



Automatic segmentation of outer retinal layers with Deep learning in ophtalmology

João Duarte do Nascimento Dias Afonso

Thesis to obtain the Master of Science Degree in

Biomedical Engineering

Supervisor(s): Prof. Daniel Simões Lopes
Diogo Gonçalo Reis Cabral, MD

Examination Committee

Chairperson: Prof. Maria Margarida Campos da Silveira

Supervisor: Prof. Daniel Simões Lopes

Member of the Committee: Prof. Carlos Jorge Andrade Mariz Santiago

December 2024

Declaration

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

The work presented in this thesis was performed at ITI/LARSyS of Instituto Superior Técnico, with the collaboration of Hospital Garcia da Horta, during the period of march-september, under the supervision of PhD Daniel Simões Lopes and MD Diogo Gonçalo Reis Cabral.

Acknowledgments

There are many people to whom I owe my deepest gratitude, and I would like to take this opportunity to express my appreciation.

First and foremost, I want to thank my family. Their tireless support has been a constant source of strength, not only during the good days but especially during the challenging ones. To my parents, my sister, my grandparents, and my aunt thank you from the bottom of my heart for always being there, for your encouragement, and for believing in me every step of the way.

I would like to express my deepest gratitude to my supervisors, Daniel Lopes and Diogo Cabral, for their invaluable guidance and expertise throughout this dissertation. Their insightful feedback not only helped me sharpen the focus of my research but also broadened the project's horizons, paving the way for new opportunities. I am deeply thankful for their dedication to the project.

A special mention goes to Bruno Pereira, and most especially to Pedro Camacho, for their significant contributions to this work. Their tireless efforts in performing the segmentation of the countless B-OCT scans were instrumental in bringing this project to life. Pedro's dedication and hard work were crucial to the success of this research, and for that, I am immensely thankful.

Finally, I want to give a huge thank you to my girlfriend and all my friends, who have been there for me through every step of this journey. Your constant support, patience, and ability to lift my spirit during tough times meant the world to me. Whether it was offering a word of encouragement or just being there when I needed a break, you all played a crucial role in helping me get to the finish line. I truly couldn't have done this without you, and I'll always be grateful for your friendship and love.

Resumo

O desenvolvimento de modelos de aprendizagem automática, e a sua aplicação nas áreas médicas como a oftalmologia, tem contribuído para a otimização da imagiologia, permitindo a identificação e monitorização de vários biomarcadores de forma rápida e reproduzível. Estes avanços na área da oftalmologia, podem facilitar a distinção entre doenças hereditárias semelhantes, detetar alterações estruturais de forma precoce, ou monitorizar a integridade e espessura de diferentes estruturas de forma rápida, aliviando o impacto financeiro sobre clínicas e hospitais.

Esta tese teve como objetivo primário desenvolver um modelo de segmentação automática das camadas externas da retina em imagens de tomografia de coerência ótica de domínio espectral (SD-OCT). Adicionalmente, pretendeu-se avaliar a reprodutibilidade e desempenho dos algoritmos desenvolvidos comparando os resultados obtidos com medições realizadas por dois especialistas. A variabilidade inter e intra-operador das medições humanas foi estudada e incluiu as seguintes estruturas: a segunda banda hiper-refletiva (Banda 2), a combinação da terceira e quarta bandas hiper-refletivas (Banda 3+4) e os segmentos externos dos fotorreceptores (POS).

Um total de 63 b-scans de 21 olhos foram incluídos no estudo. Os valores do coeficiente de correlação intraclassa (ICC) para a Banda 2, Banda 3+4 e POS foram 0.737, 0.830 e 0.878, respetivamente. Os coeficientes de correlação interclassa para estas estruturas foram 0.104, 0.810 e 0.752. Todos os modelos de rede neural convolucional profunda (dCNN) alcançaram valores de Dice superiores a 85%, com o modelo DRUNET a apresentar o melhor desempenho: 89.41% para a Banda 2, 93.25% para a Banda 3+4 e 91.42% para os POS.

Os valores de ICC demonstram uma forte consistência na segmentação manual, embora a correlação interclassa tenha revelado alguma variabilidade. O modelo DRUNET superou os outros modelos, mostrando potencial para reduzir a variabilidade interclassa e melhorar a precisão da segmentação destas estruturas.

Palavras-chave: Aprendizagem Profunda, Oftalmologia, OCT, Segmentação Semântica, Correlação Intraclassa, Distrofias Hereditárias

Abstract

With the development of machine learning models and their application in medical fields like ophthalmology, biomedical imaging has significantly evolved, enabling rapid and reproducible identification and monitoring of various biomarkers. These advancements in ophthalmology can aid in distinguishing similar hereditary diseases, detecting early structural changes, and quickly assessing the integrity and thickness of different retinal layers, thereby alleviating the financial burden on clinics and hospitals.

This thesis aimed to develop an automatic segmentation model for the outer retinal bands in spectral-domain optical coherence tomography (SD-OCT) images. Additionally, it sought to evaluate the reproducibility and performance of the developed algorithms by comparing their results with measurements performed by two retina specialists. Manual intra- and inter-observer variability was assessed for the following structures: the second hyperreflective band (Band 2), the combined third and fourth hyperreflective bands (Band 3+4), and the photoreceptor outer segments (POS).

A total of 63 B-scans from 21 eyes were included in the study. Intraclass correlation coefficient (ICC) values for Band 2, Band 3+4, and POS were 0.737, 0.830, and 0.878, respectively, while interclass correlation coefficients were 0.104, 0.810, and 0.752 for these regions. All deep convolutional neural network (dCNN) models achieved Dice scores exceeding 85%, with DRUNET delivering the best performance: 89.41% for Band 2, 93.25% for Band 3+4, and 91.42% for POS.

The high ICC values reflect strong consistency in manual segmentation, though interclass correlation revealed some variability. DRUNET exhibited the most robust performance among the tested models, demonstrating its potential to reduce interclass variability and improve segmentation accuracy for these minute structures.

Keywords: Deep Learning, Ophthalmology, OCT, Semantic Segmentation, Intraclass Correlation, Inherited Dystrophies

Contents

Acknowledgments	v
Resumo	vii
Abstract	ix
List of Tables	xiii
List of Figures	xv
1 Introduction	1
1.1 Motivation	1
1.2 Objectives	3
1.3 Thesis Outline	4
1.4 Achievements and awards	4
2 Background	7
2.1 The human eye and the importance of the retina	7
2.2 PRPH2 mutation	9
2.3 ABCA4 mutation	10
2.4 Optical Coherence Tomography (OCT)	12
3 State of the Art	17
3.1 ABCA4 and PRPH2 current therapies and detection techniques.	17
3.2 Deep Learning algorithms for segmentation.	19
3.3 Algorithms for segmentation on OCT images.	22
4 Methodological Framework	25
4.1 Data acquisition	25
4.1.1 Study Design and Settings	25
4.1.2 Patient Participation	26
4.1.3 Image characterization	27
4.1.4 Image annotation	28
4.2 Image preprocessing	30
4.3 Evaluation Metrics	31
4.4 Baseline Solution	33

4.5	Enhanced Solution	34
5	Results	37
5.1	General characterization	37
5.2	Annotation results	37
5.3	Pixel-analysis results	40
5.4	Deep Learning results	41
6	Discussion	45
6.1	Manual annotations	45
6.2	Pixel-analysis approach	47
6.3	Deep-learning approach	48
7	Conclusions	51
7.1	Findings	51
7.2	Future Work and possible improvements	53
	Bibliography	55
	Appendix	65

List of Tables

3.1	Summary of Segmentation Studies.	22
3.2	Summary of OCT Segmentation Studies.	24
4.1	Summary of main clinical data.	27
5.1	Intraclass correlation for both clinicians with Normal Images.	38
5.2	Interclass Correlation Coefficient between the two clinicians' first measurements.	39
5.3	Comparison between the manual annotations of clinician 1 and the pixel intensity results.	40
5.4	Different segmentation results for different bands in Normal OCT images.	43
1	Intraclass correlation coefficients and average differences for various regions and clinicians.	65

List of Figures

2.1	Illustration of the human eye and the respective cells in the retina.	8
2.2	Schematics of the development of the PRPH2 retinopathy.	9
2.3	Structure of the ABCA4 protein.	10
2.4	Process of NRPE recycling within the RPE and the outer segments.	11
2.5	Role of ABCA-4 in retinal diseases development.	12
2.6	Resolution and penetration depth of different imaging techniques.	13
2.7	Principle of a TD-OCT, the first implementation of an OCT scan.	13
2.8	Comparison between the SD-OCT and the SS-OCT.	14
2.9	Composition of an OCT Volume with the orientation regarding the eye.	15
2.10	Examples of different clinically used OCT devices.	16
2.11	Layers present in an OCT and their schematic representation.	16
3.1	Representation of a Convolution and a MaxPooling operation.	20
3.2	Representation of a Deconvolution operation.	20
3.3	Architecture of the first U-net developed by Ronneberger.	21
4.1	Schematic representation of the proposed pipeline of the project.	26
4.2	Example of an SD-OCT B-scan centered at fovea with (Band 3 + 4) human annotation.	28
4.3	Examples of different band structures.	29
4.4	B-scan OCT images of different eyes from the same patient.	29
4.5	Detection algorithm of the central fovea.	31
4.6	Figurative representation of the Dice Coefficient, on the left [85], and the corresponding formula on the right [86].	32
4.7	Representation of the fovea detection algorithm in a B-scan image, and the respective cropped images corresponding to the regions mentioned at the bottom of the image.	33
4.8	Representation of the segmentation masks used in the deep learning algorithm.	35
5.1	Scatter plots of ICC between the first measurements (x-axis) and the second measurements (y-axis) of the different regions.	38
5.2	Scatter plots of ICC between the first measurements (x-axis) registered by clinician 1 and the second measurements (y-axis) computed by the pixel-analysis algorithm.	41
5.3	Comparison of Clinicians' and Algorithms' Segmentations.	42

Chapter 1

Introduction

1.1 Motivation

Automatic segmentation techniques, powered by deep neural semantic models, have emerged as revolutionary tools in biomedical imaging, particularly within the ophthalmologic imaging domain. These techniques have shown remarkable capabilities in segmenting complex anatomical structures, enabling the extraction of valuable information for the diagnosis, disease progression monitoring, and treatment planning among different diseases [1–15].

Unlike conventional segmentation methods that rely on handcrafted features and algorithms, deep neural semantic models learn directly from vast amounts of labeled data, enabling the capture of complex patterns and relationships within images. This ability to learn from data and the significantly faster diagnosis allowed these models to achieve sophistication and accuracy levels beyond traditional approaches. These traits are particularly crucial in ophthalmologic imaging, where retinal diseases often involve subtle changes in the retina’s structure, which traditional segmentation methods struggle to delineate accurately.

Optical Coherence Tomography (OCT) is a non-invasive imaging technique that generates volumetric scans of eye structures, like the central macular area. This type of scan allows the capturing of cross-sectional images with varying resolutions, according to the device at hand, and also visualization in real-time and at video rate, enabling clinicians to rapidly receive an interpretable exam to assist the diagnosis of the patient and guide future interventions instantaneously.

Before the clinical integration of OCT, the diagnosis of retinal diseases relied primarily on the visual inspection of retinal fundus changes or alterations in retinal vascular angiography. The development of spectral-domain OCT (SD-OCT) has revolutionized this process by enabling direct visualization of the different retinal layers and their thicknesses in both the macula and optic nerve [16].

The clinical implementation of imaging devices with axial resolutions as fine as $3\mu m$ has enhanced the precision of clinical diagnostics and significantly increased interest among researchers in studying these minute structures of the human eye with such great detail.

ABCA4 and PRPH2-associated retinopathies are the two most common inherited retinal dystro-

phies (IRD), characterized by different disease phenotypes and rates of central vision loss [17]. While for example the most common disease caused by the mutation in the ABCA4 gene, autosomal recessive Stargardt disease (arSTGD), has a prevalence of around 1 out of 8,000 to 10,000 in the United States [18], PRPH2 is one of the most frequent non-syndromic IRD (NS-IRD) disease-causing genes, accounting for 3.4% of individuals with the disease in Japan, 3% in the USA and Canada, and 4.6% of families in the UK [19].

Recent advances in molecular diagnosis and therapeutic options, complemented by the increased recognition of new multimodal imaging biomarkers, helped predict genotype and disease progression, especially through genetic testing. Although these dystrophies have different pathophysiologies, phenotypic similitude is frequent, leading clinicians to a difficult diagnosis.

Determining the precise genetic mutation underlying certain macular dystrophy is crucial for a comprehensive understanding of the mechanism underlying these potentially blinding diseases, as this diagnosis might imply further tailored treatment strategies and personalized long-term prognosis to deal with the specific disease.

Current diagnostic approaches for inherited retinal diseases rely on blood sample testing to identify the most prevalent mutations among 307 genes. This method is extremely costly and less available in clinical practice than standard-of-care retinal imaging modalities such as OCT and color fundus photography/autofluorescence.

Recent research has unveiled distinct thickness profiles of the four existing outer retinal hyperreflective bands in OCT B-scans of PRPH2 and ABCA4 retinopathies, enabling the accurate differentiation of these overlapping phenotypes [20]. Specifically, the thickness of band 2 (ellipsoid zone of the photoreceptors), was found to be significantly higher in the PRPH2 group compared to ABCA4 patients and healthy control groups, whereas, for band 4 (retinal pigment epithelium - RPE), the opposite was observed. The thickness of the ABCA4 was significantly higher compared to the PRPH2 group and the healthy controls. This study manually measured the specific bands in 4 retinal loci with 2 independent examiners. The authors concluded that this differential thickness pattern holds promise for improving the clinical diagnosis of these retinal dystrophies, with the mentioned study reporting a 99% accuracy in differentiating these two phenotypes, based on the band 2/ band 4 ratio. Although this project retrieved significant results, the time consumed for each task and the limitation of comparing only specific locations of these bands asks for a more general and automatic implementation.

This project aims to compare different approaches for the measurements of the hyperreflective band 2, bands 3 and 4 (photoreceptors with RPE plus RPE/Bruch's membrane complex), and the photoreceptor outer segment (POS) region. The availability of a machine learning-based tool that will empower clinicians to accurately assess the risk of these two specific inherited retinal diseases in individual patients will not only reduce unnecessary tests for low-risk patients, saving time and money for both patients and clinicians, but also prioritize genetic testing for high-risk patients, enabling early disease detection, prevention, and timely treatment initiation. A high level of consistency is essential for these small measurements, and implementing automatic models shows strong potential as a reliable alternative to manual segmentation, addressing the variability problem.

1.2 Objectives

After outlining the motivation and broad goals of the project, we will focus on the specific objectives and clearly define the expected outcomes.

The project's initial phase involved a comprehensive review and summarization of pre-processing techniques, network architectures, and the associated performance metrics. Based on this analysis, and after patient recruitment and characterization, a manual examination of the patients' bands was conducted by two different clinicians. These measurements served as ground truth images for the deep learning model and terms of comparison for the thicknesses of the respective regions.

The next stage concerned the validation of these manual annotations. For that, we decided to obtain and compare the intraclass and interclass correlation coefficient values in the 3 different studied regions. The purpose of this part of the study was to evaluate the correlation and accuracy between the two clinicians assigned to this task, both in comparison to each other and to their measurements, since the regions in question are minute, and each clinician may have their interpretation regarding the frontier of each region.

After acquiring the manual delimitations of the regions, two different automatic approaches were tested. The first, denominated "conventional approach", used pixel intensity analysis. This approach compared the intensity of the pixels in each B-scan, measuring the thicknesses in each column for the bands of interest and the POS region using the outer retinal bands' profile from OCT images of the central macular area, allowing an average thickness value for each band.

This technique was implemented to eliminate the need to develop a more advanced deep learning model and, therefore, reduce the task's complexity. By focusing on a simpler, yet effective, method, we aimed to streamline the process, making it more accessible and less computationally intensive, while still using it as a foundational step.

The second approach utilized the principles of deep convolutional neural networks(dCNNs). Ground truth data, consisting of manually segmented regions by the clinician who achieved the highest intraclass correlation coefficient(ICC), was used to train and evaluate various architectures based on multiple performance metrics. This approach aimed to comprehensively segment the region of interest, providing thickness measurements for every part of the structure and enabling detailed visualization of its contours.

The conventional approach was evaluated against manually segmented results from images acquired uniquely from healthy patients. Similarly, the performance of dCNN-based segmentation models was compared, with a specialist reviewing the segmentation masks to validate their integrity and accuracy.

The primary objective was to develop a concise and reliable algorithm that can ultimately be implemented in clinical settings with clinically diagnosed patients. The secondary objective was to compare the manual measurements among the two clinicians and understand the average error in the different areas distanced from the central fovea, defining the best and worst regions to perform manual segmentations in each structure measured.

1.3 Thesis Outline

This master thesis is organized as follows. The introduction outlines the motivation behind this research, emphasizing the importance of automatic segmentation in ophthalmology and the challenges associated with diagnosing retinal diseases, particularly those linked to ABCA4 and PRPH2 mutations. It also defines the objectives and expected deliverables of the study, followed by a brief overview of the thesis structure and information on the project achievements and awards.

The background section provides a comprehensive overview of the human eye, focusing on the retina's anatomy and its significance in vision. It discusses the clinical implications of PRPH2 and ABCA4 mutations, highlighting their roles in retinal diseases. Additionally, it introduces Optical Coherence Tomography (OCT), explaining its fundamentals and importance in retinal imaging, and mentioning the devices used in the project, and the different ones used in clinical practice.

After the background chapter, the state-of-the-art reviews the existing therapies and detection techniques for ABCA4 and PRPH2-associated retinopathies. It then delves into recent advancements in automatic segmentation approaches, specifying pixel-related measurements and deep learning algorithms for OCT segmentation, comparing different techniques and their efficacy.

The methodological framework outlines the approach taken in this research. It begins by covering how the data was acquired, mentioning study design, patient participation, image characterization, and annotation processes. Image preprocessing techniques are then detailed, followed by a description of the initial baseline solution, followed by the enhanced solutions developed to improve the baseline methodologies.

The results section presents the findings of the research. It begins with a general characterization of the dataset, followed by the outcomes of the image annotation process. Detailed results from pixel-level analysis and the performance metrics of the deep learning models used for image segmentation are provided.

The thesis concludes with a summary of the key achievements and findings. It discusses the contributions made by this research to the field of retinal disease diagnosis and automatic OCT segmentation, finally outlining potential areas for future work, and suggesting possible improvements and directions for further research.

1.4 Achievements and awards

The current project has received significant support and recognition from various organizations, both nationally and internationally. Below is an outline of the key achievements and collaborations that have contributed to the progress of the research:

- The project received the "Clinical Research Fellowship" grant from the Sociedade Portuguesa de Oftalmologia (SPO). This grant was announced at the 66th Congress of the SPO in Vilamoura and aims to provide financial and structural support for the research, including funding for publication, genetic testing for model validation, the use of Dye Contrast exams, and procurement of necessary

hardware for handling large datasets. Collaborations include Centro Hospitalar e Universitário de Coimbra, São João do Porto, and University of New York.

- UNILABS provided partial funding for genetic testing related to the two inherited retinal dystrophies, contributing to future model validation.
- The project was awarded a Research Grant by the Grupo de Estudos da Retina (GER), with support from AbbVie, ensuring the continuation of research throughout 2025. (<https://www.ger-portugal.com/?p=12240>)
- The research is currently under review for publication in the Association for Research in Vision and Ophthalmology (ARVO) journal, specifically in Translational Vision Science & Technology.

Chapter 2

Background

This chapter concisely provides the theoretical concepts required to fully understand the project, specifically their anatomical fundamentals and principles of image acquisition using the OCT device. Therefore, it will focus on an overview of the anatomy of the eye, specifically focusing on the retina, its constitution (cells and their respective function), and the retinal layers relevant to the study. This exploration extends to understanding the mutations and pathologies that partially instigated this investigation, including their origins and effects on visual function. Additionally, this chapter highlights the central role of OCT as the imaging technique employed in this project, the different types of devices, their usage, and differences in resolution. This foundational knowledge sets the stage for the subsequent chapters, where there is a deeper dive into the analysis and applicability in the machine-learning world, correlating this medical knowledge with the architectures proposed.

2.1 The human eye and the importance of the retina

The eye is the primary organ responsible for vision within the visual cycle. The normal human eye has approximately 69 to 85 mm and 22 to 27 mm in circumference and anteroposterior diameter, respectively, and weighs about 7.5 grams in the adult age [21]. It is separated into three different primary layers [21], the outermost supporting layer of the eye, consisting of the clear cornea, the opaque sclera, and the limbus, the middle uveal layer, constituting the central vascular layer of the globe, formed by the iris, ciliary body, and choroid. Finally the retina, the more interior one, and the layer of interest in this study. It lies on the intermediate layer and receives nourishment from the vessels within the choroid.

Anatomically, the eye can be divided into two areas: Posterior and Anterior. The Anterior zone refers to the frontal area of the eyeball and includes the cornea, the iris, and the lens. The Posterior zone comprises the back part of the eyeball and contains the sclera, the choroid, and the retina.

The retina, responsible for capturing light, converts it into electrical signals transmitted to the brain through the optic nerve. Although the precise composition of each retinal layer remains debated, these bands provide a valuable characterization of the outer retinal region [16]. According to the same author, it comprises 16 distinct layers and houses seven key cell types that play crucial roles in the visual

process. These cell types are briefly summarized below, highlighting their significance in the visual field's formation [22]:

- Photoreceptors: the retina's outer nuclear layer houses two types of photoreceptors: rods and cones. Cones are predominantly responsible for high-acuity color vision under bright lighting conditions. They contain photopsins, specialized proteins that enable the perception of color. Rods, on the other hand, are primarily activated in low-light conditions, enabling night vision. They contain rhodopsin, a pigment that facilitates visual acuity in dark environments. While cones are densely concentrated in the fovea, the central region of the retina, rods are not found in that region;
- Bipolar cells: make up the inner nuclear layer and are responsible for transmitting an impulse to the retinal ganglion cells;
- Amacrine cells: modulate the excitation between bipolar cells and the retinal ganglion;
- Retinal ganglion cells: final receivers of the initial stimulus. They form the ganglion cell layer and transmit this stimulus to their axons in the retinal nerve fiber layer;
- Horizontal cells: modulate the communication between photoreceptors and bipolar cells;
- Muller cells: glial cells essential for proper retinal function, serve to recycle neurotransmitters, prevent glutamate toxicity, and regulate nutrient homeostasis in the retina.

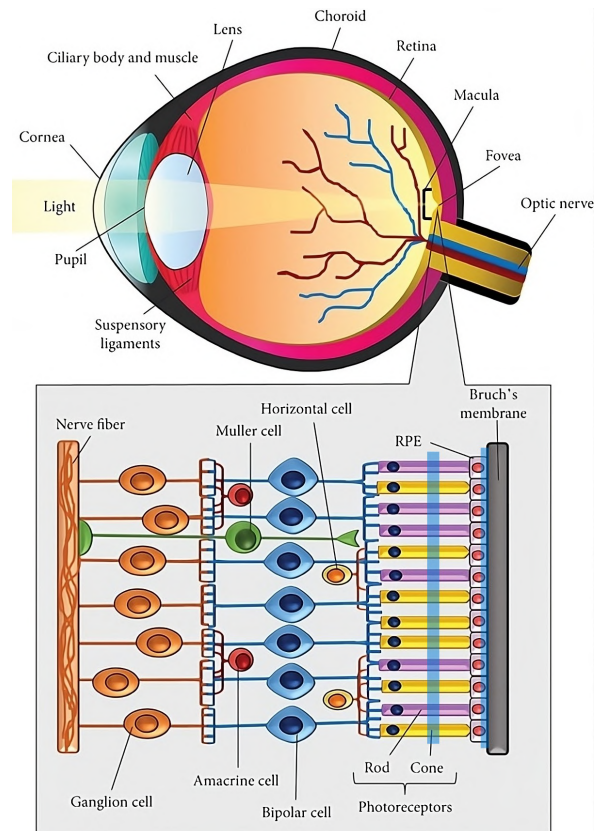


Figure 2.1: Illustration of the human eye and the respective cells in the retina. The areas in blue represent the location of the hyperreflective layers 2 and 4 of the retina, regions of interest throughout this project, as explained later. From reference [23].

2.2 PRPH2 mutation

Peripherin-2, also known as *rds*, is a gene that encodes a photoreceptor-specific tetraspanin protein, PRPH2. Tetraspanins define a family of highly conserved membrane proteins entangled in several functions. These functions include membrane organization and compartmentalization, cell signaling, adhesion, and migration. Structurally, tetraspanins consist of a short cytoplasmic N-terminus, four trans-membrane domains (TM), one small and one large extracellular loop (D1 and D2, respectively, in Figure 2.2), as well as a short cytoplasmic C-terminus [24]. These proteins can be found in all multicellular eukaryotes and can also be referred to as the transmembrane 4 superfamily (TM4SF) proteins [25].

PRPH2 is located within the POS, a specialized light-sensitive ciliary organelle that forms the outermost portion of rods and cones [26]. It is specifically concentrated in the rim region of the rod and cone discs/lamellae. In photoreceptors, PRPH2 exists as either a homo- or a hetero-tetramer alongside the rod OS membrane protein 1 (ROM1), another tetraspanin located at the disc rims. These tetramers are responsible for the morphogenesis and maintenance of the OS. While the precise role of ROM1 remains under investigation, it has been established that it is not essential for disc morphogenesis but plays a crucial role in regulating the size and structure of OS discs.

The main function of PRPH2 in the morphogenesis of rod and cone OS is to promote membrane curvature, flattening, and fusion, allowing their proper development and thus facilitating the visual cycle [24]. PRPH2 distributes to the rims of OS disc membranes as they undergo the process of disc membrane enclosure (Figure 2.2-A).

Mutations in PRPH2 prevent photoreceptor disc formation and induce a substantial release of extracellular vesicles, or ectosomes, from the photoreceptor cilium. These mutations are dominantly inherited and lead to a wide variety of disease phenotypes. 90 different disease-causing mutations in PRPH2 have been identified, which cause several forms of retinitis pigmentosa and macular degeneration [27].

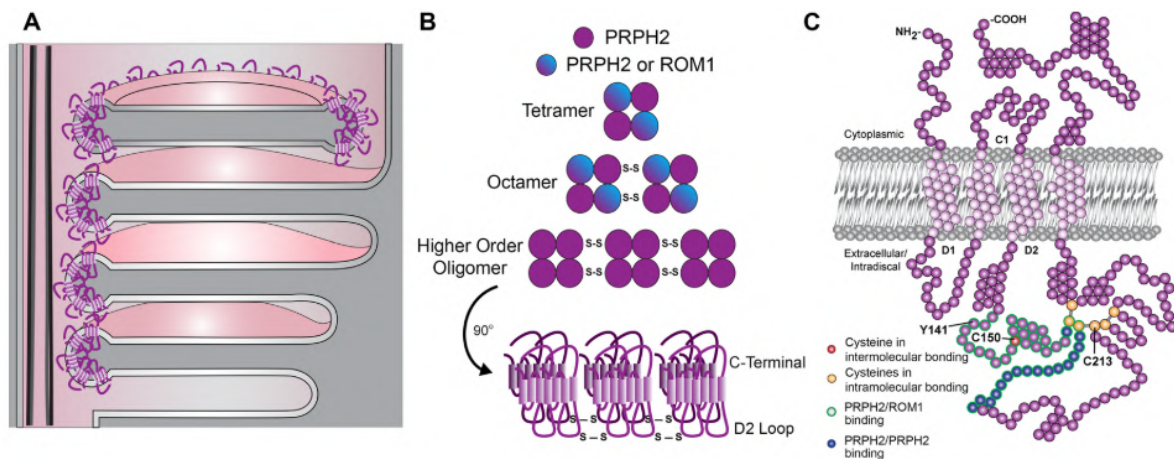


Figure 2.2: Schematics of the development of the PRPH2 retinopathy. (A) Development of photoreceptor discs. The image contains both open discs, and enclosed, mature discs separated from the OS plasma membrane. PRPH2 is located only at the disc rims adjacent to the connecting cilium in open discs. In enclosed discs, PRPH2 is located throughout the entire circumference of the disc. (B) Formation of PRPH2/ROM1 homo and hetero-oligomers. (C) Molecular structure of the PRPH2. From reference [28].

In the United Kingdom, genetic variants in the PRPH2 gene represent the fourth leading cause of IRD, accounting for 5.2% of reported cases. The Human Gene Mutation Database (HGMD) currently curates a collection of 352 disease-causing variants in PRPH2, with inheritance patterns encompassing autosomal dominant, and less frequently, autosomal recessive and digenic inheritance alongside ROM1 [29].

2.3 ABCA4 mutation

The ABCA4 protein is an ATP-binding cassette (ABC) transporter in the photoreceptor OS that functions in the visual cycle [30], playing an essential role in retinoid recycling. It is an N-retinylidene-phosphatidylethanolamine and phosphatidylethanolamine importer [31], the only known importer among mammalian ABC transporters [30].

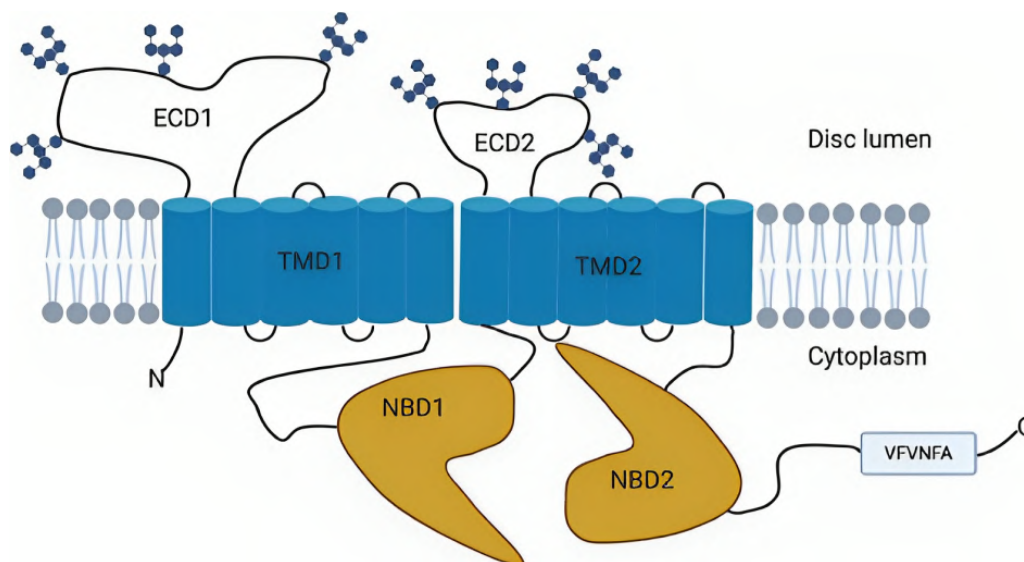


Figure 2.3: Structure of the ABCA4 protein, with its respective two transmembrane domains (TMD), nucleotide-binding domains (NBD), and exocytosmic domains (ECD). From reference [32].

Stargardt disease (STGD1), or ABCA4-retinopathy, is the most common inherited retina degeneration [33], with an incidence of 1/8000 to 1/10000 individuals, and can be normally detected in late childhood or early adulthood [32]. It is inherited in an autosomal recessive manner and is caused by variants in the ABCA4 gene. This mutation results in lipofuscin (bisretinoid fluorophores, Figure 2.5-C) accumulation in the RPE. This material originates, for the most part, from ingestion of shed POS membrane [34].

The RPE is a specialized epithelium that forms a crucial interface between the neural retina, and the capillary lamina of the choroid, serving as the outer blood-retinal barrier (BRB). It plays a fundamental role in maintaining retinal homeostasis by fulfilling essential functions such as nutrient and ion transport, light absorption to protect photoreceptors from photooxidative damage, phagocytosis of shed photoreceptor membranes, and secretion of essential factors that support retinal structural stability [35].

In a normal functioning ABCA4 transporter, N-retinylidene-PE (NRPE), formed by the release of

all-trans-retinal by photoactivated rhodopsin (Figure 2.4), is transported across the lipid bilayer to the cytoplasmic face of the disc membrane to be hydrolyzed.

When a mutation occurs in this transporter, there is a subsequent accumulation of NRPE inside the disc photoreceptor that will further react to form toxic di-retinal fluorophores (bisretinoids) [36]. To rid outer segments of these randomly produced light-absorbing molecules, the photoreceptor cell undergoes a process of outer segment membrane renewal, during which bisretinoids are transferred via phagocytosis to the RPE. This process results in an accumulation of these toxic substances in the RPE, leading to a loss of function of these cells.

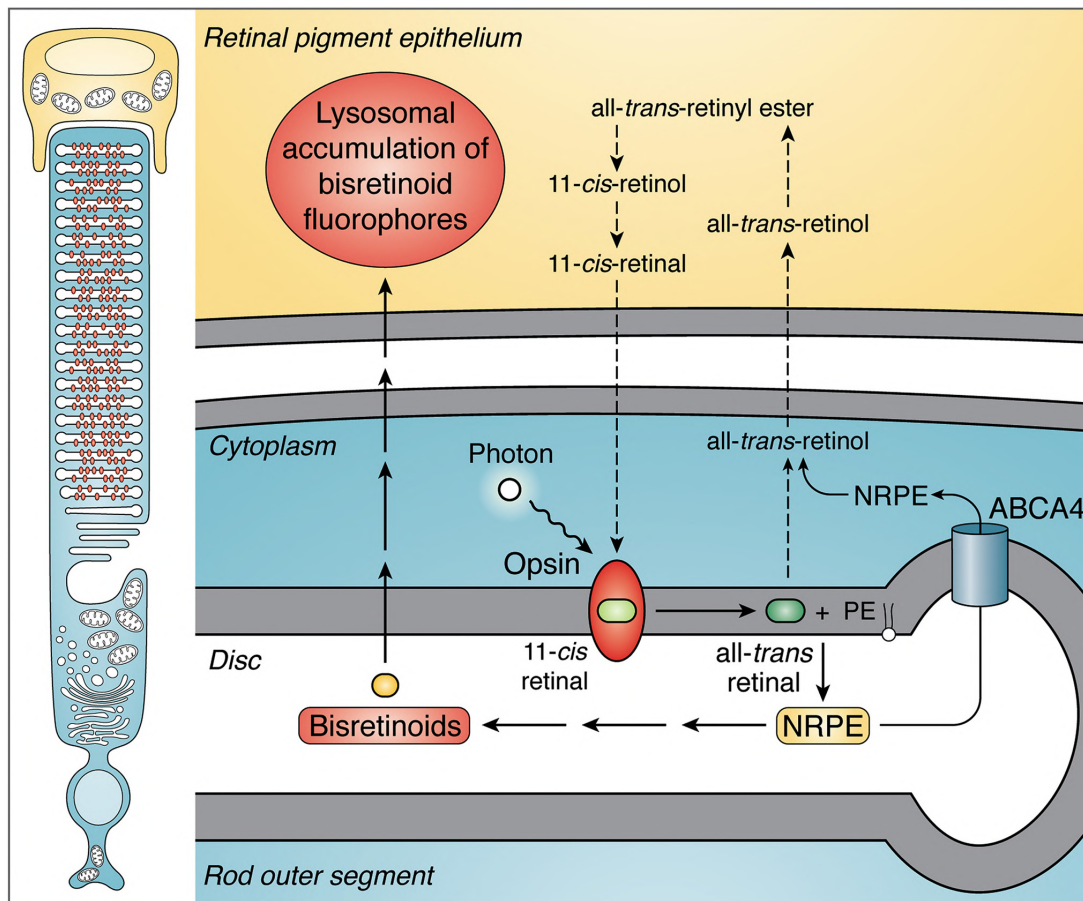


Figure 2.4: Process of NRPE recycling within the RPE and the outer segments. In the presence of the ABCA4 transporter, a degradation into all-trans-retinol will occur (green arrow). In contrast, in the case of a mutation in that transporter, bisretinoids will be formed (red arrow), causing phagocytosis to the RPE and a consequent loss of function. From reference [36].

Stargardt patients may experience a range of visual symptoms, from mild to severe, depending on the transporter's loss of function. Early symptoms often include bilateral central visual loss, photophobia, color vision abnormalities, central scotomas, and slow dark adaptation. As the disease progresses, patients may develop cone-rod dystrophy, characterized by a significant decline in both cone and rod function, and ultimately, retinitis pigmentosa, where the presence of pisciform flecks is notorious, which in turn can result in a complete loss of vision (Figure 2.5, panel A). The patient's position within this disease spectrum is determined by the total amount of residual ABCA4 function, with individuals with lower levels of function experiencing more severe and rapid disease progression.

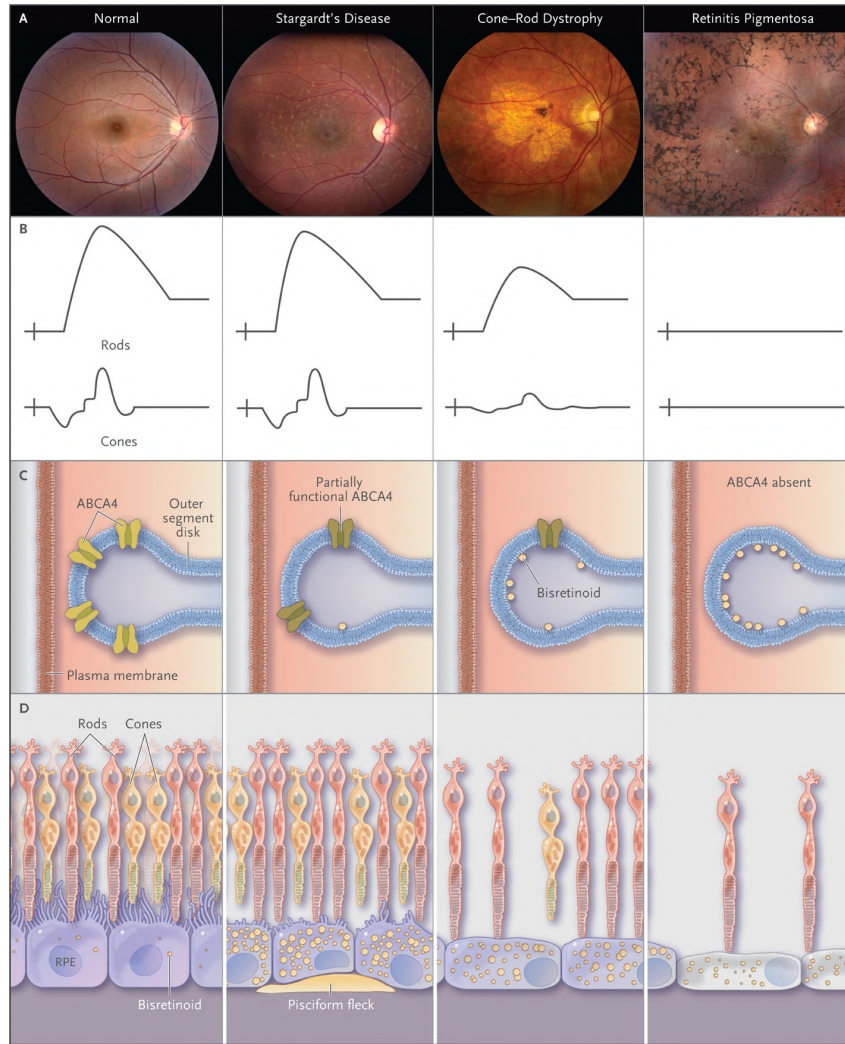


Figure 2.5: Panel A presents retinal photographs with the decrease of the ABCA-4 functioning, ranging from a normal retina to a retinitis pigmentosa patient. Panel B shows the reduction in the response of the photoreceptors in a full-field electroretinogram exam [37]. Panel C illustrates the accumulation of bisretinoid (yellow dots) caused by the reduction of ABCA4 function, on the inner leaflet of the photoreceptor outer-segment disc membranes. Panel D shows how reduced ABCA4 activity can cause photoreceptor apoptosis (especially in cones) and RPE thinning in patients with cone-rod dystrophy and retinitis pigmentosa. Role of ABCA-4 in retinal diseases development. From reference [17].

2.4 Optical Coherence Tomography (OCT)

Optical coherence tomography was employed to evaluate the structure and integrity of the retinal bands, aiming to detect the presence, type, and potential progression of retinal diseases throughout this project. OCT is a non-invasive imaging technique that generates cross-sectional images of tissues, making it particularly valuable for organs like the eye, where biopsy is not feasible, providing *in vivo* imaging without affecting the tissue itself [38]. In simple terms, this medical imaging method allows 3D reconstructions of semitransparent tissue by measuring the echo time delay and the magnitude of back-reflected light.

Typically, OCT systems have axial resolutions, or penetration depth of 20 to 5 μm , a much more

detailed resolution when compared to other medical imaging methods like ultrasound or magnetic resonance imaging (MRI). It combines an axial resolution that can reach confocal microscopy resolutions with a lateral resolution comparable to a confocal scanning laser ophthalmoscope.

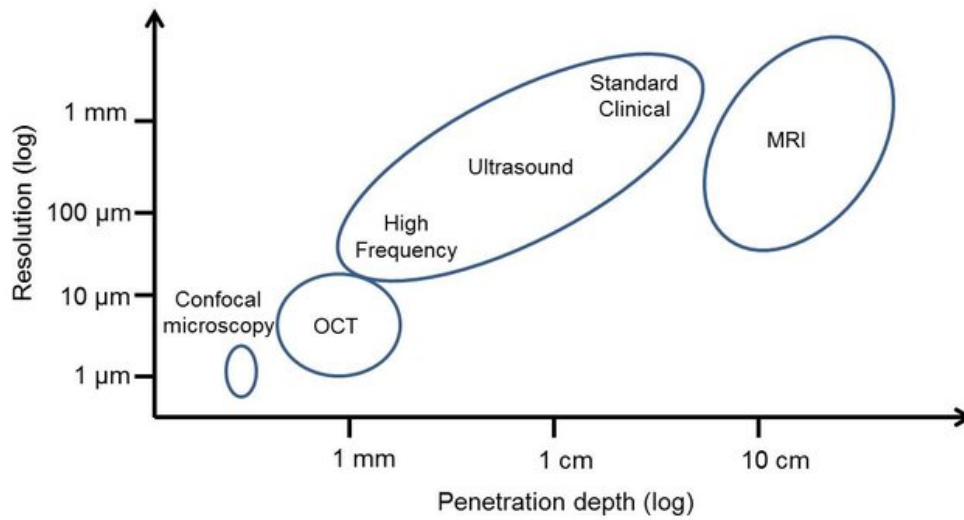


Figure 2.6: Resolution and penetration depth of different imaging techniques. From reference [39].

A simple time-domain OCT (TD-OCT), the first generation of OCT scans, uses near-infrared light waves directed through an interferometer (splits the light coming from the source and combines the light on the way back) to echo off the tissue structure, while also running through a reference path, as the source signal. It is possible then to measure the light intensity using a photodiode. However, to measure one depth profile of the sample (axial scan or A-scan) the reference arm needs to be moved (vector u in Figure 2.7), since the only way to measure the precise depth is when the optical paths of both the backscattered light and the reference mirror are the same, because of the high speed of the light, making it impossible to make a direct measurement of “optical echo”. Since TD-OCT relies on the synchronization of the light delays from the reference mirror and the target object, it necessitates using a moving mirror, increasing the acquisition process time.

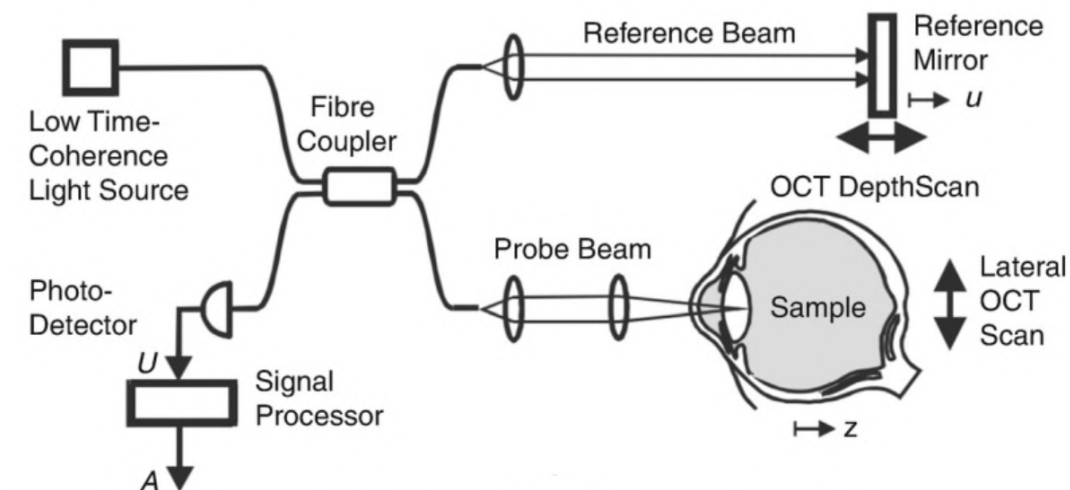


Figure 2.7: Principle of a TD-OCT, the first implementation of an OCT scan. From reference [38].

A swept-source OCT (SS-OCT), the latest generation of OCT scans, uses the Fourier domain to generate A-scans through spectral information. Thanks to its sweeping light source, whose wavelength varies with time, generated by high-speed tunable lasers, it is possible to access the interference signals from both the sample and the source while also recurring to the Fourier domain of the interference spectrum to obtain the reflectance at different depths. In other words, the path differences between sample and reference reflections cause variations in the detected beat frequency by introducing a phase difference in the two light signals, and the Fourier transform of the beat frequency spectrum gives a depth profile of the sample's reflectance. Therefore, a wider optical frequency sweep gives higher depth resolution, and a faster sweep permits rapid data acquisition to facilitate optical coherence tomography imaging [40]. This technique improved scanning rate time, providing a wide scanning range, narrow line width, high output power, and deeper penetration, which is positive when working with the thickness of retina layers.

The Spectral-Domain OCT (SD-OCT) uses Fourier domain detection technology, similar to Swept-Source OCT (SS-OCT), but with key differences in light sources: SD-OCT uses a broadband near-infrared superluminescent diode centered around 840 nm, while SS-OCT employs a swept laser with a central wavelength of approximately 1050 nm.

SS-OCT imaging outperforms SD-OCT by offering faster scan speeds, enabling more detailed scan patterns and larger areas in equivalent time frames. Additional benefits of SS-OCT include the longer wavelength and reduced sensitivity roll-off (decay in sensitivity as a function of depth), enhancing light penetration through the RPE, and improving detection of deeper retinal signals. The longer wavelength is also safer, permitting higher laser power, which improves the visibility of weaker signals from deeper retinal layers. However, the accessibility of SD-OCT in clinical settings, due to its exceptional balance between cost-effectiveness and imaging quality, made it the device of choice for this project.

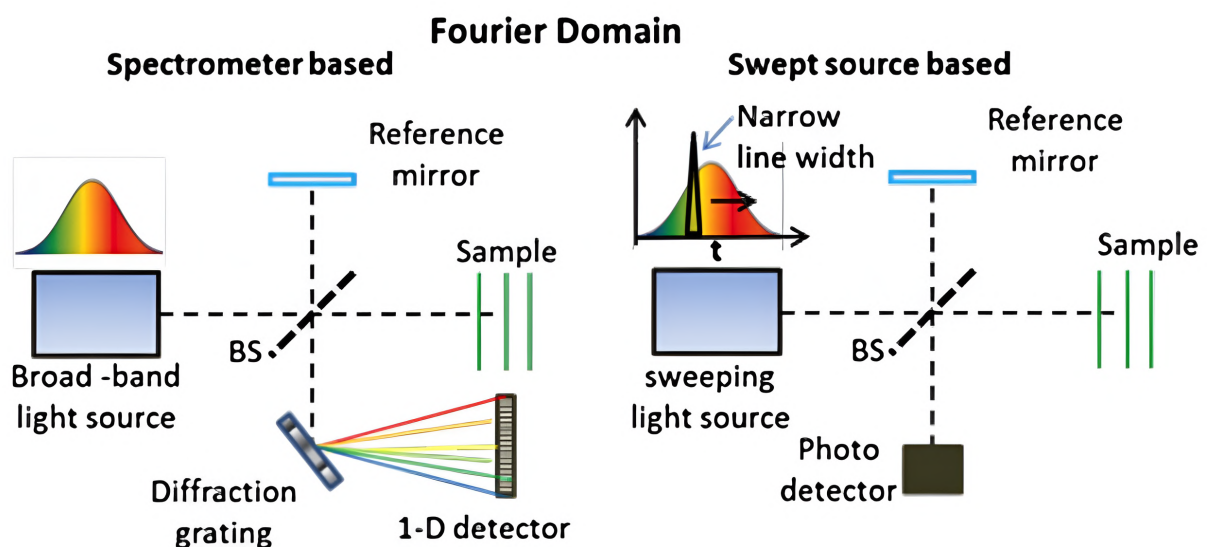


Figure 2.8: Comparison between the SD-OCT and the SS-OCT. On the left, the principle of an SD-OCT, while on the right, the principle of an SS-OCT, with different light sources. BS = Beamsplitter. From reference [41].

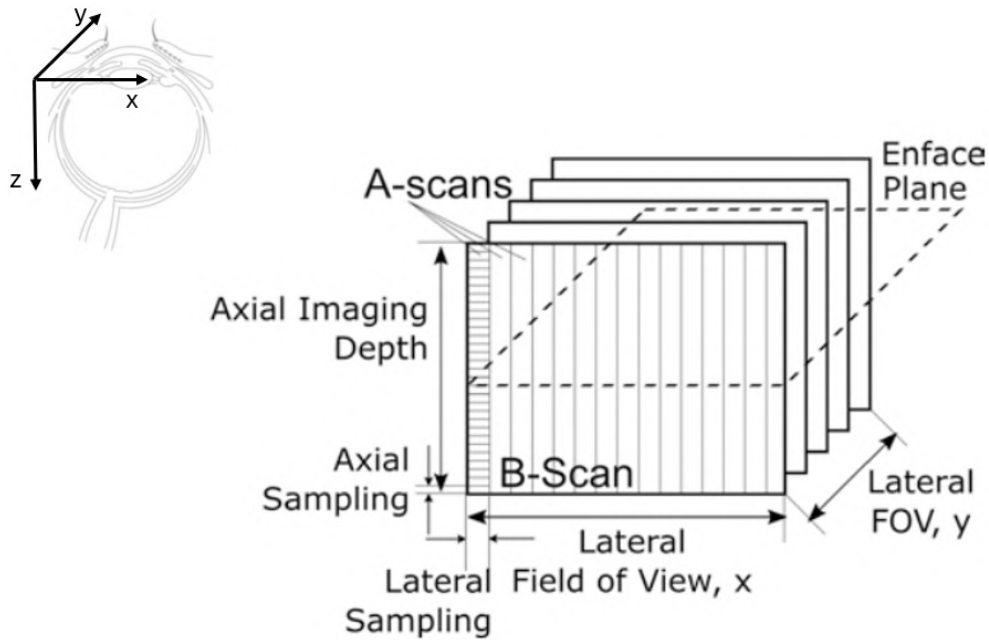


Figure 2.9: Composition of an OCT Volume with the orientation regarding the eye. From reference [38].

For clinical use, there are numerous OCT devices. For example, the ZEISS PLEX Elite 9000 [42] was the first Dual-Speed SS-OCT device, allowing it to produce up to 200,000 A-scans/sec, with resolutions up to $3.95 \mu\text{m}$. This type of resolution can be beneficial for detecting small layers, such as the ones of the retina.

The SPECTRALIS device was used to acquire the images in this project. It provides high-resolution, two-dimensional SD-OCT images of the retina and anterior segment. It has image acquisition with a scanning speed of 85 kHz, a depth resolution of $3.9 \mu\text{m}$, and a lateral resolution of $5.9 \mu\text{m}$.

For investigation purposes, other devices with increased resolutions can be utilized. The HighRes OCT instrument has an A-scan rate of 85 kHz and combines three superluminescent diodes to provide a center wavelength of 853 nm and a bandwidth of 137 nm to increase the optical axial resolution to $3 \mu\text{m}$ in tissue, although its lateral optical resolution remains unchanged relative to the SPECTRALIS platform. The HighRes OCT uses the same auxiliary technologies as the SPECTRALIS, including TruTrack Active Eye Tracking, and Noise Reduction, which synergize to enhance OCT image quality and scan positioning accuracy.

As seen in Figure 2.11, in a HighRes OCT scan of the retina, 16 layers can be differentiated. We can also observe bright hyperreflective layers and darker hyporefective layers, depending on whether they reflect or absorb light from the source.

The analysis of the 4 outer hyperreflective bands at the fovea (represented as 1st to 4th in 2.11) could inform us about the health and integrity of the cone compartments and their relationship with the RPE. Changes in the hyperreflective or hyporefective line thickness may have important clinical implications because the visual function correlates closely with band integrity after surgical or medical treatment. Several diseases, such as Age-Related Macular Degeneration, Diabetic Macular Edema, or others, show significant alterations in these exams and can be easily diagnosed by a retina specialist.

Several have arisen over the number of layers and specific cell types that the bands observed in the SD-OCT represent, especially over the 4 outer retinal bands. The nomenclature used in this project is the one proposed by Cuenca *et al.* [16]. The first and innermost one is the external limiting membrane (ELM), the second hyperreflective band corresponds to the cone ellipsoid zone (ISeZ), or the Inner Segment/Outer Segment border (IS/OS), the third band corresponds to the OS tips phagocytosed by the pigment epithelium (PhaZ), and the fourth band corresponds to the mitochondria in the basal portion of the RPE (RPEmitZ), as shown in the Figure 2.11.



Figure 2.10: Examples of different clinically used OCT devices. On the left, the ZEISS PLEX Elite 9000 and on the right the SPECTRALIS device. From reference [42, 43].

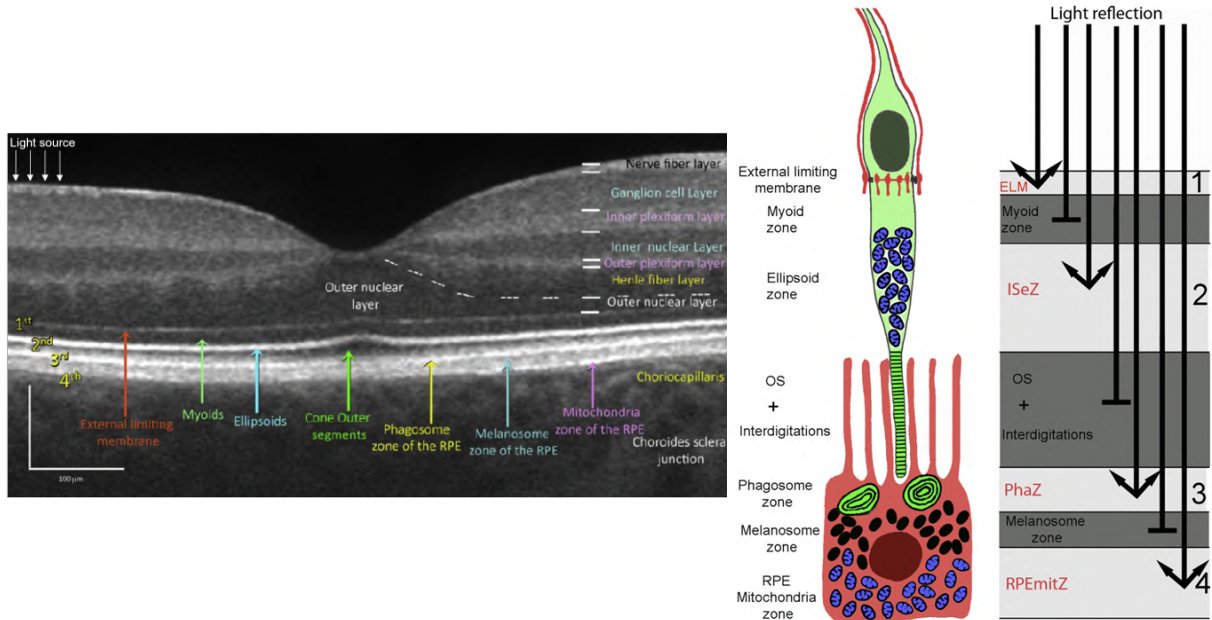


Figure 2.11: On the left is an OCT representation of the 16 layers of the retina. The layers, represented as 1st (ELM) to 4th (RPEmitZ), will be the object of study during this project, as the variations in thickness will impact the decision on the classification task, as seen in future chapters. On the right, is the interpretation of the hyper and hyporeflective layers, in this case in the sub-retinal space. From reference [16].

Chapter 3

State of the Art

This chapter explores the current diagnostic modalities and therapeutic approaches employed to accurately identify and manage retinopathologies associated with ABCA4 and PRPH-2 mutations. It begins by providing an overview of the prevalent clinical practices for diagnosing these retinal disorders, highlighting their strengths and limitations. From there, a comprehensive review of existing segmentation techniques related to deep learning models and pixel-wise approaches for OCT imaging is presented, meticulously examining the advantages and drawbacks. The chapter concludes by identifying the current gaps in the literature, which were addressed by the proposed approach.

The following literature review was conducted using search engines such as Pubmed and Google Scholar. After an initial review of the most relevant studies, a comprehensive table was created (similar to Table 3.1) summarizing each paper's motivation, research problem, hypothesis, methodology, main findings, limitations, and literature gaps. This table served as a guide for further investigation. Next, references considered relevant to the selected papers were examined, and the same procedure was repeated for the remaining papers, leaving the final tables as the ones presented in the report.

3.1 ABCA4 and PRPH2 current therapies and detection techniques.

As mentioned in section 2.3, ABCA4-retinopathy represents the most frequent inherited retinal degeneration and is directly related to lipofuscin accumulation in the RPE. This accumulation can be quantified in vivo using autofluorescence, where the increase of autofluorescence precedes the loss of function of the transporters, implying that vision could be preserved if lipofuscin accumulation is slowed [33].

Rapid and non-invasive, fundus autofluorescence (FAF) imaging is an effective technique to investigate the health and functionality of the RPE. Smaller FAF areas, conversely, are indicative of low-level RPE metabolic activity, often associated with localized atrophy and subsequent loss of photoreceptors. Therefore, FAF serves as a comprehensive tool for staging and diagnosing Stargardt disease, particularly when complemented with OCT data. An abnormal increase in FAF intensity, alongside normal parameters, suggests that the initial pathophysiological event in ABCA4-related disease may be the deposition of RPE lipofuscin.

Other approaches include a diagnosis based on family history, visual acuity, fundus examination, visual field-testing, fluorescein angiography, electroretinography, and OCT, as shown before [44].

However, to confirm this condition, genetic screening is necessary. The ABCR400 microarray is a comprehensive genetic testing tool developed to detect all known disease-associated genetic variants and a wide range of common ABCA4 polymorphisms. Overall, detection rates with these types of microarrays range between 65% and 75%, and several laboratories worldwide provide genetic testing of the ABCA4 gene. A recent report using single-molecule molecular inversion probe (smMIP)-based sequencing has demonstrated a coverage of 97.4% of the 128 kb ABCA4 gene, including all 50 exons and splice sites, and can also identify copy number variants, proving also to be a cost-effective solution for the purpose [32].

However, ABCA4 remains a complex and expensive gene to screen due to its size, highly polymorphic nature, and wide spectrum of variants identified [32]. Despite the several detection techniques, ABCA4 retinopathy remains an incurable condition. Patients are advised to avoid supplements containing vitamin A, thanks to a study that showed lipofuscin accumulation in an ABCA4 mouse, where vitamin A was injected. Additionally, it is also advised to wear protective, dark-tinted glasses in bright conditions, to reduce short wavelength light reaching the retina, thus reducing the risk of light toxicity [32].

Pharmacological therapies are mainly targeted to aspects of the visual cycle, to reduce lipofuscin deposits. Gene therapy is an ongoing, complicated procedure since the ABCA4 gene is larger than the vectors used for similar diseases.

The potential emergence of various therapeutic approaches for ABCA4 underscores the significance of early patient identification, enabling timely treatment at the early stages of the disease to prevent progression and maximize the benefits of emerging therapies.

PRPH2 in opposition to the ABCA4, has a relatively small cDNA (≈ 1.1 kb in its coding region), making it a much easier task to deliver in therapeutic approaches, such as gene replacement therapy ([45], [46], [47]), gene knockdown and replacement therapy [48], delivery of neurotrophic factors [49], and genetic treatment of disease sequelae such as neovascularization [50], however, the complex pathogenic mechanisms of PRPH2 associated macular disease and the need for a precise dose of peripherin-2 to address haploinsufficiency have delayed the development of effective gene therapy treatments.

In a study to find visual differences between these two genotypes, Jeffery *et al.* [20] evaluated, using OCT scans, the outer retinal bands both in ABCA4 and PRPH2-associated retinopathies. This study reported that the band two thickness of the outer retinal layers was significantly greater in the PRPH2 group when compared to the ABCA4 and healthy control groups.

Several factors could account for this matter, including the widening of ISeZ, the greater reflection at the IS/OS, or the merging of bands 2 and 3. PRPH2 plays a crucial role in disc formation, and when a mutation occurs, studies show elongated and disorganized discs, with compacted open discs at the base of the OS. These alterations can lead to an accumulation of light scattering subcellular structures within the IS and increased reflectivity thanks to the disorganized OS discs, which also promote the broadening of the band two profile, and therefore possibly leading to the fusion of band 2 and 3.

The same study also reported that the thickness of band 4 was significantly greater in the ABCA4

group, compared again with the PRPH2 group and healthy controls. The increased band four thickness in patients with variants in ABCA4 could be caused by the increased height of band four alone. The endogenous expression of wild-type ABCA4 in the RPE may play a role in recycling retinaldehyde in the endolysosomes. ABCA4 plays an essential role in the clearance of all-trans-retinal. Hence, lack of function results in the accumulation of bisretinoids in the OS discs and the RPE as lipofuscin. In an animal model of ABCA4, electron microscopy showed increased phagosomes with multilayered accumulation within the RPE, suggesting a likely isolated thickening of band four.

This work, reporting an altered outer retinal band profile, was best characterized by the band 2/band 4 ratio. This metric successfully discriminated between PRPH2 and ABCA4-associated retinopathy with up to 99% accuracy in a receiver operator characteristic (ROC) analysis.

This section highlights the critical importance of establishing a rapid and precise detection mechanism for these conditions. Swift detection not only allows for the implementation of treatment measures but also enables proactive interventions to prevent further developments in each disease. The differentiation of the outer retinal profiles between the two inherited retinal diseases enables a basis to create an automatic model to classify these pathologies, using segmentation techniques that can differentiate and measure the layers' thicknesses.

3.2 Deep Learning algorithms for segmentation.

CNNs have been successfully applied to various computer vision and image processing applications, including image classification, object detection, and image segmentation. The reduction of time, the detection of subtle changes, and the removal of observer bias and intra-observer variability can be of great help to improve everyday ophthalmic examinations.

Recent advancements in deep learning have revolutionized the automatic analysis of retinal pathologies from OCT images, offering promising solutions for segmentation tasks. The U-Net stands out for its consistent superiority in comparative studies among various deep learning architectures.

First introduced by Ronneberger *et al.* in 2015 [51], the U-Net rapidly gained recognition after achieving remarkable performance in the ISBI cell tracking challenge. This success has provoked the development of numerous U-Net variants, each specifically designed for a certain imaging modality. These advancements have facilitated automated retinal disease diagnosis and prognosis, among others.

The U-Net architecture is composed of two main components: an encoder and a decoder. The encoder takes the input image and progressively downsamples it, reducing its spatial resolution while increasing the number of feature channels. In simpler terms, the encoder captures more abstract and high-level features from the image, with a structure similar to a standard CNN, with several pooling and convolutional layers.

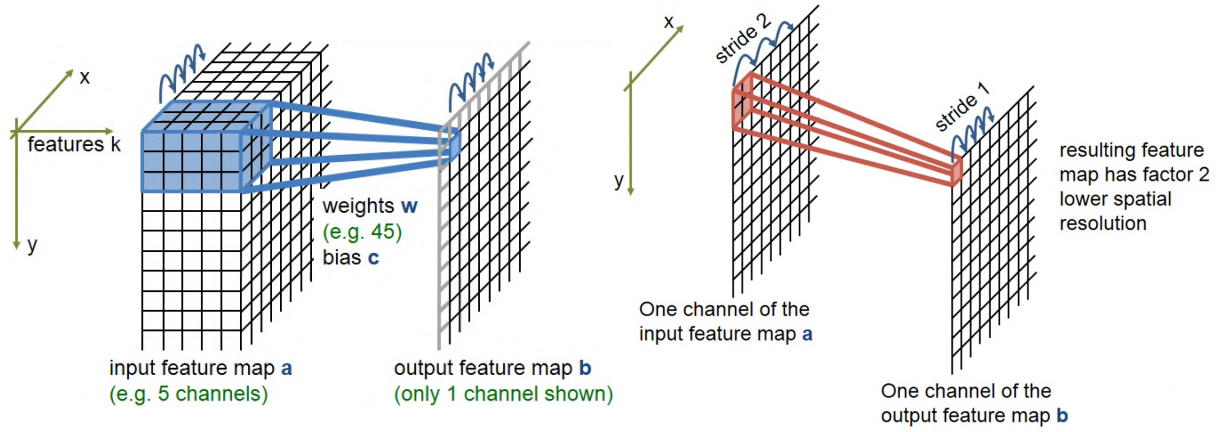


Figure 3.1: Representation of a Convolution and a MaxPooling operation. On the left, a standard 3×3 convolution, represented in Figure 3.3 by the blue arrow, followed by a non-linear activation function (ReLU) layer, while on the right the MaxPooling layer, represented by the red arrow, reduces the xy size, propagating the maximum activation in each 2×2 activation window. From reference [51].

The decoder, on the other hand, takes the downsampled feature maps from the encoder and reconstructs the original image, but with segmented regions clearly defined. It achieves this by upsampling the feature maps and using skip connections, which are "shortcuts" that allow the decoder to directly access the encoder's output. This enables the decoder to utilize both high-level and low-level features to accurately segment the image.

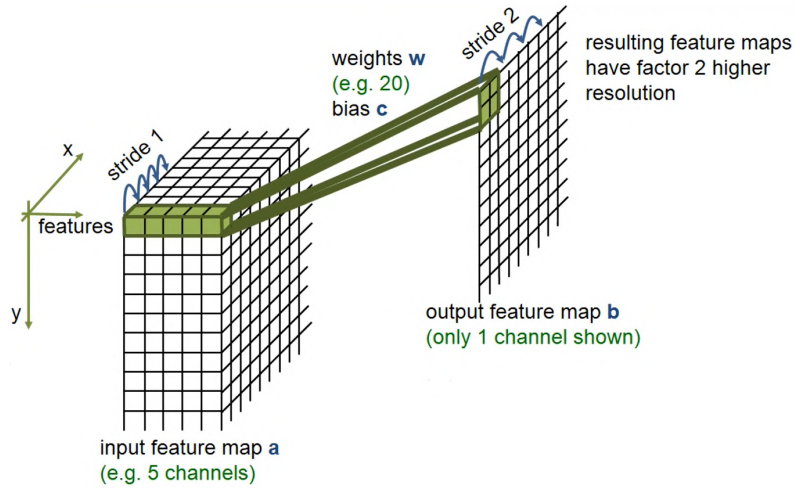


Figure 3.2: Representation of a Deconvolution operation. Deconvolution represented as the green arrows in Figure 3.3, up-sampling the space resolution by a factor of 2, followed by a ReLU activation function. From reference [51].

In summary, the main characteristics of a U-Net network are:

1. Fast and precise image segmentation technique.
2. Encoder-decoder network, able to extract both high-level and low-level features.
3. Existence of skip connections, that reintroduce detailed features of the encoder into the decoder, allowing for a precise pixel-level localization.

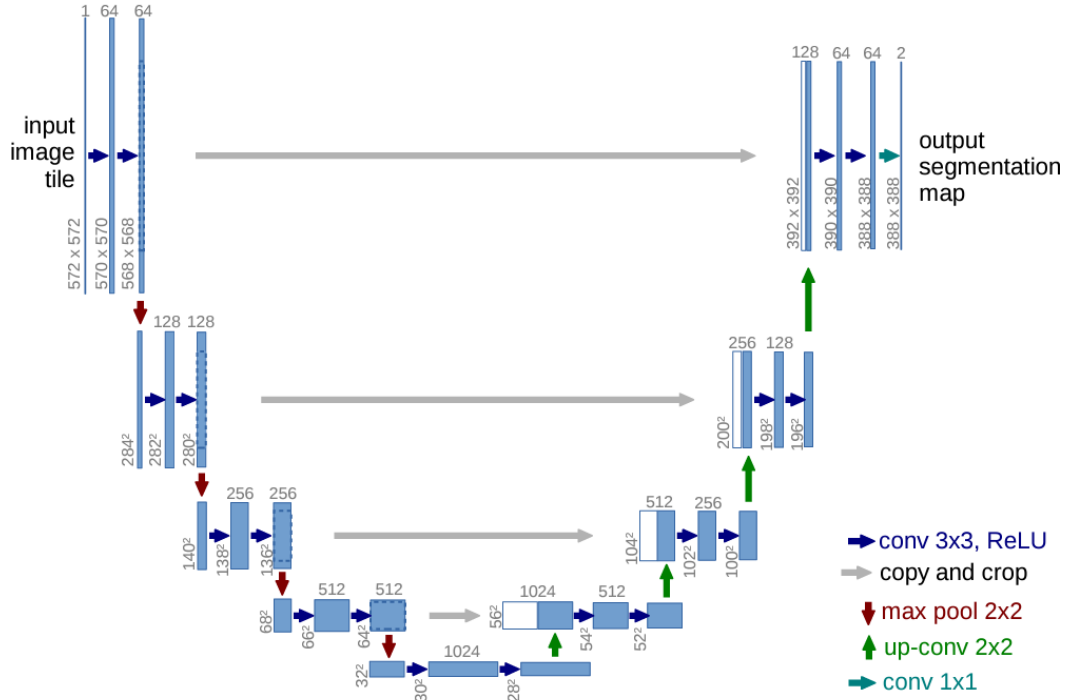


Figure 3.3: Architecture of the first U-net developed by Ronneberger. From reference [51].

In 2018, the U-Net++ was introduced by Zhou *et al.* [52]. It was originally advertised as a “new, more powerful architecture for medical image segmentation.”, where the encoder-decoder structure is composed of several nested U-Nets, where each layer receives the skip connections from the previous one (dense skip connections). The authors suggest that the network can better capture the details of foreground objects if the high-resolution feature maps from the encoder network are first enriched before they are combined with the corresponding semantically rich feature maps from the decoder network. After the introduction of this adapted network, several segmentation works have been developed with this type of network ([53], [54], [55]).

DRUNET also standing for Dilated-Residual U-Net, was also an adaptation of the U-Net developed in 2018 by Devalla *et al.* [56]. It was originally developed with the sole purpose of segmenting the optic nerve head (ONH) in OCT scans. As the name suggests, this architecture uses dilated convolutions, a type of convolution that “inflate” the kernel by inserting holes between the kernel elements, allowing CNNs to have a better receptive field, meaning that a neuron can analyse a larger portion of the image. This method allowed the capturing of both local (tissue texture) and contextual information (spatial arrangement of the tissues), showing improvements over the U-Net architecture in segmenting structures from OCT scan images.

A more recent segmentation model that has also shown significant performance improvements in segmentation tasks is the SegResNet proposed by Myronenko *et al.* [57], in 2019. This network uses ResNet blocks, where each block consists of two convolutions with normalization and ReLU, followed by additive identity skip connections [58]. In the same manner as the previous network, this one also uses dilated convolutions to capture long-range dependencies in the input image.

Despite all the models mentioned above having exceptional performances in their specific tasks,

considering OCT imaging, and more specifically segmentation of retinal layers, there have been few reports of works regarding that subject. Therefore, all the approaches are possible candidates for the final architecture of the project, depending on the overall quality, signal-to-noise ratio, and training set for the task.

The next table summarizes the most important segmentation works reviewed, with their respective contributions, limitations, and possible improvements.

Table 3.1: Summary of Segmentation Studies.

Authors	Contributions	Architecture	Results	Limitations
Hu <i>et al.</i> [59]	Proved that an increase in patch size can improve segmentation accuracy in patch segmentation techniques.	MCNN		Creates overlapping information of image patches, increasing computational overhead.
Sousa <i>et al.</i> [60]	Bilateral filter + Otsu Preprocessing algorithm. Use of UNet to reduce the search region and separate retinal layers, and DexiNed for edge detection.	U-Net + DexiNed	MAE of 0.49, 0.57, and 0.66 for the segmented layers.	Dataset with few varieties of AMD stages, limiting model's performance.
Li <i>et al.</i> [61]	Atrous Spatial Pyramid Pooling (ASPP) network captures retinal multiscale feature information. Requires few manual segmentations for training.	Xception65 + ASPP + U-Net	Mean intersection over and Se of 90.4 and 92.15%.	Need for better image enhancement processing.
Borkovkina <i>et al.</i> [62]	Demonstrated real-time retinal layer segmentation by incorporating an optimized UNet in the OCT retinal scanner.	U-NetRT	Outperforms UNet by a factor of 21 in terms of inferring speed.	Could not perform single-frame segmentation due to insufficient SNR.
Yeganeh <i>et al.</i> [63]	Useful in structures such as vessels for medical diagnosis, due to improvement in shape continuity metrics.	VGG-16+DRIU	Overall improvement in connectivity measures.	Underperforms in pixel-wise metrics.
Chen <i>et al.</i> [64]	Does not require prerequisite retinal layer segmentations.	SegResNet	Dice overlap coefficient of 0.82 ± 0.10	The training set did not cover a wide range of appearances.
Bai <i>et al.</i> [65]	Cystoid macular edema segmentation, while also providing potential to cope with a variety of symptoms.	FCN-8	Dice coefficient of 0.61 ± 0.21 , outperforming other works on the matter.	Limitations in extracting accurate locations of small target regions.
Xu <i>et al.</i> [66]	Pigment epithelium detachment segmentation in polypoidal choroidal vasculopathy patients.	S1-Net + S2-Net	Outperformed existing segmentation methods in all accuracy parameters.	Database-limited algorithm testing and performance.
Kugelman <i>et al.</i> [67]	Accurately determine the location of the choroidal boundaries of interest, while exploring semantic segmentation architectures.	DRUNET		Need to test in cases of ocular pathology and older populations.

Legend: MAE = Mean Absolute Error; AMD = Age-related Macular Degeneration ; SNR = Signal to Noise Ratio;

3.3 Algorithms for segmentation on OCT images.

With the increase in the resolution of OCT machines, the interest in the segmentation of retinal layers using different approaches has grown. For example, *Bagci et al.* [68] reported good reproducibility compared to manual annotations on 6 retinal layers, performing boundary detection based on a 2-D edge detection scheme. One limitation of this work was the layer division, which only accounted for 6 out of the 16 layers, established in the present work. This causes the layers to have a wider thickness, creating an easier environment for clinicians and the algorithm to detect.

Maidana et al. [69] developed software for analyzing DAPI-stained retinal cross-sections. The tool identifies pixels surrounding the retinal layer, then iteratively applies the AnalyzeSkeleton function [70] to track and map surrounding structures, progressively refining the layer boundary until a precise mask is created. This approach allowed for better detection of thinning layers ($p = 0.006$) than manual annotators ($p = 0.069$). However, both the DAPI-stained and the microscopic images restrained the usability of this method.

The paper presented by *Tarassoly et al.* [71] explored the use of outer retinal thickness pixel maps from SD-OCT to detect hydroxychloroquine (HCQ) retinopathy. This study noted thickness reductions as small as 10–15 μm , indicating the early stages of retinopathy before symptoms arise. However, this disease is often irreversible, and therefore the study does not address whether early intervention can fully prevent or reverse the damage.

Shahidi et al. [72] proposed a method using pixel profiling and edge-enhancing filters to analyze OCT image intensity, detecting peaks corresponding to the total retina, inner retina, outer retina, and POS. Despite its effectiveness, the method shares limitations with similar approaches, only able to detect thickness changes greater than 20 μm with 95% confidence. Moreover, local variations in the optical properties of the retina due to inhomogeneous pathologies may interfere with the peak detection algorithm's accuracy.

Various neural network-based approaches have been explored for retinal segmentation, each developed with different goals in mind, such as improving layer segmentation accuracy, enhancing disease diagnosis, or supporting clinical decision-making. *Roy et al.* [73] developed a fully convolutional deep architecture, termed ReLayNet, for semantic segmentation of retinal OCT B-scan into 7 retinal layers and fluid masses, exhibiting superior performance in standard metrics including Dice loss, retinal thickness estimation and deviation from layer contours. *Venhuizen et al.* [74] developed a robust and reliable alternative for subjective and time-consuming manual measurements of full retinal thickness. Similar studies with dCNNs also reported good segmentations in all retinal layers [75, 76]. However, as occurred with previous works, the authors do not distinguish band 2 from the POS zone, removing the possibility of evaluating the thickness of this region by itself. Others have used deep-learning segmentation in specific disease types with particular patterns, such as *Fang et al.* [77], segmenting nine retinal layer boundaries in patients with non-exudative AMD, lacking generalization for other disease types, or the work presented by *Kugelman et al.* [78] where a two-layer segmentation was obtained for patients with Stargardt disease.

The table below highlights the key findings from the aforementioned studies. A significant gap is evident in the literature, as few studies focus on the type of segmentation this work aims to achieve. While some researchers have successfully segmented larger retinal areas, none have specifically targeted the outer retinal bands as distinct entities.

In summary, the existing literature proposes various techniques, and it is important to note that the results of these papers vary depending on the specific task, dataset, and the DL technique used. The main limitations of the previous works on OCT retinal analysis, considering the main findings reported in the previous tables, are the following:

- A significant quantity of data is needed to train and test deep learning models, which may not be available.
- There is a lack of interpretability in the models, especially on the clinicians' part, lacking a comprehensive understanding of what features are important to analyse.
- There is a need to develop a concise pre-processing block to deal with noise, pixel intensities, and

regions of interest for retina segmentations.

- Few studies are reporting the segmentation of the hyperreflective bands separately;
- No studies report and compare the hyperreflective bands by themselves to manual annotations.

The main features/contributions of the proposed work developed during this master thesis are the following:

- Use of a pre-processing block to help deal with different image inputs;
- Development of a segmentation network to accurately segment the hyperreflective bands 2 and 3 plus 4 of the retina along with the POS region;
- Compare the performance of two different manual annotators and evaluate the consistency in the different regions of the fovea;
- Compare the performances of different dCNNs in the segmentation of the retinal bands;

Table 3.2: Summary of OCT Segmentation Studies.

Authors	Contributions	Results	Limitations
<i>Bagci et al.</i> [68]	Employed a 2-D filter with a wedge-shaped pass band. Utilized correlation between adjacent A-scans increasing the effect of edge detection.	Reported consistent results in 4 of the 6 retinal layers, with small discrepancies.	Only measured thicknesses of 6 out of the 16 layers.
<i>Maidana et al.</i> [69]	Developed a fast, accurate, and precise retinal thickness measurement tool for microscopic images.	Detected thinning of the outer nuclear layer better than manual annotators.	Restricted to microscopic images.
<i>Tarassoly et al.</i> [71]	Detected small thickness reductions in the retina indicating early stages of HCQ retinopathy.	10-15 μm thinning detection.	Often irreversible disease and the method only focuses on two retinal bands.
<i>Shahidi et al.</i> [72]	Reported an algorithm capable of detecting 4 regions in an OCT image using pixel profiling and edge-enhancing.	Thickness of the photoreceptor outer segment through the fovea ranged between 42 and 50 μm .	Limited to detecting larger regions, such as the full outer segment region, limiting the application to singular bands.
<i>Roy et al.</i> [73]	Reliably segment seven structures of the retina.	Dice scores of 0.92 and 0.90 in the retinal bands.	Limited accounting for the different hyperreflective bands.
<i>Venhuizen et al.</i> [74]	Total retina segmentation robust to severe retinal pathologies.	Mean absolute error of $14.0 \pm 22.1 \mu m$ in segmenting the full retina.	Only viable to segment the full retina.
<i>Pekala et al.</i> [75]	Used a DenseNet-based semantic segmentation to differentiate seven retinal layers	Mean unsigned error of 1.06 pixels, lower compared to the human error (1.10 pixels)	Should include analysis on more severe variations and severity of retinal diseases
<i>Fang et al.</i> [77]	Convolutional neural networks and graph search based method for automatic segmentation of nine layer boundaries on non exudative AMD OCT images.	Differences of less than 2 pixels compared to a manual grader.	reliance on the availability of a large annotated data set. Only suitable to non exudative AMD patients
<i>Kugelman et al.</i> [78]	Fully semantic network and graph search that delivers good performance in segmentation of OCT images of pathologic retina in Stargardt disease.	Boundary mean absolute error of 0.23 and 1.12 pixels for the ILM and the base of RPE, respectively.	Small dataset and large segmentation regions only.

Legend: HCQ = Hydroxychloroquine; AMD = Age-related Macular Degeneration

Chapter 4

Methodological Framework

As mentioned in chapter 1, the research focused on detecting the current variations in thicknesses of the hyperreflective bands in different individuals. This chapter details the methodological approach taken in this research, starting with the data acquisition process, which includes the study's design, selection of participants, and the characterization and annotation of the SD-OCT b-scans. The chapter then describes the image preprocessing techniques employed to prepare the data for analysis, ensuring accuracy and consistency in the subsequent steps. It also outlines the initial baseline solution, explaining the methodologies and algorithms applied in the early stages of the research. Finally, it discusses the enhanced solution, detailing the techniques implemented to improve the performance of the segmentation models. Figure 4.1 reports a schematic representation of the research work.

4.1 Data acquisition

4.1.1 Study Design and Settings

We conducted a cross-sectional, observational study based on foveal-centered SD-OCT scans of healthy participants based on a retrospective analysis of OCT images sourced from three primary datasets. The first source is a Portuguese-based registry that is part of the RETINA.PT platform (<https://www.retina.com.pt>), managed by the Portuguese Retina Study Group, includes data from Centro Hospitalar Universitário de Coimbra and Hospital Garcia de Orta (HGO), and the second source is the control group from the IPL/IDI&CA2024/INSYDE_AMD_ESTeSL study.

This study consisted of individuals who voluntarily participated and underwent thorough ophthalmological assessments and genetic profiling, recorded during the database query. The inclusion of patients with a clinical diagnosis of Stargardt disease or PRPH2-retinopathy was excluded from the analysis and evaluation of the metrics and algorithms presented later, due to challenges in obtaining imaging data from these individuals.

Due to constraints in recruiting patients with specific retinopathies, the project validated the model using healthy individuals. Eight patients with ABCA4 mutations have already been selected for upcoming studies and will be added to the project in the future.

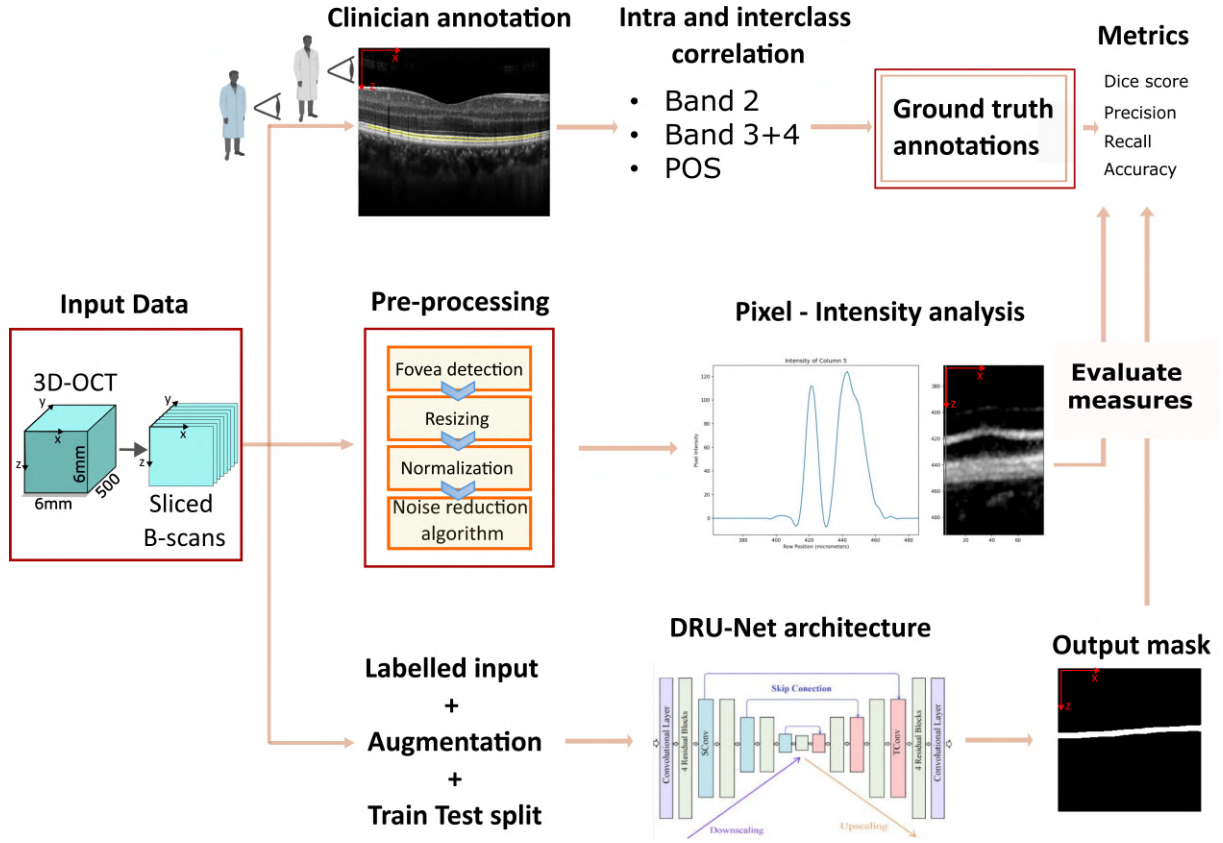


Figure 4.1: Schematic representation of the proposed pipeline of the project. After the Input Data box, represented in this chapter by section 4.1 three paths were taken, corresponding respectively to the clinical annotations (4.1.4), referencing the calculus of the correlation coefficients among clinicians, the pixel-analysis approach (4.4), and the dCNN approach (4.5), mentioned in the second and third row, respectively. The arrows comparing the results with the ground truth annotation on the right of the image refer to chapter 5.

4.1.2 Patient Participation

All procedures in this study, including data collection, adhered to the ethical principles outlined in the Declaration of Helsinki. Participation is without any associated costs, payment, or gratification. Participants can discontinue their collaboration at any time with no consequences or costs, and all enrolled patients gave their written consent at the time of recruitment.

Exclusion criteria included patients with any type of abnormality in the region of the bands of interest, including advanced stages of Drusen disease, that would make unclear the definition of the hyperreflective bands, or OCT B-scans with an inclination over 10° .

An additional date was included whenever possible: subjects, age, sex, weight, best corrected visual acuity (BCVA) using the ETDRS letter chart at the time of imaging, height, and eye laterality (right or left) (Table 4.1). The quantitative data are presented in tables as type of measurement, level, and value of the measurement for the different patient types.

Table 4.1: Summary of main clinical data.

Measurement	Level	PRPH2	ABCA4	Controls
Number of subjects	-	0	8	12
Age (years)	-	-	-	71.60
Sex	Male	-	4	6
	Female	-	4	6
Average Weight (kg)	Male	-	-	77.4
	Female	-	-	67.5
VA (ETDRS letter score)	-	-	-	79.6
Average Height (m)	Male	-	-	1.70
	Female	-	-	1.56
Study eye	RE	-	8	10
	LE	-	8	11
Total number of measurements	-	0	0	63

Legend: The column corresponding to PRPH2 annotations is informative that no subject was acquired with this pathology. Specific information for individuals with ABCA4 pathologies was also not provided; VA = Visual Acuity; RE = Right eye; LE = Left eye.

4.1.3 Image characterization

Anonymized OCT scans from an SD-OCT device (Spectralis; Heidelberg Engineering, Heidelberg, Germany) were obtained and retrospectively analyzed. The images selected for this study were originally captured during routine clinical practice and consisted of high-resolution macular volume scans (20°×20°, 49 B-scans, 7 frames per scan). These scans included 49 raster horizontal B-scans, each containing 1024 A-scans per B-scan, with a depth resolution of 3.9 μm . Three horizontal scans were analyzed for each macular volume scan - one centered on the fovea, one above, and one below the fovea. Both eyes of each patient were selected if inclusion criteria were met.

These SD-OCT images were essential for detailed work involving minute distances compared to High-speed images. This imaging technique, with a light source capable of delivering an axial resolution of up to 3 μm , provides researchers with an enhanced understanding of the retinal anatomy and vasculature, thus serving as the primary imaging method for this study.

4.1.4 Image annotation

Two retina specialists performed three different structure segmentations from normal patients on 21 of 24 eyes, totalling 63 measurements. Annotations on the three regions of interest were performed to validate and compare the developed method's performance.

Measurements were taken to assess intraclass correlation coefficient (ICC) and interclass correlation, defined as the difference in repeated measurements by the same observer and by different observers, respectively. The ICC is defined as a measure of reliability, more specifically the reliability within a class of data to measure subjects similarly [79]. Inter-rater or interclass reliability is crucial, indicating that a scale is consistent despite rater changes. Therefore, scales with high inter-rater reliability are less susceptible to measurement errors arising from differences in human judgment [80].

Each clinician performed each measurement on the same B-scan twice to evaluate intra-observer variability, while the difference between the two clinicians' measurements indicated inter-observer variability in each region.

All measurements were obtained using the inbuilt Heidelberg caliper function using 400% magnification. Band 2 was measured by identifying its internal and external hyperreflective boundaries. Band 3 and Band 4 were defined by the internal hyperreflective boundary of the interdigitation zone and the external boundary of the RPE/Bruch's complex, respectively. For the segmentation of POS, the external boundary of Band 2 and the internal boundary of Band 3 were used as reference points.

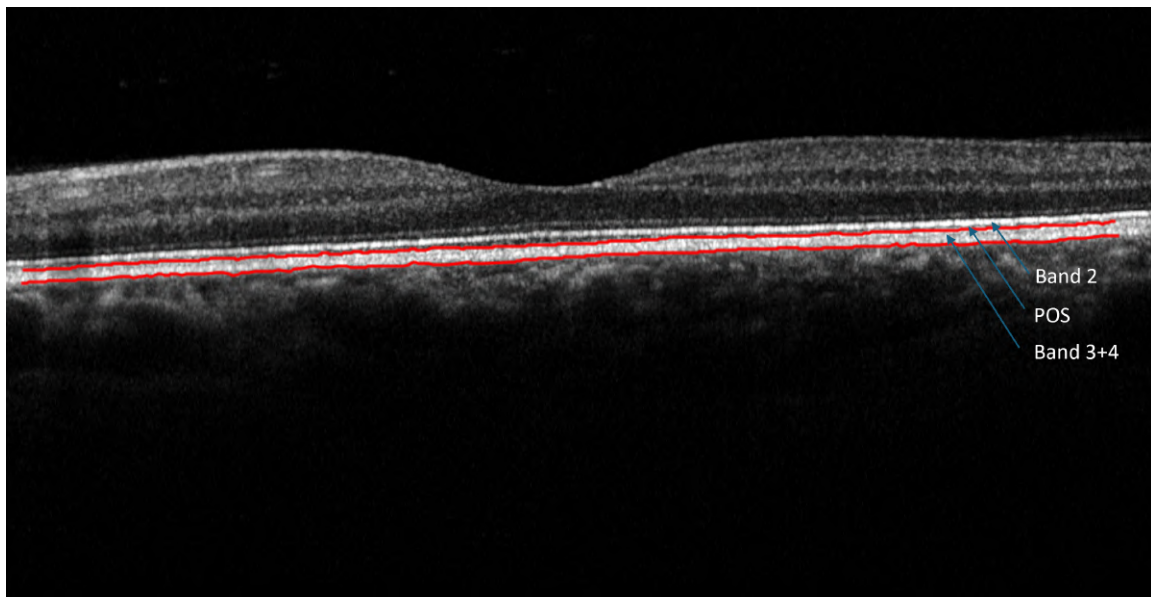


Figure 4.2: Example of an SD-OCT B-scan centered at fovea with (Band 3 + 4) human annotation. The inner red line corresponds to inner Band 3 and the outer red line corresponds to outer Band 4.

We noted that the structures between these 2 bands varied considerably between individuals and within a given scan. Some patients exhibited a single hyporeflective gap (i.e., absence of Band 3), while others displayed three distinct hyperreflective bands or cases where Band 2 was indistinguishable. These variations posed challenges for clinicians and algorithms, potentially affecting segmentation outcomes.

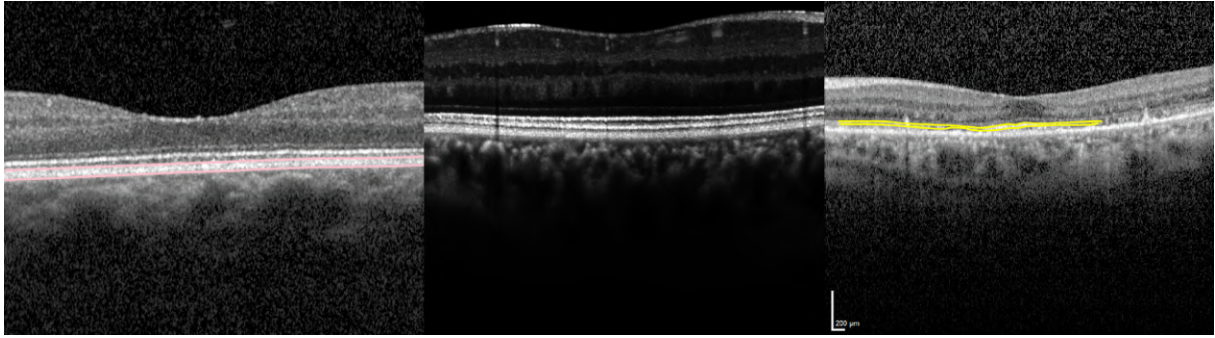


Figure 4.3: Examples of different band structures. On the left, a manually annotated image shows Bands 3 and 4 without a visible distinction between them. In the middle, a B-scan image displays all three bands of interest, with each band distinctly visible. On the right, a manually annotated Band 2 is shown, where the annotation was left incomplete due to the band vanishing.

While the first clinician performed all 63 measurements twice, the second clinician performed 24 measurements twice, totaling 174 measurements for each band. With this data, the intra- and inter-correlation coefficients between these two clinicians could be obtained, and these annotations could be used as ground truth images for the deep learning models.

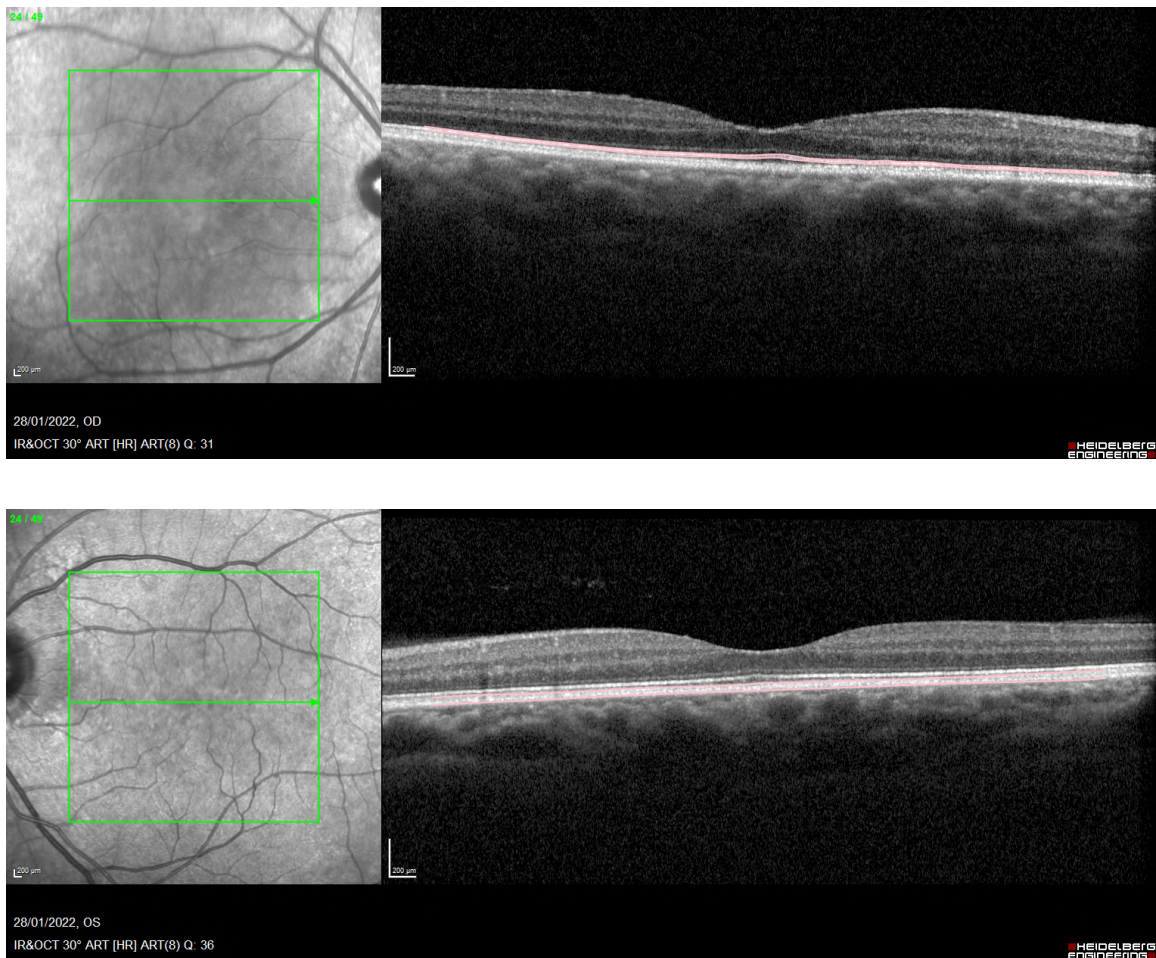


Figure 4.4: B-scan OCT images of different eyes from the same patient. The first image shows a manually segmented image of Band 2, while the second image presents an example of manually segmented images of Bands 3 and 4.

4.2 Image preprocessing

As mentioned in several papers ([6], [8], [10], [11]), a robust preprocessing algorithm is essential for ensuring optimal model performance in various image analysis tasks. Firstly, the images were converted into PNG files, as these offer smaller file sizes compared to TIFF, the original format of the received images. This conversion is crucial in reducing the computational load for all the subsequent steps in all the various approaches.

To address the need for working with RGB images, the next step involved converting the image to greyscale, using the `.convert("L")` method from the Pillow library. This transformation significantly reduced each file size, thereby accelerating the training process and simplifying the input for machine learning algorithms.

The second problem involved automatically detecting the central fovea to compare the different thicknesses in the different regions of the retina centered around the fovea, as further detailed in Figure 4.7. This was performed utilizing the reflectivity of the nerve fiber layer and the depression caused by the central foveal region. Based on a specific threshold, this algorithm focuses on the intensity of the pixels starting at the top of the image. For each column in the image, the first pixel above the threshold defined, accounting for noise in the background, is set to be the beginning of the retina. From there, tracking the retina's boundaries, it was possible to identify the central fovea. This region, identified by its characteristic depression and absence of internal layers, corresponds to the first inflection point in the retina's distance from the top of the image, or the 0 in its derivative, as shown in the bottom and middle rows of Figure 4.5.

It is crucial to highlight that, due to limitations in image resolution, a Savitzky–Golay smoothing filter was applied to reduce the noise effect, as illustrated in Figure 4.5. The Savitzky–Golay filter uses polynomial fitting to smooth noisy data while maintaining key signal features, providing an accurate method for applications that require signal preservation [81]. This filtering process allowed the detected boundary to exhibit a continuous surface, minimizing large fluctuations caused by noise, but also facilitated a more precise identification of the inflection point, thereby improving the accuracy in detecting the central fovea.

This algorithm achieved perfect accuracy in detecting the central fovea in nearly all cases, except with individuals in advanced stages of Drusen disease and OCT B-scans characterized by high noise levels or an inclination greater than 10° . In these instances, the presence of additional depressions or elevations caused by Drusen or continuous pixel distortion from various noise sources hindered the accurate detection of the fovea.

To improve the contrast between the bands and the background of the SD-OCT image, a high-pass filter of 10 kHz was applied in all the cuts, which showed an improvement as expected in the contrast, enhancing band boundaries. This high-pass filter adapted the process mentioned in Li et al. [82], where the authors found high-frequency images, which capture fine details of retinal layers, to improve the performance of deep learning models.

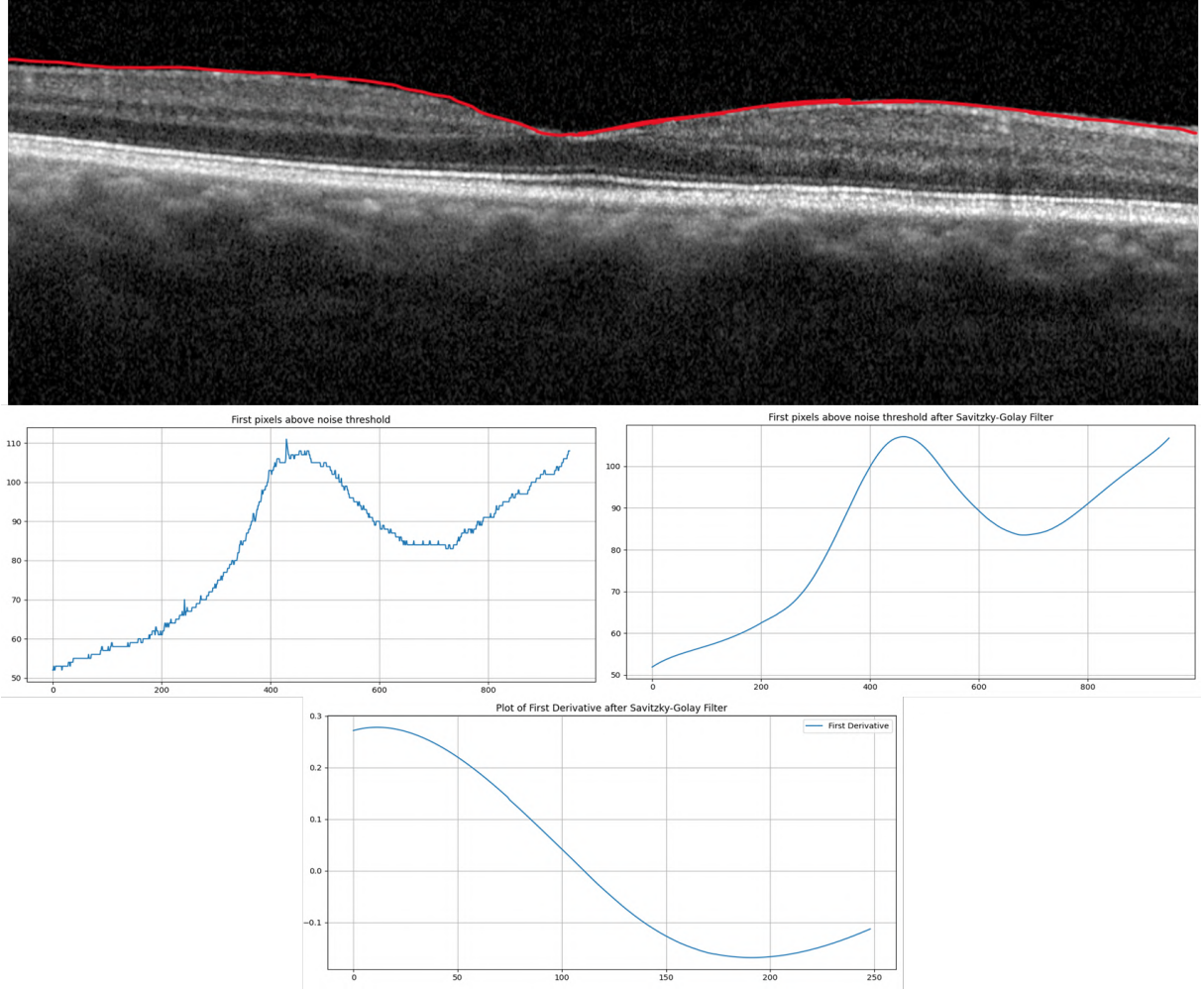


Figure 4.5: Detection algorithm of the central fovea. The top image, highlighted in red, shows the first detected pixels above the noise threshold. The middle row displays (left) the distance from the top of the image to the detected pixels, and (right) the same data after applying a smoothing filter. The bottom graph represents the first derivative, where the zero-crossing point corresponds to the central foveal depression.

4.3 Evaluation Metrics

Considering the need to develop and validate the segmentation tasks, distinct metrics were used to test performances. To evaluate the segmentation accuracy, the literature recommended the following metrics: Pixel Accuracy (PA) and Sorensen-Dice coefficient (Dice).

The PA metric denotes the percent of pixels that are accurately classified in the image when compared to the ground truth,

$$PA = \frac{\sum_{j=1}^k n_{jj}}{\sum_{i=1}^k \sum_{j=1}^k n_{ji}} \quad (4.1)$$

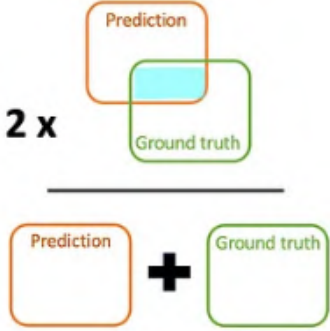
where n_{jj} denotes the number of true positives and n_{ji} the amount of pixels from class j inferred to belong to class i [83]. When considering the last statement, the previous equation can be described as

$$PA = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.2)$$

where TP, TN, FP, and FN denote the pixel-level counts of true positive, true negative, false positive, and false negative [84].

However, high PA does not directly imply superior segmentation performance, especially in unbalanced class data sets. In this case, when a class dominates the image, leaving the other class with only a small portion of the image, the correct classification of the dominant class will yield a high PA, not directly implying a good classification of the non-dominant class, which is the case in this project, where the retina layer is significantly smaller than the background.

The Dice coefficient is a more used metric in unbalanced classes since it predicts the similarity between the predicted and the true mask segmentation. It addresses some limitations of the PA, by simply focusing on the overlap between the predicted and the ground truth mask. The Dice score ranges from 0 to 1, where the last one indicates perfect overlap. It can also be simplified as twice the size of the intersection divided by the sum of the sizes of the two sets, or



$$DICE = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (4.3)$$

Figure 4.6: Figurative representation of the Dice Coefficient, on the left [85], and the corresponding formula on the right [86].

Besides these two specific segmentation metrics, others can be applied to check the model's performance accurately, which can be applied to both segmentation and classification tasks.

Precision offers valuable information about the model's accuracy when it predicts the existence of a particular class, underscoring the significance of reducing false positives. In the context of segmentation tasks, a high precision value signifies a minimal occurrence of misclassifying pixels as part of the designated target class.

$$Precision = \frac{TP}{TP + FP} \quad (4.4)$$

On the other hand, recall, or sensitivity, measures the ability of a model to correctly identify all relevant pixels belonging to a specific class. In segmentation tasks, where missing relevant pixels is critical, a higher recall can indicate a better model performance, however, for a more comprehensive assessment, it should be considered alongside the other measurements.

$$Recall = \frac{TP}{TP + FN} \quad (4.5)$$

By combining these metrics, we can have an all-around good knowledge of how the networks can perform, and therefore alter the hyperparameters, number of neurons, and layers, according to these performance values.

4.4 Baseline Solution

The baseline solution involved the separation of each original B-scan image into 4 different regions of interest with a total width of $50\ \mu m$. To compare the correlations in different areas of the retina, the cuts were considered as follows:

1. Central fovea region, based on the previous foveal detection algorithm mentioned in 4.2;
2. Parafovea region at $500\ \mu m$, one each side of the central fovea;
3. Intermediate distance between parafovea and perifovea, at $750\ \mu m$, one each side of the central fovea;
4. Perifovea region at $1000\ \mu m$, one each side of the central fovea.

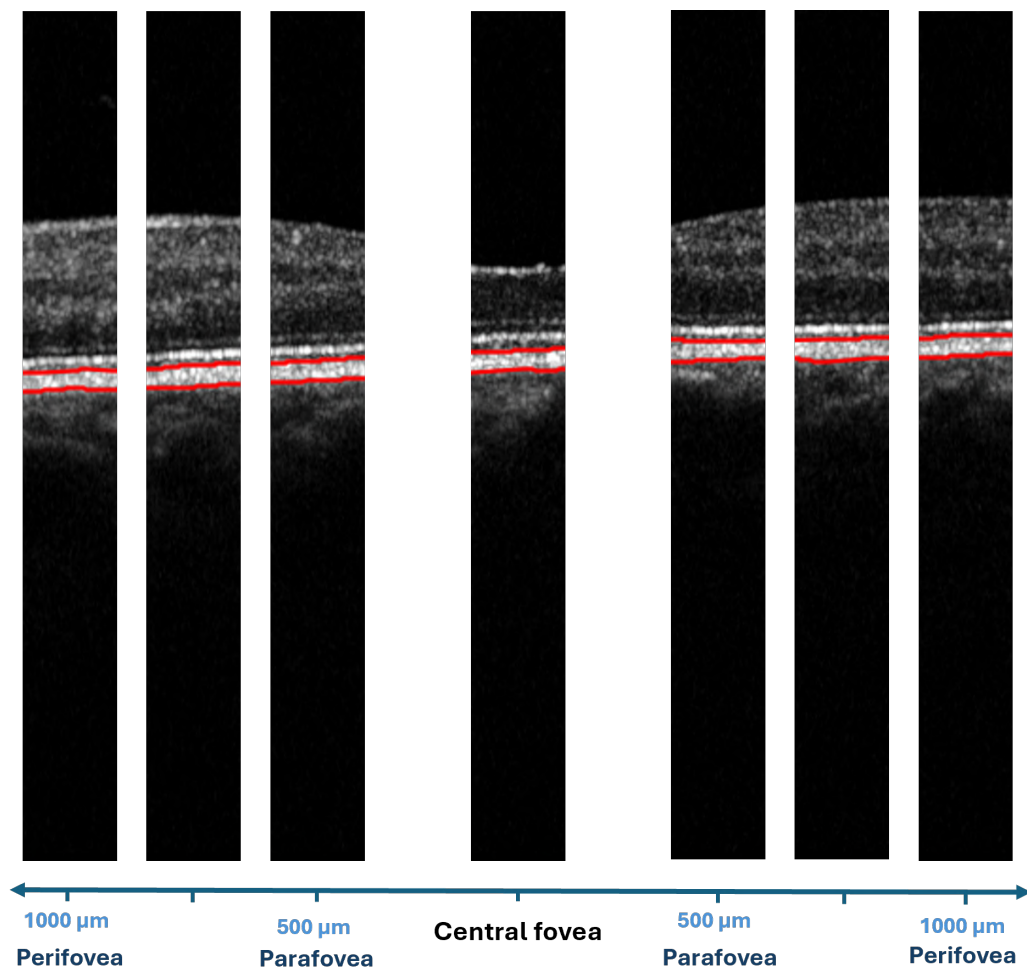


Figure 4.7: Representation of the fovea detection algorithm in a B-scan image, and the respective cropped images corresponding to the regions mentioned at the bottom of the image.

Focusing on the brightest region of the image, which corresponds to the localization of the bands, a script was employed to extract the pixel intensity for each column. Utilizing the peak detection module from the SciPy library, two high-intensity pixels were identified with a minimum separation distance. These pixels were associated with Band 2 and Bands 3 + 4, respectively.

Recognizing the limitations of pixel-based representation in digital images, where each pixel represents a discrete distance of $3.9\ \mu m$, it was necessary to implement a smoothing technique to mitigate the effects of noise and enhance the precision of the extracted data. A Savitzky-Golay filter was employed to achieve this, as it can smooth the data while preserving the underlying signal trend.

Following the identification of the peak pixels, a threshold strategy was implemented to delineate the band boundaries. This threshold was defined as the point where the intensity decreased by a specific percentage compared to the peak intensity, marking the transition from the band to the surrounding region. Subsequently, for each column within the cropped images, the average thickness of the two bands and the intervening space (corresponding to the POS) was calculated in pixels and converted to micrometers.

4.5 Enhanced Solution

A deep learning approach was developed to address the limitations of the previous solution, which prevented specific locations of the image from characterizing the bands of interest through pixel intensities.

The segmentation network was trained using ground truth images annotated by retina specialists for 21 eyes of healthy patients. These annotations were transformed into binary masks and subsequently used to train the segmentation model.

Unlike the previous solution, the images were cropped into 256x256 pixel squares centered on each side of the central fovea. Thus, for each OCT B-scan, two different images were produced: one for the left and one for the right side. Given that many pre-trained models and layers are optimized for square images and that at least 250 pixels are required to approximate $1000\ \mu m$ —the furthest region of interest in this work—this dimension was chosen. Additionally, using 512x512 images would have significantly increased computational demands, making the 256x256 size a more efficient choice.

Several architectures were tested, such as:

- UNet proposed by Ronneberger *et al.*
- UNet++ proposed by Zhou *et al.*
- DRUNET proposed by Devalla *et al.*
- SegResNet proposed by Myronenko *et al.*

Image augmentation was applied to expand the dataset. In this case, augmentations such as enhancing contrast between images, introducing Gaussian noise, and flipping the images in the y-axis introduced a whole new set of data. In total, the training set contained 640 images, the validation

dataset contained 240 images (80 images and 30 images without augmentation, respectively), and the test set contained 16 images.

Regarding common details of the networks, all were performed in a total of 150 epochs or until convergence, Adam Optimizer was utilized as an optimization technique and regularization techniques utilized included Early Stopping and Dropout in every hidden layer (0.2), with the Dice Loss employed as the loss function.

It is important to note that the segmentation masks were not compared in the same manner as the ones in the pixel intensity. Each left or right image was compared to a left or right image from the annotation data and evaluated according to the similarities to the “ground truth” mask after training.

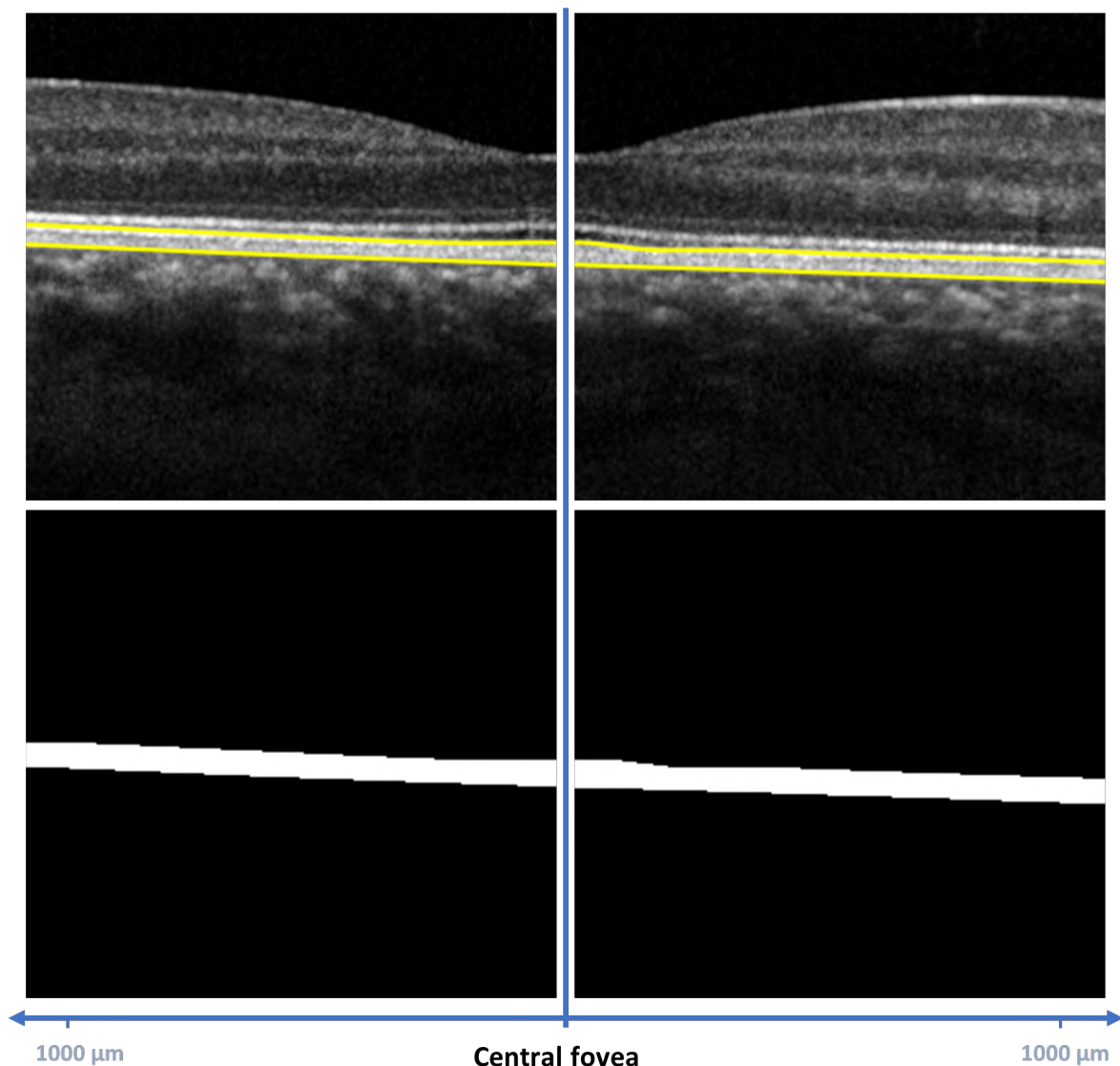


Figure 4.8: Representation of the segmentation masks used in the deep learning algorithm. At the top of the image, Clinician 1 of Band 3 + 4 performs left—and right-sided annotation. At the bottom, the respective binary masks are introduced to train the algorithm.

Chapter 5

Results

5.1 General characterization

A total of 24 eyes were initially included in the study. Of these, 21 eyes, each with three B-scans—one centered on the fovea, one above, and one below—were annotated by each grader, resulting in 63 B-OCT scans. Clinician 1 obtained an ICC of 0.737, 0.794, and 0.880 for Band 2, Band 4, and POS zone, respectively, while Clinician 2 obtained 0.563, 0.873, and 0.782 for the same regions, achieving an overall mean deviation of 1,747 \pm 0.536 pixels. The interclass correlation was 0.830 and 0.754 for Band 4 and POS, while Band 2 obtained 0.106, with a mean deviation between graders of 3.21 \pm 0.53 pixels. The Pixel-analysis thickness results showed an average difference of 4.29 \pm 1.42 pixels in an average thickness compared to the best-performing clinician in terms of ICC. Finally, the deep-learning approach achieved Dice scores varying from 87.72% to 93.25% in the 4 different architectures tested, translating into an average difference of 0.80 \pm 0.19 pixels compared to grader 1.

5.2 Annotation results

As mentioned in the previous chapter, the study was performed on 3 scans of each eye of 12 patients. One patient and one eye of another patient were excluded due to the lack of visible bands.

Two different clinicians retrieved segmentation data on two different acquisition dates. Considering the minute areas evaluated in this study, where each pixel corresponds approximately to 4 micrometers in an image, it was important to assess how closely each clinician could perform the task repeatedly and successfully. Two metrics will be considered to quantify the variability of these measurements: the ICC and the average pixel difference between each measurement.

The first clinician measured 63 OCT images on all three regions, while the second clinician performed measurements on 24 OCT images. The results are shown in the table below. The values in bold represent the best region for each clinician for the evaluated metrics.

Visualizing the measurements in scatter plots made it easier to identify outliers and trace the source of these discrepancies in the results, which will be discussed further in the analysis. The results pre-

sented in Figure 5.1 refer to Clinician 1, who achieved an overall higher ICC among all regions.

Table 5.1: Intraclass correlation for both clinicians with Normal Images.

Clinician	Region	ICC	95% Confidence Interval		Avg Diff (Pixels)
			Lower Bound	Upper Bound	
1	Band 2	0.737	0.689	0.779	1.384
	Band 3+4	0.830	0.795	0.860	1.812
	POS	0.878	0.852	0.900	1.444
2	Band 2	0.565	0.451	0.660	1.547
	Band 3+4	0.888	0.844	0.922	2.906
	POS	0.789	0.724	0.841	1.390
Average	Band 2	0.651	0.570	0.720	1.466
	Band 3+4	0.859	0.820	0.891	2.359
	POS	0.834	0.788	0.896	1.417

Legend: Avg Diff = Average Difference.

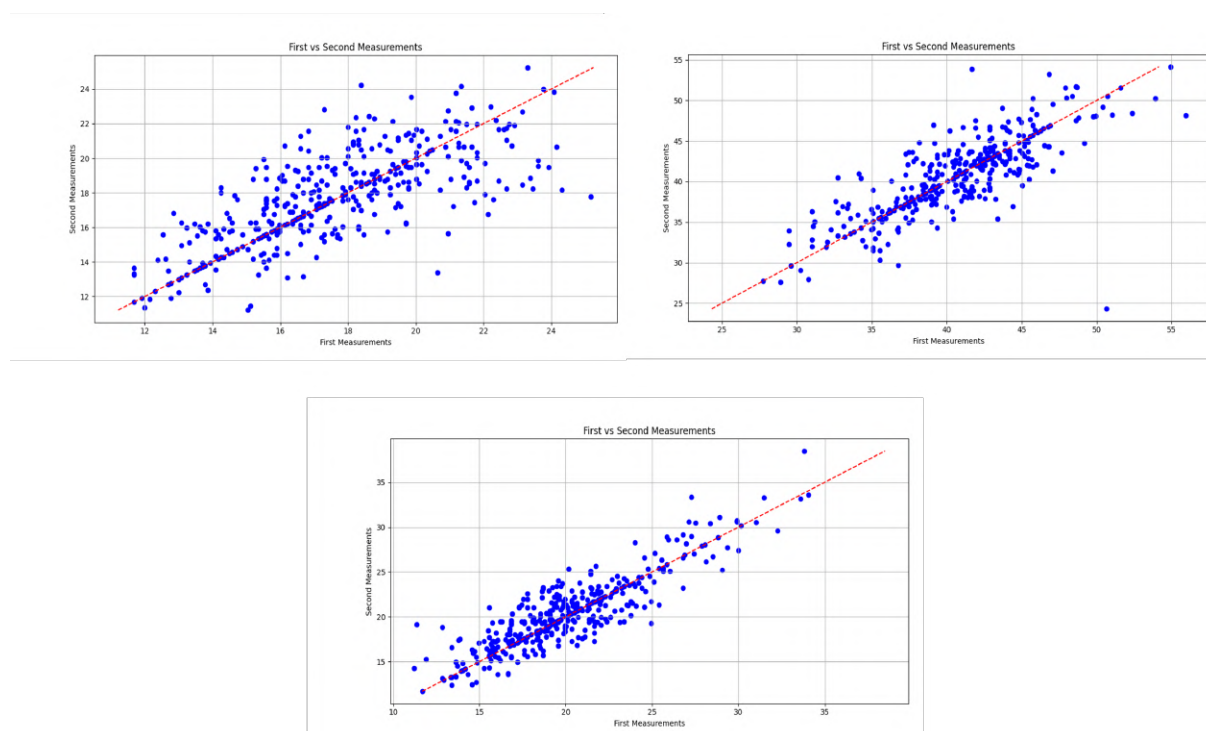


Figure 5.1: Scatter plots of ICC between the first measurements (x-axis) and the second measurements (y-axis) of the different regions. The top left plot shows measurements for the Band 2 region, the top right plot for the Band 3+4 region, and the bottom plot for the POS zone. The red line indicates the $x = y$ line, where measurement 1 equals measurement 2.

To track the difference and their respective reliability between measures, we further analyzed the four different distances from the fovea, as suggested by the study in [20], and evaluated where a higher variability of the measurements occurred, for both clinician 1 and 2.

With the previous fovea detection algorithm, we could automatically obtain the specific cuts for the central fovea, the parafovea (500 μm), the intermediate region between the parafovea and the perifovea (750 μm), and the perifovea (1000 μm). The 50 μm cropped image, as explained in section 4.4, was then calculated as the average for each thickness value detected by the algorithm in each column, and the results were arranged in the table presented in Appendix 1.

We also performed an inter-observational study to assess the differences between the two clinicians when measuring, allowing the confirmation of the precision and trustworthiness of data interpretation. For this comparison, the first measurements of both clinicians were taken into account, and only the images segmented by both clinicians were considered.

The results showed a good inter-class correlation between the two clinicians in two of the three regions studied, highlighted in the table below. Band 2 achieved a correlation of 0.106, a poor result, and to be further discussed.

Table 5.2: Interclass Correlation Coefficient between the two clinicians' first measurements.

Region	Interclass Coef	95% Confidence Interval		Avg Diff (Pixels)
		Lower Bound	Upper Bound	
Band 2	0.104	-0.049	0.253	3.818
Band 3+4	0.810	0.740	0.862	2.832
POS	0.752	0.678	0.812	2.992

Legend: Interclass Coef = Interclass Correlation Coefficient; Avg Diff = Average Difference.

5.3 Pixel-analysis results

The pixel-intensity process as mentioned above was used as a structural baseline for the final results. The compared values were the average measurements of the first clinician since it performed more annotations and achieved more consistent results.

The results presented in the following table in bold represent the ones we consider that achieved either a reasonable ICC or a small Average pixel difference. In the same way as in section 5.2, the results were also shown in a scatter plot, to visualize the dispersion between the two methods.

Table 5.3: Comparison between the manual annotations of clinician 1 and the pixel intensity results.

Location	Region	ICC	95% Confidence Interval		Avg Diff (Pixels)
			Lower Bound	Upper Bound	
Center fovea	Band 2	0.033	-0.230	0.291	2.852
	Band 3+4	0.358	0.085	0.581	4.445
	POS	0.563	0.354	0.718	3.754
500 micrometers	Band 2	0.203	0.019	0.374	3.187
	Band 3+4	0.368	0.182	0.529	3.365
	POS	0.169	-0.015	0.342	6.604
750 micrometers	Band 2	0.255	0.073	0.420	3.417
	Band 3+4	0.437	0.264	0.583	3.645
	POS	-0.049	-0.230	0.136	6.812
1000 micrometers	Band 2	0.237	0.055	0.403	3.294
	Band 3+4	0.428	0.253	0.576	3.857
	POS	-0.179	-0.351	0.004	6.228
Average	Band 2	0.268	-0.002	0.372	3.186
	Band 3+4	0.390	0.196	0.567	3.825
	POS	0.091	-0.061	0.299	5.857

Legend: Avg Diff = Average Difference.

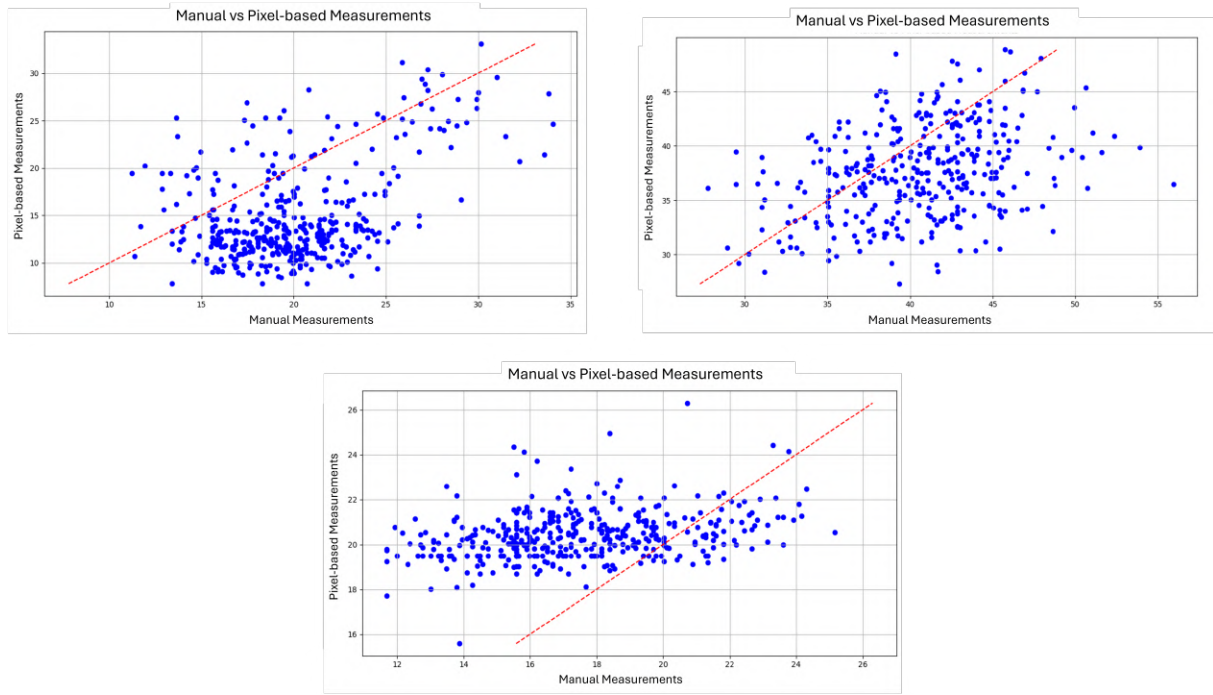


Figure 5.2: Scatter plots of ICC between the first measurements (x-axis) registered by clinician 1 and the second measurements (y-axis) computed by the pixel-analysis algorithm. Each graph represents the Band 2 region (top left plot), the Band 3 + 4 region (top right plot), and the POS zone (bottom plot). The red line represents the $x=y$ or the location where measurement 1 equals measurement 2.

5.4 Deep Learning results

Regarding the segmentation network results, the various architectures tested produced the outcomes presented in Table 5.4. These results were benchmarked against the annotation masks provided by Clinician 1, the best-performing clinician in terms of ICC scores. All models performed well, achieving excellent Dice scores. Notably, the DRUNET architecture delivered the best results in terms of both Dice score and average pixel difference. It is also worth mentioning that the simplest UNET architecture in terms of complexity, demonstrated consistent results compared to the best-performing model. All the other metrics achieved high percentages, especially the Pixel-Accuracy.

To illustrate the input/output process of the different networks, the following image compares the output masks of two DRUNET architectures and the corresponding clinician annotations, regarding the POS zone and the Band 4 region, after introducing the "original image" in the DRUNET network.

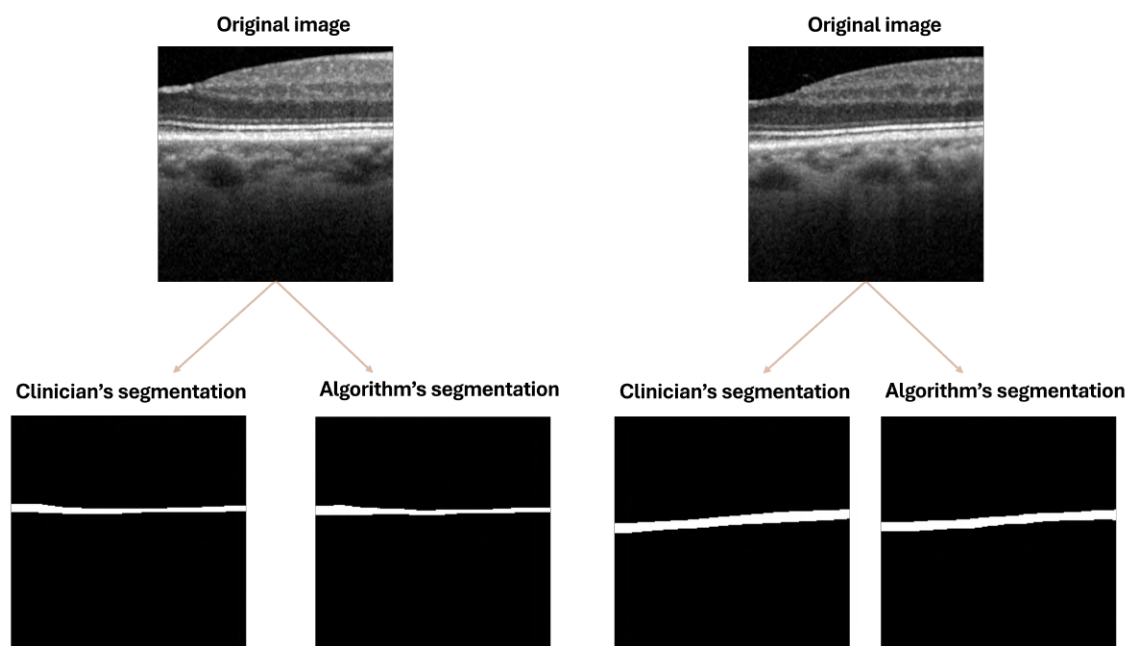


Figure 5.3: Comparison of Clinicians' and Algorithms' Segmentations. The top row shows two unannotated OCT images input into two distinct DRUNET networks. The left panel illustrates the network segmentation of the POS zone, with the corresponding segmentation shown below. The right panel displays the segmentation of the Band 4 region by a different DRUNET network, with the resulting segmentation also visualized below.

Table 5.4: Different segmentation results for different bands in Normal OCT images.

Model	Region	Dice	Accuracy	Recall	Precision	Avg Diff (Pixels)
UNET	Band 2	0.880	0.996	0.893	0.870	0.642
	Band 3+4	0.932	0.995	0.934	0.932	0.927
	POS	0.904	0.996	0.904	0.908	0.756
UNET ++	Band 2	0.877	0.996	0.838	0.923	0.729
	Band 3+4	0.912	0.993	0.967	0.868	1.286
	POS	0.913	0.996	0.893	0.936	0.700
DRUNET	Band 2	0.894	0.996	0.877	0.916	0.667
	Band 3+4	0.933	0.994	0.909	0.956	0.904
	POS	0.914	0.996	0.902	0.924	0.682
SegResNet	Band 2	0.886	0.996	0.870	0.908	0.654
	Band 3+4	0.927	0.994	0.956	0.903	0.994
	POS	0.911	0.996	0.903	0.923	0.715

Legend: The four different models' performances represent each of the four rows, while the bold results represent the best performing Dice and the lowest average pixel difference when compared to clinician 1 for each of the regions studied. Avg Diff = Average Difference.

Chapter 6

Discussion

In the analysis of 63 B-scans from 21 eyes of normal patients, the dCNNs could accurately segment the outer retinal bands. A high ICC indicated consistent manual segmentation accuracy among clinicians, though interclass correlation showed variability. The next sections explore the potential applications of these results compared with state-of-the-art studies.

6.1 Manual annotations

To evaluate the ICC, a standard procedure is to consider that values under 0.5 suggest poor reliability. At the same time, ranges of 0.5 to 0.75, 0.75 to 0.9, and above 0.9 indicate moderate, good, and excellent reliability, respectively [87].

A preliminary analysis concluded that the overall ICC for Clinicians 1 and 2 was generally good. The results presented in Table 5.1 indicate that Band 2 was the most challenging region to reproduce, as shown by the lower ICC values.

Considering the small size of the regions examined, the consistency between clinicians may be affected by various factors, such as subjective visual judgment, differences in screen resolution, and other immeasurable influences, including clinician fatigue. Nevertheless, both clinicians consistently produced comparable results for Band 4 and the POS zone, as demonstrated in Table 5.2. Band 2, as it was already predictable by the variable ICC results in Table 5.1, did not bring significant results in terms of interclass correlation.

Upon a deeper analysis of the Appendix Table 1, it is evident that the correlation coefficients obtained for the central fovea are superior to the others, especially 500 μm away from the center, where the ICC is the lowest for both clinicians.

After analysing, we can conclude a major outcome, important for future clinical evaluations in different contexts in ophthalmology. In terms of performing consistent segmentation results, the center fovea is the best region for both clinicians, indicating that this area is the most reliable for measuring the outer retinal bands. In contrast, the region located 500 μm from the fovea was the least consistent for segmentations. Both clinicians recorded their lowest overall ICC in this area, with Clinician 2 obtaining

an ICC of 0.093 in the POS zone.

Further examination of Clinician 2's segmentations revealed 11 outliers. These outliers were defined as any value exceeding a pixel difference of $4\ \mu\text{m}$ in the POS region. These outliers significantly impacted the ICC, particularly since Clinician 2 performed fewer measurements than Clinician 1. When comparing the measurer performance in a different region, only one outlier was found in Band 2, which is similar in size to the POS zone, underscoring the difficulty in reproducing consistent results for the first region.

We can argue that the overall high average pixel difference shown in Table 5.2 can be attributed to one clinician using a slightly darker marker during the annotations, causing the color to spread by one additional pixel compared to the lighter marker of the other clinician, consistently affecting the measurements without significantly impacting the interclass correlation, given the uniform nature of the deviation.

Apart from Band 2, where Clinician 2's lack of consistency and the small region led to more significant differences, the other two interclass correlation values reported good values, validating these results for future analysis.

Equivalent results have also been obtained manually identifying and segmenting retinal layers in similar topographical locations [88]. This study reported ICC for Band 2 + POS zone in the central region of the fovea with values ranging from 0.5 to 0.95, putting the results presented in the report between the expected. The same study reported the ICC of Band 3 + 4 in the same manner as our study, with ICC values ranging from 0.857 to 0.980. The results obtained in our work reveal that neither clinician could obtain an ICC level over 0.91, showing either better practice from the other clinicians, or the images that were dealt with were easier to analyse compared to the ones in this study, nevertheless, all values reported were above 0.78.

Interestingly, the highest deviation compared to our study comes from the interclass correlation coefficients. The results obtained by the clinicians in the same study achieved almost perfect correlation, with values of 0.972 (0.953; 0.984) and 0.981 (0.968; 0.989) for the same regions mentioned before, observing a big gap when evaluated against Table 5.2. A poor correlation was also not identified in the report by Shahidi et al. [72].

The major differences in this study are the considerably wider zones evaluated, with average thicknesses of $30.2 \pm 13.3\ \mu\text{m}$ and $43.2 \pm 31.6\ \mu\text{m}$ for Band 2 + POS and Band 3 + 4, respectively, compared to the $17.53 \pm 2.80\ \mu\text{m}$, $20.05 \pm 3.95\ \mu\text{m}$ and the $40.50 \pm 4.59\ \mu\text{m}$ for Band 2, POS and Band 3 + 4, respectively, in our study. The difference between the Band 2 zone and the combined Band 2 + POS region may complicate efforts to consistently define clear boundaries. However, the performance was not comparable when evaluating Band 3 + 4. Our research showed an interclass correlation of 0.810 (0.740; 0.862), reporting good reproducibility, but the literature reports a correlation of 0.981 (0.968; 0.989), an almost perfect correlation.

Upon reviewing the results, several key points must be addressed for future comparisons. While manual annotations revealed that Clinician 1 achieved a higher overall ICC than Clinician 2, it is impractical to treat the annotations from Clinician 1 as definitive 'ground truth'. The lower ICC values reflect the inherent difficulty of delineating these bands, therefore comparing the algorithms' results with these

delimitations has its uncertainty. Additionally, we cannot discard that any of the measures from Clinician 2 weren't more precise than the measures from Clinician 1. Since the masks used for accessing the algorithms' performance were the ones from the first measurement of Clinician 1, the results both from the pixel-analysis approach and the deep learning approach are subjective and must be considered with some degree of caution. The pixel analysis and deep learning results rely on the accuracy of a single clinician's measurements, which may not reflect absolute precision. Future studies should aim to validate the algorithms using a broader range of clinician annotations or standardized reference models, thereby reducing the subjectivity and improving the reliability of the segmentation outcomes.

In this process, the difficulties included important factors that should be addressed in future works concerning information and decision support systems in the clinical and management branch of biomedical engineering. A primary difficulty involved acquiring high-quality images from healthy patients, free of artifacts or dystrophies, necessitating using three separate databases. This, combined with the need for specific computing resources for segmentation and the transfer of large files, prolonged the entire process. Moreover, the segmentation platform itself was not well-designed, requiring extra manual labor and thus further extending the time clinicians spent on each task.

With the progression of the research, the lack of availability of clinically diagnosed ABCA4 and PRPH2 patients became noticeable, not only because of the obvious difficulties in detecting and reporting these diseases in their early stages but also because the databases were not all put together in one place, which made the acquisition of such images more difficult.

This obstacle narrowed the project's focus to healthy patients, which facilitated the validation of the model but simultaneously reduced its immediate clinical applicability. While the model's performance on normal data was an important step, the lack of diagnosed ABCA4 and PRPH2 patients left a critical gap in testing the model's ability to detect and analyze disease-specific patterns, limiting the project's full potential, considering the classification network idealized at the beginning of the project.

6.2 Pixel-analysis approach

Examining Table 5.3 and Figure 5.2, it becomes apparent that the results did not show a strong correlation with the expected measurements. However, a noticeable relationship between the regions, bands, and their respective outcomes was still present.

As this approach relies purely on pixel intensity, enhancing the separation between areas could lead to improved performance, as demonstrated in the POS region near the central fovea.

The natural depression of the central fovea allows for greater separation between Bands 2 and 3, facilitating more accurate measurements in the POS region. In contrast, for Bands 2 and 4, the areas further from the central fovea exhibited better differentiation, leading to slightly improved results.

Although these results were not statistically significant, the average pixel difference of around 3 pixels (equivalent to $12\ \mu m$) shows a relatively consistent margin of error.

A simple high-pass filter enhanced high-frequency transitions, such as the boundaries between the hyperreflective bands. However, some of the images, as represented in Figure 4.3, didn't provide a

clear differentiation between bands, or had vanishing bands. This problem would translate into wrongful detections by the algorithm.

Even with the algorithm calibrated to impose region-specific maximum thickness values, the presence of pixel intensity gradient discrepancies—likely caused by inherent OCT image noise and low resolution—led to detection errors, widening the gap between the algorithm’s output and manual annotators’ results.

Another approach that could be considered is the integration of adaptive filters that adjust to local intensity gradients, thereby enhancing layer separation in regions where standard filtering techniques fail. This could be particularly useful in cases where there is significant anatomical variation, such as in patients with retinal diseases.

All the previous works reported in the literature working with pixel-intensity analysis were only able to accurately measure regions with at least 40 μm of thickness, with highly sophisticated filtering, and with higher resolution imaging (depth resolution of 2 μm , compared to the 3.9 μm in our work)[68, 72].

Despite its limitations, this method should not be dismissed for use with higher-resolution OCT images. The combination of low image resolution and reliance on pixel intensity has constrained its accuracy, but applying more advanced filters, as suggested in several studies, such as the one presented by *Ross et al.* [89], could enhance its performance.

A different approach to refining the results is to incorporate each observer’s measurement error as a margin of error for accuracy. Although this method would not change the algorithm’s measurements, it could potentially strengthen the correlation between the values.

A promising direction for future improvement is the development of hybrid methods that combine pixel-intensity analysis with deep learning-based approaches. By leveraging the strengths of both methodologies, such as incorporating information from the pixel-analysis process into a segmentation network, this approach could result in a system that achieves higher segmentation accuracy while maintaining a certain computational efficiency.

This phase of the project proved to be highly time-consuming due to the necessity of manually calibrating the threshold values. Each image exhibited unique brightness and contrast characteristics, even after applying filtering techniques, which significantly complicated the process. When the thresholds were successfully optimized for one image, they often failed to generalize across others, leading to issues in band detection or recognition. Moreover, the previously mentioned pixel intensity discrepancies in certain regions of the images necessitated the imposition of several constraints on the algorithm, such as defining maximum and minimum bandwidth limits. To further mitigate the impact of potential outliers, the algorithm was restricted to only consider values within a specified range around the initial mean intensity.

6.3 Deep-learning approach

The various models tested were chosen according to those suggested in the literature for semantic image segmentation. Concerning the segmentation results, all models performed well across Dice

score, Precision, and Recall metrics. Given that the bands analyzed in this task represent only a small portion of the total image, the Pixel Accuracy metric, as expected, returned near-perfect results. This is because Accuracy accounts for both correctly identified foreground and background pixels, as highlighted in Equation 4.2.

It can be stated that DRUNET exhibited the most reliable performance across the different evaluations, as previous reports on semantic segmentation using DRUNET [67] in the choroidal region conveyed. This network outperformed the other four tested models, achieving the highest Dice scores across the various regions, while demonstrating the smallest average pixel differences in Band 3 + 4 and the POS zone compared to Clinician 1 annotations. However, despite the varying levels of complexity among the networks, their performance remained similar. This may be attributed to the high quality of the data and the relatively simple, repetitive patterns within the dataset, which allowed even less complex architectures, such as the UNET, to capture essential features needed for accurate segmentation, and perform similarly to the best-performing ones. In sum, despite having fewer parameters, simpler models performed comparably to the more complex ones, which helped reduce training times and resource consumption without sacrificing accuracy.

We can observe that the Dice scores for Band 2 and POS were lower than for Band 3 + 4, primarily due to the smaller dimensions of those regions. Consistent pixel-level errors across regions still result in lower Dice scores for smaller areas, as the Dice coefficient (Equation 4.3) is sensitive to mask overlap. Both the algorithm and manual annotations face challenges in accurately defining the boundaries of these smaller regions. Since minor boundary errors affect a greater percentage of smaller areas, their impact is more pronounced. For larger areas, the same errors represent a smaller fraction of the region, leading to a smaller effect on the Dice scores.

The present work yielded results consistent with those seen in similar approaches. The study by *Chiu et al.* [90], which used graph theory and dynamic programming to segment broader regions (Band 2 + POS and Band 3 + 4), showed larger discrepancies in thickness values when compared to a single annotator. In contrast, our method demonstrated greater accuracy and consistency in thickness measurements, producing smaller differences and more reliable segmentations.

The study by *Pekala et al.* [75], which utilized the same type of input images, reported an average pixel difference of 1.02, a result comparable to our findings, despite, again, their focus on larger surface areas. Similarly, other studies have demonstrated results in line with ours, with Dice scores of 0.92 and 0.90 in the region between Band 1 and Band 2, inclusive (see Figure 2.11), and the POS + Band 3 + Band 4 region, reported by *Roy et al.* [73]. The work by *Fang et al.* [77], also achieved a pixel difference of less than 2 units compared to manual annotations for non-exudative AMD patients.

A favorable procedure in our approach was the utilization of standardized images. The 256×256 input format facilitated the algorithm's ability to learn from left and right-sided images concerning the central fovea. This approach expanded the dataset — transforming a single B-scan annotation into 16 distinct input images for the network through data augmentation — and improved computational efficiency. By standardizing the image dimensions, the training process became more streamlined, reducing the complexity of handling varying image sizes and allowing faster processing times while

ensuring consistent input quality across the network.

A train/validation/test split of around 70/20/10 is commonly applied in studies within this domain. Alternative divisions, such as 80/10/10 or 70/15/15, are also used in various machine learning studies [74, 78]. Moreover, techniques like k-fold cross-validation and leave-one-out validation have been explored in related research [76]. Though these methods demand substantially more computational resources, they could prove useful in future models where current approaches may fail to detect retinal pathology patterns.

The image acquisition from several sources, as referred to in subsection 4.1.1, brings considerably different-sized images. Despite having the same acquisition systems, clinicians can output the images in different sizes and formats. The standardization of the different images for the networks required after-acquisition manual labour. This process involved resizing, reformatting, and ensuring specific regions were present in the image, intending to preserve important anatomical features, such as the perifovea, the furthest region analysed.

Although this setup equips the algorithm for practical applications in the "real world," where images of varying sizes from different OCT machines must be recognized for segmentation, it remains crucial to consider that either image size standardization across OCT devices will be necessary in the future, or an automated preprocessing step will need to be implemented to expedite this process.

Another issue arose from the clinicians' use of the delimitation pen. To measure the various thicknesses, a binary mask was generated between the coloured lines in each image. This would give a mask of the full retina, and from there, at each location, the specific thickness would be calculated considering the pixels identified as one. However, we noted that for darker colours, such as the red in Figure 4.7, the colour would spread through more pixels, increasing the masks' size, when compared to a brighter colour such as the yellow in Figure 4.8, creating a uniform deviation of the results.

The limited availability of clinical data and the extremely time-consuming process of manually segmenting each region significantly restricted the network's training potential. Despite the considerable efforts of both clinicians, the dataset size could have been substantially larger, which would likely have reduced segmentation errors, increased DICE scores, and enhanced the generalizability of these dCNNs, if only healthy patients were supplied. However, we believe that the introduction of patients with dystrophies would in turn decrease the overall results.

Chapter 7

Conclusions

7.1 Findings

The significance of automating the segmentation of retinal bands extends beyond distinguishing between the two diseases discussed earlier in this work. It is a critical tool for clinicians, enabling rapid and accurate identification of structural differences and assessing the integrity of retinal layers. Automated segmentation also allows for precise comparison of band thickness across different regions of the fovea, facilitating the early detection of abnormalities. Moreover, it minimizes the variability and subjectivity often associated with manual segmentation, enhancing diagnostic consistency. By reducing the time required for analysis and improving accuracy, automated segmentation could be instrumental in clinical decision-making and long-term monitoring of disease progression. In the context of inherited retinal diseases, it also enables more efficient screening for potential genetic therapies, ensuring that patients receive timely and appropriate intervention.

The results demonstrate that, although reproducibility was achieved, several factors — such as the limited resolution of the OCT images, the small size of the regions under study, the use of different pens for the clinicians' delimitations, and the subjectivity in the definition of the regions, among others — impacted the ICC and interclass correlation coefficients, preventing them from reaching exceptional levels. These limitations likely introduced variability in boundary detection and measurement precision. However, despite these constraints, the ICC results still showed strong repeatability and reproducibility, even when compared to other manual segmentation studies of larger structures.

The fovea detection algorithm was a critical step in effectively identifying central foveal depression in most cases, as it allowed clinicians to pinpoint the optimal regions for accurately measuring retinal thickness. Additionally, this algorithm facilitated the division of the images to introduce in the segmentation networks.

Due to inconsistent agreement on band thickness, clinicians' segmentations cannot be considered the definitive ground truth when evaluating automatic approaches. Variability between clinicians, often resulting from differences in interpretation, annotation tools, or image resolution, introduces subjectivity into the segmentation process, as mentioned before. This variability can influence the outcomes and

make manual segmentations challenging as an absolute standard. However, despite this limitation, clinicians' segmentations remain valuable as a reference point for comparison, serving as a benchmark to assess the performance of automated algorithms.

The pixel-analysis algorithm demonstrated no significant correlation between thickness values and manual annotations, with certain regions displaying large pixel differences. This discrepancy could be attributed to the algorithm's inherent limitations in dealing with noise and low-resolution images. However, despite these inconsistencies, the average pixel difference of approximately 3 pixels suggests a relatively agreeing error margin, indicating that the method maintains some reliability. Given this, the pixel-analysis approach should not be disregarded for future use, particularly with higher-resolution OCT images, where the increased resolution could mitigate some of the algorithm's current limitations. In such cases, the algorithm could potentially provide more accurate segmentation results, especially when supplemented by more advanced filtering techniques or combined with other automated methods.

The dCNNs architectures' results further underscore these models' potential in clinical applications. With an average pixel difference of 0.80 ± 0.19 , the correlation with manual annotations is notably strong, reflecting the capability of these networks to achieve a high degree of accuracy.

Compared to other state-of-the-art projects, the results of the dCNN models align well with expectations, particularly in terms of Dice scores, even when evaluated against models segmenting larger retinal regions.

Furthermore, integrating deep learning models addresses a critical limitation in manual segmentation—clinician variability. By minimizing discrepancies between manual annotations, these models provide a more standardized approach to segmentation, leading to more consistent results across different cases and practitioners. This is particularly relevant in clinical settings, where variability between clinicians can influence diagnosis and treatment decisions.

In addition, these networks' ability to accurately monitor retinal layer thickness makes them especially valuable for detecting patterns linked to specific retinal diseases, such as ABCA4 and PRPH2-associated retinopathies. These patterns are often subtle and can be easily missed in manual segmentation, especially in the early stages of disease progression. Using deep learning models offers a way to detect these early signs more reliably, improving the chances of timely intervention and better outcomes for patients.

The consistent performance of these networks also suggests their applicability in routine clinical practice. By automating the segmentation process, deep learning models can significantly reduce the time required for image analysis, allowing clinicians to focus more on patient care and decision-making. Moreover, as these models continue to evolve and incorporate larger and more diverse datasets, their accuracy and generalizability are likely to improve further.

In conclusion, the dCNN architectures evaluated in this study demonstrate significant potential for clinical integration. Their ability to produce accurate, consistent results, coupled with their capacity to reduce variability and enhance diagnostic precision, makes them a valuable tool in the ongoing effort to improve retinal disease detection and treatment.

7.2 Future Work and possible improvements

In the future, we aim to expand the deep-learning dataset by incorporating images of patients with various retinal diseases, such as ABCA4 and PRPH2 mutations, or AMD. This expansion intends to enhance the network's ability to learn from a more diverse patient population, ultimately improving segmentation accuracy across different conditions. Also, gathering images across diverse patient populations and demographics could further improve the algorithm's generalizability and clinical relevance.

Also, a possible alternative to address the current dataset limitations involves exploring transfer learning or incorporating semi-supervised learning. The first could provide a valuable strategy, where the model is pre-trained on larger, publicly available datasets and fine-tuned for retinal segmentation tasks. Additionally, incorporating semi-supervised learning techniques could enable the network to leverage labeled and unlabelled data, further enhancing performance without requiring extensive manual annotation.

Using the segmentation masks generated for each image, we plan to develop a classification network focused on the three regions of interest. This network would aim to differentiate between ABCA4, PRPH2, and healthy patients, in line with the work of *Heath et al.* [20], which demonstrated that band thickness can be a distinguishing feature between these conditions. A multiclass classification network using this study's three different segmentation masks could classify patients based on the likelihood of having any of the targeted diseases, guiding them toward appropriate genetic testing.

We expect to develop a model that could be useful for clinics to guide their patients, according to the probability reported by the network of possessing one of the phenotypes, for clinical testing, and from there, to perform pharmacological and even gene therapies. Therefore, patients could have personalized treatment according to the evolution of the bands' integrity and dimensions, receiving tailored interventions based on their unique genetic and phenotypic profiles.

Pilot studies in real clinical environments are essential to validate the applicability of these models. Allowing clinicians to utilize these models during routine patient assessments would provide crucial feedback on usability, accuracy, and workflow integration, offering valuable insights from a medical point of view for further refinement. Therefore, the integration of a repository, along with a clear user guide and relevant examples could be shared with hospitals, clinics, and research centers worldwide.

Furthermore, the integration of immersive VR technology, by redesigning conventional desktop SD-OCT image annotation and segmentation interfaces, along with the need to enhance manual annotation tasks and improve the visualization of 2D and 3D OCT data, makes VR an excellent tool for transforming traditional retinal data analysis workstations.

Conventional interfaces to visualize, annotate, classify, and segment OCT images follow desktop-based approaches that are stuck to the Windows-icons-menus-pointer (WIMP) paradigm, which is heavily based on mouse and keyboard input. Despite being a 3D volume, OCT experts deal mostly with 2D data, which demands cognitive load to mentally reconstruct 3D representation from a set of 2D slices and 2D projection maps. Experts routinely analyze 2D slices of the data, but volume rendering representations are also analyzed with care as 3D images provide anatomical context to 2D slice visualization

and also an overview of identifiable anomalies [91].

Therefore, a future direction could involve designing a novel multimodal VR interface to streamline diagnosis time and accelerate manual annotation and segmentation tasks. We envision an interactive system integrating VR with pen-based input, touchscreens, head gaze, hand gestures, and voice control. In this proposed multimodal VR OCT workspace, an expert would be immersed in a virtual environment, seated at an ergonomic desk with a pen or stylus in the dominant hand, while the non-dominant hand interacts with the screen. Commands and actions could be executed using a combination of pen input, head gaze, hand gestures, or voice.

An immersive VR OCT workstation could be developed by accessing specific equipment (e.g., Quest 2, Wacom Cintiq Pro) and the required software development expertise (e.g., Unity 3D, C#). Additionally, this setup could be used to explore the possibility of VR OCT being employed by ophthalmologists to interactively explain complex eye conditions or surgical procedures to patients through storytelling techniques or to enhance the teaching of OCT radiologic anatomy.

Previous works using the pixel analysis approach have observed challenges in accurately measuring small regions. Applying more sophisticated filtering methods during pre-processing and using higher-resolution OCT images could improve segmentation accuracy in these regions, especially when the objective is to differentiate minute variations.

Regarding the deep learning approach, one limitation regards the relatively small dataset. Although the study utilized over 100 images for training and validation, it lacks diversity in terms of other disease types, which restricts the broader applicability and generalizability of the project we intend to achieve in the future. The applicability of data augmentation increases the diversity learned by the algorithm, however more images of patients are necessary.

The second improvement that could be made to this network is to train the algorithm with high-resolution images. These images can reduce noise and other types of artifacts, helping clinicians perform the annotations, since compared to the normal resolution scans, these images possess double the optical axial resolution. Therefore, a better ground truth representation could result in a better performance for the network.

The use of the Dice Loss function is particularly important for the regions under study; however, variations of this loss function, such as Log-Cosh Dice Loss and generalized Wasserstein Dice Loss, can be applied in similar segmentation tasks [92]. These variants have slightly different characteristics and could potentially improve efficiency in this specific segmentation challenge.

This study highlights the promising role of deep learning in advancing retinal OCT image analysis, providing a reliable alternative to manual segmentation of hyperreflective bands. By integrating OCT images from a wider range of retinal diseases, there is potential to further refine diagnostic accuracy, enhancing clinical decision-making and patient care. In the future, segmentation and classification algorithms hold the potential to improve the quality of patient-clinician feedback, particularly for identifying complex and subtle retinal conditions.

Bibliography

- [1] Z. Ma, Q. Xie, P. Xie, F. Fan, X. Gao, et al. Hctnet: A hybrid convnet-transformer network for retinal optical coherence tomography image classification. *Biosensors*, 12(7):542, Jul 2022. doi: <https://doi.org/10.3390/bios12070542>. URL <https://pubmed.ncbi.nlm.nih.gov/35884345/>.
- [2] S. Sotoudeh-Paima, A. Jodeiri, F. Hajizadeh, and H. Soltanian-Zadeh. Multi-scale convolutional neural network for automated amd classification using retinal oct images. *Computers in Biology and Medicine*, 144:105368, May 2022. doi: <https://doi.org/10.1016/j.compbimed.2022.105368>. URL <https://www.sciencedirect.com/science/article/pii/S0010482522001603?via%3Dihub>.
- [3] S. A. Kamran, S. Saha, A. Sabbir, and A. Tavakkoli. Optic-net: A novel convolutional neural network for diagnosis of retinal diseases from optical tomography images. *arXiv (Cornell University)*, Dec 2019. doi: <https://doi.org/10.1109/icmla.2019.00165>. URL <https://ieeexplore.ieee.org/document/8999264>.
- [4] L. Fang, C. Wang, S. Li, H. Rabbani, X. Chen, et al. Attention to lesion: Lesion-aware convolutional neural network for retinal optical coherence tomography image classification. *IEEE Transactions on Medical Imaging*, 38(8):1959–1970, Aug 2019. doi: <https://doi.org/10.1109/tmi.2019.2898414>. URL <https://pubmed.ncbi.nlm.nih.gov/30763240/>.
- [5] O. Akinniyi, M. M. Rahman, H. S. Sandhu, A. El-Baz, and F. Khalifa. Multi-stage classification of retinal oct using multi-scale ensemble deep architecture. *Bioengineering*, 10(7):823, Jul 2023. doi: <https://doi.org/10.3390/bioengineering10070823>. URL <https://www.mdpi.com/2306-5354/10/7/823>.
- [6] K. J. Choi, J. E. Choi, H. C. Roh, J. S. Eun, J. M. Kim, et al. Deep learning models for screening of high myopia using optical coherence tomography. *Scientific Reports*, 11(1), Nov 2021. doi: <https://doi.org/10.1038/s41598-021-00622-x>. URL <https://pubmed.ncbi.nlm.nih.gov/34737335/>.
- [7] D. Kermany, M. H. Goldbaum, W. Cai, C. C. S. Valentim, H. Liang, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 172(5):1122–1131.e9, Feb 2018. doi: <https://doi.org/10.1016/j.cell.2018.02.010>. URL <https://pubmed.ncbi.nlm.nih.gov/29474911/>.
- [8] J.-W. Kim and L. Tran. Retinal disease classification from oct images using deep learning al-

- gorithms. *IEEE.org*, Oct 2021. doi: <https://doi.org/10.1109/cibcb49929.2021.9562919>. URL <https://ieeexplore.ieee.org/document/9562919>.
- [9] S. Diao, J. Su, C. Yang, W. Zhu, D. Xiang, et al. Classification and segmentation of oct images for age-related macular degeneration based on dual guidance networks. *Biomedical Signal Processing and Control*, 84:104810, Jul 2023. doi: <https://doi.org/10.1016/j.bspc.2023.104810>. URL <https://www.sciencedirect.com/science/article/abs/pii/S1746809423002434?via%3Dihub>.
- [10] H. S. Sandhu, A. ElTanboly, A. Shalaby, R. Keynton, S. Schaal, et al. Automated diagnosis and grading of diabetic retinopathy using optical coherence tomography. *Investigative Ophthalmology Visual Science*, 59(7):3155, Jun 2018. doi: <https://doi.org/10.1167/iovs.17-23677>. URL <https://pubmed.ncbi.nlm.nih.gov/30029278/>.
- [11] A. ElTanboly, M. Ismail, A. Shalaby, A. E. Switala, A. El-Baz, et al. A computer-aided diagnostic system for detecting diabetic retinopathy in optical coherence tomography images. *Medical Physics*, 44(3):914–923, Mar 2017. doi: <https://doi.org/10.1002/mp.12071>. URL <https://pubmed.ncbi.nlm.nih.gov/28035657/>.
- [12] M. Ghazal, S. Ali, A. Mahmoud, A. Shalaby, and A. El-Baz. Accurate detection of non-proliferative diabetic retinopathy in optical coherence tomography images using convolutional neural networks. *IEEE Access*, 8:34387–34397, Jan 2020. doi: <https://doi.org/10.1109/access.2020.2974158>. URL <https://ieeexplore.ieee.org/document/9011786>.
- [13] M. Elgafi, A. Sharafeldeen, A. Elnakib, A. Elgarayhi, N. S. Alghamdi, et al. Detection of diabetic retinopathy using extracted 3d features from oct images. *Sensors*, 22(20):7833, Oct 2022. doi: <https://doi.org/10.3390/s22207833>. URL <https://www.mdpi.com/1424-8220/22/20/7833>.
- [14] P. Zang, T. T. Hormel, X. Wang, K. Tsuboi, D. Huang, et al. A diabetic retinopathy classification framework based on deep-learning analysis of oct angiography. *Translational Vision Science Technology*, 11(7):10, Jul 2022. doi: <https://doi.org/10.1167/tvst.11.7.10>. URL <https://tvst.arvojournals.org/article.aspx?articleid=2783476>.
- [15] M. Elsharkawy, A. Sharafeldeen, A. Soliman, F. Khalifa, M. Ghazal, et al. A novel computer-aided diagnostic system for early detection of diabetic retinopathy using 3d-oct higher-order spatial appearance model. *Diagnostics*, 12(2):461, Feb 2022. doi: <https://doi.org/10.3390/diagnostics12020461>. URL <https://www.mdpi.com/2075-4418/12/2/461>.
- [16] N. Cuenca, I. Ortuño-Lizarán, and I. Pinilla. Cellular characterization of oct and outer retinal bands using specific immunohistochemistry markers and clinical implications. *Ophthalmology*, 125(3):407–422, Mar 2018. doi: <https://doi.org/10.1016/j.ophtha.2017.09.016>. URL <https://www.sciencedirect.com/science/article/pii/S0161642017316743>.
- [17] S. R. Sadda, A. P. Schachat, C. P. Wilkinson, D. R. Hinton, P. Wiedemann, et al. *Ryan's retina*. EU Elsevier Health, 2022. URL <https://www.eu.elsevierhealth.com/ryans-retina-9780323722131.html>.

- [18] R. Allikmets. Stargardt disease. *Humana Press eBooks*, page 105–118, Nov 2007. doi: https://doi.org/10.1007/978-1-59745-186-4_5. URL https://link.springer.com/chapter/10.1007/978-1-59745-186-4_5.
- [19] L. Fernández-Caballero, I. Martín-Merida, F. Blanco-Kelly, A. Avila-Fernandez, E. Carreño, et al. Prph2-related retinal dystrophies: Mutational spectrum in 103 families from a spanish cohort. *International Journal of Molecular Sciences*, 25(5):2913, Mar 2024. doi: <https://doi.org/10.3390/ijms25052913>. URL <https://www.mdpi.com/1422-0067/25/5/2913>.
- [20] R. C. Heath, J. Lo, J. A. Thompson, T. M. Lamey, T. L. McLaren, et al. Analysis of the outer retinal bands in abca4 and prph2-associated retinopathy using oct. *Ophthalmology Retina*, May 2023. doi: <https://doi.org/10.1016/j.oret.2023.05.010>. URL <https://pubmed.ncbi.nlm.nih.gov/37209970/>.
- [21] B. D. Kels, A. Grzybowski, and J. M. Grant-Kels. Human ocular anatomy. *Clinics in Dermatology*, 33(2):140–146, Mar 2015. doi: <https://doi.org/10.1016/j.clindermatol.2014.10.006>. URL <https://www.sciencedirect.com/science/article/pii/S0738081X1400234X?via%3Dihub>.
- [22] C. Joyce, P. H. Le, and N. M. Sadiq. Histology, retina, Aug 2023. URL <https://www.ncbi.nlm.nih.gov/books/NBK546692/>.
- [23] C. Mahendra, T. Teng Hern, P. Pusparajah, T. Htar, L. Chuah, et al. Detrimental effects of uvb on retinal pigment epithelial cells and its role in age-related macular degeneration. *Oxidative Medicine and Cellular Longevity*, 2020:1–29, 08 2020. doi: 10.1155/2020/1904178.
- [24] L. Tebbe, M. Kakakhel, M. S. Makia, M. R. Al-Ubaidi, and M. I. Naash. The interplay between peripherin 2 complex formation and degenerative retinal diseases. *Cells*, 9(3):784, Mar 2020. doi: <https://doi.org/10.3390/cells9030784>. URL <https://pubmed.ncbi.nlm.nih.gov/32213850/>.
- [25] T. R. Lewis, M. S. Makia, M. Kakakhel, M. R. Al-Ubaidi, V. Y. Arshavsky, et al. Photoreceptor disc enclosure occurs in the absence of normal peripherin-2/rds oligomerization. *Frontiers in Cellular Neuroscience*, 14, Apr 2020. doi: <https://doi.org/10.3389/fncel.2020.00092>. URL <https://www.frontiersin.org/articles/10.3389/fncel.2020.00092/full>.
- [26] S. Seo and P. Datta. Photoreceptor outer segment as a sink for membrane proteins: hypothesis and implications in retinal ciliopathies. *Human Molecular Genetics*, 26(R1):R75–R82, Apr 2017. doi: <https://doi.org/10.1093/hmg/ddx163>. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5886464/>.
- [27] S. M. Conley and M. I. Naash. Gene therapy for prph2-associated ocular disease: Challenges and prospects. *Cold Spring Harbor Perspectives in Medicine*, 4(11):a017376, Aug 2014. doi: <https://doi.org/10.1101/cshperspect.a017376>. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4208711/>.
- [28] T. R. Lewis, M. S. Makia, M. Kakakhel, M. R. Al-Ubaidi, V. Y. Arshavsky, et al. Photoreceptor disc enclosure occurs in the absence of normal peripherin-2/rds oligomerization. *Frontiers in Cellular*

- Neuroscience*, 14, Apr 2020. doi: <https://doi.org/10.3389/fncel.2020.00092>. URL <https://www.frontiersin.org/articles/10.3389/fncel.2020.00092/full>.
- [29] H. Jeffery, J. Thompson, J. Lo, E. Chelva, S. Armstrong, et al. Retinal dystrophies associated with peripherin-2: Genetic spectrum and novel clinical observations in 241 patients. *Investigative ophthalmology visual science*, 65(5), 2024. doi: <https://doi.org/10.1167/iov.65.5.22>. URL <https://pubmed.ncbi.nlm.nih.gov/38743414/>.
- [30] F. P. Cremers, W. Lee, R. W. Collin, and R. Allikmets. Clinical spectrum, genetic complexity and therapeutic approaches for retinal disease caused by *abca4* mutations. *Progress in Retinal and Eye Research*, 79:100861, Nov 2020. doi: <https://doi.org/10.1016/j.preteyeres.2020.100861>. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7544654/>.
- [31] F. Quazi, S. Lenevich, and R. S. Molday. *Abca4* is an n-retinylidene-phosphatidylethanolamine and phosphatidylethanolamine importer. *Nature Communications*, 3(1), Jun 2012. doi: <https://doi.org/10.1038/ncomms1927>. URL <https://pubmed.ncbi.nlm.nih.gov/22735453/>.
- [32] S. Al-Khuzaei, S. Broadgate, C. Foster, M. Shah, J. Yu, et al. An overview of the genetics of *abca4* retinopathies, an evolving story. *Genes*, 12(8):1241, Aug 2021. doi: <https://doi.org/10.3390/genes12081241>. URL <https://www.mdpi.com/2073-4425/12/8/1241>.
- [33] M. Pfau, W. M. Zein, L. A. Huryn, C. A. Cukras, B. G. Jeffrey, et al. Genotype–phenotype association in *abca4*-associated retinopathy. *Advances in Experimental Medicine and Biology*, page 289–295, Jan 2023. doi: https://doi.org/10.1007/978-3-031-27681-1_42. URL <https://pubmed.ncbi.nlm.nih.gov/37440047/>.
- [34] J. R. Sparrow, E. M. Gregory-Roberts, K. Yamamoto, A. Blonska, S. K. Ghosh, et al. The bis-retinoids of retinal pigment epithelium. *Progress in Retinal and Eye Research*, 31(2):121–135, Mar 2012. doi: <https://doi.org/10.1016/j.preteyeres.2011.12.001>. URL <https://pubmed.ncbi.nlm.nih.gov/22209824/>.
- [35] R. Simó, M. Villarroel, L. Corraliza, C. Hernández, and M. García-Ramírez. The retinal pigment epithelium: Something more than a constituent of the blood-retinal barrier—implications for the pathogenesis of diabetic retinopathy. *Journal of Biomedicine and Biotechnology*, 2010:1–15, Jan 2010. doi: <https://doi.org/10.1155/2010/190724>. URL <https://www.hindawi.com/journals/bmri/2010/190724/>.
- [36] H. J. Kim and J. R. Sparrow. Bisretinoid phospholipid and vitamin a aldehyde: shining a light. *Journal of Lipid Research*, 62:100042–100042, Jan 2021. doi: <https://doi.org/10.1194/jlr.tr120000742>. URL <https://www.sciencedirect.com/science/article/pii/S0022272521000225>.
- [37] S. Asanad and R. Karanjia. Full-field electroretinogram, Aug 2023. URL <https://www.ncbi.nlm.nih.gov/books/NBK557483/>.

- [38] S. Aumann, S. Donner, J. Fischer, and F. Müller. Optical coherence tomography (oct): Principle and technical realization. *Springer eBooks*, page 59–85, Jan 2019. doi: https://doi.org/10.1007/978-3-030-16638-0_3. URL <https://www.ncbi.nlm.nih.gov/books/NBK554044/>.
- [39] A. Malm. Oct velocimetry and x-ray scattering rheology of complex fluids, Oct 2015. URL https://www.researchgate.net/publication/296695687_OCT_Velocimetry_and_X-ray_Scattering_Rheology_of_Complex_Fluids.
- [40] S. R. Chinn, E. A. Swanson, and J. G. Fujimoto. Optical coherence tomography using a frequency-tunable optical source. *Optics Letters*, 22(5):340–342, Mar 1997. doi: <https://doi.org/10.1364/OL.22.000340>. URL <https://opg.optica.org/ol/abstract.cfm?uri=ol-22-5-340>.
- [41] A. Gunalan and L. S. Mattos. Towards oct-guided endoscopic laser surgery—a review. *Diagnostics*, 13(4):677, Feb 2023. doi: <https://doi.org/10.3390/diagnostics13040677>. URL https://www.researchgate.net/publication/368491910_Towards_OCT-Guided_Endoscopic_Laser_Surgery-A_Review.
- [42] Zeiss. ZEISS PLEX Elite 9000 Swept-Source OCT Angiography, 2020. URL <https://www.zeiss.com/meditec/en/products/optical-coherence-tomography-devices/plex-elite-9000-swept-source-oct.html#specifications>. Last accessed on 15 May 2024.
- [43] H. E. Inc., 2024. URL <https://business-lounge.heidelbergengineering.com/us/en/products/spectralis/spectralis/>.
- [44] M. Nuno Filipe Aguiar Silva, April 2023. URL https://eyewiki.aao.org/Stargardt_Disease/Fundus_Flavimaculatus.
- [45] M. Nour, S. J. Fliesler, and M. I. Naash. Genetic supplementation of rds alleviates a loss-of-function phenotype in c214s model of retinitis pigmentosa. *Advances in Experimental Medicine and Biology*, page 129–138, Jan 2008. doi: https://doi.org/10.1007/978-0-387-74904-4_14. URL https://link.springer.com/chapter/10.1007/978-0-387-74904-4_14.
- [46] S. M. Conley, M. Nour, S. J. Fliesler, and M. I. Naash. Late-onset cone photoreceptor degeneration induced by r172w mutation in rds and partial rescue by gene supplementation. *Investigative Ophthalmology Visual Science*, 48(12):5397, Dec 2007. doi: <https://doi.org/10.1167/iovs.07-0663>. URL <https://pubmed.ncbi.nlm.nih.gov/18055786/>.
- [47] F. C. Schlichtenbrede, L. da Cruz, C. Stephens, A. J. Smith, A. Georgiadis, et al. Long-term evaluation of retinal function in prph2rd2/rd2 mice following aav-mediated gene replacement therapy. *The Journal of Gene Medicine*, 5(9):757–764, May 2003. doi: <https://doi.org/10.1002/jgm.401>. URL <https://pubmed.ncbi.nlm.nih.gov/12950066/>.
- [48] H. Petrs-Silva, D. Yasumura, M. T. Matthes, M. M. LaVail, A. S. Lewin, et al. Suppression of rds expression by sirna and gene replacement strategies for gene therapy using raav vector. *Advances in Experimental Medicine and Biology*, page 215–223, Nov 2011. doi: https://doi.org/10.1007/978-1-4419-9888-8_14.

//doi.org/10.1007/978-1-4614-0631-0_29. URL https://link.springer.com/chapter/10.1007/978-1-4614-0631-0_29.

- [49] M. Miyazaki, Y. Ikeda, Y. Yonemitsu, Y. Goto, R. I. Kohno, et al. Synergistic neuroprotective effect via simian lentiviral vector-mediated simultaneous gene transfer of human pigment epithelium-derived factor and human fibroblast growth factor-2 in rodent models of retinitis pigmentosa. *The Journal of Gene Medicine*, 10(12):1273–1281, Oct 2008. doi: <https://doi.org/10.1002/jgm.1257>. URL <https://pubmed.ncbi.nlm.nih.gov/18837062/>.
- [50] A. L. Askou, J.-A. Pournaras, M. Pihlmann, J. D. Svalgaard, Y. Arsenijević, et al. Reduction of choroidal neovascularization in mice by adeno-associated virus-delivered anti-vascular endothelial growth factor short hairpin rna. *The Journal of Gene Medicine*, 14(11):632–641, Nov 2012. doi: <https://doi.org/10.1002/jgm.2678>. URL <https://pubmed.ncbi.nlm.nih.gov/23080553/>.
- [51] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation, 2015. URL <https://arxiv.org/abs/1505.04597>.
- [52] Z. Zhou, M. Rahman, N. Tajbakhsh, and J. Liang. Unet++: A nested u-net architecture for medical image segmentation, 2018. URL <https://arxiv.org/abs/1807.10165v1>.
- [53] M. P. Priyadarsini, S. S. A. Jabeena, G. Rajini, G. Subramanian, and E. B. Clinton S. Retinal vessel segmentation using unet++. In *2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN)*, pages 1–5, 2023. doi: 10.1109/ViTECoN58111.2023.10157589.
- [54] K.-W. Huang, Y. Yang, Z.-H. Huang, Y. Liu, and S.-H. Lee. Retinal vascular image segmentation using improved unet based on residual module. *Bioengineering*, 10(6):722, Jun 2023. doi: <https://doi.org/10.3390/bioengineering10060722>. URL <https://www.mdpi.com/2306-5354/10/6/722>.
- [55] K. Ren, L. Chang, M. Wan, G. Gu, and Q. Chen. An improved u-net based retinal vessel image segmentation method. *Heliyon*, 8(10):e11187, Oct 2022. doi: <https://doi.org/10.1016/j.heliyon.2022.e11187>. URL <https://www.sciencedirect.com/science/article/pii/S2405844022024756>.
- [56] S. K. Devalla, P. K. Renukanand, B. K. Sreedhar, S. Perera, J.-M. Mari, et al. Drunet: A dilated-residual u-net deep learning network to digitally stain optic nerve head tissues in optical coherence tomography images, 2018. URL <https://arxiv.org/abs/1803.00232>.
- [57] A. Myronenko. 3d mri brain tumor segmentation using autoencoder regularization. In A. Crimi, S. Bakas, H. Kuijff, F. Keyvan, M. Reyes, and T. van Walsum, editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 311–320, Cham, 2019. Springer International Publishing. ISBN 978-3-030-11726-9.
- [58] M. M. R. Siddique, D. Yang, Y. He, D. Xu, and A. Myronenko. Automated ischemic stroke lesion segmentation from 3d mri, 2022. URL <https://arxiv.org/abs/2209.09546>.

- [59] K. Hu, B. Shen, Y. Zhang, C. Cao, F. Xiao, et al. Automatic segmentation of retinal layer boundaries in oct images using multiscale convolutional neural network and graph search. *Neurocomputing*, 365:302–313, Nov 2019. doi: <https://doi.org/10.1016/j.neucom.2019.07.079>. URL <https://www.sciencedirect.com/science/article/abs/pii/S0925231219310860?via%3Dihub>.
- [60] J. Alves, Anselmo, A. Silva, D. Sousa, G. Bráz, et al. Automatic segmentation of retinal layers in oct images with intermediate age-related macular degeneration using u-net and dexined. *PLOS ONE*, 16(5):e0251591, May 2021. doi: <https://doi.org/10.1371/journal.pone.0251591>. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0251591>.
- [61] Q. Li, S. Li, Z. He, H. Guan, R. Chen, et al. Deepretina: Layer segmentation of retina in oct images using deep learning. *Translational Vision Science Technology*, 9(2):61, Dec 2020. doi: <https://doi.org/10.1167/tvst.9.2.61>. URL <https://tvst.arvojournals.org/article.aspx?articleid=2772051>.
- [62] S. Borkovkina, A. Camino, W. Janponsri, M. V. Sarunic, and Y. Jian. Real-time retinal layer segmentation of oct volumes with gpu accelerated inferencing using a compressed, low-latency neural network. *Biomedical Optics Express*, 11(7):3968, Jun 2020. doi: <https://doi.org/10.1364/boe.395279>. URL <https://opg.optica.org/boe/fulltext.cfm?uri=boe-11-7-3968&id=432926>.
- [63] Y. Yeganeh, A. Farshad, G. Guevercin, A. Abu-zer, R. Xiao, et al. Scope: Structural continuity preservation for medical image segmentation, 2023. URL <https://arxiv.org/abs/2304.14572>.
- [64] M. Chen, J. Wang, I. Oguz, B. L. VanderBeek, and J. C. Gee. Automated segmentation of the choroid in edi-oct images with retinal pathology using convolution neural networks. *Lecture Notes in Computer Science*, page 177–184, Jan 2017. doi: https://doi.org/10.1007/978-3-319-67561-9_20. URL <https://pubmed.ncbi.nlm.nih.gov/29757338/>.
- [65] F. Bai, M. J. Marques, and S. J. Gibson. Cystoid macular edema segmentation of optical coherence tomography images using fully convolutional neural networks and fully connected crfs, 2017. URL <https://arxiv.org/abs/1709.05324>.
- [66] Y. Xu, K. Yan, J. M. Kim, X. Wang, C. Li, et al. Dual-stage deep learning framework for pigment epithelium detachment segmentation in polypoidal choroidal vasculopathy. *Biomedical Optics Express*, 8(9):4061, Aug 2017. doi: <https://doi.org/10.1364/boe.8.004061>. URL <https://opg.optica.org/boe/fulltext.cfm?uri=boe-8-9-4061&id=370563>.
- [67] J. Kugelman, D. Alonso-Caneiro, S. A. Read, J. Hamwood, S. J. Vincent, et al. Automatic choroidal segmentation in oct images using supervised deep learning methods. *Scientific Reports*, 9(1), Sep 2019. doi: <https://doi.org/10.1038/s41598-019-49816-4>. URL <https://www.nature.com/articles/s41598-019-49816-4>.
- [68] A. M. Bagci, M. Shahidi, R. Ansari, M. Blair, N. P. Blair, et al. Thickness profiles of retinal layers by optical coherence tomography image segmentation. *American Journal of Ophthalmology*, 146(5):

- 679–687.e1, Nov 2008. doi: <https://doi.org/10.1016/j.ajo.2008.06.010>. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2590782/>.
- [69] D. E. Maidana, S. Notomi, T. Ueta, T. Zhou, D. Joseph, et al. ThicknessTool: automated imagej retinal layer thickness and profile in digital images. *Scientific Reports*, 10(1), Oct 2020. doi: <https://doi.org/10.1038/s41598-020-75501-y>. URL <https://www.nature.com/articles/s41598-020-75501-y#Sec17>.
- [70] I. Arganda-Carreras, R. Fernández-González, A. Muñoz-Barrutia, and C. Ortiz-De-Solorzano. 3d reconstruction of histological sections: Application to mammary gland tissue. *Microscopy Research and Technique*, 73(11):1019–1029, Mar 2010. doi: <https://doi.org/10.1002/jemt.20829>. URL <https://pubmed.ncbi.nlm.nih.gov/20232465/>.
- [71] K. Tarassoly, A. Miraftehi, A. Sedaghat, and M. M. Parvaresh. Analysis of the outer retinal thickness pixel maps for the screening of hydroxychloroquine retinopathy. *International Ophthalmology*, 43(5):1737–1743, Nov 2022. doi: <https://doi.org/10.1007/s10792-022-02573-2>. URL <https://link.springer.com/article/10.1007/s10792-022-02573-2>.
- [72] M. Shahidi, Z. Wang, and R. Zelkha. Quantitative thickness measurement of retinal layers imaged by optical coherence tomography. *American Journal of Ophthalmology*, 139(6):1056–1061, Jun 2005. doi: <https://doi.org/10.1016/j.ajo.2005.01.012>. URL [https://www.ajo.com/article/S0002-9394\(05\)00020-6/abstract](https://www.ajo.com/article/S0002-9394(05)00020-6/abstract).
- [73] A. G. Roy, S. Conjeti, K. Karri, D. Sheet, A. Katouzian, et al. RelayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks. *Biomedical Optics Express*, 8(8):3627, Jul 2017. doi: <https://doi.org/10.1364/boe.8.003627>. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5560830/>.
- [74] F. G. Venhuizen, B. van Ginneken, B. Liefers, M. J.J.P, S. Fauser, et al. Robust total retina thickness segmentation in optical coherence tomography images using convolutional neural networks. *Biomedical Optics Express*, 8(7):3292, Jun 2017. doi: <https://doi.org/10.1364/boe.8.003292>. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5508829/>.
- [75] M. Pekala, N. Joshi, T. A. Liu, N. Bressler, D. C. DeBuc, et al. Deep learning based retinal OCT segmentation. *Computers in Biology and Medicine*, 114:103445, Nov 2019. doi: <https://doi.org/10.1016/j.compbiomed.2019.103445>. URL <https://www.sciencedirect.com/science/article/pii/S0010482519303221?via%3Dihub>.
- [76] Y. He, A. Carass, Y. Liu, B. M. Jedynek, S. D. Solomon, et al. Deep learning based topology guaranteed surface and mme segmentation of multiple sclerosis subjects from retinal OCT. *Biomedical Optics Express*, 10(10):5042, Sep 2019. doi: <https://doi.org/10.1364/boe.10.005042>. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6788619/>.
- [77] L. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. Li, et al. Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and

- graph search. *Biomedical Optics Express*, 8(5):2732, Apr 2017. doi: <https://doi.org/10.1364/boe.8.002732>.
- [78] J. Kugelman, D. Alonso-Caneiro, Y. Chen, S. Arunachalam, D. Huang, et al. Retinal boundary segmentation in stargardt disease optical coherence tomography images using automated deep learning. *Translational Vision Science Technology*, 9(11):12, Oct 2020. doi: <https://doi.org/10.1167/tvst.9.11.12>. URL <https://tvst.arvojournals.org/article.aspx?articleid=2770902>.
- [79] C. A. Bobak, P. J. Barr, and A. J. O'Malley. Estimation of an inter-rater intra-class correlation coefficient that overcomes common assumption violations in the assessment of health measurement scales. *BMC Medical Research Methodology*, 18(1), September 2018. doi: [10.1186/s12874-018-0550-6](https://doi.org/10.1186/s12874-018-0550-6). URL <https://bmcmmedresmethodol.biomedcentral.com/articles/10.1186/s12874-018-0550-6>.
- [80] P. E. Shrout and J. L. Fleiss. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86:420–428, 1979. URL <https://psycnet.apa.org/doiLanding?doi=10.1037/2F0033-2909.86.2.420>.
- [81] A. Savitzky and M. J. E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8):1627–1639, 1964. doi: [10.1021/ac60214a047](https://doi.org/10.1021/ac60214a047). URL <https://doi.org/10.1021/ac60214a047>.
- [82] A. Li, J. P. Winebrake, and K. D. Kovacs. Facilitating deep learning through preprocessing of optical coherence tomography images. *BMC Ophthalmology*, 23(1), Apr 2023. doi: <https://doi.org/10.1186/s12886-023-02916-2>. URL <https://bmcophthalmol.biomedcentral.com/articles/10.1186/s12886-023-02916-2>.
- [83] A. Maier. Lecture notes in deep learning: Segmentation and object detection - part 1 - pattern recognition lab, Aug 2020. URL <https://lme.tf.fau.de/lecture-notes/lecture-notes-dl/lecture-notes-in-deep-learning-segmentation-and-object-detection-part-1/>.
- [84] L. Yang, R. P. Ghosh, J. M. Franklin, C. You, et al. Nuset: A deep learning tool for reliably separating and analyzing crowded cells. *bioRxiv (Cold Spring Harbor Laboratory)*, Aug 2019. doi: <https://doi.org/10.1101/749754>. URL <https://www.biorxiv.org/content/10.1101/749754v1.full>.
- [85] N. Huynh. Understanding evaluation metrics in medical image segmentation, Mar 2023. URL https://medium.com/@nghi.huynh_37300/understanding-evaluation-metrics-in-medical-image-segmentation-d289a373a3f.
- [86] Invicro, Jul 2019. URL <https://invicro.com/case-studies/rhesus-teeth-segmentation-by-use-of-multi-atlas-library/>.
- [87] T. K. Koo and M. Y. Li. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2):155–163, Jun 2016. doi: <https://doi.org/10.1016/j.jcm.2016.06.001>.

//doi.org/10.1016/j.jcm.2016.02.012. URL <https://www.sciencedirect.com/science/article/abs/pii/S1556370716000158>.

- [88] P. Camacho, M. Dutra-Medeiros, L. Salgueiro, S. Sadio, and P. C. Rosa. Manual segmentation of 12 layers of the retina and choroid through sd-oct in intermediate amd: Repeatability and reproducibility. *Journal of Ophthalmic and Vision Research*, Jul 2021. doi: <https://doi.org/10.18502/jovr.v16i3.9435>. URL <https://pubmed.ncbi.nlm.nih.gov/34394867/>.
- [89] D. Ross, M. Clark, P. Godara, et al. Refmob: A reflectivity feature model-based automated method for measuring four outer retinal hyperreflective bands in optical coherence tomography. *Investigative Ophthalmology Visual Science*, 2015. doi: 10.1167/iovs.14-15256. URL https://www.researchgate.net/publication/279731574_RefMoB_a_Reflectivity_Feature_Model-Mased_Automated_Method_for_Measuring_Four_Outer_Retinal_Hyperreflective_Bands_in_Optical_Coherence_Tomography.
- [90] S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, et al. Automatic segmentation of seven retinal layers in sdopt images congruent with expert manual segmentation. *Optics Express*, 18(18):19413–19419, August 2010. doi: 10.1364/oe.18.019413. URL <https://opg.optica.org/oe/fulltext.cfm?uri=oe-18-18-19413&id=205410>.
- [91] P. M. Maloca, S. Feu-Basilio, J. Schottenhamml, P. Valmaggia, H. P. N. Scholl, et al. Reference database of total retinal vessel surface area derived from volume-rendered optical coherence tomography angiography. *Scientific Reports*, 12(1), Mar 2022. doi: <https://doi.org/10.1038/s41598-022-07439-2>. URL <https://www.nature.com/articles/s41598-022-07439-2#citeas>.
- [92] R. Azad, M. Heidary, K. Yilmaz, M. Hüttemann, S. Karimijafarbigloo, et al. Loss functions in the era of semantic segmentation: A survey and outlook, Dec 2023. URL <https://arxiv.org/html/2312.05391v1>.

Appendix

Table 1: Intraclass correlation coefficients and average differences for various regions and clinicians.

Location	Clinician	Region	ICC	95% Confidence Interval		Avg Diff (Pixels)
				Lower Bound	Upper Bound	
Center fovea	1	Band 2	0.764	0.629	0.855	1.227
		Band 3+4	0.848	0.745	0.912	1.965
		POS	0.864	0.778	0.918	1.602
	2	Band 2	0.698	0.417	0.857	1.854
		Band 3+4	0.776	0.549	0.897	3.243
		POS	0.749	0.503	0.883	2.130
500 μm	1	Band 2	0.730	0.630	0.806	1.361
		Band 3+4	0.715	0.601	0.800	1.992
		POS	0.746	0.652	0.817	1.380
	2	Band 2	0.598	0.379	0.753	1.544
		Band 3+4	0.722	0.523	0.841	3.405
		POS	0.093	-0.130	0.330	2.711
750 μm	1	Band 2	0.697	0.588	0.781	1.551
		Band 3+4	0.852	0.787	0.898	1.632
		POS	0.800	0.723	0.857	1.316
	2	Band 2	0.539	0.305	0.712	1.411
		Band 3+4	0.860	0.719	0.927	2.897
		POS	0.316	0.052	0.543	2.074
1000 μm	1	Band 2	0.732	0.633	0.807	1.345
		Band 3+4	0.837	0.766	0.887	1.658
		POS	0.730	0.630	0.806	1.477
	2	Band 2	0.539	0.301	0.714	1.411
		Band 3+4	0.870	0.748	0.931	2.917
		POS	0.577	0.352	0.739	1.756

Avg Diff = Average Difference.

