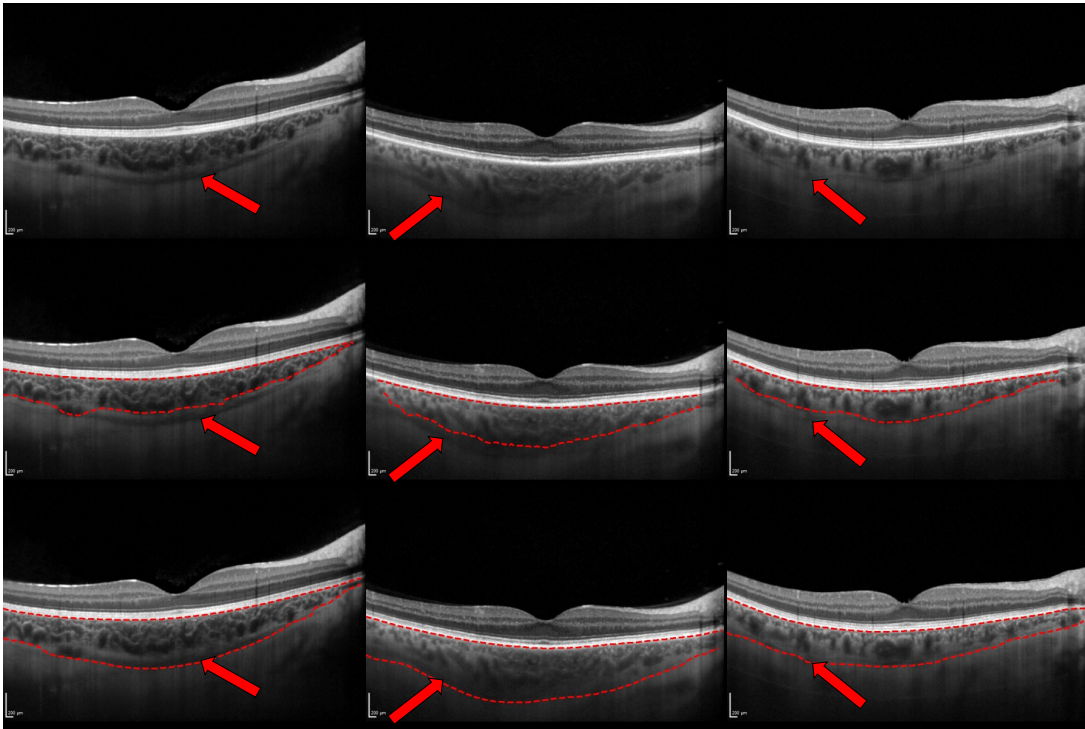
from section 1.7), while setting a very low threshold ( $< 0.05$ ) allows Choroidalyzer to segment past Haller's layer to the inner scleral boundary (following definition (3) from section 1.7). Thus, Choroidalyzer appears to be reasonably flexible in permitting different definitions of the Choroid-Sclera boundary. However, if attempting this, we recommend the end-user inspects the resulting segmentations to ensure Choroidalyzer is not attempting to segment both the choroid and suprachoroid separately and simultaneously (resulting in a zig-zag like Choroid-Sclera boundary).

## 5.5 Choroidalyzer's Reproducibility

### 5.5.1 Statistical analysis

Similar to MMCQ's reproducibility, we investigated Choroidalyzer's reproducibility using the i-Test and GCU Topcon repeated samples. For all repeated pairs, Choroidalyzer segmented the choroidal space and vessels and detected the fovea to allow standardised, fovea-centred choroid-derived measurements.

Briefly, for the i-Test sample, average choroid thickness and CVI were mea-



**Figure 5.14:** (Top) Selection of choroids from section 6.3 with visible supra-choroidal spaces. Red boundaries delineate the choroid segmented by Choroidalyzer and according to definitions (1) and (3) from section 1.7, i.e. to exclude (Middle) / include (Bottom) the suprachoroid.

sured for the nine sub-fields in the ETDRS grid for each OCT volume pair. Thus, there were 1080 comparisons made ( $9 \text{ values} \times 120 \text{ eyes}$ ). Choroidalyzer’s fovea detection was used to centre each spatial map onto the macula. For the GCU Topcon sample, 12-stack radial scans were collected throughout the day for all 33 eyes. To prevent sampling bias and remain objective in our analyses, we selected 13 repeated pairs from the morning, 12 from the afternoon and 8 from the evening. Thus, there were 396 comparisons made ( $33 \text{ eyes} \times 12 \text{ B-scans}$ ). Choroidalyzer was used to measure subfoveal choroidal thickness (SFCT), choroidal area and CVI in a 6 mm, fovea-centred ROI. All derived metrics have been defined in section 1.7.

We measure Choroidalyzer’s reproducibility at the population-level, reporting population mean and standard deviation (SD), as well as mean absolute error (MAE), Pearson and Spearman correlation coefficients. Additionally, we



show correlation plots and use Bland-Altman plots to assess agreement and error in the repeated samples. We also report Choroidalizer's reproducibility at the eye-level using measurement noise  $\lambda$  [200]. All performance metrics have been described previously in section 1.7.4.

## 5.5.2 Results

### 5.5.2.1 Population-level

Table 5.11 presents the reproducibility performance of Choroidalizer for macular OCT volume and B-scan data. Choroidalizer had excellent correlation across both cohorts (Pearson/Spearman for ETDRS choroid thickness: 0.9933/0.9969, B-scan SFCT: 0.9858/0.9889, B-scan choroid area: 0.9835/0.9870). Reproducibility for CVI was slightly lower but still good (ETDRS CVI: 0.9669/0.9655, B-scan CVI: 0.9090/0.9145).

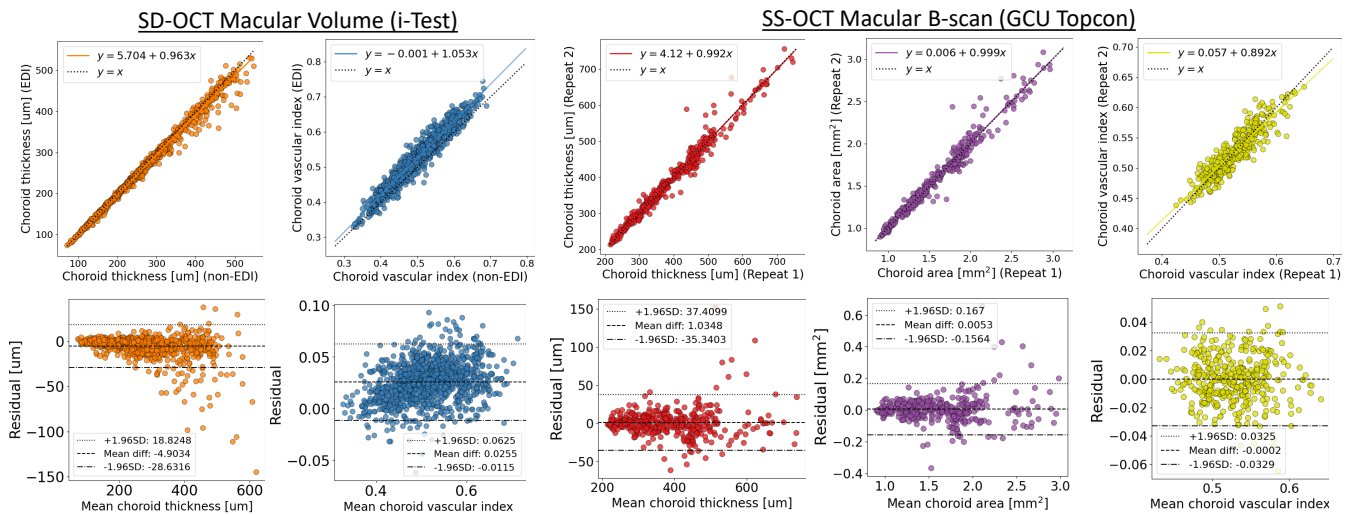
Measurements for the GCU Topcon data suggested these individuals had much thicker subfoveal choroids (mean  $\pm$  SD of  $386.9 \pm 110.9 \mu\text{m}$ ), compared to the i-Test measurements (mean  $\pm$  SD of  $275.7 \pm 91.8 \mu\text{m}$ ). While the i-Test sample was older than the GCU Topcon sample, the GCU Topcon study did target hyperopes during recruitment, thus resulting in smaller eyes and thicker choroids. Moreover, the i-Test measurements represent ETDRS sub-field averages allowing for extremal outliers, and potentially measurement error, to be averaged over unlike one-dimensional, point-source SFCT.

These observations explain why we see higher MAE in the GCU Topcon cohort (MAE:  $11.6 \mu\text{m}$ ) compared with the ETDRS data (MAE:  $6.7 \mu\text{m}$ ). Additionally,

| Dataset                           | Metric [unit]                       | Mean (SD)       | MAE     | Pearson | Spearman |
|-----------------------------------|-------------------------------------|-----------------|---------|---------|----------|
| i-Test<br>(ETDRS)                 | Choroid thickness [ $\mu\text{m}$ ] | 274.09 (91.80)  | 6.7342  | 0.9933  | 0.9969   |
|                                   | CVI                                 | 0.51 (0.03)     | 0.0271  | 0.9669  | 0.9655   |
| GCU Topcon<br>(Fovea-<br>centred) | SFCT [ $\mu\text{m}$ ]              | 392.63 (110.94) | 11.5535 | 0.9858  | 0.9889   |
|                                   | Choroid area [ $\text{mm}^2$ ]      | 1.62 (0.46)     | 0.0509  | 0.9835  | 0.9870   |
|                                   | CVI                                 | 0.53 (0.03)     | 0.0130  | 0.9090  | 0.9145   |

**Table 5.11:** Reproducibility performance of Choroidalizer. All Pearson and Spearman correlations were statistically significant with  $P < 0.0001$ .

## Chapter 5. Choroidalalyzer: End-to-end choroidal analysis in optical coherence tomography



**Figure 5.15:** Correlation and Bland-Altman plots for assessing the reproducibility of Choroidalalyzer for macular OCT data using Choroidalalyzer.

this also highlights the importance of reporting values over an area through averaging, rather than taking a single point-source measurement. Nevertheless, these errors are still below any change expected due to diurnal variation [52, 53, 54, 55, 56, 57], which changes by approximately  $30 \mu\text{m}$  over the course of the day. These errors are also below the threshold expected to exceed manual rater agreement in SFCT [106], and previously reported effect sizes in choroid thickness, such as in myopia progression which can be as low as  $20 - 30 \mu\text{m}$  [24, 110].

For CVI, we report an MAE of 0.0271 for SD-OCT ETDRS CVI (EDI and non-EDI acquisition), and 0.0130 for SS-OCT B-scan CVI. The MAE in the i-Test sample was double that in the GCU Topcon sample. The MAE was worse in the i-Test sample because of the additional speckle noise appearing in each non-EDI member of each i-Test sample pair in the ETDRS comparison, which can make vessel wall boundaries ambiguous. Nevertheless, it is unlikely that these MAEs are clinically significant. Breher, et al. [235] tested the reproducibility of Sonoda's Niblack method [27] across different sub-fields of the ETDRS macular region and reported a mean difference ranging from 3.9% to 5.1%. Additionally, Agrawal's major review on CVI as a biomarker in retinal pathology reported changes between healthy and disease eyes between 2%

and 6% [28].

Figure 5.15 show the correlation and Bland-Altman plots for the SD-OCT i-Test and SS-OCT GCU Topcon macular OCT samples. Generally we see that for larger choroids there is larger error in regional metrics. In the case for i-Test, this is entirely expected because the optical signal for non-EDI OCT acquisition degrades the deeper through the eye it penetrates, and the visibility of the deeper choroidal vasculature decays rapidly also. This is also the case for larger choroids in general, and thus why we see similarly larger error in the GCU Topcon sample — the largest choroid had an SFCT of  $756 \mu\text{m}$  (approximately 3.25 standard deviations away from the mean).

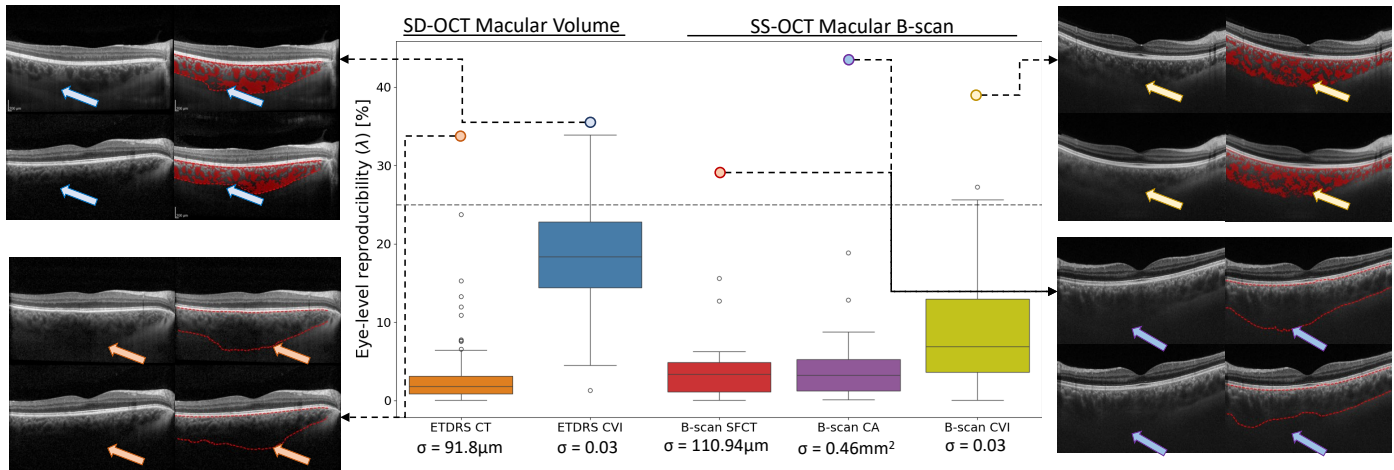
The distributions in the Bland-Altman plots for SD-OCT macular volume data in figure 5.15 show under-prediction of choroid thickness and over-prediction of CVI for the non-EDI scans of each pair, while the distributions for macular SS-OCT B-scan data are perfectly centred. The under-prediction of choroid thickness in the non-EDI measurements was likely due to optical signal decay in non-EDI scans, which had a direct consequence on CVI, as this metric is normalised by the size of the choroid, leading to over-prediction of CVI in non-EDI measurements. However, this systematic bias of 0.026 (2.6%) is unlikely to be clinically significant because it sits at the lower bound of previously reported effect sizes in retinal pathology [28].

### 5.5.2.2 Eye-level

Figure 5.16 presents the reproducibility of Choroidalalyzer for macular OCT data at the eye-level. We present box-plots to describe the distribution of measurement noise  $\lambda$ , expressed in terms of the overall population variability.

For region metrics, Choroidalalyzer's measurement error was very low compared to the population's variability, with the upper quartile of the box-plot distributions for ETDRS thickness/B-scan SFCT/B-scan area sitting below 5% of each metric's population variability. We observed higher variability in CVI at the eye-level, with B-scan CVI measurement error primarily sitting below 15% of the population variability ( $\sigma=0.03$ ) and 25% for ETDRS CVI ( $\sigma=0.03$ ). This sug-





**Figure 5.16:** Choroidalyzer’s reproducibility at the eye-level. We show distributions of eye-level measurement noise  $\lambda$  [200] for both the i-Test macular SD-OCT volume data and GCU Topcon macular SS-OCT B-scan data. Representative B-scans, with segmentations overlaid in red, are selected and shown as major outliers with arrows indicating the source of the error. The between-eye standard deviations are shown below each box-plot.

gests that CVI is less reliable than regional metrics like area given the population’s variability, a conclusion drawn also by Breher, et al. [235].

ETDRS CVI could have higher measurement variability than B-scan CVI because of the image quality difference between EDI and non-EDI volume pairs (ART 50 for EDI volume data vs. ART 12 for non-EDI volume data), as well as the unregistered nature of the volume scans which have different B-scan sampling rates (31 B-scans per EDI volume compared to 61 B-scans per non-EDI volume) and slightly different regions of interest up to some rotation. These differences ultimately lead to different stacks of cross-sections of the dense, heterogeneous choroidal vasculature being analysed, although previous research suggests this should not make a difference [184, 185, 186].

The major outliers for the macular OCT B-scans shown in figure 5.16 are mainly thick choroids whose lower boundaries (arrows) have poor visibility primarily due to decaying optical signal. This has a direct consequence for the CVI outliers, as CVI is normalised by the estimated size of the choroid. Moreover, image quality plays a clear role in poor quality segmentation, as the major

outlier for B-scan CVI shows one member of the repeated pair with significantly poorer contrast and thus poorer vessel wall definition. Overall, it is encouraging to see that major outliers are coming from challenging cases of thicker choroids with obscure Choroid-Sclera boundary and poor quality vessel wall definition.

## **5.6 Discussion**

### **5.6.1 Evaluation**

We developed Choroidalyzer, an end-to-end pipeline for choroidal analysis. Choroidalyzer shows excellent performance on the internal and external test sets. Where Choroidalyzer produced the highest errors were primarily cases of imperfect ground truth labels and Choroidalyzer was generally preferred by a blinded adjudicating ophthalmologist (supervisor Dr. Ian J.C. MacCormick), further indicating robustness and good performance. Its agreement with manual segmentations, which demand substantial time and attention from a human expert, is comparable to the inter-grader agreement. This suggests that Choroidalyzer performs well compared to laborious manual segmentation and also highlights the subjectivity introduced by manual graders. Choroidalyzer not only produces results similar to that of a skilled manual grader, but also does so fully-automatically without introducing subjectivity. Choroidalyzer is deterministic, meaning it is 100% repeatable when applied to the same image more than once, unlike certain semi-automatic approaches like GPET and especially manual approaches. Thus, researchers using Choroidalyzer would provide much more consistent, reliable and comparable results to other studies also using Choroidalyzer than if different manual graders were used in each case. Therefore, Choroidalyzer can help to improve the standardisation of choroid-derived metrics in OCT image sequences.

Additionally, Choroidalyzer saves a substantial amount of time per B-scan over manual segmentation, freeing up researcher time and enabling large scale analyses that otherwise would not be possible. Even compared to the current state-of-the-art for automatic methods, DeepGPET and Niblack, Choroidalyzer

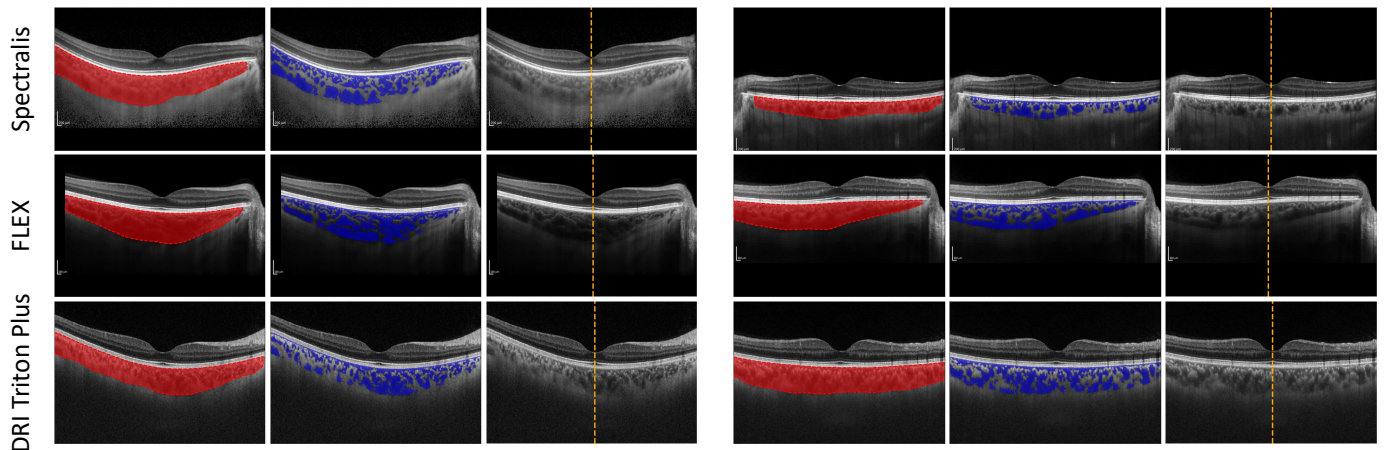
can do the analysis in roughly a quarter of the time, suggesting its feasibility for large-scale ophthalmic image analysis, even on a standard laptop CPU.

AlzEye [274] is a recently collected, large-scale dataset of retinal images from the Moorefield's Eye Hospital NHS Foundation Trust. AlzEye contains 1,567,358 OCT volume scans from 154,830 patients. This OCT image set could be processed by Choroidal analyzer in approximately 1 year using a low-grade CPU, compared to approximately 3.5 years using Niblack and DeepGPET combined. These estimates do not take into account the fact that Niblack requires parameter specification which may not be suitable across the entire OCT image set, and that the inference time for Choroidal analyzer (and DeepGPET, albeit) could be significantly improved with GPU hardware.

Importantly, Choroidal analyzer also provides an end-to-end pipeline which makes it easier to implement and use than having to combine multiple methods like Niblack and DeepGPET. Ease-of-use is often underappreciated in the literature but key in saving researchers time and allowing them to focus on the science. Additionally, Choroidal analyzer works similarly across different devices (Heidelberg and Topcon) and OCT scan types (SD-OCT and SS-OCT). Figure 5.17 shows some notable examples of Choroidal analyzer applied to two OCT B-scans from each of the OCT imaging devices used to develop it. This is a major advancement in the field as previous methods (tables 1.1 – 1.3) tend to only train/evaluate their methods on only on OCT imaging device.

Choroidal analyzer performed well against manual graders relative to the state-of-the-art methods, reaching or surpassing the levels of agreement even between the two manual graders, particularly for CVI, a far more difficult metric to calculate accurately than area and thickness. The inter-grader agreement between manual graders for these metrics indicate a potential lower bound of what effect sizes we might expect from these metrics. This has important downstream impact on the statistical confidence of results from cohort studies, particularly when assessing the choroidal vasculature.

It is often difficult to visualise the choroid due to imaging noise, poor eye tracking and patient fixation, or operator inexperience. Thus, in some cases vessel boundaries can be hard to discern. This is why we proposed to use a soft ver-



**Figure 5.17:** Examples of Choroidalyzer being applied to B-scans from different imaging devices for region segmentations (left), vessel segmentations (middle) and fovea column location (right).

sion of CVI where the probabilities that Choroidalyzer outputs are used instead of thresholded, binarised segmentations. The probabilities capture uncertainty about the precise location of the vessel wall and thus is more robust than using a single, somewhat arbitrary threshold. End-users could also tune the binarization threshold for their own images, if desired, which might help in instances of poor visibility of the choroidal vasculature.

The dataset in the present work was substantially larger than the one used for DeepGPET and the majority of other methods from the literature (tables 1.1 – 1.3), allowing Choroidalyzer to learn from a vast number (5,600 B-scans from 385 eyes in 233 participants) of unique OCT B-scans. The dataset also included manual ground truth labels of particularly challenging SS-OCT choroids. As a result, we found that Choroidalyzer may have the ability to segment challenging choroids with poor quality Choroid-Sclera boundary, unlike DeepGPET (figure 5.13). Moreover, Choroidalyzer was trained on region and vessel ground truths generated by fully automatic DeepGPET and semi-automatic MMCQ — the latter of which had fixed parameters to promote reproducibility and consistency in its segmentation procedure. Maloca, et al. [189] argues that such approaches to generating ground truth labels are preferable as they reduce subjectivity and thus bias and inconsistency, representing an additional benefit of Choroidalyzer over previous approaches in the literature which

use manual graders for generating vessel segmentation ground truth labels [168, 169, 170, 172].

Our model detected the fovea well, and the largest errors were cases where ground-truths were incorrectly labelled with the model correctly identifying the fovea location as confirmed by masked adjudication. Thus, the model performed even better than what the quantitative results suggest. Moreover, the error observed in both internal and external test sets did not result in significantly different choroid-derived metrics of thickness, area and CVI even after simulating random, per-sample deviations of  $\pm 6\text{px}$ , twice the Median absolute error on a subset of the dataset (figure 5.10).

We did observe that small errors in the lateral fovea position can in rare cases lead to large error in choroid-derived metrics (figure 5.10). These are where the local curvature of the segmented RPE-Choroid boundary shifts significantly (figure 5.10(C), left - red arrow), where the choroid is extremely heterogeneous in terms of size around the fovea (figure 5.10(C), middle - red arrow), or the unfortunate combination of superficial retinal vessel shadowing and vessel heterogeneity (figure 5.10(C), right - red arrow). Therefore, while on average small perturbations in the lateral position of the fovea does not impact choroid-derived metrics, this is not the case across all choroids.

In the present work, we have focused on identifying the lateral position of the fovea which is needed to define the fovea-centred ROI. The axial position is typically less important than the lateral position because the chorioretinal structures predominantly lie laterally. However, in cases related to poor image acquisition or high myopia, the choroid can exhibit skew or curvature relative to the image axis (figure 5.12). In these cases, it is not only important to make measurements which are choroid-aligned, i.e. locally perpendicular to the RPE-Choroid boundary as in section 1.7.1, but it is also pertinent to know the axial location of the fovea so that the resulting ROI is accurately accounting for this skew/curvature, which Choroidalizer is able to do successfully already (figure 5.12).

Segmentation performance for peripapillary scans was reasonable but much worse than for other scan patterns and locations. This could be due to those

scans being relatively rare in our dataset and showing parts of the retina on the nasal side of the optic disc that are not captured in macular B-scans. Although, we found that Choroidalyzer was less robust than DeepGPET at segmenting the peripapillary choroidal space (figure 5.7). DeepGPET's model was previously trained on a large natural image dataset, ImageNet [238], and only fine tuned for one task rather than training a network from scratch on three tasks like Choroidalyzer. Thus, DeepGPET's fine-tuning procedure may have instilled better generalisable features for choroid region segmentation. For example, DeepGPET's model weights were averaged over during training using exponential weight averaging which is known to improve generalisability [251].

Nevertheless, Choroidalyzer's lightweight model still outperforms DeepGPET on macular OCT data in terms of execution time and functionality. More peripapillary training data would likely increase performance, as well as a more rigorous training procedure, such as use of exponential moving average of the model weights. At present, Choroidalyzer can be used for peripapillary B-scan but requires subsequent manual inspection and potential correction. Furthermore, adjusting the binarization threshold for the vessel predictions may improve results. Regardless, DeepGPET is likely to provide more reliable choroid region segmentations.

The suprachoroidal space remains an elusive structure often not seen on OCT images because of how thin it is [275]. As people age, their choroid thins and their sclera stiffens, so there is a higher chance of this opening being seen on OCT [18]. However, there are clinical situations where it may be important to quantify this space, such as in post-operative care from surgery, where severely low eye pressure can sometimes cause the suprachoroidal space to bloat [36]. For Choroidalyzer's model development, we imposed a strict definition of the Choroid-Sclera boundary for Choroidalyzer to learn which follows definition (1) from section 1.7, i.e. exclusion of the suprachoroidal space. However, we found that the end-user may fine-tune the threshold used to binarise the region segmentation map to detect the Choroid-Sclera boundary according to all three definitions from section 1.7 (figure 5.14). This may have important consequences in scenarios where the suprachoroidal space is clinically relevant.

## 5.6.2 Reproducibility

Choroidalalyzer had very high correlations across repeated measurements (Pearson and Spearman correlations for regional metrics were all greater than or equal to 0.98, and 0.91 for CVI in the macular data). Encouragingly, MAEs from all paired samples in the repeated macular OCT data were below the values of previously reported reproducibility, agreement thresholds or effect sizes [24, 28, 106, 235].

The systematic bias observed in the macular OCT volume data was a result of comparing repeated scans where EDI mode was toggled on and off, increasing the potential for the non-EDI scan of each pair having poor Choroid-Sclera boundary visibility. The residuals for CVI were well balanced and appeared to show no apparent trend, apart from the minor systematic bias due to the aforementioned non-EDI optical signal decay. Major outliers from eye-level reproducibility analysis for both regional and vessel metrics in macular OCT data highlighted the challenge larger choroids pose for segmentation, given the propensity for optical signal degradation.

Nevertheless, the majority of measurement error of regional metrics observed for macular OCT B-scan data were within 5% of the population's variability for regional metrics, 25% for volumetric CVI and 15% for B-scan CVI. However, the population variability is notably smaller for CVI measurements, and the image quality, sampling rate and unregistered nature of the EDI and non-EDI volume pairs all likely played a role in producing greater variability in volumetric CVI. It's also very possible that the high measurement variability in CVI, relative to the regional metrics, is due to the exacerbation of measurement error in both vessel area and region area, as observed in chapter 3.

The higher measurement variability in CVI indicates the need to interpret this choroid-derived metric with care [235]. In chapter 3 we considered the importance of reporting purely vascular metrics like vessel area and volume alongside CVI, but this was not done here. It is possible that Choroidalalyzer-derived metrics of vessel area and volume may be more reliable than CVI, and thus should also be reported alongside. Ultimately, the macular regional metrics were more reliable than vessel metrics, but they were both still reasonable for



the homogeneous (young and healthy) cohorts analysed.

Thus, Choroidalizer is able to provide fully automatic choroid-derived measurements of the choroidal space and vessels which are standardised according to a robust, fovea-detected ROI. Importantly, the reproducibility analysis (both at the population- and eye-level) highlight the ability for Choroidalizer to output reliable and clinically meaningful measurement of the choroid, and provide end-users with interpretable metrics to help differentiate measurement error from true biological change in future studies which adopt Choroidalizer for OCT image analysis.

### 5.6.3 Limitations and future work

There were some limitations associated with Choroidalizer. Choroidalizer was built using only OCT data related to systemic disease and has not been evaluated on data related to eye pathology. While we anticipate Choroidalizer's value for ophthalmology in systemic health, its performance in the context of ocular health is yet unknown. In particular, stress testing Choroidalizer with examples of extreme choroid thinning or thickening such as in pathological myopia [255] or central serous chorioretinopathy [256], respectively, or optical disturbance from photoreceptor degeneration in geographic atrophy [257], would be suitable next steps. Nevertheless, we anticipate Choroidalizer will be a particularly useful tool in the field of oculomics, where it has demonstrated potential for generalisability.

Regarding Choroidalizer's reproducibility, and similar to DeepGPET's reproducibility limitations (section 4.6.2), our paired macular OCT volume data were not registered between acquisitions. Thus, the i-Test sample does not provide a perfect comparison to which segmentation could be the direct cause of measurement error. Moreover, there exists potential data leakage in our reproducibility analysis. Choroidalizer's training and evaluation had a distinct overlap with the two reproducibility populations (using ground truth labels previously segmented by DeepGPET). 22% of the i-Test sample (13 / 60) were used for training, but only the EDI counterpart of their OCT volume pair was used. Thus, while Choroidalizer had potentially seen 26 eyes (of 120 used



for reproducibility analysis) during training, the image artefacts introduced by deactivating EDI mode still posed a significant challenge to Choroidalyzer's reliability as a method. For the GCU Topcon sample, 20 eyes featured in at least one or more B-scans for training Choroidalyzer. However, the majority of B-scans from 62% of participants in the reproducibility cohort (13 / 21) were excluded from Choroidalyzer's dataset because of DeepGPET's failure to generate ground-truth segmentation labels during data curation, with only a few of their B-scans being added back after manual annotation (figure 5.5).

Choroidalyzer can process a single raw OCT B-scan inputted as an image file into a set choroid-derived measurements. Thus, Choroidalyzer is set up to only process single B-scans at a time, and would require additional configuration to process and measure OCT volumes (for example, to generate ETDRS sub-field measurements as in section 1.7.2). Moreover, it is common for researchers to collect OCT image data saved using proprietary file formats which store the image and metadata direct from the imaging device — a common proprietary file format for Heidelberg imaging devices are `.vol` RAW files. To facilitate truly end-to-end and accessible choroidal analysis to the general researcher, it should be wrapped into a more automatic pipeline.

This is why we are working presently on such a toolkit, OCTolyzer, which folds Choroidalyzer into a segmentation module (as well as DeepGPET, as alluded to in section 4.6.3) with the aim of processing macular OCT scan patterns (single-line, radial and volume) to provide reproducible and clinically meaningful measurements of the choroid in the macula. In this way, DeepGPET and Choroidalyzer provide fully automatic analysis of the posterior pole, further broadening the impact of these methods to the wider research community. While there are still hurdles preventing fully automatic large-scale ophthalmic image analysis due to the additional overhead of automating data extraction from the imaging device, OCTolyzer's ability to be run from the terminal eliminates certain pre-requisite knowledge in computer science or the Python programming language which Choroidalyzer requires to enable efficient and clinically meaningful analysis of the macular choroid.

### 5.6.4 Outputs

In this chapter, there have been two publication outputs, one of which has been published and peer-reviewed (by December 2024) and another which is currently in peer-review, but has been uploaded as a pre-print on Arxiv. Author of thesis is in bold type, with any co-lead authors underlined.

- Engelmann, Justin, **Jamie Burke**, Charlene Hamid, Megan Reid-Schachter, Dan Pugh, Neeraj Dhaun, Diana Moukaddem et al. “Choroidalyzer: An open-source, end-to-end pipeline for choroidal analysis in optical coherence tomography.” *Investigative Ophthalmology & Visual Science* 65, no. 6 (2024): 6-6. (Mathematical and Computational Ophthalmology Special Issue)
- **Burke, Jamie**, Justin Engelmann, Charlene Hamid, Diana Moukaddem, Dan Pugh, Neeraj Dhaun, Niall Strang et al. “OCTolyzer: Fully automatic analysis toolkit for segmentation and feature extracting in optical coherence tomography (OCT) and scanning laser ophthalmoscopy (SLO) data.” arXiv preprint arXiv:2407.14128 (2024). Submitted and in peer-review at Elsevier’s Medical Image Analysis.

The software associated with this method, Choroidalyzer, has been published as open-source and is freely available on GitHub [here](#). Similarly, OCTolyzer is freely available on GitHub [here](#).

### 5.6.5 Executive summary

In this chapter, we have developed a fully automatic, deep learning based method, Choroidalyzer, for generating automatic measurements of the choroid from raw SD-OCT or SS-OCT B-scans. Choroidalyzer segments the region and vessels of the choroid in an OCT B-scan and detects the lateral position of the fovea, enabling standardised measurements such as thickness, area, vessel area and CVI to be computed automatically in <0.3 seconds per B-scan

on a standard laptop CPU. Choroidalyzer performed similarly well across internal and external test sets, suggesting Choroidalyzer’s potential to generalise to new datasets in systemic disease. Choroidalyzer improved on current state-of-the-art approaches DeepGPET and Niblack for region and vessel detection — according to comparison with manual ground truth labels — and was preferred by a clinical ophthalmologist after blinded manual adjudication on examples exhibiting high error. We also found Choroidalyzer to output reproducible and clinically meaningful choroid-derived metrics at both the population- and eye-level, with regional metrics like thickness and area reporting lower measurement variability than CVI.

Choroidalyzer’s functionality potentially extends upon its original design, such as being able to segment challenging SS-OCT B-scans — unlike DeepGPET — and capturing the inherent uncertainty of the choroidal vessels in OCT B-scans by outputting a ‘soft’ CVI. Moreover, Choroidalyzer has the ability to consider different definitions of the Choroid-Sclera boundary (include/exclude suprachoroidal space). Moreover, while we anticipate Choroidalyzer to be particularly useful in datasets related to systemic health, Choroidalyzer has not been validated on examples related to retinal pathology. Furthermore, external validation and benchmarking of Choroidalyzer’s deep learning model on publicly available datasets of abnormal retinochoroidal structures would further facilitate the model’s applicability in wider ophthalmology, separate from its use in oculomics. As Choroidalyzer currently operates ‘per B-scan’, future work would look to wrap Choroidalyzer within an automatic pipeline, OCTolyzer, capable of performing large-scale choroidal image analysis processing for batch-processing of different macular OCT scan patterns.

# Chapter 6

## Applications of choroidal image analysis in OCT image sequences

### 6.1 The OCTANE Study: Chorioretinal changes in a longitudinal cohort of kidney donor and recipients

#### 6.1.1 Introduction

Chronic kidney disease (CKD) is a major cause of morbidity and mortality, affecting over 800 million people worldwide [276, 277, 278]. CKD is a gradual loss of kidney function, resulting in a failure to filter blood effectively and is associated with alterations in microvascular structure and function [279]. Assessment of microvascular function is therefore helpful in the diagnosis, prognosis and treatment of renal disease. Current clinical evaluation is based on blood and urine testing and invasive kidney biopsy.

The eye and the kidney show a striking resemblance in anatomy, physiology and response to disease [60], and in particular the choroidal microcirculation reflects the renal microcirculation [61]. Developmentally, the choroid (specifically, the choriocapillaris) and cortical (glomerular) capillaries have similar vascular endothelium with similarly sized fenestrated vessel walls allowing subreti-

nal fluid exchange in the choroid and blood filtration in the kidneys. Moreover, the circulatory systems in the choroid and renal cortex have similar proportions of blood flow in contrast to their retinal and renal medullary counterparts [7, 280]. These similarities suggest the potential for choroidal microvasculature to reflect information about kidney function for the care of patients with CKD.

Choroid thickness has previously been shown to correlate strongly with systemic inflammation and renal dysfunction in patients with CKD, compared with sex- and age-matched healthy and hypertensive individuals, suggesting that choroid thinning may reflect systemic microvascular injury in general and renal injury in particular [59]. An overactive sympathetic drive may contribute to disease progression in CKD [281], and could explain choroidal thinning since choroidal perfusion is influenced by its autonomic supply [59]. These observations suggest the potential clinical utility of choroidal biomarkers in renal disease. Moreover, a recent study reported that choroidal thinning significantly associates with decline in estimated glomerular filtration rate (eGFR), and is modifiable during treatment of CKD [16].

Microvascular dysfunction and systemic inflammation is observed in CKD, and it is important to measure microvascular function through renal biopsies during therapy. With the striking similarities in circulation, structure and pathogenic pathways between the kidney and eye [60, 61], and with the advent of OCT imaging, examining the microvascular structures in the choroid may provide a more practical and less invasive way to monitor treatment for CKD. We hypothesised that the observed effect of renal injury on the choroid at end-stage CKD is also similarly reflected during treatment, whereby improved renal function corresponds to an increase in choroid thickness and area. Thus, we prospectively analysed a longitudinal cohort of patients with end-stage CKD undergoing living donor kidney transplantation, to investigate the correspondence between the choroid and improving renal function/reduced systemic inflammation. Moreover, as a secondary analysis, we also explored whether a loss in renal function after unilateral nephrectomy results in thinning of the choroid in donors.

## 6.1.2 Data

### 6.1.2.1 Study population

Longitudinal data was prospectively collected on patients with end-stage CKD undergoing living donor kidney transplantation and healthy kidney donors undergoing unilateral nephrectomy (NCT0213274) [225]. Data collection and subsequent analyses were conducted after ethical approval from the South East Scotland Research Ethics Committee, in accordance with the principles of the Declaration of Helsinki and all participants gave informed consent to recruitment. Eligibility criteria for recruitment were (1) donors must be living and healthy throughout the period of analysis, (2) recipients with end-stage CKD have a functional kidney transplant and (3) must be aged 18 or over. To prevent ocular issues confounding our results, our exclusion criteria were (1) any ocular pathology pre-transplant, (2) any previous eye surgery, (3) a refractive error exceeding  $\pm 6$  dioptres or (4) a diagnosis of diabetes mellitus. We judged image quality by the OSCAR-IB criteria [226] and excluded images with B-scan signal quality  $\leq 15$ , using the imaging device's built-in SNR (Heidelberg Engineering, Heidelberg, Germany), or whose Choroid-Sclera boundary was invisible due to speckle noise or partial image cropping.

Data was sampled when possible at 8 time points: before transplant surgery (baseline), and seven time points post-transplant (1 week, 2 weeks, 4 weeks, 8 weeks, 16 weeks, 28 weeks and 52 weeks). Clinical data at each time point included standard biomarkers of renal function, including high sensitivity C-reactive protein (hsCRP), serum urea and creatinine, eGFR and urine protein to creatinine ratio (Urine P:Cr). At baseline, standard clinical measurements of systolic, diastolic and mean arterial blood pressure (BP) and body mass index (BMI) were also collected.

### 6.1.2.2 Image collection and analysis

Horizontal-line, fovea-centred, EDI-OCT B-scans of each participant's right eye was collected using the Heidelberg SD-OCT Spectralis OCT1 Standard Module. Each B-scan covered a 30° FOV (approximately 9 mm laterally), and was

extracted as a  $768 \times 768$  (pixel height $\times$ width) image for analysis. Patients were examined between 9am and 5pm and were endeavoured to be followed up at around the same time of day at each time point to limit the effects of diurnal variation on the choroid. Each scan was captured using active eye tracking with an ART of 100. Although refractive error and axial length were not collected for this population, the built-in scan focus parameter (Heidelberg Engineering, Heidelberg, Germany) is used to compensate for the eye's refractive error [102], and was extracted from the acquisition metadata and used as a surrogate value for refractive error.

Semi-automatic GPET was used to measure choroidal thickness and area using definitions described in section 1.7.1. We measured choroid thickness at three distinct locations: subfoveal and 2 mm temporally and nasally. Choroid area was measured in a 6 mm, fovea-centred ROI centred. Note that DeepGPET, MMCQ and Choroidalizer were not used here as these approaches were developed after this work was conducted.

### 6.1.2.3 Statistical analysis

We investigated how choroid thickness and area measurements changed over time in donors and recipients. To improve the robustness of our thickness measurements and prevent spurious associations, we considered only a single measurement by averaging choroid thickness across all three macular locations [23], resulting in a more representative value of thickness for each scan. We examined the difference in choroid thickness and area between measurements at baseline and one year post-transplant, testing for differences in their means longitudinally via the Student's t-test for dependent, paired samples. We use a P-value of 0.05 as the threshold for statistical significance.

We also investigated the average longitudinal evolution of relative change that choroidal measures underwent over time, and tested for linear associations between choroid image measures and markers of renal function. Pairwise associations were estimated at the patient-level for each cohort using repeated measures Pearson correlation coefficient,  $r_{rm}$  [282, 283], which accounts for variability across and within patients.

## 6.1. The OCTANE Study: Chorioretinal changes in a longitudinal cohort of kidney donor and recipients

---

The nested structure within the data was also leveraged to fit linear mixed-effects models, setting an individual random effect to model the variation among individuals, accounting for age, sex, approximate refractive error (measured in dioptres, D) and the time of day the scan was taken, relative to when their baseline scan was taken (measured in decimal hours). Linear mixed-effect models were used to account for variation in measurements across time and across individuals. For each clinical measure of renal function, we fitted a linear mixed-effect model for each choroidal image measure in turn as one of the associated independent variables. Linear mixed-effect modelling was carried out using the `lmer` package in base R (version 4.2.2) [284]. We identified independent variables as statistically significant predictors if their associated 95% confidence interval excluded 0.

### 6.1.3 Results

#### 6.1.3.1 Participants

Figure 6.1 shows a flowchart on how the population was selected for this study. At the time of data extraction, 22 donors and 23 recipients were eligible for analysis, after excluding a total of 65 recipients and 5 donors. The majority of exclusions were because there were no eye scans performed for these individuals. This was because, once recruited, they were scheduled for transplantation and then immediately returned to their referring institution after transplant away from Edinburgh, thus making it impractical to retain in the study and re-scan. Each participant was assessed on the day of transplant and then attempts were made at 7 other time periods over the following year (8 time points in total). 20 donors and 16 recipients for eligible for association estimation with clinical variables related to renal function, after excluding 2 donors and 10 recipients due to missing clinical measurements/eye scans post-transplant, or poor quality EDI-OCT images.

Tables 6.1 and 6.2 present the baseline demographics and clinical variables collected longitudinally for both cohorts, respectively. There were a number of missing values for certain clinical variables in both cohorts, which rendered



## Chapter 6. Applications of choroidal image analysis in OCT image sequences

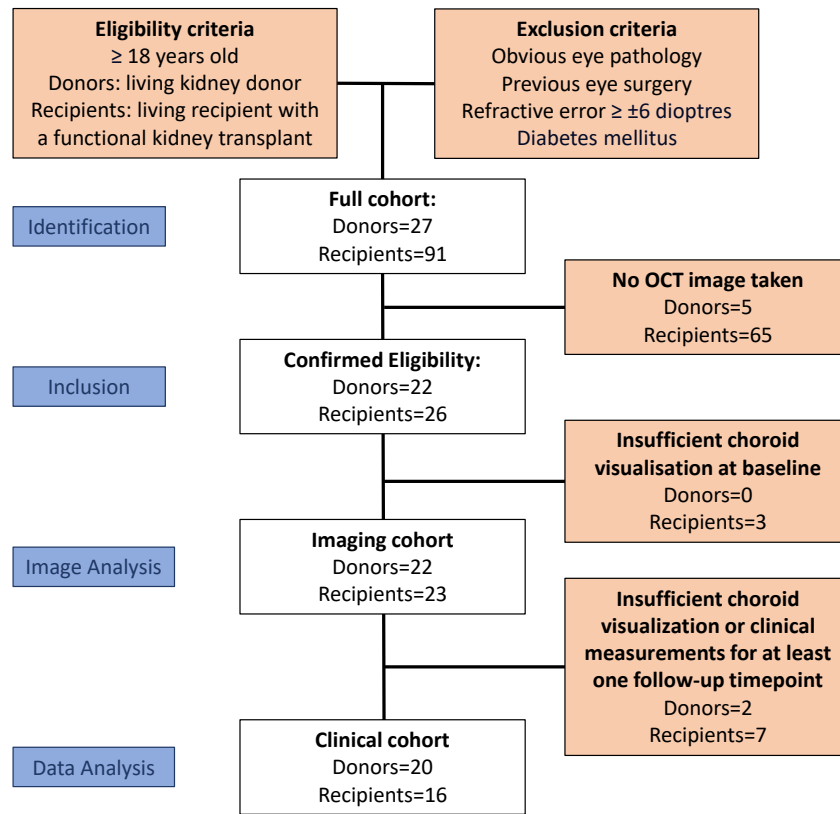
|                                     | Cohort        |                 |
|-------------------------------------|---------------|-----------------|
|                                     | Donors        | Recipients      |
| Sample, n (%)                       | 20            | 16              |
| Age, years                          | 51 ± 10       | 45 ± 14         |
| Male sex, n (%)                     | 8 (40)        | 10 (63)         |
| Approx. Refractive error (D)        | 0.49 ± 1.17   | -0.33 ± 1.73    |
| Daytime (Hr:Min)                    | 13:50 ± 1:41  | 13:59 ± 2:19    |
| BMI, kg/m <sup>2</sup> (%)          | 25 ± 3 (95)   | 27 ± 4 (100)    |
| Systolic BP, mmHg (%)               | 113 ± 14 (85) | 136 ± 15 (88)   |
| Diastolic BP, mmHg (%)              | 73 ± 18 (85)  | 83 ± 8 (88)     |
| MAP (%)                             | 86 ± 16 (85)  | 100 ± 9 (88)    |
| hsCRP, mg/l (%)                     | 1 ± 2 (60)    | 3 ± 4 (44)      |
| Serum urea, mmol/l (%)              | 5 ± 1 (95)    | 18 ± 7 (100)    |
| Creatinine, μmol/l (%)              | 66 ± 8 (95)   | 678 ± 187 (100) |
| eGFR, ml/min/1.73m <sup>2</sup> (%) | 91 ± 8 (95)   | 8 ± 2 (100)     |
| Urine P:Cr, mg/mmol (%)             | 1 ± 5 (38)    | 403 ± 394 (35)  |

**Table 6.1:** Baseline demographics collected for donors and recipients eligible. Values are shown as mean ± SD, and those in parentheses represent proportion of data completeness relative to baseline populations.

|                                     | Cohort     | Timepoint       |                |               |                |                |                |               |                |
|-------------------------------------|------------|-----------------|----------------|---------------|----------------|----------------|----------------|---------------|----------------|
|                                     |            | Baseline        | 1 Week         | 2 Weeks       | 4 Weeks        | 8 Weeks        | 18 Weeks       | 28 Weeks      | 52 Weeks       |
| Sample, n (%)                       | Donors     | 20 (100)        | 10 (50)        | 2 (10)        | 8 (40)         | 10 (50)        | 8 (40)         | 10 (50)       | 11 (55)        |
|                                     | Recipients | 16 (100)        | 12 (75)        | 4 (25)        | 5 (31)         | 9 (56)         | 5 (31)         | 7 (44)        | 7 (44)         |
| Daytime (Hr:Min)                    | Donors     | 13:50 ± 1:41    | 13:48 ± 1:20   | 12:04 ± 1:34  | 12:48 ± 3:47   | 13:43 ± 2:12   | 13:54 ± 2:00   | 13:30 ± 2:03  | 13:48 ± 2:07   |
|                                     | Recipients | 13:59 ± 2:19    | 12:44 ± 1:42   | 11:45 ± 1:50  | 12:23 ± 2:01   | 11:31 ± 1:13   | 12:29 ± 1:55   | 11:47 ± 1:39  | 12:25 ± 1:31   |
| Time from baseline (wks)            | Donors     | 0 ± 0           | 1 ± 0          | 2 ± 0         | 4 ± 1          | 8 ± 2          | 16 ± 4         | 29 ± 3        | 53 ± 5         |
|                                     | Recipients | 0 ± 0           | 1 ± 0          | 2 ± 0         | 4 ± 1          | 8 ± 3          | 16 ± 4         | 28 ± 3        | 52 ± 4         |
| hsCRP, mg/l (%)                     | Donors     | 1 ± 2 (60)      | 58 ± 28 (70)   | 2 ± 1 (100)   | 23 ± 34 (63)   | 3 ± 3 (70)     | 5 ± 9 (75)     | 1 ± 1 (80)    | 2 ± 4 (55)     |
|                                     | Recipients | 3 ± 4 (44)      | 17 ± 20 (42)   | 3 ± 3 (50)    | 7 ± 10 (100)   | 3 ± 2 (33)     | 2 ± 2 (60)     | 2 ± 0 (29)    | 25 ± 52 (86)   |
| Serum urea, mmol/l (%)              | Donors     | 5 ± 1 (95)      | 5 ± 2 (90)     | 6 ± 1 (100)   | 6 ± 1 (88)     | 6 ± 1 (100)    | 7 ± 2 (75)     | 6 ± 1 (90)    | 7 ± 1 (64)     |
|                                     | Recipients | 18 ± 7 (100)    | 6 ± 2 (75)     | 6 ± 1 (100)   | 6 ± 1 (100)    | 7 ± 2 (89)     | 7 ± 2 (100)    | 8 ± 2 (86)    | 6 ± 1 (100)    |
| Creatinine, μmol/l (%)              | Donors     | 66 ± 8 (95)     | 102 ± 17 (90)  | 102 ± 3 (100) | 89 ± 7 (88)    | 102 ± 24 (100) | 101 ± 25 (75)  | 100 ± 22 (90) | 91 ± 9 (64)    |
|                                     | Recipients | 678 ± 187 (100) | 114 ± 36 (75)  | 95 ± 26 (100) | 104 ± 25 (100) | 123 ± 20 (89)  | 119 ± 27 (100) | 124 ± 37 (86) | 109 ± 27 (100) |
| eGFR, ml/min/1.73m <sup>2</sup> (%) | Donors     | 91 ± 8 (95)     | 63 ± 7 (90)    | 64 ± 9 (100)  | 71 ± 10 (88)   | 68 ± 18 (100)  | 67 ± 14 (75)   | 66 ± 12 (90)  | 67 ± 9 (64)    |
|                                     | Recipients | 8 ± 2 (100)     | 62 ± 15 (65)   | 77 ± 18 (100) | 71 ± 14 (100)  | 61 ± 13 (89)   | 63 ± 16 (100)  | 58 ± 18 (86)  | 69 ± 13 (100)  |
| Urine P:Cr, mg/mmol (%)             | Donors     | 1 ± 5 (38)      | 165 ± 329 (50) | 1 ± 0 (50)    | 3 ± 3 (25)     | 1 ± 2 (60)     | 1 ± 2 (50)     | 1 ± 1 (60)    | 1 ± 1 (27)     |
|                                     | Recipients | 403 ± 394 (35)  | 257 ± 28 (25)  | 12 ± 0 (25)   | 89 ± 57 (60)   | 31 ± 19 (44)   | 21 ± 8 (100)   | 29 ± 20 (57)  | 9 ± 13 (57)    |

**Table 6.2:** Population statistics for donors and recipients consistently collected over period of analysis. Where appropriate, values are shown as mean ± SD, and values in parentheses represent proportion of data completeness relative to baseline populations.

## 6.1. The OCTANE Study: Chorioretinal changes in a longitudinal cohort of kidney donor and recipients



**Figure 6.1:** Population derivation for longitudinal renal cohort.

them too incomplete to be used for downstream modelling (BMI, BP, hsCRP, Urine P:Cr). Thus, based on the level of complete data (relative to baseline sample size) in table 6.2, only clinical variables eGFR, serum creatinine and serum urea were selected for downstream statistical analysis.

### 6.1.3.2 Choroidal variation

Table 6.3 shows the average choroidal measurements and markers of renal function at baseline and one year post-transplant. All measurements were significantly different one year post-transplant, relative to baseline measurements in both cohorts. Figure 6.2 shows the average percentage change in choroidal measurements (with standard errors annotated as vertical lines at each time point) from baseline measurements in both cohorts. In transplant recipients, there was a significant increase in choroidal thickness of  $12.8 \pm 4.8\%$  (me-

## Chapter 6. Applications of choroidal image analysis in OCT image sequences

| Measurement                       | Donors      |             |                  | Recipients  |             |                  |
|-----------------------------------|-------------|-------------|------------------|-------------|-------------|------------------|
|                                   | Baseline    | 1 year PT   | <i>P</i> -value  | Baseline    | 1 year PT   | <i>P</i> -value  |
| Choroid area [mm <sup>2</sup> ]   | 1.59 ± 0.41 | 1.49 ± 0.41 | <b>0.01</b>      | 1.58 ± 0.41 | 1.74 ± 0.46 | <b>0.01</b>      |
| Choroid thickness [μm]            | 278 ± 69    | 259 ± 68    | <b>0.01</b>      | 271 ± 70    | 301 ± 75    | <b>0.01</b>      |
| eGFR (ml/min/1.73m <sup>2</sup> ) | 91 ± 5      | 64 ± 7      | <b>&lt;0.001</b> | 8 ± 3       | 69 ± 14     | <b>&lt;0.001</b> |
| Creatinine (μmol/l)               | 69 ± 9      | 92 ± 10     | <b>&lt;0.001</b> | 659 ± 196   | 109 ± 29    | <b>&lt;0.001</b> |
| Urea (mmol/l)                     | 5.0 ± 0.8   | 6.8 ± 1.3   | <b>0.002</b>     | 18.5 ± 9.7  | 6.7 ± 1.7   | <b>0.005</b>     |

**Table 6.3:** Choroidal measurements at baseline and one year post-transplant.

|                   | Donors |            |       | Recipients |            |        |
|-------------------|--------|------------|-------|------------|------------|--------|
|                   | eGFR   | Creatinine | Urea  | eGFR       | Creatinine | Urea   |
| Choroid area      | -0.01  | 0.04       | -0.14 | 0.68***    | -0.68***   | -0.42* |
| Choroid thickness | -0.09  | 0.13       | -0.27 | 0.51**     | -0.51**    | -0.26* |

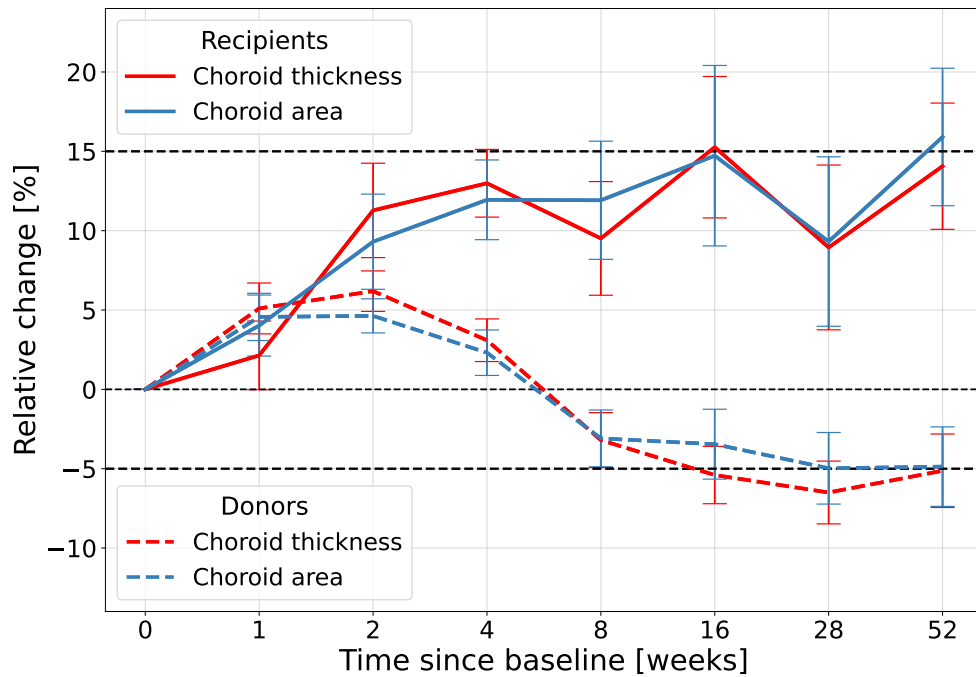
**Table 6.4:** Unadjusted, repeated Pearson correlation coefficient,  $r_{rm}$  [283] between choroidal thickness, area and markers of renal function eGFR, serum creatinine and serum urea. \*\*\*:  $P < 0.0001$ , \*\*:  $P < 0.001$ , \*:  $P < 0.05$

dian: 11.4%) from baseline after 4 weeks post-transplant ( $P=0.02$ ). After 1 year post-transplant, we observed a significant increase of  $14.1 \pm 11.4\%$  (median: 10.7%) from baseline ( $P=0.01$ ). Interestingly, we observed a significant increase of  $5.1 \pm 5.6\%$  (median: 5.4%) 1 week post-transplant in the choroid of donors ( $P=0.01$ ). The average choroid remained enlarged relative to baseline measurements and at some point between 4 and 8 weeks post-transplant the choroid deflated. After 1 year post-transplant, we observed a significant decrease of  $-4.9 \pm 7.9\%$  (median: -3.5%) in automatic choroid thickness ( $P=0.01$ ).

Table 6.4 presents the repeated Pearson correlation coefficients between choroidal measurements and clinical variables. In the recipient cohort, all choroidal measurements linearly correlated with eGFR, serum creatinine and serum urea with statistical significance ( $P < 0.05$ ). However, there were no significant, linear associations found between choroidal measurements and renal function in the donor cohort. Choroid area appeared to have stronger linear correspondence with markers of renal function in recipients compared to choroid thickness ( $r_{rm}$  between eGFR and choroid area, 0.68; choroid thickness, 0.51).

Table 6.5 shows the model summaries for predicting eGFR, serum creatinine

## 6.1. The OCTANE Study: Choriorretinal changes in a longitudinal cohort of kidney donor and recipients



**Figure 6.2:** Longitudinal change in choroid thickness and area over time, measured relative to baseline choroid measurements. Error bars refer to standard error measurements. Horizontal dashed lines at +15% and -5% are superimposed to aid readability.

and serum urea using each choroidal measurement in turn for transplant recipients (top) and donors (bottom). In transplant recipients linear mixed-effects modelling showed evidence of statistically significant positive linear association with eGFR, and negative associations with serum creatinine and urea levels.

Interestingly, the time difference ('daytime from baseline' in table 6.5) between follow-up and baseline image acquisition was a statistically significant predictor of all markers of renal functions for models for transplant recipients. This covariate had the opposite effect that choroid thickness and area had in each model, i.e.  $\hat{\beta}$  for the model using choroid area in transplant recipients was +1.54, and for daytime from baseline this was -0.40. There were no statistically significant associations found between any choroidal measurement and marker of renal function in the donor cohort. However, sex was associated with serum creatinine, albeit with weak statistical significance.

## Chapter 6. Applications of choroidal image analysis in OCT image sequences

| Transplant recipients        | eGFR          |                    |                  | Creatinine    |                    |                  | Urea          |                    |                  |
|------------------------------|---------------|--------------------|------------------|---------------|--------------------|------------------|---------------|--------------------|------------------|
|                              | $\hat{\beta}$ | 95% CI             | P-value          | $\hat{\beta}$ | 95% CI             | P-value          | $\hat{\beta}$ | 95% CI             | P-value          |
| <u>Choroid area</u>          |               |                    |                  |               |                    |                  |               |                    |                  |
| Intercept                    | -0.10         | (-1.38, 1.17)      | 0.84             | 0.09          | (-1.18, 1.36)      | 0.86             | 0.03          | (-0.49, 0.55)      | 0.90             |
| Age                          | 0.98          | (-0.47, 2.43)      | 0.15             | -1.11         | (-2.54, 0.31)      | 0.10             | -0.30         | (-0.95, 0.34)      | 0.29             |
| Sex (Male)                   | -0.24         | (-1.55, 1.07)      | 0.66             | 0.27          | (-1.03, 1.57)      | 0.60             | 0.22          | (-0.31, 0.76)      | 0.34             |
| <b>Daytime from Baseline</b> | -0.40         | (-0.59, -0.21)     | <b>&lt;0.001</b> | 0.59          | (0.39, 0.80)       | <b>&lt;0.001</b> | 0.57          | (0.32, 0.81)       | <b>&lt;0.001</b> |
| Approx. Refraction           | -0.86         | (-2.17, 0.46)      | 0.114            | 0.92          | (-0.37, 2.21)      | 0.12             | 0.31          | (-0.28, 0.89)      | 0.24             |
| <b>Choroid area</b>          | 1.54          | (0.97, 2.10)       | <b>&lt;0.001</b> | -1.52         | (-2.12, -0.92)     | <b>&lt;0.001</b> | -0.60         | (-1.09, -0.11)     | <b>0.02</b>      |
| Patient Std. Dev.            | 1.63          | Cond. $R^2$ : 0.93 |                  | 0.60          | Cond. $R^2$ : 0.91 |                  | 0.55          | Cond. $R^2$ : 0.57 |                  |
| <u>Choroid thickness</u>     |               |                    |                  |               |                    |                  |               |                    |                  |
| Intercept                    | -0.10         | (-1.19, 0.99)      | 0.83             | 0.07          | (-0.91, 1.05)      | 0.84             | 0.03          | (-0.48, 0.53)      | 0.90             |
| Age                          | 0.74          | (-0.51, 1.99)      | 0.20             | -0.80         | (-1.92, 0.31)      | 0.12             | -0.27         | (-0.90, 0.36)      | 0.33             |
| Sex (Male)                   | -0.31         | (-1.43, 0.81)      | 0.49             | 0.31          | (-0.69, 1.30)      | 0.44             | 0.28          | (-0.25, 0.81)      | 0.24             |
| <b>Daytime from Baseline</b> | -0.46         | (-0.67, -0.24)     | <b>&lt;0.001</b> | 0.65          | (0.42, 0.88)       | <b>&lt;0.001</b> | 0.55          | (0.30, 0.80)       | <b>&lt;0.001</b> |
| Approx. Refraction           | -0.43         | (-1.55, 0.68)      | 0.37             | 0.46          | (-0.54, 1.46)      | 0.27             | 0.18          | (-0.35, 0.71)      | 0.42             |
| <b>Choroid thickness</b>     | 1.15          | (0.51, 1.78)       | <b>&lt;0.001</b> | -1.01         | (-1.62, -0.40)     | <b>0.004</b>     | -0.54         | (-1.06, -0.02)     | <b>0.04</b>      |
| Patient Std. Dev.            | 1.37          | Cond. $R^2$ : 0.88 |                  | 1.05          | Cond. $R^2$ : 0.81 |                  | 0.50          | Cond. $R^2$ : 0.52 |                  |
| <b>Transplant donors</b>     |               |                    |                  |               |                    |                  |               |                    |                  |
| <u>Choroid area</u>          |               |                    |                  |               |                    |                  |               |                    |                  |
| Intercept                    | 0.04          | (-0.28, 0.36)      | 0.78             | -0.03         | (-0.34, 0.28)      | 0.83             | -0.05         | (-0.36, 0.26)      | 0.72             |
| Age                          | -0.34         | (-0.76, 0.08)      | 0.10             | 0.07          | (-0.34, 0.47)      | 0.73             | 0.08          | (-0.33, 0.48)      | 0.69             |
| <b>Sex (Male)</b>            | 0.06          | (-0.29, 0.40)      | 0.73             | 0.36          | (0.02, 0.69)       | <b>0.04</b>      | 0.11          | (-0.22, 0.45)      | 0.48             |
| Time from Baseline           | 3.5e-3        | (-0.24, 0.25)      | 0.98             | -0.02         | (-0.25, 0.21)      | 0.86             | -0.06         | (0.0-0.31, 0.18)   | 0.60             |
| Approx. Refraction           | -0.05         | (-0.47, 0.36)      | 0.79             | 0.13          | (-0.27, 0.53)      | 0.51             | 0.22          | (-0.18, 0.62)      | 0.26             |
| Choroid area                 | 6.7e-4        | (-0.33, 0.33)      | 0.99             | -0.11         | (-0.42, 0.21)      | 0.49             | -0.07         | (-0.39, 0.24)      | 0.64             |
| Patient Std. Dev.            | 0.41          | Cond. $R^2$ : 0.27 |                  | 0.45          | Cond. $R^2$ : 0.38 |                  | 0.35          | Cond. $R^2$ : 0.21 |                  |
| <u>Choroid thickness</u>     |               |                    |                  |               |                    |                  |               |                    |                  |
| Intercept                    | 0.04          | (-0.28, 0.36)      | 0.78             | -0.03         | (-0.34, 0.28)      | 0.84             | -0.05         | (-0.35, 0.25)      | 0.73             |
| Age                          | -0.34         | (-0.76, 0.36)      | 0.10             | 0.06          | (-0.35, 0.46)      | 0.78             | 0.07          | (-0.33, 0.47)      | 0.72             |
| <b>Sex (Male)</b>            | 0.05          | (-0.30, 0.40)      | 0.75             | 0.36          | (0.02, 0.70)       | <b>0.04</b>      | 0.12          | (-0.22, 0.45)      | 0.47             |
| Time from Baseline           | 3.9e-3        | (-0.24, 0.25)      | 0.98             | -0.02         | (-0.25, 0.21)      | 0.86             | -0.06         | (-0.31, 0.18)      | 0.60             |
| Approx. Refraction           | -0.05         | (-0.47, 0.36)      | 0.80             | 0.13          | (-0.27, 0.54)      | 0.49             | 0.22          | (-0.17, 0.62)      | 0.24             |
| Choroid thickness            | -0.01         | (-0.34, 0.31)      | 0.95             | -0.08         | (-0.39, 0.23)      | 0.60             | -0.07         | (-0.38, 0.25)      | 0.66             |
| Patient Std. Dev.            | 0.41          | Cond. $R^2$ : 0.27 |                  | 0.45          | Cond. $R^2$ : 0.37 |                  | 0.34          | Cond. $R^2$ : 0.21 |                  |

**Table 6.5:** Linear mixed-effects modelling summary tables for predicting eGFR, serum creatinine and serum urea in transplant recipients (top) and donors (bottom) for choroid thickness and choroid area. Statistically significant covariates are shown in bold.

### 6.1.4 Discussion

Correlation analyses and linear mixed-effects modelling reported significant, linear correspondence between choroidal image measures and markers of renal function over time. All choroidal measurements changed substantially over time for all study participants (figure 6.2), as did eGFR, serum creatinine and serum urea (table 6.2). This is consistent with the choroid reflecting renal function during treatment of CKD. The choroid does vary naturally during the day due to diurnal variation, and Tan et al. [53] described the average change in choroid thickness across daytime hours to be approximately  $8.5 \pm 5.2\%$ . The

## 6.1. The OCTANE Study: Choriorretinal changes in a longitudinal cohort of kidney donor and recipients

---

change reported here in choroid thickness in transplant recipients one year post-transplant was  $14.1 \pm 11.4\%$  ( $14.7 \pm 11.2\%$  for choroid area), suggesting that choroidal inflation in transplant recipients cannot be explained fully through diurnal variation, but potentially through improved renal function [62]. For donors, saw an average decrease in choroid thickness of  $-4.9 \pm 7.9\%$  ( $-5.3 \pm 7.7\%$  for area) one year post-transplant (figure 6.2). While we cannot rule out diurnal variation contributing to the reduced choroidal space, from a sub-clinical perspective the reduced choroidal measurements may reflect relative reduction in renal function, as observed in the eGFR levels in table 6.2.

In our linear mixed-effects models for transplant recipients we accounted for covariates to quantify the impact of myopia and diurnal variation ('approximate refraction' and 'daytime from baseline' in table 6.5, respectively). Interestingly, relative daytime proved to be a significant predictor variable for markers of renal function alongside the choroid in transplant recipients, while approximate refractive error did not. In every transplant recipient linear mixed-effects model, there was an opposite effect of relative daytime compared with the choroid — model coefficients for these covariates had opposite signs. This suggests that the size of the choroid is negatively offset such that the choroid thickness and choroid area in a follow-up scan taken later in the day (relative to the baseline acquisition time) is decreased in absolute value — and vice versa for follow-up scans taken earlier than the original baseline scan. This makes sense given we know that the choroid naturally decreases throughout the course of the day [53] from diurnal variation. Note however that the magnitude of effect (standardised model coefficient  $\hat{\beta}$ ) for 'daytime from baseline' is consistently lower in absolute value, relative to the model coefficient for choroid thickness and area in each transplant recipient linear mixed-effects model (table 6.5). Therefore, it remains important that daytime be recorded to account for within-patient choroidal fluctuation in downstream longitudinal analysis, or image acquisition should be performed around the same time of day so as not to diminish any longitudinal signal present in the data.

Depending on whether dialysis had been performed prior to transplant, and whether the recipient's new kidney is able to start making urine immediately, patients may need to undergo (haemo)dialysis post-transplant [285], which can

contribute to dynamic fluid shifts during and after transplantation [65]. However, we have not been able to account for any change in body fluid volume post-transplant and what impact that may have had on the choroid. Cho, et al. [63] found that the choroid thinned after haemodialysis, and that this was more pronounced in those patients with diabetes mellitus (exclusion criteria for transplant recipients in this analysis). However, eligibility criteria assumed that the renal transplant was functional, so temporary dialysis may only have lasted a few of days. Thus, while we cannot account for any potential choroidal fluctuation in response to post-operative dialysis in the short-term, it is unlikely these fluid shifts would have any long-term effects on body fluid volume and the choroid.

eGFR is a high-level aggregate measurement of kidney function, derived from taking into account demographic factors such as age, sex and how much waste is in the blood via creatinine. Serum creatinine and urea are excreted by the kidney and are natural waste products. Thus, in the case for eGFR, higher values indicate good kidney function, and vice versa for serum creatinine and urea. Therefore, the significant correspondence observed between improved kidney function (increased eGFR, lowered creatinine and urea) and increased choroidal measures over time further indicates the potential for the choroid to reflect systemic microvascular injury in the context of kidney health. The progressively increasing choroidal area and thickness over time may indicate the progressive return to normal of its perfusion, which is influenced by its autonomic supply, a regulatory system which was overactive during end-stage CKD [59].

Our results are consistent with the literature. Aksoy, et al. [21] similarly found progressively increased choroidal thickness one week and one month post-transplant in recipients. Moreover, Farrah, et al. [16] recently conducted a study with a larger cohort (of which this dataset is a subsample of) in OCT chorioretinal image analysis in CKD patients and similarly observed significant choroidal thinning in the presence of CKD, which was indeed reversible after kidney transplantation. Moreover, they also reported long-term, gradual choroidal thinning in healthy kidney donors, also consistent with this work. Thus, high-resolution and non-invasive OCT scans of the choroidal microvas-

## 6.1. The OCTANE Study: Choriorretinal changes in a longitudinal cohort of kidney donor and recipients

---

culature may be informative in the monitoring and prediction of renal injury [16], but a larger and more diverse cohort is needed to further statistically validate these initial findings.

In the donor cohort, we observed eGFR to be at approximately 66% of baseline function, which increased very marginally to approximately 75% one year post-transplant (table 6.2). This is mostly consistent with current literature on kidney function after unilateral nephrectomy [286], which report immediate post-nephrectomy kidney function to be 50% of baseline, which marginally rises to 60% as a result of compensatory renal hypertrophy. Moody, et al. [287] reported that healthy kidney donors thus had a greater risk of cardiovascular mortality, observing higher levels of microalbuminuria. While Mule, et al. [288] reported choroidal thinning with lower levels of eGFR and microalbuminuria, the longitudinal change in an average donor's choroid from our sample did not reflect the immediate decline and subsequent plateau consistent with their eGFR levels, and neither through linear mixed-effects modelling, ultimately suggesting a more complex and very likely non-linear relationship between these variables, if any at all.

It may be that the immediate inflation of the choroid in the first 4 weeks after kidney donation could be a result of a prompt systemic, cardiovascular response to the unilateral nephrectomy, as it is within the first month in a donor's recovery period that compensatory renal hypertrophy takes place [289, 290]. The relative change after 8 weeks post-transplant showed gradual choroidal thinning, which was consistent with Mule et al. [288], and indeed with Choi and Kim [62], who cautiously suggest that a decrease in choroid thickness is associated with a decline in renal function but its increase is associated with renal hypertrophy, albeit in patients with diabetic retinopathy.

There were a number of limitations in this study. Firstly, our two cohorts were limited in sample size and not all participants or clinical variables were collected at every time point. Moreover, of the 20 donors and 16 recipients included for clinical evaluation, only 11 donors and 8 recipients were followed up 1 year post-transplant, which potentially introduced bias into our clinical evaluation, and contributed to a reduction of statistical power. Furthermore, using a sub-cohort for clinical evaluation could potentially introduce selection bias when



estimating associations between the choroid and renal function. Therefore, further studies with a larger and more complete dataset would permit a higher confidence regarding the significance of our analyses. While Farrah, et al. [61] has been able to do so at a larger-scale and confirm these initial findings, image analysis was conducted manually. Thus, reproducing these findings using automated systems which are now readily available would be an obvious next step.

Additionally, from an image analysis perspective, we have only provided an analysis on regional quantities of the choroid and their links to CKD treatment and kidney transplantation. With the introduction of improved automated systems for characterising the choroid, such as Choroidalizer, further work would aim to quantify any relationships with the choroidal vasculature. This would enable measurement of choroid vessel area and CVI, which would provide a more representative measurement of the choroid. However, Aksoy, et al. [21] did not observe CVI to change significantly pre- and post-transplant, which may be due to improper use of Niblack's method, or simply that the increase in choroidal area was of a similar magnitude of increase as choroidal vessel area, thus stabilising reported CVI values pre- and post-transplant. This necessitates the need to also report choroid vessel area alongside CVI, as highlighted in chapters 3 and 5.

#### 6.1.4.1 Outputs

In this section, there has been one publication output:

- **Burke, Jamie**, Dan Pugh, Tariq Farrah, Charlene Hamid, Emily Godden, Thomas J. MacGillivray, Neeraj Dhaun, J. Kenneth Baillie, Stuart King, and Ian J.C. MacCormick. “*Evaluation of an automated choroid segmentation algorithm in a longitudinal kidney donor and recipient cohort.*” *Translational Vision Science & Technology* 12, no. 11 (2023): 19-19.

#### **6.1.4.2 Executive summary**

In this section, we assessed 22 healthy kidney donors and 23 patients requiring renal transplantation up to 1 year post-transplant. We measured choroidal thickness and area and estimated associations between choroidal measurements and markers of renal function (estimated glomerular filtration rate (eGFR), serum creatinine and urea). We observed significant choroidal thickening in recipients 1-year post-transplant and the choroid corresponded to improved eGFR, serum creatinine and urea levels. In donors, we observed choroidal inflation in the first 4 weeks after the unilateral nephrectomy, followed by significant thinning 1-year post-transplant.

In recipients, the relationship between choroid and kidney reported here adds further evidence to the growing literature around the choroid reflecting microvascular injury in CKD. In donors, initial inflation of the choroid could possibly be related to compensatory renal hypertrophy, while long-term thinning may be linked to long-term impairment of renal function. Future work should aim at a more complete analysis with a larger and more diverse sample in order to increase statistical power.

## 6.2 The PREVENT Dementia Study: Association between choroidal microvasculature and Alzheimer's disease risk

### 6.2.1 Introduction

Dementia is a multi-factorial neurological disorder characterised by progressive cognitive decline that significantly impacts daily functioning and quality of life. As the global population ages, the prevalence of dementia continues to rise, posing a significant public health challenge [291, 292]. While the pathophysiology of dementia is complex, accumulating evidence suggests that genetic and vascular factors play a crucial role in its aetiology and progression [292]. Genetic factors such as the apolipoprotein E4 (APOE4) allele [293], and family history of dementia [294, 295], have been identified as important determinants of dementia risk.

The retina is part of the central nervous system [33], which means the retinal microvasculature might reflect changes to similar vessels in the brain. Their similar embryonic origin and cellular composition has been leveraged previously to understand brain health through the lens of the ocular microvasculature [67, 296]. While the primary goal of OCT has often been to image and analyse the cross-sectional retinal layers (a *neuronal* measure and marker of neurodegeneration), advances in EDI-OCT [11] now permit exploration of possible associations between the choroid and brain diseases. In comparison to the retinal circulation, the choroidal circulation has significantly greater blood flow and perfusion pressure [297, 298] and is innervated by the central autonomic network [31].

However, there have been conflicting reports of choroidal changes in established Alzheimer's disease in older populations [70, 299, 300, 301, 71, 72] raising questions about the underlying mechanisms and stages of the disease that might influence choroidal vasculature. One possibility is that genetic factors like APOE4 and family history of dementia (FH) may play a role. To our knowledge, only one study [14] directly assessed the relationship between

## 6.2. The PREVENT Dementia Study: Association between choroidal microvasculature and Alzheimer's disease risk

---

choroidal measures and genetic risk (APOE4) in cognitively normal, old-age individuals but found no significant evidence of differences between carriers and non-carriers. Choroidal changes at the very beginning of the Alzheimer's disease trajectory remain unexplored. This could provide valuable insights into the prognostic value of the eye in reflecting the underlying processes linking vascular pathology to prospective cognitive decline.

The *PREVENT Dementia study* [271, 302] investigates concurrent cerebral and retinal vascular changes in a cohort of mid-life individuals, half of whom are at increased risk of Alzheimer's disease. While most research on ocular microvascular changes in Alzheimer's disease focuses on older age groups [70, 299, 300, 301, 71, 72, 14], the PREVENT cohort presents a unique opportunity to examine the microvasculature in individuals who may be in the very early prodromal stage of disease. Accordingly, the study presented here aimed to explore the associations between the choroidal vasculature and dementia risk in cognitively healthy midlife adults at baseline.

### 6.2.2 Data

#### 6.2.2.1 Study population and image capture

The protocol for the PREVENT Dementia programme has been outlined previously [271, 302]. The study aimed to recruit participants aged between 40 and 59 from five different research institutes, and to include a significant portion with a family history of dementia (approximately 50%). The Edinburgh site recruited 224 participants, and 132 participants provided written informed consent to the retinal sub-study at baseline. The study adhered to the principles of the Declaration of Helsinki.

Eligibility criteria for the Edinburgh arm was competent and consenting adults recruited from the wider PREVENT study who had the capacity to manoeuvre themselves to the retinal imaging machines unaided and follow instructions to facilitate patient fixation during imaging at baseline. Exclusion criteria were those with a clinical diagnosis of dementia, those without capacity to consent at baseline or individuals not able to fully understand written and verbal English.

Exclusion criteria also included participants with current or previous ocular disease such as glaucoma, macular degeneration, diabetic retinopathy, uveitis, vitreous haemorrhage, ischaemic optic neuropathy, optic neuritis or other optic nerve diseases, or those who had undergone previous ocular surgery such as cataract surgery or retinal surgery.

OCT capture was performed with the SD-OCT Heidelberg Engineering Spectralis OCT1 Standard Module (Heidelberg Engineering, Heidelberg, Germany) for both eyes. Single, horizontal- and vertical-line scans were taken using active eye tracking with an ART of 100 to reduce speckle noise and covered a 30° angle, resulting in approximately a 9 mm FOV. Posterior pole volumetric scans were also collected, with 61 parallel line-scans covering the macula and spaced approximately 125 microns apart without EDI mode activated, also using active eye tracking but with an ART of 9. B-scans with an SNR index <15 were excluded, according to the OSCAR-IB criteria [226] for retinal OCT quality assessment.

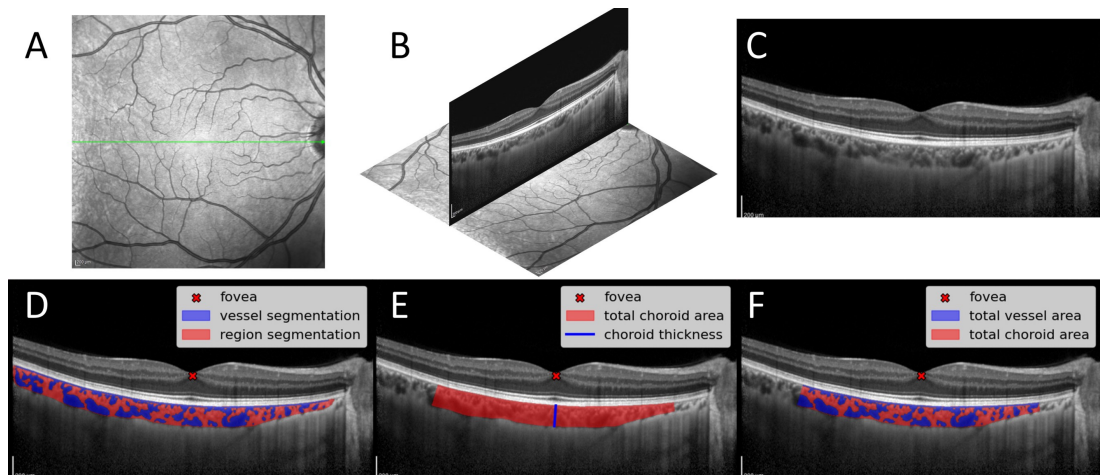
### 6.2.2.2 Image analysis

Measurements of fovea-centred subfoveal choroid thickness, region area, vessel area and CVI were computed using Choroidalyzer (chapter 5), which detected the fovea, and segmented the choroid region and vessels in every OCT B-scan. Figure 6.3 shows an OCT B-scan with the corresponding localiser SLO (figure 6.3(A – C)) and segmentations of the choroidal space, vessels and fovea using Choroidalyzer (figure 6.3(D – F)). Details on how these metrics were computed are described in section 1.7.1. Choroid measurements were computed for three distinct ROIs defined according to the different ETDRS circular grids (diameters 1 mm, 3mm and 6 mm) [197].

### 6.2.2.3 Statistical analysis

Participants in the PREVENT cohort [271] were prospectively stratified into three risk groups based on APOE4 presence and family history (FH) by design: ‘high’ risk was defined where APOE4 and FH were both present, ‘medium’ risk

## 6.2. The PREVENT Dementia Study: Association between choroidal microvasculature and Alzheimer's disease risk



**Figure 6.3:** Example derivation of choroidal measurements. (A) En face SLO image of the retina with location of B-scan shown in green. (B) OCT B-scan cross-section overlaid onto the en face SLO image. (C) OCT B-scan. (D) Automatic segmentation of region, vessels and foveal pit detection using Choroidalyzer. (E) Measurements of choroid thickness and choroid area and (F) vessel area and CVI.

where either APOE4 or FH were present and 'low' risk where neither were present.

As this was an exploratory study with a relatively small sample size, our primary objective was to report descriptive statistics and investigate potential trends and associations. We first graphed choroidal measurements, testing for normality using the Shapiro-Wilks test, and performed univariate hypothesis testing on the mean differences between choroidal measures stratified for each risk for dementia. Alongside univariate statistical tests with individual risk markers, we assumed a biological ordering between risk groups and used ordinal logistic regression with each choroidal measure in turn as a single covariate in an unadjusted model, as well as in adjusted models controlling for age, sex, and mean arterial blood pressure [303, 304, 305]. Unadjusted models were estimated so as to retain maximum statistical power in a relatively low sample size.

Because much of the posterior pole data was unusable due to the combined effect of acquisition without EDI mode and low ART, and to minimise the risk

of spurious associations from analyses of multiple ROIs per eye, we only selected choroidal measurements from horizontal-line scans with a 1.5mm fovea-centred ROI with a preference for the left eye. The left eye tends to have a thinner choroid than the right [306], helping overcome issues with image quality from optical signal degradation. Any missing values were removed at the point of analysis. Analyses were performed using Python's statsmodels (version 0.14.0).

## 6.2.3 Results

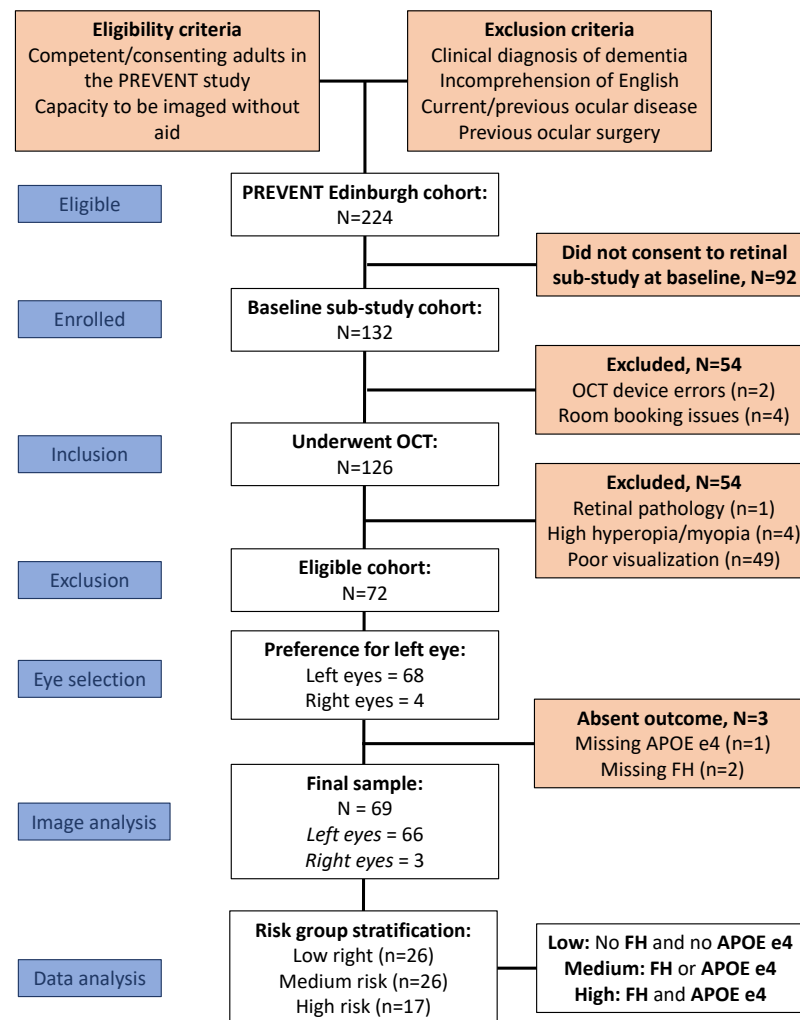
### 6.2.3.1 Participants

Figure 6.4 describes the derivation of our final sample (N=69). Of the 224 participants enrolled at the Edinburgh arm of the PREVENT dementia study, 132 consented to the Edinburgh site's retinal sub-study. Of 132 consented, 126 participants underwent OCT capture (OCT device/computer error, N=3; room booking issues, N=4). Furthermore, as the primary aim of OCT in the study protocol was to only image the retina (which does not require EDI mode activated), EDI-OCT capture was not collected for many of the earlier participants, which rendered the choroids of some eyes too poor quality for analysis (N=49).

As described in section 1.5.3, EDI-OCT is an important functionality in SD-OCT image capture for improved visualisation of the choroid. Without it, particularly for older devices and hardware, can render the choroids of some eyes too poor quality for analysis. Figure 6.5 shows a selection of 8 eyes which were excluded from the analysis in this cohort due to imaging without EDI mode, which ultimately led to only the retinal layers being visible. In figure 6.6 we show paired examples where EDI mode was toggled on and off.

Alongside the 49 participants excluded due to poor choroid visualisation, participants with retinal pathology (N=1) were excluded from our analysis to prevent abnormal retinae from confounding our results. While ocular pathology was part of the exclusion criteria, the image technician identified one participant with macular degeneration in both eyes during acquisition and after recruitment. Moreover, participants who were highly myopic/hyperopic ( $\geq \pm 6$

## 6.2. The PREVENT Dementia Study: Association between choroidal microvasculature and Alzheimer's disease risk



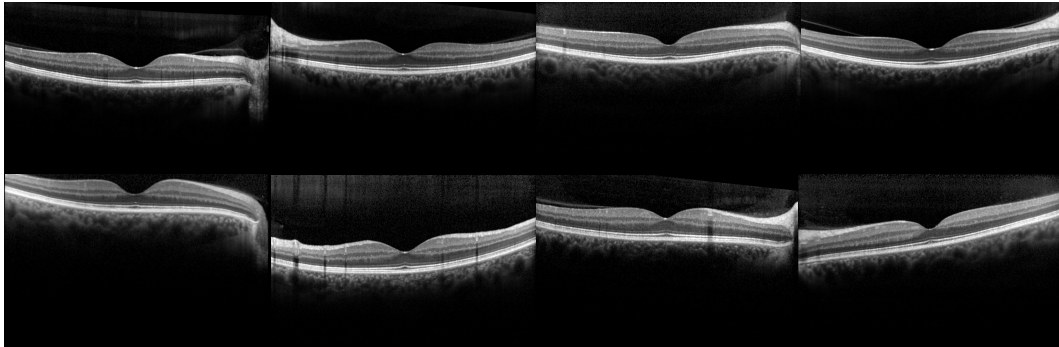
**Figure 6.4:** Sample derivation flowchart.

dioptries, measured using the scan focus on the Heidelberg imaging device [307]) were also excluded (N=4), since extreme myopia/hyperopia is known to shrink/enlarge the choroidal space due to the elongation/compression that the eye undergoes [308].

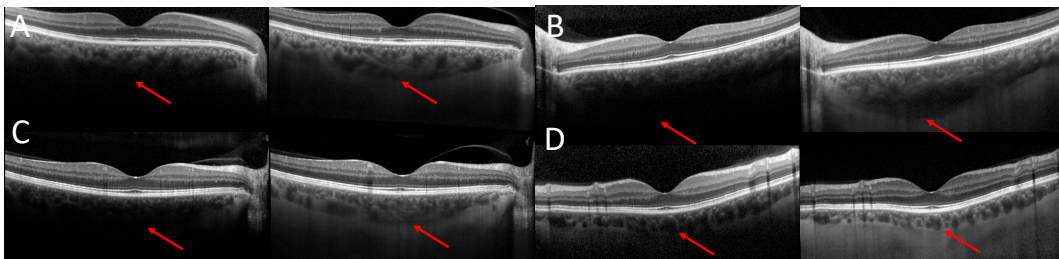
After selecting predominantly left eyes (68 left eyes, 4 right eyes where left was not available), and excluding participants with missing APOE4 and FH (N=3), this left a final sample of 69 participants (66 left eyes, 3 right eyes), stratified into low (N=26), medium (N=26) and high (N=17) risk groups.

Table 6.6 presents the demographics of the entire PREVENT Dementia Study





**Figure 6.5:** Example OCT B-scans from the PREVENT retinal sub-study which were captured without EDI mode activated.



**Figure 6.6:** Paired examples of OCT B-scans where EDI mode was toggled off and on, with red arrows indicating the lowest point of the Choroid-Sclera junction in each instance.

cohort, stratified into whether a participant was included into our final sample or not. We ran standard statistical tests to compare the distributions between those participants included and excluded from the final sample and found no evidence of significant differences between baseline demographics and clinical variables. We did not run any statistical tests between the entire cohort and the final sample as this would violate the independence assumption underlying most statistical tests, due to overlapping samples.

Demographics and study variables are summarised in table 6.7. The overall population had a mean age of 51.6 years (standard deviation, SD = 5.71) and a slight female majority (N=41, 59.4%). Mean arterial blood pressure was slightly higher in the low-risk group compared to the medium and high-risk groups which was statistically significant (low vs. medium,  $P=0.001$ ; low vs.

## 6.2. The PREVENT Dementia Study: Association between choroidal microvasculature and Alzheimer's disease risk

|                                 | Overall          | Included in final sample |                  | P-Value |
|---------------------------------|------------------|--------------------------|------------------|---------|
|                                 |                  | No                       | Yes              |         |
| <b>Participants</b>             | 224              | 155                      | 69               |         |
| <b>Age</b> (years)              | 51.347 (5.532)   | 51.255 (5.467)           | 51.551 (5.707)   | 0.718*  |
| <b>Sex</b> (female)             | 128 (57.658)     | 87 (56.863)              | 41 (59.420)      | 0.834†  |
| <b>Hypertension</b>             | 23 (10.407)      | 15 (9.804)               | 8 (11.765)       | 0.840†  |
| <b>Blood pressure</b> (mmHg)    | 128.205 (14.733) | 127.121 (14.204)         | 130.594 (15.678) | 0.119*  |
| <b>BMI</b> (kg/m <sup>2</sup> ) | 28.892 (6.207)   | 28.782 (6.415)           | 29.143 (5.741)   | 0.680*  |
| <b>Smoking</b>                  |                  |                          |                  | 0.704†  |
| Current                         | 10 (4.545)       | 8 (5.298)                | 2 (2.899)        |         |
| Ex                              | 78 (35.455)      | 54 (35.762)              | 24 (34.783)      |         |
| None                            | 132 (60.000)     | 89 (58.940)              | 43 (62.319)      |         |
| <b>Diabetes</b>                 | 8 (3.620)        | 5 (3.289)                | 3 (4.348)        | 0.707†  |
| <b>Risk factors</b>             |                  |                          |                  |         |
| APOE4 status                    | 89 (40.455)      | 64 (42.384)              | 25 (36.232)      | 0.475†  |
| Family History                  | 104 (47.489)     | 69 (46.000)              | 35 (50.725)      | 0.614†  |

**Table 6.6:** Demographics and study variables of the entire PREVENT Dementia study cohort at the Edinburgh site, stratified by participant inclusion in this study's final sample. BMI (body mass index). Notes: All values are N (%) or Mean (standard deviation). Missing data (number of eyes, percentage): Sex (2, 0.01%), Age (2, 0.01%), Hypertension (3, 0.01%); Blood Pressure (3, 0.01%), BMI (4, 0.02%); Smoker (4, 0.02%), Diabetes (3, 0.01%), APOE4 (4, 0.02%) and Family History (5, 0.02%). \*, Two-sample *t*-test; †, Chi-squared test.

high,  $P=0.01$ ). No other evidence of significant differences in demographic or cardiovascular variables were observed between risk groups.

### 6.2.3.2 Choroidal variation

Figure 6.7 shows the histogram distributions for each choroidal measurement. Using the Shapiro-Wilks normality test, we found that each choroid measurement was approximately normal (test statistic (P-value) for choroid thickness, 0.984 (0.50); choroid area: 0.979 (0.31); CVI: 0.984 (0.52); choroid vessel area: 0.987 (0.72)).

Figure 6.8 shows box-plots and distributions of each choroidal measurement stratified by the presence of each individual risk marker for Alzheimer's disease. We observed that choroids had larger vascular tissue and appeared larger in area for APOE4 carriers and FH individuals, compared to non-carriers

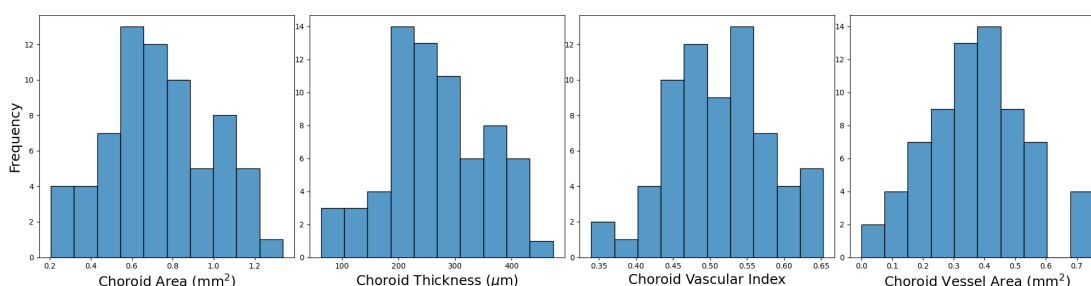
|                                 | Risk group      |                 |                 | Overall         |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|
|                                 | Low             | Medium          | High            |                 |
| <b>Participants</b>             | 26              | 26              | 17              | 69              |
| <b>Age</b> (years)              | 50.615 (6.32)   | 52.423 (4.54)   | 51.647 (6.42)   | 51.551 (5.71)   |
| <b>Sex</b> (female)             | 14 (53.9%)      | 17 (65.4%)      | 10 (58.8%)      | 41 (59.4%)      |
| <b>Hypertension</b>             | 4 (15.4%)       | 3 (11.5%)       | 1 (6.3%)        | 8 (11.8%)       |
| <b>Blood Pressure</b> (mmHg)    | 138.603 (13.02) | 124.654 (16.37) | 127.431 (13.73) | 130.594 (15.68) |
| <b>BMI</b> (kg/m <sup>2</sup> ) | 30.362 (6.87)   | 28.103 (5.27)   | 28.748 (4.23)   | 29.143 (5.74)   |
| <b>Smoking</b>                  |                 |                 |                 |                 |
| Current                         | 1 (3.9%)        | 1 (3.8%)        | 0 (0%)          | 2 (2.9%)        |
| Ex                              | 8 (30.8%)       | 7 (26.9%)       | 9 (52.9%)       | 24 (34.8%)      |
| Non                             | 17 (65.4%)      | 18 (69.2%)      | 8 (47.1%)       | 43 (62.3%)      |
| <b>Diabetes</b>                 | 1 (3.9%)        | 2 (7.7%)        | 0 (0%)          | 3 (4.3%)        |
| <b>Risk factors</b>             |                 |                 |                 |                 |
| APOE4 status                    | 0 (0%)          | 8 (30.8%)       | 17 (100.0%)     | 25 (36.2%)      |
| Family History                  | 0 (0%)          | 18 (69.2%)      | 17 (100.0%)     | 35 (50.7%)      |
| <b>Choroidal measures</b>       |                 |                 |                 |                 |
| Area (mm <sup>2</sup> )         | 0.68 (0.25)     | 0.71 (0.194)    | 0.87 (0.30)     | 0.74 (0.26)     |
| Thickness (μm)                  | 250.39 (84.48)  | 257.00 (78.37)  | 303.35 (105.56) | 265.93 (89.32)  |
| CVI                             | 0.477 (0.06)    | 0.533 (0.07)    | 0.528 (0.08)    | 0.511 (0.07)    |
| Vessel area (mm <sup>2</sup> )  | 0.307 (0.14)    | 0.364 (0.14)    | 0.459 (0.18)    | 0.366 (0.16)    |

**Table 6.7:** Demographics and study variables, stratified by risk group. BMI (body mass index); CVI (choroidal vascular index). All values are N (%) or Mean (standard deviation). Missing data (number of eyes, percentage): Hypertension (1, 1.5%); BMI (2, 2.9%); CVI (2, 2.9%); vessel area (2, 2.9%).

and those with no FH. These differences were supported by independent *t*-tests for choroid vessel area ( $P=0.022$ ,  $P=0.009$ ). This was also observed for choroid area ( $P=0.020$ ) and CVI ( $P=0.022$ ) for APOE4 carriers and FH individuals, respectively (Table 6.8).

Figure 6.9 presents the choroidal measurements plotted against the covariates chosen for ordinal logistic modelling (age, sex and mean arterial blood pressure). Choroidal measurements appeared to decrease weakly with age (figure 6.9, top) but the observed trend was much clearer for the low-risk group. Choroid vessel area and CVI appeared lower in males (figure 6.9, middle), but had very similar distributions for choroid area and thickness. For low-/medium-risk groups, we observed weak trends of smaller choroidal measures with increasing blood pressure (figure 6.9, bottom). However, in the high-risk group, these trends appeared much more pronounced, and blood pressure was lower in higher-risk groups in general as seen from table 6.7. These observed trends

## 6.2. The PREVENT Dementia Study: Association between choroidal microvasculature and Alzheimer's disease risk



**Figure 6.7:** Histogram distributions of each choroidal measure.

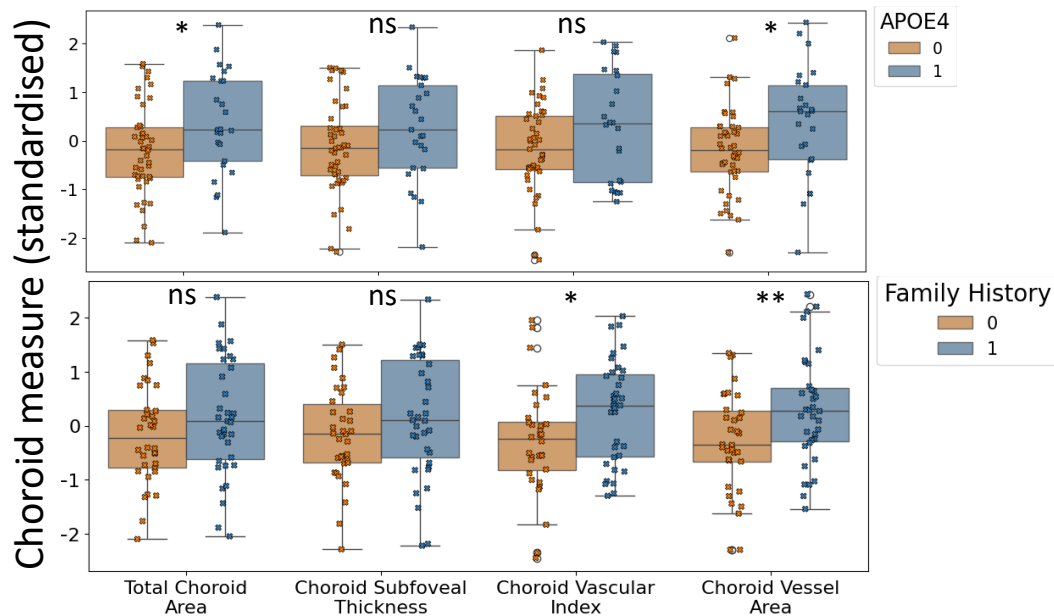
|                                | APOE4 carrier status       |                | P-Value      |
|--------------------------------|----------------------------|----------------|--------------|
|                                | Non-carrier                | Carrier        |              |
| N                              | 47                         | 25             |              |
| Area [mm <sup>2</sup> ]        | 0.67 (0.23)                | 0.83 (0.27)    | <b>0.020</b> |
| Thickness [ $\mu$ m]           | 248.40 (85.36)             | 288.64 (94.13) | 0.081        |
| CVI                            | 0.50 (0.06)                | 0.53 (0.08)    | 0.104        |
| Vessel area [mm <sup>2</sup> ] | 0.33 (0.14)                | 0.43 (0.18)    | <b>0.022</b> |
|                                | Family History of Dementia |                | P-Value      |
|                                | None                       | Yes            |              |
| N                              | 34                         | 36             |              |
| Area [mm <sup>2</sup> ]        | 0.69 (0.23)                | 0.79 (0.28)    | 0.122        |
| Thickness [ $\mu$ m]           | 252.03 (78.16)             | 283.50 (99.83) | 0.146        |
| CVI                            | 0.49 (0.07)                | 0.53 (0.07)    | <b>0.022</b> |
| Vessel area [mm <sup>2</sup> ] | 0.32 (0.15)                | 0.42 (0.16)    | <b>0.009</b> |

**Table 6.8:** Distribution of choroidal measures by dementia risk with P-values for *t*-tests. Values are reported as mean (SD).

in our sample justified our selected covariates when reporting adjusted ordinal logit models for regressing choroidal measurements on Dementia risk-group classification.

Figure 6.10 presents grouped box-plots of choroidal measurements and prospectively defined Alzheimer's disease risk groups, showing a biological gradient between increasing risk and increasing choroidal measurements, particularly for choroid vessel area.

After adjusting for age, sex, and blood pressure, all choroidal measurements were significantly associated with increased combined risk for Alzheimer's disease. Table 6.9 presents the odds ratios (OR) from the ordinal logistic models. Specifically, the difference between having no markers and at least one or



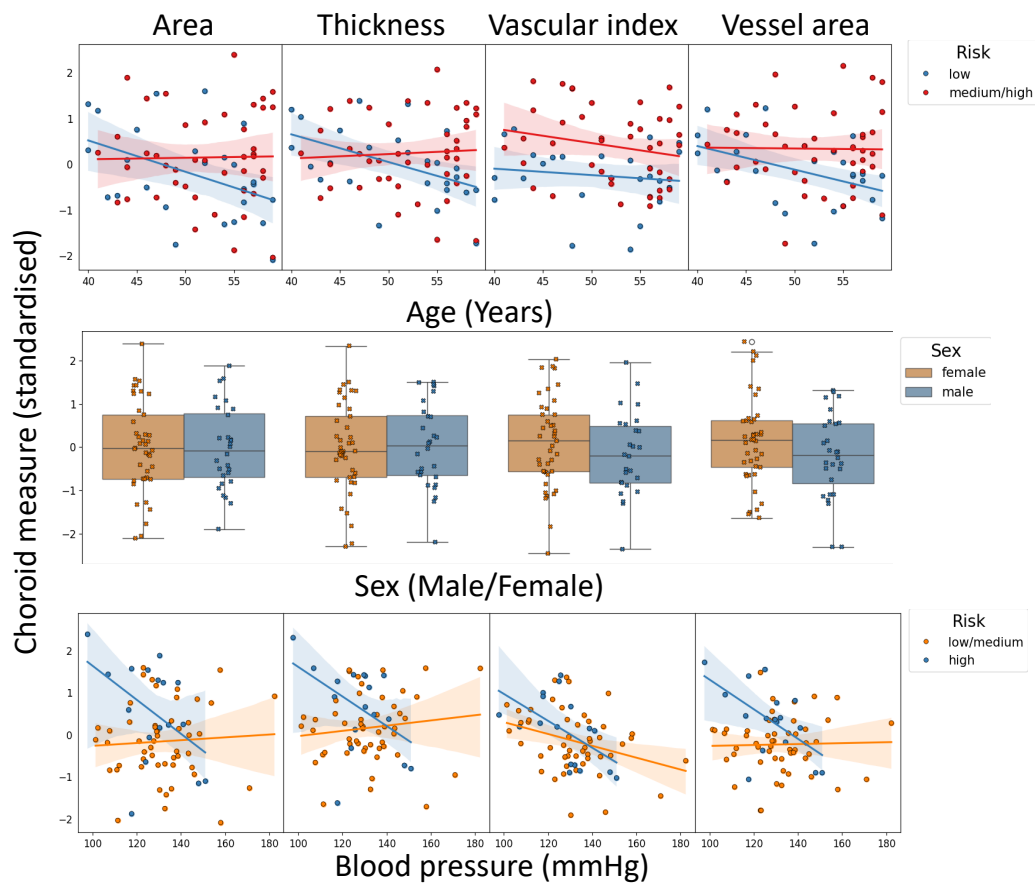
**Figure 6.8:** Box-plots showing the relationship between (standardised) choroidal measures and APOE4 (top) and FH presence (bottom). \*\*:  $P < 0.01$ ; \*:  $P < 0.05$ ; ns:  $P > 0.05$ .

both was significantly associated with increased choroidal area (OR per SD increase = 1.87, 95% CI:[1.12 – 3.11],  $P = 0.016$ ), choroidal thickness (OR per SD increase = 1.69, 95% CI:[1.03 – 2.79],  $P = 0.039$ ), CVI (OR per SD increase = 1.70, 95% CI: [1.02 – 2.82],  $P = 0.041$ ), and vessel area (OR per SD increase = 2.25, 95% CI:[1.33 – 3.80],  $P = 0.003$ ). The magnitude of effect for choroidal vessel area was the largest of all measures. Put another way, for every unit standard deviation increase in choroidal vessel area ( $0.16 \text{ mm}^2$ ), on average it is 2.25 times more likely that this individual has at least one marker of risk for Alzheimer’s disease (APOE4, FH) rather than none, after holding all other covariates fixed.

## 6.2.4 Discussion

We explored the association between the choroidal vasculature and risk factors for Alzheimer’s disease and observed a significant, positive association

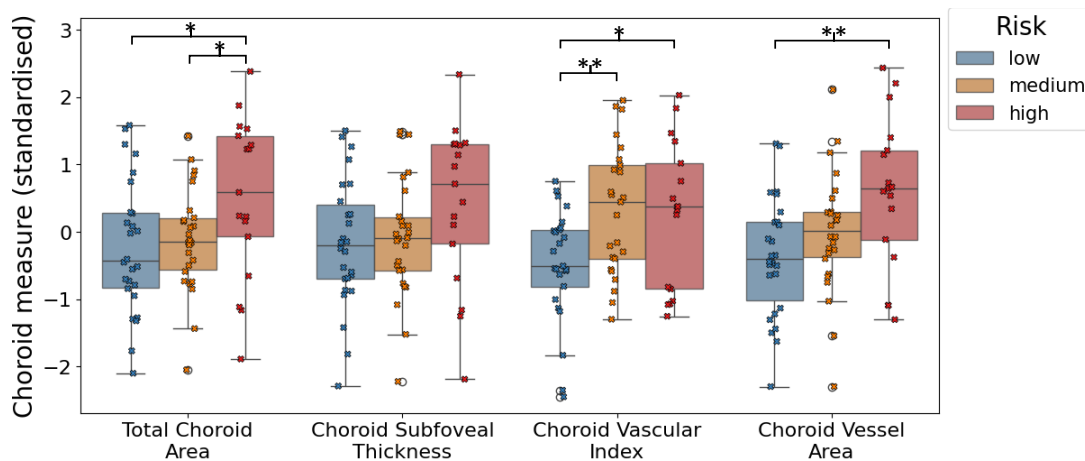
## 6.2. The PREVENT Dementia Study: Association between choroidal microvasculature and Alzheimer's disease risk



**Figure 6.9:** Observed trends between choroidal measures and chosen covariates for ordinal logistic modelling, with each choroidal measure listed column-wise, and each covariate row-wise. Age (top), sex (middle) and mean arterial blood pressure (bottom).

between each choroidal measure and degree of combined risk (APOE4, FH). To our best knowledge, this is the first exploration of the choroidal vasculature in cognitively healthy midlife adults at risk of Alzheimer's disease. Previous, related studies have focused only on retinal difference between groups [309].

In this exploratory study of asymptomatic people at risk of Alzheimer's disease, we observed trends which, independent of covariates, indicate an increased retinal choroid vasculature in participants who carry the APOE4 genotype or have a family history of dementia, relative to those who do not. This reached statistical significance in choroid vessel area and choroid area. Sub-



**Figure 6.10:** Boxplots showing the relationship between choroidal measures and risk groups, with individual data points overlaid in a swarm-plot. \*\*:  $P < 0.01$ ; \*:  $P < 0.05$ .

foveal choroid thickness is a one-dimensional point-source measurement, and is less robust in measuring the choroid relative to two-dimensional measurements such as area and CVI.

The relationship between CVI and vessel area/choroid area likely played a role in any significant, distributional differences observed between groups. Both choroidal area and vessel area were significantly higher in APOE4 carriers relative to non-carriers. Area and vessel area are directly and inversely proportional to CVI, respectively, and so these two measures directly contribute toward estimated associations with CVI. Thus, the increase in CVI (mean increase of 0.03) did not reach significance likely because of the greater mean increase in region area (mean increase of  $0.16 \text{ mm}^2$ ) relative to vessel area (mean increase of  $0.10 \text{ mm}^2$ ). For individuals with a family history of dementia, we observed a smaller magnitude of increase in total choroid area relative to APOE4 carriers (mean increase of  $0.10 \text{ mm}^2$ ). Put in the context of the significant increase observed in choroidal vessel area (mean increase of  $0.10 \text{ mm}^2$ ), this very likely contributed to the significant increase observed in CVI (mean increase of 0.04).

While these observations highlight the interplay between these representative choroidal measurements, they ultimately suggest that choroidal vessel area



## 6.2. The PREVENT Dementia Study: Association between choroidal microvasculature and Alzheimer's disease risk

|                   | Choroid area      |               | Choroid thickness |               | CVI               |               | Choroid vessel area |               |
|-------------------|-------------------|---------------|-------------------|---------------|-------------------|---------------|---------------------|---------------|
|                   | OR (95% CI)       | P-value       | OR (95% CI)       | P-value       | OR (95% CI)       | P-value       | OR (95% CI)         | P-value       |
| <b>Covariates</b> |                   |               |                   |               |                   |               |                     |               |
| Age               | 1.33 (0.83, 2.15) | 0.2395        | 1.31 (0.81, 2.11) | 0.2664        | 1.34 (0.83, 2.18) | 0.2288        | 1.36 (0.84, 2.21)   | 0.2160        |
| Blood Pressure    | 0.5 (0.3, 0.85)   | <b>0.0095</b> | 0.49 (0.29, 0.83) | <b>0.0076</b> | 0.57 (0.34, 0.98) | <b>0.0415</b> | 0.5 (0.29, 0.85)    | <b>0.0100</b> |
| Sex (Male)        | 1.13 (0.42, 3.01) | 0.8140        | 1.09 (0.41, 2.9)  | 0.8687        | 1.35 (0.49, 3.7)  | 0.5662        | 1.43 (0.52, 3.91)   | 0.4907        |
| Choroid measure   | 1.87 (1.12, 3.11) | <b>0.0162</b> | 1.69 (1.03, 2.79) | <b>0.0390</b> | 1.7 (1.02, 2.82)  | <b>0.0408</b> | 2.25 (1.33, 3.8)    | <b>0.0025</b> |
| <b>Intercept</b>  |                   |               |                   |               |                   |               |                     |               |
| Low/Medium        | 0.6 (0.31, 1.16)  | 0.1317        | 0.61 (0.32, 1.17) | 0.1344        | 0.67 (0.35, 1.27) | 0.2194        | 0.65 (0.33, 1.26)   | 0.2031        |
| Medium/High       | 1.91 (1.35, 2.7)  | <b>0.0002</b> | 1.88 (1.33, 2.65) | <b>0.0003</b> | 1.87 (1.31, 2.66) | <b>0.0005</b> | 2.0 (1.42, 2.83)    | <b>0.0001</b> |

**Table 6.9:** Odds ratios (with 95% confidence intervals and corresponding P-values) for each ordinal logistic regression model, predicting risk group with each choroidal measure in turn. All models were adjusted for age (standardised), blood pressure (standardised) and sex. To provide stable confidence intervals in logistic models, all choroidal measures were standardised. P-values lower than 0.05 are in bold type.

had the strongest signal among measurements, and that the vasculature itself might be the driving factor for increased choroidal area and thickness, rather than surrounding extravascular tissue (for our sample). Yet again, like in previous chapters, we highlight the importance of reporting choroidal vascular measures like area rather than only CVI.

Encouragingly, measurement error and diurnal variation are unlikely to have had a major effect on our results. The differences in mean choroid measurements between APOE4 carriers and non-carriers (thickness, 53 microns; area, 0.19 mm<sup>2</sup>; CVI, 0.05; and vessel area, 0.15 mm<sup>2</sup>), were greater than expected from approximate choroidal fluctuation due to diurnal variation [52, 54, 310, 55, 56, 57], (thickness, 30 microns; area, 0.035 mm<sup>2</sup>; vascular index, 0.015; vessel area, 0.02 mm<sup>2</sup>). Moreover, our results are greater than the threshold expected to exceed Choroidalizer's estimated measurement error for single-line OCT B-scans (thickness, 11.5 microns; area, 0.05 mm<sup>2</sup>; CVI, 0.013). This was also the case for individuals with and without a family history of dementia (table 6.8).

The ordinal models give us an indication of the size of the observed associations found in our sample and can be easily interpreted. For example, consider two males of equal age, measuring the same blood pressure. If the first male's choroidal vascular tissue was greater in area by at least 0.16 mm<sup>2</sup>, they are



on average over two times more likely to have at least one (or both) of the risk markers, assuming the comparative individual had no risk markers, thus placing them in a higher risk group for developing Alzheimer's disease in later life. Whether this effect exists within the general population should be investigated in cohorts designed to test this hypothesis.

The APOE4 variant is a major risk factor for Alzheimer's disease [293] and has been associated with a wide range of negative health-related outcomes or features including cardiovascular disease [311], blood-brain barrier dysfunction [312] and inflammation [313]. However, there is also some debate around its potential role in reducing/increasing risk for ocular disorders like age-related macular degeneration [314, 315], a disease of the choroid. Thus, it could be that these observed associations are simply reflecting an effect of the APOE4 allele (or related genotypes) on the choroid itself, rather than a direct association with brain health.

An alternative may be related to autonomic dysfunction, of which the APOE4 variant has been associated with [316]. Lohman, et al. [316] reports that APOE4 carriers exhibit central autonomic dysfunction in early-stage Alzheimer's disease, including the parasympathetic control of cardiovascular functions [317]. Dillon et al. [318] found that mildly cognitively impaired patients were on average 5.6 times more likely to have autonomic dysfunction, showing significant parasympathetic deficits, which may be involved in the pathogenesis of hypotension in dementia. Given that choroidal blood flow is regulated to some extent by autonomic input, particularly parasympathetic-mediated vasodilation, and systemic hypotension does not cause sympathetic input to the choroid [31], dysregulation of this autonomic supply could have played a role in the increased choroidal measurements observed in our sample. While this would correspond with larger calibre, the choroid is a highly heterogeneous vascular compartment, and the current device resolution and cross-sectional nature of OCT capture prevents us from definitively concluding whether increased vasculature corresponds to larger calibre, or simply more vessels.

Despite its promise as a marker of disease risk for Alzheimer's disease [319], we found just one other study that used EDI-OCT choroid measures to investigate asymptomatic individuals with known APOE4 status and family history

## 6.2. The PREVENT Dementia Study: Association between choroidal microvasculature and Alzheimer's disease risk

---

[14]. However, in contrast with our findings, Ma et al. [14] found no evidence of significant differences in choroidal measures between carriers and non-carriers. One reason for the conflicting findings could be that Ma et al. [14] reported choroidal measurements in pixel units. By contrast, we converted choroidal measurements from pixel units into physical units (i.e. microns) according to the unique axial and lateral pixel length-scales corresponding to each B-scan, so that measurements across the population could be compared more appropriately. Moreover, the average participant age in Ma et al. [14] was around 20 years older than ours, suggesting that age differences may also play a role in the contrasting findings.

EDI-OCT has previously been leveraged to investigate the link between diagnosed dementia and the choroid, with mixed results. Some have found *reduced* choroidal thickness in Alzheimer's disease and mild cognitive impairment compared to controls [300, 70, 299, 320]. Bulut et al. [300] and Gharbiya et al. [70] showed significant choroidal thinning in Alzheimer's disease, but selected both eyes per participant for their statistical analyses without accounting for intra-participant correlation, violating a core statistical principle. Gharbiya et al. [70] and Bayhan et al. [299] analysed smaller cohorts than ours and performed measurement of the choroid manually, which has the potential for significant measurement error [106]. Recently, Kwapong et al. [320] found reduced choriocapillaris density in OCT-Angiography imaging of early-age onset Alzheimer's disease patients compared with healthy controls, and was able to discriminate between such groups. However, compared to the PREVENT Dementia study, the Kwapong et al. [320] cohort was older and already presenting with symptoms sufficient for clinical diagnosis— whether the same approach can be leveraged at the preclinical stage of Alzheimer's disease still requires further validation.

On the other hand, other works have found *increased* choroidal measurements in Alzheimer's disease [72, 71]. Asanad et al. [71] found an increased choroidal thickness in post mortem human tissue in Alzheimer's disease patients relative to controls using histopathology, while Robbins et al. [72] found increased choroidal area and vessel area in Alzheimer's disease compared with controls in EDI-OCT. However, these measurements were also reported in pixel

space. Moreover, the authors paradoxically observed significantly decreased choroidal thickness in their sample, but these distances were measured as vertical lines without accounting for choroidal skew or curvature. Thus, where our results fit into the disease trajectory of Alzheimer's disease is yet to be realised.

However, the Dementia study is ongoing, collecting longitudinal data on this mid-life cohort. This allows future temporal modelling that may further validate our findings. Nevertheless, our preliminary results in this relatively small sample suggest choroidal differences may exist between prospective risk groups for Alzheimer's disease, and may potentially take place long before the onset of disease.

Strengths of the current study include the use of Choroidalyzer as a tool to extract measurements from the choroidal vasculature using a standardised, fovea-centred ROI across the whole cohort, reporting values in physical units rather than pixel units [14], aiding interpretability and helping mitigate spurious associations. Another strength is the uniqueness of the PREVENT cohort, permitting characterisation of the cerebral and ocular microvasculature in a cohort of asymptomatic individuals, at increased risk and 20 years younger than most individuals in similar studies.

There were some limitations to this study. Firstly, of 132 consented at baseline, 49 participants had too poor choroid visualisation for image analysis and all OCT volume data was unusable because the OCT acquisition protocol was originally designed for retinal OCT and hence did not prioritise EDI imaging. Therefore, much of the data collected was unsuitable for analysis due to poor choroid visualisation, resulting in a low sample size. Thus, statistical analysis was limited by low power due to small sample size, and we cannot be certain that selection bias did not play a role in our results. Additionally, participants were primarily white in ethnicity, therefore, generalisability of our findings to other ethnic groups or the wider population may be limited. Thus, while our associations could relate to mechanistic links between the choroid microcirculation and pathogenesis of Alzheimer's disease, more research is needed. However, future temporal modelling of this ongoing, longitudinal study may enable more impactful results by tracking evolution from mere risk, to cognitive impairment and potential clinical diagnosis of Alzheimer's disease.

## 6.2. The PREVENT Dementia Study: Association between choroidal microvasculature and Alzheimer's disease risk

---

### 6.2.4.1 Outputs

In this section, there has been one publication output which is currently under peer-review at *Alzheimer's & Dementia: Diagnosis, Assessment and Disease Monitoring*. The manuscript has been uploaded to medRxiv as a pre-print:

- **Burke, Jamie**, [Samuel Gibbon](#), Audrey Low, Charlene Hamid, Megan Reid-Schachter, Graciela Muniz-Terrera, Craig W. Ritchie, Baljean Dhillon, John T. O'Brien, Stuart King, Ian J.C. MacCormick and Thomas J. MacGillivray "Association between choroidal microvasculature in the eye and Alzheimer's disease risk in cognitively healthy midlife adults." medRxiv 2024.08.27.24312649 (2024).

### 6.2.4.2 Executive summary

We investigated associations between choroidal measurements obtained with EDI-OCT, with risk factors (APOE4 and parental family history) for Alzheimer's disease in 69 healthy, mid-life individuals from the PREVENT Dementia study. We found that the choroidal space and vascular tissue was significantly larger in participants who were both APOE4 carriers and had a positive family history of dementia, and this increase was consistent and ordinal between risk groups (vascular tissue in those with neither risk factor < only one risk factor < both risk factors). In our sample, a 0.16 mm<sup>2</sup> increase in choroidal vascular area was associated with a two-fold increase in the chance of having one or more markers of Alzheimer's disease risk, compared with none. Thus, our results suggest a potential link between the choroidal vasculature and risk of Alzheimer's disease. However, these findings are exploratory and should be replicated in a larger and more diverse sample.

## 6.3 D-RISC II: A phase two evaluation of retinal choroid biomarkers for shock resuscitation in critically ill adults

### 6.3.1 Introduction

Circulatory shock is a non-specific pathophysiological state constituting central organ hypoperfusion and can result from hypovolaemic, cardiogenic, obstructive, distributive, or endocrine aetiologies. Shock represents a large proportion of morbidity and mortality in acute care settings, with Sakr, et al. [321] estimating in 2010 that 1 in 3 patients experience some kind of circulatory shock throughout the course of intensive care. More recently, circulatory shock has been a key driver of multi-organ failure and cause of COVID-19 mortality within intensive care [322].

Current management of circulatory shock targets improved vascular perfusion through inotropes and vasopressors, with fluid resuscitation to replete intravascular volume. These early goal-directed therapies are effective at restoring measurements of large vessel function, such as systemic blood pressure, but assume that measurements of macrovascular function reflect the microvascular environment and thus informs vital organ perfusion and clinical outcomes.

However, macrovascular function does not necessarily correspond to microvascular perfusion [323]. This macrophysical perspective of treatment may explain why early goal-directed therapies to restore large vessel parameters appear to be ineffective [324, 325, 326, 327] and have the potential to mislead clinicians on the state of the patient and increase risk of mortality. Byrne, et al. [328] showed that fluid resuscitation in an ovine model increased morbidity through altering the function and permeability of the vascular endothelium. Furthermore, Maitland, et al. [329] highlighted the potential for overly aggressive fluid resuscitation to lead to negative patient outcomes after reporting the increased 48-hour mortality in critically ill children with impaired perfusion after fluid boluses.

While macrovascular function does not necessarily correspond to microvascu-

### 6.3. D-RISC II: A phase two evaluation of retinal choroid biomarkers for shock resuscitation in critically ill adults

---

lar perfusion, microvascular function within central organs *is* directly relevant to perfusion. It's been observed that microvascular perfusion of sublingual capillaries appears to be a better predictor of multi-organ failure in haemorrhagic shock than serum lactate or systolic blood pressure [330]. Moreover, strategies targeting microcirculatory dysfunction by protecting the endothelium and preventing vascular leakage may prove more effective therapeutic pathways for patients [331, 332, 333]. With fluid resuscitation playing a key role in sepsis care, being able to monitor potential microvascular injury by consistently reassessing patient response could help guide treatment at the patient-level, reduce mortality and prevent unnecessary deaths from fluid overload.

However, the microvascular circulation is a highly heterogeneous environment and it is generally difficult to measure the microvascular perfusion of central organs consistently in humans, making interpretation in the clinical setting challenging. The retinal and choroidal circulations are an exception, since they are both directly accessible with non-invasive optical imaging. Recent examination of the retina using OCT highlight its potential as a non-invasive, safe, feasible and valid approach to observe and quantify intraocular structures in supine, critically ill patients [75, 78].

Moreover, the choroidal vascular bed represents a novel microvascular sample site which is preserved as part of the physiological response to shock. This is because it is part of the central nervous system and adapts dynamically to systemic physiology [31]. Recent pre-clinical evidence [77] in a septic rat model reported reduced choroidal blood flow during shock, and we already observed it's potential to reflect the perfusion of other central organs, such as the kidney (section 6.1).

Thus, we hypothesised that choroidal microvascular blood flow may represent the microvascular environment in vital organs such as the brain, heart and lungs in patients experiencing circulatory shock. Accordingly, we aimed to establish the feasibility of bedside OCT within a heterogeneous group of critically ill intensive care and high dependency unit patients, assess whether the intra-ocular vasculature is responsive to progression within a patient's disease course, and identify any physiological correlates with this change.

## 6.3.2 Data

### 6.3.2.1 Study population

Between March 2023 and July 2023, patients were recruited to the Direct Retinal Imaging for Shock Resuscitation in Critically Ill Adults- II (D-RISC-ii) study — a single-centre cohort study within tertiary intensive (ITU) and high-dependency care units (HDU) at the Royal Infirmary of Edinburgh. Local research and developmental approval (2022/0190) and national ethical approval (22/SS/0055) were obtained and tenets of the Declaration of Helsinki followed. Patients were screened for eligibility by the departmental research nurses. Eligible patients were aged 16 years or older and receiving care corresponding to Intensive Care Society Level 2 or 3 [334]. Exclusion criteria included anticipated survival time < 24 hours, pregnancy, obvious ocular pathology or facial trauma, previous clinical exclusion for pupillary dilatation, if OCT equipment was not available, if study enrolment would be disruptive to care provision, or if consent was not given. Due to ethical approval, follow-up imaging was only done within ITU/HDU. We aimed to recruit 15 participants to assess feasibility of choroidal imaging.

We recorded several clinical parameters prospectively from clinical records. Only clinical variables which had at most 25% missing values at baseline were used for downstream analysis and are listed in table 6.10, with units in square brackets and completeness at baseline and follow-up in round brackets.

### 6.3.2.2 Image acquisition

We sampled patients at two time points, with follow-up examination occurring 12 – 72 hours post-baseline. We performed retinal imaging as soon as feasibly possible once recruited. In each instance, and with preference to the right eye, retinal imaging was done after pharmacological pupil dilation (Tropicamide 1% weight-by-volume) 30 minutes before examination to facilitate visualisation [335]. Pupil dilation within a shallow ocular anterior chamber can cause acute angle closure, so anterior chamber depth was screened before dilation using the oblique illumination test (OIT).

### 6.3. D-RISC II: A phase two evaluation of retinal choroid biomarkers for shock resuscitation in critically ill adults

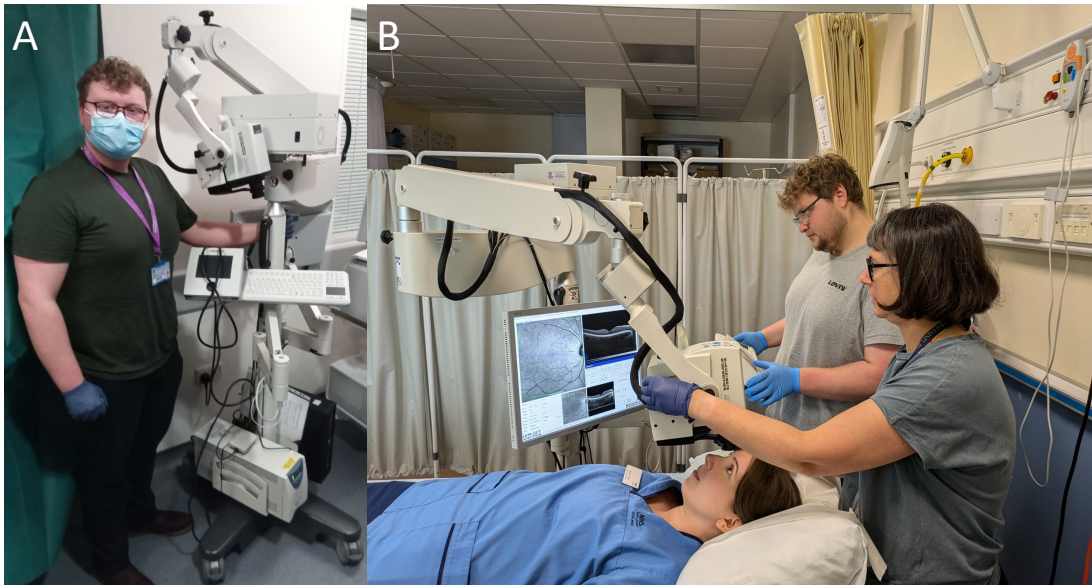
|   |           |  |  |
|---|-----------|--|--|
| Age (Years)<br>FU:100%]                               | [BL:100%, | Fluid 24 Hour Output (ml)<br>[BL:83%, FU:100%]               | Neutrophils (x10 <sup>3</sup> /mL)<br>[BL:100%, FU:100%]         |
| Albumin (g/L)<br>FU:100%]                             | [BL:83%,  | Fraction Inspired Oxygen (%)<br>[BL:92%, FU:83%]             | Oxygen saturation (%)<br>[BL:100%, FU:100%]                      |
| APACHE2 (N)<br>FU:67%]                                | [BL:75%,  | Glasgow Coma Score (N)<br>[BL:100%, FU:100%]                 | Peripheral Capillary Refill Time (seconds)<br>[BL:100%, FU:100%] |
| Bicarbonate (mmol/L)<br>FU:83%]                       | [BL:92%,  | Glucose (mmol/L) [BL:92%,<br>FU:83%]                         | Peripheral Temperature (Celsius)<br>[BL:100%, FU:100%]           |
| Calcium (mmol/L)<br>FU:50%]                           | [BL:75%,  | Haematocrit (%) [BL:100%,<br>FU:100%]                        | Phosphate (mg/dL) [BL:92%,<br>FU:100%]                           |
| Creatinine (mmol/L)<br>FU:100%]                       | [BL:100%, | Haemoglobin (g/L) [BL:100%,<br>FU:100%]                      | Platelets (x10 <sup>3</sup> /mL) [BL:100%,<br>FU:100%]           |
| Cumulative Fluid Balance (ml)<br>[BL:100%, FU:100%]   |           | Heart Rate (bpm) [BL:100%,<br>FU:100%]                       | Potassium (mmol/kg) [BL:100%,<br>FU:100%]                        |
| Cumulative Fluid Input (ml)<br>[BL:100%, FU:100%]     |           | Highest Lactate (24 Hour)<br>(mmol/L) [BL:92%, FU:83%]       | Respiratory Rate (bpm)<br>[BL:100%, FU:100%]                     |
| Cumulative Fluid Output (ml)<br>[BL:100%, FU:100%]    |           | Lowest Perfusion Pressure of<br>Oxygen (kPa) [BL:92%, FU:0%] | Sodium (mmol/L) [BL:100%,<br>FU:100%]                            |
| Diastolic Blood Pressure<br>(mmHg) [BL:100%, FU:100%] |           | Lymphocytes (x10 <sup>3</sup> /mL)<br>[BL:100%, FU:100%]     | Systolic Blood Pressure<br>(mmHg) [BL:100%, FU:100%]             |
| Fluid 24 Hour Balance (ml)<br>[BL:83%, FU:100%]       |           | Magnesium (mmol/L) [BL:75%,<br>FU:67%]                       | Urea (mmol/L) [BL:100%,<br>FU:100%]                              |
| Fluid 24 Hour Input (ml)<br>[BL:83%, FU:100%]         |           | Mean Arterial Blood Pressure<br>(mmHg) [BL:100%, FU:100%]    | White Cell Count (x10 <sup>3</sup> /mL)<br>[BL:100%, FU:100%]    |

**Table 6.10:** Clinical variables collected and used in downstream statistical analysis. Individual elements are listed as ‘variable name (units) [completeness at baseline (BL), completeness at follow-up (FU)]’. APACHE2, Acute Physiological and Chronic Health Evaluation.

All scans collected were macula-centred with a 30-degree FOV with B-scans covering approximately 9 mm<sup>2</sup>. All image capture was collected with EDI mode toggled on to improve visualisation of the Choroid-Sclera boundary and vasculature. Active eye tracking with ART was also used to help improve image quality, longitudinal registration and reduce speckle noise.

Our imaging protocol is enumerated below in order of image capture: A fovea-centred, horizontal-line B-scan with an ART of 100 followed by a vertical-line B-scan with similar parameters. A fovea-centred posterior pole volume scan, consisting of equally spaced B-scans approximately 240 microns apart using





**Figure 6.11:** (A) Image of the author (Jamie Burke) with the OCT Spectralis FLEX module in tertiary care ITU/HDU unit at the Royal Infirmary of Edinburgh. Image taken by Dr Ian J.C. MacCormick. (B) Demonstration of the OCT Spectralis FLEX operated by two trained technicians, the author and Charlene Hamid, on a healthy volunteer, Rachel Woods. One technician operated the camera head, while the other operated the computer. Image taken by Rosie Keane.

an ART of 9. Dependent upon patient cooperation, the volume scan either contained 31 B-scans covering a  $9.0 \times 6.6$  mm FOV, 25 B-scans covering a  $9.0 \times 5.3$  mm FOV or 25 B-scans covering a  $5.3 \times 5.3$  mm FOV.

OCT scans were collected using the SD-OCT Heidelberg Spectralis FLEX OCT1 module (Heidelberg Engineering, Heidelberg, Germany). The Spectralis FLEX module is a mobile version of the table-mounted Spectralis Standard module, with the camera head mounted on a boom to facilitate retinal imaging of supine patients. This was operated by two trained technicians: one to position the camera, and one to operate the computer. In some cases a third person helped to stabilise the head and eyelids. OCT imaging was performed by three trained operators, Charlene Hamid, (C.H.), Jamie Burke (the author) (J.B.) and Emily Godden (E.G.).

Figure 6.11(A) shows the FLEX folded into resting position next to the author

### 6.3. D-RISC II: A phase two evaluation of retinal choroid biomarkers for shock resuscitation in critically ill adults

---

after successful OCT capture of the first patient recruited to D-RISC-ii. In panel (B) we show a demonstration of the FLEX in use on a healthy volunteer being handled by two persons: the author operating the computer during acquisition, and Charlene Hamid operating the camera head.

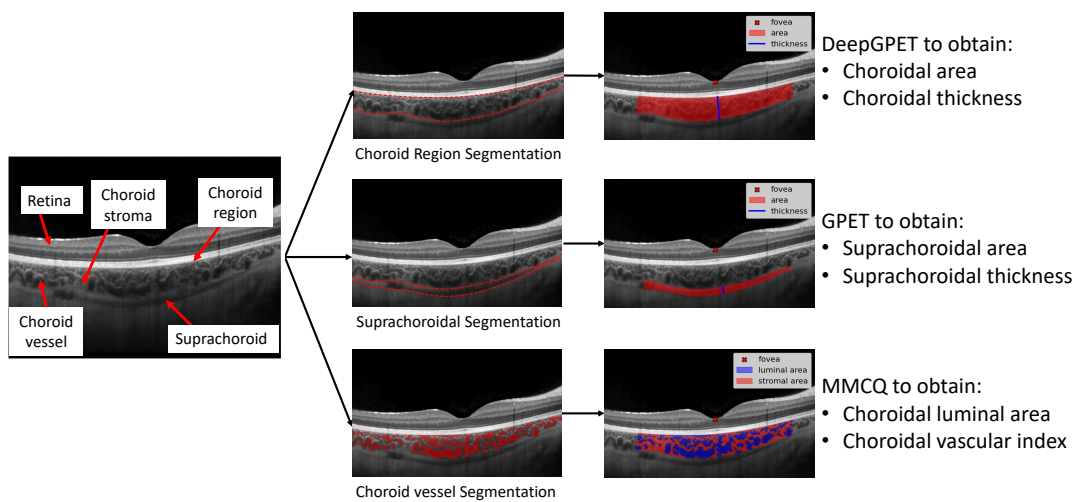
#### 6.3.2.3 Image analysis

Figure 6.12 shows the anatomical layers of interest and the methods used and measurements taken for this study. Delineation of the choroidal space in OCT B-scans was performed using DeepGPET (chapter 4) and the supra-choroid, if observed, was delineated using GPET (chapter 2) which allowed human-assisted pixel-level segmentation (due to its size and co-location to the Choroid-Sclera boundary). Choroidal vessels were segmented using MMCQ (chapter 3) — Choroidalizer (chapter 5) was not developed until after data collection and image analysis. For each OCT B-scan, we measure subfoveal choroid thickness, area, vessel area and CVI using a fovea-centred ROI. The definition of the choroid, suprachoroid and choroidal measurements have been described previously in section 1.7.1.

#### 6.3.2.4 Statistical analysis

Qualitative feasibility was assessed by collecting several process and safety outcomes to help guide future protocols. These include anterior chamber depth estimated using the OIT [335], application of topical mydriatic agent for pupil dilation, adverse event reporting, patient positions, and case-specific feedback from the image team for each patient imaged. We also leveraged metadata from the image data to quantitatively assess feasibility in terms of acquisition time, image quality index (reported from the imaging device's built-in SNR estimate as an unbounded value starting at 0 and plateauing around 50 (Heidelberg Engineering, Heidelberg, Germany) [307]) and the number of repeated scans needed for protocol completion.

We sought to generate hypotheses around how patients' physiological and clinical status might correlate with choroidal measurements. For this analysis, we



**Figure 6.12:** An OCT B-scan with landmarks of interest annotated, and pixel-level anatomical segmentations and measurements of the choroidal and supra-choroidal spaces, and choroidal vessels overlaid.

selected only horizontal-line, fovea-centred OCT B-scans. A 4 mm, fovea-centred ROI was chosen to include all patients, as larger ROIs precluded challenging patients from analysis.

We reviewed the distributions of choroidal and clinical data graphically both at baseline and across time points and explored potential pairwise associations. Since these associations are exploratory and hypothesis-generating we did not report P-values associated with the correlation coefficients.

At baseline, we graphed choroidal metrics against clinical variables using scatter plots and fitted curves to plots that suggested linear or non-linear trends. After graphing, we estimated associations using Pearson’s linear correlation coefficient ( $r$ ), or using Spearman’s monotonic correlation coefficient ( $s$ ) to assess non-linear association.

For those successfully imaged at follow-up, we estimated the overall pairwise associations at the patient-level using repeated measures Pearson correlation coefficient ( $r_{rm}$ ) [282, 283], accounting for variability within and across patients.

### 6.3.3 Results

#### 6.3.3.1 Participants

Figure 6.13 summarises patient recruitment. 21 patients were screened and found eligible for the study and 15 were recruited (declined consent, N=5; ITU/HDU discharged, N=1). Of 15 recruited, 14 underwent OCT capture (end-of-life care, N=1). Follow-up capture was attempted on 6 patients, with 8 patients unavailable (ITU/HDU discharged, N=4; end-of-life care, N=1; withdrew consent, N=1; poor baseline imaging, N=2).

Demographics of participants who had retinal imaging are reported in Table 6.11. At baseline, the cohort had a mean age of 56.2 years, with 71% male. All patients were receiving Level 3 care for a mean of 9.1 days. Elective surgical patients contributed approximately half of the cohort (7/15). Mean 24-hour fluid balance was +460ml with a haemoglobin of 95.2g/L (115 – 180), albumin of 18.3 g/L (normal range 36 – 47) and APACHE2 score 20.4 (5.9). Mean time to follow-up was  $46.3 \pm 31.7$  hours.

#### 6.3.3.2 Feasibility

Table 6.12 presents case-specific feasibility data. Baseline imaging was attempted in 14/15 recruited participants (the exception was owing to clinical deterioration and end-of-life care) and follow-up imaging attempted in six of these. Cumulatively, the imaging protocol was attempted twenty times, with eighteen of these (18/20, 90%) producing satisfactory outputs for analysis. In particular, baseline imaging was successful in 12 of 14 (86%) participants and follow-up imaging was successful in all six attempts (100%). All recruitment and imaging took place in ITU. There were no adverse events from imaging, and case-specific comments largely cite patient difficulty fixating or camera head-related issues due to in-situ ventilators. However, these did not negatively impact image acquisition, implying there was no significant issue for those whose protocol was successfully completed.

Figure 6.14 presents bar plots on acquisition metadata related to each suc-

## Chapter 6. Applications of choroidal image analysis in OCT image sequences

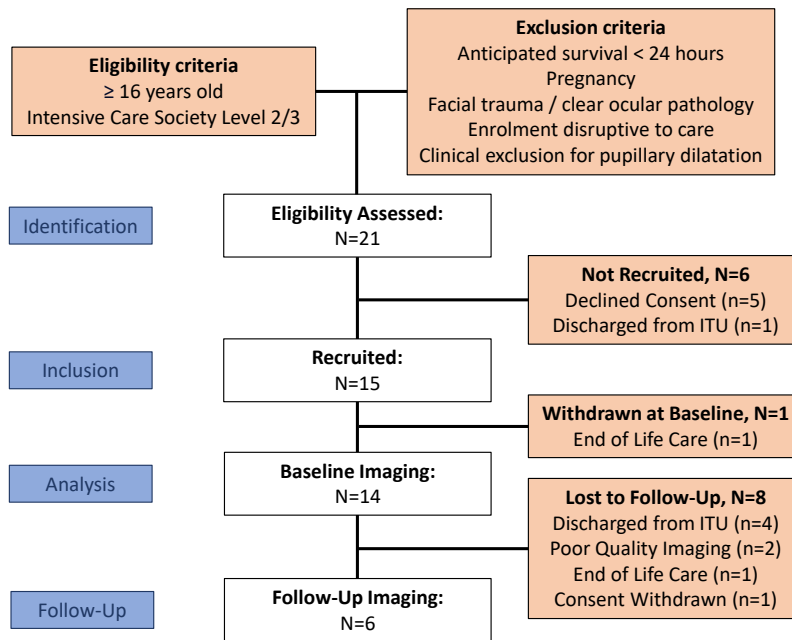
| Patient Characteristics                                  | Time point    |              | Laboratory Reference¶ |
|--|---------------|--------------|-----------------------|
|  | Baseline      | Follow-Up    |                       |
| N (Patients)   | 14            | 6            |                       |
| Age (Mean, Years):                                       | 56.2±15.3     | 47.0±19.0    | -                     |
| Sex (male)   | 10 (71)       | 3 (50)       | -                     |
| Body Mass Index (kg/m <sup>2</sup> ):                    | 28.4±5.9      | 27.3±7.3     | -                     |
| APACHE2* Score at Critical Care Admission                | 20.4±5.9      | 19±6.6       | -                     |
| Days Since Escalation to Level 2 or Level 3 Care**       | 9.1±8.8       | 11.7±11.0    | -                     |
| Proportion of Patients Receiving Level 3 Care**          | 14 (100)      | 6 (100)      | -                     |
| Primary Diagnosis on Admission                           |               |              |                       |
| Cardiac Arrest:  | 1 (7)         | 0 (0)        | -                     |
| Type 1 Respiratory Failure†                              | 2 (14)        | 2 (34)       | -                     |
| Overdose   | 1 (7)         | 1 (17)       | -                     |
| Toxic Shock Syndrome                                     | 1 (7)         | 1 (17)       | -                     |
| Acute Biliary Disease‡                                   | 2 (14)        | 1 (17)       | -                     |
| Elective Igor-Lewis Oesophagectomy                       | 2 (14)        | 0 (0)        | -                     |
| Elective Whipple's Procedure                             | 1 (7)         | 1 (17)       | -                     |
| Elective Liver Transplantation                           | 3 (21)        | 0 (0)        | -                     |
| Elective Thoraco-aortic Aneurysm Repair                  | 1 (7)         | 0 (0)        | -                     |
| <b>Indicators of Circulatory Function</b>                |               |              |                       |
| Heart Rate (min-1)                                       | 87.9±16.9     | 87.8±18.2    | -                     |
| Mean Arterial Pressure (mmHg)                            | 84.0±15.3     | 87.6±14.3    | -                     |
| 24 Hour Fluid Balance (mL)                               | 460.4±1931.2  | -366.5±560.1 | -                     |
| Cumulative Fluid Balance Since Treatment Escalation (mL) | 5376.8±8461.5 | 6423±11896.1 | -                     |
| <b>Predictors of Circulatory Function</b>                |               |              |                       |
| Highest Lactate 24-Hours Preceding Sampling (mmol/L)     | 1.54±1.00     | 0.98±0.30    | 0.5-1.6               |
| Creatinine Preceding Sampling (mmol/L)                   | 123.6±83.3    | 117.2±136.7  | 50-111                |
| Haemoglobin Preceding Sampling (g/L)                     | 95.2±20.1     | 92.7±21.3    | 115-180               |
| Haematocrit Preceding Sampling                           | 0.28±0.06     | 0.28±0.06    | 0.36-0.52             |
| Albumin Preceding Sampling (g/L)                         | 18.3±4.2      | 17.5±3.8     | 36-47                 |
| <b>Receipt of Organ Supportive Therapies</b>             |               |              |                       |
| Receipt of Invasive Mechanical Ventilation               | 4 (29)        | 0 (0)        | -                     |
| FiO <sub>2</sub> Preceding Sampling                      | 31.0±19.5     | 23.8±5.6     | -                     |
| SaO <sub>2</sub> Preceding Sampling                      | 96.7±2.2      | 98.0±1.4     | -                     |
| Receipt of Vasopressor Therapy                           | 3 (21)        | 1 (17)       | -                     |
| Receipt of Intra-Aortic Balloon Pump                     | 6 (43)        | 2 (33)       | -                     |
| Receipt of Renal Replacement Therapy                     | 0 (0)         | 0 (0)        | -                     |
| Receipt of Intravenous Sedation§                         | 3 (21)        | 1 (17)       | -                     |
| Glasgow Coma Scale Preceding Sampling (Median, IQR)      | 15, 3         | 15, 0        | -                     |
| <b>Choroid measurements (Median, IQR)</b>                |               |              |                       |
| Subfoveal choroid thickness (µm)                         | 249 (134)     | 309 (150)    | -                     |
| Area (mm <sup>2</sup> )                                  | 0.82 (0.62)   | 1.15 (0.68)  | -                     |
| Luminal area (mm <sup>2</sup> )                          | 0.36 (0.35)   | 0.60 (0.39)  | -                     |
| CVI  | 0.44 (0.05)   | 0.52 (0.07)  | -                     |
| <b>Suprachoroidal measurements (Median, IQR)</b>         |               |              |                       |
| Subfoveal suprachoroidal thickness (µm)                  | 64 (34)       | 48 (26)      | -                     |
| Suprachoroidal area (mm <sup>2</sup> )                   | 0.22 (0.16)   | 0.21 (0.13)  | -                     |

**Table 6.11:** Patient summary characteristics at baseline and follow-up. All values are N (%) or Mean (standard deviation), unless specified otherwise. \*, Acute Physiology and Chronic Health Evaluation Score 2 (APACHE2); \*\*, Intensive Care Society Level 2 (HDU) or Level 3 (ITU) care; †, Type 1 Respiratory Failure: one case secondary to pneumonia (concomitant bacterial and viral); ‡, Biliary Disease: one case of sepsis secondary to cholangitis, one patient with acute kidney injury following cholecystitis; ¶, Biochemical reference ranges as indicated by local health board (NHS Lothian); §, Sedation including either intravenous propofol or clonidine infusion.

### 6.3. D-RISC II: A phase two evaluation of retinal choroid biomarkers for shock resuscitation in critically ill adults

| Patient ID**                      | Captured | GCS | Tropicamide  | Laterality | Position | Time (m:s)* | Operator\$ | SCS visible | Case-Specific Notes  |
|-----------------------------------|----------|-----|--------------|------------|----------|-------------|------------|-------------|--|
| Patients Imaged at Baseline: N=14 |          |     |              |            |          |             |            |             |  |
| 1                                 | Yes      | 4   | Applied      | Right      | Supine   | 20:27       | A          | Yes         | Sedated, unable to fixate, large supra-choroid, image artefacts. Camera position affected by ventilator in-situ.   |
| 2                                 | No       | 11  | Applied      | Right      | Upright  | N/A¶        | A, B       | No          | Unable to fixate, cooperate, or follow instructions. Camera position affected by ventilator in-situ.   |
| 3                                 | Yes      | 15  | Applied      | Right      | Upright  | 04:32       | A          | Yes         | Required suctioning of airway secretions during protocol. Camera position affected by ventilator in-situ.  |
| 4                                 | Yes      | 15  | Not applied§ | Right      | Supine   | 01:59       | B          | Yes         |  |
| 5                                 | No       | 3   | Applied      | Both       | Supine   | 13:32       | A          | No          | Sedated, required saline eye drops for dry eyes. Difficulty locating macula due to previous ophthalmic conditions (AMD). Camera position affected by ventilator in-situ.                 |
| 7                                 | Yes      | 15  | Applied      | Right      | Supine   | 02:53       | A          | Yes         | Oxygen face mask in-situ.  |
| 8                                 | Yes      | 15  | Applied      | Right      | Upright  | 05:44       | A          | Yes         | Camera position affected by nebuliser in-situ. Required suctioning mid-protocol. Bedside access blocked ipsilaterally.   |
| 9                                 | Yes      | 15  | Applied      | Right      | Upright  | 03:26       | A          | Yes         | Lack of head support due to upright positioning, lead to difficulty fixating.  |
| 10                                | Yes      | 15  | Applied      | Right      | Upright  | 02:19       | A          | Yes         | Patient hearing impaired with difficulty fixating.   |
| 11                                | Yes      | 10  | Applied      | Right      | Supine   | 07:15       | A          | No          | Required saline eye drops for dry eyes. Partially sedated. Highly myopic eyes, large area of peripapillary atrophy. Difficulty fixating. Camera position affected by ventilator in-situ. |
| 12                                | Yes      | 15  | Applied      | Right      | Upright  | 01:11       | B          | No          | Poor cooperating and fixation due to patient discomfort.   |
| 13                                | Yes      | 15  | Applied      | Right      | Upright  | 04:09       | A          | No          | Patient drowsy with difficulty fixating.   |
| 14                                | Yes      | 15  | Applied      | Right      | Supine   | 01:30       | A          | Yes         |  |
| 15                                | Yes      | 15  | Applied      | Right      | Supine   | 03:04       | A          | Yes         | Difficulty imaging deep choroid with enhanced depth OCT.   |
| Patients Imaged at Follow-Up: N=6 |          |     |              |            |          |             |            |             |  |
| 1                                 | Yes      | 11  | Applied      | Right      | Supine   | 28:40       | A          | Yes         | Sedated, unable to fixate, Camera position affected by ventilator in-situ  |
| 3                                 | Yes      | 15  | Applied      | Right      | Supine   | 05:27       | A          | Yes         |  |
| 4                                 | Yes      | 15  | Applied      | Right      | Supine   | 01:52       | A          | Yes         | Protocol completed despite OCT power supply issues.  |
| 8                                 | Yes      | 15  | Applied      | Right      | Supine   | 04:14       | B          | Yes         | Horizontal imaging repeated to improve capture quality.  |
| 9                                 | Yes      | 15  | Not applied§ | Right      | Supine   | 01:32       | A          | Yes         | OCT imaging facilitated by supine position.  |
| 14                                | Yes      | 15  | Applied      | Right      | Upright  | 01:04       | A          | Yes         |  |

**Table 6.12:** Feasibility data for OCT in critical care settings. SCS: supra-choroidal space, m:s, minutes:seconds. GCS, Glasgow Coma Scale. \*\* : Patient 6 was excluded before baseline imaging due to clinical deterioration and end of life care; \* : Time was the intra-session difference in acquisition time between first and last scan acquired; § : Patient right eye was clinically dilated, with no requirement for mydriasis. ¶ : Only one scan was collected so timing could not be estimated; \$ : Operators A were Charlene Hamid and Jamie Burke (the author), and B were Emily Godden and Jamie Burke (the author).



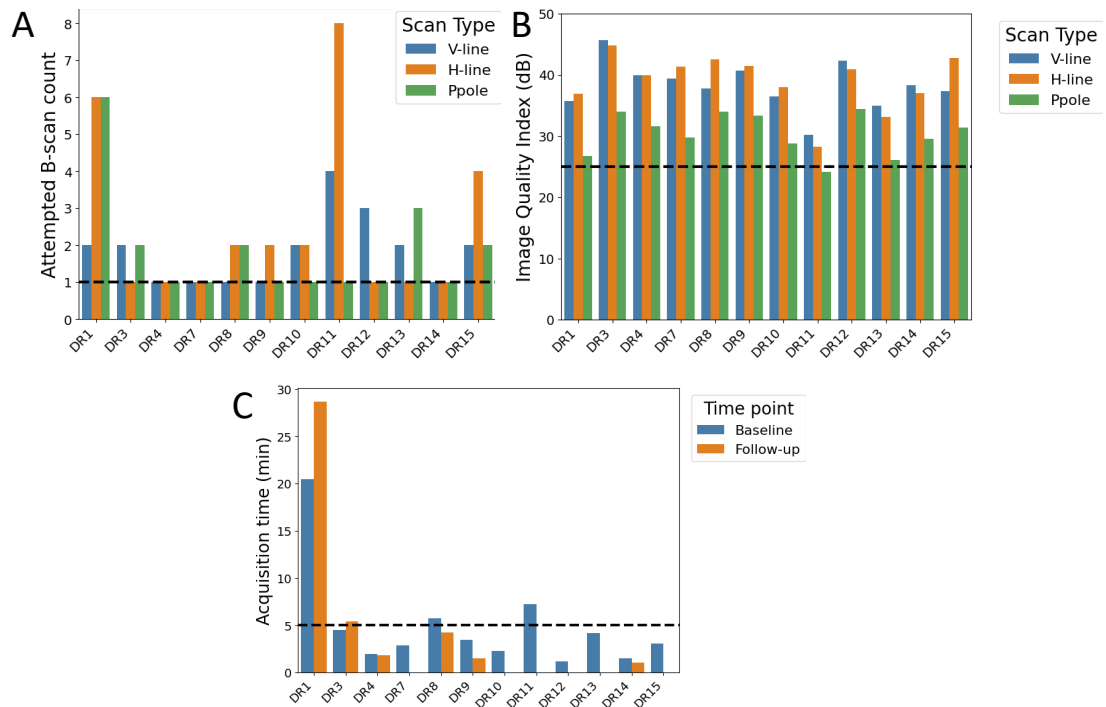
**Figure 6.13:** Flowchart summarising patient recruitment. From twenty-one eligible patients, fourteen patients were imaged at baseline and six at follow-up.

successfully imaged patient. Median (IQR) acquisition time per visit was 02:09 (3:30) minutes:seconds, with thirteen out of eighteen attempted imaging sessions completed within 5 minutes. Average (SD) scores for SNR across the cohort was  $38.9 \pm 4.6$ ,  $38.3 \pm 3.9$  and  $30.3 \pm 3.3$  for horizontal-line, vertical-line and posterior pole volume scans, respectively, with all image quality strengths exceeding the minimum score outlined by the manufacturer (25dB) [307]. Median (IQR) number of scan attempts needed per visit was 1.5 (1).

Two patients (ID2, ID5) failed the imaging protocol. For Patient ID2, who developed respiratory failure after a planned surgical admission, imaging was attempted in an upright position with topical mydriasis applied to the right eye. The patient, partially sedated, became agitated and was unable to fixate or follow instructions. A second imaging attempt on the fellow eye was unsuccessful, leading to their exclusion from the study with only one scan collected overall. Patient ID5 was highly myopic with advanced AMD in the right eye and vitreo-macular traction in the fellow eye, which made locating the macula centre a significant challenge. The patient's sedation further hindered cooperation,



### 6.3. D-RISC II: A phase two evaluation of retinal choroid biomarkers for shock resuscitation in critically ill adults



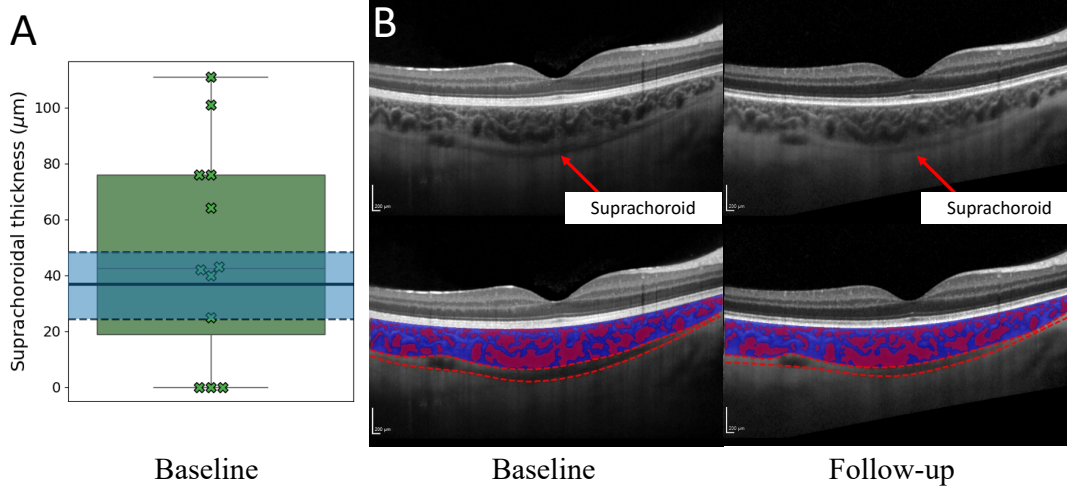
**Figure 6.14:** OCT acquisition information for each successfully imaged patient. (A) Acquisition time, stratified by scan type, with a black dashed line at 1. (B) OCT image quality index, stratified by scan type, with a black dashed line at 25dB. (C) Acquisition time between first and last B-scan for each time point, with a black dashed line at 5 minutes. Dashed lines are used as a visual aid.

and after 13.5 minutes of attempted acquisition, the protocol was abandoned. Any collected scans were excluded due to insufficient choroidal visualisation and poor macula-centring.

#### 6.3.3.3 Choroidal variation

Baseline suprachoroidal space thickness is shown in figure 6.15(A). The distribution of baseline suprachoroidal thickness had an interquartile range notably larger than previously reported in a healthy, elderly cohort aged 55 – 85 [18]. In nine of the twelve (75%) successfully imaged patients at baseline, the suprachoroidal space was visible and markedly enlarged, and was largest in the younger patients of our cohort — the suprachoroid of the youngest patient (aged 16) decreased in subfoveal thickness by 26 microns between baseline





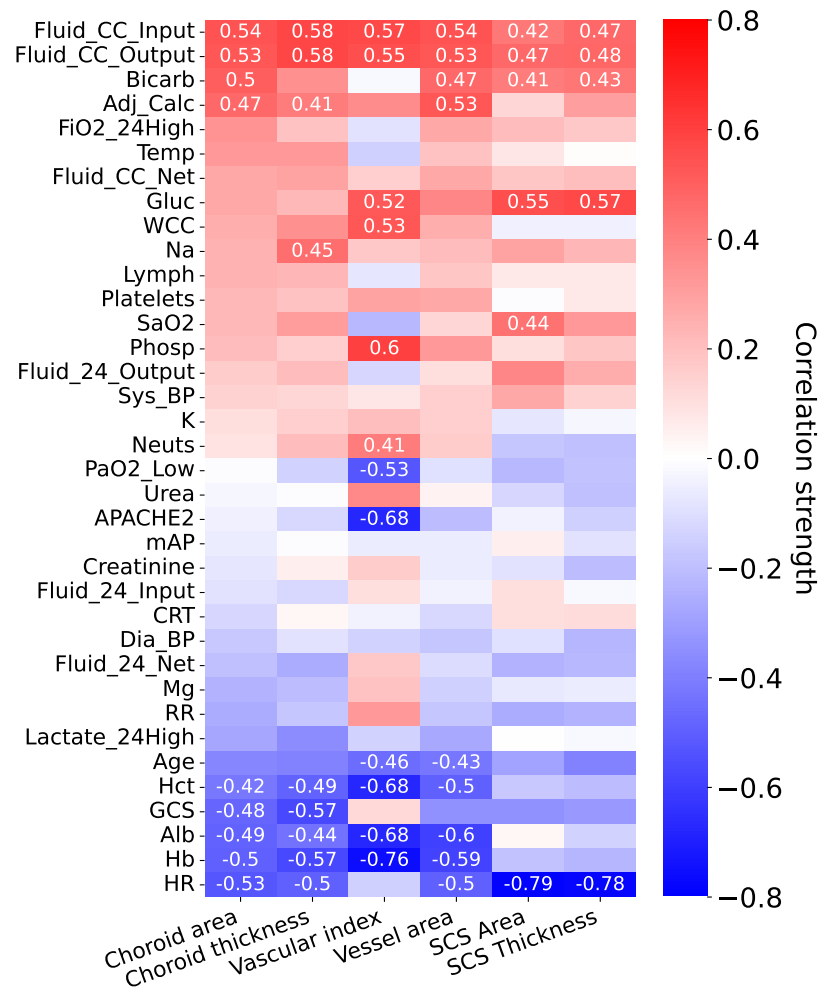
**Figure 6.15:** (A) Choroidal variation at baseline for suprachoroidal thickness, with shaded blue region representing median and IQR estimated from Yiu, et al. [18]. (B) Visible suprachoroidal space in youngest patient (age: 16) at baseline and at follow-up 22 hours later. (Top) suprachoroidal space highlighted by the red arrows. (Bottom) suprachoroidal space outlined in dotted red.

and follow-up after 22 hours (figure 6.15(B)). Of these nine, six had follow-up scans with the suprachoroid still visible which decreased from baseline median thickness of 64 microns to 48 microns (table 6.11).

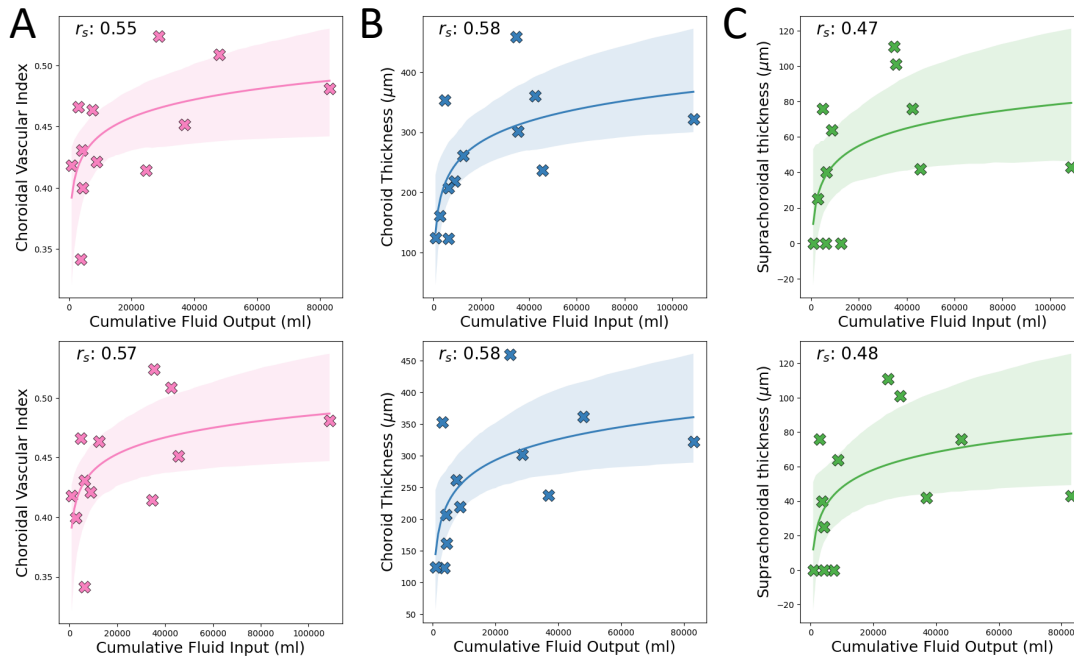
Figure 6.16 presents a correlation heat map between choroid and clinical measurements at baseline, and is colour coordinated according to the sign and strength of the correlation coefficient. All correlations represent Pearson's linear correlation coefficient,  $r$ , with exception of cumulative fluid input/output/net which was estimated using Spearman's monotonic correlation coefficient,  $s$ , due to the observed non-linear trend after graphing. The correlation matrix is ordered according to choroid area, and only correlation coefficients  $\geq 0.4$  in absolute value are annotated.

All baseline choroidal measurements appeared to have positive non-linear, monotonic (Spearman  $s$ ) associations with cumulative fluid input and output (input fluid: choroidal thickness,  $s=+0.58$ ; suprachoroidal space thickness,  $s=+0.47$ ; CVI  $s=+0.57$ ; output fluid: choroidal thickness,  $s=+0.58$ ; suprachoroidal space thickness,  $s=+0.48$ ; CVI  $s=+0.55$ ). These non-linear trends are shown in figure 6.17.

### 6.3. D-RISC II: A phase two evaluation of retinal choroid biomarkers for shock resuscitation in critically ill adults



**Figure 6.16:** Correlation heat map between choroidal measurements (horizontal axis) and clinical measurements (vertical axis) at baseline, with correlation coefficients superimposed and ordered by choroid area (far left). All correlations are Pearson correlation coefficients, with exception of cumulative fluid input/output/net which use the Spearman correlation coefficient. Only correlation coefficients  $\geq 0.4$  in absolute value are annotated. CC, cumulative; Bicarb, bicarbonate; Adj\_Calc: Calcium; FiO2, fraction inspired oxygen; Temp, peripheral temperature; Gluc, glucose; WCC, white cell count; Na, sodium; Lymph, lymphocytes; SaO2, oxygen saturation; Phosp, phosphate; Sys\_BP, systolic blood pressure; K, potassium; Neuts, Neutrophils; PaO2\_Low, lowest partial pressure of oxygen; APACHE2, Acute Physiology and Chronic Health Evaluation Score 2; mAP, mean arterial pressure; CRT, capillary refill time; Dia\_BP, diastolic blood pressure; Mg, magnesium; RR, respiratory rate; CRP, C-reactive protein; Hct, haematocrit; GCS, Glasgow coma scale; Alb, albuminuria; Hb, haemoglobin; HR, heart rate; Hb, haemoglobin; SCS, suprachoroidal space.

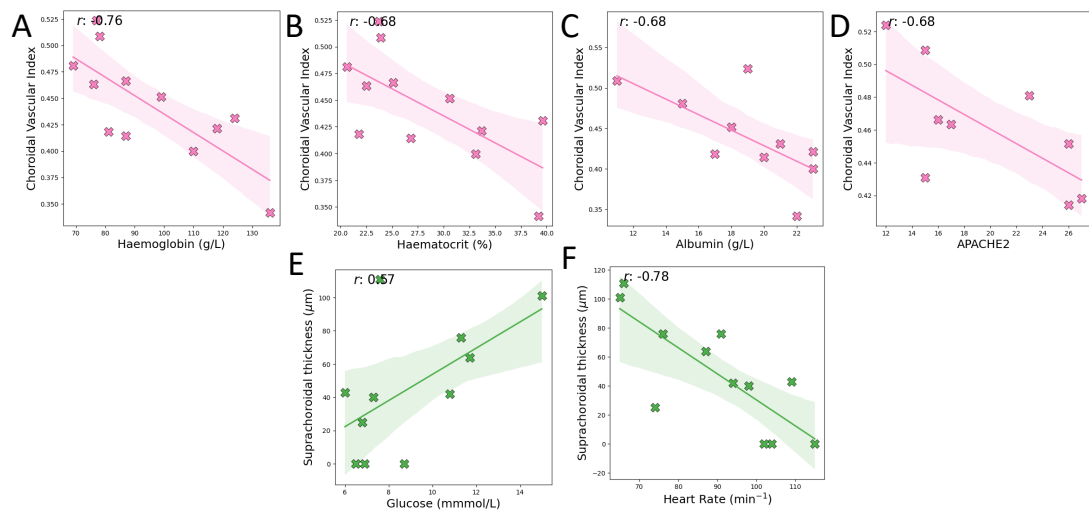


**Figure 6.17:** Observed non-linear and monotonic trends between choroidal measurements and cumulative fluid input (top) and output (bottom). (A) CVI, (B) subfoveal choroid thickness and (C) subfoveal suprachoroidal thickness.

Baseline CVI appeared to have many strong, linear associations with the clinical variables collected, in particular blood markers. Figure 6.18(A – D) presents the strongest linear associations estimated between baseline CVI and the clinical variables selected for analysis. Baseline CVI appeared negatively associated with Acute Physiology and Chronic Health Evaluation (APACHE) II score ( $r = -0.68$ ), albumin ( $r = -0.68$ ), haematocrit ( $r = -0.68$ ), haemoglobin ( $r = -0.76$ ). A weaker, negative association was observed with partial pressure oxygen ( $r = -0.53$ ) and positively with phosphate ( $r = +0.60$ ), glucose ( $r = +0.52$ ) and white cell count ( $r = +0.53$ ).

For those blood markers, haemoglobin, haematocrit and albumin, we observed relatively strong associations with measurements of thickness, area and vessel area. Suprachoroidal thickness and area appeared to be positively associated with glucose (thickness  $r = +0.57$ , area  $r = +0.55$ ) and negatively associated with heart rate (thickness  $r = -0.78$ , area  $r = -0.79$ ), as shown in figure 6.18(E – F). There were no obvious associations at baseline with systemic blood pres-

### 6.3. D-RISC II: A phase two evaluation of retinal choroid biomarkers for shock resuscitation in critically ill adults



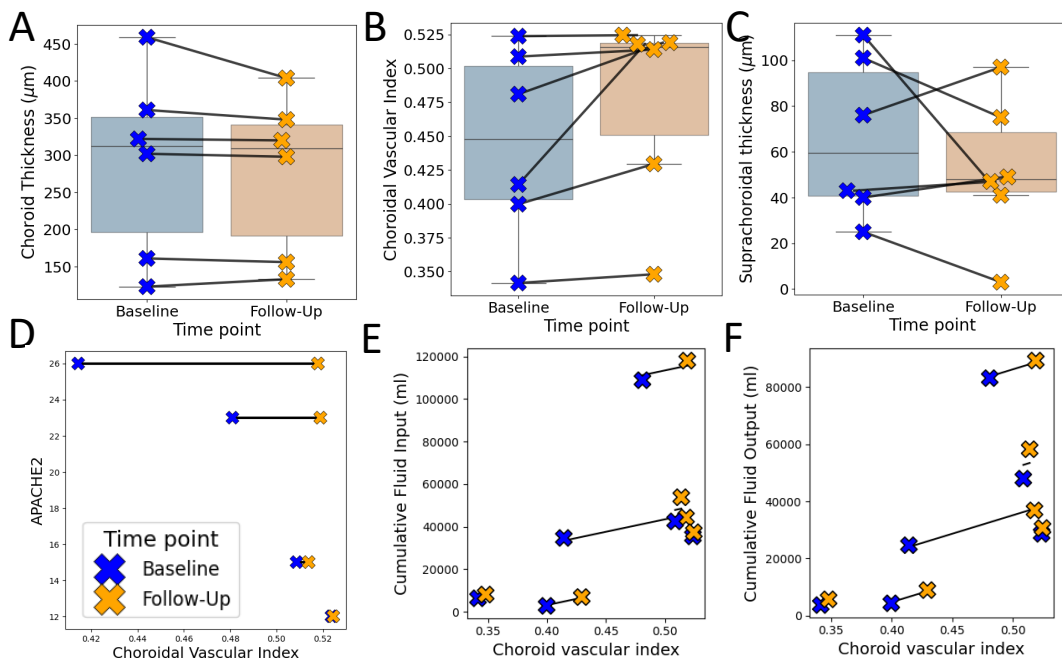
**Figure 6.18:** (A – D) Observed linear trends between CVI and (A) haemoglobin, (B) haematocrit, (C) albumin and (D) APACHE2 score. (E – F) Observed linear trends with subfoveal suprachoroidal space thickness and (E) glucose and (F) heart rate. APACHE2, Acute Physiology and Chronic Health Evaluation Score 2;

sure, nor was there an obvious association between these rheological markers and systemic blood pressure (mean arterial blood pressure and haemoglobin  $r = +0.02$ , haematocrit  $r = -0.08$  and albumin  $r = -0.03$ ).

Figure 6.19 shows longitudinal change in choroidal measurements. Subfoveal choroidal thickness did not vary substantially, but CVI and suprachoroidal thickness did in some individuals (figure 6.19(A – C)). The largest increases in CVI appeared to be in patients with the highest APACHE II scores (Figure 6.19(D)), however two of the six patients who were followed-up did not have an APACHE2 score. Finally, CVI appeared to linearly correlate with cumulative fluid markers over time and within patients (repeated measures Pearson correlation for CVI with input fluid,  $r_{rm} = +0.74$ ; and output fluid  $r_{rm} = +0.83$ ).

#### 6.3.4 Discussion

This study has demonstrated that bedside OCT is feasible and effective in diverse clinical presentations from a single-centre, tertiary critical care unit. The



**Figure 6.19:** (A – C) Observed change in measurements between time points for (A) CVI, (B) subfoveal choroid thickness and (C) subfoveal suprachoroidal thickness. (D – F) Longitudinal change, with common slope in black, between CVI and clinical markers across time points for (D) APACHE2, (E) cumulative input fluid and (F) cumulative output fluid. 2 patients of the 6 patients who were followed-up did not have an APACHE2 score. APACHE2, Acute Physiology and Chronic Health Evaluation Score 2.

majority (60%) of encounters were completed with patients positioned supine, supporting feasible imaging of hospitalised patients. The need to perform multiple, repeated OCT scans of the same location per visit were primarily for patients who were sedated, had thick choroids, or had poor fixation. Nevertheless, 50% of all examinations (and macular OCT scan types) only required one attempt, and 81% within two attempts. Repeated attempts at scanning were also associated with elongated acquisition times, but the majority (72%) of examinations were finished within 5 minutes which, considering the time configuring the device at bedside (~ 5 minutes), is a reasonable length of time to ensure any disruption to provisional care is minimised.

Our exploratory results in this small sample suggest that choroidal measurements may reflect systemic fluid status and rheology, and that EDI-OCT is

### 6.3. D-RISC II: A phase two evaluation of retinal choroid biomarkers for shock resuscitation in critically ill adults

---

sensitive to changes within individuals over time. In particular, we did observe a potential monotonic relationship between choroidal measures and cumulative input/output fluid, both at baseline and over time, while CVI appeared to associate negatively with blood markers haemoglobin, haematocrit and albumin. These observations are particularly interesting given the apparent absence of relationship with macrovascular markers of systemic blood pressure. Moreover, there was no obvious association with systemic blood pressure and these blood markers. This may suggest the choroidal vascular bed is more informative at reflecting the microvascular environment and could be a potentially useful sampling site for assessing organ perfusion.

This study is also the first to report the visibility of the suprachoroid in critically ill patients, which was present in 75% of successfully imaged patients. This frequency and magnitude of suprachoroidal space we found is unusual. The suprachoroid is a potential space between the choroid and sclera traversed by vessels and nerves, and ranges in thickness from 10 to 35 microns [36], and is often not visible due to limited OCT device resolution. Although we observed this space as open for younger patients, a thin suprachoroid may be visible in up to 44% of healthy people aged 55 – 85 [18], but an obviously visible space is usually associated with ocular disease such as severely low eye pressure or retinochoroidal inflammation [36, 275, 336]. Whilst we did not measure eye pressure, the presence of this space in these ITU patients, and its ability to change during admission, might indicate a response to fluid build-up and choroidal effusion [337]. Further work in this space is needed to elucidate the potential role of the suprachoroid during fluid therapy, while considering eye pressure.

Additionally, to the authors knowledge, this study is the first to successfully quantify choroidal variation in both cross-sectional and longitudinal settings using EDI-OCT in a heterogeneous cohort of critically ill patients across multiple diagnoses. Previous studies have shown OCT feasibility in tertiary critical care units [78, 79], but this study is the first to correlate intra-ocular changes with patients' acute physiology across heterogeneous aetiologies. Liu et al. [78] demonstrated OCT and OCT-Angiography (OCT-A) feasibility in a similar setting, and our study extends this by investigating a marginally larger cohort

and suggesting the feasibility of EDI-OCT of the choroid. Additionally, Courtie et al. [79] found changes in retinal blood flow using OCT-A in patients undergoing elective oesophagostomy, supporting our findings of abnormal intra-ocular anatomy in critical illness. We hypothesise that acute changes in patients' intravascular compartment and physiological status are linked to microcirculatory adaptations in intra-ocular structures. Our study is a first step towards understanding how choroidal responses to systemic stress might be used to monitor circulatory shock progression in critically ill adults.

The main limitation of this study was the small and heterogeneous study population. While the wide inclusion criteria are *advantageous* to assess bedside OCT feasibility, they also introduce *confounding* factors into the statistical analysis. The limited number of participants, due to the exploratory nature and resource constraints of this preliminary study, prevented application of more comprehensive statistical analysis methods or reporting of statistical significance of estimated associations. A larger cohort may help further statistically validate our exploratory findings.

Moreover, whilst 14 patients were imaged at baseline, only 6 (43%) were followed-up per-protocol. Of the 8 patients lost, 4 (29%) were discharged from critical care before follow-up imaging could be arranged — a result of limited resources to implement and continue the study protocol outside of the tertiary critical care unit. Thus, future studies designed to test hypotheses about associations between choroidal metrics and systemic physiology should follow patients up after discharge.

Finally, we also note that while it is feasible, effective imaging with the Heidelberg Engineering OCT Spectralis FLEX required 2 – 3 operators and cooperation of nurses. However, the size, cost, and ease of use of portable OCT equipment is likely to improve, creating the potential for investigating retinal microcirculatory biomarkers of multi-organ failure and shock further [80, 81], especially since retinochoroidal OCT images can now be analysed with open-source tools like Choroidalizer.

Future work will seek to perform OCT and OCT-A imaging of heterogeneous intensive care and high dependency unit patients, to validate the initial findings

### 6.3. D-RISC II: A phase two evaluation of retinal choroid biomarkers for shock resuscitation in critically ill adults

---

presented in this study, and explore the utility of the suprachoroid and choroidal vascular bed as biomarkers for circulatory shock. Ultimately, it is hoped that this research will enable patient-level physiological optimisation of fluid resuscitation and to prevent excess iatrogenic morbidity from fluid resuscitation within critically ill patients.

#### 6.3.4.1 Outputs

In this section, there is a manuscript in preparation but this has not been submitted (as of December 2024). Current plans are to submit to BMC Critical Care. The manuscript has not yet been uploaded to medRxiv. The title and authorship is listed below for completeness, with the author of this thesis in bold and any co-lead author contributions in italic.

- George Cooper, **Burke, Jamie**, Charlene Hamid, Emily Godden, Neeraj Dhaun, Stuart King, Tom MacGillivray, J. Kenneth Baillie, David M. Griffith and Ian J.C. MacCormick. “*D-RISC II: Direct Retinal Imaging for Shock Resuscitation in Critical Ill Adults II.*” Unpublished/unsubmitted as of December 2024.

#### 6.3.4.2 Executive summary

We tested the feasibility of measuring the choroid in intensive care and explored associations between choroidal measurements and severity of disease. We performed OCT capture on patients admitted to ITU/HDU and repeated imaging once 12 – 72 hours later. We measured the choroid and compared this to prospectively collected clinical variables. Attempted OCT imaging on fourteen patients was successful in twelve at baseline, and all six attempts at follow-up was successful. At baseline, choroidal measurements were positively associated with fluid balance, and negatively with Acute Physiology and Chronic Health Evaluation (APACHE) II score, haematocrit, albumin and haemoglobin. A measurable suprachoroidal space was seen in 75% (9) patients and was inversely associated with heart rate. There was substantial intra-individual



variation in choroidal measurements over time. We found that measuring the choroid was feasible in patients with critical illness. Exploratory associations with systemic variables suggest that the choroid may provide information about the microvascular function of other major organs. Size and change of choroidal measurements may reflect perfusion pressure or inflammation. To the best of our knowledge, this is the first report of choroidal imaging in intensive care.

# Chapter 7

## Discussion

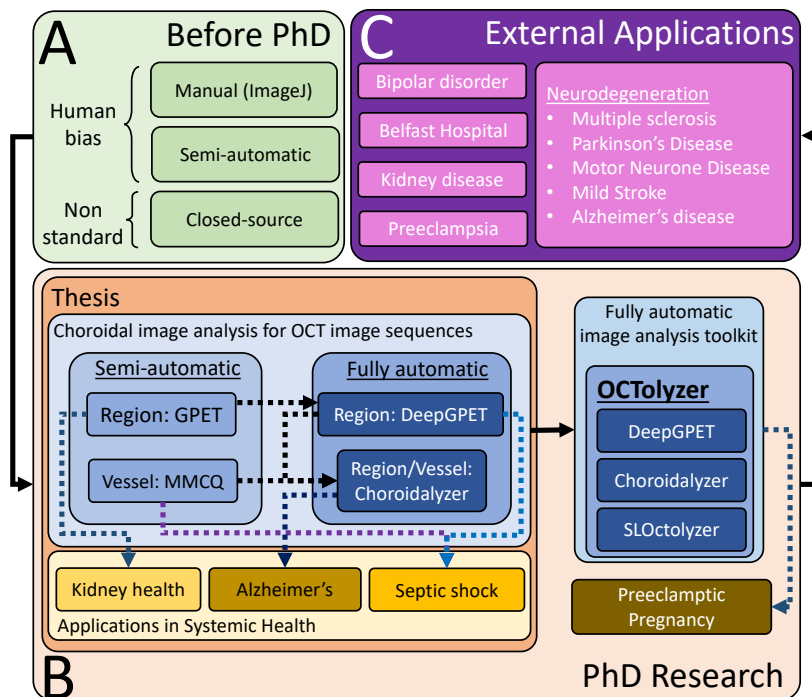
### 7.1 Core contributions to the field

Figure 7.1 presents the diagram of the PhD research conducted during the study period, highlighting the accomplishments of this thesis in addressing the lack of standardised tools for measuring the choroid in OCT image sequences.

In this thesis, we described and evaluated the accuracy and reproducibility of several new approaches for measuring the choroid in OCT image sequences. These approaches address the issue of reproducibility in choroidal image analysis, enabling the research community to better interpret their results when using one of these methods.

The approaches are freely available online in user-friendly formats, allowing the research community to adopt them for their own research studies, thereby tackling the significant lack of open-source tools for choroid measurement. Fully automatic segmentation methods, such as DeepGPET and Choroidalyzer, eliminate the need for manual intervention or semi-automatic alternatives, do not require domain-specific knowledge or specialised training, and provide consistent definitions and protocols for the segmentation and measurement of the choroidal space. This, in turn, contributes to the standardisation of choroidal measurements in OCT image sequences.

Moreover, the quantitative methods developed in this thesis evolved in a series,



**Figure 7.1:** (A) Previous approaches to choroidal image analysis. (B) Work conducted during research period. (C) Past/current/future applications of developed approaches.

reflecting the iterative nature of the research process. Specifically, the discussion and recognition of the biases and limitations inherent in each method led to the development of the next method as an improvement, addressing some of the shortcomings of its predecessors. This iterative research process promoted transparency, critical self-reflection, and scientific rigour, ensuring the development of more accurate and reliable choroidal measurements for the research community.

Finally, we demonstrated the use of these methods in three specific applications of choroidal image analysis in systemic disease: renal health, Alzheimer’s disease risk, and circulatory shock. These applications not only illustrate the potential role of the choroid as a biomarker for systemic disease, but also highlight the importance of accurate measurements of the choroid in understanding this role better.

## 7.2 The technical work accomplished in this thesis

The quantitative methods developed in this work were tested rigorously in our evaluation. We performed population-level comparisons against ground truth labels generated using deterministic, semi-automatic approaches. We also provided transparent and detailed analyses of errors, both quantitatively and qualitatively. In the latter, strict protocols were used to educate manual adjudicators and image graders on image quality, segmentation quality and manual segmentations to ensure consistency (appendix A).

Moreover, we assessed the reproducibility of our methods to provide researchers with the ability to differentiate measurement error from true biological change. For choroidal metrics or applications where effect sizes are likely to be small, such as measuring CVI between healthy and diseased eyes (2% – 6%) [28] or tracking myopia progression with choroidal thickness (20 – 30  $\mu\text{m}$ ) [24, 110], respectively, it's imperative that clinical measurements are interpreted by accounting for the expected effect size and potential measurement variability [235]. Fortunately, we have assessed each method's measurement error at the eye-level [200], which allows end-users to better interpret their study results if using these new approaches to measure the choroid.

This thesis has contributed new approaches to the field which are easier to interpret than predictive models which output risk or classification of disease. The approaches introduced and validated in chapters 2 – 5 output visual delineation of a particular anatomical tissue, which is something that can be easily assessed for accuracy, and excluded or corrected if desired. Conversely, deep learning-based disease classification models encode any tissue quantification step within their 'black-box' architecture as a latent feature which is then subsequently used for prediction. This does not allow for any assessment of the working which contributes to the overall low explainability of any decision made from these models. While ultimately the output segmentations are only one step of the pipeline from bedside to biological interpretation, the visual output of such methods are easier to verify than a continuous risk score of disease.

### 7.2.1 Outline of thesis progression

We first developed GPET (chapter 2), a semi-automatic approach to general purpose edge tracing which was designed around choroid region segmentation in OCT B-scans. This allowed flexible edge modelling with human intervention, and removed measurement error present in manual measurements. However, it was dependent on critical pre-processing and manual initialisation which resulted in poor execution time ( $\approx 35$  seconds/B-scan), thus making it unfeasible for large-scale OCT image analysis. While the underlying theory of the approach was sound, GPET was not able to be used routinely in research because its semi-automatic pipeline required domain-specific knowledge. This precluded GPET's accessibility for those unequipped with the prerequisite knowledge in image analysis or mathematics.

The human bias, poor execution time and challenging end-user accessibility motivated a more fully automatic approach, DeepGPET (chapter 4), which was developed using deep learning as a fast and significantly more accessible approach to choroid region segmentation. We found DeepGPET to be highly reproducible for SD-OCT scans both across the macula and also in the peripapillary at the population- and eye-level. Additionally, DeepGPET reduced the execution time for region segmentation by 96% relative to GPET and, qualitatively, it was preferred in cases of large disagreement with ground truth labels generated by GPET. However, using DeepGPET to generate standardised, fovea-centred ROI measurements of the choroid still required an element of manual intervention to detect the fovea. DeepGPET was also not robust to thick choroids in SS-OCT data, preventing its wider use in other OCT data types. Importantly, DeepGPET was unable to segment the choroidal vessels.

For choroid vessel segmentation we developed a semi-automatic procedure MMCQ for multi-scale vessel enhancement and segmentation (chapter 3). MMCQ was developed as a domain-specific solution to address the heterogeneity of the choroidal vasculature in terms of size and shape, and the poor contrast of which the choroid often suffers. Developing this method also prevented the need to have large amounts of manually labelled data for downstream deep learning model development, given that grading is hard to do unambiguously

for a human and does not represent a gold standard.

We showed that the multi-scale nature of MMCQ's enhancement procedure addressed the heterogeneity and contrast issues related to the choroidal vasculature. We also showed that it was reproducible across SD-OCT and SS-OCT data, and was able to better preserve the fidelity of the choroidal vasculature than local thresholding approaches such as the Niblack method outlined by Agrawal, et al. [158]. MMCQ is a semi-automatic, open-source tool which should require no parameter tuning or manual intervention, and only takes approximately 2 seconds/B-scan. Thus, MMCQ is a suitable replacement for choroid vessel segmentation to popularised approaches like Niblack [27, 158].

At the time of writing, local thresholding of choroidal vessels using the Niblack method is still at the forefront of choroid vessel segmentation. This is despite issues related to agreement among different versions of Niblack's method [173], improper usage across various themes such as reproducibility [235], disease [177, 179], robustness across devices [176, 183], brightness sensitivity [175], scale sensitivity [174] and parameter sensitivity [163]. All of these contribute to Niblack's variability in providing optimal vessel segmentations of the choroid.

While we showed that the Niblack method had similar levels of reproducibility to MMCQ for macular OCT data on a fixed setting of parameters, Niblack's segmentations were of poorer quality across choroids of varying size. This is a problem which could be solved by tuning Niblack's parameters according to the size of choroidal vasculature, thus injecting subjectivity into the vessel segmentation procedure and making Niblack a human-driven methodology. We demonstrated earlier that the range of acceptable values for Niblack's internal parameters can vary significantly when tuned for optimal segmentation agreement to different manual graders (section 3.1.1).

Regardless of the ability for MMCQ to provide more precise vessel segmentation than the current reference standard, the method still fell short of providing automated software to the research community. This was because of its multi-stage pipeline which has the ability to be further tuned by the end-user and requiring manual fovea detection to generate standardised measurements of the choroidal vasculature. This unfortunately injects human bias into the seg-

mentation procedure. The ultimate goal was to develop an algorithm capable of producing high quality choroid vessel segmentations, which was parameter-free to the end-user, and worked on a variety of choroids.

In realising the bias and limitations of GPET, DeepGPET and MMCQ, and identifying a need for a standardised and fully automatic approach to segment and measure the choroid, we developed Choroidalyzer.

Choroidalyzer has several benefits to the research community, satisfying many of the key element missing from much of the current literature (tables 1.1 – 1.3). Namely, it was developed using a substantially larger dataset than most previous methods, and whose ground-truth labels were generated by semi-automatic methods rather than manual ones, it permits fully-automatic measurement of the choroid for a single B-scan, it is generalisable across different OCT imaging devices and data types, and has been rigorously evaluated to provide interpretable reproducibility metrics for downstream choroidal image analysis in OCT image sequences.

### 7.2.2 Choroidalyzer's key contributions

Choroidalyzer combined all four stages of choroidal image analysis into a single procedure (figure 5.1) and was trained on 5,600 OCT B-scans from two market-leading OCT device manufacturers (Heidelberg Engineering and Topcon) and OCT data types (SD-OCT and SS-OCT). Additionally, the ground truth labels generated from these B-scans used semi-automatic methods rather than manual grading, mitigating any potential human error [189]. Besides, manual ground truth labelling was simply not practical given the amount of time it takes for a human grader to annotate choroidal vessels (section 5.4.2.2).

Choroidalyzer's functionality suggests its feasibility for large-scale ophthalmic image analysis, given its low CPU execution time (0.3 seconds/B-scan) and demonstrated successful segmentation of large and poor quality choroids (figure 5.13) from SS-OCT. Choroidalyzer also demonstrated generalisability on an external test set in systemic disease, which performed in parallel with the internal test set, and detailed error analyses showed errors within ground truth

labels rather than error on Choroidalyzer alone. The model also outperformed the current reference standard (Niblack for vessels and DeepGPET for region segmentation) used by most in the research community, and was preferred in cases of high error by a qualitative adjudicator.

Choroidalyzer also reported strong levels of reproducibility across SD-OCT and SS-OCT data, with examples of major outliers representing challenging cases of poor vessel visualisation and invisible Choroid-Sclera boundary. In particular, its eye-level reproducibility enables the research community to differentiate measurement error from true biological effect. Additionally, this reproducibility analysis demonstrated Choroidalyzer's ability to process volumetric data and successfully measure the choroid when EDI mode is toggled on and off, making it an attractive approach for fully automatic analysis of the macular choroid in OCT image sequences.

Choroidalyzer's deterministic output ensures a strict segmentation and measurement protocol is followed, preventing any subjectivity around the definition of the Choroid-Sclera boundary. Although, it may also be possible to tune Choroidalyzer's probabilistic output according to the research problem which may consider different definitions of the Choroid-Sclera boundary, permitting segmentation of the choroid and/or suprachoroidal space (figure 5.14, further exploration required). Choroidalyzer also outputs a soft CVI to account for the inherent uncertainty around choroidal vessels in OCT. This is particularly advantageous over semi-automatic approaches which consider the problem of vessel segmentation a binary one, or still have human bias injected into the segmentation and measurement procedures from end-users performing manual pre-processing or selecting optimal parameters per B-scan [27, 158].

While these attributes demonstrate the rigour in Choroidalyzer's model development and performance, it is not without its limitations. The model struggled with peripapillary OCT B-scans (figure 5.7) and can presently only process single B-scans inputted as an image, requiring minimal interfacing with Python programming for the end-user. Additionally, its lack of validation across examples related to retinal pathology highlight the need for external validation. Nevertheless, we anticipate the significant benefit this toolkit will have for choroidal image analysis in OCT image sequences in systemic disease.



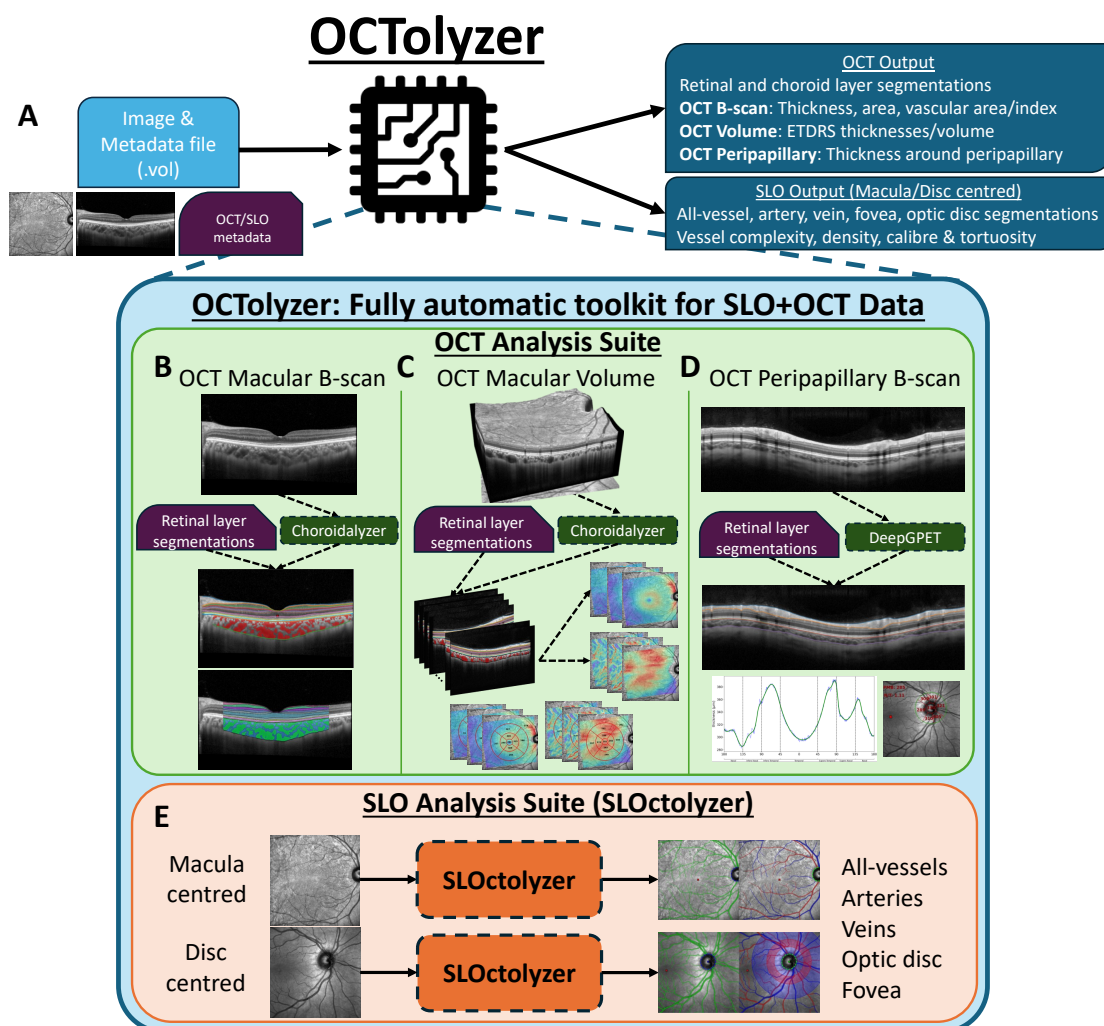
### 7.2.3 From imaging device to choroid measurement

A significant limitation of Choroidalyzer was that it worked on a single B-scan presented as an image file, and was not equipped to handle proprietary file formats or different scan patterns. To ensure methods like Choroidalyzer and DeepGPET are accessible to the general researcher and require no technical background in image analysis or indeed computer science, Choroidalyzer and DeepGPET have been wrapped into a fully automatic analysis toolkit for OCT data, OCTolyzer [5] — the first of its kind. Figure 7.2 shows a schematic diagram of this toolkit.

OCTolyzer is a Python-based toolkit which (currently) supports processing `.vol` RAW files (native to Heidelberg Engineering) which contain the OCT (and localiser SLO) image data and necessary file metadata. Due to lack of data availability at the time of development, other file formats, such as `.fda`, `.fds`, `.img` and `.oct` are not supported. Additionally, `.e2e` (also native to Heidelberg Engineering) is not supported because current python-based file readers [339, 340] are unable to locate the necessary pixel length-scales for converting from pixel space into physical space. However, it would be desirable for OCTolyzer to be fully compatible with DICOM (`.dcm`) format for purposes of vendor neutrality. It is hoped that continued development from the author or future collaborators may facilitate this.

OCTolyzer has an OCT analysis suite which supports processing of macular single-line/radial-line B-scans, macular volume scans and peripapillary B-scans. OCTolyzer can automatically process and measure the choroid using the methods discussed in section 1.7.1.3. As previously discussed in sections 4.6.3 and 5.6.3, Choroidalyzer is used for processing macular OCT data, while DeepGPET is used for processing peripapillary OCT data.

Although outside the scope of this thesis (figure 7.1), OCTolyzer also has the ability to measure the retina through layer segmentations saved within the input file metadata, since most OCT imaging devices have built-in software to process retinal layer segmentations. Moreover, OCTolyzer is equipped to segment and measure the retinal vessels on the en face localiser SLO using OCTolyzer's SLO analysis suite, SLOctolyzer [338]. Thus, OCTolyzer is capa-



**Figure 7.2:** OCTolyzer’s pipeline [5]. (A) OCTolyzer expects a file containing the OCT (and SLO) image data, and necessary OCT file metadata. (B – E) A deeper dive into OCTolyzer’s internal structure. (B – D) OCT analysis suite: analysis pipelines for when OCTolyzer detects a single/radial macular B-scan (B), macular volume scan (C) or peripapillary B-scan (D). (E) SLO analysis suite for macula- or disc-centred SLO localiser images, which accompany OCT capture, using SLOctolyzer [338].

ble of converting a raw proprietary OCT file into a set of clinically meaningful and reproducible measurements of the choroid, alongside cross-sectional and en face measurements of the retina. The work on developing OCTolyzer and SLOctolyzer has resulted in two articles [5, 338]. At the time of writing (December 2024), one has been published [338] and the other is in peer review [5]. The codebases for OCTolyzer and SLOctolyzer have been released as open-source and can be found [here](#) and [here](#), respectively.

The development and subsequent release of toolkits like OCTolyzer and SLOctolyzer has started to streamline much of the OCT image analysis conducted at the author's home research institution (University of Edinburgh). Researchers (Miracle Ozzoude, Borja Marin, Francesca Pentimalli) and technicians (Charlene Hamid, Rosie Keane) are currently using these tools for their own applications, such as in neurodegeneration, bipolar disorder, mild stroke, Alzheimer's disease and preeclampsia. There has also been interest outside of the research institute in using these tools, such as with fellow collaborators at University College London and at Queen's University Belfast. This not only highlights the continued interest for the choroid (and retina) in systemic disease, but also demonstrates the potential utility that open-source methods OCTolyzer (and SLOctolyzer) can and will have to the research community.

### 7.2.4 Closing the gap between research and clinical practice

There is a growing recognition in healthcare-oriented artificial intelligence (AI) of the need for publicly available methods that have undergone rigorous evaluation [341]. The distinct lack of such methods contributes to the significant gap between academic research and clinical AI applications [342].

Despite the abundance of deep learning and AI methods applied to healthcare problems, there is a distinct lack of transparency in the field surrounding their reproducibility, validation strategies and publicly available source code. McDermott, et al. [343] surveyed these themes for 511 machine learning studies presented at machine learning conferences between 2017 and 2019, of which

211 were in the context of healthcare. Of these 211, they found that only 44% reported their measurement variability and 21% made their code and model available. In addition to model availability and reproducibility, external validation has also become a key element of algorithm validation [272].

Model and measurement reproducibility are key to understanding measurement variability and model reliability, and are core elements of a method's validation. The work in this thesis endeavoured to follow current guidelines and reporting standards which exist around machine learning reproducibility [344]. We provided sufficient model and hyperparameter specification for model reproduction, as well as detailed description of our training procedures to help facilitate reproducibility and transparency for both DeepGPET and Choroidalyzer. Additionally, we assessed the measurement reproducibility of these tools using SD-OCT and SS-OCT at both the population-level and eye-level. However, we were unable to share the datasets used for model development, and nor were there publicly available datasets to benchmark model performance with.

Open-source tools can play a critical role in helping bridge the gap between research-only software and their use-case in research studies by making models more accessible to clinicians and researchers [345]. This is why we have made a concerted effort to make the tools developed in this PhD publicly available, which we hope will facilitate future collaboration to help externally validate, extend and maintain. Importantly, without proper long-term management and maintenance, their impact may be ephemeral at most [345].

While the approaches introduced during this research period will make broader impact to the research community by helping improve our understanding of the choroid in disease, their entry into clinical practice would reach the clinician and patient more directly. However, introducing automated tools into clinical practice is no easy feat, particularly those using AI given their black-box nature resulting in poor interpretability [346, 342]. Nevertheless, Choroidalyzer and DeepGPET advantageously only provide visual, anatomical labelling rather than disease detection which removes a level of abstraction that the latter models encode in their decision making.

The ability for a method to generalise across out-of-sample distributions is a key concern among the healthcare community [347], contributing to the difficulty of adopting and integrating these technologies within the medical community. The current landscape is dominated by academic research where methods are often developed and tailored to specific in-house datasets, are closed-source or held behind a membership-service, and thus may not generalise well to broader clinical settings [341]. For example, Liu, et al. [347] found a distinct lack of external validation of deep learning models for disease detection, despite finding their performance equivalent to that of healthcare professionals. While we acknowledge the importance of external validation, this was only performed with Choroidalizer on data related to one systemic disease — containing examples with relatively normal retinæ — on prospectively collected research data. Thus, further out-of-sample validation is an important next step, particularly from external datasets which were not collected in-house and which also reflect image acquisition of real-world clinical care.

Overcoming these barriers around reproducibility, validation and model availability requires significant interdisciplinary collaboration from researchers, clinicians, regulators, and technicians. However, the research community has acknowledged this and attempted to report guidelines for the use of AI in healthcare [348]. For example, the CONSORT-AI [348] guidelines aim to improve the reporting standards, explainability and reproducibility of AI in healthcare and hopes that these will facilitate increased trust in clinical practice. A clear framework for AI in clinical trial reports may also help regulatory bodies better evaluate these models as they attempt to keep pace with the ever-evolving nature of AI [349].

The development of these methods in a healthcare environment is not just about creating tools but also about fostering a collaborative environment where researchers, clinicians, and technicians can work together to conceptualise, develop, refine and validate models [350]. This is a sentiment which the work during this research period aimed to align with. The methods developed in this thesis were developed not only in consultation with OCT imaging technicians, but also clinicians who were available for crucial image annotation and model adjudication. The use of AI in healthcare must leverage multi-disciplinary

teams and collaboration throughout the developmental process [351]. This cross-functional effort is what drives continuous improvement and adaptation to the evolving needs of healthcare systems.

Choroidal image analysis in OCT is still an evolving field, and we hope that tools like Choroidalyzer will motivate other researchers to consider the importance of distributing, evaluating and validating transparent, open-source methods. In the field of choroid analysis in OCT — one which has grown substantially with the advancements in AI, OCT technology and the observed associations between the choroid and systemic disease — standardisation of choroidal measurements has become a necessity. Fully automatic, deep learning-based tools like Choroidalyzer may help mitigate measurement error introduced from manual or semi-automatic methods, and make a positive step toward standardisation of ocular parameters associated with the choroid. With additional validation from external sources, we anticipate Choroidalyzer to make further impact to the field.

## 7.3 Current limitations

The primary limitation of this work surround the datasets used for model development which were all related to systemic health from prospective studies. Thus, not only did the tools developed during the research period did not undergo any real internal or external validation on ocular pathology, but the development of all models was conducted on high-quality, idealised datasets whose image acquisition was conducted by a trained and experienced image technician.

Real-world data from routine clinical care is often marred by virtue of unpredictable or uncontrollable elements such as patient cooperation, camera operator experience and incidental findings. In contrast, research data from prospective studies is collected within a controlled clinical environment, which can often correct for the aforementioned elements impacting real-world data. Thus, it would be a natural next step to quantify the performance of Choroidalyzer on real-world data from clinical care.

Moreover, the majority of researchers in ophthalmology will be investigating the choroid from OCT in chorioretinopathy, and the generalisability of these models on these datasets are currently unknown, which represents a significant limitation. This is especially pertinent to extreme cases of choroidal thinning due to pathological myopia [255], extreme choroidal enlargement such as central serous chorioretinopathy [256], or conditions affecting the photoreceptors and RPE leading to signal hypertransmission into the choroidal space, such as geographic atrophy [257].

However, organs are not closed systems, and the emerging evidence of the choroid's role in systemic disease motivated much of the clinical applications (chapter 6). Thus, the rigorous evaluation on the research data related to systemic health presented in this thesis highlight the feasibility of Choroidalizer and DeepGPET being used for such applications by the research community. Thus, we anticipate them to still be of significant value to the nascent field of ophthalmology for systemic disease, or "oculomics" [49, 67, 274].

A significant limitation associated with our clinical applications is their relatively small sample sizes. In some cases, larger-scale image analysis was precluded due to historical imaging protocols capturing OCT images without EDI mode resulting in poor choroid visualisation. This had a direct impact on the amount of OCT data that was available for image analysis and downstream statistical analysis, and thus impacted the statistical power to which we could draw our conclusions with. This was particularly relevant to the applications in CKD (section 6.1) and Alzheimer's disease (section 6.2), as much of the OCT volume data was not usable because OCT volume scans were collected without EDI mode activated. Consequently, statistical analysis was simplified to fovea-centred single line-scans.

This ultimately introduced an element of selection bias, since the usability of non-EDI OCT for choroid analysis is correlated with the size of the choroid, i.e. thicker choroids imaged with non-EDI OCT will be at higher risk of poor Choroid-Sclera boundary visualisation. Moreover, it's likely that poor OCT acquisition is associated with degree of sickness (as shown by the feasibility of OCT capture in ITU from section 6.3). However, while this thesis has been impacted by OCT protocols neglecting the choroid during OCT capture, the work



has motivated a change of protocols at the Edinburgh Imaging Facility (Royal Infirmary of Edinburgh) to consider EDI mode as an important attribute in many of the ongoing research studies related to preeclampsia, Alzheimer’s disease, stroke and kidney disease. Moreover, better training is being provided so that imaging technicians are well equipped with the ability to image deep into the eye and visualise the choroid.

Additionally, at the time of model development there were no publicly available datasets for choroid segmentation in either systemic disease or retinal pathology, which precluded benchmarking performance. Benchmarking tools are essential in research as they enable objective performance comparison between different methods, ensuring transparency and reproducibility by providing a standardised framework for evaluating performance and replicating results. While this does not yet exist in OCT choroid analysis, it does for other popular modalities like retinal CFP which have many publicly available datasets such as the DRIVE [352], ORIGA [353], STARE [354], and RIM-ONE [355].

While Choroidalyzer was developed to be robust to both SD-OCT and SS-OCT data, older methods were developed for SD-OCT data which resulted in DeepGPET’s poor generalisation to large choroids in SS-OCT data (figure 4.8). Even with Choroidalyzer’s robustness to both OCT data types, only two OCT imaging device manufacturers were used (Heidelberg Engineering, Topcon), and thus the performance of Choroidalyzer on data from other device manufacturers, such as Zeiss, Optovue and Nidek is unknown. Nevertheless, Topcon and Heidelberg are market-leading OCT device manufacturers and providing tools which are fully compatible with these devices is an important advance.

## 7.4 Future work

### 7.4.1 Large-scale analysis, external validation and benchmarking

Choroidalyzer’s accessibility can enable reproducible and clinically meaningful analysis of the choroid which, coupled with its strict measurement proto-



col, permits standardisation of choroidal measurements. This has potential to improve consistency in and impact of future research studies which adopt Choroidalyzer. Additionally, Choroidalyzer presents massive opportunities for large-scale automation of choroidal image analysis. Applications presented in this thesis were on relatively small datasets, but future work would seek to apply Choroidalyzer to large-scale, real-world datasets from clinical care such as AlzEye [274] or UK Biobank [356]. However, while the development of OCTolyzer mitigates any requirement of Python programming and computer science, challenges remain surrounding efficient, large-scale data extract protocols from different machine, as well as Choroidalyzer's lack of validation on these kinds of datasets.

Future work would also seek external validation of Choroidalyzer on datasets related to retinochoroidal pathology to evaluate its generalisability to abnormal retinae. A recent dataset of publicly available OCT B-scans and corresponding segmentations related to macular holes was published, which included choroid region segmentation labels [357]. Not only could this be used to measure Choroidalyzer's generalisability, but could also enable transparent benchmarking of its performance for future iterations (or different models entirely) to be evaluated against in the public domain. Additionally, Choroidalyzer should be re-trained with the original dataset plus these additional 3,859 OCT images, with more superior, transformer-based architectures [147, 172] to leverage more state-of-the-art AI techniques in the field of computer vision.

In time, the availability of public datasets of OCT choroid data will improve, as exemplified by Arian, et al. [167], who have released their OCT data (albeit without annotations) related to diabetic retinopathy and pachychoroid spectrum disorder. This dataset, and others in the future, will permit external validation and benchmarking of open-source algorithms such as the ones developed in this thesis. With external validation comes re-evaluation of the model's limitations and identification of where improvements should be made. This cyclical process will hopefully continue to facilitate accurate and reliable measurements of the choroid, thus continuing to improve the standardisation of ocular parameters for choroidal image analysis in OCT image sequences.

### 7.4.2 Towards choroidal vessel standardisation

As discussed in chapter 3, choroid vessel segmentation is a challenging problem, and there is currently no gold standard. Choroidalizer outputs a soft CVI to account for the oblique presentation of vessels and ambiguity of their vessel walls seen in OCT data. To begin to address this ambiguity it would be pertinent to quantify it, agreed upon by experts in the field (from both those who acquire the images, i.e. technicians, and those who interpret them, i.e. clinicians). We showed in chapter 3 that choroid-derived vessel measurements may have the potential to vary significantly across raters using a small sample of B-scans, but a more systematic approach would be to have several experienced raters tackle a larger and more diverse dataset.

Fortunately, Choroidalizer outputs vessel segmentation masks as probabilistic maps which can be binarised given a threshold, if desired. Thus, future work would consider asking multiple experts to apply Choroidalizer in a large, varied OCT image set and select the optimal threshold (using an analogue scale from 0 to 1). We hypothesise that the distribution of selected thresholds across raters would differ significantly, indicating that human ground truth labels are poor reference standards to benchmark automatic models for choroid vessel segmentation against, or train with. Laurik-Feuerstein, et al. [358] did a similar procedure with colour fundus photography in the context of ground-truth label generation for image quality labels.

### 7.4.3 Context-dependent image quality

As with any large-scale image analysis pipeline, the ability to reject images based on image quality or segmentation accuracy is crucial for forming a meaningful dataset of clinical measurements. SNR estimates built into commercial OCT imaging devices do not characterise the quality of the image in the context to which it will be analysed. Such context-dependent image quality methods are essential in large-scale ophthalmic image analysis, such that automatic rejection plays a precursory role in the image analysis pipeline.

However, definitions for image quality and taxonomy generation require effective collaboration from clinicians, technicians and researchers because of the different, but plausible, definitions surrounding quality. For example, quality with respect to choroid image segmentation may be oriented around clear visualisation of the Choroid-Sclera boundary and vessels, while quality with respect to chorioretinopathy may simply be the ability to observe the disease manifested on the OCT scan. Therein lies the significant challenge which requires technicians, researchers and clinicians collaborating on a taxonomy that is context-dependent. Encouragingly, such context-dependent quality methods are currently being developed for CFPs [359].

### 7.4.4 Choroid depth inference from en face imaging

With the availability of methods like Choroidalyzer enabling large-scale OCT choroid image analysis, choroid quantification across the macula can be done reasonably fast across large volumes of data. Advancements in deep learning have enabled the inference of three dimensional information from two dimensional input, such as depth-maps for natural images [360]. In ophthalmic images, this has also been attempted [361]. Sun, et al. [361] was able to use en face CFP and SLO images to estimate cross-sectional retinal thickness maps across the macula, typically measured from OCT. This could enable more scalable image capture to take place where less focus is needed on OCT capture which is typically more challenging and time-intensive than CFP.

Translating this task to estimating choroid thickness maps across the macula using a combination of CFP and SLO images would also have significant benefit. Importantly, the choroid has historically been poorly imaged, and it is likely that vast volumes of historical OCT data have been captured without EDI mode activated, or before the advent of SS-OCT. Thus, there is potential to recover lost information on the choroid from large OCT image sets which have been plagued from historic data collection, such as the UK Biobank's focus on conventional SD-OCT [356]. This is also applicable in research institutions where protocols for OCT capture neglected the choroid, such as in the initial PREVENT dementia and OCTANE OCT imaging protocols.

## 7.5 Concluding remarks

This thesis has addressed several key issues within the field of choroidal image analysis for OCT image sequences, namely around accuracy, reproducibility and accessibility. We developed several tools of which the latest, Choroidalyzer, now enables end-to-end choroidal analysis in OCT images. Ultimately, Choroidalyzer is the culmination of the previous methods introduced in this research period and was developed to equip the general researcher with the ability to measure the choroid more accurately, consistently and reliably. Ultimately, we hope the dissemination of tools like Choroidalyzer into the research community will facilitate more accurate and reliable measurements of the choroid than existing tools, helping aid interpretability in future research studies.

A major deliverable of these tools was their public release in user-friendly frameworks for end-to-end choroidal analysis. This was to promote accessibility, ensuring that researchers and clinicians alike can adopt these off-the-shelf approaches to measure the choroid in their own studies.

Importantly, these tools remove the need for manual and semi-automatic measurement of the choroid which, due to the distinct lack of available and fully automatic methods, are still the current reference standard despite their subjective approach to segmentation. This open-source approach is key to standardising choroidal measurements.

This research has also demonstrated the application of choroidal analysis in systemic disease, continuing to provide evidence of the links between the microvasculature of the eye and renal health, brain health and critical illness. This illustrates the potential importance of SS-OCT and EDI-OCT imaging of the choroid outside ocular health.

Although current limitations remain in their generalisability to retinal pathologies and validation on real-world data from clinical care, future work aims to address this and benchmark their performance. Regardless, these contributions are anticipated to play a crucial role in advancing large-scale choroidal image analysis and accelerate progress in the nascent field of oculomics.

# Bibliography

- [1] Jamie Burke and Stuart King. Edge tracing using gaussian process regression. *IEEE Transactions on Image Processing*, 31:138–148, 2021.
- [2] Jamie Burke, Dan Pugh, Tariq Farrah, Charlene Hamid, Emily Godden, Thomas J MacGillivray, Neeraj Dhaun, J Kenneth Baillie, Stuart King, and Ian JC MacCormick. Evaluation of an automated choroid segmentation algorithm in a longitudinal kidney donor and recipient cohort. *Translational Vision Science & Technology*, 12(11):19–19, 2023.
- [3] Jamie Burke, Justin Engelmann, Charlene Hamid, Megan Reid-Schachter, Tom Pearson, Dan Pugh, Neeraj Dhaun, Amos Storkey, Stuart King, Tom J MacGillivray, et al. An open-source deep learning algorithm for efficient and fully automatic analysis of the choroid in optical coherence tomography. *Translational Vision Science & Technology*, 12(11):27–27, 2023.
- [4] Justin Engelmann, Jamie Burke, Charlene Hamid, Megan Reid-Schachter, Dan Pugh, Neeraj Dhaun, Diana Moukaddem, Lyle Gray, Niall Strang, Paul McGraw, et al. Choroidalyzer: An open-source, end-to-end pipeline for choroidal analysis in optical coherence tomography. *Investigative Ophthalmology & Visual Science*, 65(6):6–6, 2024.
- [5] Jamie Burke, Justin Engelmann, Samuel Gibbon, Charlene Hamid, Diana Moukaddem, Dan Pugh, Tariq Farrah, Niall Strang, Neeraj Dhaun, Tom MacGillivray, Stuart King, and Ian J.C. MacCormick. Octolyzer: Fully automatic analysis toolkit for segmentation and feature extracting in

- optical coherence tomography (oct) and scanning laser ophthalmoscopy (slo) data. *arXiv preprint arXiv:2407.14128*, 2024.
- [6] Jamie Burke, Samuel Gibbon, Audrey Low, Charlene Hamid, Megan Reid-Schachter, Graciela Muniz Terrera, Craig Ritchie, Baljean Dhillon, John T O'Brien, Stuart King, Ian J.C. MacCormick, and Tom J. MacGillivray. Association between choroidal microvasculature in the eye and alzheimer's disease risk in cognitively healthy midlife adults. *medRxiv*, 2024.
- [7] Debora L Nickla and Josh Wallman. The multifunctional choroid. *Progress in retinal and eye research*, 29(2):144–168, 2010.
- [8] Sarah Mrejen and Richard F Spaide. Optical coherence tomography: imaging of the choroid and beyond. *Survey of ophthalmology*, 58(5):387–429, 2013.
- [9] David Huang, Eric A Swanson, Charles P Lin, Joel S Schuman, William G Stinson, Warren Chang, Michael R Hee, Thomas Flotte, Kenton Gregory, Carmen A Puliafito, et al. Optical coherence tomography. *science*, 254(5035):1178–1181, 1991.
- [10] Bingyao Tan, Jacqueline Chua, Damon Wong, Xinyu Liu, Munirah Ismail, and Leopold Schmetterer. Techniques for imaging the choroid and choroidal blood flow in vivo. *Experimental Eye Research*, page 110045, 2024.
- [11] Richard F Spaide, Hideki Koizumi, and Maria C Pozonni. Enhanced depth imaging spectral-domain optical coherence tomography. *American journal of ophthalmology*, 146(4):496–500, 2008.
- [12] Javier Mazzaferri, Luke Beaton, Gisèle Hounye, Diane N Sayah, and Santiago Costantino. Open-source algorithm for automatic choroid segmentation of oct volume reconstructions. *Scientific reports*, 7(1):42112, 2017.

- [13] KX Cheong, LW Lim, KZ Li, and CS Tan. A novel and faster method of manual grading to measure choroidal thickness using optical coherence tomography. *Eye*, 32(2):433–438, 2018.
- [14] Justin P Ma, Cason B Robbins, Jia Min Lee, Srinath Soundararajan, Sandra S Stinnett, Rupesh Agrawal, Brenda L Plassman, Eleonora M Lad, Heather Whitson, Dilraj S Grewal, et al. Longitudinal analysis of the retina and choroid in cognitively normal individuals at higher genetic risk of alzheimer disease. *Ophthalmology Retina*, 2022.
- [15] Xiangcong Xu, Xuehua Wang, Jingyi Lin, Honglian Xiong, Mingyi Wang, Haishu Tan, Ke Xiong, and Dingan Han. Automatic segmentation and measurement of choroid layer in high myopia for oct imaging using deep learning. *Journal of Digital Imaging*, 35(5):1153–1163, 2022.
- [16] Tariq E Farrah, Dan Pugh, Fiona A Chapman, Emily Godden, Craig Balmforth, Gabriel C Oniscu, David J Webb, Baljean Dhillon, James W Dear, Matthew A Bailey, et al. Choroidal and retinal thinning in chronic kidney disease independently associate with egfr decline and are modifiable with treatment. *Nature Communications*, 14(1):7720, 2023.
- [17] Meng Xuan, Wei Wang, Danli Shi, James Tong, Zhuoting Zhu, Yu Jiang, Zongyuan Ge, Jian Zhang, Gabriella Bulloch, Guankai Peng, et al. A deep learning–based fully automated program for choroidal structure analysis within the region of interest in myopic children. *Translational Vision Science & Technology*, 12(3):22–22, 2023.
- [18] Glenn Yiu, Paula Pecun, Neeru Sarin, Stephanie J Chiu, Sina Farsiou, Prithvi Mruthyunjaya, and Cynthia A Toth. Characterization of the choroid-scleral junction and suprachoroidal layer in healthy individuals on enhanced-depth imaging optical coherence tomography. *JAMA ophthalmology*, 132(2):174–181, 2014.
- [19] Taha Sezer, Muhammet Altınışik, İbrahim Arif Koytak, and Mehmet Hakan Özdemir. The choroid and optical coherence tomography. *Turkish journal of ophthalmology*, 46(1):30, 2016.

- [20] Sumit Randhir Singh, Kiran Kumar Vupparaboina, Abhilash Goud, Kunal K Dansingani, and Jay Chhablani. Choroidal imaging biomarkers. *Survey of Ophthalmology*, 64(3):312–333, 2019.
- [21] Mustafa Aksoy, Leyla Asena, Mustafa Agah Tekindal, Ebru Hatice Ayvazoğlu Soy, Gürsel Yılmaz, and Mehmet Haberal. Choroidal vascularity index and choroidal thickness changes following renal transplantation. *Turkish Journal of Ophthalmology*, 53(2):97, 2023.
- [22] Nattapon Boonarpha, Yalin Zheng, Alexandros N Stangos, Huiqi Lu, Ankur Raj, Gabriela Czanner, Simon P Harding, and Jayashree Nair-Sahni. Standardization of choroidal thickness measurements using enhanced depth imaging optical coherence tomography. *International journal of ophthalmology*, 8(3):484, 2015.
- [23] Rui Xie, Bingjie Qiu, Jay Chhablani, and Xinyuan Zhang. Evaluation of choroidal thickness using optical coherent tomography: a review. *Frontiers in Medicine*, 8:783519, 2021.
- [24] Katharina Breher, Arne Ohlendorf, and Siegfried Wahl. A metrological approach to the analysis of choroidal thickness by optical coherence tomography 3d scans in myopia research. *Scientific Reports*, 9(1):20322, 2019.
- [25] Raluca Brehar, Delia-Alexandrina Mitrea, Flaviu Vancea, Tiberiu Marita, Sergiu Nedevschi, Monica Lupsor-Platon, Magda Rotaru, and Radu Ioan Badea. Comparison of deep-learning and conventional machine-learning methods for the automatic recognition of the hepatocellular carcinoma areas from ultrasound images. *Sensors*, 20(11):3085, 2020.
- [26] Sarah Patterson. Oct-tools. <https://github.com/sarastokes/OCT-tools>, 2019.
- [27] Shozo Sonoda, Taiji Sakamoto, Takehiro Yamashita, Makoto Shirasawa, Eisuke Uchino, Hiroto Terasaki, and Masatoshi Tomita. Choroidal structure in normal eyes and after photodynamic therapy determined by bina-



- rization of optical coherence tomographic images. *Investigative ophthalmology & visual science*, 55(6):3893–3899, 2014.
- [28] Rupesh Agrawal, Jianbin Ding, Parveen Sen, Andres Rousselot, Amy Chan, Lisa Nivison-Smith, Xin Wei, Sarakshi Mahajan, Ramasamy Kim, Chitaranjan Mishra, et al. Exploring choroidal angioarchitecture in health and disease using choroidal vascularity index. *Progress in retinal and eye research*, 77:100829, 2020.
- [29] Johannes Schindelin, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, Curtis Rueden, Stephan Saalfeld, Benjamin Schmid, et al. Fiji: an open-source platform for biological-image analysis. *Nature methods*, 9(7):676–682, 2012.
- [30] Pukhraj Rishi, Zeeshan Akhtar, Rupesh Agrawal, Ashutosh Agrawal, and Ekta Rishi. Analysis of choroidal structure and vascularity indices with image binarization of swept source optical coherence tomography images: A prospective study of 460 eyes. *Oman Journal of Ophthalmology*, 15(1):49–55, 2022.
- [31] Anton Reiner, Malinda EC Fitzgerald, Nobel Del Mar, and Chunyan Li. Neural control of choroidal blood flow. *Progress in retinal and eye research*, 64:96–130, 2018.
- [32] Leonard M Parver, Charles Auker, and David O Carpenter. Choroidal blood flow as a heat dissipating mechanism in the macula. *American journal of ophthalmology*, 89(5):641–646, 1980.
- [33] Dale Purves, G Augustine, D Fitzpatrick, L Katz, A LaMantia, J McNamara, and S Williams. Neuroscience 2nd edition. sunderland (ma) sin-aer associates. *Types of Eye Movements and Their Functions*, 2001.
- [34] Sohan Singh Hayreh and Shelagh Bell Hayreh. Uveal vascular bed in health and disease: uveal vascular bed anatomy. paper 1 of 2. *Eye*, 37(13):2590–2616, 2023.

- [35] Gordon L Ruskell. Peripapillary venous drainage from the choroid: a variable feature in human eyes. *British journal of ophthalmology*, 81(1):76–79, 1997.
- [36] Shohista Saidkasimova and Thomas H. Williamson, editors. *Supra-choroidal Space Interventions*. Springer International Publishing, Cham, 2021.
- [37] Erandi Chandrasekera, Evan N Wong, Danuta M Sampson, David Alonso-Caneiro, and Fred K Chen. Posterior choroidal stroma reduces accuracy of automated segmentation of outer choroidal boundary in swept source optical coherence tomography. *Investigative ophthalmology & visual science*, 59(11):4404–4412, 2018.
- [38] GA Lutty, T Hasegawa, T Baba, R Grebe, I Bhutto, and DS McLeod. Development of the human choriocapillaris. *Eye*, 24(3):408–415, 2010.
- [39] Raphael Lejoyeux, Jennifer Benillouche, Joshua Ong, Marie-Hélène Errera, Ethan A Rossi, Sumit R Singh, Kunal K Dansingani, Susana da Silva, Debasish Sinha, José-Alain Sahel, et al. Choriocapillaris: fundamentals and advancements. *Progress in Retinal and Eye Research*, 87:100997, 2022.
- [40] Christophe Delaey and Johan Van de Voorde. Regulatory mechanisms in the retinal and choroidal circulation. *Ophthalmic research*, 32(6):249–256, 2000.
- [41] Albert Alm and Anders Bill. Ocular and optic nerve blood flow at normal and increased intraocular pressures in monkeys (macaca irus): a study with radioactively labelled microspheres including flow determinations in brain and some other tissues. *Experimental eye research*, 15(1):15–29, 1973.
- [42] Leopold Schmetterer and Jeffrey Kiel. *Ocular blood flow*. Springer Science & Business Media, 2012.
- [43] Chui Ming Gemmy Cheung and Richard F Spaide. Watersheds and mini-watersheds. *Eye*, 35(9):2449–2450, 2021.

- [44] SOHAN SINGH Hayreh. Segmental nature of the choroidal vasculature. *British Journal of Ophthalmology*, 59(11):631–648, 1975.
- [45] Sohan Singh Hayreh. In vivo choroidal circulation and its watershed zones. *Eye*, 4(2):273–289, 1990.
- [46] Norbert D Wangsa-Wirawan and Robert A Linsenmeier. Retinal oxygen: fundamental and clinical aspects. *Archives of ophthalmology*, 121(4):547–557, 2003.
- [47] Alon Harris, Thomas A Ciulla, Hak Sung Chung, and Bruce Martin. Regulation of retinal and optic nerve blood flow. *Archives of ophthalmology*, 116(11):1491–1495, 1998.
- [48] Zuyuan He and Kazuo Hotate. Synthesized optical coherence tomography for imaging of scattering objects by use of a stepwise frequency-modulated tunable laser diode. *Optics letters*, 24(21):1502–1504, 1999.
- [49] Siegfried K Wagner, Dun Jack Fu, Livia Faes, Xiaoxuan Liu, Josef Huemer, Hagar Khalid, Daniel Ferraz, Edward Korot, Christopher Kelly, Konstantinos Balaskas, et al. Insights into systemic disease through retinal imaging-based oculomics. *Translational vision science & technology*, 9(2):6–6, 2020.
- [50] Kara-Anne Tan, Preeti Gupta, Aniruddha Agarwal, Jay Chhablani, Ching-Yu Cheng, Pearse A Keane, and Rupesh Agrawal. State of science: choroidal thickness and systemic health. *Survey of ophthalmology*, 61(5):566–581, 2016.
- [51] Takeshi Iwase, Kentaro Yamamoto, Eimei Ra, Kenta Murotani, Shigeyuki Matsui, and Hiroko Terasaki. Diurnal variations in blood flow at optic nerve head and choroid in healthy eyes: diurnal variations in blood flow. *Medicine*, 94(6):e519, 2015.
- [52] Ranjay Chakraborty, Scott A. Read, and Michael J. Collins. Diurnal Variations in Axial Length, Choroidal Thickness, Intraocular Pressure, and Ocular Biometrics. *Investigative Ophthalmology & Visual Science*, 52(8):5121–5129, July 2011.

- [53] Colin S Tan, Yanling Ouyang, Humberto Ruiz, and Srinivas R Sadda. Diurnal variation of choroidal thickness in normal, healthy subjects measured by spectral domain optical coherence tomography. *Investigative ophthalmology & visual science*, 53(1):261–266, 2012.
- [54] Shinichi Usui, Yasushi Ikuno, Masahiro Akiba, Ichiro Maruko, Tetsuju Sekiryu, Kohji Nishida, and Tomohiro Iida. Circadian changes in subfoveal choroidal thickness and the relationship with circulatory factors in healthy subjects. *Investigative Ophthalmology & Visual Science*, 53(4):2300–2307, April 2012.
- [55] Takamasa Kinoshita, Yoshinori Mitamura, Kayo Shinomiya, Mariko Egawa, Akiko Iwata, Akiko Fujihara, Yoko Ogushi, Kentaro Semba, Kei Akaiwa, Eisuke Uchino, Shozo Sonoda, and Taiji Sakamoto. Diurnal variations in luminal and stromal areas of choroid in normal eyes. *The British Journal of Ophthalmology*, 101(3):360–364, March 2017.
- [56] Sumit Randhir Singh, Mohammed Abdul Rasheed, Abhilash Goud, Niroj Kumar Sahoo, Kiran Kumar Vupparaboina, and Jay Chhablani. Diurnal variation in subfoveal and peripapillary choroidal vascularity index in healthy eyes. *Indian Journal of Ophthalmology*, 67(10):1667–1672, 2019.
- [57] Lisa A. Ostrin, Elise Harb, Debora L. Nickla, Scott A. Read, David Alonso-Caneiro, Falk Schroedl, Alexandra Kaser-Eichberger, Xiangtian Zhou, and Christine F. Wildsoet. IMI-The Dynamic Choroid: New Insights, Challenges, and Potential Significance for Human Myopia. *Investigative Ophthalmology & Visual Science*, 64(6):4, May 2023.
- [58] Konstantinos A Papathanasiou, Dimitrios Kazantzis, Dimitrios A Vrachatis, Sotiria G Giotaki, Eva Papaconstantinou, Menelaos Kanakis, Dimitrios Avramides, Spyridon Deftereos, Irini Chatziralli, and Ilias Georgalas. Choroidal thickness in patients with systemic arterial hypertension: a systematic review and meta-analysis. *Therapeutic Advances in Ophthalmology*, 14:25158414221132825, 2022.

- [59] Craig Balmforth, Job JMH van Bragt, Titia Ruijs, James R Cameron, Robert Kimmitt, Rebecca Moorhouse, Alicja Czopek, May Khei Hu, Peter J Gallacher, James W Dear, et al. Choriorretinal thinning in chronic kidney disease links to inflammation and endothelial dysfunction. *JCI insight*, 1(20), 2016.
- [60] Chee Wai Wong, Tien Yin Wong, Ching-Yu Cheng, and Charumathi Sabanayagam. Kidney and eye diseases: common risk factors, etiological mechanisms, and pathways. *Kidney international*, 85(6):1290–1302, 2014.
- [61] Tariq E Farrah, Baljean Dhillon, Pearse A Keane, David J Webb, and Neeraj Dhaun. The eye, the kidney, and cardiovascular disease: old concepts, better tools, and new horizons. *Kidney international*, 98(2):323–342, 2020.
- [62] Min Gyu Choi and Jee Taek Kim. Strong correlation of renal function with choroidal thickness in patients with type 2 diabetes: retrospective cross-sectional study. *Journal of clinical medicine*, 9(7):2171, 2020.
- [63] Heeyoon Cho, Yong Un Shin, Doo Roo Kim, Joo Hark Yi, et al. Evaluation of changes in choroidal thickness and the choroidal vascularity index after hemodialysis in patients with end-stage renal disease by using swept-source optical coherence tomography. *Investigative Ophthalmology & Visual Science*, 60(9):1545–1545, 2019.
- [64] Fariba Shaikhi Shoshtari, Sasan Biranvand, Leila Rezaei, Nader Salari, and Nasrin Aghaei. The impact of hemodialysis on retinal and choroidal thickness in patients with chronic renal failure. *International ophthalmology*, 41:1763–1771, 2021.
- [65] Bernard Canaud, Charles Chazot, Jeroen Koomans, and Allan Collins. Fluid and hemodynamic management in hemodialysis patients: challenges and opportunities. *Brazilian Journal of Nephrology*, 41:550–559, 2019.

- [66] Gizem Ashraf, Myra McGuinness, Muhammad Azaan Khan, Czarina Obtinalla, Xavier Hadoux, and Peter van Wijngaarden. Retinal imaging biomarkers of alzheimer's disease: A systematic review and meta-analysis of studies using brain amyloid beta status for case definition. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 15(2):e12421, 2023.
- [67] Siegfried Karl Wagner, David Romero-Bascones, Mario Cortina-Borja, Dominic J. Williamson, Robbert R. Struyven, Yukun Zhou, Salil Patel, Rimona S. Weil, Chrystalina A. Antoniadou, Eric J. Topol, Edward Korot, Paul J. Foster, Konstantinos Balaskas, Unai Ayala, Maitane Barrenechea, Iñigo Gabilondo, Anthony H.V. Schapira, Anthony P. Khawaja, Praveen J. Patel, Jugnoo S. Rahi, Alastair K. Denniston, Axel Petzold, Pearse Andrew Keane, and for UK Biobank Eye & Vision Consortium. Retinal Optical Coherence Tomography Features Associated With Incident and Prevalent Parkinson Disease. *Neurology*, 101(16):e1581–e1593, October 2023. Publisher: Wolters Kluwer.
- [68] Alex Suh, Gilad Hampel, Aditya Vinjamuri, Joshua Ong, Sharif Amit Kamran, Ethan Waisberg, Phani Paladugu, Nasif Zaman, Prithul Sarker, Alireza Tavakkoli, et al. Oculomics analysis in multiple sclerosis: Current ophthalmic clinical and imaging biomarkers. *Eye*, pages 1–10, 2024.
- [69] VIH Kwa, JJ Van der Sande, J Stam, N Tijmes, and JL Vrooland. Retinal arterial changes correlate with cerebral small-vessel disease. *Neurology*, 59(10):1536–1540, 2002.
- [70] Magda Gharbiya, Alessandro Trebbastoni, Francesco Parisi, Silvia Manganiello, Filippo Cruciani, Fabrizia D'Antonio, Umberto De Vico, Letizia Imbriano, Alessandra Campanelli, and Carlo De Lena. Choroidal thinning as a new finding in alzheimer's disease: evidence from enhanced depth imaging spectral domain optical coherence tomography. *Journal of Alzheimer's Disease*, 40(4):907–917, 2014.
- [71] Samuel Asanad, Fred N Ross-Cisneros, Ernesto Barron, Marco Nasisi, William Sultan, Rustum Karanjia, and Alfredo A Sadun. The reti-

- nal choroid as an oculo-vascular biomarker for Alzheimer's dementia: A histopathological study in severe disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 11:775–783, 2019.
- [72] Cason B. Robbins, Dilraj S. Grewal, Atalie C. Thompson, James H. Powers, Srinath Soundararajan, Hui Yan Koo, Stephen P. Yoon, Bryce W. Polascik, Andy Liu, Rupesh Agrawal, and Sharon Fekrat. Choroidal Structural Analysis in Alzheimer Disease, Mild Cognitive Impairment, and Cognitively Healthy Controls. *American Journal of Ophthalmology*, 223:359–367, March 2021.
- [73] Liang Wang, Serena Shah, Cristina Nicole Llaneras, and Raquel Goldhardt. Insight into the brain: Application of the retinal microvasculature as a biomarker for cerebrovascular diseases through optical coherence tomography angiography. *Current Ophthalmology Reports*, 12(1):1–11, 2024.
- [74] Delia Cabrera DeBuc, Magdalena Gaca-Wysocka, Andrzej Grzybowski, and Piotr Kanclerz. Identification of retinal biomarkers in Alzheimer's disease using optical coherence tomography: recent insights, challenges, and opportunities. *Journal of clinical medicine*, 8(7):996, 2019.
- [75] E. Courtie, T. Veenith, A. Logan, A. K. Denniston, and R. J. Blanch. Retinal blood flow in critical illness and systemic disease: a review. *Annals of Intensive Care*, 10(1):152, November 2020.
- [76] Ella Courtie, Ahmed Gilani, Tonny Veenith, and Richard J Blanch. Optical coherence tomography angiography as a surrogate marker for end-organ resuscitation in sepsis: A review. *Frontiers in Medicine*, 9:1023062, 2022.
- [77] Jang Ryul Park, ByungKun Lee, Min Ji Lee, Kyuseok Kim, and Wang-Yuhl Oh. Visualization of three-dimensional microcirculation of rodents' retina and choroid for studies of critical illness using optical coherence tomography angiography. *Scientific Reports*, 11(1):14302, 2021.

- [78] Xiaoxuan Liu, Aditya Uday Kale, Nicholas Capewell, Nicholas Talbot, Sumiya Ahmed, Pearse A. Keane, Susan Mollan, Antonio Belli, Richard J. Blanch, Tonny Veenith, and Alastair K. Denniston. Optical coherence tomography (OCT) in unconscious and systemically unwell patients using a mobile OCT device: a pilot study. *BMJ Open*, 9(11):e030882, November 2019. Publisher: British Medical Journal Publishing Group Section: Ophthalmology.
- [79] Ella F Courtie, Aditya U Kale, Benjamin TK Hui, Xiaoxuan Liu, Nicholas I Capewell, Jonathan RB Bishop, Tony Whitehouse, Tonny Veenith, Ann Logan, Alastair K Denniston, et al. Stability of oct and octa in the intensive therapy unit setting. *Diagnostics*, 11(8):1516, 2021.
- [80] Reena Chopra, Siegfried K. Wagner, and Pearse A. Keane. Optical coherence tomography in the 2020s—outside the eye clinic. *Eye*, 35(1):236–243, January 2021. Publisher: Nature Publishing Group.
- [81] Ge Song, Evan T. Jelly, Kengyeh K. Chu, Wesley Y. Kendall, and Adam Wax. A review of low-cost and portable optical coherence tomography. *Progress in Biomedical Engineering*, 3(3):032002, May 2021. Publisher: IOP Publishing.
- [82] Charles E Riva, Stephen D Cranstoun, Juan E Grunwald, and Benno L Petrig. Choroidal blood flow in the foveal region of the human ocular fundus. *Investigative ophthalmology & visual science*, 35(13):4273–4281, 1994.
- [83] Young Ho Kim and Jaeryung Oh. Choroidal thickness profile in chorioretinal diseases: beyond the macula. *Frontiers in Medicine*, 8:797428, 2021.
- [84] SA Hewick, AC Fairhead, JC Culy, and HR Atta. A comparison of 10 mhz and 20 mhz ultrasound probes in imaging the eye and orbit. *British journal of ophthalmology*, 88(4):551–555, 2004.
- [85] Robert W Flower and Bernard F Hochheimer. A clinical technique and apparatus for simultaneous angiography of the separate retinal and



- choroidal circulations. *Investigative Ophthalmology & Visual Science*, 12(4):248–261, 1973.
- [86] Ingeborg Stalmans, Evelien Vandewalle, Douglas R Anderson, Vital P Costa, Ronald EP Frenkel, Gerhard Garhofer, Juan Grunwald, Konstantin Gugleta, Alon Harris, Christopher Hudson, et al. Use of colour doppler imaging in ocular blood flow research. *Acta ophthalmologica*, 89(8):e609–e630, 2011.
- [87] Leopold Schmetterer, Susanne Dallinger, Oliver Findl, Ursula Graselli, Hans-Georg Eichler, and Michael Wolzt. A comparison between laser interferometric measurement of fundus pulsation and pneumotonometric measurement of pulsatile ocular blood flow 2. effects of changes in pco<sub>2</sub> and po<sub>2</sub> and of isoproterenol. *Eye*, 14(1):46–52, 2000.
- [88] James G Fujimoto. Optical coherence tomography. *Comptes Rendus de l'Académie des Sciences-Series IV-Physics*, 2(8):1099–1111, 2001.
- [89] David Stifter. Chapter 20 - optical coherence tomography for the characterization of micro-parts and -structures. In Yi Qin, editor, *Micro-Manufacturing Engineering and Technology*, Micro and Nano Technologies, pages 324–330. William Andrew Publishing, Boston, 2010.
- [90] Wolfgang Drexler, Mengyang Liu, Abhishek Kumar, Tschackad Kamali, Angelika Unterhuber, and Rainer A Leitgeb. Optical coherence tomography today: speed, contrast, and multimodality. *Journal of biomedical optics*, 19(7):071412–071412, 2014.
- [91] Michelle L Gabriele, Gadi Wollstein, Hiroshi Ishikawa, Larry Kagemann, Juan Xu, Lindsey S Folio, and Joel S Schuman. Optical coherence tomography: history, current status, and laboratory work. *Investigative ophthalmology & visual science*, 52(5):2425–2436, 2011.
- [92] Christopher Mody. The past, present and future of oct imaging. *Optician*, 2016(7):143397–1, 2016.
- [93] Maciej Wojtkowski, Rainer Leitgeb, Andrzej Kowalczyk, Tomasz Bajraszewski, and Adolf F Fercher. In vivo human retinal imaging by fourier

- domain optical coherence tomography. *Journal of biomedical optics*, 7(3):457–463, 2002.
- [94] R. Paschotta. Optical coherence tomography. RP Photonics Encyclopedia.
- [95] Mehreen Adhi and Jay S Duker. Optical coherence tomography—current and future applications. *Current opinion in ophthalmology*, 24(3):213–221, 2013.
- [96] Shoji Kishi. Impact of swept source optical coherence tomography on ophthalmology. *Taiwan journal of ophthalmology*, 6(2):58–68, 2016.
- [97] MM Teussink, S Donner, T Otto, K Williams, and A Tafreshi. State-of-the-art commercial spectral-domain and swept-source oct technologies and their clinical applications in ophthalmology. *Heidelberg Engineering*, pages 1–19, 2019.
- [98] Topcon Healthcare. Dri oct triton brochure. [https://topconhealthcare.eu/en\\_EU/productdownload/1180/1454](https://topconhealthcare.eu/en_EU/productdownload/1180/1454).
- [99] Heidelberg Engineering. Spectralis oct2 module brochure. [https://media.heidelbergengineering.com/uploads/Products-Downloads/200240-002\\_SPECTRALIS\\_Brochure\\_OCT2-Insert\\_EN.pdf](https://media.heidelbergengineering.com/uploads/Products-Downloads/200240-002_SPECTRALIS_Brochure_OCT2-Insert_EN.pdf).
- [100] Optopol Technology. Revo-nx130 brochure. [https://optopol.com/wp-content/uploads/2023/03/ONLINE\\_0208221-Revo-NX130-207.pdf](https://optopol.com/wp-content/uploads/2023/03/ONLINE_0208221-Revo-NX130-207.pdf).
- [101] Josef F Bille. High resolution imaging in microscopy and ophthalmology: new frontiers in biomedical optics. 2019.
- [102] Heidelberg Engineering. Spectralis hra+oct hardware instructions. [https://cdn.ymaws.com/www.opsweb.org/resource/collection/8C09696C-D8CA-4647-A6BF-B0B83369CA24/Spectralis\\_Hardware\\_Operating\\_Manual.pdf](https://cdn.ymaws.com/www.opsweb.org/resource/collection/8C09696C-D8CA-4647-A6BF-B0B83369CA24/Spectralis_Hardware_Operating_Manual.pdf), 2007.

- [103] Mahsa Sohrab, Katherine Wu, and Amani A Fawzi. A pilot study of morphometric analysis of choroidal vasculature in vivo, using en face optical coherence tomography. *PloS one*, 7(11):e48631, 2012.
- [104] Lauren A Branchini, Mehreen Adhi, Caio V Regatieri, Namrata Nandakumar, Jonathan J Liu, Nora Laver, James G Fujimoto, and Jay S Duker. Analysis of choroidal morphologic features and vasculature in healthy eyes using spectral-domain optical coherence tomography. *Ophthalmology*, 120(9):1901–1908, 2013.
- [105] Zaixing Mao, Atsuya Miki, Song Mei, Ying Dong, Kazuichi Maruyama, Ryo Kawasaki, Shinichi Usui, Kenji Matsushita, Kohji Nishida, and Kin-pui Chan. Deep learning based noise reduction method for automatic 3d segmentation of the anterior of lamina cribrosa in optical coherence tomography volumetric scans. *Biomedical optics express*, 10(11):5832–5851, 2019.
- [106] Waheeda Rahman, Fred Kuanfu Chen, Jonathan Yeoh, Praveen Patel, Adnan Tufail, and Lyndon Da Cruz. Repeatability of manual subfoveal choroidal thickness measurements in healthy subjects using the technique of enhanced depth imaging optical coherence tomography. *Investigative ophthalmology & visual science*, 52(5):2267–2271, 2011.
- [107] Lei Shao, Liang Xu, Chang Xi Chen, Li Hong Yang, Kui Fang Du, Shuang Wang, Jin Qiong Zhou, Ya Xing Wang, Qi Sheng You, Jost B Jonas, et al. Reproducibility of subfoveal choroidal thickness measurements with enhanced depth imaging by spectral-domain optical coherence tomography. *Investigative ophthalmology & visual science*, 54(1):230–233, 2013.
- [108] Dawn A Sim, Pearse A Keane, Hemal Mehta, Simon Fung, Javier Zarranz-Ventura, Marcus Fruttiger, Praveen J Patel, Catherine A Egan, and Adnan Tufail. Repeatability and reproducibility of choroidal vessel layer measurements in diabetic retinopathy using enhanced depth optical coherence tomography. *Investigative ophthalmology & visual science*, 54(4):2893–2901, 2013.

- [109] Vivian S Vuong, Elad Moisseiev, David Cunefare, Sina Farsiu, Ala Moshiri, and Glenn Yiu. Repeatability of choroidal thickness measurements on enhanced depth imaging optical coherence tomography using different posterior boundaries. *American journal of ophthalmology*, 169:104–112, 2016.
- [110] Ignacio Flores-Moreno, Francisco Lugo, Jay S Duker, and José M Ruiz-Moreno. The relationship between axial length and choroidal thickness in eyes with high myopia. *American journal of ophthalmology*, 155(2):314–319, 2013.
- [111] Fred K Chen, Jonathan Yeoh, Waheeda Rahman, Praveen J Patel, Adnan Tufail, and Lyndon Da Cruz. Topographic variation and interocular symmetry of macular choroidal thickness using enhanced depth imaging optical coherence tomography. *Investigative ophthalmology & visual science*, 53(2):975–985, 2012.
- [112] AR Cho, YJ Choi, and YT Kim. Influence of choroidal thickness on subfoveal choroidal thickness measurement repeatability using enhanced depth imaging optical coherence tomography. *Eye*, 28(10):1151–1160, 2014.
- [113] Edsger W Dijkstra. A note on two problems in connexion with graphs. In *Edsger Wybe Dijkstra: his life, work, and legacy*, pages 287–290. Association for Computing Machinery, 2022.
- [114] Huiqi Lu, Nattapon Boonarpa, Man Ting Kwong, and Yalin Zheng. Automated segmentation of the choroid in retinal optical coherence tomography images. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5869–5872. IEEE, 2013.
- [115] Saleha Masood, Bin Sheng, Ping Li, Ruimin Shen, Ruogu Fang, and Qiang Wu. Automatic choroid layer segmentation using normalized graph cut. *IET Image Processing*, 12(1):53–59, 2018.

- [116] Bahareh Salafian, Rahele Kafieh, Abdolreza Rashno, Mohsen Pourazizi, and Saeid Sadri. Automatic segmentation of choroid layer in edi oct images using graph theory in neutrosophic space. *arXiv preprint arXiv:1812.01989*, 2018.
- [117] Vedran Kajić, Marieh Esmaeelpour, Boris Považay, David Marshall, Paul L Rosin, and Wolfgang Drexler. Automated choroidal segmentation of 1060 nm oct in healthy and pathologic eyes using a statistical model. *Biomedical optics express*, 3(1):86–103, 2012.
- [118] David Alonso-Caneiro, Scott A Read, and Michael J Collins. Automatic segmentation of choroidal thickness in optical coherence tomography. *Biomedical optics express*, 4(12):2795–2812, 2013.
- [119] Md Akter Hussain, Alauddin Bhuiyan, Hiroshi Ishikawa, R Theodore Smith, Joel S Schuman, and Ramamohanrao Kotagiri. An automated method for choroidal thickness measurement from enhanced depth imaging optical coherence tomography images. *Computerized Medical Imaging and Graphics*, 63:41–51, 2018.
- [120] Qiang Chen, Sijie Niu, Wangyi Fang, Yuanlu Shuai, Wen Fan, Songtao Yuan, and Qinghuai Liu. Automated choroid segmentation of three-dimensional sd-oct images by incorporating edi-oct images. *Computer Methods and Programs in Biomedicine*, 158:161–171, 2018.
- [121] Zhihong Hu, Xiaodong Wu, Yanwei Ouyang, Yanling Ouyang, and Srinivas R Sadda. Semiautomated segmentation of the choroid in spectral-domain optical coherence tomography volume scans. *Investigative ophthalmology & visual science*, 54(3):1722–1729, 2013.
- [122] Hajar Danesh, Raheleh Kafieh, Hossein Rabbani, and Fedra Hajizadeh. Segmentation of choroidal boundary in enhanced depth imaging octs using a multiresolution texture based modeling in graph cuts. *Computational and mathematical methods in medicine*, 2014, 2014.
- [123] Teresa Torzicky, Michael Pircher, Stefan Zotter, Marco Bonesi, Erich Götzinger, and Christoph K Hitzenberger. Automated measurement of

- choroidal thickness in the human eye by polarization sensitive optical coherence tomography. *Optics express*, 20(7):7564–7574, 2012.
- [124] Lian Duan, Masahiro Yamanari, and Yoshiaki Yasuno. Automated phase retardation oriented segmentation of chorio-scleral interface by polarization sensitive optical coherence tomography. *Optics express*, 20(3):3353–3366, 2012.
- [125] Qiang Chen, Wen Fan, Sijie Niu, Jiajia Shi, Honglie Shen, and Songtao Yuan. Automated choroid segmentation based on gradual intensity distance in hd-oct images. *Optics express*, 23(7):8974–8994, 2015.
- [126] Chuang Wang, Ya Xing Wang, and Yongmin Li. Automatic choroidal layer segmentation using markov random field and level set method. *IEEE journal of biomedical and health informatics*, 21(6):1694–1702, 2017.
- [127] Kiran Kumar Vupparaboina, Srinath Nizampatnam, Jay Chhablani, Ashutosh Richhariya, and Soumya Jana. Automated estimation of choroidal thickness distribution and volume based on oct images of posterior visual section. *Computerized Medical Imaging and Graphics*, 46:315–327, 2015.
- [128] Neetha George and CV Jiji. Two stage contour evolution for automatic segmentation of choroid and cornea in oct images. *Biocybernetics and biomedical Engineering*, 39(3):686–696, 2019.
- [129] Niall O’Mahony, Sean Campbell, Anderson Carvalho, Suman Harapanahalli, Gustavo Velasco Hernandez, Lenka Krpalkova, Daniel Riordan, and Joseph Walsh. Deep learning vs. traditional computer vision. In *Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC), Volume 1 1*, pages 128–144. Springer, 2020.
- [130] Min Chen, Jiancong Wang, Ipek Oguz, Brian L VanderBeek, and James C Gee. Automated segmentation of the choroid in edi-oct images with retinal pathology using convolution neural networks. In *Fetal, Infant and Ophthalmic Medical Image Analysis: International Workshop*,

- FIFI 2017, and 4th International Workshop, OMIA 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 4*, pages 177–184. Springer, 2017.
- [131] Shingo Tsuji, Tetsuju Sekiryu, Yukinori Sugano, Akira Ojima, Akihito Kasai, Masahiro Okamoto, and Satoshi Eifuku. Semantic segmentation of the choroid in swept source optical coherence tomography images for volumetrics. *Scientific Reports*, 10(1):1088, 2020.
- [132] Xiaodan Sui, Yuanjie Zheng, Benzhen Wei, Hongsheng Bi, Jianfeng Wu, Xuemei Pan, Yilong Yin, and Shaoting Zhang. Choroid segmentation from optical coherence tomography with graph-edge weights learned from deep convolutional neural networks. *Neurocomputing*, 237:332–341, 2017.
- [133] Baidaa Al-Bander, Bryan M Williams, Majid A Al-Tae, Waleed Al-Nuaimy, and Yalin Zheng. A novel choroid segmentation method for retinal diagnosis using deep learning. In *2017 10th International Conference on Developments in eSystems Engineering (DeSE)*, pages 182–187. IEEE, 2017.
- [134] Saleha Masood, Ruogu Fang, Ping Li, Huating Li, Bin Sheng, Akash Mathavan, Xiangning Wang, Po Yang, Qiang Wu, Jing Qin, et al. Automatic choroid layer segmentation from optical coherence tomography images using deep learning. *Scientific reports*, 9(1):3058, 2019.
- [135] Fang He, Rachel Ka Man Chun, Zicheng Qiu, Shijie Yu, Yun Shi, Chi Ho To, and Xiaojun Chen. Choroid segmentation of retinal oct images based on cnn classifier and l2-lq fitter. *Computational and mathematical methods in medicine*, 2021(1):8882801, 2021.
- [136] Hung-Ju Chen, Yu-Len Huang, Siu-Lun Tse, Wei-Ping Hsia, Chung-Hao Hsiao, Yang Wang, and Chia-Jen Chang. Application of artificial intelligence and deep learning for choroid segmentation in myopia. *Translational Vision Science & Technology*, 11(2):38–38, 2022.

- [137] Wei Ping Hsia, Siu Lun Tse, Chia Jen Chang, and Yu Len Huang. Automatic segmentation of choroid layer using deep learning on spectral domain optical coherence tomography. *Applied Sciences*, 11(12):5488, 2021.
- [138] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.
- [139] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [140] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.
- [141] Jason Kugelman, David Alonso-Caneiro, Scott A Read, Jared Hamwood, Stephen J Vincent, Fred K Chen, and Michael J Collins. Automatic choroidal segmentation in oct images using supervised deep learning methods. *Scientific reports*, 9(1):13298, 2019.
- [142] Xiaoqian Mao, Yitian Zhao, Bang Chen, Yuhui Ma, Zaiwang Gu, Shenshen Gu, Jianlong Yang, Jun Cheng, and Jiang Liu. Deep learning with skip connection attention for choroid layer segmentation in oct images. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1641–1645. IEEE, 2020.
- [143] Sripad Krishna Devalla, Prajwal K Renukanand, Bharathwaj K Sreedhar, Giridhar Subramanian, Liang Zhang, Shamira Perera, Jean-Martial Mari, Khai Sing Chin, Tin A Tun, Nicholas G Strouthidis, et al. Drunet: a dilated-residual u-net deep learning network to segment optic nerve head tissues in optical coherence tomography images. *Biomedical optics express*, 9(7):3244–3265, 2018.



- [144] Zhenjie Chai, Kang Zhou, Jianlong Yang, Yuhui Ma, Zhi Chen, Shenghua Gao, and Jiang Liu. Perceptual-assisted adversarial adaptation for choroid segmentation in optical coherence tomography. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 1966–1970. IEEE, 2020.
- [145] Dheo AY Cahyo, Ai Ping Yow, Seang-Mei Saw, Marcus Ang, Michael Girard, Leopold Schmetterer, and Damon Wong. Multi-task learning approach for volumetric segmentation and reconstruction in 3d oct images. *Biomedical Optics Express*, 12(12):7348–7360, 2021.
- [146] Huihong Zhang, Jianlong Yang, Kang Zhou, Fei Li, Yan Hu, Yitian Zhao, Ce Zheng, Xiulan Zhang, and Jiang Liu. Automatic segmentation and visualization of choroid in oct with knowledge infused deep learning. *IEEE Journal of Biomedical and Health Informatics*, 24(12):3408–3420, 2020.
- [147] Yaqi Wang, Zehua Yang, Xindi Liu, Zhi Li, Chengyu Wu, Yizhen Wang, Kai Jin, Dechao Chen, Gangyong Jia, Xiaodiao Chen, et al. Pgcd-net: Prior-guided and knowledge diffusive network for choroid segmentation. *Artificial Intelligence in Medicine*, 150:102837, 2024.
- [148] Jing Tian, Pina Marziliano, Mani Baskaran, Tin Aung Tun, and Tin Aung. Automatic segmentation of the choroid in enhanced depth imaging optical coherence tomography images. *Biomedical optics express*, 4(3):397–411, 2013.
- [149] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.
- [150] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
- [151] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE con-*

- ference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [152] Kiran Kumar Vupparaboina, Ashutosh Richhariya, Jay Chhablani, and Soumya Jana. Optical coherence tomography imaging: Automated binarization of choroid for stromal-luminal analysis. In *2016 International Conference on Signal and Information Processing (IConSIP)*, pages 1–5. IEEE, 2016.
- [153] Li Zhang, Kyungmoo Lee, Meindert Niemeijer, Robert F Mullins, Milan Sonka, and Michael D Abramoff. Automated segmentation of the choroid from clinical sd-oct. *Investigative ophthalmology & visual science*, 53(12):7510–7519, 2012.
- [154] Vedran Kajić, Marieh Esmaelpour, Carl Glittenberg, Martin F Kraus, Joachim Honegger, Richu Othara, Susanne Binder, James G Fujimoto, and Wolfgang Drexler. Automated three-dimensional choroidal vessel segmentation of 3d 1060 nm oct retinal data. *Biomedical optics express*, 4(1):134–150, 2012.
- [155] Marieh Esmaelpour, Vedran Kajić, Behrooz Zabihian, Richu Othara, Siamak Ansari-Shahrezaei, Lukas Kellner, Ilse Krebs, Susanne Nemetz, Martin F Kraus, Joachim Honegger, et al. Choroidal haller’s and sattler’s layer thickness measurement using 3-dimensional 1060-nm optical coherence tomography. *PloS one*, 9(6):e99690, 2014.
- [156] Nagaraj R Mahajan, Ravi Chandra Reddy Donapati, Sumohana S Channappayya, Sivaramakrishna Vanjari, Ashutosh Richhariya, and Jay Chhablani. An automated algorithm for blood vessel count and area measurement in 2-d choroidal scan images. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3355–3358. IEEE, 2013.
- [157] Alejandro F Frangi, Wiro J Niessen, Koen L Vincken, and Max A Viergever. Multiscale vessel enhancement filtering. In *Medical Image*

- Computing and Computer-Assisted Intervention—MICCAI'98: First International Conference Cambridge, MA, USA, October 11–13, 1998 Proceedings 1*, pages 130–137. Springer, 1998.
- [158] Rupesh Agrawal, Preeti Gupta, Kara-Anne Tan, Chui Ming Gemmy Cheung, Tien-Yin Wong, and Ching-Yu Cheng. Choroidal vascularity index as a measure of vascular status of the choroid: measurements in healthy eyes from a population-based study. *Scientific reports*, 6(1):21090, 2016.
- [159] Xiaoxiao Liu, Lei Bi, Yupeng Xu, Dagan Feng, Jinman Kim, and Xun Xu. Robust deep learning method for choroidal vessel segmentation on swept source optical coherence tomography images. *Biomedical Optics Express*, 10(4):1601–1612, 2019.
- [160] Bjorn Kaijun Betzler, Jianbin Ding, Xin Wei, Jia Min Lee, Dilraj S Grewal, Sharon Fekrat, Srinivas R Sadda, Marco A Zarbin, Aniruddha Agarwal, Vishali Gupta, et al. Choroidal vascularity index: a step towards software as a medical device. *British Journal of Ophthalmology*, 106(2):149–155, 2022.
- [161] Wayne Niblack. *An introduction to digital image processing*. Strandberg Publishing Company, 1985.
- [162] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.
- [163] Joshua Muller, David Alonso-Caneiro, Scott A Read, Stephen J Vincent, and Michael J Collins. Application of deep learning methods for binarization of the choroid in optical coherence tomography images. *Translational Vision Science & Technology*, 11(2):23–23, 2022.
- [164] Nizampatnam Srinath, A Patil, V Kiran Kumar, Soumya Jana, Jay Chhablani, and Ashutosh Richhariya. Automated detection of choroid boundary and vessels in optical coherence tomography images. In *2014 36th*

- Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 166–169. IEEE, 2014.
- [165] Junmeng Li, Lei Zhu, Ruilin Zhu, Yanye Lu, Xin Rong, Yadi Zhang, Xiaopeng Gu, Yuwei Wang, Zhiyue Zhang, Qiushi Ren, et al. Automated analysis of choroidal sublayer morphologic features in myopic children using edi-oct by deep learning. *Translational Vision Science & Technology*, 10(13):12–12, 2021.
- [166] Gu Zheng, Yanfeng Jiang, Ce Shi, Hanpei Miao, Xiangli Yu, Yiyi Wang, Sisi Chen, Zhiyang Lin, Weicheng Wang, Fan Lu, et al. Deep learning algorithms to segment and quantify the choroidal thickness and vasculature in swept-source optical coherence tomography images. *Journal of Innovative Optical Health Sciences*, 14(01):2140002, 2021.
- [167] Roya Arian, Tahereh Mahmoudi, Hamid Riazi-Esfahani, Hooshang Faghihi, Ahmad Mirshahi, Fariba Ghassemi, Alireza Khodabande, Rahel Kafieh, and Elias Khalili Pour. Automatic choroid vascularity index calculation in optical coherence tomography images with low-contrast sclerochoroidal junction using deep learning. In *Photonics*, volume 10, page 234. MDPI, 2023.
- [168] Tin Tin Khaing, Takayuki Okamoto, Chen Ye, Md Abdul Mannan, Hirotaka Yokouchi, Kazuya Nakano, Pakinee Aimmanee, Stanislav S Makhanov, and Hideaki Haneishi. Choroidnet: a dense dilated u-net model for choroid layer and vessel segmentation in optical coherence tomography images. *IEEE Access*, 9:150951–150965, 2021.
- [169] Lei Zhu, JunMeng Li, Ruilin Zhu, Xiangxi Meng, Pei Rong, Yibao Zhang, Zhe Jiang, Mufeng Geng, Bin Qiu, Xin Rong, et al. Synergistically segmenting choroidal layer and vessel using deep learning for choroid structure analysis. *Physics in Medicine & Biology*, 67(8):085001, 2022.
- [170] Xuehua Wang, Rui Li, Junyan Chen, Dingan Han, Mingyi Wang, Honglian Xiong, Wenzheng Ding, Yixu Zheng, Ke Xiong, and Yaguang Zeng. Choroidal vascularity index (cvi)-net-based automatic assessment

- of diabetic retinopathy severity using cvi in optical coherence tomography images. *Journal of Biophotonics*, 16(6):e202200370, 2023.
- [171] Jeya Maria Jose Valanarasu, Poojan Oza, Ilker Hacihaliloglu, and Vishal M Patel. Medical transformer: Gated axial-attention for medical image segmentation. In *Medical image computing and computer assisted intervention—MICCAI 2021: 24th international conference, Strasbourg, France, September 27–October 1, 2021, proceedings, part I 24*, pages 36–46. Springer, 2021.
- [172] Yang Wen, Yi-Lin Wu, Lei Bi, Wu-Zhen Shi, Xiao-Xiao Liu, Yu-Peng Xu, Xun Xu, Wen-Ming Cao, and David Dagan Feng. A transformer-assisted cascade learning network for choroidal vessel segmentation. *Journal of Computer Science and Technology*, 39(2):286–304, 2024.
- [173] Xin Wei, Shozo Sonoda, Chitaranjan Mishra, Neha Khandelwal, Ramasamy Kim, Taiji Sakamoto, and Rupesh Agrawal. Comparison of choroidal vascularity markers on optical coherence tomography using two-image binarization techniques. *Investigative Ophthalmology & Visual Science*, 59(3):1206–1211, 2018.
- [174] Ahmet Elbay, Furkan Kırık, Kamile Melis Ertürk, and Hakan Ozdemir. Comparison of the choroidal vascularity index measurements of images acquired with different optical coherence tomography scan scale types. *Retina*, pages 10–1097, 2022.
- [175] Nicola Rosa, Marco Gioia, Rachele Orlando, Martina De Luca, Eleonora D’Aniello, Isabella Fioretto, Ciro Sannino, and Maddalena De Bernardo. Impact of brightness on choroidal vascularity index. *Journal of Clinical Medicine*, 13(4):1020, 2024.
- [176] Feiyan Ma, Yifan Bai, Jialiang Duan, Yuchen Liang, and Qingli Shang. Validation of reliability, repeatability and consistency of three-dimensional choroidal vascular index. *Scientific Reports*, 14(1):1576, January 2024. Publisher: Nature Publishing Group.

- [177] Shulin Liu, Liping Du, Qingyun Zhou, Qi Zhang, Ke Hu, Jian Qi, Liang Liang, Chunjiang Zhou, Aize Kijlstra, and Peizeng Yang. The choroidal vascularity index decreases and choroidal thickness increases in vogt–koyanagi–harada disease patients during a recurrent anterior uveitis attack. *Ocular immunology and inflammation*, 26(8):1237–1243, 2018.
- [178] Hiroki Kawano, Shozo Sonoda, Takehiro Yamashita, Ichiro Maruko, Tomohiro Iida, and Taiji Sakamoto. Relative changes in luminal and stromal areas of choroid determined by binarization of edi-oct images in eyes with vogt-koyanagi-harada disease after treatment. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 254:421–426, 2016.
- [179] Rupesh Agrawal, Lilian Koh Hui Li, Vikram Nakhate, Neha Khandelwal, and Padmamalini Mahendradas. Choroidal vascularity index in vogt-koyanagi-harada disease: an edi-oct derived tool for monitoring disease progression. *Translational vision science & technology*, 5(4):7–7, 2016.
- [180] Durgasri Jaisankar, Rajiv Raman, Hitesh R Sharma, Neha Khandelwal, Muna Bhende, Rupesh Agrawal, Sudharshan Sridharan, and Jyotirmay Biswas. Choroidal and retinal anatomical responses following systemic corticosteroid therapy in vogt–koyanagi–harada disease using swept-source optical coherence tomography. *Ocular Immunology and Inflammation*, 27(2):235–243, 2019.
- [181] Marco Pellegrini, Giuseppe Giannaccare, Federico Bernabei, Fabiana Moscardelli, Costantino Schiavi, and Emilio C Campos. Choroidal vascular changes in arteritic and nonarteritic anterior ischemic optic neuropathy. *American Journal of Ophthalmology*, 205:43–49, 2019.
- [182] Abhilash Guduru, Mohammed Abdul Rasheed, Abhilash Goud, Mohamed Ashik, Vupparaboina Kiran Kumar, Jay Chhablani, Akshay Badakere, Ramesh Kekunnaya, and Preeti Patil-Chhablani. Choroidal vascularity in non-arteritic anterior ischaemic optic neuropathy. *Neuro-Ophthalmology*, 43(5):305–309, 2019.
- [183] Rupesh Agrawal, Sophia Seen, Shiva Vaishnavi, Kiran Kumar Vupparaboina, Abhilash Goud, Mohammed Abdul Rasheed, and Jay Chh-

- ablani. Choroidal vascularity index using swept-source and spectral-domain optical coherence tomography: a comparative study. *Ophthalmic Surgery, Lasers and Imaging Retina*, 50(2):e26–e32, 2019.
- [184] Rupesh Agrawal, Xin Wei, Abhilash Goud, Kiran Kumar Vupparaboina, Soumya Jana, and Jay Chhablani. Influence of scanning area on choroidal vascularity index measurement using optical coherence tomography. *Acta ophthalmologica*, 95(8):e770–e775, 2017.
- [185] Abhilash Goud, Sumit Randhir Singh, Niroj Kumar Sahoo, Mohammed Abdul Rasheed, Kiran Kumar Vupparaboina, Samatha Ankireddy, Marco Lupidi, and Jay Chhablani. New insights on choroidal vascularity: a comprehensive topographic approach. *Investigative Ophthalmology & Visual Science*, 60(10):3563–3569, 2019.
- [186] Yung Hui Kim, Hyung Nam Jin, Hyun Jee Kim, Jong Hoon Lee, and Yong-Sok Ji. Influence of scan direction on subfoveal choroidal vascularity index using optical coherence tomography. *Scientific Reports*, 12(1):16626, 2022.
- [187] Sumit Randhir Singh, Alessandro Invernizzi, Mohammed Abdul Rasheed, Carlo Cagini, Abhilash Goud, Kiran Kumar Vupparaboina, Mariano Cozzi, Marco Lupidi, and Jay Chhablani. Wide-field choroidal vascularity in healthy eyes. *American Journal of Ophthalmology*, 193:100–105, 2018.
- [188] Zhongdi Chu, Yuxuan Cheng, Qinqin Zhang, Hao Zhou, Yining Dai, Yingying Shi, Giovanni Gregori, Philip J Rosenfeld, and Ruikang K Wang. Quantification of choriocapillaris with phansalkar local thresholding: pitfalls to avoid. *American journal of ophthalmology*, 213:161–176, 2020.
- [189] Peter M Maloca, Maximilian Pfau, Lucas Janeschitz-Kriegl, Michael Reich, Lukas Goerd, Frank G Holz, Philipp L Müller, Philippe Valmaggia, Katrin Fasler, Pearse A Keane, et al. Human selection bias drives the linear nature of the more ground truth effect in explainable deep learning

- optical coherence tomography image segmentation. *Journal of Biophotonics*, page e202300274, 2023.
- [190] Alexander Brandt. Oct-marker. <https://github.com/neurodial/OCT-Marker/tree/master>, 2018.
- [191] Ron Margolis and Richard F Spaide. A pilot study of enhanced depth imaging optical coherence tomography of the choroid in normal eyes. *American journal of ophthalmology*, 147(5):811–815, 2009.
- [192] Shohista Saidkasimova. Anatomy and Physiology of the Suprachoroidal Space. In Shohista Saidkasimova and Thomas H. Williamson, editors, *Suprachoroidal Space Interventions*, pages 1–30. Springer International Publishing, Cham, 2021.
- [193] John V. Forrester, Andrew D. Dick, Paul G. McMenamin, Fiona Roberts, and Eric Pearlman. Chapter 1 - anatomy of the eye and orbit. In John V. Forrester, Andrew D. Dick, Paul G. McMenamin, Fiona Roberts, and Eric Pearlman, editors, *The Eye (Fourth Edition)*, pages 1–102.e2. W.B. Saunders, fourth edition edition, 2016.
- [194] Heidelberg Engineering. How to take the perfect image. [https://www.academy.heidelbergengineering.com/hedata/e-learning/Totara/Dateien/pdf-tutorials/99913-005\\_Spectralis\\_HRA+OCT\\_Acquisition\\_Brochure.pdf](https://www.academy.heidelbergengineering.com/hedata/e-learning/Totara/Dateien/pdf-tutorials/99913-005_Spectralis_HRA+OCT_Acquisition_Brochure.pdf), 2019.
- [195] Dan Jouma A Maman Lawali, Guanrong Wu, Yunxiang Guo, Zhangjie Lin, Qiaowei Wu, Abdou Amza, Zijing Du, Yun Ren, Ying Fang, Xinran Dong, et al. Measurement of foveal retinal thickness in myopic patients using different display modes on optical coherence tomography: A retrospective, cross-sectional study. *Ophthalmology and Therapy*, 12(1):167–178, 2023.
- [196] Alessandro Arrigo, Emanuela Aragona, Lorenzo Bianco, Alessio Antropoli, Andrea Saladino, Francesco Bandello, and Maurizio Battaglia Parodi. The clinical role of the choroidal assessment in high myopia:



- Characteristics and association with neovascular and atrophic complications. *Investigative Ophthalmology & Visual Science*, 64(12):16–16, 2023.
- [197] Early Treatment Diabetic Retinopathy Study Research Group et al. Grading diabetic retinopathy from stereoscopic color fundus photographs—an extension of the modified airle house classification: Etdrs report number 10. *Ophthalmology*, 98(5):786–806, 1991.
- [198] Heidelberg Engineering. Bmo positioning and oct interpretation in glaucoma. [https://www.academy.heidelbergengineering.com/hedata/e-learning/Totara/Dateien/pdf-tutorials/210023-002\\_SPECTRALIS\\_PDF%20Tutorial\\_BMO%20Positioning%20and%20OCT%20Interpretation%20in%20Glaucoma\\_EN.pdf](https://www.academy.heidelbergengineering.com/hedata/e-learning/Totara/Dateien/pdf-tutorials/210023-002_SPECTRALIS_PDF%20Tutorial_BMO%20Positioning%20and%20OCT%20Interpretation%20in%20Glaucoma_EN.pdf), 2019.
- [199] Zerrin Tuncer and Mitat Altuğ. Does foveal position relative to the optic disc affect optical coherence tomography measurements in glaucoma? *Turkish journal of ophthalmology*, 48(4):178, 2018.
- [200] Justin Engelmann, Diana Moukaddem, Lucas Gago, Niall Strang, and Miguel O Bernabeu. Applicability of oculomics for individual risk prediction: Repeatability and robustness of retinal fractal dimension using dart and automorph. *Investigative Ophthalmology & Visual Science*, 65(6):10–10, 2024.
- [201] Fernando C Monteiro and Aurélio C Campilho. Performance evaluation of image segmentation. In *International Conference Image Analysis and Recognition*, pages 248–259. Springer, 2006.
- [202] Pasi Fränti and Radu Marinescu-Istodor. Soft precision and recall. *Pattern Recognition Letters*, 167:115–121, 2023.
- [203] Dominik Müller, Iñaki Soto-Rey, and Frank Kramer. Towards a guideline for evaluation metrics in medical image segmentation. *BMC Research Notes*, 15(1):210, 2022.

- [204] Yukun Zhou, Siegfried K Wagner, Mark A Chia, An Zhao, Moucheng Xu, Robbert Struyven, Daniel C Alexander, Pearse A Keane, et al. Automorph: automated retinal vascular morphology quantification via a deep learning pipeline. *Translational vision science & technology*, 11(7):12–12, 2022.
- [205] Ian JC MacCormick, Bryan M Williams, Yalin Zheng, Kun Li, Baidaa Al-Bander, Silvester Czanner, Rob Cheeseman, Colin E Willoughby, Emery N Brown, George L Spaeth, et al. Accurate, fast, data efficient and interpretable glaucoma diagnosis with automated spatial analysis of the whole cup to disc profile. *PloS one*, 14(1):e0209409, 2019.
- [206] Dirk B Walther, Barry Chai, Eamon Caddigan, Diane M Beck, and Li Fei-Fei. Simple line drawings suffice for functional mri decoding of natural scene categories. *Proceedings of the National Academy of Sciences*, 108(23):9661–9666, 2011.
- [207] Peter Junteng Liu. *Using Gaussian process regression to denoise images and remove artefacts from microarray data*. University of Toronto, 2007.
- [208] Ashish Kapoor, Kristen Grauman, Raquel Urtasun, and Trevor Darrell. Gaussian processes for object categorization. *International journal of computer vision*, 88(2):169–188, 2010.
- [209] Frank Hakeem. Gaussian process models for computer vision. Master's thesis, Department of Mathematics and Statistics, 2020.
- [210] He He and Wan-Chi Siu. Single image super-resolution using gaussian process regression. In *CVPR 2011*, pages 449–456. IEEE, 2011.
- [211] Alexander Freytag, Björn Fröhlich, Erik Rodner, and Joachim Denzler. Efficient semantic segmentation with gaussian processes and histogram intersection kernels. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pages 3313–3316. IEEE, 2012.

- [212] Zhuowen Tu and Song-Chun Zhu. Image segmentation by data-driven markov chain monte carlo. *IEEE Transactions on pattern analysis and machine intelligence*, 24(5):657–673, 2002.
- [213] Kyle Simek and Kobus Barnard. Gaussian process shape models for bayesian segmentation of plant leaves. *Proceedings of the Computer Vision Problems in Plant Phenotyping (CVPPP)*, pages, pages 4–1, 2015.
- [214] Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.
- [215] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International journal of computer vision*, 1(4):321–331, 1988.
- [216] Azriel Rosenfeld. Picture processing by computer. *ACM Computing Surveys (CSUR)*, 1(3):147–176, 1969.
- [217] Azriel Rosenfeld. *Digital picture processing*. Academic press, 1976.
- [218] Per-Erik Danielsson and Olle Seger. Generalized and separable sobel operators. In *Machine vision for three-dimensional scenes*, pages 347–379. Elsevier, 1990.
- [219] Lawrence G Roberts. *Machine perception of three-dimensional solids*. PhD thesis, Massachusetts Institute of Technology, 1963.
- [220] Judith MS Prewitt. Object enhancement and extraction. *Picture processing and Psychopictorics*, 10(1):15–19, 1970.
- [221] John Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, 8(6):679–698, 1986.
- [222] Stephanie J Chiu, Xiao T Li, Peter Nicholas, Cynthia A Toth, Joseph A Izatt, and Sina Farsiu. Automatic segmentation of seven retinal layers in sdoct images congruent with expert manual segmentation. *Optics express*, 18(18):19413–19428, 2010.

- [223] Richard O Duda and Peter E Hart. Experiments in scene analysis. Technical report, STANFORD RESEARCH INST MENLO PARK CA, 1970.
- [224] Richard Szeliski. *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.
- [225] Neeraj Dhaun. Optical coherence tomography and nephropathy: The octane study. <https://clinicaltrials.gov/ct2/show/NCT02132741>, 2014. ClinicalTrials.gov identifier: NCT02132741. Updated April 14, 2021. Accessed June 29, 2022.
- [226] Prejaas Tewarie, Lianne Balk, Fiona Costello, Ari Green, Roland Martin, Sven Schippling, and Axel Petzold. The oscar-ib consensus criteria for retinal oct quality assessment. *PloS one*, 7(4):e34823, 2012.
- [227] J Martin Bland and DouglasG Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*, 327(8476):307–310, 1986.
- [228] H Passing and W Bablok. A new biometrical procedure for testing the equality of measurements from two different analytical methods. application of linear regression procedures for method comparison studies in clinical chemistry, part i. 1983.
- [229] Paul A. Yushkevich, Joseph Piven, Heather Cody Hazlett, Rachel Gimpel Smith, Sean Ho, James C. Gee, and Guido Gerig. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage*, 31(3):1116–1128, 2006.
- [230] Stephen M Pizer, E Philip Amburn, John D Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B Zimmerman, and Karel Zuiderveld. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3):355–368, 1987.
- [231] Paul Heckbert. Color image quantization for frame buffer display. *ACM Siggraph Computer Graphics*, 16(3):297–307, 1982.

- [232] Joel S Schuman, James G Fujimoto, Jay Duker, and Hiroshi Ishikawa. *Optical coherence tomography of ocular diseases*. CRC Press, 2024.
- [233] Mingui Kong, Da Ye Choi, Gyule Han, Yun-Mi Song, Sung Yong Park, Jooheon Sung, Sungsoo Hwang, and Don-Il Ham. Measurable range of subfoveal choroidal thickness with conventional spectral domain optical coherence tomography. *Translational vision science & technology*, 7(5):16–16, 2018.
- [234] Diana Moukaddem, Niall Strang, Lyle Gray, Paul McGraw, and Chris Scholes. Comparison of diurnal variations in ocular biometrics and intraocular pressure between hyperopes and non-hyperopes. *Investigative Ophthalmology & Visual Science*, 63(7):1428–F0386, 2022.
- [235] Katharina Breher, Louise Terry, Thomas Bower, and Siegfried Wahl. Choroidal biomarkers: a repeatability and topographical comparison of choroidal thickness and choroidal vascularity index in healthy eyes. *Translational vision science & technology*, 9(11):8–8, 2020.
- [236] Justin Engelmann, Ana Villaplana-Velasco, Amos Storkey, and Miguel O Bernabeu. Robust and efficient computation of retinal fractal dimension through deep approximation. In *Ophthalmic Medical Image Analysis: 9th International Workshop, OMIA 2022, Held in Conjunction with MICCAI 2022, Singapore, Singapore, September 22, 2022, Proceedings*, pages 84–93. Springer, 2022.
- [237] Adria Perez-Rovira, T MacGillivray, Emanuele Trucco, KS Chin, K Zutis, C Lupascu, Domenico Tegolo, Andrea Giachetti, Peter J Wilson, A Doney, et al. Vampire: Vessel assessment and measurement platform for images of the retina. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3391–3394. IEEE, 2011.
- [238] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. IEEE, 2009.

- [239] Andrej Karpathy. tsnejs - t-sne visualization algorithm implementation in javascript. <https://github.com/karpathy/tsnejs>, 2016.
- [240] Christopher Williams, Fabian Falck, George Deligiannidis, Chris C Holmes, Arnaud Doucet, and Saifuddin Syed. A unified framework for u-net design and analysis. *Advances in Neural Information Processing Systems*, 36:27745–27782, 2023.
- [241] Veronika Cheplygina, Marleen De Bruijne, and Josien PW Pluim. Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Medical image analysis*, 54:280–296, 2019.
- [242] Meenal V Narkhede, Prashant P Bartakke, and Mukul S Sutaone. A review on weight initialization strategies for neural networks. *Artificial intelligence review*, 55(1):291–322, 2022.
- [243] Thomas Pearson, Yingdi Chen, Baljean Dhillon, Siddharthan Chandran, Jano van Hemert, and Tom MacGillivray. Multi-modal retinal scanning to measure retinal thickness and peripheral blood vessels in multiple sclerosis. *Scientific Reports*, 12(1):20472, 2022.
- [244] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1314–1324, 2019.
- [245] Pavel Iakubovskii. Segmentation models pytorch. [https://github.com/qubvel/segmentation\\_models.pytorch](https://github.com/qubvel/segmentation_models.pytorch), 2019.
- [246] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.
- [247] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

- [248] Ning Qian. On the momentum term in gradient descent learning algorithms. *Neural networks*, 12(1):145–151, 1999.
- [249] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7), 2011.
- [250] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.
- [251] Daniel Morales-Brotons, Thijs Vogels, and Hadrien Hendriks. Exponential moving average of weights in deep learning: Dynamics and benefits. *Transactions on Machine Learning Research*, 2024.
- [252] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of big data*, 6(1):1–48, 2019.
- [253] Priya Ranganathan, CS Pramesh, and Rakesh Aggarwal. Common pitfalls in statistical analysis: Measures of agreement. *Perspectives in clinical research*, 8(4):187, 2017.
- [254] Hongli Yang, Haomin Luo, Stuart K Gardiner, Christy Hardin, Glen P Sharpe, Joseph Caprioli, Shaban Demirel, Christopher A Girkin, Jeffrey M Liebmann, Christian Y Mardin, et al. Factors influencing optical coherence tomography peripapillary choroidal thickness: a multicenter study. *Investigative ophthalmology & visual science*, 60(2):795–806, 2019.
- [255] Kyoko Ohno-Matsui, Pei-Chang Wu, Kenji Yamashiro, Kritchai Vutipongsatorn, Yuxin Fang, Chui Ming Gemmy Cheung, Timothy YY Lai, Yasushi Ikuno, Salomon Yves Cohen, Alain Gaudric, et al. Idiopathic pathologic myopia. *Investigative ophthalmology & visual science*, 62(5):5–5, 2021.
- [256] Adrian T Fung, Yi Yang, and Andrew W Kam. Central serous chorioretinopathy: a review. *Clinical & Experimental Ophthalmology*, 51(3):243–270, 2023.

- [257] Veronica Vallino, Alessandro Berni, Andrea Coletto, Sonia Serafino, Francesco Bandello, Michele Reibaldi, and Enrico Borrelli. Structural oct and oct angiography biomarkers associated with the development and progression of geographic atrophy in amd. *Graefe's Archive for Clinical and Experimental Ophthalmology*, pages 1–16, 2024.
- [258] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [259] Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. Segment anything in medical images. *Nature Communications*, 15(1):654, 2024.
- [260] Yukun Zhou, Mark A Chia, Siegfried K Wagner, Murat S Ayhan, Dominic J Williamson, Robbert R Struyven, Timing Liu, Moucheng Xu, Mateo G Lozano, Peter Woodward-Court, et al. A foundation model for generalizable disease detection from retinal images. *Nature*, 622(7981):156–163, 2023.
- [261] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16000–16009, 2022.
- [262] Justin Engelmann and Miguel O Bernabeu. Training a high-performance retinal foundation model with half-the-data and 400 times less compute. *arXiv preprint arXiv:2405.00117*, 2024.
- [263] Reza Alizadeh Eghtedar, Mahdad Esmaeili, Alireza Peyman, Mohammadreza Akhlaghi, and Seyed Hossein Rasta. An update on choroidal layer segmentation methods in optical coherence tomography images: a review. *Journal of Biomedical Physics & Engineering*, 12(1):1, 2022.



- [264] Amrish Selvam, Matthew Driban, Joshua Ong, Sandeep Chandra Bollepalli, Jay Chhablani, and Kiran Kumar Vupparaboina. Artificial intelligence in choroid through optical coherence tomography: A comprehensive review. 2023.
- [265] Jason Walsh, Alice Othmani, Mayank Jain, and Soumyabrata Dev. Using u-net network for efficient brain tumor segmentation in mri images. *Healthcare Analytics*, 2:100098, 2022.
- [266] RV Manjunath and Karibasappa Kwadiki. Modified u-net on ct images for automatic segmentation of liver and its tumor. *Biomedical Engineering Advances*, 4:100043, 2022.
- [267] Wufeng Liu, Jiaxin Luo, Yan Yang, Wenlian Wang, Junkui Deng, and Liang Yu. Automatic lung segmentation in chest x-ray images using improved u-net. *Scientific Reports*, 12(1):8649, 2022.
- [268] Vahid Ashkani Chenarlogh, Mostafa Ghelich Oghli, Ali Shabanzadeh, Nasim Sirjani, Ardavan Akhavan, Isaac Shiri, Hossein Arabi, Morteza Sanei Taheri, and Mohammad Kazem Tarzamni. Fast and accurate u-net model for fetal ultrasound image segmentation. *Ultrasonic imaging*, 44(1):25–38, 2022.
- [269] Ignacio A Viedma, David Alonso-Caneiro, Scott A Read, and Michael J Collins. Deep learning in retinal optical coherence tomography (oct): A comprehensive survey. *Neurocomputing*, 507:247–264, 2022.
- [270] Vincent Loos, Rohit Pardasani, and Navchetan Awasthi. Demystifying the effect of receptive field size in u-net models for medical image segmentation. *arXiv preprint arXiv:2406.16701*, 2024.
- [271] Craig W. Ritchie and Karen Ritchie. The PREVENT study: a prospective cohort study to identify mid-life biomarkers of late-onset Alzheimer’s disease. *BMJ Open*, 2(6):e001893, January 2012. Publisher: British Medical Journal Publishing Group Section: Geriatric medicine.
- [272] Federico Cabitza, Andrea Campagner, Felipe Soares, Luis García de Guadiana-Romualdo, Feyissa Challa, Adela Sulejmani, Michela

- Seghezzi, and Anna Carobene. The importance of being external. methodological insights for the external validation of machine learning models in medicine. *Computer Methods and Programs in Biomedicine*, 208:106288, 2021.
- [273] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. pmlr, 2015.
- [274] Siegfried Karl Wagner, Fintan Hughes, Mario Cortina-Borja, Nikolas Pontikos, Robbert Struyven, Xiaoxuan Liu, Hugh Montgomery, Daniel C Alexander, Eric Topol, Steffen Erhard Petersen, et al. Alzeve: longitudinal record-level linkage of ophthalmic imaging and hospital admissions of 353 157 patients in london, uk. *BMJ open*, 12(3):e058552, 2022.
- [275] Elad Moisseiev, Anat Loewenstein, and Glenn Yiu. The suprachoroidal space: from potential space to a space with potential. *Clinical Ophthalmology (Auckland, N.Z.)*, 10:173–178, January 2016.
- [276] Csaba P Kovesdy. Epidemiology of chronic kidney disease: an update 2022. *Kidney International Supplements*, 12(1):7–11, 2022.
- [277] Clare MacRae, Stewart W Mercer, Bruce Guthrie, and David Henderson. Comorbidity in chronic kidney disease: a large cross-sectional study of prevalence in scottish primary care. *British Journal of General Practice*, 71(704):e243–e249, 2021.
- [278] Kitty J Jager, Csaba Kovesdy, Robyn Langham, Mark Rosenberg, Vivekanand Jha, and Carmine Zoccali. A single number for advocacy and communication—worldwide more than 850 million individuals have kidney diseases, 2019.
- [279] Alfons JHM Houben, Remy JH Martens, and Coen DA Stehouwer. Assessing microvascular function in humans from a chronic disease perspective. *Journal of the American Society of Nephrology*, 28(12):3461–3472, 2017.

- [280] Simon C Satchell and Filip Braet. Glomerular endothelial cell fenestrations: an integral component of the glomerular filtration barrier. *American Journal of Physiology-Renal Physiology*, 296(5):F947–F956, 2009.
- [281] Jasdeep Kaur, Benjamin E Young, and Paul J Fadel. Sympathetic overactivity in chronic kidney disease: consequences and mechanisms. *International journal of molecular sciences*, 18(8):1682, 2017.
- [282] J Martin Bland and Douglas G Altman. Correlation, regression, and repeated data. *BMJ: British Medical Journal*, 308(6933):896, 1994.
- [283] Jonathan Z Bakdash and Laura R Marusich. Repeated measures correlation. *Frontiers in psychology*, 8:456, 2017.
- [284] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.
- [285] Deepika Jain, Danny B Haddad, and Narender Goel. Choice of dialysis modality prior to kidney transplantation: does it matter? *World journal of nephrology*, 8(1):1, 2019.
- [286] Arthur G Krohn, David A Ogden, and Joseph H Holmes. Renal function in 29 healthy adults before and after nephrectomy. *Jama*, 196(4):322–324, 1966.
- [287] William E Moody, Charles J Ferro, Nicola C Edwards, Colin D Chue, Erica Lai Sze Lin, Robin J Taylor, Paul Cockwell, Richard P Steeds, and Jonathan N Townend. Cardiovascular effects of unilateral nephrectomy in living kidney donors. *Hypertension*, 67(2):368–377, 2016.
- [288] Giuseppe Mulè, Maria Vadalà, Tiziana La Blasca, Rossella Gaetani, Giuseppe Virone, Marco Guarneri, Massimo Castellucci, Giulia Guarasi, Micol Terrasi, and Santina Cottone. Association between early-stage chronic kidney disease and reduced choroidal thickness in essential hypertensive patients. *Hypertension Research*, 42(7):990–1000, 2019.

- [289] Robert G Anderson, Anton J Bueschen, L Keith Lloyd, Eva V Dubovsky, and John R Burns. Short-term and long-term changes in renal function after donor nephrectomy. *The Journal of urology*, 145(1):11–13, 1991.
- [290] Darling M Rojas-Canales, Jordan Y Li, Leek Makuei, and Jonathan M Gleadle. Compensatory renal hypertrophy following nephrectomy: When and how? *Nephrology*, 24(12):1225–1232, 2019.
- [291] Bengt Winblad, Philippe Amouyel, Sandrine Andrieu, Clive Ballard, Carol Brayne, Henry Brodaty, Angel Cedazo-Minguez, Bruno Dubois, David Edvardsson, Howard Feldman, et al. Defeating alzheimer’s disease and other dementias: a priority for european science and society. *The Lancet Neurology*, 15(5):455–532, 2016.
- [292] Gill Livingston, Jonathan Huntley, Andrew Sommerlad, David Ames, Clive Ballard, Sube Banerjee, Carol Brayne, Alistair Burns, Jiska Cohen-Mansfield, Claudia Cooper, et al. Dementia prevention, intervention, and care: 2020 report of the lancet commission. *The lancet*, 396(10248):413–446, 2020.
- [293] Hannah Stocker, Laura Perna, Korbinian Weigl, Tobias Möllers, B Schötter, Hauke Thomsen, Bernd Holleczek, Dan Rujescu, and Hermann Brenner. Prediction of clinical diagnosis of alzheimer’s disease, vascular, mixed, and all-cause dementia by a polygenic risk score and apoe status in a community-based cohort prospectively followed over 17 years. *Molecular psychiatry*, 26(10):5812–5822, 2021.
- [294] Daniela Scarabino, Giuseppe Gambina, Elisabetta Broggio, Franca Pelliccia, and Rosa Maria Corbo. Influence of family history of dementia in the development and progression of late-onset alzheimer’s disease. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 171(2):250–256, 2016.
- [295] Elijah Mak, Maria-Eleni Dounavi, Audrey Low, Stephen F Carter, Elizabeth McKiernan, Guy B Williams, P Simon Jones, Isabelle Carriere, Graciela Terrera Muniz, Karen Ritchie, et al. Proximity to dementia onset and

- multi-modal neuroimaging changes: The prevent-dementia study. *Neuroimage*, 229:117749, 2021.
- [296] Samuel Chiquita, Ana C Rodrigues-Neves, Filipa I Baptista, Rafael Carecho, Paula I Moreira, Miguel Castelo-Branco, and António F Ambrósio. The retina as a window or mirror of the brain changes detected in alzheimer's disease: critical aspects to unravel. *Molecular neurobiology*, 56:5416–5435, 2019.
- [297] Chiara Villa. Ocular vascular changes: Choroidal thickness as an early biomarker for alzheimer's disease?, 2021.
- [298] Elena Salobar García Martín. Glia and blood retinal barrier: Effects of ocular hypertension. In *Cardiovascular Disease II*, pages 123–162. iConcept Press Ltd, 2014.
- [299] Hasan Ali Bayhan, Seray Aslan Bayhan, Asuman Celikbilek, Nermin Tanık, and Canan Gürdal. Evaluation of the chorioretinal thickness changes in a lzheimer's disease using spectral-domain optical coherence tomography. *Clinical & experimental ophthalmology*, 43(2):145–151, 2015.
- [300] Mehmet Bulut, Aylin Yaman, Muhammet Kazim Erol, Fatma Kurtuluş, Devrim Toslak, Berna Doğan, Deniz Turgut Çoban, and Ebru Kaya Başar. Choroidal thickness in patients with mild cognitive impairment and alzheimer's type dementia. *Journal of ophthalmology*, 2016(1):7291257, 2016.
- [301] Joao Paulo Cunha, Rita Proença, Arnaldo Dias-Santos, Diana Melancia, Rita Almeida, Helena Aguas, Bruno Oliveira Santos, Marta Alves, Joana Ferreira, Ana Luísa Papoila, et al. Choroidal thinning: Alzheimer's disease and aging. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 8:11–17, 2017.
- [302] Craig W Ritchie, Katie Bridgeman, Sarah Gregory, John T O'Brien, Samuel O Danso, Maria-Eleni Dounavi, Isabelle Carriere, David Driscoll, Robert Hillary, Ivan Koychev, Brian Lawlor, Lorina Naci, Li Su, Audrey

- Low, Elijah Mak, Paresh Malhotra, Jean Manson, Riccardo Marioni, Lee Murphy, Georgios Ntalianis, William Stewart, Graciela Muniz-Terrera, Karen Ritchie, and on behalf of the PREVENT dementia programme group. The PREVENT dementia programme: baseline demographic, lifestyle, imaging and cognitive data from a midlife cohort study investigating risk factors for dementia. *Brain Communications*, 6(3):fcae189, June 2024.
- [303] K Polak, E Polska, A Luksch, G Dorner, G Fuchsjäger-Mayrl, O Findl, HG Eichler, M Wolzt, and L Schmetterer. Choroidal blood flow and arterial blood pressure. *Eye*, 17(1):84–88, 2003.
- [304] Luke Thomas Sansom, Carlo A Suter, and Martin McKibbin. The association between systolic blood pressure, ocular perfusion pressure and subfoveal choroidal thickness in normal individuals. *Acta Ophthalmologica*, 94(2):e157–e158, 2016.
- [305] Shalaka R Waghmare, Sunita Mittal, Monika Pathania, Ramanuj Samanta, Devesh Kumawat, Neeti Gupta, and Sanjeev K Mittal. Comparison of choroidal thickness in systemic hypertensive subjects with healthy individuals by spectral domain optical coherence tomography. *Indian journal of ophthalmology*, 69(5):1183–1188, 2021.
- [306] Jie Lu, Hao Zhou, Yingying Shi, James Choe, Mengxi Shen, Liang Wang, Kelly Chen, Qinqin Zhang, William J Feuer, Giovanni Gregori, et al. Interocular asymmetry of choroidal thickness and vascularity index measurements in normal eyes assessed by swept-source optical coherence tomography. *Quantitative imaging in medicine and surgery*, 12(1):781, 2022.
- [307] Heidelberg Engineering. *SPECTRALIS Product Family User Manual sw ver 7.0*. Number 230131-008 INT.AE22. Version 7 edition, June 2022.
- [308] Yilin Liu, Lijun Wang, Yanyun Xu, Zuoxiang Pang, and Guoying Mu. The influence of the choroid on the onset and development of myopia: from perspectives of choroidal thickness and blood flow. *Acta Ophthalmologica*, 99(7):730–738, 2021.

- [309] Inés López-Cuenca, Lidia Sánchez-Puebla, Elena Salobrar-García, María Álvarez-Gutierrez, Lorena Elvira-Hurtado, Ana Barabash, Federico Ramírez-Toraño, José A Fernández-Albarral, José A Matamoros, Alberto Nebreda, et al. Exploratory longitudinal study of ocular structural and visual functional changes in subjects at high genetic risk of developing alzheimer's disease. *Biomedicines*, 11(7):2024, 2023.
- [310] Colin S. Tan, Yanling Ouyang, Humberto Ruiz, and Srinivas R. Sadda. Diurnal Variation of Choroidal Thickness in Normal, Healthy Subjects Measured by Spectral Domain Optical Coherence Tomography. *Investigative Ophthalmology & Visual Science*, 53(1):261–266, January 2012.
- [311] Carlos Lahoz, Ernst J Schaefer, L Adrienne Cupples, Peter WF Wilson, Daniel Levy, Doreen Osgood, Stefanos Parpos, Juan Pedro-Botet, Jennifer A Daly, and Jose M Ordovas. Apolipoprotein e genotype and cardiovascular disease in the framingham heart study. *Atherosclerosis*, 154(3):529–537, 2001.
- [312] Axel Montagne, Daniel A Nation, Abhay P Sagare, Giuseppe Barisano, Melanie D Sweeney, Ararat Chakhoyan, Maricarmen Pachicano, Elizabeth Joe, Amy R Nelson, Lina M D'Orazio, et al. Apoe4 leads to blood–brain barrier dysfunction predicting cognitive decline. *Nature*, 581(7806):71–76, 2020.
- [313] Qiushan Tao, Ting Fang Alvin Ang, Charles DeCarli, Sanford H Auerbach, Sheral Devine, Thor D Stein, Xiaoling Zhang, Joseph Massaro, Rhoda Au, and Wei Qiao Qiu. Association of chronic low-grade inflammation with risk of alzheimer disease in apoe4 carriers. *JAMA network open*, 1(6):e183597–e183597, 2018.
- [314] Katrine L Rasmussen, Anne Tybjærg-Hansen, Børge G Nordestgaard, and Ruth Frikke-Schmidt. Associations of alzheimer disease–protective apoe variants with age-related macular degeneration. *JAMA ophthalmology*, 141(1):13–21, 2023.
- [315] Perttu J Liuska, Joel T Rämö, Susanna Lemmelä, Kai Kaarniranta, Hannu Uusitalo, Elisa Lahtela, Mark J Daly, Mika Harju, Aarno Palotie,

- Joni A Turunen, et al. Association of apoe haplotypes with common age-related ocular diseases in 412,171 individuals. *Investigative Ophthalmology & Visual Science*, 64(14):33–33, 2023.
- [316] Trevor Lohman, Arunima Kapoor, Allison C Engstrom, Fatemah Shenasa, John Paul M Alitin, Aimee Gaubert, Kathleen E Rodgers, David Bradford, Mara Mather, S Duke Han, et al. Central autonomic network dysfunction and plasma alzheimer's disease biomarkers in older adults. *Alzheimer's Research & Therapy*, 16(1):124, 2024.
- [317] Trevor Lohman, Isabel Sible, Arunima Kapoor, Allison C Engstrom, Fatemah Shenasa, John Paul M Alitin, Aimee Gaubert, Kathleen E Rodgers, David Bradford, Mara Mather, et al. Blood pressure variability, central autonomic network dysfunction, and cerebral small-vessel disease in apoe4 carriers. *Journal of the American Heart Association*, 13(9):e034116, 2024.
- [318] Orla Collins, Shiela Dillon, Ciaran Finucane, Brian Lawlor, and Rose Anne Kenny. Parasympathetic autonomic dysfunction is common in mild cognitive impairment. *Neurobiology of aging*, 33(10):2324–2333, 2012.
- [319] Mirna Safieh, Amos D Korczyn, and Daniel M Michaelson. Apoe4: an emerging therapeutic target for alzheimer's disease. *BMC medicine*, 17:1–17, 2019.
- [320] William Robert Kwapong, Fei Tang, Peng Liu, Ziyi Zhang, Le Cao, Zijuan Feng, Shiyun Yang, Yang Shu, Heng Xu, Ying Lu, et al. Choriocapillaris reduction accurately discriminates against early-onset alzheimer's disease. *Alzheimer's & Dementia*, 2024.
- [321] Yasser Sakr, Veronika Gath, Janaina Oishi, Stefan Klinzing, Tim-Phillip Simon, Konrad Reinhart, and Gernot Marx. Characterization of buccal microvascular response in patients with septic shock. *European Journal of Anaesthesiology*, 27(4):388–394, April 2010.



- [322] Sefer Elezkurtaj, Selina Greuel, Jana Ihlow, Edward Georg Michaelis, Philip Bischoff, Catarina Alisa Kunze, Bruno Valentin Sinn, Manuela Gerhold, Kathrin Hauptmann, Barbara Ingold-Heppner, et al. Causes of death and comorbidities in hospitalized patients with covid-19. *Scientific reports*, 11(1):4263, 2021.
- [323] Raúl J. Gazmuri and Cristina Añez de Gomez. From a pressure-guided to a perfusion-centered resuscitation strategy in septic shock: Critical literature review and illustrative case. *Journal of Critical Care*, 56:294–304, April 2020.
- [324] null null. A Randomized Trial of Protocol-Based Care for Early Septic Shock. *New England Journal of Medicine*, 370(18):1683–1693, 2014. Publisher: Massachusetts Medical Society \_eprint: <https://www.nejm.org/doi/pdf/10.1056/NEJMoa1401602>.
- [325] Mouncey Paul R., Osborn Tiffany M., Power G. Sarah, Harrison David A., Sadique M. Zia, Grieve Richard D., Jahan Rahi, Harvey Sheila E., Bell Derek, Bion Julian F., Coats Timothy J., Singer Mervyn, Young J. Duncan, and Rowan Kathryn M. Trial of Early, Goal-Directed Resuscitation for Septic Shock. *New England Journal of Medicine*, 372(14):1301–1311, 2015. Publisher: Massachusetts Medical Society \_eprint: <https://www.nejm.org/doi/pdf/10.1056/NEJMoa1500896>.
- [326] null null. Early, Goal-Directed Therapy for Septic Shock — A Patient-Level Meta-Analysis. *New England Journal of Medicine*, 376(23):2223–2234, 2017. Publisher: Massachusetts Medical Society \_eprint: <https://www.nejm.org/doi/pdf/10.1056/NEJMoa1701380>.
- [327] Tine S Meyhoff, Peter B Hjortrup, Jørn Wetterslev, Praleene Sivapalan, Jon H Laake, Maria Cronhjort, Stephan M Jakob, Maurizio Cecconi, Marek Nalos, Marlies Ostermann, et al. Restriction of intravenous fluid in icu patients with septic shock. *New England Journal of Medicine*, 386(26):2459–2470, 2022.
- [328] Liam Byrne, Nchafatso G Obonyo, Sara D Diab, Kimble R Dunster, Margaret R Passmore, Ai-Ching Boon, Louise See Hoe, Sanne Pedersen,

- Mohd Hashairi Fauzi, Leticia Pretti Pimenta, et al. Unintended consequences: fluid resuscitation worsens shock in an ovine model of endotoxemia. *American Journal of Respiratory and Critical Care Medicine*, 198(8):1043–1054, 2018.
- [329] Maitland Kathryn, Kiguli Sarah, Opoka Robert O., Engoru Charles, Olupot-Olupot Peter, Akech Samuel O., Nyeko Richard, Mtove George, Reyburn Hugh, Lang Trudie, Brent Bernadette, Evans Jennifer A., Tibenderana James K., Crawley Jane, Russell Elizabeth C., Levin Michael, Babiker Abdel G., and Gibb Diana M. Mortality after Fluid Bolus in African Children with Severe Infection. *New England Journal of Medicine*, 364(26):2483–2495, 2011. Publisher: Massachusetts Medical Society \_eprint: <https://www.nejm.org/doi/pdf/10.1056/NEJMoa1101549>.
- [330] Sam D. Hutchings, David N. Naumann, Philip Hopkins, Clare Mellis, Paul Riozzi, Stefano Sartini, Jasna Mamuza, Tim Harris, Mark J. Midwinter, and Julia Wendon. Microcirculatory Impairment Is Associated With Multiple Organ Dysfunction Following Traumatic Hemorrhagic Shock: The MICROSHOCK Study. *Critical Care Medicine*, 46(9):e889, September 2018.
- [331] Toshiaki Iba, Cheryl L Maier, Julie Helms, Ricard Ferrer, Jecko Thachil, and Jerrold H Levy. Managing sepsis and septic shock in an endothelial glycocalyx-friendly way: from the viewpoint of surviving sepsis campaign guidelines. *Annals of Intensive Care*, 14(1):64, 2024.
- [332] Ross R McMullan, Daniel F McAuley, Cecilia M O’Kane, and Jonathan A Silversides. Vascular leak in sepsis: physiological basis and potential therapeutic advances. *Critical Care*, 28(1):97, 2024.
- [333] JERROLD H Levy and IBA Toshiaki. Endothelial glycocalyx protection in sepsis. *Juntendo Medical Journal*, 70(1):23–25, 2024.
- [334] ICS. Levels of care.

- [335] Thomas Kuriakose. The Slit Lamp Examination. In Thomas Kuriakose, editor, *Clinical Insights and Examination Techniques in Ophthalmology*, pages 55–62. Springer, Singapore, 2020.
- [336] Srinivas R. Sadda, Andrew P. Schachat, Charles P. Wilkinson, David R. Hinton, Peter Wiedemann, K. Bailey Freund, and David Sarraf. *Ryan's Retina E-Book*. Elsevier Health Sciences, April 2022. Google-Books-ID: 6H9qEAAAQBAJ.
- [337] Nadia K. Waheed, Luísa S. M. Mendonça, and Lucy H. Young. Choroidal Effusions and Detachments. In Daniel M. Albert, Joan W. Miller, Dimitri T. Azar, and Lucy H. Young, editors, *Albert and Jakobiec's Principles and Practice of Ophthalmology*, pages 3857–3874. Springer International Publishing, Cham, 2022.
- [338] Jamie Burke, Samuel Gibbon, Justin Engelmann, Adam Threlfall, Ylenia Giarratano, Charlene Hamid, Stuart King, Ian JC MacCormick, and Thomas J MacGillivray. Sloctolyzer: Fully automatic analysis toolkit for segmentation and feature extracting in scanning laser ophthalmoscopy images. *Translational Vision Science & Technology*, 13(11):7–7, 2024.
- [339] Olivier Morelle. Eyepy. <https://github.com/MedVisBonn/eyepy/tree/master>, 2023.
- [340] Mark Graham. Eyepy. <https://github.com/marksgraham/OCT-Converter>, 2024.
- [341] Olga Ciobanu-Caraus, Anatol Aicher, Julius M Kernbach, Luca Regli, Carlo Serra, and Victor E Staartjes. A critical moment in machine learning in medicine: on reproducible and interpretable learning. *Acta Neurochirurgica*, 166(1):14, 2024.
- [342] E Hope Weissler, Tristan Naumann, Tomas Andersson, Rajesh Ranganath, Olivier Elemento, Yuan Luo, Daniel F Freitag, James Benoit, Michael C Hughes, Faisal Khan, et al. The role of machine learning in clinical research: transforming the future of evidence generation. *Trials*, 22:1–15, 2021.

- [343] Matthew BA McDermott, Shirly Wang, Nikki Marinsek, Rajesh Ranganath, Luca Foschini, and Marzyeh Ghassemi. Reproducibility in machine learning for health research: Still a ways to go. *Science Translational Medicine*, 13(586):eabb1655, 2021.
- [344] Joelle Pineau, Philippe Vincent-Lamarre, Koustuv Sinha, Vincent Larivière, Alina Beygelzimer, Florence d’Alché Buc, Emily Fox, and Hugo Larochelle. Improving reproducibility in machine learning research (a report from the neurips 2019 reproducibility program). *Journal of machine learning research*, 22(164):1–20, 2021.
- [345] Keerthi B Harish, W Nicholson Price, and Yindalon Aphinyanaphongs. Open-source clinical machine learning models: critical appraisal of feasibility, advantages, and challenges. *JMIR Formative Research*, 6(4):e33970, 2022.
- [346] Sandeep Reddy. Explainability and artificial intelligence in medicine. *The Lancet Digital Health*, 4(4):e214–e215, 2022.
- [347] Xiaoxuan Liu, Livia Faes, Aditya U Kale, Siegfried K Wagner, Dun Jack Fu, Alice Bruynseels, Thushika Mahendiran, Gabriella Moraes, Mohith Shamdas, Christoph Kern, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The lancet digital health*, 1(6):e271–e297, 2019.
- [348] Xiaoxuan Liu, Samantha Cruz Rivera, David Moher, Melanie J Calvert, Alastair K Denniston, Hutan Ashrafian, Andrew L Beam, An-Wen Chan, Gary S Collins, Ara Darzi, Jonathan J Deeks, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the consort-ai extension. *The Lancet Digital Health*, 2(10):e537–e548, 2020.
- [349] Sven Van Laere, Katoo M Muylle, and Pieter Cornu. Clinical decision support and new regulatory frameworks for medical devices: are we ready for it?-a viewpoint paper. *International Journal of Health Policy and Management*, 11(12):3159, 2022.

- [350] William Boag, Alifia Hasan, Jee Young Kim, Mike Revoir, Marshall Nichols, William Ratliff, Michael Gao, Shira Zilberstein, Zainab Samad, Zahra Hoodbhoy, et al. The algorithm journey map: a tangible approach to implementing ai solutions in healthcare. *NPJ Digital Medicine*, 7(1):87, 2024.
- [351] Monica M Bertagnolli. Advancing health through artificial intelligence/-machine learning: the critical importance of multidisciplinary collaboration, 2023.
- [352] Joes Staal, Michael D Abràmoff, Meindert Niemeijer, Max A Viergever, and Bram Van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE transactions on medical imaging*, 23(4):501–509, 2004.
- [353] Zhuo Zhang, Feng Shou Yin, Jiang Liu, Wing Kee Wong, Ngan Meng Tan, Beng Hai Lee, Jun Cheng, and Tien Yin Wong. Origa-light: An online retinal fundus image database for glaucoma analysis and research. In *2010 Annual international conference of the IEEE engineering in medicine and biology*, pages 3065–3068. IEEE, 2010.
- [354] AD Hoover, Valentina Kouznetsova, and Michael Goldbaum. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical imaging*, 19(3):203–210, 2000.
- [355] Francisco Fumero, Silvia Alayón, José L Sanchez, Jose Sigut, and M Gonzalez-Hernandez. Rim-one: An open retinal image database for optic nerve evaluation. In *2011 24th international symposium on computer-based medical systems (CBMS)*, pages 1–6. IEEE, 2011.
- [356] Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, et al. Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine*, 12(3):e1001779, 2015.

- [357] Xin Ye, Shucheng He, Xiaxing Zhong, Jiafeng Yu, Shangchao Yang, Yingjiao Shen, Yiqi Chen, Yaqi Wang, Xingru Huang, and Lijun Shen. Oimhs: An optical coherence tomography image dataset based on macular hole manual segmentation. *Scientific Data*, 10(1):769, 2023.
- [358] Kornélia Lenke Laurik-Feuerstein, Rishav Sapahia, Delia Cabrera De-Buc, and Gábor Márk Somfai. The assessment of fundus image quality labeling reliability among graders with different backgrounds. *PloS one*, 17(7):e0271156, 2022.
- [359] Charlene Hamid, Justin Engelmann, Fiona Buckmaster, Fabian Yii, and Miguel Bernabeu. Finequal: A taxonomy for fine-grained, functional quality assessment of retinal colour fundus images and triple-graded assessment of a random sample of uk biobank images. *Investigative Ophthalmology & Visual Science*, 65(7):5952–5952, 2024.
- [360] Yue Ming, Xuyang Meng, Chunxiao Fan, and Hui Yu. Deep learning for monocular depth estimation: A review. *Neurocomputing*, 438:14–33, 2021.
- [361] Yihua Sun, Dawei Li, Seongho Kim, Ya Xing Wang, Jinyuan Wang, Tien Yin Wong, Hongen Liao, and Su Jeong Song. Retinal thickness prediction from multi-modal fundus photography. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 585–595. Springer, 2023.

# Appendices





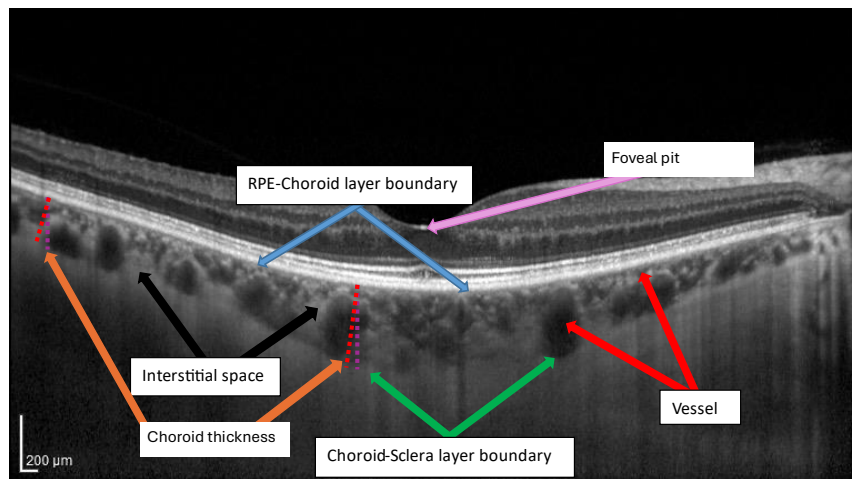
# Appendix A

## Protocol Documentation

### A.1 OCT B-scan definitions

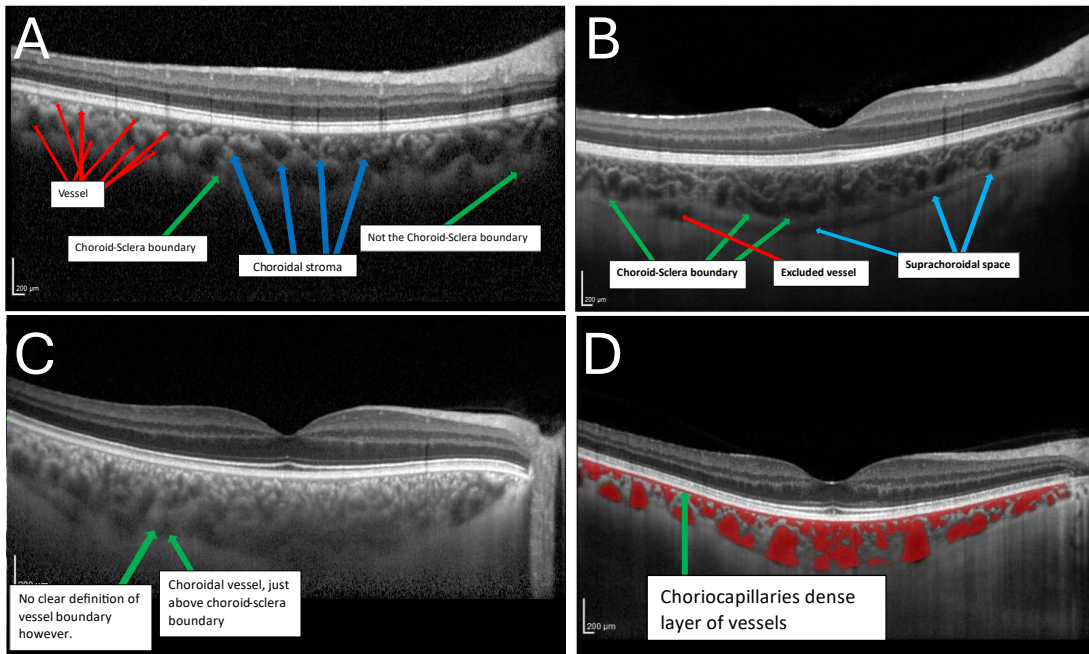
Below are a list of definitions of various landmarks related to the choroid, as seen on an OCT B-scan. The definitions relate to the features on the B-scan, rather than the biology. See figure [S1](#) for an exemplar OCT B-scan with regions of the choroid labelled.

- **RPE-Choroid boundary:** The lower surface of the junction between the hyperreflective RPE layer and Bruch's membrane complex (figure [S1](#), blue).
- **Choroid-sclera boundary:** The upper surface of the junction between the choroid and sclera. A predominantly smooth and concave line posterior to the posterior-most choroidal vessels (figure [S1](#), green).
- **Choroid thickness:** This measures a straight-line extending from the RPE-Choroid boundary to the Choroid-Sclera boundary, measured locally perpendicular to the RPE-Choroid boundary (figure [S1](#), red-dashed line), not vertically (figure [S1](#), purple-dashed line).
- **Choroidal space:** Region between the RPE-Choroid and Choroid-Sclera boundaries.



**Figure S1:** OCT B-scan with arrows directed at various features of the choroid.

- **Choriocapillaris:** A small strip below and parallel to the RPE-Choroid boundary, which extends across the choroid seen in OCT B-scans. Typically  $\approx 3$  pixels deep for Heidelberg Engineering Spectralis OCT1 devices. This is typically included in segmentations as a thin strip below the RPE-Choroid boundary (figure S2(D)).
- **Choroidal medium/large vessel:** Large, dark and irregularly shaped and sized blob seen with lower intensity relative to locally adjacent and higher intensity strips in the choroid. This is located within the Sattler and Haller's layers (figure S2(A)).
- **Choroidal small vessel:** Smaller, but still irregularly shaped and sized blobs than larger choroidal vessels and appear below the choriocapillaris (figure S2(A)). These smaller vessels are darker in intensity than adjacent regions, but their contrast is much poorer than larger choroidal vessels.
- **Choroid vessel wall/boundary:** Edges of choroidal vasculature are inherently more difficult to define as they blend into adjacent interstitial fluid – their exact definition is image-dependent based on local contrast, but is generally brighter than the centre of the vessel itself (figure S2(C)).
- **Choroidal vasculature:** Non-uniform distribution of irregularly sized and shaped blobs constituting the choriocapillaris, medium/large and small vessels. The vasculature can suffer from poor contrast and visualisa-



**Figure S2:** (A) OCT B-scan with arrows indicating Choroid-Sclera boundary (green), vessels (red) and stroma (blue). (B) OCT B-scan with visible supra-choroidal space (blue) and specification of Choroid-Sclera boundary (green) and excluded vessel (red). (C) Large choroid with poor contrast vessel visualisation (green arrow). (D) OCT B-scan with vessel segmentation overlaid with choriocapillaris band (green arrow).

tion, making vessel walls difficult to define, even for larger vessels (figure S2(C)).

- **Choroidal stroma/interstitial space:** Brightly illuminated regions distributed around choroidal vasculature in the choroidal space (figure S2(A)).
- **Suprachoroid:** A potential space posterior to the Choroid-Sclera boundary, and is visualised as a faint, homogeneous strip parallel with the Choroid-Sclera boundary between the brighter sclera and heterogeneous (in intensity) choroidal space. This, and any vessels traversing it, is **not** considered part of the choroidal space and should be avoided during segmentation or adjudication (figure S2(B)).
- **Foveal pit:** Pixel of highest illumination at the deepest point of the foveola depression in the B-scan (figure S1, pink), typically vertically-aligned with

a ridge forming at the anterior point of the outer retina.

## A.2 Equipment check

Be it adjudication or segmentation, the following list describes essential equipment. Figure S3 shows the ideal monitor and zoom setting for viewing an OCT B-scan for adjudication or segmentation.

- You will need a separate mouse and mouse pad, not a laptop trackpad.
- You will need a large monitor with high resolution (greater than the standard  $1366 \times 768$  pixel resolution, i.e.  $1920 \times 1080$  pixels) so that you can ideally see the whole B-scan and detailed segmentation together (figure S3).
- Two monitors together are also ideal for pairwise, visual comparisons, but not strictly necessary.



**Figure S3:** Example setup with ITK-Snap open covering the entire screen of a standard HP monitor. Image courtesy of Dr. Ian J.C. MacCormick.

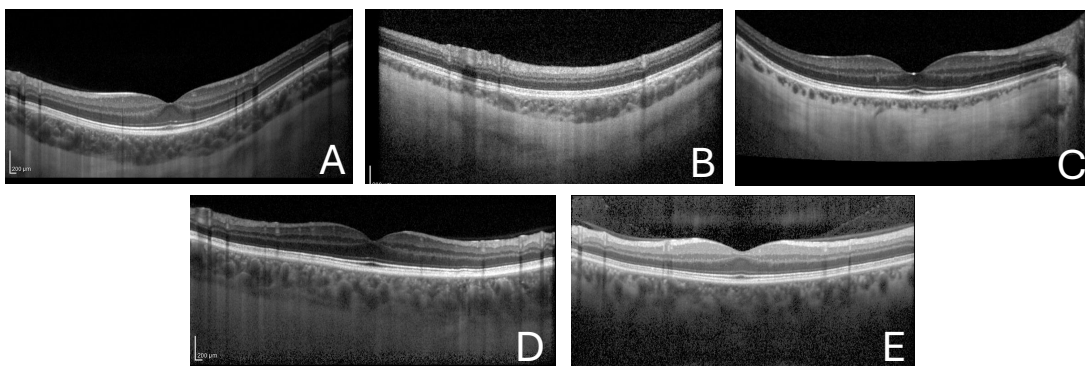
- Use the zoom function at your discretion for assessing smaller-scale regions.
- Perform adjudication in a darkly lit room, for example, the lights off on a dull day without much sunlight coming in.

### A.3 OCT B-scan quality guide

Below is a list of definitions for each scale when assessing the quality of the choroidal region in OCT B-scans. Accompanying each scale is a representative OCT B-scan, shown in figure S4.

#### A.3.1 Choroidal space OCT image quality

- **2 Very good:** The definition of the Choroid-Sclera boundary is excellent, and is particularly enhanced in regions by darker, posterior-most vessels above it. Its definition is clear across the whole OCT B-scan (figure S5(A)).
- **1 Good:** The contrast of the Choroid-Sclera boundary is still greater than Haller's layer of the choroid above it, but elements of poor signal (i.e. speckle noise), or interstitial fluid is obscuring parts across the OCT B-scan (relative to Very good) (figure S5(B)).



**Figure S4:** Exemplar OCT B-scans to accompany each scale to assess image quality for the choroidal space.

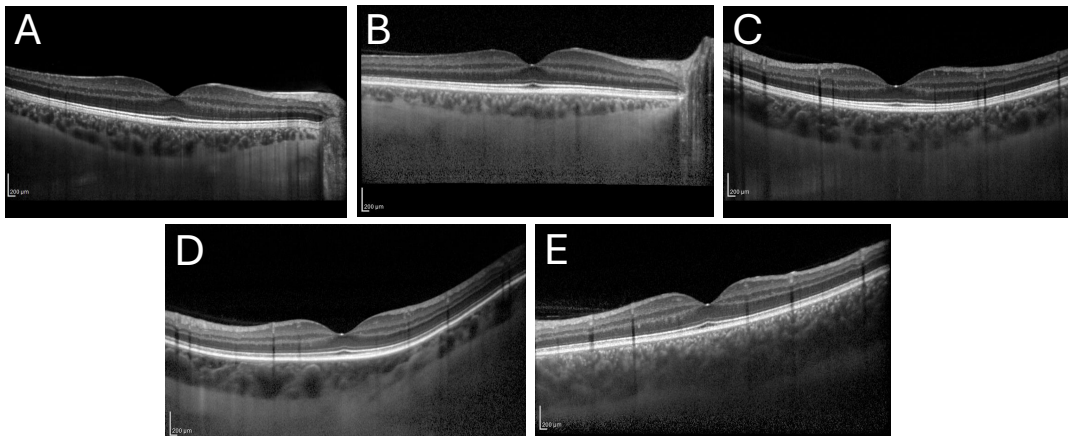
- **0 Okay:** Choroid-Sclera boundary is still visible enough to be traced, but the contrast between it and the choroidal space is poor (relative to Good) (figure S5(C)).
- **-1 Bad:** Exact definition of Choroid-Sclera boundary is uncertain and made “fuzzy” by poor signal penetration and speckle noise — posterior-most vessels unable to be used as a reference/guide to trace boundary (figure S5(D)).
- **-2 Very bad:** The choroid region is cropped in parts, or there is little signal penetrating deep enough to make the Choroid-Sclera boundary visible in most locations along the OCT B-scan (figure S5(E)).

### A.3.2 Choroidal vasculature OCT image quality

Below is a list of definitions for each scale when assessing the quality of the choroidal vasculature in OCT B-scans. Accompanying each scale is a representative OCT B-scan, shown in figure S5.

- **2 Very good:** Large contrast between large vessels and interstitial fluid, with clear vessel wall definition. Smaller vessels are easily identifiable relative to interstitial fluid (figure S5(A)).
- **1 Good:** Large vessels still significantly darker in intensity than adjacent interstitial fluid, but vessel wall definition has poorer contrast. Small vessels are still easily identifiable. Some medium vessels aren’t as dark as larger ones, leading to some uncertainty in their vessel wall definition (figure S5(B)).
- **0 Okay:** Good definition across all vessels, but many vessels clumped together so vessel wall separation challenging, with speckle noise corrupting vessels at all scales (figure S5(C)).
- **-1 Bad:** Low contrast choroid and speckle noise corrupting vessel wall definition, the centre of larger vessels still visible, but smaller vasculature difficult to confidently define (figure S5(D)).





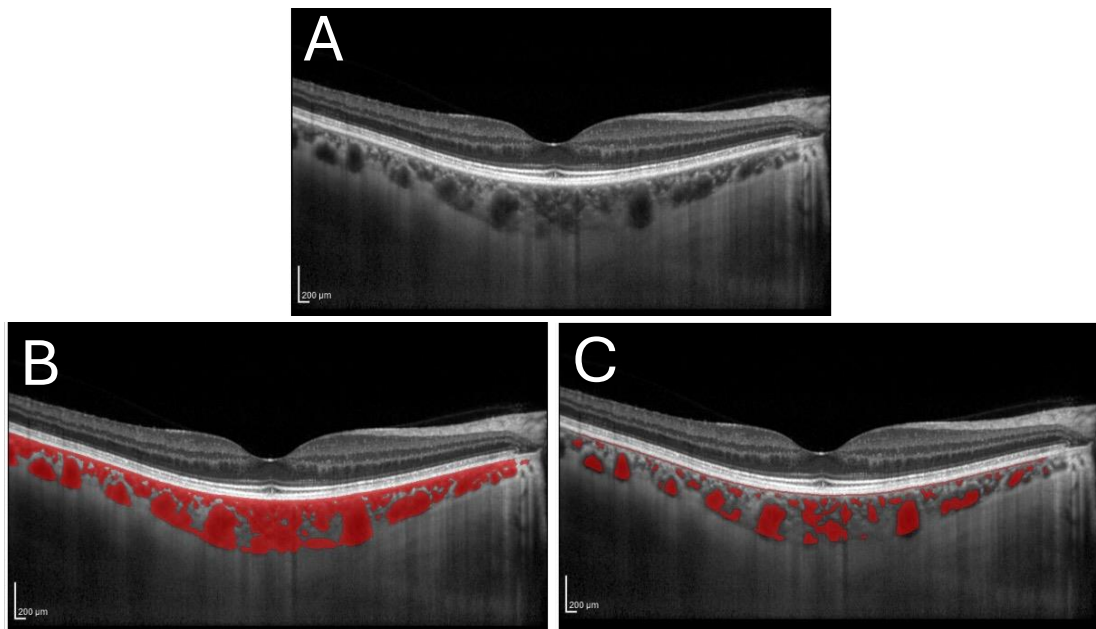
**Figure S5:** Exemplar OCT B-scans to accompany each scale to assess image quality for the choroidal vasculature.

- **-2 Very bad:** Choroidal region is cropped and/or speckle noise destroys any clear vessel wall definition for larger vessels, and smaller vessels are indescribable relative to adjacent interstitial fluid (figure S5(E)).

## A.4 OCT B-scan vessel guide

The example in figure S6 highlights the effect of over- and under-segmentation of the choroidal vasculature. In figure S6(B), over-segmentation leads to all vessel pixels correctly segmented, as well as much of the interstitial space. This represents good *completeness of vessels* but poor *interstitial segmentation*. Conversely, in figure S6(C), under-segmentation leads to only vessel pixels segmented, but many pixels representing vessels were miss-classified as interstitial space (not coloured red). This represents poor *completeness of vessels* but good *interstitial segmentation*. Figure S2(D) shows the ideal segmentation for this OCT B-scan.

The two examples above highlight an inherent problem with choroidal vessel segmentation. There are two goals in mind: successfully segmenting every vessel in the intravascular space, and successfully excluding all pixels representing interstitial fluid/choroidal stroma. In reality, contrast issues and noise corruption can make this problem a very difficult one, leading to uncertainty in



**Figure S6:** (A) OCT B-scan. (B) Over-segmentation and (C) under-segmentation of the choroidal vasculature.

what constitutes a vessel boundary or not. This results in two design choices made on the objective behind choroidal vessel segmentation:

1. Segment all vessels at the cost of segmenting interstitial fluid as well (over-segmentation)
2. The segmentation only identifies vessel in order to exclude the interstitial space, at the cost of potentially missing pixels representing vessel boundaries/low contrast vessels.

We believe (2) is the preferred approach, so that the fidelity of the vessels are preserved. Nevertheless, it's important a segmentation technique be consistent and repeatable, using the same definition/design choice across all images (which is something quite difficult for a human grader to apply).



## A.5 OCT B-scan manual segmentation protocol

### A.5.1 ITK-SNAP setup

1. Go to ITK-SNAP Version 3.x Downloads and click the link for ITK-SNAP 3.8.0 (stable version) windows Self-Install Package (64-bit).
2. Fill in the form and select next. On the next page and select the download. It will take you to another page and the download should start.
3. Navigate to the downloaded `.exe` file and double click to begin the install process.
4. Select next, agree to the terms and conditions and select next. Tick the box next to 'Do not add SNAP to the system PATH', keep selecting next until the setup is complete, press finish.
5. Navigate to the SNAP program and open, it will prompt you if you would like to enable automatic updates, select yes.
6. **(Optional)** Go to [RSNA 2016 ITK-SNAP Training - Intro & Manual Segmentation - YouTube](#) and watch the 15min video. This is a training video to learn how to use SNAP, they give an example in 3D but we will be annotating in 2D. The slides for the training video can be found [here](#).

### A.5.2 Load in image, label definition, main tools

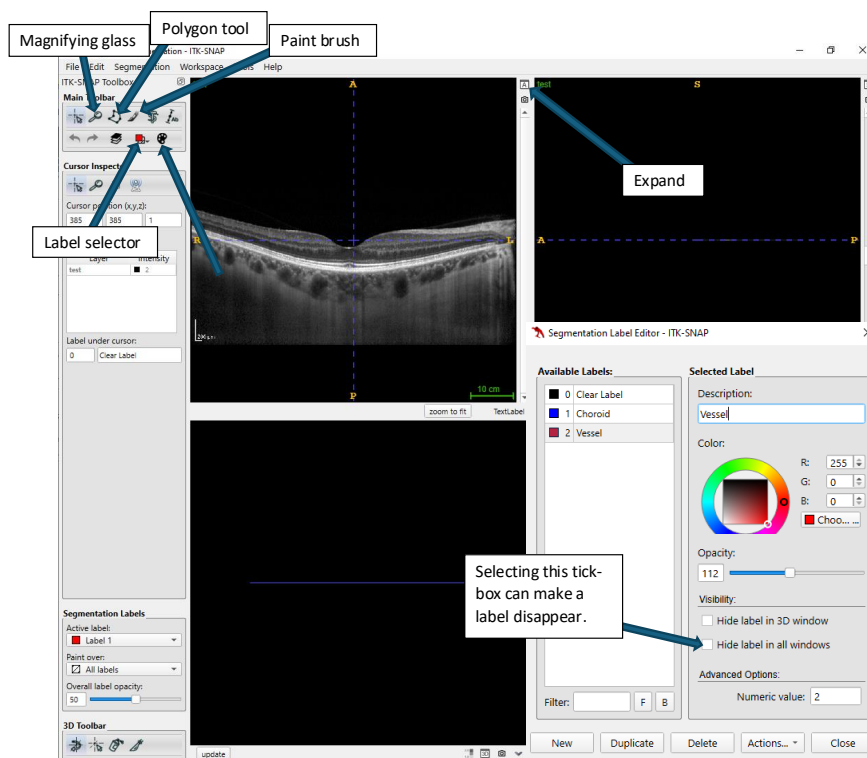
1. Go to File > Open Main Image. For image file name use the browse button to navigate to the image. Under File Format select 'Generic ITK image'. Select Next and Finish.
2. Click the 'A' in the top right corner of the top left window to expand the view to full size and set the window to full screen (figure [S7](#)).
3. Next, we need to set up some labels. Click on the label editor which allows you to define separate segmentation labels (figure [S7](#)).
4. We will have three separate labels (figure [S7](#)):

- (a) Label 1, called 'Choroid' represents segmenting the choroidal region, in blue.
  - (b) Label 2, called 'Vessel' represents labelling vessels within the choroidal space, in red. Here, this will have an opacity of 113 to aid segmentation.
  - (c) Label 0 is the 'Clear' label which is your 'rubber'.
5. On the Main Toolbar, the label selector allows you to toggle between different segmentation labels.
    - (a) Alternatively, you can do the same on the 'Segmentation Labels' toolbar, and there is also an option to change the colour opacity. There is a drop-down list of all the labels in the 'Active label' section.
  6. On the Main Toolbar, the magnifying glass can be used (up to 4x) to zoom into the portion of the image you would like to annotate. If you hover your mouse over this icon it provides instructions on how to do that
  7. Label 1 segmentation, defining the choroidal region segmentation, are done using the Polygon tool seen on the Main Toolbar.
  8. Label 2 segmentations, defining the choroidal vessel segmentation, are done using the Paintbrush tool seen on the Main Toolbar.
  9. Image enhancement and brightness tool features will not be used in this protocol.

### **A.5.3 Segmentation**

#### **A.5.3.1 Region segmentation**

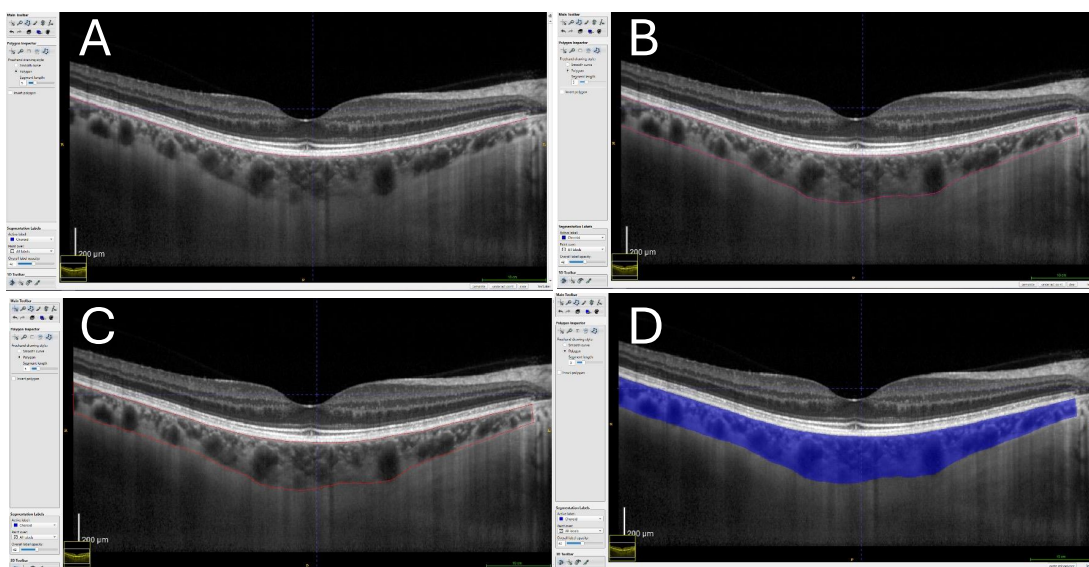
1. Select the Polygon tool. This allows you to draw lines by either selecting pixels to connect linearly, or by holding the left-mouse clicker to draw a smooth line.
2. Select the "Choroid" segmentation label and make sure you are painting over All Labels.



**Figure S7:** Screenshot of computer screen of ITK-Snap after loading an OCT B-scan as a generic ITK image, with various functions annotated. The label editor also shows the creation of labels for the choroid region (blue) and vessels (red).

3. Select the Polygon style button within the Freehand drawing style window. Choose a segment length of 5.
4. You may use the magnifying glass freely. Holding the right mouse-click will zoom in and out for you
5. There are buttons on the bottom-right of the window to allow you to undo the most recently selected pixel, or to clear all selected pixels entirely.
6. Core instructions:
  - (a) Start at the top-left most pixel of the RPE-Choroid layer boundary and move left-to-right until all of the visible layer boundary has been traced (figure S8(A)).
  - (b) On the same polygon segmentation, select the bottom-right most

## A.5. OCT B-scan manual segmentation protocol



**Figure S8:** (A) Manual selection of pixels from left–right along RPE-Choroid boundary using Polygon tool. (B) Within the same Polygon tool selection, select pixels from right–left along Choroid-Sclera boundary. (C) Clicking ‘Complete’ forms polygonal shape representing choroidal space, linking first and last selected pixels. (D) Clicking ‘Accept’ converts polygonal shape into segmentation object.

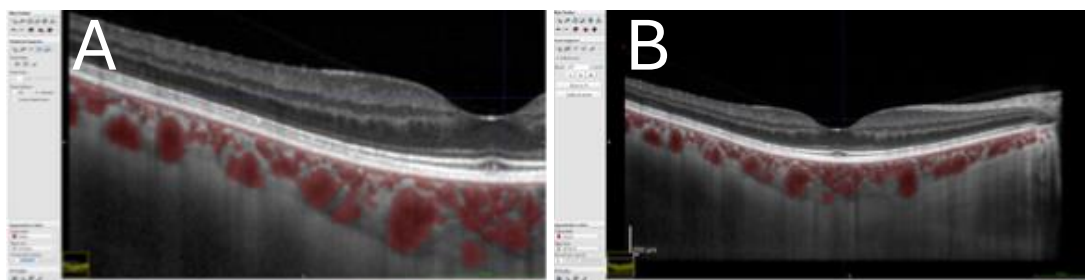
pixel of the Choroid-Sclera layer boundary and trace right-left ((figure S8(B)). You should see a polygonal shape begin to form, defining the whole choroidal space.

- (c) When you reach the left-most pixel of the Choroid-Sclera layer boundary, click Complete, this will join the first and last pixels you have selected (figure S8(C)).
- (d) Here, you can move (or remove) individually selected pixels around to better fit the polygonal shape to the choroidal region.
- (e) When you are finished, click Accept, this will turn your polygonal shape into a Blue segmented region (figure S8(D)).

7. If you wish, you can go straight to section A.5.4 to save your region segmentation, please save as ‘filename\_RegionAnnotations’.

### A.5.3.2 Vessel segmentation

1. In order for the region segmentation above to not block your view, click the Label Editor and select the box 'Hide label in all windows' for Choroid Label 1. This removes the blue, choroidal space segmentation, giving you a clearer view of the vessels.
2. Select the Paintbrush tool. This is a paint brush that allows you to 'colour in' the vessels. You can change the size and shape of the brush.
3. Change the 'Active label' to 'Vessel' or the 'Segmentation Labels' toolbar and make sure you select the 'All labels' layer in the 'Paint over' dropdown list.
4. You may alter the opacity to aid segmentation.
5. You may use up to a brush size between 2 and 12 (for small and large vessel labelling), selecting the circular shape and ticking the isotropic box (figure [S9](#)).
6. You may use the magnifying glass to zoom freely for small regions.
7. Some tips:
  - (a) You can select the 'Clear label' as a 'rubber', in case you label parts of the image mistakenly. This can also be done with the right mouse-click.
  - (b) Ctrl+Z will undo the most recent Paintbrush annotation.
  - (c) Press 2 or 4 to toggle between pan/zoom (2) and brush tool (4)
  - (d) Regularly reduce the 'Overall label opacity' on the 'Segmentation Labels' toolbar to sanity check any selected regions to make sure you are happy with your selections.
8. If you wish, you can go straight to (5) to save your vessel segmentation, please save as 'filename\_VesselAnnotations'.



**Figure S9:** (A) Zoomed in temporal side of choroid, as we progress through segmenting with half-opacity using the Paintbrush. (B) completed vessel segmentation. Note the choriocapillaries (approximately 3 pixels deep below the RPE-Choroid layer boundary on Heidelberg Engineering Spectralis OCT1 data) is always assumed to be vessel due to its dense capillary structure, hence why it is highlighted also.

### A.5.4 Saving segmentations

1. Once completed the region and vessel segmentations, you can view your final segmentation and make any amendments by using the right-click 'rubber' i.e. use the default Clear label (below).
  - (a) Unselect the 'Hide label in all windows' for Choroid Label 1 on the Label Editor to see the choroid region segmentation, alongside the choroid vessel segmentation.
  - (b) Note that you must select the 'Choroid' label to make any amendments to the choroid region segmentation, and likewise for amendments to the choroid vessel segmentation.
2. Once happy, to save the segmentation go to Segmentation > Save Segmentation Image and call the filename the same as the beginning of the image's filename and add '\_Annotations' to the end of it. For example, 'test\_Annotations'. Make sure the format is NiFTI and select Finish.
3. Save the labels to a `.txt` file by selecting Segmentation > Export Label Descriptions and save the name with the beginning of the image file name and add 'labels.txt' – e.g. 'test\_Labels'.
4. You can do step 2 & 3 as you go along. If you need to finish an annotation later follow step 2 and 3 and close SNAP. When you want to come back

to it, load the image as in section [A.5.2](#) and load the segmentation by going to Segmentation > Open Segmentation and select the file to open and go to Segmentation > Import Label Descriptions and select the `.txt` file to open and continue annotating where you left off.