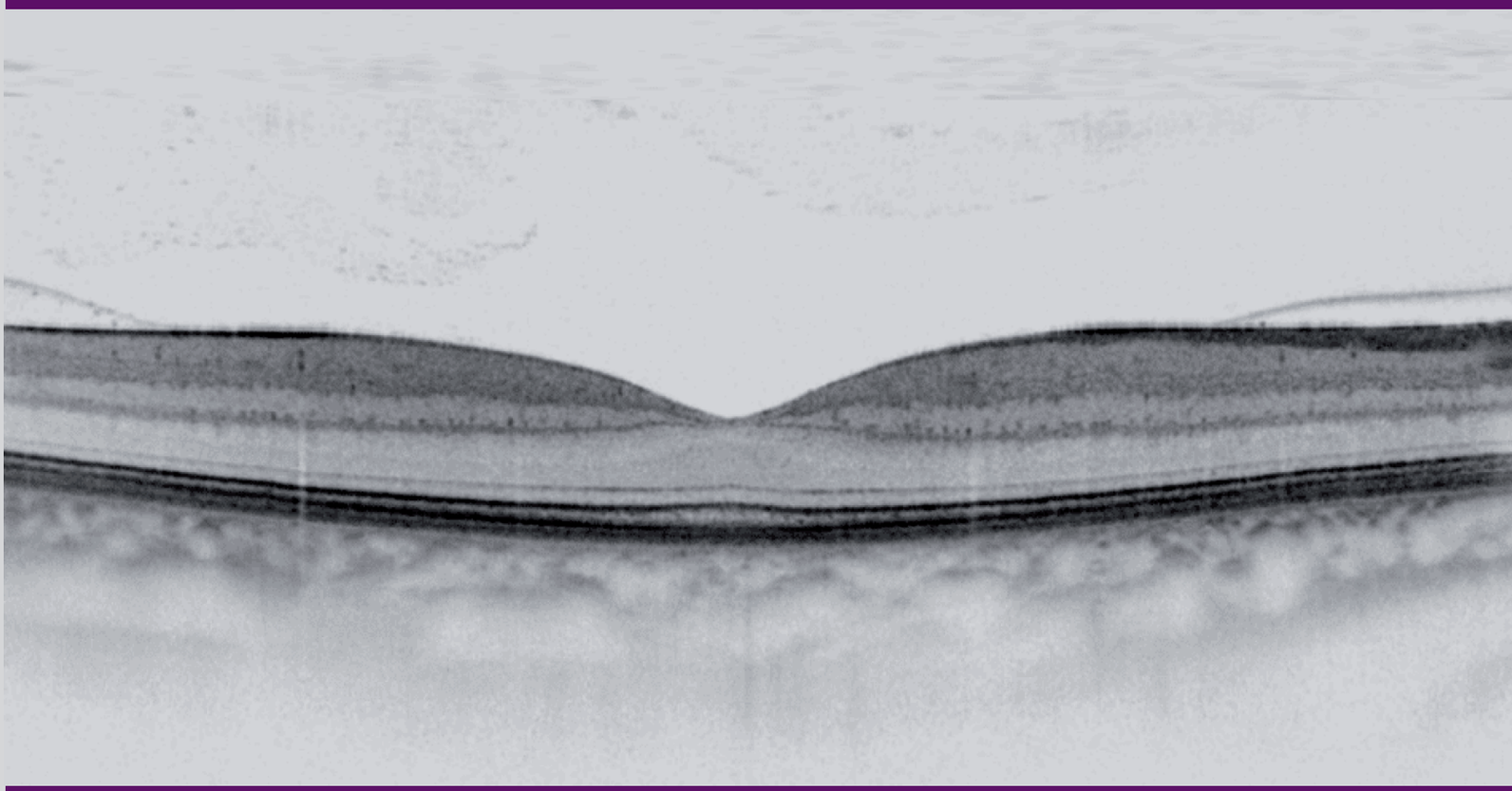
Mondragon Unibertsitatea

DOCTORAL THESIS

ADVANCING RETINAL OCT IMAGE ANALYSIS AS A BIOMARKER FOR PARKINSON'S DISEASE



DAVID ROMERO BASCONES | Arrasate-Mondragón, 2023

DAVID ROMERO BASCONES | Advancing retinal OCT image analysis as a biomarker for Parkinson's disease

Advancing retinal OCT image analysis as a biomarker for Parkinson's disease

David Romero Bascones

Supervisors:

Unai Ayala Fernández

Maitane Barrenetxea Carrasco

Iñigo Gabilondo Cuellar

A thesis submitted for the degree of
Doctoral Program in Engineering

Department of Biomedical Engineering
Mondragon Unibertsitatea
September 12, 2023

We need less research,
better research,
and research done for the right reasons.

DOUG ALTMAN

Agradecimientos

En primer lugar, me gustaría agradecer a Mondragon Goi Eskola Politeknikoa y al Grupo de Investigación de Teoría de la Señal y Comunicaciones el haberme brindado la oportunidad de realizar esta tesis doctoral. Han sido cuatro años en los que he podido desarrollarme como investigador y llevar a cabo mi trabajo con una gran libertad.

Asimismo, me gustaría dar las gracias de forma especial a mis directores por su inestimable contribución a este proyecto y por permitirme explorar ideas más allá de la tesis. A Unai Ayala, agradezco el haber confiado en mí para esta tesis y haberme guiado a lo largo de su desarrollo desde la cercanía. Sin duda, su disposición para poner a prueba mis ideas ha contribuido notablemente a mejorar este trabajo. A Maitane Barrenetxea, le agradezco sus siempre constructivas críticas, su implicación y su excepcional empatía, elementos que han contribuido de manera significativa a enriquecer esta investigación. A Iñigo Gabilondo, le estoy sumamente agradecido por haber concebido esta línea de investigación y por motivarme a formular mejores preguntas y buscar respuestas más completas. Su papel ha sido fundamental para lograr el difícil enlace entre lo técnico y lo clínico en este trabajo.

No quisiera olvidarme de muchas otras personas a las que debo agradecer una gran parte de este trabajo. A Asier Erramuzpe, el abrirme la puerta al mundo open-source y enseñarme el valor del desarrollo de software científico. A Ane Murueta, el haber sentado las bases de esta tesis y haber guiado parte de la misma. A Sara Teijeira, la paciencia y el buen ánimo mostrados las infinitas veces en las que he necesitado ayuda con las imágenes.

Durante este proyecto he tenido el privilegio de trabajar junto a grupos excepcionales a los que también debo dar las gracias. Al equipo de Indiana University Bloomington y DIPY: Eleftherios Garyfallidis, Serge Koudoro, Shreyas Fadnavis y Bramsh Chandio, por seleccionarme para formar parte de su proyecto y enseñarme cómo la vocación de unos pocos puede cambiar un campo de investigación para beneficio de muchos. Al equipo de Moorfields Eye Hospital: Siegfried Wagner, Robbert Struyven, Dominic Williamson, Pearse Keane, y Axel Petzold, por integrarme en su equipo como a uno más y permitirme rodearme de gente verdaderamente brillante.

Mi más profundo agradecimiento a Daniel Lakens, Ben Fulcher y Richard Feynman, cuyo pensamiento y compromiso con la excelencia han moldeado mi forma de ver la ciencia, la estadística y el mundo.

Mucho más allá de lo académico, he de dar las gracias a aquellos que me han acompañado más de cerca durante este proceso. Aita, ama y Andrea, quiero daros las gracias por haberme apoyado desde un comienzo y hacer de nuestra familia un refugio inmejorable. Sin vuestro apoyo jamás hubiera podido abordar este proyecto. A ti, Maitane, gracias por ser parte del azar que desencadenó esta tesis y haber compartido tanto los éxitos como las decepciones sin perder la fe ni el optimismo. Has sido, eres y siempre serás la mejor compañera de equipo que hubiera podido tener.

Statement of originality

I hereby declare that the research recorded in this thesis and the thesis itself were developed by myself at the Department of Biomedical Engineering, Faculty of Engineering (MU-ENG), Mondragon Unibertsitatea.

Arrasate-Mondragón, September 2023

David Romero Bascones

Abstract

The number of people worldwide affected by Parkinson's disease (PD) is increasing every year. This is a worrying trend, which requires new biomarkers to better diagnose and monitor the disease. Research to date has demonstrated that, in addition to brain neurodegeneration, there is a dysfunction of the retina in PD patients.

Promisingly, several studies have shown that retinal changes in PD can be detected using optical coherence tomography (OCT) imaging. However, there is no conclusive agreement on the potential of OCT as a reliable biomarker for PD. As a key limitation, most research has focused on a small set of features and standard OCT image analysis. Applying more advanced methods could help identify the specific aspects of the retina affected in PD and potentially uncover new biomarkers.

Given this context, the present body of work has two main objectives: 1) to improve and advance existing OCT image analysis techniques and 2) to apply the developed methods to be able to study the retina of PD patients in more detail.

First, novel methods were developed for OCT image alignment, quality control, and feature extraction. The developed methods were integrated into an open-source toolbox called RETIMAT, which forms the backbone of this work and is freely available to the community.

Subsequently, data from healthy controls was utilized to evaluate the impact of age and sex on retinal morphology and create a normative dataset. The potential of OCT features for the diagnosis, severity assessment, and monitoring of PD was then evaluated, employing both conventional and novel features. Data for this purpose was obtained from two of the largest longitudinal cohorts to date.

Our findings reveal that the explored OCT features contain information related to cognitive and motor impairment. Furthermore, the evidence suggests a differential evolution of the retina of PD patients over time. However, accurate diagnosis, severity assessment, and monitoring at an individual level using OCT features does not appear feasible. In can thus be concluded that OCT may be a better tool for understanding general aspects of PD than for guiding clinical decisions on individual patients.

Resumen

El número de personas en todo el mundo afectadas por la enfermedad de Parkinson (EP) aumenta cada año. Ante esta tendencia, es vital desarrollar nuevos biomarcadores para diagnosticar y monitorizar mejor la enfermedad. Gracias a investigaciones previas se sabe que, además de la neurodegeneración cerebral, existe una disfunción de la retina en los pacientes con EP.

Prometedoramente, varios estudios han mostrado que los cambios retinianos en la EP pueden detectarse utilizando imágenes de tomografía de coherencia óptica (OCT, por sus siglas en inglés). Sin embargo, no existe un consenso en cuanto al potencial de la OCT como un biomarcador robusto. Como limitación clave, gran parte de las investigaciones se han centrado en un conjunto pequeño de características y en el análisis estándar de imágenes de OCT. Así, aplicar métodos más avanzados podría ayudar a identificar los aspectos específicos de la retina afectados en la EP y potencialmente descubrir nuevos biomarcadores.

Dado este contexto, el presente trabajo tiene dos objetivos principales: 1) extender las técnicas existentes de análisis de imágenes de OCT y 2) aplicar los métodos desarrollados para estudiar la retina de los pacientes con EP con mayor detalle.

En primer lugar, se han desarrollado nuevos métodos para la alineación de imágenes de OCT, el control de calidad y la extracción de características. Los métodos desarrollados se han integrado en un software de código abierto llamado RETIMAT, que constituye la columna vertebral de este trabajo y está disponible de forma libre para la comunidad.

Posteriormente, se han utilizado datos de controles sanos para evaluar el impacto de la edad y el sexo en la morfología retiniana y establecer valores normativos. Finalmente, se ha evaluado el potencial de las características de OCT para el diagnóstico, la evaluación de la severidad y el monitoreo de la enfermedad, empleando tanto características convencionales como novedosas. Para este propósito se han empleado dos de las cohortes longitudinales de pacientes más grandes hasta la fecha.

Los hallazgos revelan que las características de OCT exploradas contienen información relacionada con el deterioro cognitivo y motor. Además, la evidencia sugiere una evolución diferencial de la retina de los pacientes a lo largo del tiempo. Sin embargo, el diagnóstico, la evaluación de la gravedad y el monitoreo preciso a nivel individual utilizando las características de OCT no parecen ser factibles. Por lo tanto, se puede concluir que la OCT puede ser una mejor herramienta para comprender aspectos generales de la enfermedad que para guiar decisiones clínicas sobre pacientes individuales.

Laburpena

Munduan, Parkinson gaixotasuna duten pertsonen kopurua handitzen ari da urtero. Joera horren aurrean, premiazkoa da biomarkatzaile berriak garatzea gaixotasuna zehaztasunez diagnostikatzeko eta monitorizatzeko. Gaur egun, jakina da garuneko neuroendekapenaz gain gaixoek erretinaren disfuntzioa garatzen dutela.

Izan ere, hainbat ikerketa lanek erakutsi dute gaixotasunak erretinan sorrarazten dituen aldaketak koherentzia optikoko tomografia (OCT, ingelesez) irudien bidez detektatu daitezkeela. Hala ere, OCT irudiak biomarkatzaile fidagarri gisa erabil daitezkeela frogatzear dago oraindik. Hobekuntza puntu gisa, ikerketa gehienek erretinaren ezaugarrien multzo txiki bat baino ez dute ikertu eta OCT irudi analisirako teknika estandarrak erabili dituzte gehienetan. Hemen, ezaugarri eta metodo aurreratuagoak erabiltzeak gaixotasunak eragindako erretina aldaketa konkretuak identifikatzen eta biomarkatzaile berriak garatzen lagun lezake.

Aurrekoa kontuan hartuta, lan honek bi helburu nagusi ditu: 1) OCT irudi analisirako teknikak hobetu eta aurreratu, eta 2) garatutako metodoak aplikatu Parkinson gaixotasuna dutenen erretina xehetasun gehiagoz aztertzeko.

Lehenik, metodo berriak garatu dira OCT irudiak lerrokatzeko, irudi kalitate ona bermatzeko eta erretina ezaugarri berriak kalkulatzeko. Sortutako metodoak RETIMAT izeneko softwarean integratu dira. Software horrek ikerketa lan honen oinarria osatzen du, eta era librean dago eskuragarri komunitate zientifikoarentzat.

Ondoren, gizaki osasuntsuen datuak erabili dira adinak eta sexuak erretinan duten eragina aztertzeko eta datu base normatibo bat sortzeko. Azkenik, OCT ezaugarri konbentzional eta berriak balioztatu dira gaixotasunaren hiru aplikazio esparruetan: diagnostikoa, larritasunaren ebaluazioa eta monitorizazioa. Azken kasuan, gaur arte jarraipen luzeena duten bi datu base longitudinalak erabili egin dira.

Emaitzen arabera OCT ezaugarriek Parkinson gaixotasunak eragindako kalte kognitibo eta motorrarekin loturiko informazioa dute. Gainera, gaixoen erretinak aldaketa diferentzialak izaten ditu denborak aurrera egin ahala. Hala ere, norbanako bakoitza zehaztasunez diagnostikatzea eta ebaluatzea OCT ezaugarriak erabiliz ez dirudi bideragarria. Horrela, gaixo indibidualen erabaki klinikoak gidatu baino gehiago, OCTk gaixotasunaren alderdi orokorrak ulertzeko tresna hobea izan daitekeela ondorioztatu daiteke.

Contents

List of Figures

List of Tables

Acronyms

AD Alzheimer's disease.

AIC Akaike information criterion.

AMD age-related macular degeneration.

AUC area under the curve.

BAC balanced accuracy.

BHRI Biobizkaia Health Research Institute.

BM Bruch's membrane.

CFT Central foveal thickness.

CI confidence interval.

COV coefficient of variation.

ELM-BM external limiting membrane-Bruch's membrane.

ETDRS Early Treatment Diabetic Retinopathy Study.

FD-OCT fourier-domain OCT.

FDR false discovery rate.

FOV field of view.

GCIPL ganglion cell-inner plexiform layer.

GCL ganglion cell layer.

GLCM Gray-level co-occurrence matrix.

GUI graphical user interface.

HC healthy control.

HES Hospital Episode Statistics.

HEYEX Heidelberg Eye Explorer.

HY Hoehn-Yahr.

ICC intra-class correlation coefficient.

ICD International Classification of Diseases.

ILM inner limiting membrane.

INL inner nuclear layer.

iPD idiopathic Parkinson's disease.

IPL inner plexiform layer.

LBP Local binary patterns.

LDA linear discriminant analysis.

LOESS locally estimated scatterplot smoothing.

LR logistic regression.

MAD mean absolute difference.

MEH Moorfields Eye Hospital NHS Foundation Trust.

MoCA Montreal Cognitive Assessment.

MRI magnetic resonance imaging.

MS multiple sclerosis.

NFL nerve fiber layer.

OCT optical coherence tomography.

OCTA OCT angiography.

ONL outer nuclear layer.

ONPL outer nuclear and plexiform layer.

OPL outer plexiform layer.

PD Parkinson's disease.

pfGCIPL parafoveal GCIPL.

PHR photorecepthor layer.

pNFL peripapillary NFL.

RETIMAT Retinal Image Analysis in MATLAB.

RF random forest.

RMSE root mean square error.

ROC receiver operating characteristic.

RPE retinal pigment epithelium.

SNR signal to noise ratio.

SVM support vector machine.

TABS Topcon Advanced Boundary Segmentation.

TRT total retinal thickness.

UKBB UK Biobank.

UPDRS Unified Parkinson's Disease Rating Scale.

1 Introduction

The hereby presented research work has one high level purpose: to investigate new imaging biomarkers for neurodegenerative diseases with Parkinson's disease (PD) as a specific case. That is, to use medical images to derive quantitative measurements of the anatomy and explore their link with the disease.

Although research on neurodegenerative diseases has been traditionally focused on the brain, as a particularity of this work, we use images of the retina instead of neuroimaging. This perhaps surprising idea stems from two facts: 1) the retina contains neurons and is directly connected to the brain, and 2) there is solid research evidence demonstrating an involvement of the retina in neurodegenerative diseases such as PD.

Importantly, the examination of the retina in-vivo is nowadays possible by means of optical coherence tomography (OCT) imaging. This technique is capable of imaging the retina in a non-invasive fashion and with high-resolution. For these reasons, it was the imaging method used in this research.

In broad terms, the work presented in this thesis involved developing a robust pipeline for OCT feature extraction, and evaluating the potential of the computed features as a biomarker for PD.

To further contextualize the research subject, this chapter introduces the fundamental concepts underlying the thesis. First, Section 1.1 describes the retina and its connection with the brain. Secondly, the working principle of OCT imaging and existing image processing methodologies are presented in Section 1.2. Finally, Section 1.3 provides a general description of PD, and establishes the connection between the retina, PD and OCT imaging.

1.1 The retina and the visual system

The retina is located on the inner surface of the posterior part of the eye (Figure 1.1). It is a photosensitive layered tissue that captures incident light and transforms it into nerve impulses that are transmitted to the brain via the optic nerve.

To achieve the 120° field of view (FOV) of the human eye [1], the retina must cover most of the ocular globe. This is because incoming light hits the retina at a different point depending on the angle of entry. However, visual capability is not uniform across the entire FOV and is diminished in the peripheral areas which are further away from the central region [2]. These differences are a consequence of the non-uniform distribution of the two main photoreceptor cells in the retina: cones and rods (Figure 1.2). Cones are related to color perception and visual acuity in daylight conditions, while rods are highly sensitive photoreceptors specialized in night vision [3].

The photoreceptors are located all across the retina. These generate signals in response to incident light, that are ultimately transmitted to the brain by neuronal axons that merge together forming the optic nerve. This nerve is directly connected to the brain and is located at 17°

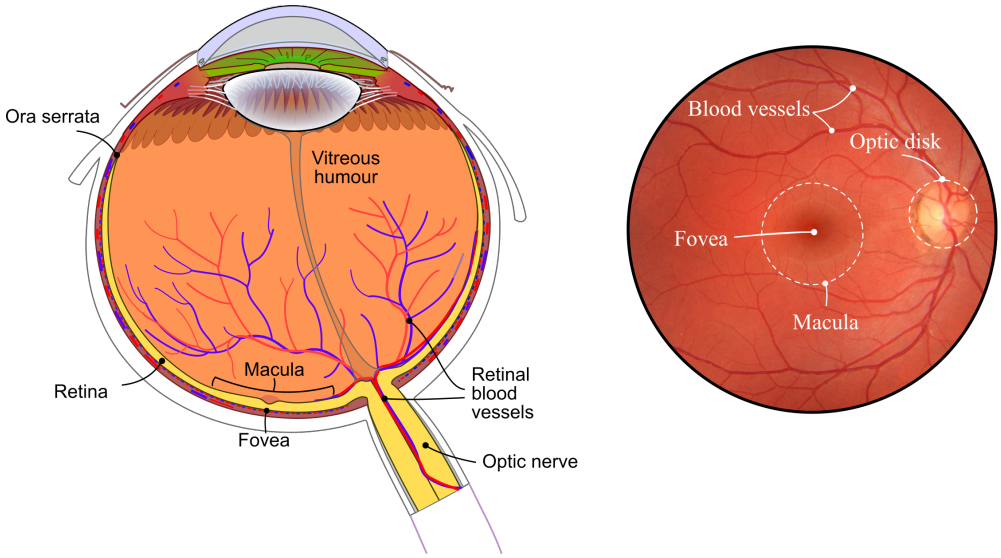


Figure 1.1: Eye and retina. Left[1]: axial drawing of the eye and its main landmarks. Right: color photography of the eye fundus (i.e., interior surface of the back of the eye).

horizontally from the fovea.

The number of cones is noticeably higher for the central 17° of our FOV [4]. Anatomically, this range of our vision field corresponds to the macula, a 6 mm diameter circular region that surrounds the center of vision and plays a major role in visual function (Figure 1.1). The central subregion of the macula, where the cone density reaches its maximum, is called the fovea and is a highly specialized region where the cell layers that are usually on top of photoreceptors are not present. This architecture allows light to reach photoreceptors without obstacles and gives the fovea its convex shape.

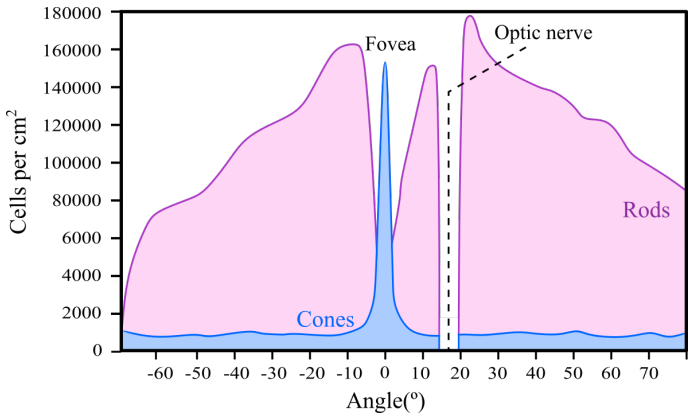


Figure 1.2: Distribution of cone and rod density[2]. The horizontal axis represents the angle of the field of view measured from the center of vision, which is located at the fovea.

The distribution of rods follows a different pattern. The greatest difference is that there are far fewer rods than cones in the center of vision. This clearly shows that, in the fovea, color and daylight vision prevails over low contrast visual acuity [5]. In fact, the point of highest visual sensitivity under poor illumination reaches its maximum at 7° from the fovea. Altogether, the joint distribution of both cones and rods is the underlying reason the thickness of the retina ranges

[1] Adapted from *Diagram of the human eye in English* by Rhcastilhos and Jmarchn. CC BY-SA 3.0 license.
[2] Adapted from *Distribution of Cones and Rods on Human Retina* by Jochen Burghardt. CC BY-SA 3.0 license.

from 0.08 *mm* at the ora serrata up to 0.32 *mm* close to the center of vision [6], as illustrated in Figure 1.1.

At a microscopic level, the retina possesses a complex architecture comprised of numerous neurons responsible for processing and transmitting information. Each neuron is composed of a cell body, dendrites that receive signals from other neurons, and an axon that carries signals to other neurons or target cells. The neurons in the retina are organized into multiple horizontal layers of two main categories: cell body layers, made of neurons bodies, and synapse layers, which contain neuron axons and interconnect cell body layers.

Figure 1.3 shows a cross-sectional representation of the layers that comprise the retina. Each layer has different cells that fulfill a specific function. The communication between layers takes place in a bottom-up fashion, i.e., signals generated by photoreceptors are transmitted upwards, layer by layer, by means of several synapse processes until they are finally sent to the brain.

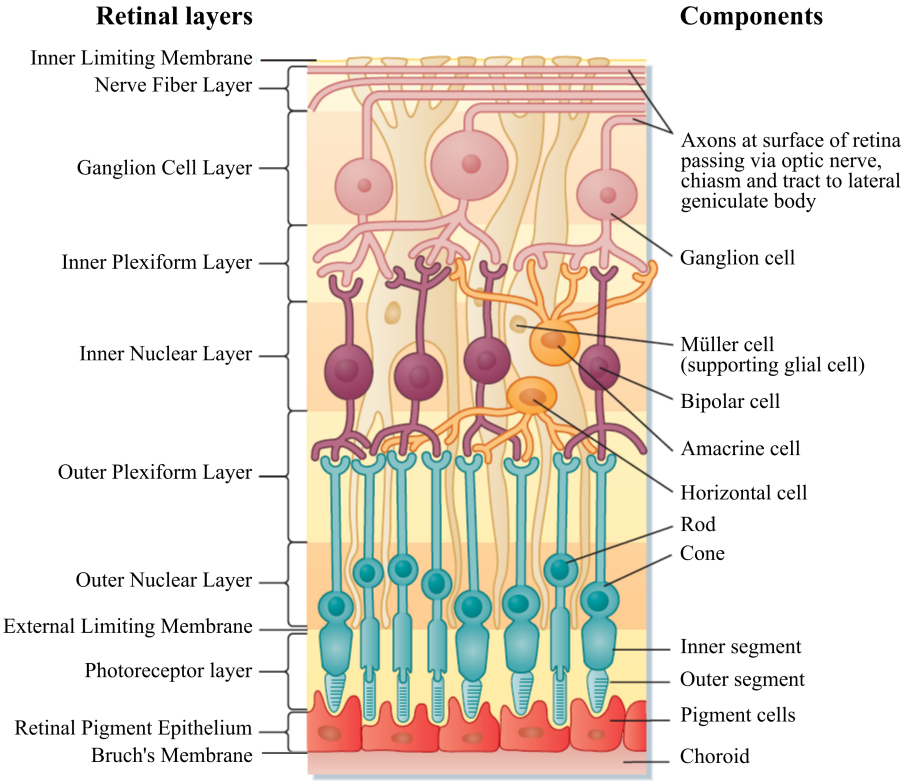


Figure 1.3: Retinal layers and cells[3]. Retinal layers are listed in the left while the cellular elements that comprise each layer are labeled in the right.

The bottom boundary of the retina is the Bruch's membrane (BM) and is located just below the retinal pigment epithelium (RPE), the outermost cell layer of the retina. Epithelial cells serve as ground tissue for cones and rods, and reflect light before it reaches the photorecepthor layer (PHR), which contains the inner and outer segments of both cones and rods. Photoreceptor cell bodies are located just above, in the so-called outer nuclear layer (ONL). The junction between PHR and ONL is called the outer limiting membrane, and contains adherents between photoreceptor cell inner segments and Müller cells. The latter are a special type of glial cells that prevent retinal detachment by keeping photoreceptors in their position.

[3]Reprinted with permission from [7] © Elsevier (2017).

The layer on top of the ONL is a synapse layer called the outer plexiform layer (OPL), in which axons from photoreceptors, bipolar and horizontal neurons are intertwined and communicate with each other. Every bipolar cell receives information from one or more photoreceptors, while horizontal cells help the information integration process. Both bipolar and horizontal neurons, together with amacrine cells and Müller cell bodies, are located in the inner nuclear layer (INL). The layer where synaptic connections between bipolar, amacrine and ganglion neurons take place is the synapse layer called the inner plexiform layer (IPL). In this information exchange, bipolar cells directly communicate with ganglion cells with the help of amacrine cells. These ganglion cells comprise the ganglion cell layer (GCL), the innermost cell layer of the retina that ultimately sends the information forward to the brain via the nerve fiber layer (NFL), a layer composed only of ganglion cell axons running from every point of the retina to the optic nerve. Finally, just on top of the NFL, Müller cell endpoints create the inner limiting membrane (ILM), an upper boundary between the retina and the vitreous humor.

It is important to note that the retina is just one of the building blocks of the human visual system. The whole system (Figure 1.4) comprises several anatomical elements and a series of complex steps until the signal generated by the retina reaches the cerebral cortex. Electrical signals sent by NFL axons travel via the optic nerve to the optic chiasm. At that point optic nerves from both eyes intersect. After that, the signals continue through the so-called optic radiations to the lateral geniculate nucleus and the visual cortex, located at the occipital lobe of the brain.

As a distinctive characteristic, the left visual field is processed by the right hemisphere of the brain (red and yellow in Figure 1.4, and the right visual field is processed by the left hemisphere. This is because the nasal fibers of each retina (i.e., closer to the nose) cross over to the opposite side of the brain, while the temporal fibers remain on the same side. This arrangement allows for the integration of visual information from both eyes and enables depth perception and a more complete understanding of the visual environment.

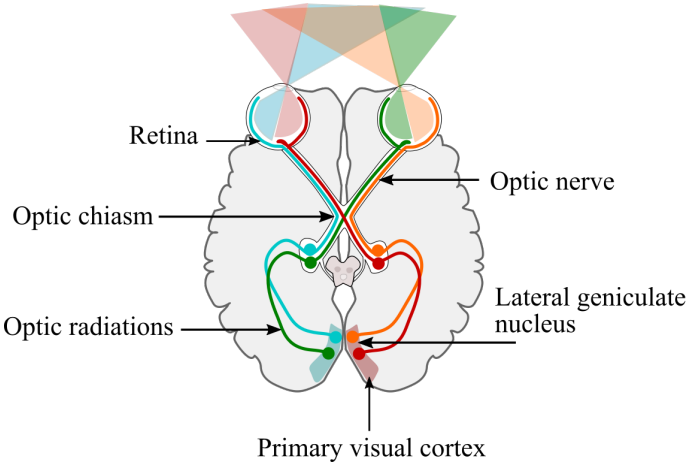


Figure 1.4: Axial view of the human visual system[4]. The retina transforms the light entering the eye into nerve impulses that are transmitted to the brain via the optic nerve. In the optic chiasm optic nerves partially cross over so that the information from the left visual field of both eyes is sent to the right brain hemisphere. Accordingly, the information from the right visual field is processed by the left hemisphere.

The entire visual system can be visualized with medical imaging techniques with different levels of resolution. While neuroimaging is commonly used to evaluate cortical and subcortical structures, retinal layers can be imaged with high resolution by using OCT.

[4]Adapted from *Human visual pathway* by Miquel Perello Nieto. CC BY-SA 4.0 license.

1.2 Optical Coherence Tomography

Working principle

OCT is a non-invasive imaging technique that generates high resolution depth images of tissues or materials. It uses low coherence light and is based on the echo-location principle. As in ultrasound imaging, a tissue is scanned in-depth by measuring the round-trip delay and intensity of successively sent pulses. In OCT, however, near-infrared light pulses are normally used due to their remarkable penetration, high resolution, and harmless properties [8].

The technique was conceived in the 1980s and the first in-vivo retina image was published in 1991 [9]. Since then, it has undergone constant development and has proved useful for many applications, of which ophthalmology is one of the most significant. The key advantage of OCT is the ability to acquire high quality images of the retinal layers in a non-invasive manner and within a few minutes. The micrometric resolution of these images enables detailed qualitative and quantitative examination of the retina. For these reasons, OCT is one of the most widely used tools for both research and clinical ophthalmology.

Figure 1.5 illustrates the OCT scanning process, which is carried out in a point-wise fashion.

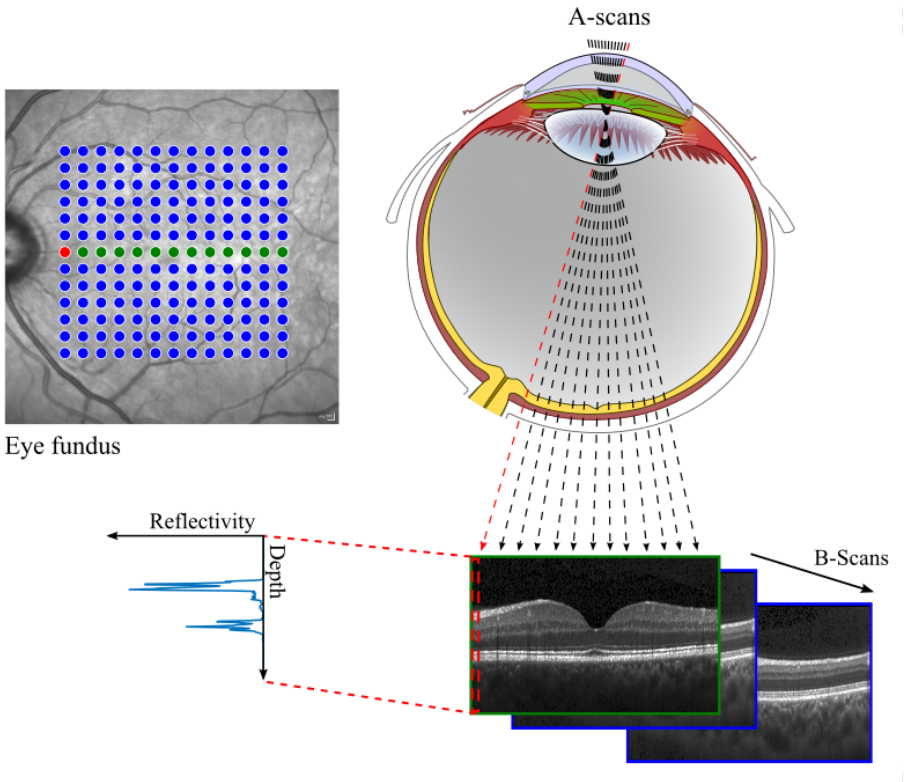


Figure 1.5: OCT scanning process[5]. The top-left image shows the eye fundus (i.e., the interior surface of the back of the eye) with several OCT scanning points. Each of the points is known as A-scan and is transformed into a grayscale image column to create 2D images known as B-scans.

Each scanning point begins with the transmission of several pulse that penetrates the tissue and travels through different cellular layers. This signal generates reflections at each layer boundary.

[5]Created using *Diagram of the human eye in English* by Rhcastilhos and Jmarchn. CC BY-SA 3.0 license.

From the delay and intensity of each reflection it is possible to obtain a reflectivity profile that encodes the depth and reflectivity of each tissue layer (see bottom-left of Figure 1.5). This one-dimensional signal, also known as A-scan, can then be mapped into a grayscale as a column in an image. When several A-scans are stacked in parallel, a cross-sectional image of the retina, called a B-scan, is generated. By acquiring distinct B-scans a volume scan of the tissue can be obtained. The entire scanning process comprises the acquisition of multiple A-scans following a certain spatial sampling distribution. Although multiple acquisition patterns exist, the most commonly employed is the horizontal raster pattern, in which horizontal B-scans are acquired sequentially.

As a consequence of the propagation speed of near-infrared pulses, the time delays of reflected signals are extremely short (in the order of nano seconds) and cannot be measured directly by conventional electronic devices. To overcome this problem OCT technology relies on interferometry and low coherence light which, operating together, permit the receptor to measure the reflected time delays indirectly.

The whole OCT system involves several stages of reflection, interference, and reception. To fully grasp its operation, it is essential to analyze the evolution of the signal through the acquisition process from an analytical perspective. The explanation presented here closely follows a more comprehensive description from the reference work by Drexler et al. [10].

The most common OCT set-up is a Michelson interferometer and includes a light source, a beam splitter, a mirror, and a detector (Figure 1.6).

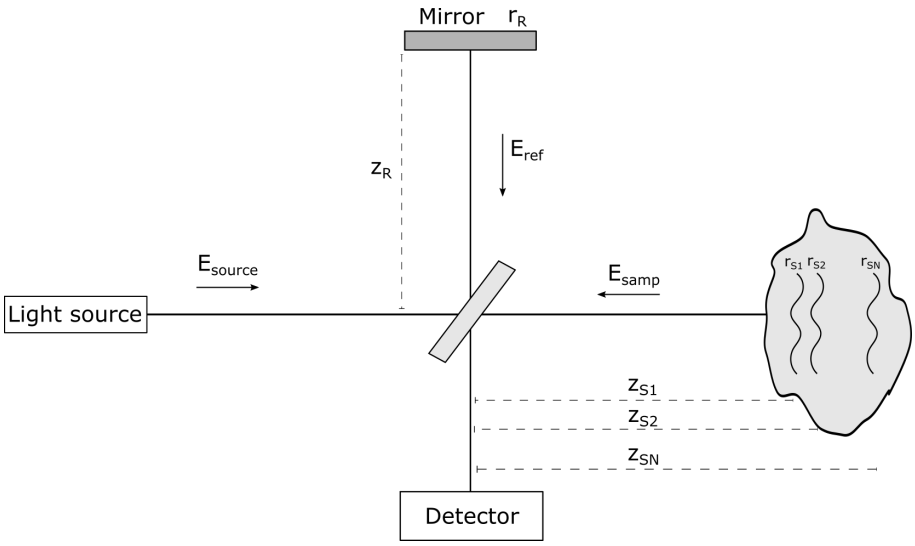


Figure 1.6: Basic OCT set-up. Shown elements: light source (left), beam splitter (middle), mirror (top), tissue sample (right), and detector (bottom).

The scanning process begins with a continuous low coherence light source, usually a laser, which emits an electromagnetic plane wave whose electric field is E_{source}. A beam splitter is used to divide the wave into two equal beams, which are sent to two arms:

- **Reference Arm:** the wave follows a distance z_R until it is completely reflected in a mirror with reflectivity r_R. After reflecting, the signal that reaches the beam splitter is E_{ref}.
- **Sample Arm:** the wave follows this path until it reaches the tissue sample. As the light travels through the tissue, differences in the refraction index between layers result in different reflections. The tissue can be modeled as a set of N layers with a particular depth (z_{Sn}) and reflectivity (r_{Sn}). The signal that returns to the splitter (E_{samp}) is the sum of the N reflections caused by every tissue layer.

Light from both arms is then recombined in the beam splitter resulting in an interference

phenomenon, from which it is possible to determine the reflectivity and depth of the layers that form the tissue sample. The photocurrent generated at the detector ($I_D(k)$) in the frequency domain (i.e., as a function of the wave-number k) can be described as:

$$I_D(k) \propto S(k) \sum_{n=1}^{N} r_R r_{Sn} cos(2k(z_R - z_{Sn})) \tag{1.1}$$

Essentially, the result of the whole process described above is that the source spectrum $S(k)$ is modulated by a series of sinusoids, each corresponding to a reflection in a certain tissue layer. The amplitude and frequency of these sinusoids encode both the reflectivity and depth of each layer, respectively.

As an example, Figure 1.7 shows how interference modifies the source spectrum depending on the reflective layers of the tissue sample. In the simplest scenario, a sample with only one layer is scanned. There is only one reflection and thus the fringe pattern has a single frequency. The amplitude of the oscillations is proportional to the reflectivity of the layer.

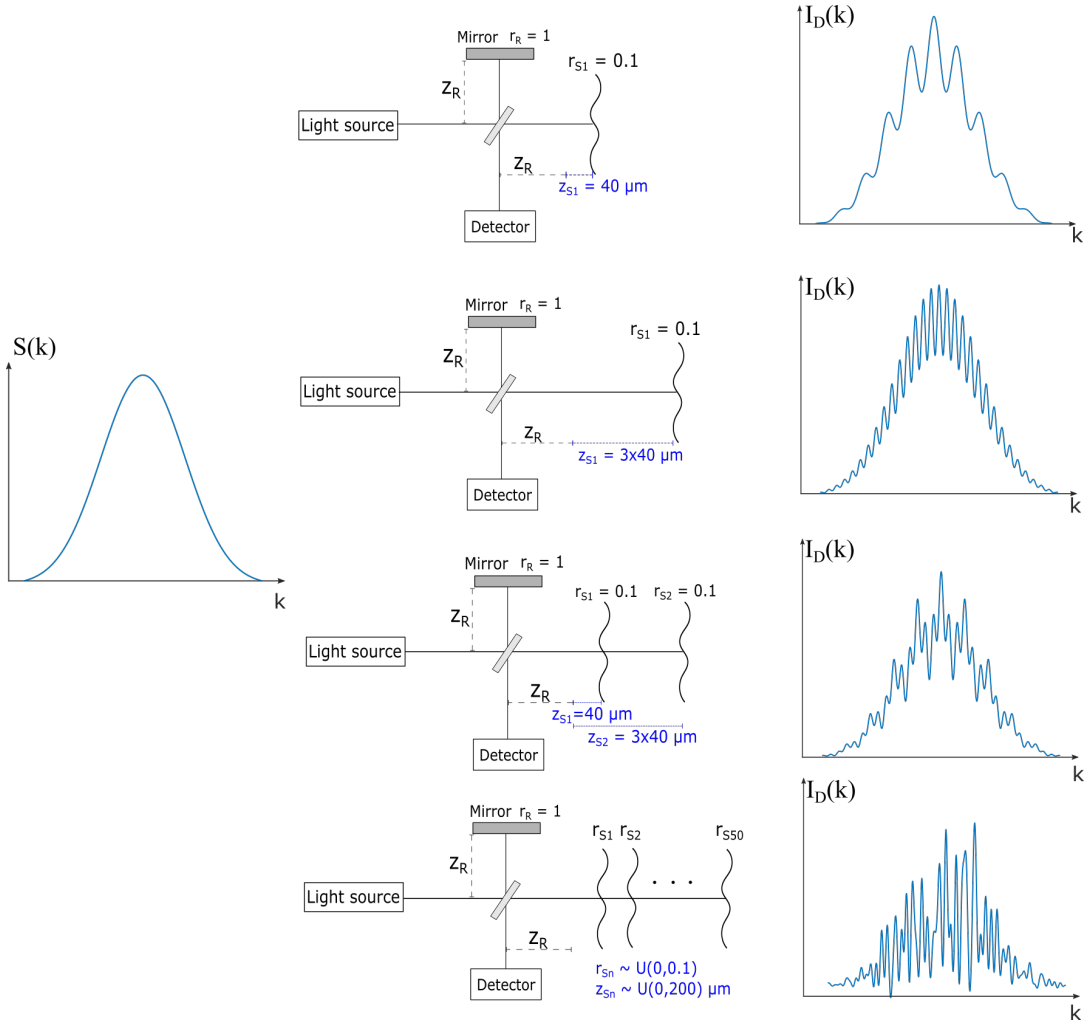


Figure 1.7: OCT signal formation examples. The gaussian spectrum of the light source ($S(k)$ in the left) gets modulated differently depending on the image tissue layers generating a different photocurrent $I_D(k)$ in each scenario.

As the scanned layer becomes deeper, the resulting modulation pattern increases its frequency. The second row of the figure shows that a reflective layer located at three times the first layer

generates oscillations of three times higher frequency. When two or more layers are scanned, modulations generated by each layer are added up to form the resulting pattern.

Finally, the last row depicts a more realistic example in which 50 layers are randomly located at depths $z_{Sn} \sim \mathrm{U}(0,\, 200\,\mu m)$ with reflectivity following $r_{Sn} \sim U(0, 0.1)$. The measured spectrum is the composite of all modulations.

At this point, it is important to note that all the previous reasoning is only possible due to the low coherence of light. This means that the light source is not monochromatic but consists of a set of frequencies usually forming a gaussian spectrum, as shown in Figure 1.7:

Without this property, the modulation patterns of Figure 1.7 would not be generated, and the received spectrum would comprise a single frequency from which it would not be possible to retrieve any depth or reflectivity information.

Modern OCT scanning devices decode this information by means of fourier-domain OCT (FD-OCT) [11,12], a technology that aims to first recover the whole received spectrum and then derive a layer reflectivity profile from the modulation pattern. Depending on the method used for measuring the received spectrum, FD-OCT systems can be divided into two categories:

- **Spectral Domain**: the light source has is broadband and covers a wide spectrum. The received signal is then decomposed based on a spectrometer which splits the signal into different frequencies.
- **Swept Source**: the light source consists of a single tunable frequency that sweeps across frequencies. With each sweep, a single frequency is sent and measured. The received spectrum is then recovered by combining all independent measurements acquired during the sweeping process.

Using either of these methods, once the spectrum $I_D(k)$ is measured, the corresponding intensity profile $I_D(z)$ is obtained with the inverse Fourier transform. Then, the final intensity profile (i.e., A-scan) is retrieved by further processing steps.

Image analysis

OCT images allow to analyze the retina both qualitatively, by inspecting the image, and quantitatively, by computing features such as thickness and volume. This quantitative analysis usually comprises three image processing tasks: preprocessing, layer segmentation, and feature extraction (Figure 1.8). In addition to these, quality assessment can be added after every step to detect errors and prevent unreliable results.

Preprocessing

Initial image preprocessing tasks intend to improve or simplify posterior processing steps. There are three main preprocessing steps in OCT: display distortion correction, image denoising, and image registration.

Traditionally, OCT images are displayed and analyzed in a rectangular format even though A-scans are not acquired in parallel but in a fan-beam angular pattern. This rectangular display introduces a bias known as display distortion, which affects eye curvature and thickness measurements, especially for larger eccentricities [13]. Display distortion can be corrected by using optical ray-tracing simulations to relocate each A-scan position according to its real optical path. Problematically, these simulations require information about eye biometry (axial length, corneal curvature, etc.) and optical properties of the device [14,15]. This information is often not available and therefore display distortion correction is not always applied in OCT research.

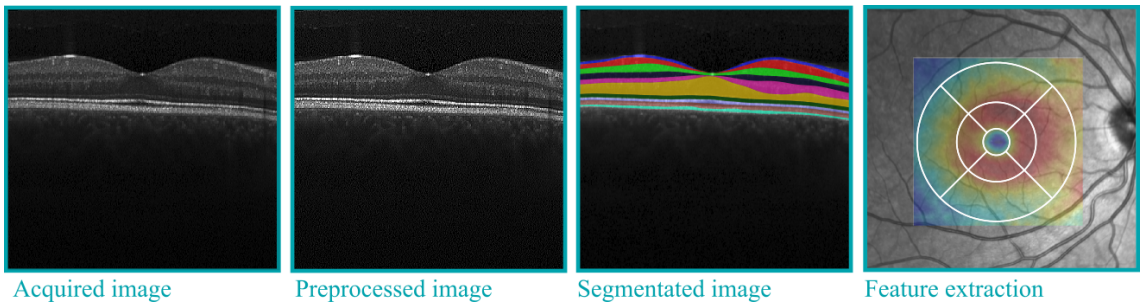


Figure 1.8: OCT image analysis pipeline. The acquired image is enhanced by means of preprocessing. Then, the retinal layers are segmented and used to compute thickness en-face maps. An en-face map (shown in color in the right image) is a 2D map containing the thickness values of a certain layer computed for each scanning point or A-scan. From the thickness maps multiple retinal features can be computed.

Image denoising aims at reducing the speckle noise present in OCT images as a consequence of the interference between multiple forward and backward scattering of light waves that are mutually coherent [16]. The reason behind this technique is that enhanced image quality results in better performance of subsequent tasks such as grading or segmentation. Denoising is a complex task that can be performed by statistical methods [17, 18] as well deep learning models [19, 20]. The benefits of denoising are dependent on the original image quality and the intended task and it is therefore not always necessary [17, 19]).

Image registration serves to align images so that anatomically equivalent locations overlap. Accurate alignment is key for reliable groupwise analysis and change detection. In the latter, the goal is to compare images of the same eye acquired in different moments. For this purpose, both classical image registration methods [21] and custom OCT algorithms [22] have shown good performance. In addition, the built-in software of scanning devices often provides registration functionalities. For instance, Heidelberg devices include a follow-up registration function that ensures that the device images the same location in every visit, which has proven to reduce measurement variability [23].

On the other hand, groupwise registration is a more difficult task that is especially challenging in OCT. Part of the difficulty lies in the scarcity of anatomical landmarks in the retina. Apart from the fovea and the optic disk, there are no anatomical structures that are logically coherent across the population and which could be used to aid registration. Existing solutions for groupwise registration involve complex 2D and 3D non-rigid registration models that are still under research [24, 25]. Alternatively, a coarse alignment of images can be attained by aligning the foveal center of all subjects under study. This is partially achieved during acquisition by centering the image at the fovea. However, fixation errors can cause the center of the fovea and the center of the image to become misaligned. The solution to this is to locate the foveal center and use it as a new origin of coordinates. For this, automatic centering algorithms are a very compelling option to correct centering errors in a fast and reproducible way.

Segmentation

After preprocessing, the first step is usually retinal layer segmentation. Images can be manually segmented, but this is a time-consuming process subject to inter-rater variability. To overcome this problem, the development of automatic segmentation algorithms has been a subject of much study. Algorithms based on graph-search have shown particularly good performance [26–28]. These methods exploit the contrast difference between retinal layers and consist of two steps: 1) vertical image gradients are computed to enhance the boundaries between layers and accordingly assign weights to pixels, and 2) layer boundaries are sequentially segmented by finding horizontal paths following pixels with the highest gradient value. In addition to graph-search, approaches based on machine learning and boundary classification have also been proposed [29]. More recently, research on retinal segmentation has shifted to a data-centric approach, in which deep learning

models are directly trained from annotations [30–33]. Most of these models are based on U-Net, a neural network architecture that has demonstrated outstanding performance across multiple domains [34].

Automatic segmentation methods have gained traction in recent years, achieving a performance close to inter-rater accuracy [35]. For this reason, they are commonly preferred over manual segmentation. Indeed, most scanning OCT devices incorporate built-in retinal layer segmentation software. For instance, Heidelberg and Topcon devices rely on Heidelberg Eye Explorer (HEYEX) [36] and Topcon Advanced Boundary Segmentation (TABS) [37], respectively. Although the inner workings of these proprietary algorithms are not disclosed, they have demonstrated good performance [35] and are easy to use as they are usually integrated with the device.

Quality assurance

Ensuring both images and segmentation results are of sufficient quality is crucial to obtain reliable results. In the context of OCT, at least six issues should be considered in quality assurance:

- **Poor image contrast:** acquired B-scans may show poor contrast between layers (Figure 1.9a) due to ocular conditions such as cataracts, vitreous floaters, or system misconfiguration. Poor contrast images can complicate downstream tasks such as segmentation. Devices usually provide both a quality metric as well as a recommended minimum threshold for an image to be considered of acceptable quality.
- **Motion artifacts:** when the subject under examination does not maintain a constant fixation, the imaged area may include overlapping regions. For instance, Figure 1.9b shows an example of a subject who changed fixation mid-acquisition which resulted in the foveal region being imaged twice.
- **Blink artifact:** if a subject blinks during acquisition, some B-scans may be not image the retina but appear totally empty as shown by black horizontal B-scans in Figure 1.9c. The severity of this artifact depends on the speed at which B-scans are acquired and the mechanisms of the acquisition device to detect and correct the problem.
- **Off-center artifact**: it is caused by a mismatch between the scan depth range and the depth of the eye under examination. As a result, the retina is cropped and cannot be analyzed completely (Figure 1.9d). In fact, often a mirror artifact appears in the cropped region, which is a consequence of the FD-OCT device not being able to distinguish between positive and negative reflectance delays [38].
- **Segmentation errors:** automatic segmentation algorithms may fail because of poor image contrast or ocular lesions. Detecting and correcting these errors is a difficult task that often needs to be done manually. An example error is shown in Figure 1.9e.
- **Ocular pathologies:** there is a plethora of ocular lesions that can affect the integrity of the retina. When such lesions are not part of the research question, they can heavily bias measurements as depicted in Figure 1.9f.

Correcting these errors can be difficult for segmentation failures or even impossible in images with poor contrast or artifacts. Therefore, to ensure that computed retinal features are reliable, images with errors need to be identified and excluded. Detecting these problems, however, requires accounting for multiple sources of errors, which is not straightforward. Although solutions based on automatic algorithms are being investigated [39,40], image quality assurance still requires visual inspection of the data in many cases.

Feature extraction

The main goal of quantitative OCT analysis is to compute features that describe the retina. These features range from simple thickness values to more complex metrics such as fractal dimen-

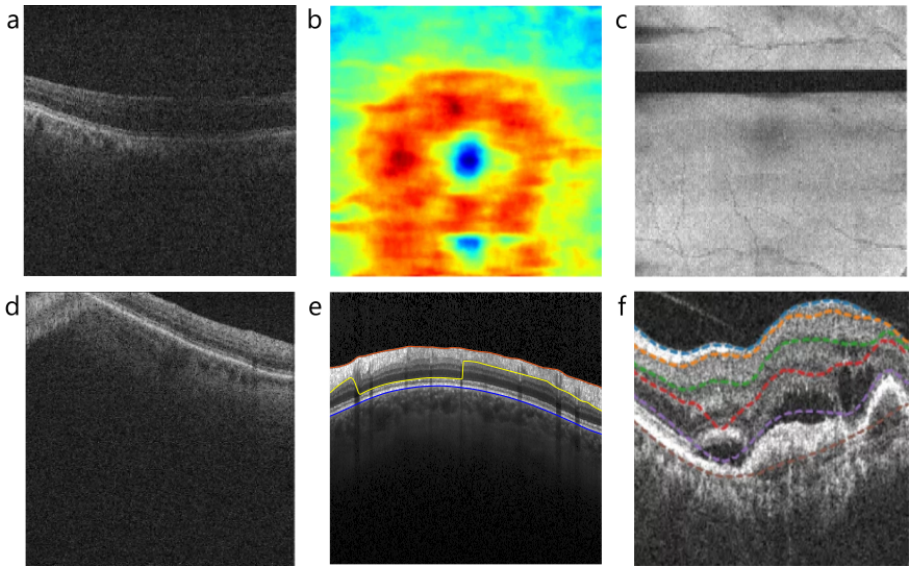


Figure 1.9: Common problems affecting OCT images. a) B-scan with poor contrast. **b)** En-face thickness map with the fovea duplicated due to a motion artifact. **c)** En-face reflectance map showing a few B-scans totally black (empty) due to blinking. **d)** B-scan cropped at the top left corner due to an off-center artifact. **e)** Segmentation error of the NFL-GCL boundary (yellow). **f)** Ocular pathology affecting the retinal structure.

sion. Each feature describes a concrete aspect of the retina and involves a specific set of processing steps. Depending on the aspect of the retina under study, features can be grouped into three main categories:

- **Thickness**: reflects the spatial distribution of each retinal layer. Changes in thickness are of high interest as they might capture variations in the inner cellular structure. For instance, neurodegeneration may result in thinning [41] while pathologies such as age-related macular degeneration (AMD) can lead to thickening [42].
- **Foveal pit morphology:** the foveal pit is a singular region whose morphology can be geometrically analyzed. These type of features describe high-level properties of the fovea such as width, depth, or slope.
- **Texture**: accounts for local statistical and structural properties of a thickness map or image volume. These features provide more complex information that can be used to identify patterns or structures within an image.

From OCT features it is possible to perform statistical analyses to determine how a certain pathology affects the retina. In the particular case of this thesis, multiple features are used to investigate retinal changes in PD.

1.3 Parkinson's Disease

PD is a heterogeneous neurodegenerative disease that causes a severe movement disorder as well as a broad range of non-motor symptoms. Its motor symptoms—resting tremor, rigidity and bradykinesia—were described for the first time by James Parkinson in 1817 and constitute a syndrome called parkinsonism [43]. The disease is typically manifested as a slow deteriorating process [44]. As the disease progresses, PD patients can develop numerous non-motor clinical features related with cognition, visual function and mental health [44]. For instance, cognitive decline is associated with PD and results in around 40% of PD patients developing dementia [43]. Depression and sleep disorders are also common in PD, and visual impairment affecting color vision

and contrast sensitivity has also been described [45]. In addition to a worsening in the quality of life, the life expectancy of PD patients is shorter and can be reduced by more than 6 years at 65 years-old [46].

Only surpassed by Alzheimer's disease (AD), PD is currently the second most common neurodegenerative disease [47], with a prevalence of around 0.5% among 70-year-old people [48]. While men are 1.4 times more likely to develop PD, the principal risk factor is age [48]. In fact, as a consequence of the progressive ageing of the population, there has been a 144% increase in the total number of PD patients between 1990 and 2016 [48], and the number of people with PD could dramatically rise up to 14 million by 2040 [49]. This makes PD diagnosis and treatment a worldwide healthcare challenge.

Except for a few variants associated with specific gene mutations, the underlying cause of PD remains unknown [44]. As a result, most PD cases are typically classified as idiopathic Parkinson's disease (iPD), a category that encompasses patients with a wide variety of symptoms and factors. Although there are no established causal mechanisms for iPD, the current research view is that the cause is probably an intricate combination of genetic and extrinsic factors [44].

Unlike the etiology, the pathophysiology of the disease is better understood. Parkinsonian symptoms are deemed to be a consequence of a loss of dopaminergic neurons in the substantia nigra [43], a small midbrain region with a high density of dopaminergic neurons (Figure 1.10). These neurons secrete a dopamine neurotransmitter through the so-called nigrostriatal dopaminergic pathway that extends across the brain and is associated with motor control [50]. Neuronal death in the substantia nigra disrupts this dopaminergic pathway resulting in a malfunction. Interestingly, motor symptoms are not experienced until 50%-80% of the neurons are lost [43]. The neurodegeneration process is caused by protein accumulations called Lewy bodies and Lewy neurites that grow inside neurons and ultimately cause their death. These bodies are mainly made up of α-synuclein and have been found not only in the substantia nigra, but in other regions of the brain such as the cortex or the amygdala [43]. Moreover, in addition to the nigrostratal pathway, several other motor and non-motor circuits are affected in PD [44].

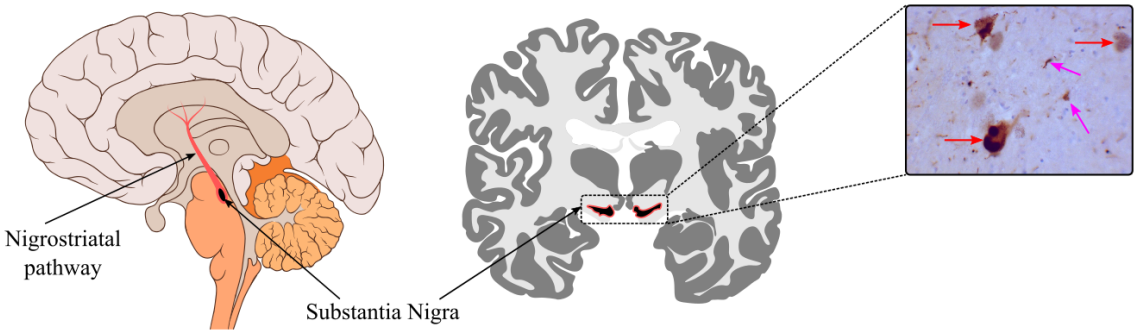


Figure 1.10: Substantia nigra and Lewy bodies[6]. Left and middle drawings show the location of the substantia nigra in coronal and sagital axes, respectively. The right image shows Lewy bodies (red) and Lewy neurites (pink) observed under the microscope.

Regarding treatment, the drug that has proven to be the most effective is an amino acid called Levodopa that increases dopamine concentrations in the brain [51]. However, in severe cases with no response to drugs, highly-invasive surgical treatments such as thalamotomy and deep brain stimulation can be used [44, 52].

One of the main challenges of PD is its diagnosis, which nowadays is entirely clinical. After excluding alternative causes of motor symptoms, clinicians employ a clinical criteria and a grading

[6] Adapting and combining *Basal Ganglia* by Mikael Häggström and Andrew Gillies, *Dopaminergic pathways* by Patrick J. Lynch, and *Lewy bodies (alpha synuclein inclusions) 2* by Suraj Rajan. CC BY-SA 3.0 license.

scale that measures the diagnosis certainty based on observed symptoms and response to treatment [44, 53]. However, definite diagnosis can only be made by autopsy [44]. As a consequence, diagnostic errors are common in PD, specially in early disease stages or in presence of comorbidity complicating the diagnosis [44]. In addition, not all patient evolve equally, which makes prognosis difficult. Thus, the development of new biomarkers for PD is critical to aid diagnosis and monitor the evolution of the disease more accurately. In this sense, the use of neuroimaging and other imaging techniques has been extensively investigated and many promising biomarkers have been proposed [54–56]. However, most of these biomarkers have not been fully validated for clinical use or are only used as ancillary tests [54, 56].

As an alternative, imaging modalities that do not look directly into the brain are also being explored. In recent years, retinal OCT imaging has gained attention as a new tool for studying PD. The examination of the retina and the visual pathway in PD is motivated by the known visual impairment associated with the disease. In fact, PD patients often experience visual symptoms such as a decline in contrast sensitivity, impaired color discrimination, poorer visual acuity, and perception problems such as blurring [57].

Importantly, research suggests that visual symptoms are related to a functional dysfunction of the retina. More concretely, early studies examining the electrical function of the retina found abnormal evoked potentials and a specific malfunction of ganglion cells in PD patients [58,59]. Further research has also shown this retinal dysfunction to be improved in patients under medication, suggesting a direct correlate with the disease [60].

The involvement of the retina in PD is further supported by post-mortem studies that have observed a decreased dopamine levels in human retinal tissue [61]. Later work has reported similar findings in animal models and has observed a retinal dopaminergic cell degeneration [57,62]. These findings are in line with one of the primary pathological hallmarks in PD: a dopaminergic neuron loss. Notably, a presence of abnormally phosphorylated α-synuclein as well as Lewy neurite structures has also been found in PD retina [63]. These discoveries are of great importance as they suggest that retinal changes parallel brain pathology.

Building upon this histological evidence, several studies have sought to examine the retina by means of OCT imaging. The first wave of OCT studies in PD has provided accumulated evidence of specific retinal changes in PD [41]. Despite these encouraging results, it remains uncertain whether the observed changes can be reliably used as biomarkers. This is because there is a high heterogeneity in the literature regarding the spatial location and the effect size of the changes. Moreover, a clear link between retinal changes measured by OCT and clinical progression in PD is yet to be ascertained.

Accordingly, there is still research to be conducted to establish or dismiss OCT as a clinical tool for PD. As a key point of improvement, it is crucial that new research linking OCT and PD makes use of richer datasets. To this end, in the present body of research we: 1) explore new OCT features to enhance the description of retinal changes, and 2) use larger datasets with longitudinal follow-up and clinical information to evaluate the clinical utility of OCT in PD.

2 | State Of The Art

This chapter provides a literature review of the topics addressed by the main contributions of this work. In Section 2.1, the elements of the OCT analysis pipeline investigated in the thesis are examined, including automatic foveal center location, quality assurance, feature extraction, foveal pit morphology analysis, and OCT analysis software. Section 2.2 summarizes studies investigating how demographic factors affect the retina. Then, a comprehensive review of studies linking OCT with PD is presented in sections 2.3 and 2.4.

The aim of this chapter is twofold: 1) to identify areas of improvement in OCT processing, and 2) describe the findings and limitations of existing OCT studies in both healthy and PD cohorts. The main conclusions of the literature review as well as existing gaps in knowledge are summarized in Section 2.5. Finally, Section 2.6 states the specific objectives of this research project.

2.1 OCT image analysis

Automatic foveal center location

Correctly locating the foveal center reduces the impact of misalignment errors and improves the reliability of thickness measurements [64]. Although the foveal center can be located manually, it is highly desirable to perform this step automatically to increase reproducibility and reduce human involvement.

The simplest automatic approach involves locating the fovea at the point of minimum thickness [64]. However, this approach is not reliable in retinas where the absolute minima does not lie at the foveal center, for instance, due to ocular pathologies or imaging artifacts [65]. To account for this, Niu et al. proposed an algorithm based on saliency maps computed from both local and global features [65]. The algorithm showed good performance in both healthy eyes and eyes with AMD. As a limitation, it is technically complex and requires fine-tuning of hyperparameters, which makes it difficult to implement. Moreover, the used test dataset included only 64 different eyes imaged with a high-resolution protocol with 128 B-scans. It is therefore not clear how it would perform in datasets with a larger variety of ocular conditions and acquisition protocols.

Alternatively, Liefers et al. proposed a deep learning model based on pixel-wise classification and a convolutional neural network that showed good performance in a cohort of 400 AMD eyes [66]. More recently, in [67], the authors developed a pixel-wise distance regression model based on spatial location priors and a U-net architecture. The model uses a whole OCT image volume as input and generates a distance map from which the fovea is located. The model was trained and tested in a large dataset of 5586 OCT volumes including cases with AMD and macular edema. According to the results, the model worked well in all cases and outperformed the previous deep learning model by Liefers et al. [66]. As a limitation, the model was trained only on Heidelberg images and it is unknown how it would extrapolate to other devices. In addition, since it is not openly available, using such a model would require a new model to be trained from scratch following the

architecture of the paper.

Quality assurance

To address OCT quality problems, the community has adopted a quality control standard criteria called OSCAR-IB [68] that provide validated guidelines for image quality assurance in research linking OCT and multiple sclerosis (MS). The criteria account for: (O) obvious problems, (S) poor signal strength, (C) centering problems, (A) algorithm failure, (R) retinal pathology not related with MS, (I) illumination, and (B) beam placement. Although the criteria were developed for MS, they can be directly extended to diseases measuring similar changes such as PD or AD.

For the criteria to be as robust as possible, one would want to reduce subjectivity by setting quantitative inclusion rules that can be automatically applied. Such a rule can only be directly defined for image contrast. Indeed, OSCAR-IB criteria establish a signal strength > 15 as a threshold for image inclusion. This threshold, however, was derived from and for Heidelberg devices and cannot be applied to other devices. Moreover, the actual derivation of the quality metrics of each imaging device is often not disclosed, which makes it difficult to interpret them. As an alternative, a few research works have tried to develop and validate other image contrast metrics. For instance, in [69], the authors proposed a quality index that showed to be closely related to human grading. Later, Huang et al. developed a metric called maximum tissue contrast index that generalized well to multiple devices [70].

Despite the utility of these metrics, poor image contrast is only one of the problems in OCT, and additional solutions are needed for full quality assurance. In this direction, in recent years several automated solutions have been proposed to account for more than one problem. For instance, Wang et al. used deep learning models to classify B-scans as good, off-center (cropped retina), signal-shielded (poor or no image contrast), and other (problems not covered by previous cases) [39]. The best performing model was a ResNet-50 and obtained an overall accuracy of 96.25%. Additionally, the authors also showed that using this model for quality exclusion improved the performance of diabetic retinopathy detection in the same dataset. More recently, Kauer-Bonin et al. have proposed a method based on modular neural networks [40]. Their approach uses dedicated deep learning models to 1) detect image fixation (macula, disc, other), 2) evaluate if images are aligned, and 3) detect image contrast and cropping problems. The proposed model achieved an accuracy above 96% for all three tasks. In addition, they also describe a fourth algorithm to flag potential segmentation errors based on segmentation plausibility. This work is now part of a commercial solution offer by a company called Nocturne.

These two methods have demonstrated the potential of deep learning for quality assurance. However, relying on neural networks has shortcomings. First, the proposed algorithms cannot be directly reproduced by other researchers unless an entirely new model is trained from scratch. More problematically, deep learning models are known to be very sensitive to distribution shifts, and may not generalize well to images acquired from with different devices and protocols. To overcome this problem, Kauer-Bonin et al. used transfer learning to fine tune a model trained on Heidelberg images to perform well with Zeiss images. However, fine tuning a model to every possible image device and protocol is hard to envision and further evidence is needed to demonstrate that the models work for devices and protocols not used during training.

Feature extraction

Retinal layer thicknesses are the most analyzed OCT features. After segmentation, thickness calculation only involves computing the distance between the boundaries of the layer of interest for each A-scan. Computed thickness values are then used to build an en-face thickness map like the one colored in the left of Figure 2.1. At this point, it is common to add an extra sectorization step where point thickness values are averaged over several predefined sectors. This helps summarize

thickness information into a set of values that are more amenable for analysis. The most common sectorization is the Early Treatment Diabetic Retinopathy Study (ETDRS) [71], which consists of nine sectors arranged in a central region and two rings divided in four angular quadrants (nasal, superior, temporal and inferior) (Figure 2.1).

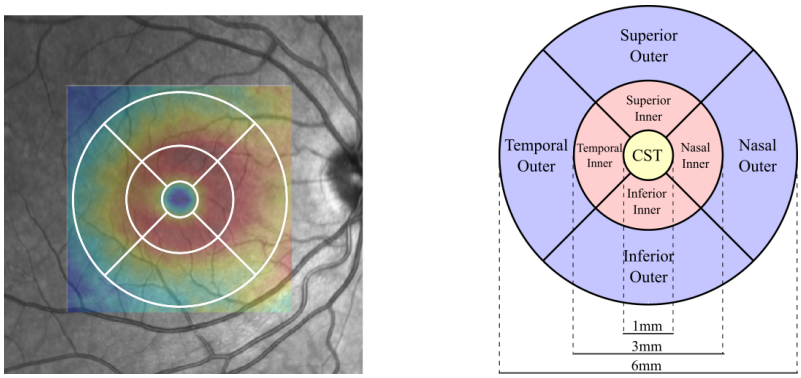


Figure 2.1: ETDRS sectorization of the right eye.. Left: a grayscale fundus image and a thickness map in colors. A thickness map is obtained by computing the thickness of a certain layer after segmenting its boundaries in all OCT images. Right: dimensions and naming convention of ETDRS sectors.

The foveal pit can be quantitatively analyzed by computing features that describe its morphology. These features describe geometrical properties of the fovea such as slope or depth. Alternatively, custom mathematical models can also be employed to model the foveal pit shape and use the fitted equation coefficients as features.

Foveal pit features can be classified into radial and global features. Radial features are computed separately for each angular direction and, as shown in Figure 2.2a, include four types of features:

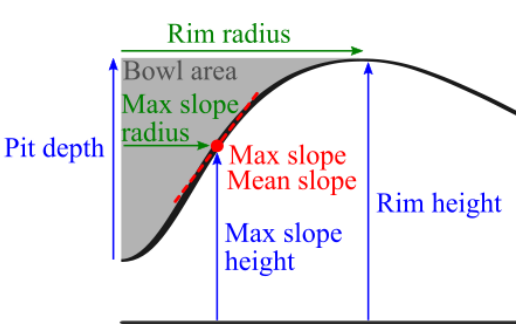
- *Height*: vertical measurements at anatomically meaningful landmarks such as the point of maximum slope or the rim.
- *Width*: horizontal measurements at different foveal heights measured from total retinal thickness (TRT) values.
- *Slope*: measurements of foveal steepness between the foveal center and the rim (e.g., maximum or mean slope).
- *Pit area*: the surface covered by the the foveal pit area.

Combining estimations from different directions, radial maps of variation can be generated (Figure 2.2b). These maps characterize the foveal pit radially visualizing differences between angular directions.

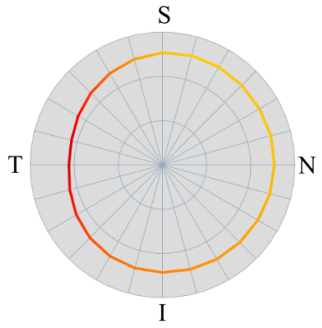
On the other hand, global features describe the foveal pit as a whole with a single value and include:

- *Central foveal thickness (CFT)*: retinal thickness at the foveal center.
- *Disk area*: area of the foveal disk generated at different heights (Figure 2.2c).
- *Pit volume*: measures the volume of the foveal pit.
- *Average values*: obtained by averaging radial parameters across angular directions.

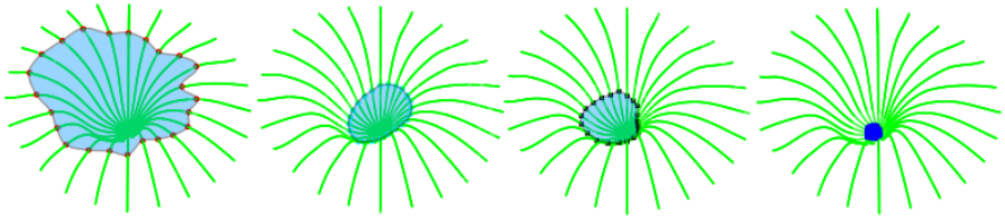
The analysis of the foveal pit involves three main steps: flattening, model fitting and feature calculation. Flattening is achieved by calculating the TRT from previously segmented ILM and BM boundaries. This step sets the BM as a horizontal reference thus removing the difficulty of



(a) Radial geometrical features.



(b) Radial variation map of the rim height.



(c) Foveal disk at different depths[7].

Figure 2.2: Foveal pit morphology features. The geometrical features shown in **a)** can be computed for multiple angular directions to derive a radial map of variation such as the one shown in **b)**. Additionally, features can also derived from the disks formed by the fovea at different heights illustrated in **c)**.

defining width and height in a curved retina.

At this point, to properly analyze the foveal pit in various directions, all B-scans must cover the foveal center. Therefore, if B-scans were not acquired radially, it is necessary to introduce an intermediate interpolation step to transform TRT values into a radial pattern.

The second main step is mathematical model fitting. This step is not strictly necessary to compute geometrical features, but allows to 1) reduce segmentation noise, and 2) use the fitted coefficients of the mathematical model as features. Mathematical models were first applied to study the foveal pit in 2009 by Dubis et al. [73], who used a model based on a difference of Gaussians to analyze the slope, depth, and width of the fovea. After that, several models have been proposed (see Table 2.1). There are two main differences between models: fitted region and mathematical principle. Regarding the former, the fitting process can be performed at different scales. For example, the model proposed by Ding et al. models the whole TRT surface [74], while others fit each B-scan separately [73,75,76]. The latter involves a higher number of fittings but a presumably higher fitting accuracy. There are also models that go beyond and perform the fitting for half a B-scan [77] or even quarter of a B-scan separately [72]. These approaches are able to characterize the foveal pit radially by fitting single angular directions individually [77].

The mathematical principle is an important feature of each model. Due to the convex shape of the fovea, Gaussian curves are behind most models. For instance, the model by Dubis et al. relied on the difference between a narrow and a wide Gaussian [73]. A problematic aspect of this model is that it enforces symmetry and thus is not able to capture well-known foveal pit asymmetries (e.g., nasal and temporal asymmetry) [72]. As an alternative, Ding et al. added a polynomial term and modeled the foveal pit as the subtraction between a second order polynomial surface and a bivariate Gaussian [74]. In a similar vein, Liu et al. proposed a sloped piecemeal Gaussian equation that models each B-scan by a first order polynomial and a Gaussian [75]. The piecemeal nature

[7]Reprinted with permission from [72]. © The Optical Society (2017).

Model	Number of parameters	Modeled region	Mathematical principle	Formula
Dubis, 2009 [73]	6/B-scan	B-scan	Difference of Gaussians	$y(x) = g - a\left[e^{-(x-\mu)^2/\sigma^2} - de^{-(x-\mu)^2/\sigma^2}\right]$
Ding, 2014 [74]	8	2D TRT map	Bivariate Gaussian and 2nd order polynomial surface	$y(x_1, x_2) = A_0 + A_{11}x_1 + A_{21}x_2 + A_{12}x_1^2 + A_{22}x_2^2 + Ke^{\left(-\frac{x_1^2}{2\sigma_1^2} - \frac{x_2^2}{2\sigma_2^2}\right)}$
Scheibe, 2014 [77]	8/B-scan	$\frac{1}{2}$ B-scan	Second derivative of Gaussian	$y(x) = \mu\sigma^2 r^\gamma e^{-\mu r^\gamma} + \alpha(1 - e^{-\mu r^\gamma})$
Liu, 2016 [75]	6/B-scan	B-scan	Sloped piecemeal Gaussian	$y(x) = g - (a\bar{G} + fx)$ $\bar{G} = \begin{cases} e^{-(x-\mu)^2/\sigma^2} & x < \mu - \lambda/2 \\ max(e^{-(x-\mu)^2/\sigma^2}) & \mu - \lambda/2 \leq x \leq \mu + \lambda/2 \\ e^{-(x-\mu)^2/\sigma^2} & x > \mu + \lambda/2 \end{cases}$
Yadav, 2017 [72]	8/B-scan	$\frac{1}{4}$B-scan	Bezier Curves	$Q(t) = \sum_{i=0}^{n} P_{i,n} B_{i,n}(t), 0 \leq t \leq 1$
Breher, 2019 [76]	9/B-scan	B-scan	Sum of three Gaussians	$y(x) = \sum_{i=1}^{3} a_i e^{-(\frac{x-b_i}{c_i})^2}$

Table 2.1: List of mathematical models describing foveal pit morphology. Existing models are compared in terms of the number of parameters of the equation defining the model, the region of the retina modeled, and the underlying mathematical principle. TRT: total retinal thickness.

of this model makes it possible to fit foveas with a flat bottom [75]. Another alternative to model foveal asymmetries is the inclusion of more parameters as in the model presented by Breher et al., based on the sum of three Gaussians [76]. Finally, it is also possible to model each angular direction individually, as it is done in [77] and [72]. The former proposes a four parameter model based on a second derivative of Gaussian where each parameter is related to a certain morphological feature of the fovea. The approach of Yadav et al., on the contrary, uses Bezier Curves to model the foveal pit. This approach aims for a high accuracy but fitted coefficients lack any intuitive meaning [72].

Although there is no comprehensive model comparison in the literature, Liu et al. [75] reported that their model outperformed the one by Dubis et al. [73] by more than $3\,\mu m$ in root mean square error (RMSE). Similarly, Yadav et al. reported an RMSE value of their model of $2.5\,\mu m$, smaller than the $4\,\mu m$ and $8\,\mu m$ measured for Ding et al. [74] and Dubis et al. [73], respectively. Importantly, there has been no research investigating the benefits of mathematical modeling to compute geometrical features of the fovea such as radius and slope.

Texture analysis is a broad topic and spans a large set of features aimed at characterizing high-order local properties of images. In OCT, it can used to analyze en-face thickness maps. Within the large range of texture features, the following methods have been used on OCT images:

- *Gray-level co-occurrence matrix (GLCM)*: pixels in a gray-scale image are compared to their neighbors to build a GLCM matrix. This matrix is effectively a 2D histogram representing the prevalence of each pair of values. From that matrix, several features can be computed, the most common being Haralick features [78].
- *Local binary patterns (LBP)*: for every neighbor of a pixel, a binary value of 0 or 1 is computed if the value of the pixel of interest is larger or smaller than that of its neighbor. These binary digits are used to build a histogram from which to compute classical statistical measures such as mean, standard deviation, or skewness [79].
- *Wavelet analysis*: an image is decomposed into smaller components by means of the two-dimensional discrete wavelet transform. The estimated coefficients of its transformation can be used as features describing image properties at different scales [80].
- *Fractal analysis*: fractal features try to describe the complexity of the thickness map surface. While fractal dimension is related to roughness [81], lacunarity [82] describes the distribution of gaps in the thickness pattern.

It is worth mentioning that, although texture analysis is almost always applied to en-face thickness maps, 3D volumes can also be studied by texture features, for instance, to detect fluid-filled regions in macular images [83].

OCT analysis software

Computational OCT image analysis is a complex process that involves multiple steps between image acquisition and formal analysis. To facilitate the process, it is key to have high-quality software libraries that can be used to perform tasks such as file reading, preprocessing, and segmentation. Unlike in other imaging modalities such as magnetic resonance imaging (MRI), there are very few openly available libraries for OCT image analysis, and existing tools have a strong focus on retinal layer segmentation rather than building end-to-end analysis pipelines. Existing software for OCT analysis is listed in Table 2.2.

One of the most used tools is OCT Explorer [26, 87–89], a graphical user interface (GUI) developed by the Iowa Institute for Biomedical Imaging. The GUI incorporates a layer segmentation algorithm and works with multiple file formats. Nevertheless, generated output is limited to segmentation values and computing additional features is left to the user. Moreover, the code is not directly accessible via an application programming interface and, therefore, it cannot be used to build custom pipelines.

Software	Type	Last release	Supported formats			
			Heidelberg	Zeiss	Topcon	RGB*
ReLayNet [84]	Algorithm	2018				✓
Dufour [85, 86]	GUI (.exe)	2012	✓			
OCT Explorer [26, 87–89]		2017	✓	✓	✓	✓
AMPAO [90]	GUI (MATLAB)	2019	✓	✓		
Caserel [91]		2021				✓
OCTSEG [92]		2016	✓			✓
Livelayer [93, 94]		2021	✓			
AURA tools [29, 95]	MATLAB toolbox	2015	✓	✓		

*Image formats such as tiff or jpg.

Table 2.2: Openly available OCT analysis software tools. The table only lists software libraries that can be freely downloaded. The last release date refers to the date in which the last major update was made to the code. GUI: graphical user interface.

Another important software is AUtomated Retinal Analysis tools (AURA tools), a MATLAB toolbox that incorporates a boundary classification segmentation algorithm as well as basic thickness calculation capabilities [29, 95]. As a limitation, it only supports *vol* (Heidelberg) and *img* (Zeiss) file formats, and it does not have a comprehensive documentation explaining how to use it. More recently, Kafieh et al. released another MATLAB software called Automatic Multifaceted Matlab Package for Analysis of Ocular Images (AMPAO) that includes a GUI as well as several functions for file reading, denoising, and feature extraction [90]. The proposed software was released in bulk within the paper without extensive documentation. In addition to these, there are other less used software libraries developed exclusively for retinal layer segmentation [84, 85, 91–93].

Overall, existing software solutions are limited in terms of supported features. Except for OCT Explorer, there is no software able to read all the main proprietary file formats. Moreover, none of the existing libraries includes a full range of functions for file reading, preprocessing, and feature extraction that are documented in detail. More worryingly, most of the tools are not actively maintained, which may cause them to become obsolete and cease to function properly in the near future.

2.2 Retinal changes in healthy subjects

Although a big part of OCT research is focused on pathological retina, it is also of interest to study how the retina is structured in healthy subjects. At the microscopic level, this is typically conducted using histology, which provides high-resolution data but is limited in terms of sample size and the area covered. These limitations are overcome by OCT images, which are an efficient method of analyzing greater regions of the retina in a larger cohort of subjects. Based on OCT, it is possible to investigate how the retinal structure evolves along lifespan and how it varies as a function of other demographic factors such as sex or ethnicity. This kind of research aims to establish a normative description of the retina in order to better understand pathologies. The following section describes previous research investigating the effects of age and sex on retinal thickness and foveal pit morphology.

Thickness

Multiple studies have analyzed the effect of aging on the TRT. A systematic review of 49 studies concluded that the TRT decreases with age [96]. Moreover, this thinning effect is not uniform

across the macula: while it is more noticeable for eccentricities larger than 0.5 mm, the central region appears to remain unchanged or slightly thickened as a consequence of age [96]. A similar thinning pattern was also observed in a study of 40000 subjects from the UK Biobank (UKBB) dataset, which measured a central and whole macular annual thinning rates of -0.17 $\mu m/year$ and -0.19 $\mu m/year$, respectively [97]. More recent publications have also demonstrated an overall thinning in both inner and outer rings of the macula [98–101]. The estimated thinning rates range from -0.19 $\mu m/year$ [99] to -0.61 $\mu m/year$ [101].

Of greater interest than the TRT analysis, is the determination of the individual layers—with specific cellular architectures and spatial distributions—that drive the thinning effect. Regarding the innermost layers of the retina, histological studies have observed an age-related decrease in the number of fibers that conform the optic nerve and the NFL [102,103]. However, not all OCT studies analyzing the NFL agree with these findings. While some studies have reported an NFL thinning in the macula [104–106], others have observed no statistically significant changes [99,100], or even thickening [98,107]. Among these, the largest study reported a small yet statistically significant thinning rate of -0.06 $\mu m/year$ [106]. These discrepancies might be explained by the very low thickness of the NFL in the macular region, which may be more susceptible to segmentation errors. In fact, studies examining the NFL in the optic nerve region have shown more robust results. For instance, Demirkaya et al. observed a significant negative correlation ($r = $-0.3) of peripapillary NFL (pNFL) thickness and age [108], a finding also found by other cross-sectional and longitudinal studies [109,110]. The longitudinal study reported a loss of -0.14 $\mu m/year$ equivalent to a 0.61% thickness loss per year [110].

Regarding the GCL, histological evidence points to an age-related decrease of ganglion cell density [103], which is in agreement with most OCT studies [104–106,108]. However, a few studies reported that there is no statistically significant age-related effect [111,112]. Observed thinning rates range from -0.07 $\mu m/year$ to -0.33 $\mu m/year$. Differences between macular regions do exist, being the inner and outer rings much more affected than the central region. In this regard, several authors suggest that these regional differences only reflect differences in absolute GCL thickness [113,114]. More concretely, they reported a 0.28% decrease in GCL thickness uniform across the whole macula. Despite that, results of a longitudinal study did observe differences in the relative thinning rates of the ganglion cell-inner plexiform layer (GCIPL) [110].

The IPL parallels GCL results closely. In fact, many studies chose to analyze both layers as a single complex (GCIPL). Measured thinning rates range from -0.33 $\mu m/year$ [98] to -0.15 $\mu m/year$ from a study including 42044 UKBB subjects [106]. Nevertheless, no effect was found in [115].

As an important remark, most studies rely on the assumption of linearity to perform both regression and correlation analyses and do not report any check of the validity of this assumption. However, as stated in [113], a segmented linear regression fits the data better, and GCL thickness might not start to decline until the 40s. The lack of long-term longitudinal studies, however, makes it difficult to conclude on this matter.

As for the INL, previous work has also reported a thinning effect [98,99,101,116]. Although outermost layers of the retina have been less studied, reported results point to a thickening effect [96].

The effect of sex on the retinal structure has also been investigated. Several studies have found a thicker retina in men [97,99,117,118]. Differences in average TRT range from 2.7 μm [97] to 7.1 μm [100]. In all studies the differences were higher for the inner macular ring and diminished for the outer ring, where a large cohort study even measured a 0.1 μm greater TRT in females [97].

As with the aging effect, it is of interest to analyze which layers might explain differences in TRT. In this respect, the macular NFL has been studied with diverse results. Some studies have measured a thicker NFL in males [99,101,105,119], while others have observed it to be thicker in females [100,104,106].

Regarding the GCL, most studies observed a higher thickness in males [99–101,105,119]. How-

ever, a study on 42000 subjects measured a $1\,\mu m$ thicker GCL in females [106]. Results for the IPL layer are similar. While several studies showed a thicker IPL in males [100, 101, 105], others observed it to be thinner [106]. The few studies analyzing the INL measured it to be thicker in males [100]. Similarly, sex differences have been reported for the outer retinal layers [100, 105].

In addition to sectorized thickness, Nunes et al. showed that texture can also help discriminate female and male retina [120]. More concretely, GLCM and wavelet features of the GCL, IPL, OPL, and ONL were able to discriminate sexes. This may indicate that sexual differences exist beyond average thickness.

As pointed out by previous work, thickness measurements are influenced by the axial length based on the so-called ocular magnification problem [121]. In longer eyes, light entering the eye travels a longer distance, which increases the separation between adjacent A-scans (i.e. the lateral scale) [122]. As a consequence, each eye has a specific lateral scale that must be considered when computing sectorized thickness values. However, not all commercial scanners correct it automatically and the fact that females have a slightly shorter axial length [123] may introduce a bias in sex comparisons.

As an important limitation, most studies studying either age or sex have relied on the standard ETDRS sectorization [97–101, 105, 106, 112, 117, 118]. This reduces the description of the macula to nine values and might undermine the ability to describe age and sex effects with high spatial detail. As an alternative, recent studies have begun to use smaller sectors. For instance, an 8 x 8 square grid has been used to determine clustered spatial patterns of age changes and establish a normative database of macular thickness in [119, 124]. Additionally, a radial grid with 61 sectors has also been used to examine age changes in retinal thickness [116].

Foveal pit morphology

Although to a lesser extent than retinal thickness, several studies have analyzed foveal morphology in healthy populations. In a study investigating normative foveal pit shape, Scheibe et al. evidenced the asymmetric shape of the foveal pit [125]. The nasal sector has both the widest radius and the highest rim, while superior and inferior sectors show the highest slope and the smallest radius. These differences are a direct consequence of the presence of the optic nerve and the distribution of nerve fibers (Figure 2.3).

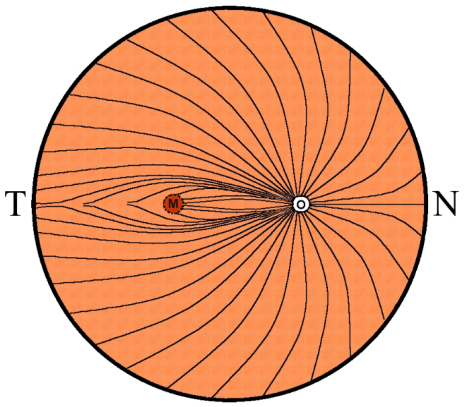


Figure 2.3: NFL fiber distribution[8]. The macula and the optic disk are depicted in red and white, respectively. Temporal and nasal sides are encoded from left to right. Black lines depict the direction of NFL fibers running from the optic nerve to each retinal region.

[8] Adapted from *Pattern of Retinal Nerve Fibers* by Maria Sieglinda von Nudeldorf. CC BY-SA 3.0 license.

In addition, foveal morphology has also been related to other macular features. For instance, Barak et al. found that foveae with steeper slope, less symmetry and higher mathematical modeling complexity are associated with a higher risk of developing macular holes [126]. On the other hand, Dubis et al. reported that the pit depth and diameter are two independent factors that influence the size of the foveal avascular zone [127]. The weak correlation between pit depth and diameter might point to different factors driving the development of the foveal pit.

On another note, only a few studies have investigated the effect age on foveal morphology (see Table 2.3). A direct comparison between studies is complicated by the diversity of metrics and processing steps among studies. Tick et al. did not find any correlation between age and foveal morphology [128]. However, only 57 subjects of 45 years-old or younger were included in the study so any effects after that age were not assessed. In a larger study Nesmith et al. found age to be associated with an increased maximal slope, a higher mathematical complexity, and a decrease in foveal volume [129]. Contrary to these results, in a study including only subjects older than 40 years-old, slope was found to decrease with age [130]. These discrepancies might be explained by differences in the slope calculation procedure. In fact, while in [129] slope was measured in 50 μm eccentricity steps, in [130] it was calculated for a predefined 500 μm eccentricity. Moreover, neither of these studies used a mathematical model to fit the foveal surface. In a later study with only 30 subjects [131], a model presented in [77] was used but no age effect was observed. However, foveal width was estimated to decrease with age although the differences were not statistical significant. Finally, a study that mixed ghanaians and caucasian subjects did observe changes in the foveal pit associated with age, specially in ghanaians [132].

Regarding sex, studies that have compared both male and female fovea are listed in Table 2.4. This topic was addressed for the first time in 2001 by Wagner-Schuman et al. [133]. In this reference work, the model proposed by Dubis et al. [73] was used to compare foveal depth, width and slope between males and females. Observed differences, despite not reaching statistical significance, estimated a sharper, narrower, and deeper foveal pit in males. In a posterior study using the same model, similar slope differences were observed [127]. However, diverging results were reported as well, as males were found to have a much larger pit depth and a wider fovea. The small sample size of these studies might explain those differences. In the most exhaustive study so far, Scheibe et al. used their previously proposed model to study a larger cohort of 109 males and 111 females [125]. In this case, clear differences were observed regarding height, width, slope, and area, pointing to a sharper foveal pit in males. Results in this direction were also recently found by Zouache et al. [132]. More recently, a large UKBB study conducted by Olvera-Barrios et al. also found sex differences in foveal curvature [134].

As with thickness measurements, the axial length can have a confounding effect on foveal pit morphology metrics [133]. In fact, it has been reported to affect horizontal metrics such as pit diameter [128]. However, not all studies assessed that factor equally. While some studies corrected the lateral scale based on each subject's axial length [127, 133], no explicit correction was reported in others [125, 132]. Alternatively, Zouache et al. carried out a Monte Carlo simulation to ensure the robustness of the results despite not having performed any axial length correction [132].

2.3 Retinal changes in Parkinson's disease

The measurement of retinal changes in PD aims to find out if the well-known brain neuronal death caused by PD is also present in the retina. This is supported by both histological and physiological studies that have found a retinal involvement in PD [57].

OCT images were first used to study PD in 2004 by Inzelberg et al. [135]. In that work, a significant pNFL thinning was reported. Those results motivated a large number of studies investigating retinal thickness in PD. These studies have been aggregated by several meta-analyses concluding that there is a thinning of the TRT, GCIPL, and pNFL in PD [41, 136–139]. More concretely, the estimated standardized mean differences (i.e., Cohen's d) are: -0.22 (TRT), -0.53

Study	Group	N	Age		CFT	Rim height	Pit depth (μm/10 years)	Rim radius (μm/10 years)	Slope (°/10 years)
			Mean± σ	Range					
Tick, 2011 [128]		57	-	18-45	NAF	NAF	NAF	NAF	-
Nesmith, 2014 [129]		390	52.6	13-97	-	-	-	-	Increase*
Gella, 2015 [130]		668		≥ 40	-	-	-	-	Decrease*
Sepulveda, 2016 [131]	young old	20 10		24-33 62-76	NAF	NAF	-	Decrease	-
Zouache, 2020 [132]	ghanaian caucasian	84 37	65.1 ± 11.5 56.0 ± 8.0	41-85 45-82	- -	- -	-4.8 -1.5	-46* +30	-0.054* -0.012
Olvera-Barrios, 2022 [134]		63939	56 ± 8.0	40-69	-	-	-	-	Increase in females*

* statistically significant ($p < 0.05$)

Table 2.3: List of studies linking age and foveal morphology. For each study the sample size (N), the age range of the participants, and the estimated age effect for several foveal features are shown. CFT: central foveal thickness. NAF: no association found.

Study	Group	N_{male}	N_{female}	Age		CFT	Rim height	Pit depth	Rim radius	Slope
				Mean ± σ	Range	(μm)	(μm)	(μm)	(μm)	(°)
Wagner-Schuman, 2011 [133]		47	43	27.8 ± 9.0	≥ 18	-	-	+1	−15	+0.4*
Dubis, 2012 [127],2012		26	16	26.5	18-67	-	-	+13	+79	+0.5
Scheibe, 2016 [125]		109	111	44 ± 13	21-77	+4.2	+6.0*	-	−27.5*	+0.5*
Zouache, 2020 [132]	ghanaian caucasian	30 9	54 28	65.1 ± 9.4 61.9 ± 11.5	45-82 41-85	- -	- -	+7 +10	-145 -180	+2.26 +1.05
Olvera-Barrios, 2022 [134]	28842	35097		56 ± 8.0	40-69	+6.8*	-	-	-	+0.87*

* statistically significant ($p < 0.05$)

Table 2.4: List of studies linking sex and foveal morphology. For each study the sample size (N), the sex distribution of the participants, and the estimated age effect for several foveal features are shown. CFT: central foveal thickness. NAF: no association found.

(pNFL), and -0.43 (GCIPL) [139]. These differences correspond to small-medium effect sizes and, therefore, it would be difficult to spot them with a naked eye. The TRT thinning has been found in all macular sectors, suggesting that the effect is not limited to a single small region. Nevertheless, differences appear to be smaller in the central 0.5 mm radius circular region, where inner retinal layers are not present [41]. In fact, GCIPL thinning is present in both inner and outer macular rings [137]. The pNFL, however, appears to be more clearly affected in superior, inferior and temporal sectors rather than nasal [41]. Existing meta-analyses could not conclude on the INL and outer retinal layers, mostly because of the scarcity of studies investigating these layers.

Despite the meta-analytical evidence, discrepancies between studies do exist. Not all studies measured the same effect size nor agreed on the retinal region with more prominent changes [140,141]. Observed discrepancies might be attributed to the heterogeneity of the literature, where studies exhibit differences regarding scanning device, inclusion criteria, racial differences, and age ranges [140]. Moreover, the cross sectional nature of most of these studies prevents from reaching definite conclusions out of their findings.

Trying to address those limitations, a few longitudinal studies have already been carried out [142–145] (see Table 2.5). Although with different thinning rate estimations, all studies investigating the TRT found a higher thinning rate in patients. On the other hand, a pNFL thinning effect was observed in [142,145] but not in [144]. This last study, however, had a shorter follow-up time and a smaller dataset. On the contrary, both [145] and [142]—with a large follow-up time—measured pNFL thinning rates up to four times higher in patients. Finally, the only study that analyzed the GCIPL [145] concluded that patients lose thickness three times faster (-0.67 $\mu m/year$ vs -0.23 $\mu m/year$) than healthy control (HC) subjects. These differences were higher in the parafoveal region (i.e., 1-3 mm inner ring).

Study	**Layer**	**Subjects**		**Follow-up (years)**	**Rate of change ($\mu m/year$)**	
		HC	PD		HC	PD
Satue, 2017 [142]	TRT	30	30	5	−0.36	−1.28
Ma, 2018 [143]		-	22	2.5	-	−2.80
Hasanov, 2019 [144]		19	19	1.75	+0.94	−4.63
Satue, 2017 [142]	pNFL	30	30	5	−0.28	−0.50
Ma, 2018 [143]		-	22	2.5	-	−2.40
Hasanov, 2019 [144]		19	19	1.75	−0.98	+0.84
Murueta-Goyena, 2020 [145]		17	50	3	−0.15	−0.55
Murueta-Goyena, 2020 [145]	GCIPL	17	50	3	−0.23	−0.67

Table 2.5: List of longitudinal studies linking OCT and PD. Studies are grouped by the investigated retinal layer. The number of subjects was relatively small in all studies with a maximum of 50 patients. HC: healthy control. PD: Parkinson's disease. TRT: total retinal thickness. pNFL: peripapillary nerve fiber layer. GCIPL: ganglion cell-inner plexiform layer.

Although a big part of previous work has relied on inference, there has also been research trying to develop diagnostic models based on retinal features. Here, rather than concluding if an effect is present or not, the aim is to build a classification model based on OCT features. Existing studies of this kind are listed in Table 2.6. One of the first models of this kind was presented by Garcia-Martin et al. in 2014 [146]. In their work, the authors trained a linear discriminant analysis (LDA) model on pNFL and TRT thickness features that achieved an area under the curve (AUC) of 0.902. Later studies, however, were not able to achieve such a good performance despite using similar features [147–150]. For instance, in [147] and [148], thickness features only showed classification power when combined with electrophysiology and visual function, respectively. Alternatively, in [149], thickness features reached an AUC of 0.718 that was further improved to 0.849 when adding OCT angiography (OCTA) vasculature features. The most recent study reported an AUC of 0.796 using thickness features and linear discriminant analysis [150].

Study	Subjects		Features	Model	Accuracy (%)	AUC
	HC	PD				
Garcia-Martin, 2014 [146]	200	111	Thickness	LDA	-	0.90
Ding, 2014 [74]	27	27	Foveal pit	LR	74	0.70
Slotnick, 2015 [154]	72	24	Foveal pit	PCA	65	0.65
Miri, 2016 [147]	8	10	Thickness	LR	-	0.43
Huang, 2018 [148]	41	53	Thickness	ROC curve	-	0.40
Nunes, 2019 [157]	27	28	Thickness Reflectance	SVM	> 90	-
Young, 2019 [155]	40	33	Thickness Foveal pit	LR	-	0.80*
Pinkhardt, 2020 [156]	176	184	Foveal pit	ROC curve	62	0.60
Zou, 2020 [149]	35	35	Thickness	ROC curve	-	0.72
Satue, 2022 [150]	146	42	Thickness	-	80	

* including sex

Table 2.6: List of PD diagnosis models based on OCT features. or each model, the sample size, the type of features, the model, and the reported results are described. LDA: linear discriminant analysis. LR: logistic regression. PCA: principal component analysis. ROC: receiver operating characteristic. SVM: support vector machine.

In addition to thickness features, the foveal pit morphology might provide valuable information in PD. This hypothesis is supported by the thickness decrease observed in the literature. As a consequence of this thinning, it has been proposed that the foveal pit is remodeled in PD, resulting in thinner and wider pits [151, 152]. Furthermore, the visual impairment associated with the disease [153] might be related with changes in the foveal structure, which is located at the center of the vision field.

To date, there have been a few works exploring foveal pit features in PD. Ding et al. presented a 2D model based on a polynomial surface and a bivariate Gaussian [74]. Adjusted parameters of the model were used for classification, reporting a 26% error rate using seven model parameters and logistic regression (LR). A subsequent study carried out by the same research group applied the same model on a new dataset this time achieving a minimum error rate of 33% and an AUC of 0.7 [154]. It must be pointed out that no cross-validation procedure was applied in either studies and thus reported results might be overoptimistic. More recently, Young et al. applied the mathematical model developed by Dubis et al. [73] to calculate metrics of slope, width and height [155]. Group comparisons, however, did not yield any significant result and the performance of a diagnostic model only improved when adding sex as a feature. Similarly, Pinkhardt et al. used the model proposed by Ding et al. in a larger dataset and obtained a best accuracy of 61.86%, not fully reproducing the original results [156].

As an alternative to structural features, Nunes et al. used texture features derived from thickness and reflectance en-face maps to train support vector machine (SVM) diagnostic models for PD and AD [157]. More concretely, the authors used GLCM and wavelet features to train several SVM models achieving an accuracy of 80.8%. Similar approaches have also been applied in other neurodegenerative diseases with good results. For instance, GLCM, wavelet, and LBP features have shown to differentiate well between MS patients and HC SUBJECTS [158, 159]. Similarly, fractal dimension has already been used in AD [160]. Finally, lacunarity has also been used in OCTA analysis in PD [161].

2.4 Linking OCT and disease severity

In addition to disease diagnosis, the greatest potential of OCT lies in using it as a biomarker to assess and monitor the disease. That is, to use OCT to support clinical decisions regarding a

patient. Here two main use cases can be distinguished: 1) clinical assessment, to determine the disease severity of a patient, and 2) prognosis, to predict the clinical evolution of the disease. In both cases, clinical variables related with motor impairment, cognition, and visual impairment need to be measured, and their relationship with OCT features needs to be demonstrated.

Here it is important to describe the main clinical tests used to assess the status of a patient. First, the Unified Parkinson's Disease Rating Scale (UPDRS) comprises of a set of standard tests evaluating motor symptoms, mental health, behavior, and complications of therapy in PD patients [162]. From this battery of tests, the third part (UPDRS-III) focuses on motor impairment and is used as a measure of motor disability. Alternatively, motor impairment can also be evaluated by the Hoehn-Yahr (HY) scale, which classifies PD patients according to the following stages: 1 (unilateral affection), 2 (bilateral affection without balance problems), 3 (bilateral affection with balance problems), 4 (severe impairment), and 5 (confinement to bed or wheelchair) [163]. As for cognition, a common test is the Montreal Cognitive Assessment (MoCA), a cognitive impairment screening test that assesses attention, memory, language, visuospatial abilities, executive functions, and orientation to time and place [164]. The test provides a result in a scale of 0 to 30 and was designed as an screening tool for cognitive impairment. In fact, a MoCA score below 26 has been used as cut-off value for mild cognitive impairment in the literature [165]. Similarly, even lower values of MoCA can be used as an indicator for dementia (MoCA < 22). The choice of cut-off values results in different sensitivity and specificity values and establishing normative thresholds for different groups of subjects is a research topic in itself [165, 166].

Several works have tried to relate thickness measurements to clinical features. In the first series of OCT studies, Altintaş et al.—in a study of 17 PD patients—reported an inverse correlation of -0.66 between foveal TRT and UPDRS-III [167]. Later on, Jiménez et al. measured a similarly strong relationship between UPDRS III and pNFL thickness, and proposed an estimation formula based on it [168]. In much larger databases, an inverse relationship between HY score and the thickness of both pNFL [169] and GCL [146] has also been found. The relationship between retinal thickness and HY has been subsequently reported for the TRT [142,170,171], pNFL [138,152,170–172], GCIPL [138, 145, 170–172], and more recently, the INL and OPL [171]. The latter results were derived by Wang et al. in a total of 397 PD patients, the largest study of this kind to date. Nevertheless, there are also studies that have not been able to replicate these results [156,173,174].

In addition to motor impairment, several works have linked OCT features and cognitive function. For instance, a positive relationship between scores in MoCA and thickness features has been reported for the TRT [174], pNFL [175], and GCIPL [145, 174]. In the large dataset mentioned before, Wang et al. found a similar positive relationship between cognition evaluated by the Mini-Mental State Examination and pNFL, GCL, and photoreceptor layer [171]. On the other hand, associations between retinal thinning and visual function have also been reported by a few studies. For example, in [176] the authors found GCIPL parafoveal thickness to be a good predictor for visual acuity. Moreover, baseline GCIPL thickness was found to be a good discrminator of visual impairment. In a similar fashion, Pinkhardt et al. also found a correlation between the restructuring of the fovea and visual contrast sensitivity loss [156].

Perhaps the most interesting use of a biomarker is disease prognosis. Predicting the evolution of a patient is key for deciding treatment. The development of prognostic models is challenging as it requires longitudinal data to prove their reliability, which is especially difficult in diseases with a reduction in life expectancy. Moreover, the evolution of a disease is rarely determined by a single parameter and establishing the causal mechanisms behind disease progression is not straightforward. Despite this difficulties, there is some evidence of the potential of OCT for PD prognosis. First, Satue et al. reported a moderate association between NFL thickness changes and worsening of visual function and motor impairment [142]. Additionally, Murueta-Goyena et al. recently showed that baseline retinal thickness can predict cognitive decline in PD [145]. More concretely, patients with lower parafoveal GCIPL and pNFL thickness had a relative risk of 3.49 and 3.28 of cognitive decline, respectively. These effects, however, were not found for motor impairment.

2.5 Critical analysis of the state of the art

After surveying the existing literature, here we summarize and identify existing gaps of knowledge regarding OCT image analysis and its application to research with healthy individuals and PD patients.

First, although retinal OCT has existed for some decades, certain technical parts of OCT image processing are not fully developed. There is no standard end-to-end analysis pipeline, and existing software tools have limitations: most of them only work for specific devices and features, and are not actively maintained. Moreover, existing algorithms for automatic foveal location and quality assurance are difficult to implement and seldom used by the community. Similarly, multiple approaches to analyze the foveal pit morphology have been proposed, but there has been no research into evaluating which is the most appropriate methodology. New research should target these issues.

On another note, retinal layer thickness has been extensively studied in healthy populations. Based on current evidence, inner retinal layer thickness decreases with aging and the male retina is slightly thicker. However, most studies relied on the ETDRS sectorization, which limits the description of the macula to only nine regions. Increasing the number of sectors could substantially improve the understanding of age and sex effects on the retina by visualizing a more detailed pattern.

Additionally, foveal pit morphology analysis is an emerging OCT image analysis technique that, despite its potential, has received little practical attention. The relationship between the foveal pit and demographic factors such as age and sex has been little investigated and is not fully understood. Existing works point towards a sharper and narrower foveal pit in males, but more studies are needed to confirm this relationship and explore age effects.

Regarding the study of PD, there has been a large effort on analyzing retinal layer thicknesses, and current evidence supports a PD-related thinning effect of the TRT, pNFL, and GCIPL. Nevertheless, there is a large heterogeneity across studies and a lack of consistency regarding the specific region affected in PD. Notably, retinal thinning has also been reported in AD [177] and MS [140] and it is crucial to determine specific hallmarks related to PD that could provide a differential diagnosis. In spite of this need, most studies relied on small sample sizes and focused on a relatively small set of retinal features, namely thickness values averaged over the ETDRS grid. Although this choice simplifies the analysis, it may mask smaller changes happening in early disease stages. Moreover, there may be features with a higher discriminative power that have not been fully explored yet. In this sense, there is some evidence pointing to foveal pit and texture features as an alternative, but new research is needed to evaluate their potential.

The relationship between OCT features and disease severity is not entirely established. Several works have measured a correlation between retinal thicknesses and motor impairment, cognitive function, and visual impairment. However, not all the studies agreed on the findings, and it is not clear if the measured effect is sufficient to be reliably used in a clinical setting. To overcome this, new research should evaluate conventional and novel OCT features to determine the potential of OCT for disease assessment.

Finally, very few longitudinal studies have investigated changes in the retina of PD patients. Preliminary studies support an accentuated thinning rate in patients, but are limited in terms of sample size and the number of retinal features explored. In fact, longitudinal research so far has only explored changes in a few retinal layers without looking into other retinal features. More importantly, it is not clear whether longitudinal changes in the retina hold information about disease progression. Larger and richer longitudinal datasets could shed light into this matter.

2.6 Objectives of the thesis

The main goal of this doctoral research project is to evaluate the potential of advanced OCT retinal features as a biomarker for PD. This is, to study how and to what extent OCT images can provide valuable information on disease status and progression beyond what has been shown to date.

In pursuit of this objective, it is also essential to advance in both the way in which OCT images are analyzed, and our understanding of the link between the retina and demographic factors. Including these two aspects, the scope of the thesis is further defined by the following objectives:

1. **Improve OCT processing algorithms**: this includes four smaller objectives aimed at establishing the best strategies to develop a robust OCT feature extraction pipeline:
   - Build an OCT quality assessment model that flags an image volume to be excluded in presence of poor image contrast, ocular lesions, or artifacts.
   - Determine the best method for an accurate automatic foveal location.
   - Compare existing methodologies to analyze foveal pit morphology from a quantitative point of view.
   - Develop an open-source OCT image analysis toolbox capable of reading different proprietary file formats and extracting a comprehensive set of retinal features.
   - Apply the developed methods to compute a comprehensive set of features from the datasets used in this work.
2. **Describe the effect of age and sex in the macula with high spatial detail**: analyze how retinal thicknesses and foveal pit morphology vary as a function of age and sex. This provides the baseline knowledge needed to differentiate the normal evolution of the retina from abnormal changes caused by PD. The improvement over previous work lies in the use of smaller macular sectors and a radial foveal analysis.
3. **Evaluate the potential of advanced OCT features for PD evaluation**: this involves evaluating a comprehensive set of OCT features including both conventional (i.e., ETDRS thickness) and novel features (e.g., foveal shape and texture analysis) for two tasks:
   - *Diagnosis*: training and evaluation of PD diagnostic models based on OCT features.
   - *Severity assessment*: using OCT as a biomarker for cognitive function and motor impairment in a regression setting.
4. **Evaluate OCT as a tool for disease monitoring in PD**: the first goal is to identify the retinal layers and regions primarily affected in PD by measuring longitudinal retinal changes as the disease evolves. Additionally, we seek to assess the potential of OCT features as a reliable tool for monitoring disease progression by establishing a link between retinal changes measured by OCT and the clinical progression of patients

The next part of the thesis is structured as follow: Chapter 3 describes the datasets used throughout this research work. Then, Chapters 4, 5, 6, and 7 describe the work carried out to address each of the aforementioned objectives. Finally, Chapter 8 summarizes the main contributions and conclusions.

3 | Databases

Sections 3.1 and 3.2 provide an overview of the two databases used throughout this research work: Biobizkaia Health Research Institute (BHRI) and AlzEye. Then, the specific data subsets used in each analysis are described in Section 3.3.

3.1 Biobizkaia Health Research Institute

Over the past decade, extensive efforts have been made to collect OCT images at BHRI in Barakaldo, Spain. Subject recruitment and image acquisition have been funded through several research projects, each with a specific duration and cohort type. A summary of the total number of PD patients and HC subjects imaged can be found in Table 3.1.

Project	**Duration**	**Center**	**Device**	**Subjects**	
				HC	PD
Michael J. Fox	2015 - 2018	BHRI	Heidelberg	37	49
Virtual Games	2018 - 2018	BHRI	Heidelberg	10	4
EiTB maratoia	2018 - present	BHRI	Heidelberg	439	26
Begipark	2020 - present	BHRI	Heidelberg	42	56
		Bioaraba	Heidelberg	-	27
		Biodonostia	Topcon	-	26
Total				528	188

Table 3.1: Subjects enrolled in each BHRI project. Each project was funded separately and included a different number of subjects. Notably, *Begipark* included images acquired at three hospitals.

The first project was funded by the Michael J. Fox Foundation and aimed at building the largest longitudinal dataset of OCT images of PD patients. Later, in the *Virtual Games* project a small cohort of patients and controls was also acquired. Additionally, *EiTB maratoia* project received public funding to build a large dataset of controls. Ongoing efforts are now focused on *Begipark*, a multi-centric project with PD patients recruited in three health research centers of the Basque Country: Biobizkaia, Bioaraba, and Biodonostia.

The aggregated database includes a total of 528 controls and 188 PD patients. Healthy controls were 62.4% female and had an age of 55.4 ± 12.4 years in range [21, 88]. Conversely, PD patients were older (64.8 ± 8.6 [41, 80]) and only 35.6% female. The average disease duration was 5.9 ± 4.5 with a maximum of 22.8 years.

In addition to the cross-sectional dataset, a total of 169 subjects were followed-up longitudinally (Table 3.2). The average follow-up time was 2.89 years with a maximum of 5.7 years for subjects from the *Michael J. Fox* project. Importantly, subjects from this project were scheduled to be

Group	Baseline			Longitudinal		
	N	Age* (years)	Male (%)	N	Follow time (years)*	Visits*
Control	528	55.4 ± 12.4 [21, 88]	37.6	58	2.8 ± 1.7 [1, 5]	2.4 ± 0.5 [2, 3]
Parkinson	188	64.8 ± 8.6 [41, 80]	64.4	111	2.6 ± 1.7 [1, 5]	2.4 ± 0.5 [2, 3]

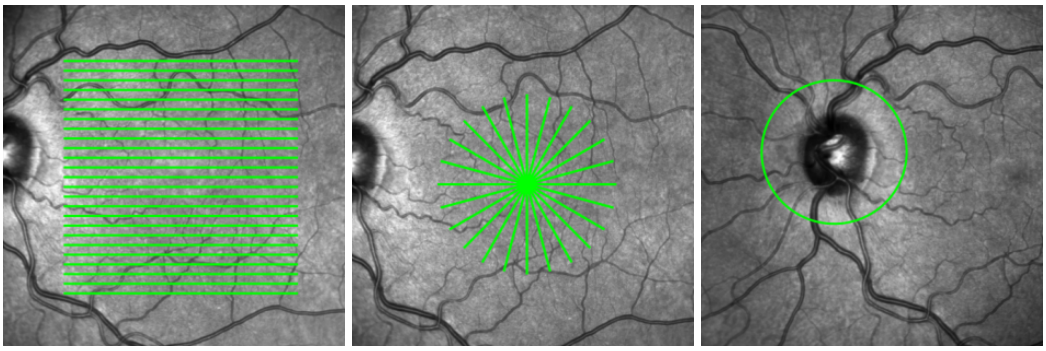
* mean ± σ [range].

Table 3.2: Subjects with OCT imaging in BHRI dataset. The table shows first the age and sex statistics of the subjects with at least one visit (Baseline). From that dataset some subjects (N) were imaged more than once and followed-up more than two and a half years on average.

imaged three times: baseline, year 3, and year 5.

All subjects were Caucasian and underwent a screening process that consisted of an ophthalmological examination and a comprehensive questionnaire on neurological, systemic, and eye-related diseases. Exclusion criteria were: history of heavy smoking (more than 20 cigarettes/day), heavy alcohol use (more than 4 drinks/day for men or 3 drinks/day for women), diagnosis of any type or grade of diabetes, uncontrolled or resistant elevated blood pressure, obesity (body mass index higher than 30), history of consumption of drugs or medications known to induce retinal toxicity, chronic inflammatory systemic diseases, history of traumatic brain injury, or neurological diseases. Additionally, subjects with spherical equivalent refractive error higher to 4.00 diopters, lower than -4.00 diopters, and higher than 3.00 diopters of astigmatism were excluded. In cases where only one of the eyes of a participant was excluded, the other eye was included. Following the tenets of the Declaration of Helsinki, all participants gave written informed consent prior to their participation.

In acquisitions made with Heidelberg devices (Heidelberg Engineering, Heidelberg, Germany), three main OCT acquisition protocols were used: macular raster, macular star, and peripapillar (Figure 3.1).



(a) Macular raster **(b)** Macular star **(c)** Peripapillar

Figure 3.1: OCT protocols used in BHRI dataset. The macular raster protocol is comprised of 25 horizontal B-scans of 512 A-scans imaging a 6 x 6 mm^2 region. Macular star images image a slightly smaller region with 12 radial B-scans. The peripapillary protocol is used to image the area surrounding the optic disc with a single B-scan.

Each acquisition protocol consisted of the set of B-scans described in Table 3.3 and a grayscale fundus image. While both macular raster and optic disk peripapillar scans were always acquired, not all subjects underwent macular star images acquisition. Additionally, two subsets of 10 and 12 healthy subjects were imaged twice the same day with both the standard and a higher resolution macular raster protocol, respectively. These smaller datasets were used to evaluate test-retest repeatability and the impact of using different resolutions. In all cases, for each final B-scan a total of 49 slices were averaged to improve image quality.

Protocol	Region	B-scan	A-scan	Subjects		Images	
				HC	PD	HC	PD
Macular raster	6 x 6 mm^2	25	512	514	159	1149	592
Macular star	⌀ 4.3 mm	12	768	264	109	604	367
Peripapillar	⌀ 1.7 mm	1	768	519	159	1168	602
Macular raster (test-retest)	6 x 6 mm^2	25	512	10	-	20	-
Macular raster (high density)	6 x 6 mm^2	97	1024	12	-	24	-

Table 3.3: Number of images and subjects in BHRI by protocol. In addition to the three main protocols two smaller datasets were acquired to evaluate test-retest repeatability and the impact of scanning resolution. HC: healthy control. PD: Parkinson's disease.

All the images were visually inspected and those not fulfilling OSCAR-IB criteria were excluded [68]. Additionally, the images were segmented by HEYEX 1.9.10.0 software. Segmentations were reviewed by three specialists and evident errors within the 3 mm radius macular region were manually corrected.

In addition to OCT imaging and demographic data, additional variables were collected. These included the years of education in all subjects as well as the Levodopa equivalent daily dose in patients. Moreover, some of the subjects undergone an assessment of motor and cognitive function in which the UPDRS-III, HY, and MoCA were measured.

3.2 AlzEye

AlzEye is a retrospective dataset of 353157 subjects who attended Moorfields Eye Hospital NHS Foundation Trust (MEH) between January 2008 and April 2018 [178]. Included individuals were aged 40 years and over and attended the glaucoma, retina, neuro-ophthalmology, or emergency services of any of the clinical sites part of MEH. The dataset includes three sources of information:

- *Ophthalmic health variables*: patient-level data variables extracted from MEH data warehouse including: date of birth, sex, ethnicity, socioeconomic status measured by the Index of Multiple Deprivation decile, and both clinical appointment and operation dates.
- *Retinal imaging*: color fundus photographs as well as OCT images. Two algorithms were used to segment fundus images and extract vasculature features: The Vascular Assessment and Measurement Platform for Images of the Retina [179,180], and AutoMorph [181]. OCT images were also automatically segmented as described later in this section.
- *Systemic health variables*: labels regarding systemic diseases were obtained by linking MEH data with the larger National Health Service database of Hospital Episode Statistics (HES). This HES database includes routinely collected data following patient admission to any site part of the healthcare system. These data are translated to International Classification of Diseases (ICD) codes by clinical coders (e.g., G20 and G30 are used for PD and AD, respectively). These codes can ultimately be used as labels for data analysis.

From the entire cohort, AlzEye currently contains data from 149108 subjects. The drop in subject number is related with subjects who did not have a HES or opted out of using their data for research. From the included subjects, 1404 have an ICD code for PD (Table 3.4). Some of the subjects included in AlzEye visited MEH more than once and have longitudinal imaging acquired. There is a large heterogeneity in the number of visits and total follow-up time of these subjects. This is because each subject attended MEH for a different clinical reason and had a specific appointment schedule based on the severity of their case.

Regarding OCT imaging, the vast majority of images were acquired with Topcon devices (Top-

Group	Baseline			Longitudinal		
	N	Age (years)*	Male (%)	N	Follow time (years)*	Visits*
All	149108	66.1 ± 13.7 [30, 110]	47.7	81904	4.7 ± 2.9 [0.002, 11]	7.7 ± 10.2 [2, 153]
Parkinson	1404	76.5 ± 8.5 [37, 97]	57.7	828	3.0 ± 2.1 [0.01, 8.6]	7.3 ± 9.3 [2, 80]

*mean $\pm$ σ [range].

Table 3.4: AlzEye subject summary. From the 149108 subjects included in AlzEye 1404 had a label for Parkinson's disease from which 518 had more than one visit.

con Corporation, Tokyo, Japan). The acquisition protocol most often employed and intended to be used for analysis is a macular cube (Figure 3.2). In total, there were 1348934 images of this kind. It consists of 128 horizontal B-scans covering an area of 6 x 6 mm^2. Each B-scan consisted of 512 A-scans and was automatically segmented by TABS algorithm [37]. This proprietary algorithm is able to segment 10 retinal layers based on dual-scale gradient graph search. In addition to the segmentation, TABS also estimates the foveal center location and computes a set of quality metrics related with image contrast, artifacts and segmentation failures.

Figure 3.2: Macular cube protocol used in AlzEye. In each Alzeye acquisition a color fundus image is acquired along with the OCT protocol. The latter is shown in green and conists of 128 horizontal B-scans with 512 A-scans each.

3.3 Selected databases

From the aforementioned large databases, different data subsets have been used throughout the thesis. The datasets used in each analysis are listed in Table 3.5 and described below:

- *Automatic image assessment*: a total of 4000 images from AlzEye were used to develop a model to detect OCT quality problems. The AlzEye database was selected for this task because of its high prevalence of quality issues.
- *Automatic foveal center location*: a dataset including all BHRI subjects was used to validate automatic foveal location algorithms. Additionally, a smaller test-retest dataset was employed

to evaluate the improvement of the repeatability when locating the foveal center correctly.

- *Robust foveal pit morphology analysis*: this dataset included 185 BHRI healthy subjects imaged twice in the same day with both macular raster and macular star protocols. These two acquisitions were used to evaluate different automatic methodologies for foveal pit morphology analysis.
- *Validation or raster acquisition protocol*: a small dataset of 12 subjects imaged twice using different resolution protocols was used to evaluate the bias introduced when using the standard raster protocol with 25 B-scans instead of an higher-density acquisition pattern.
- *Retinal morphology in a healthy population*: 444 healthy subjects used to evaluate the impact of age and sex on the retina.
- *Diagnosis and severity assessment*: HC subjects were matched with PD patients based on age and sex. This dataset was used to develop a diagnostic model for PD and evaluate regression models for disease severity prediction.
- *Longitudinal analysis*: subjects from both BHRI and AlzEye datasets were used to evaluate how the retina evolves over time in PD. The Table 3.5 reflects the finally included number of AlzEye subjects after applying the inclusion criteria detailed in Chapter 7.

As an important detail, in the datasets derived from BHRI only images derived from Spectralis devices were included, thus excluding images acquired with a Topcon device at Biodonostia center. This was decided to prevent the potential bias resulting from adding 26 subjects that were all PD patients and were acquired with a device that differed notably from the rest of the BHRI dataset (refer to Table 3.2 for a summary).

Analysis	Section	Dataset	Protocol	Group	Subjects/ Eyes	Age*	Male (%)
Automatic image assessment	Section 4.1	AlzEye	Macular raster	-	4000/4000	67.6 ± 13.8	47.9
Automatic foveal center location	Section 4.2	BHRI	Macular raster	- HC	724/1388 10/20†	57.8 ± 12.4 35.2 ± 9.9	44.8 30.0
Robust foveal pit morphology analysis	Section 4.3	BHRI	Macular raster Macular star	HC	185/185	54.8 ± 11.9	40.0
Validation or raster acquisition protocol	Section 4.6	BHRI	Macular raster Macular raster‡	HC	12/24	37.0 ± 13.3	25
Retinal morphology in a healthy population	Chapter 5	BHRI	Macular raster	HC	444/855	54.9 ± 12.7	36.7
Diagnosis and severity assessment	Chapter 6	BHRI	Macular raster	HC PD	174/334 174/341	65.0 ± 8.5 64.9 ± 8.5	61.5 59.8
Longitudinal analysis	Chapter 7	BHRI	Macular raster Peripapillar	HC PD	72/138 158/309	61.3 ± 7.6 64.9 ± 8.6	42.3 65.2
		AlzEye	Macular cube	HC PD	873/1134 167/217	75.7 ± 9.4 76.7 ± 8.6	60.6 55.7

* mean ± σ (years)

† acquired twice with the same protocol (test-retest)

‡ high density protocol

Table 3.5: Datasets used in each analysis. From the two large datasets described in this chapter (BHRI and AlzEye) several smaller subsets of data were used in each of the analyses carried out as part of this research work. HC: healthy control. PD: Parkinson's disease.

4 | OCT PROCESSING ALGORITHMS

In this chapter the technical work carried out to develop a robust OCT image analysis pipeline is described. First, Section 4.1 describes the development of an automatic quality control model for OCT images. Secondly, a new method for automatic foveal center location is presented in Section 4.2. Then, different strategies for foveal pit morphology analysis are compared in Section 4.3. Next, Section 4.4 presents RETIMAT, an open-source toolbox designed to facilitate OCT image analysis that integrates all previously developed methods. Subsequently, Section 4.5 describes how we applied RETIMAT to process the databases described in Chapter 3. Finally, a sensitivity analysis carried to validate the standard macular raster acquisition protocol used in BHRI is described in Section 4.6. Some of the results presented in Sections 4.2, 4.3, and 4.6 have already been published by the author in [182, 183].

4.1 Automatic image assessment

Quality assurance is a key step in OCT imaging. Before formal analysis, images need to be visually inspected to identify quality problems. This step is crucial as poor quality data can dramatically impact any statistical analysis or model trained on such images. With large databases, however, rather than manually inspecting the data it would be beneficial to develop an algorithm capable of automatically excluding invalid images. A difficulty in this regard is the highly multifaceted nature of the problem. That is, images can be unsuitable for analysis for a variety of reasons such as poor contrast, artifacts, segmentation errors, or ocular lesions. Moreover, the intricate relationship between these factors makes it difficult to develop an independent solution for each factor. To address this problem, we developed a more general machine learning model that could screen OCT volumes and detect anomalies caused by quality problems.

Image processing and labeling

A random subset of 4000 macular images from the AlzEye database were included (Table 4.1).

Images	Subjects/Eyes	Age (years)*	Male (%)
4000	4000/4000	67.6 ± 13.8	47.9

*mean ± σ

Table 4.1: Subjects used for foveal pit model comparison. In this analysis a single image was included for each subject.

The images were derived from 4000 randomly selected subjects. For each subject only one image was randomly included to avoid the statistical dependence between left and right eyes. The

images had been previously segmented by the Topcon proprietary TABS algorithm, and included a wide range of cases with poor contrast, artifacts, and ocular pathologies.

The images were visually inspected by a biomedical engineer and those showing obvious macular anomalies (i.e., fovea not clearly present) or artifacts affecting the central macular region were labeled as *exclude*. The labeling process was carried out with the MATLAB GUI shown in Figure 4.1. Using this GUI, images can be successively annotated based on both raw and range-normalized TRT maps. TRT maps were used instead of actual B-scans because they provide a quick general overview of an entire OCT volume and help identify overall quality problems that could be missed by looking into single B-scans.

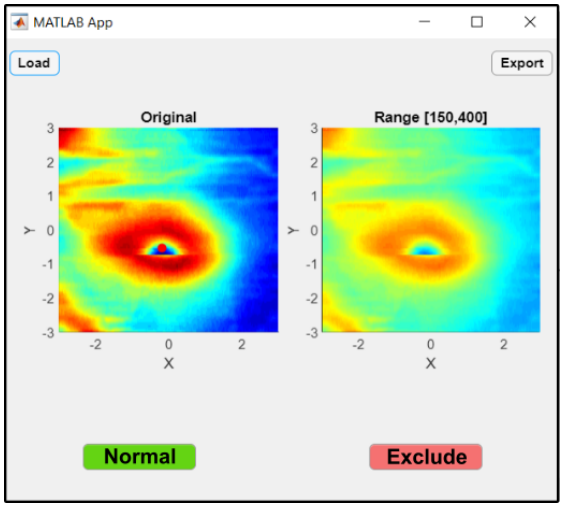
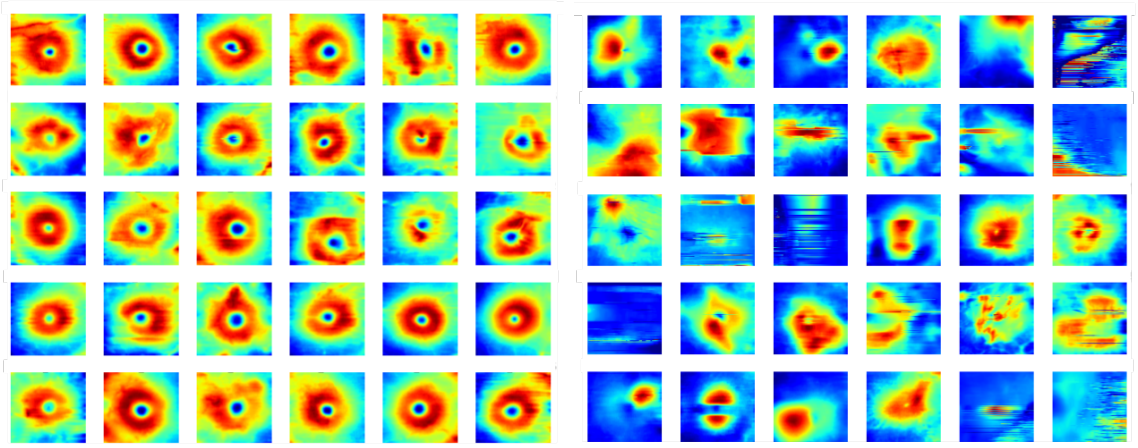


Figure 4.1: GUI used for quality labeling. The layout displays the TRT thickness map in the original and normalized scales. To speed up the process a new image is automatically shown after labeling an image. GUI: graphical user interface.

As a result of the labeling process, a total of 1760 images (44%) were labeled as *exclude*. This high value of bad images can be attributed to the nature of the AlzEye dataset, which consists of images routinely acquired in an ophthalmological hospital. Examples TRT maps of each class are shown in Figure 4.2.



(a) Normal **(b)** Abnormal

Figure 4.2: Examples of normal and abnormal TRT maps. Normal thickness maps show a clearly visible fovea. Conversely, poor image quality, artifacts, and ocular pathologies result in highly distorted maps in which the fovea is not clearly recognizable.

Data analysis

For every annotated OCT volume, 7 numerical features were derived (Table 4.2). The first 5 are part of the image metadata provided by the scanning OCT device (Topcon). According to technical documentation, the *quality* metric is related to image contrast. The *ilm indicator* measures the edge strength around the ILM boundary over the weakest region of the scan and is useful to identify blinks. Similarly, *max motion delta*, *min motion correlation*, and *max motion factor* are derived from thickness values (NFL and TRT), and are used to identify blinks, motion artifacts, and segmentation failures. These features have been used as part of the data curation pipeline in research analyzing OCT images from the UKBB [37, 97].

Feature	**Description**
quality	Topcon metric
ilm indicator	Topcon metric
max motion delta	Topcon metric
min motion correlation	Topcon metric
max motion factor	Topcon metric
pit depth	Average radial TRT
pit depth range	Radial foveal pit depth range

Table 4.2: Tabular features used to build the anomaly detection model. The first 5 features are part of Topcon metadata and are used to detect problems related to poor contrast and artifacts. Conversely, we designed the last two features to characterize foveal abnormalities. TRT: total retinal thickness.

The last two features (foveal *pit depth* and *pit depth range*) were custom designed for the task and are related to the foveal pit structure. They were included under the assumption that anomalous maculae have abnormal foveal shapes. To compute them, the TRT maps were resampled into a radial grid with 24 angular directions. For each direction the *pit depth* was computed as the difference between the highest and the lowest TRT values. An illustration of the highest TRT values is shown in Figure 4.3. From these radial values, the two final features were computed as the mean (final *pit depth* feature) and the total range (*pit depth range*) of all 24 values.

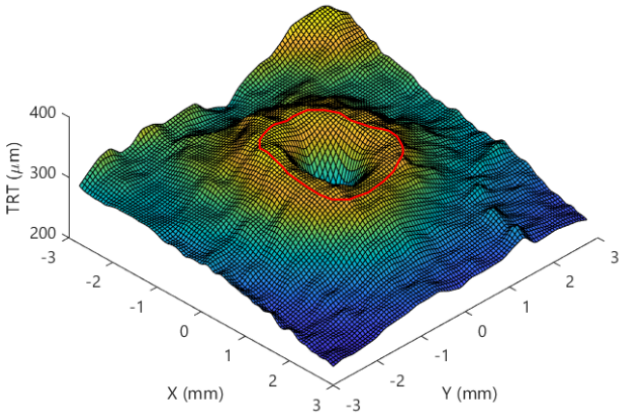


Figure 4.3: 3D view of the foveal rim used to compute pit depth. The surface is a 3D visualization of the thickness maps shown in Figure 4.2. The red contour delineates the foveal rim (i.e., perimeter of highest retinal thickness) used to compute *pit depth* and *pit depth range features*. TRT: total retinal thickness.

As a first exploratory analysis, the distributions of all features were inspected to find missing values and outliers. Similarly, the relationship between each pair of features was examined to detect highly correlated features.

Then, the computed features were used first to train a LR classifier with individual features. Then, LR, SVM and random forest (RF) models were trained combining all features. In both univariate and multivariate settings, 80% and 20% of the images were used for training and testing, respectively. Both SVM and RF models were fine tuned by optimizing several hyperparameters by means of randomized search and 5-fold cross-validation. The optimization procedure was configured to minimize the balanced accuracy (BAC) to account for the existing class imbalance. Model performance was assessed based on the AUC, BAC, sensitivity, and specificity. The latter three are defined in terms of the true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) as:

$$sensitivity = \frac{TP}{TP + FN} \tag{4.1}$$

$$specificity = \frac{TN}{TN + FP} \tag{4.2}$$

$$balanced\,accuracy = \frac{sensitivity + specificity}{2} \tag{4.3}$$

To reduce the variability caused by data partition, the model training and evaluation was repeated 500 times with different partitions. For each model performance matric, a 95% confidence interval (CI) was obtained as the 5 and 95 percentiles of the 500 score distribution. Images labeled as *exclude* were considered as the positive class.

Finally, the calibration of the trained multivariate LR, SVM, and RF models was evaluated by inspecting individual calibration curves. The curves were computed using uniform bins and for test samples only. More concretely, a 5-fold cross-validation procedure was employed to obtain probabilistic predictions for the entire dataset. These predictions were then compared against the fraction of positives in each bin.

Results

Individual feature distributions are shown in Figure 4.4. The data revealed that *max motion delta*, *max motion factor*, and *min motion correlation* followed a skewed distribution. Therefore, these features were normalized using the Yeoh-Johnson transformation to eliminate any potential bias [184].

The *max motion factor* feature followed a zero-inflated distribution (68.8% of the values were 0) and strongly correlated with *max motion delta* (Pearson correlation $r = 0.995$). For that reason, it was excluded from further analyses.

The performance of each model is reported in Table 4.3. The best individual predictor was *pit depth range* (AUC: 0.88 [0.86, 0.90], BAC: 0.82 [0.80, 0.84]), followed by *max motion delta* (AUC: 0.84 [0.81, 0.86], BAC: 0.66 [0.64, 0.68]). On the other hand, *quality* feature obtained the worst results with an AUC of 0.64 [0.61, 0.68] and a BAC of 0.61 [0.58, 0.63]. All the LR models trained on individual features showed a notable difference in the performance for each class. Indeed, the specificity was always higher than the sensitivity, likely as a consequence of the existing class imbalance.

When using all the features together, the performance improved for all three LR, SVM and RF models. The best point estimate was obtained by the latter with an AUC of 0.95 [0.93, 0.96]

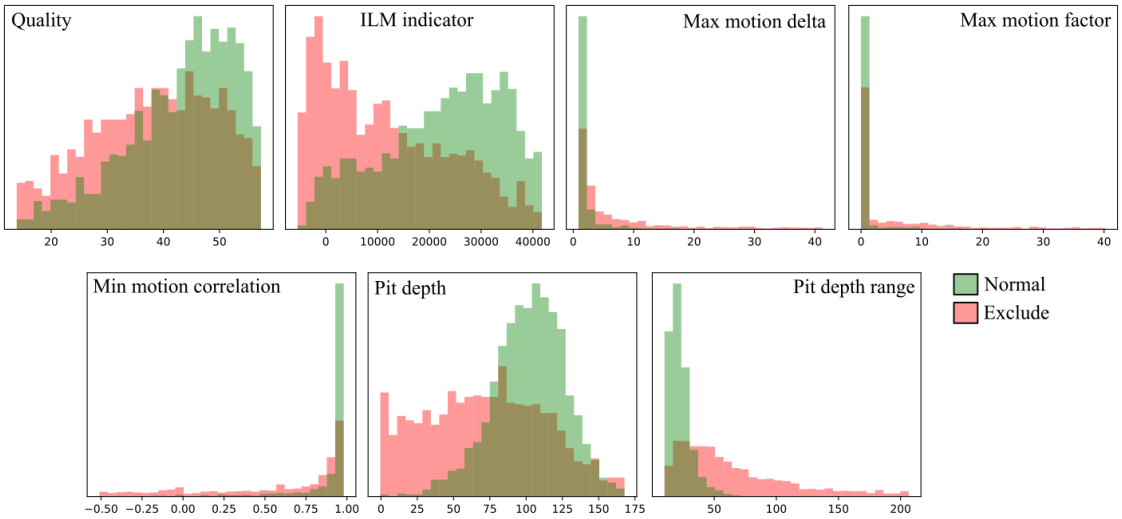


Figure 4.4: Distribution of each feature used for quality assessment. The distributions are normalized to represent a probability density function. *Max motion delta*, *max motion factor*, and *min motion correlation* features follow a highly skewed distribution that was normalized later. As observed, small values of *pit depth* are associated with an abnormal shape. Similarly, a large *pit depth range* is a clear indicator of an eye to be excluded.

and a BAC of 0.88 [0.85, 0.90]. As shown in Figure 4.5 all three multivariate models showed good calibration. Although the RF tended to overestimate low (0.1-0.35) and high probabilities (0.65-0.95), the differences with a perfectly calibrated model were always below 0.1 points.

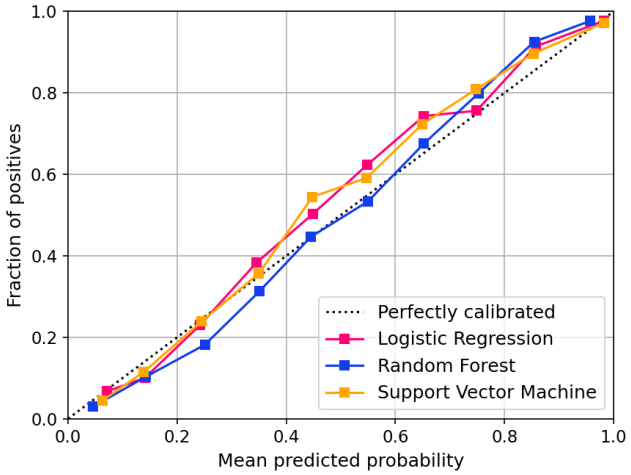


Figure 4.5: Model calibration curves. The curves show the relationship between the average score (i.e., predicted probability of being an image to be excluded) and the actual fraction of positives with that score. All three models showed good calibration.

Model	Features	AUC	BAC	Sensitivity (label=*exclude*)	Specificity (label=*include*)
LR	quality	0.64 [0.61, 0.68]	0.61 [0.58, 0.63]	0.41 [0.37, 0.46]	0.80 [0.75, 0.84]
	ilm indicator	0.74 [0.71, 0.78]	0.70 [0.67, 0.73]	0.62 [0.57, 0.66]	0.78 [0.74, 0.81]
	max motion delta	0.84 [0.81, 0.86]	0.66 [0.64, 0.68]	0.38 [0.33, 0.43]	0.95 [0.92, 0.96]
	min motion correlation	0.81 [0.79, 0.84]	0.68 [0.65, 0.70]	0.45 [0.40, 0.49]	0.90 [0.88, 0.93]
	pit depth	0.68 [0.64, 0.71]	0.68 [0.66, 0.71]	0.45 [0.40, 0.49]	0.92 [0.88, 0.94]
	pit depth range	0.88 [0.86, 0.90]	0.82 [0.80, 0.84]	0.71 [0.66, 0.75]	0.94 [0.92, 0.96]
LR	all	0.92 [0.90, 0.94]	0.85 [0.83, 0.87]	0.77 [0.73, 0.81]	0.93 [0.90, 0.95]
RF	all	0.95 [0.93, 0.96]	0.88 [0.85, 0.90]	0.85 [0.80, 0.89]	0.90 [0.87, 0.93]
SVM	all	0.94 [0.92, 0.95]	0.87 [0.85, 0.89]	0.83 [0.78, 0.87]	0.92 [0.89, 0.94]

Table 4.3: Model performance in anomaly detection. First 6 rows display the performance of individual features. Last three rows show how this performance was improved when combining all the features. LR: logistic regression. RF: random forest. SVM: support vector machine. AUC: are under the curve. BAC: balanced accuracy.

Based on the results it can be concluded that the RF achieved the best performance without undermining calibration. To further investigate its behavior, a new RF model was trained using the entire dataset and the hyperparameter tuning approach described previously. From this model, the feature importances were extracted based on impurity decrease. These values, shown in Figure 4.6, confirmed that the *pit depth range* was the feature contributing the most to accurate prediction. In a second level of significance, *max motion delta*, *min motion correlation*, and *pit depth* features reported moderate importance. Finally, both *quality* and *ilm indicator* features demonstrated low influence on prediction.

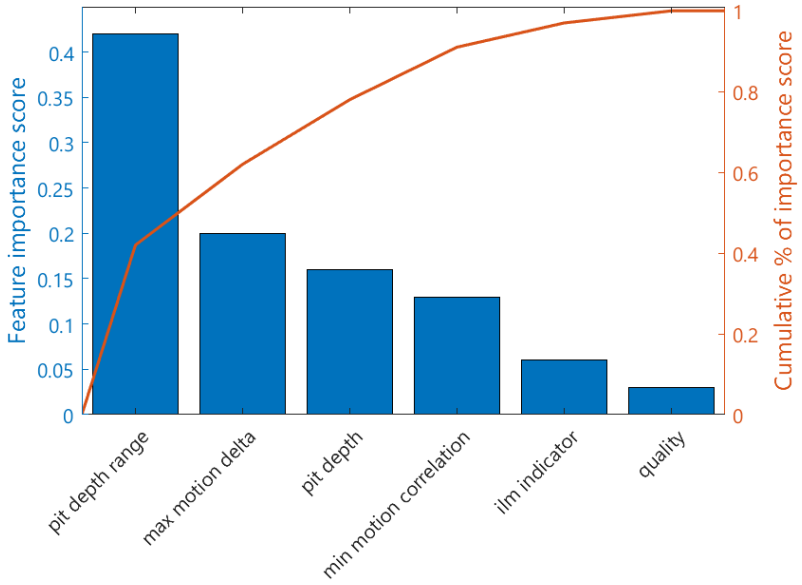


Figure 4.6: Feature importance for automatic quality control models. Left vertical axis depicts the importance of each feature. Right vertical axis shows the cumulative importance as features are combined.

Discussion

The results show that a relatively simple model can be used to automatically detect quality problems in OCT images. Interestingly, there were differences between the performance of each individual feature. The *pit depth range* feature obtained the best results, which could indicate that this feature is especially suitable for anomaly detection as it captures localized changes in the foveal structure. For instance, ocular lesions affecting concrete macular regions may not impact the overall foveal pit depth, but may be more clearly reflected by the *pit depth range*. This would also explain the poor individual performance of the *pit depth* feature by itself.

On the other hand, pure image contrast measured by the *quality* feature did not yield good results. This highlights that low signal to noise ratio (SNR) is only one aspect of OCT image quality control. In fact, combining all the features improved the performance, thus showing that more than one feature are needed to capture all sources of problems. In this regard, a RF model achieved the best performance. In any case, differences between multivariate models were not huge, which highlights that even a simple LR model can capture most of the signal. Importantly, all models showed good calibration on their probabilistic predictions. This observation opens up possibilities for utilizing these models beyond a binary classification. For instance, researchers working with large datasets may utilize model scores to exclude images only when the model exhibits high confidence in their exclusion (e.g., $p > 0.8$). Similarly, from a more conservative perspective, one could decide to directly include all images with high probability of being correct (e.g., $p < 0.2$) and review the rest.

Previous work has addressed OCT quality control before. As a common trait, state of the art methods have mostly relied on deep learning models to classify single B-scans [39, 40]. In contrast, we trained a simpler model to perform quality control at a volume level. This approach parallels a standard research workflow in which the goal is deciding if an entire volume needs to be excluded. In addition, existing approaches have a strong focus on poor image contrast and cropping artifacts, and disregard other sources of problems such as movement artifacts and ocular lesions. Admittedly, by limiting the scope of quality control to those factors, they were able to achieve an accuracy equal to or higher than 96% [39, 40].

Despite its benefits, the simplicity of our method could also present limitations. We only used the information contained in TRT maps under the assumption that it captures most of the problems affecting both inner and outer retina. However, smaller ocular lesions and segmentation errors may not be captured by TRT maps and would not be flagged by our model. Finally, most of the features used to train the model are part of Topcon device metadata and cannot be extrapolated to other devices directly. Nonetheless, we engineered a general (*pit depth range*) feature that outperformed all Topcon features and can be used with macular OCT images from any device.

4.2 Automatic foveal center location

During macular OCT acquisition, the scanning beam is placed roughly at the foveal center by the device operator. However, the actual scanning center may not lay precisely at the foveal center due to fixation errors or incorrect centering (see Figure 4.7 for an example of incorrect centering).

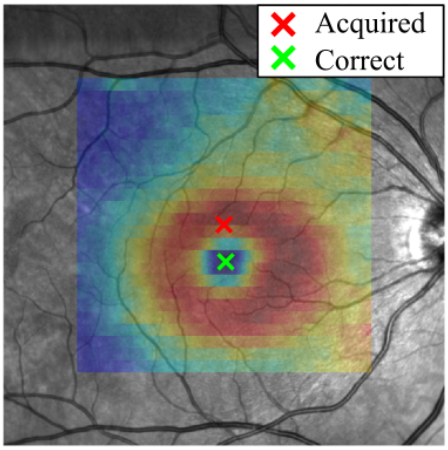


Figure 4.7: Example of incorrect centering. The center of the acquired image was 0.6 mm higher than the actual foveal center. Assuming the acquired center to be the fovea would result in an incorrect estimation of retinal features.

Such centering errors can bias any computation that relies on an accurate foveal center location (e.g., sectorized thickness measurements). As an alternative, automatic foveal center location algorithms can provide a fast and reproducible solution to this problem. However, existing methods for this task are either complex to implement or rely on deep learning models tailored to specific acquisition protocols. Hence, in this section we present a novel automatic foveal center location algorithm and evaluate its performance.

Images and data annotation

A total of 1388 images from 724 subjects in BHRI dataset were used (Table 4.4). The subjects included both healthy individuals and patients with PD. All the images were acquired using a

macular raster acquisition protocol. Although most of the subjects had both eyes included, a total of 60 eyes were excluded from analysis due to ocular pathologies.

Subjects	Images	Age (years)*	Male (%)
724	1388	57.8 ± 12.4 [21, 88]	44.8

* mean ± σ [range].

Table 4.4: Demographic characteristics of subjects used in foveal analysis.

From the retinal layer segmentation, TRT maps were computed as the difference between the ILM and the BM. These TRT maps were the input for the subsequently developed foveal location algorithm. Before data annotation, all TRT maps were resampled to generate images of 256 x 256 pixels using 2D cubic interpolation. These images were then used to manually locate the foveal center with Label Studio platform. More concretely, the fovea was located at the center of the darkest blue region surrounded by the foveal rim (see Figure 4.7). The obtained coordinates of the manually located foveal center were considered as the ground truth and used to evaluate the performance of the algorithm.

In addition, a smaller subset of images was included to evaluate the improvement on test-retest repeatability as a result of locating the foveal center correctly. This dataset contained images of both eyes of 10 subjects that were acquired twice in the same week following the exact same macular raster acquisition protocol.

Foveal center location algorithm

The developed algorithm leverages the convex shape of the foveal pit and locates the foveal center at the position with the deepest point compared to the foveal rim. The entire process consists of two steps.

First, a coarse location of the foveal center is achieved by means of the flooding algorithm described in Algorithm 1. Initially, several particles are randomly placed around the macula. Then, the location of each particle is iteratively modified by finding the neighbor with the highest negative TRT gradient. This process forces the particles to descend the foveal surface towards a local minima. After several iterations, the particles either reach the foveal center or a location near the edge of the TRT map (Figure 4.8). Particles close to the edge are assumed not to be the fovea and are discarded. The remaining particle locations are then used to build a 2D histogram and the fovea is coarsely located at the pixel with the highest histogram value.

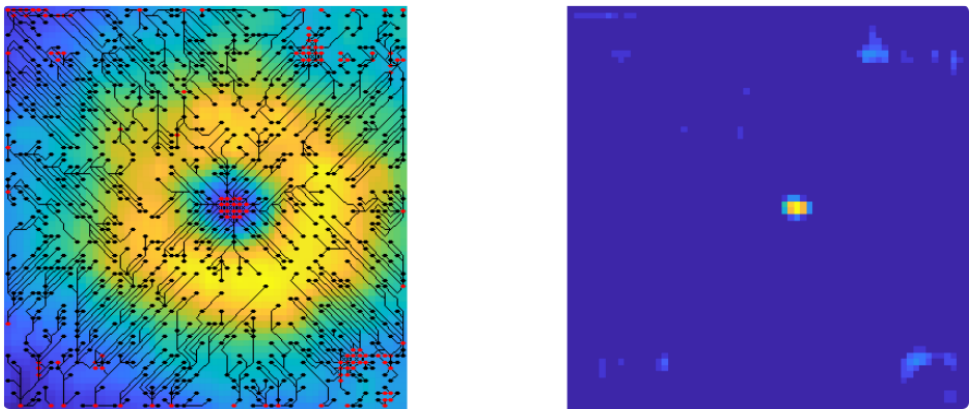


Figure 4.8: Foveal location flooding algorithm. *Left*: particle trajectories from starting point (black dots) to end points (red dots). *Right*: 2D histogram of particle end location.

Algorithm 1 Flooding algorithm: coarse foveal center location

Inputs:
TRT map: $\mathbf{Z}$ (N x N matrix)
Number of particles: n_p
Number of steps: n_s
Margin gap: m
Procedure:
$\mathbf{H} = h_{i,j} = 0$ (N x N matrix)
for $p = 1 : n_p$ **do**
Random pixel location: i, j
for $s = 1 : n_s$ **do**
$\mathbf{i}' = [i-1,\ i,\ i+1,\ i-1,\ i+1,\ i-1,\ i,\ i+1]$
$\mathbf{j}' = [j-1,\ j-1,\ j-1,\ j,\ j,\ j+1,\ j+1,\ j+1]$
for $n = 1 : 8$ **do**
if $\mathbf{i}'[n] < 1$ **or** $\mathbf{i}'[n] > N$ **or** $\mathbf{j}'[n] < 1$ **or** $\mathbf{j}'[n] > N$ **then**
$\mathbf{d}[n] = 0$
else
$\mathbf{d}[n] = \mathbf{Z}[\mathbf{i}'[n], \mathbf{j}'[n]] - \mathbf{Z}[i, j]$
end if
end for
if $min(\mathbf{d}) >= 0$ **then**
break
else
$n_{next} = \arg\min_n(\mathbf{d})$
$i = \mathbf{i}'[n_{next}]$
$j = \mathbf{j}'[n_{next}]$
end if
end for
if $i \leq m$ **or** $i > N - m$ **or** $j \leq m$ **or** $j > N - m$ **then**
break
else
$\mathbf{H}[i, j] = \mathbf{H}[i, j] + 1$
$\mathbf{Z}[i, j] = \mathbf{Z}[i, j] + 1$
end if
end for
$i_{fovea}, j_{fovea} = \arg\max_{i,j}(\mathbf{H})$

The code implementation of the flooding algorithm has four inputs: the TRT map resampled into a 64 x 64 pixel grid, the number of particles to generate ($n_p = 3000$), the maximum number of steps before stopping ($n_s = 30$), and the margin gap used to exclude particles with an end location at the edge of the image ($m = 15$).

In the second stage, the location of the foveal center obtained by the flooding procedure is refined. To this end, the original TRT map is resampled into a high density grid and smoothed with a filter with a $0.05\,mm$ radius circular kernel. The foveal center is located at the point of minimum TRT in a circular $0.5\,mm$ radius region surrounding the coarse foveal center position.

Performance evaluation of the algorithm

The error of the algorithm was measured as the euclidean distance between the located foveal center and the ground truth. This metric was computed for every image to obtain an error distribution.

To evaluate the improvement of the algorithm over simpler strategies, four additional foveal location methods were evaluated:

- **None**: consider the acquisition center as the foveal pit center.
- **Min**: locate the foveal center at the A-scan point of minimum TRT in the central $0.85\,mm$ radius region.
- **Interpolation + min**: resample the central part of the TRT map to a regular grid of 0.85 x $0.85\,mm^2$. Then, locate the foveal center at the grid point with minimum TRT.
- **Smooth + min**: resample the central part of the TRT map to a regular grid of 0.85 x $0.85\,mm^2$, and smooth it before locating the foveal center at the grid point with minimum TRT. We used the implementation of AURA tools (*foveaFinder.m* function) [29,95] to smooth the resampled TRT map by applying a filter with a $0.05\,mm$ radius circular kernel (same smoothing procedure as that used in the second stage of the proposed algorithm).

The error distribution of each method is plotted in Figure 4.9 and summary results are presented in Table 4.5. The average error when assuming the images to be correctly centered (*None* method) was $109\,\mu m$ with a maximum of $0.6\,mm$. This error was reduced by all the other methods to different extents.

The proposed algorithm outperformed all simpler strategies and reduced the default misalignment by 72%. The difference with the second best strategy (*smooth + min*) was $7.8\,\mu m$ (Mann-Whitney U test, $p = 10^{-14}$), which corresponded to a 20% reduction in average error. This difference remained significant when including a single image per subject (Mann-Whitney U test, $p = 10^{-7}$), which confirmed that the potential statistical dependence between both eyes of each subject did not affect the results.

Method	Error (μm)*
None	109.9 ± 74.5 [4.0, 600.3]
Min	58.3 ± 39.3 [2.9, 293.7]
Interpolation + min	52.3 ± 37.2 [3.3, 291.4]
Smooth + min	38.3 ± 19.3 [1.7, 115.1]
Proposed	30.5 ± 16.8 [0.4, 112.3]

* mean ± σ [min, max]

Table 4.5: Foveal location error of each method. The error is defined as the euclidean distance between the located foveal center and the labeled ground truth.

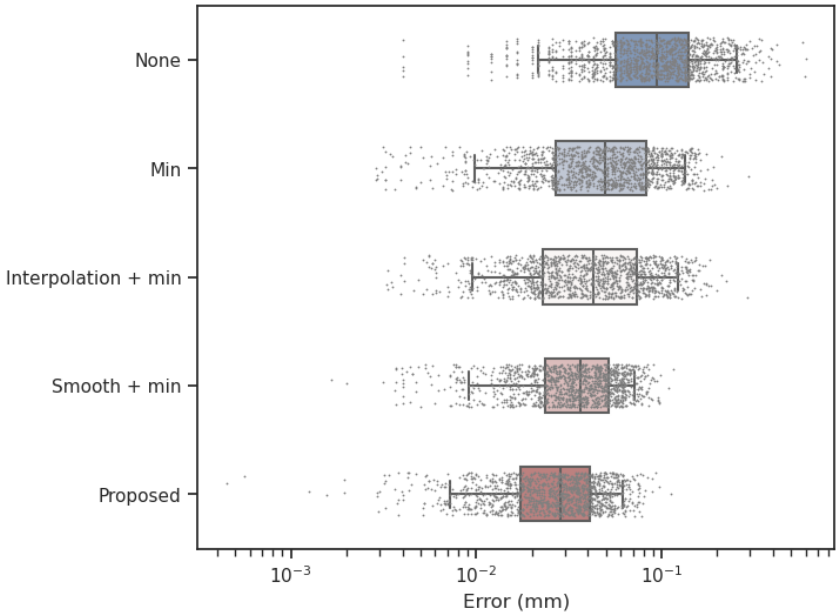


Figure 4.9: Foveal location error distribution of each method. The error is defined as the euclidean distance between the located foveal center and the labeled ground truth. For comparison, the error is given in logarithmic scale.

Repeatability improvement

In this second analysis we investigated the impact of correctly locating the foveal center in terms of test-retest repeatability. We computed TRT and GCIPL thicknesses averaged over both the whole macula and ETDRS sectors. This computation was done twice with and without locating the foveal center using the proposed algorithm.

The results are displayed in Table 4.6 and highlight how the mean absolute difference (MAD) between test and retest measurements diminished for all sectors when correctly locating the fovea.

For instance, properly locating the fovea reduced the expected MAD from $1.7\,\mu m$ to $1.5\,\mu m$ ($\Delta = 12\%$) for macular TRT. This effect was even more noticeable for the central and outer sectors, which showed an overall improvement above 20% with a maximum of 45.5%. Although the improvement effect appears to be evident it should be mentioned that the limited sample size ($N = 10$) prevented any formal statistical test.

Applicability and comparison with previous work

Macular OCT images are not always perfectly centered at the fovea. In the analyzed dataset the misalignment error ranged up to $0.6\,mm$. It is important to note that this error is present even though the images were acquired as part of a research study and careful attention was devoted into ensuring an accurate alignment. This suggests that images acquired in a clinical setting would likely have more severe alignment problems and benefit more from the foveal location step.

Importantly, we showed that misalignment errors can be considerably reduced by automatic foveal center location methods, which suggests it is advisable to incorporate such a step in any quantitative OCT analysis pipeline. More specifically, we developed and evaluated a novel two-stage algorithm that outperformed simpler methods and whose improved performance can be attributed to three main aspects:

1. **Two-stage procedure**: the first stage of the proposed algorithm aims to locate the foveal

Layer	Sector	MAD (μm)		Δ (%)
		None	Proposed	
TRT	Macula	1.70	1.50	12.1
	Central	3.28	2.65	19.2
	Inner nasal	1.94	1.62	16.5
	Inner superior	1.99	1.65	17.1
	Inner temporal	2.06	1.65	19.8
	Inner inferior	3.17	2.36	25.4
	Outer nasal	1.97	1.95	0.9
	Outer superior	2.63	1.89	28.2
	Outer temporal	3.04	2.27	25.1
	Outer inferior	2.74	1.57	42.6
GCIPL	Macula	0.76	0.73	4.2
	Central	2.07	1.21	41.5
	Inner nasal	1.76	1.30	26.1
	Inner superior	2.01	1.87	6.7
	Inner temporal	1.71	1.56	8.69
	Inner inferior	2.51	1.79	28.7
	Outer nasal	1.82	1.41	22.4
	Outer superior	1.92	1.04	45.5
	Outer temporal	2.11	1.57	25.4
	Outer inferior	2.57	1.86	27.5

Table 4.6: Test-retest results of foveal location methods. Repeatability is measured as the mean average difference (MAD) between test and retest acquisition. The last column shows the improvement atained by correctly locating the foveal center.

center approximately. The search space of the second stage can thus be limited to a small region of interest and the foveal center can be located more accurately. Moreover, contrary to the simpler methods—which limit the search space a priori to a small 0.85 mm radius region—, the proposed algorithm explores a larger macular region and could locate the foveal center in cases with a misalignment much larger than that observed in our dataset.

2. **Interpolation**: when the number of B-scans is relatively low (in this case 25), resampling the data into a higher density grid might help locate the foveal center in cases where the central B-scan does not capture it.

3. **Smoothing**: the filtering operation aggregates information across adjacent pixels and is likely more robust against segmentation errors.

It is worth mentioning that locating the fovea precisely has a positive impact on thickness measurements. In fact, we measured an up to 45% improvement in the test-retest repeatability of certain ETDRS thickness measurements. This suggests that an accurate centering reduces measurement error, which would likely result in a higher statistical power.

While some previous researchers have also studied automatic foveal location, there remain some differences. For instance, Niu et al. [65] proposed a method based on saliency maps that conceptually bears some resemblance to our approach. Rather than relying on the complex mathematical modeling featured in their work, our approach stems from the intuitive idea of *flooding* and does not require fine-tuning, which may make it easier to understand and implement. Moreover, we used a considerably larger database.

Deep learning approaches have also been proposed [66, 67]. Although these approaches have shown great potential, we opted not to explore them for two reasons: 1) since our proposed algorithm successfully fulfilled the task, there was little incentive to pursue more complex models, and 2) our algorithm does not rely on either image intensity or TRT scaling, which arguably results in better generalizability. Furthermore, differences in scanning devices and protocols are known to

hinder the performance of deep learning models. It is true, however, that locating the foveal center in pathological cases with a distorted fovea may require a model as complex as a neural network.

As mentioned, the subjects included in the analysis did not present any ocular pathology. Therefore, the obtained results may not extrapolate to subjects with ocular lesions affecting the macular structure (e.g., when the fovea is not clearly recognizable). This factor makes it difficult to numerically compare the results with previous research, which did include pathological cases [65–67]. We did not address this point in the present work because the main objective was to develop a robust foveal location method for the BHRI dataset, which does not include any cases with ocular pathology.

4.3 Robust foveal pit morphology analysis

The foveal pit morphology can be studied by either computing geometrical features (e.g., mean slope, depth, etc.) and/or by fitting mathematical models to the foveal shape. The mathematical models can potentialy be used in two ways: 1) to smooth the foveal pit before computing geometrical parameters, and 2) to use estimated equation coefficients as features describing the fovea. With this aim, multiple models have been proposed [72–77]. However, geometrical features can be directly computed from raw TRT maps [128, 130, 185], and there is no conclusive evidence of the benefits of using mathematical models as a previous smoothing step. Moreover, to the best our knowledge, existing models have not been compared quantitatively.

To address this, here we investigate the advantages and limitations of introducing a modeling step prior to the computation of geometrical parameters. Specifically, we quantitatively compared six mathematical models and two smoothing approaches in terms of their ability to harmonize measurements from different acquisitions of the same eye.

Dataset and image processing

The dataset used is described in Table 4.7 and consisted of a total of 185 eyes from 185 healthy subjects imaged twice in the same day using raster and star acquisition protocols. Raster and star scans were treated as test and retest, respectively.

Subjects/Eyes	Age (years)*	Male (%)
185	54.8 ± 11.9	46.8

* mean ± σ [range].

Table 4.7: Subjects used for foveal pit model comparison.

All the images were processed following the pipeline illustrated in Figure 4.10. First, the TRT maps is computed. This step is useful to define a common flat reference from which to compute geometrical parameters and disregard the effect of the eye curvature. The obtained TRT profile can be considered as the raw curve delineating the foveal surface. This signal, usually noisy, is the one that will be subsequently modeled. Secondly, the foveal center was automatically detected (see Section 4.2) and used to align TRT maps by means of 2D translation. From centered TRT maps, raster scans were transformed into a star grid with 24 angular directions and a 2.5 mm radius. This was necessary to radially characterize the foveal shape.

Next, six mathematical models developed by Dubis et al. [73], Ding et al. [74], Scheibe et al. [77], Liu et al. [75], Yadav et al. [72], and Breher et al. [76] were fitted to the data (refer to Table 2.1 for a detailed list). This fitting process employed the non-linear least squares method, allowing a maximum of 1000 iterations and setting a tolerance of 10^{-6} for both the residuals and

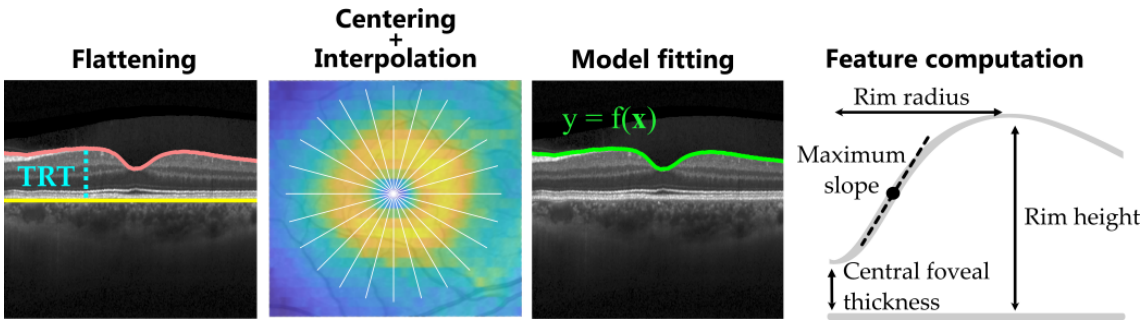


Figure 4.10: Model comparison pipeline. In the first step the retina is flattened by computing the total retinal thickness (TRT). Then the foveal center is located and used to interpolate the thickness map into a radial grid. Finally, mathematical models are fitted to the data and used to derive geometrical features. Adapted from [182].

model coefficients. The initial coefficient values were manually adjusted, and the option yielding the lowest fitting error was chosen. In addition, locally estimated scatterplot smoothing (LOESS) was also used based on a second-degree polynomial and a span ranging from 1% to 50%. It is important to note that the smoothing procedure was applied individually to each B-scan.

From the fitted data, four geometrical parameters were derived for each OCT volume: rim height (highest TRT value), rim radius (horizonal distance from the fovea to the point of maximum TRT), maximum slope (maximum of the firs derivative of the TRT profile between the fovea and the rim), and CFT (TRT at the located foveal center). The first three—with a different value for each angular direction—were averaged across all 24 angular directions to compute a single value per image.

Model comparison

Both mathematical models and smoothing methods were compared based on the following metrics:

- **Fitting error:** measured as the RMSE between the raw TRT maps and the TRT maps obtained after model fitting.
- **Absolute agreement between raster and star**: used to determine the ability of each strategy to enhance agreement between two distinct acquisitions of the same eye (raster and star). This assessment was carried out for each morphological parameter and involved the use of the intra-class correlation coefficient (ICC) based on a single measurement and a 2-way mixed-effects model (ICC (2,1), [186]). In addition to the mean ICC, 95% confidence intervals were computed using the percentile bootstrap method resampling the data 10^4 times.
- **Estimation bias**: used to evaluate the effect of the modeling/smoothing step on each parameter estimation. This was calculated as the relative difference between the estimation of each parameter before (x_{raw}) and after model fitting or smoothing (x_{model}):

$$Bias(\%) = 100\,\frac{x_{model} - x_{raw}}{x_{raw}} \tag{4.4}$$

Separate calculations were performed to determine the RMSE and estimation bias for both raster and star scans.

The results of the model comparison can be found in Tables 4.8 and 4.9. Two representative instances of LOESS are showcased, corresponding to a low degree of smoothing (span = 20%) and a high degree of smoothing (span = 50%).

Model	RMSE		ICC			
	Raster	Star	CFT	Rim height	Rim radius	Maximum slope
None	-	-	0.976 [0.966, 0.983]	0.990 [0.987, 0.992]	0.894 [0.865, 0.919]	0.307 [0.236, 0.381]
Dubis et al. [73]	3.6 ± 0.7	4.1 ± 0.7	0.988 [0.984, 0.992]	0.995 [0.994, 0.996]	0.949 [0.934, 0.962]	0.968 [0.957, 0.977]
Ding et al. [74]	5.3 ± 0.9	5.9 ± 0.9	0.988 [0.984, 0.992]	0.995 [0.994, 0.997]	0.957 [0.945, 0.966]	0.969 [0.958, 0.977]
Scheibe et al. [77]	2.6 ± 0.6	3.2 ± 0.6	-	0.995 [0.994, 0.997]	0.949 [0.933, 0.962]	0.956 [0.939, 0.969]
Liu et al. [75]	11.5 ± 2.7	11.5 ± 2.7	0.987 [0.983, 0.991]	0.994 [0.992, 0.996]	0.961 [0.949, 0.970]	0.959 [0.944, 0.971]
Yadav et al. [72]	1.6 ± 0.3	2.5 ± 0.4	-	-	-	0.958 [0.943, 0.970]
Breher et al. [76]	2.9 ± 0.6	3.6 ± 1.3	0.986 [0.979, 0.990]	0.995 [0.993, 0.996]	0.941 [0.924, 0.955]	0.958 [0.942, 0.971]
LOESS (20%)	0.9 ± 0.1	1.7 ± 0.3	0.985 [0.980, 0.989]	0.994 [0.992, 0.996]	0.901 [0.875, 0.924]	0.953 [0.936, 0.966]
LOESS (50%)	5.9 ± 1.5	6.5 ± 1.6	0.989 [0.984, 0.993]	0.995 [0.994, 0.997]	0.960 [0.947, 0.970]	0.986 [0.981, 0.990]

Table 4.8: Fitting error and agreement of each foveal pit modeling approach. The error is measured as the root mean square error (RMSE) between the raw and modeled foveal shape. The agreement is evaluated by the intraclass correlation coefficient (ICC). CFT: central foveal thickness. From [182].

Model	Bias (%)							
	CFT		Rim height		Rim radius		Maximum slope	
	Raster	Star	Raster	Star	Raster	Star	Raster	Star
Dubis et al. [73]	1.3 ± 1.3	1.4 ± 1.9	-0.2 ± 0.2	-0.5 ± 0.3	-7.8 ± 3.7	-8.2 ± 4.1	-14.1 ± 4.1	-34.0 ± 9.7
Ding et al. [74]	1.1 ± 1.4	1.2 ± 2.1	-0.5 ± 0.3	-0.8 ± 0.3	-7.8 ± 3.8	-8.1 ± 4.1	-13.9 ± 3.9	-33.9 ± 9.7
Scheibe et al. [77]	-	-	-0.1 ± 0.3	-0.3 ± 0.3	-3.8 ±2.4	-3.5 ± 2.4	-19.8 ± 4.2	-38.6 ± 7.8
Liu et al. [75]	-1.1 ± 1.2	-1.1 ± 1.8	-3.6 ± 0.9	-3.9 ± 0.9	35.0 ± 7.4	36.4 ± 8.0	-5.3 ± 4.6	-27.1 ± 9.8
Yadav et al. [72]	-	-	-	-	-	-	-9.1 ± 4.8	-29.7 ± 11.9
Breher et al. [76]	0.8 ± 1.1	0.9 ± 1.8	-0.4 ± 0.2	-0.6 ± 0.2	-6.5 ± 2.9	-6.6 ± 3.2	-11.9 ± 3.4	-32.1 ± 9.4
LOESS (20%)	0.3 ± 0.5	0.4 ± 1.4	-0.1 ± 0.1	-0.4 ± 0.1	-0.1 ± 0.9	-0.1 ± 1.5	-9.1 ± 2.3	-29.2 ± 10
LOESS (50%)	6.0 ± 2.7	6.6 ± 3.3	-0.3 ± 0.3	-0.5 ± 0.3	2.2 ± 2.7	2.5 ± 2.8	-28.8 ± 6.1	-46.6 ± 8.2

Table 4.9: Estimation bias of each foveal pit modeling approach. The bias is computed as the relative difference between the estimations obtained by the raw and modeled data. The results are presented separately for raster and star patterns. CFT: central foveal thickness. From [182].

When no model was utilized, there was a high level of agreement between raster and star estimations for both the CFT (ICC = 0.976) and the rim height (ICC = 0.990). The rim radius also exhibited a good level of agreement (ICC = 0.894), while the maximum slope yielded the poorest results (ICC = 0.307). The limited agreement observed for the maximum slope was attributed to a consistent overestimation in the star scans (Figure 4.11).

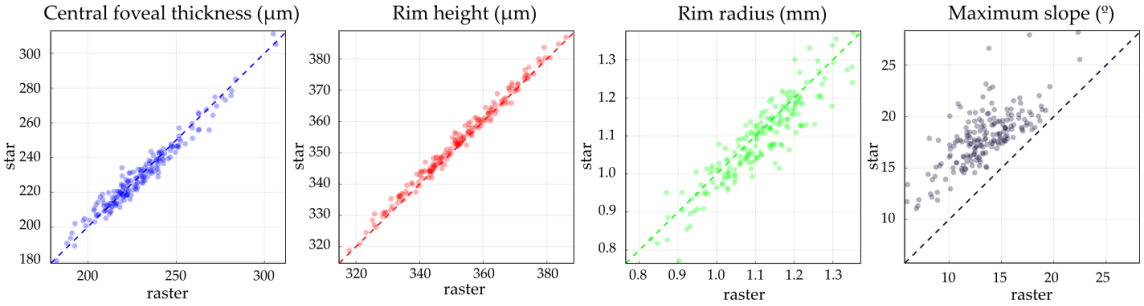


Figure 4.11: Raster vs. star estimation without using any model. A systematic overestimation of the maximum slope was observed only for star acquisitions. From [182].

With the exception of the model proposed by Liu et al. [75], all the other models successfully fit the data with an RMSE below 6 μm. Of these, the model presented by Yadav et al. [72] fitted the data best. Figure 4.12 displays representative examples of a low, medium, and high fitting errors.

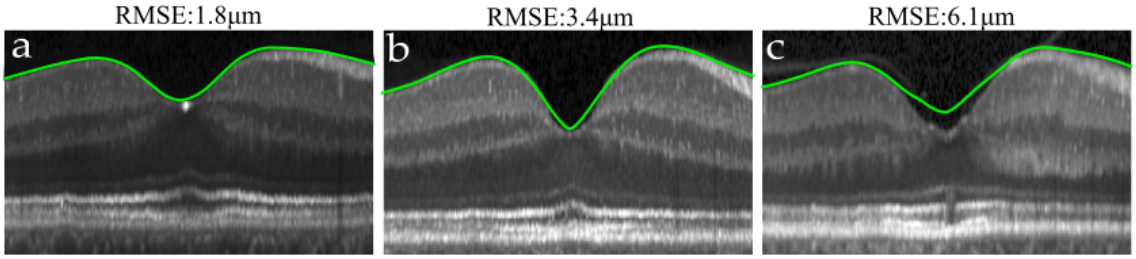


Figure 4.12: Examples of fitting errors. While a small error reduces the noise without altering the foveal structure, larger errors derived from inadequate modeling misrepresent the actual foveal shape. From [182].

It is worth noting that, on average, star scans presented a higher fitting error and exhibited more noise as depicted in Figure 4.13.

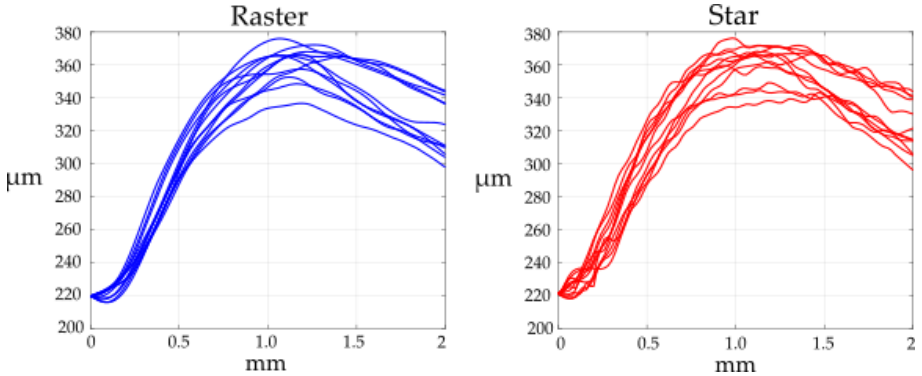


Figure 4.13: Raster and star TRT profiles. Star acquisitions are noisier, which can result in an overestimation of the slope. From [182].

The introduction of the modeling step resulted in an overall improvement in the agreement between raster and star scans. However, this was primarily notable for the rim radius and, partic-

ularly, the maximum slope, where the agreement substantially increased beyond an ICC of 0.95. The improvement of the ICC came at the cost of an estimation bias that varied between parameters: a slight overestimation of the CFT, a minimal underestimation of the rim height, and a more substantial underestimation of both the rim radius and maximum slope.

Figure 4.14 depicts the relationship between the introduced bias and the ICC through the LOESS curve. There was a rapid increase in ICC, particularly for the maximum slope, as the bias increased until reaching a point where the improvement became marginal. In this bivariate comparison, models introducing higher biases did not consistently result in proportionally improved ICC. For example, the model proposed by Scheibe et al. [77] introduced the highest bias in estimating the maximum slope (-19.8%), yet it achieved lower ICC agreement compared to approaches with smaller biases.

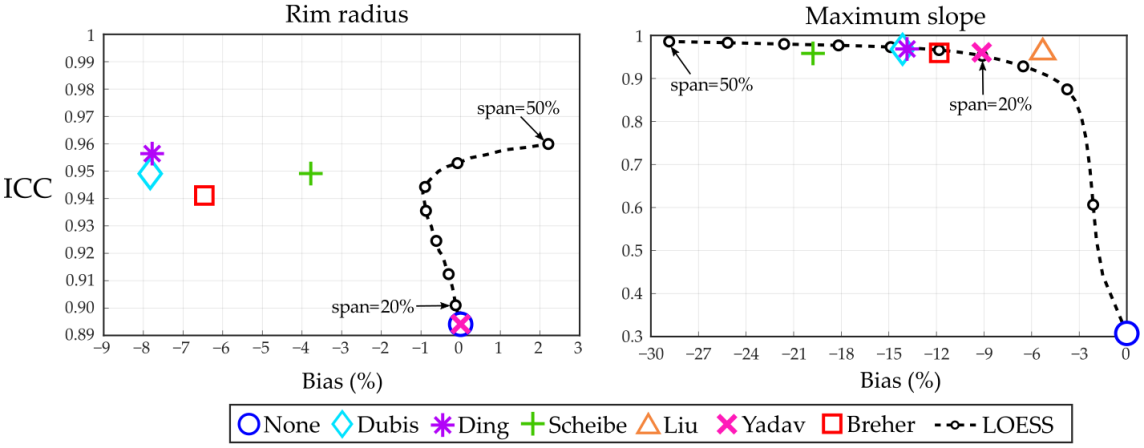


Figure 4.14: ICC as a function of the bias in raster scans. Each modeling strategy improves the agreement between acquisition at the cost of a certain bias in the estimation of foveal features. Adapted from [182].

On the other hand, LOESS smoothing demonstrated similar performance to most of the models. Specifically, even a mild degree of smoothing significantly improved the ICC of the maximum slope, while a higher degree of smoothing led to an overestimation of the CFT.

Choosing the right methodology

The suitability of mathematical models for characterizing the foveal pit is a topic of debate, and not all researchers opt to employ them. In principle, the rationale for the introduction of a mathematical model is twofold: firstly, it serves to reduce noise and obtain a smoother representation of the data. Secondly, it enables parametrization, allowing for the characterization of foveal pit morphology through the coefficients derived from the fitted equation.

As regarding the former, the strong agreement observed for both the CFT and rim height suggests that these metrics are sufficiently robust to characterize the foveal pit without the need for denoising. The slightly lower but still satisfactory agreement observed for the rim radius can be attributed to its susceptibility to segmentation errors. As there is minimal thickness variation at the foveal rim, even slight noise-induced irregularities can significantly affect the point of maximum thickness, and consequently impac radius measurements.

On the other hand, the poor agreement observed for the maximum slope suggests that slope metrics inherently exhibit more noise. Star scans, in particular, exhibited higher levels of irregularities, resulting in an overestimation of the maximum slope and thus a lower agreement. This behavior could potentially stem from interpolation errors during the resampling of the star pattern, whose sampling density decreases for larger eccentricities.

Importantly, the application of mathematical modeling or smoothing techniques improved the agreement, supporting the justification for incorporating a smoothing/modeling step when calculating slope metrics for noisy data. This improvement came at the expense of introducing fitting errors and estimation bias, however. Previous studies have compared models based on their fitting errors, aiming to achieve better accuracy than existing models. For example, studies by Liu et al. [75], Breher et al. and [76] demonstrated lower fitting errors than the model proposed by Dubis et al. [73]. Similarly, Yadav et al. [72], showed superior fitting of their model compared to those of Dubis et al. [73] and Ding et al. [74]. These differences were also observed in our study.

Nevertheless, in the context of denoising, the objective is not necessarily to achieve the best possible fit, but rather to introduce the smallest possible bias that enhances reliability. Given the absence of a ground truth reference, we adopted the following premise: when similar levels of agreement are achieved, the method with the lowest bias is preferable.

The first model proposed in the literature (Dubis et al. [73]) exhibited a notable bias in estimating both the rim radius and maximum slope, consistent with its known limitations in accurately capturing foveal asymmetries [75]. The model presented by Ding et al. [74] relied on a restricted set of eight parameters to model the entire TRT map, resulting in an underestimation of both the rim radius and maximum slope due to the inherent limitations of the model.

In the case of the radial model proposed by Scheibe et al. [77], although it achieved a low fitting error, it also exhibited the highest underestimation of the maximum slope. This could be attributed to the lack of flexibility of the model in capturing various foveal shapes effectively.

The model proposed by Liu et al. [75] yielded the best results for the maximum slope, with a nearly maximum ICC value and the smallest bias among the models. However, it also displayed the highest fitting error and a significant bias for the rim radius. This is likely because the model was specifically designed to account for flat pit bottoms by focusing on fitting the foveal pit region using a piecewise model. Consequently, its performance decreases when fitting data far from the foveal center.

The model with the highest fitting accuracy, as proposed by Yadav et al. [72], performed well by fitting the inner parts of each side of the B-scan separately. However, it should be noted that this model uses the foveal center and rim as references, which means that metrics derived solely from those landmarks, such as the CFT, rim height, or rim radius, are estimated as if no model were applied. Additionally, the fitting of cubic Bézier curves in this model introduces more complexity than the simpler equation fitting required by other models.

The model proposed by Breher et al. [76], which utilized the sum of three Gaussians, exhibited rigidity in both the rim radius and maximum slope. We observed that the fitting process of this model was highly sensitive to the initial coefficient estimation, possibly due to the higher number of coefficients (nine) involved.

Interestingly, our findings indicated that a simple LOESS smoothing approach could significantly reduce noise without introducing significant bias. The trade-off between bias and agreement was evident in the ICC vs. bias curve of LOESS, as excessive smoothing could distort the estimation while achieving high agreement.

When considering the utilization of model coefficients as parameters for characterizing foveal pit morphology, it is often preferable for these parameters to correspond to specific features of the foveal pit, so as to ensure clear interpretation of the analyses. In this regard, the coefficients employed in the model proposed by Scheibe et al. [77] can be regarded as the most intuitive, as they describe aspects of the fovea such as steepness. On a secondary level, certain coefficients defined in the models presented by Ding et al. [74], Dubis et al. [73], and Liu et al. [75] still retain interpretability. In contrast, the values defining Bézier curves [72] or the sum of three Gaussians [76] can be more complex to interpret.

It is worth noting that the analysis of the foveal pit encompasses various potential approaches,

and those reviewed in this study do not represent the entire spectrum. Therefore, future research should aim to expand the analyses presented here to include other parameters related to the foveal pit, as well as explore alternative modeling and smoothing techniques.

When analyzing foveal pit morphology using mathematical models, it is essential to approach the task with careful consideration. In this regard, the following guidelines can be employed: if the objective is to characterize the foveal pit using model coefficients, it is advisable to select a model whose coefficients are easy to interpret or align closely with the research question at hand. Alternatively, in cases where parametrization is not desired, such as when examining fundamental parameters like maximum slope or rim height, it is important to assess whether denoising is necessary. This can be determined by visually inspecting the data. If denoising is required, it is advisable to try first a simple smoothing technique and consider how it may bias parameter estimation.

4.4 RETIMAT Toolbox development

Introduction

OCT image analysis involves multiple steps and usually requires the use of multiple software tools. For instance, to compute thickness features it might be necessary to use different software for file reading, segmentation, and feature extraction. Problematically, existing software libraries may not be able to directly communicate between each other and custom code is often needed to bridge this gap. Moreover, using more than one tool complicates the process and makes it difficult to build reproducible pipelines (i.e., creating full end to end workflows that always produce the same result).

Hence, there exists a need to develop a single software toolbox that could perform all or most of the processing steps. To this end, we integrated all previously described OCT processing algorithms into an open-source MATLAB toolbox called Retinal Image Analysis in MATLAB (RETIMAT) (`https://github.com/drombas/retimat`). The toolbox has already been presented to the community in conferences [183, 187] and the design is described below.

High-level design

An overview of RETIMAT is shown in Figure 4.15. It was designed as an easy-to-use application programming interface composed of independent functions that fulfill specific tasks in the OCT analysis pipeline. These functions are grouped into modules with similar logic (input/output, visualization, etc.) so that custom OCT processing pipelines can be built depending on the use case.

Target users are expected to be familiar with scripting languages to make use of advanced features. However, the toolbox includes a detailed documentation in the form of tutorials to guide less experienced users in the most common tasks.

Implemented functions

File reading

RETIMAT provides an interface to read proprietary OCT files from the main vendors: Heidelberg (*e2e* and *vol*), Topcon (*fda*), and Zeiss (*img*). It also supports segmentation data obtained by the IOWA Reference Algorithm implemented in OCTExplorer 3.8.0 software (*xml*) [26, 87–89].

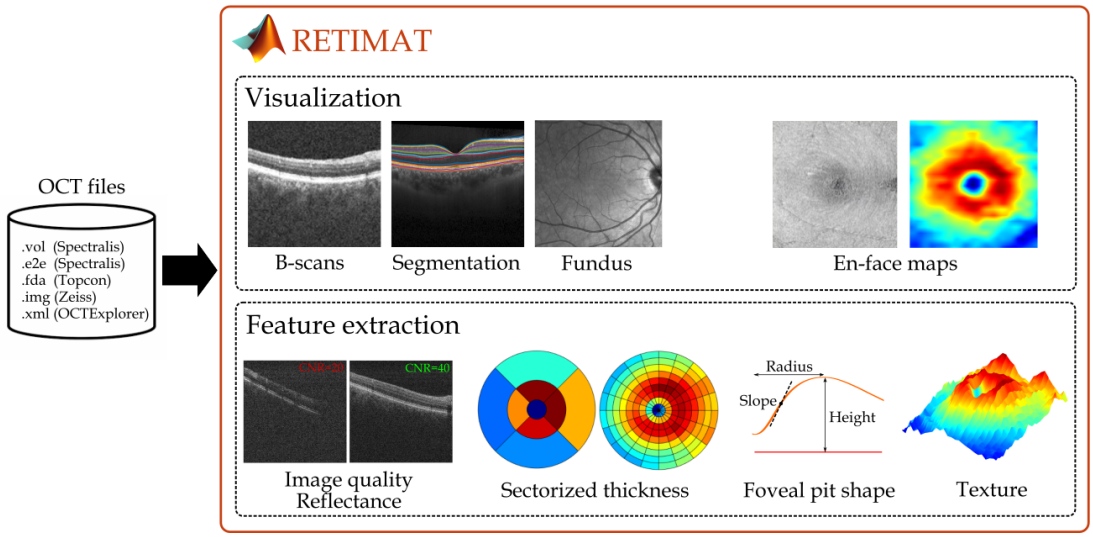


Figure 4.15: High-level design of RETIMAT. RETIMAT reads files from different devices and includes different modules for visualization and feature extraction. Each module comprises a set of independent functions that can be used to create custom processing pipelines.

Data read from any of the supported formats is loaded into MATLAB with the same structure, which facilitates building multi-vendor pipelines.

It is important to mention that most of the aforementioned file formats are not disclosed and the implementation in RETIMAT has only been possible as a result of previous laudable efforts to parse every file format: AURA tools [29,95], Unified OCT explorer [188], and OCT-Converter [189].

Visualization

Summary figures can be created with the visualization module as a quick way of looking into an OCT volume to identify poor image quality, ocular lesions, or segmentation errors. The report consists of a fundus image, several en-face reflectance and thickness maps, and actual B-scans with segmentation. The number of layers shown can be fully specified by the user and the figure can be either visualized interactively or saved into a *png* file. In addition, there are also functions to visualize sectorized thickness or radial foveal pit parameters (see Figure 4.17).

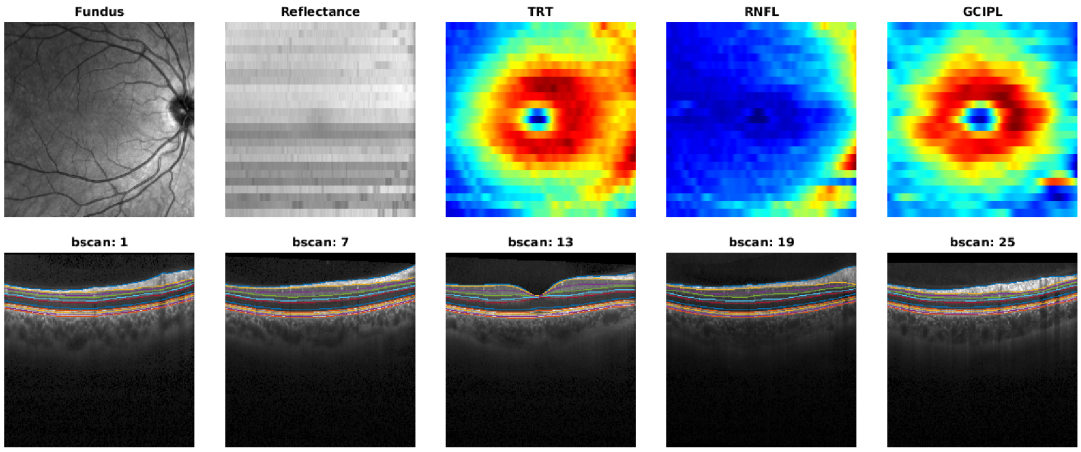


Figure 4.16: Example summary figure generated by RETIMAT. The figure shows the fundus image, a reflectance map, three thickness maps and 5 B-scans. The number of thickness maps and B-scans can be customized. The figure can be either rendered or saved into memory.

Preprocessing

The preprocessing part consists of functions to segment retinal layers, detect the foveal center automatically (Section 4.2), compute thickness from segmentation data, and resample 2D maps into different grids (e.g., transforming a raster grid into a star grid for radial analysis). Retinal layer segmentation is supported via an interface with the AURA tools segmentation algorithm [29, 95]. This algorithm was developed in MATLAB and has been widely used by the research community since its validation. Using the developed interface, segmentation can be integrated into a single processing pipeline without the need for external tools. In addition, several utility functions are also available for internal file handling and data format conversion.

Feature extraction

The largest module of the toolbox includes functions to compute numerical features that describe different aspects of retinal images:

- **Image quality:** image contrast metrics computed from segmentation and pixel intensity (SNR, contrast to noise ratio, etc.).
- **Sectorized thickness:** point thickness values averaged into arbitrary parcellations (see Section 4.5).
- **Foveal pit morphology:** features from two categories: 1) geometrical features such as maximum slope or rim height, and 2) mathematical coefficients derived from fitting the models from the literature [72–77].
- **Texture features**: four types of metrics computed from 2D en-face maps: fractal dimension, lacunarity, GLCM features, and LBP features.
- **Reflectance:** image intensity or reflectance used to build en-face maps from which sectorized and texture features can be computed.

Applications of RETIMAT and future work

RETIMAT is the backbone of every analysis presented in chapters 5, 6 and 7. In particular, it has been used to analyze all the images from the BHRI dataset and extract the features used in data analyses. Additionally, part of RETIMAT has been translated into Python and used to build the main OCT feature extraction pipeline for more than 1 million OCT images in the AlzEye database [178]. The computed features have facilitated an important body of research investigating multiple clinical conditions, part of which has already been published [190]. Both BHRI and AlzEye feature extraction pipelines are described in Section 4.5.

These examples illustrate how quality scientific software facilitates research. Indeed, by releasing RETIMAT as open-source software we intend to contribute to the community by providing a tool that advances OCT image analysis.

As future lines of work, one of the main goals would be to integrate already implemented functions into a single workflow that could be run without coding experience. This workflow would automatically read the images, perform quality screening, generate exports for visualization, preprocess the images, and compute a set of features for further analysis. As a limitation, RETIMAT does not currently incorporate state of the art retinal layer segmentation models and relies on the segmentation data stored in proprietary OCT images and AURA segmentation algorithm [29, 95]. Implementing such a model is one of the future improvements that would make RETIMAT a more efficient standalone tool.

4.5 Enhanced parametrization

In this section we describe how we used RETIMAT to process the images from both BHRI and AlzEye datasets to extract numerical features for data analysis. It is important to note that this section describes the entire spectrum of computed features, but different subsets of them were used throughout the thesis depending on the research question.

BBHRI

Preprocessing

To ensure proper alignment, all macular scans were automatically aligned by locating the foveal center automatically (refer to Section 4.2). The lateral scale of macular images was adjusted for ocular magnification by means of the built-in Spectralis software, who adjusts lateral resolution of each eye based on an estimation of the refractive error [122]. Left eye scans were flipped horizontally to match the orientation of the right eyes. From layer segmentation data provided by the device, thickness values were computed for each A-scan and the following layers: TRT, NFL, GCIPL, INL, outer nuclear and plexiform layer (ONPL), and external limiting membrane-Bruch's membrane (ELM-BM). Point thickness values were then interpolated onto a regular grid of 200 x 200 points covering the same macular region to ensure isotropic sampling.

From these thickness maps, features from three categories were computed: sectorized thickness, foveal pit morphology, and texture.

Sectorized thickness

Retinal thickness is typically studied by averaging A-scan values across a certain macular region to reduce the number of values to analyze. The most commonly employed sectorization is the so-called ETDRS. Although using this approach facilitates comparison across studies, it limits the description of the macular region to only nine parameters. To obtain a more comprehensive characterization of the macula, here we used 6 different macular grids with varying spatial resolution (see Figure 4.17):

- *Macula*: fovea centered disk of 3 mm radius.
- *ETDRS*: nine sectors split into four quadrants and two rings of 0.5, 1.5 and 3 mm radii (9 features).
- *ETDRS rings*: parafovea (inner ring with 0.5 to 1.5 mm radii), and perifovea (outer ring with 1.5 to 3 mm radii) (3 features).
- *5 rings*: a set of five concentric rings with a 0.5 mm spacing between radii.
- *12 angles*: radial sectorization with 12 angular directions.
- *Regular*: 20 x 20 square sectors.

The choice of sectorization depends on the research question and the initial hypothesis. For instance, if the effect under study is expected to be localized in a concrete region it may be more effective to use smaller sectors. All 6 grids were used to compute average thickness values for previously listed retinal layers. The total number of obtained thickness values per layer was 430.

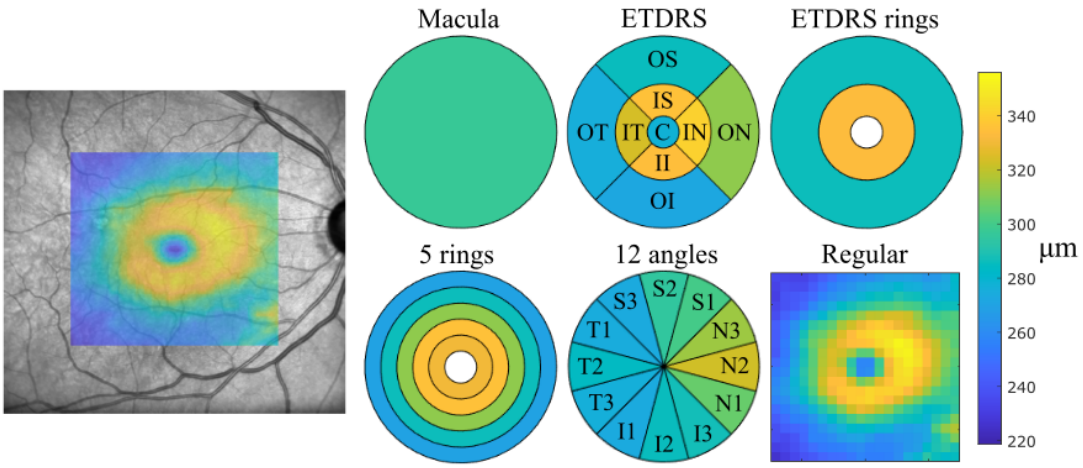


Figure 4.17: Sectorizations of the macula. Left image shows a fundus image with a colored macular thickness map. Right plots depict different sectorization grids used to analyze the raw thickness map.

Foveal pit morphology

Foveal pit features can provide complementary information about the retinal structure. There are two types of foveal features: geometrical parameters describing intuitive aspects of the fovea (e.g., slope or depth), and equation coefficients derived by fitting a mathematical model to the foveal pit shape. Here the shape of the foveal pit formed by both TRT and GCIPL layers was analyzed. This is possible because the thickness distribution of both layers follows a convex shape that can be analyzed by existing approaches. In either case, point thickness values were interpolated onto a radial grid consisting of 24 angular directions, a radius of 2.5 mm, and 100 points per direction. This was necessary to characterize the foveal pit radially.

Initially, the CFT was computed as the thickness value at the foveal center. Then, to reduce segmentation ripple, LOESS smoothing (span = 50%) was applied individually to the thickness profiles for each radial direction. These smoothed profiles were used to determine the following features:

- *Rim height*: highest thickness value.
- *Rim radius*: lateral distance from the foveal center to the rim.
- *Rim disk perimeter*: perimeter of the disk created by all rim height points.
- *Rim disk area*: area of the disk created by all rim height points.

Finally, original thickness profiles were smoothed again using a lighter smoothing (span = 15%). The difference in the smoothing is in line with the results of Section 4.3 that highlight how a larger smoothing is beneficial for radius determination but highly biases slope-related estimations. Accordingly, this lightly-smoothed thickness profiles were used to estimate the next features:

- *Mean slope*: the average first derivative between the foveal center and the rim.
- *Max slope*: the maximum first derivative between the foveal center and the rim.
- *Max slope height*: thickness at the point of maximum slope.
- *Max slope radius*: lateral distance from the foveal center to the point of maximum slope.
- *Pit area*: area filled by the foveal depression.
- *Max slope disk perimeter*: perimeter of the disk created by all rim height points.

- *Max slope disk area*: area of the disk created by all maximum slope height points.

Global features (i.e., CFT and any disk area/perimeter feature) were computed once for the whole fovea. Conversely, radial features such as slope or radius were derived for each direction separately as illustrated in Figure 4.18. These were then averaged over both the entire macula and 4 ETDRS quadrants. In addition to the mean, the standard deviation of the radial features was also computed.

To complement geometrical features, the mathematical models developed by Dubis et al. [73], Ding et al. [74], Scheibe et al. [77], and Breher et al. [76] were fitted to the data (refer to Table 2.1 for a detailed model description). The fitted coefficients of each model were extracted and used as features. Although some of these features do not have a direct morphological interpretation, they can potentially describe more subtle aspects of the foveal shape. As with geometrical features, radial model coefficients were averaged across the same angular directions used for geometrical features.

Texture

Texture features are effective to describe high-order statistical properties of thickness maps. Here, we computed 5 types of features that have already been used in OCT analysis:

- *Fractal dimension*: computed using the box-counting algorithm [81].
- *Lacunarity*: computed for box sizes of 2, 4, 8, 16 and 32 following the procedure described by Roy and Perfect [82].
- *Standard deviation*: derived as a simple measurement of the variability of thickness values in a certain sector.
- *LBP*: the LBP histogram was built considering the surrounding 8 neighbors of each pixel. Then, seven statistical measurements were derived from the histogram: mean, median, standard deviation, interquartile range, kurtosis, skewness, and entropy [79].
- *GLCM*: we computed a total of 21 GLCM features [78, 191, 192] from the average of four GLCM matrices derived by comparing adjacent pixels horizontally, vertically, and in both diagonals. The number of grayscale levels was set to 60. The derivation of each GLCM feature is described below.

As a first step to compute GLCM features, the GLCM matrix (G) was transformed into a 2D probability distribution as:

$$P_{i,j} = \frac{G_{i,j}}{\sum_{i=1}^{N}\sum_{j=1}^{N} G_{i,j}} \tag{4.5}$$

where N=60, and i and j are the row and column indexes, respectively. From this matrix we can already compute a first set of 11 features:

$$Autocorrelation = \sum_{i=1}^{N}\sum_{j=1}^{N} P_{i,j}\, i\, j \quad [191]$$

$$Contrast = \sum_{i=1}^{N}\sum_{j=1}^{N} P_{i,j}(i-j)^2 \quad [78]$$

$$Dissimilarity = \sum_{i=1}^{N}\sum_{j=1}^{N} P_{i,j}|i-j| \quad [191]$$

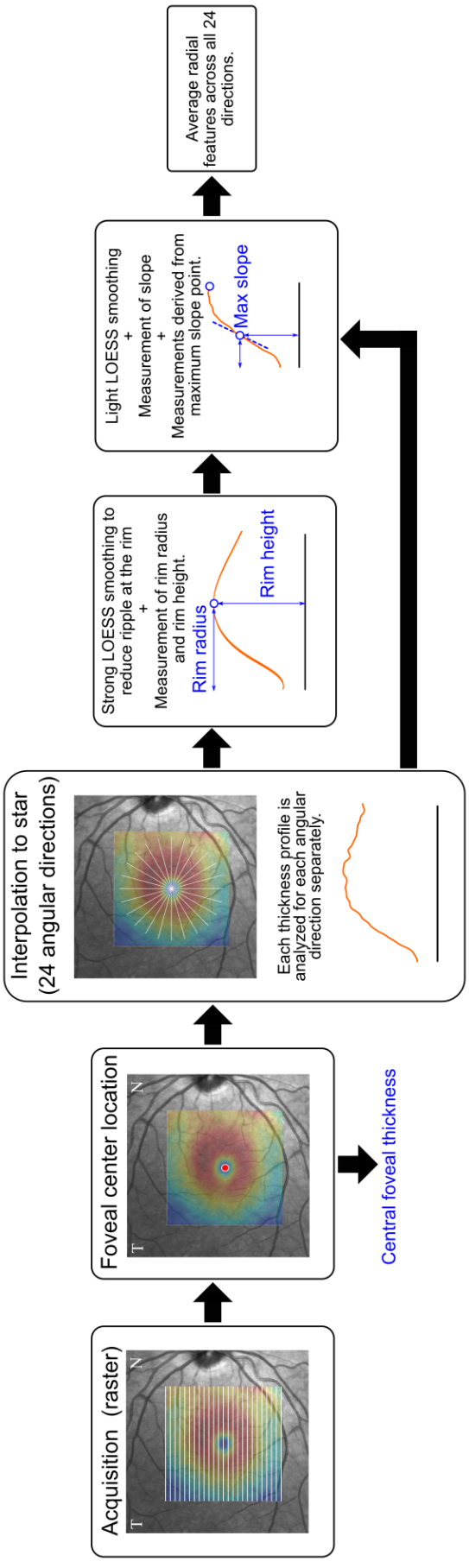


Figure 4.18: Foveal pit morphology analysis pipeline. After acquiring the images the foveal center is located. From this the central foveal thickness (CFT) is directly estimated. Then, the original thickness map is interpolated into a star pattern. Obtained radial thickness profiles are smoothed by strong and light LOESS. The former is used to estimate the position of the rim and derive features related to it. The latter is employed for slope measurements. Radial parameters are finally averaged over multiple angles. Adapted from [183].

$$Energy = \sum_{i=1}^{N}\sum_{j=1}^{N} P_{i,j}^2 \quad [78]$$

$$Entropy = entropy\ of\ P_{i,j} \quad [78]$$

$$Inverse\ Difference = \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{P_{i,j}}{1+|i-j|} \quad [192]$$

$$Inverse\ Difference\ Normalized = \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{P_{i,j}}{1+\frac{|i-j|}{N}} \quad [192]$$

$$Inverse\ Difference\ Moment = \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{P_{i,j}}{1+(i-j)^2} \quad [78]$$

$$Inverse\ Difference\ Moment\ Normalized = \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{P_{i,j}}{1+(\frac{i-j}{N})^2} \quad [192]$$

$$Joint\ Variance = \sum_{i=1}^{N}\sum_{j=1}^{N} P_{i,j}(i-autocorrelation)^2 \quad [78]$$

$$Maximum\ Probability = max(P_{i,j}) \quad [191]$$

The derivation of a second set of features requires first the computation of the average index for both rows (μ_i) and columns (μ_j), as well as the corresponding standard deviations (σ_i, σ_j):

$$\mu_i = \sum_{i=1}^{N}\sum_{j=1}^{N} P_{i,j}\ i \quad (4.6)$$

$$\mu_j = \sum_{i=1}^{N}\sum_{j=1}^{N} P_{i,j}\ j \quad (4.7)$$

$$\sigma_i = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} P_{i,j}\ (i-\mu_i)^2} \quad (4.8)$$

$$\sigma_j = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} P_{i,j}\ (j-\mu_j)^2} \quad (4.9)$$

From these quantities we derive 4 additional features as:

$$Cluster\ prominence = \sum_{i=1}^{N}\sum_{j=1}^{N} P_{i,j}(i+j-\mu_i-\mu_j)^4 \quad [191]$$

$$Cluster\ shade = \sum_{i=1}^{N}\sum_{j=1}^{N} P_{i,j}(i+j-\mu_i-\mu_j)^3 \quad [191]$$

$$Correlation = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} P_{i,j}ij - \mu_i\mu_j}{\sigma_i\sigma_j} \quad [78]$$

$$Sum\ Of\ Squares = \sum_{i=1}^{N}\sum_{j=1}^{N} P_{i,j}(i-\mu_i)^2 \quad [191]$$

In a third category, we derive features from the distribution of index sum (P_{i+j}) and difference (P_{i-j}). That is from two 1D histograms describing the occurrence of each possible values of $|i-j|$ and $i+j$:

$$Sum\ Average = average\ of\ P_{i+j} \quad [78]$$

$$Sum\ Entropy = entropy\ of\ P_{i+j} \quad [78]$$

$$Difference\ Variance = variance\ of\ P_{i-j} \quad [78]$$

$$Difference\ Entropy = entropy\ of\ P_{i-j} \quad [78]$$

The final two GLCM features are derived from an information theory framework and require computing the entropy of row index probability (H_i), column index probability (H_j), $P_{i,j}$ matrix ($H_{i,j}$), and the product distribution ($H'_{i,j}$):

$$Informational\ Measure\ of\ Correlation\ 1 = \frac{H_{i,j} - H'_{i,j}}{max(H_i, H_j)} \quad [78]$$

$$Informational\ Measure\ of\ Correlation\ 2 = \sqrt{1 - e^{-2(H'_{i,j} - H_{i,j})}} \quad [78]$$

All the texture features were extracted for all the 490 sectors and retinal layers listed before when describing sectorized thickness computation.

AlzEye

AlzEye dataset contains 1348934 macular images that were processed with a similar pipeline. In this case, the foveal center was automatically located by Topcon proprietary TABS software. Further preprocessing was the same as for BHRI dataset. As an important difference, Topcon metadata did not provide any method to account for ocular magnification and, therefore, a fixed lateral scale was assumed for all images. More importantly, in AlzEye only the following subset of features were computed due to compute and memory constrains consequence of the large number of images:

- *Sectorized thickness*: the analyzed layers were: TRT, NFL, GCIPL, INL, ONPL, and ELM-BM. The used sectorizations were: macula, ETDRS, ETDRS rings, and *5 Rings* (see Figure 4.17).
- *Foveal pit morphology*: four parameters were analyzed for both TRT and GCIPL layers: rim height, rim radius, pit depth, and mean slope.

The obtained features were employed to investigate longitudinal changes in PD (see Chapter 7). In addition, they have also been used to investigate multiple research conditions by other AlzEye project members [190, 193].

4.6 Validation of the macular raster acquisition protocol

An important factor that may affect OCT features is the sampling density of the acquisition protocol. The standard macular protocol used in BHRI consists of 25 B-scans, a number that is significantly lower than existing high-density protocols. In this section we investigate the potential bias introduced into OCT features by using this standard protocol.

A subset of 12 healthy subjects underwent an imaging session where both eyes were captured twice in the same day using standard and high-density protocols (Table 4.10). All other scan parameters remained unchanged.

Protocol	Region	B-scan	A-scan
Standard	6 x 6 mm^2	25	512
High-density	6 x 6 mm^2	97	1024

Table 4.10: Comparison of standard vs high-density protocol. The high density protocol uses a larger number of sampling points at the cost of a longer acquisition time.

The images acquired from both protocols were used to compute the next OCT features as described in Section 4.5.

- *Sectorized thickness*: computed for the TRT, GCIPL, INL, ONPL, and ELM-BM. We evaluated both the average thickness using both the entire macular region and a 20 x 20 square grid.
- *Foveal pit features*: CFT, rim height, rim radius, and mean slope computed from the TRT.

To measure the bias introduced by using the standard protocol, a mixed-effects model linear regression was fitted to each parameter with a fixed term for the bias (β_{bias}) and a random intercept ($\gamma_{subject}$) to account for inter-eye correlation:

$$y = \beta_0 + \beta_{bias} isStandard + \gamma_{subject} \tag{4.10}$$

The estimated bias introduced in each average thickness and foveal parameter due to the standard protocol is shown in Table 4.11. Corresponding results for the 20 x 20 sectorization are presented in Figure 4.19.

The findings reveal a small bias for average macular thicknesses ($< 1\%$) and the rim height. However, in the case of the CFT and particularly the foveal slope, the bias was found to be non-negligible for (3.02% and -6.57%, respectively). It should be noted that the use of smaller sectors (i.e., 20 x 20 grid) resulted in an overall bias below 5%, with the greatest effect observed in the center and outer regions of the NFL, INL, and GCIPL.

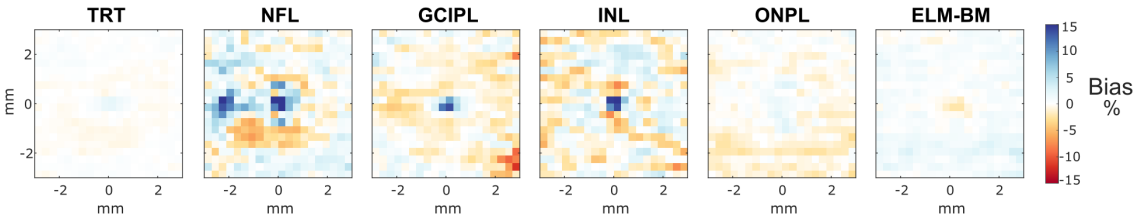


Figure 4.19: Sensitivity analysis results for the 20 x 20 grid. Top row shows the relative bias introduced in each retinal layer and region.The NFL, GCIPL, and INL layers are more affected by using a standard protocol. Adapted from [183].

Layer / Parameter	**Absolute bias**	**Relative bias (%)**	**p-value**
TRT (μm)	0.1 [-0.93, 1.14]	0.03 [-0.31, 0.37]	0.84
NFL (μm)	0.11 [-0.37, 0.59]	0.36 [-1.19, 1.9]	0.64
GCIPL (μm)	-0.4 [-0.88, 0.08]	-0.57 [-1.26, 0.11]	0.099
INL (μm)	0.02 [-0.27, 0.32]	0.07 [-0.83, 0.97]	0.87
ONPL (μm)	-0.18 [-0.59, 0.24]	-0.19 [-0.64, 0.26]	0.4
ELM-BM (μm)	0.55 [0.11, 0.98]	0.69 [0.14, 1.24]	0.015*
CFT (μm)	6.72 [4.88, 8.57]	3.02 [2.19, 3.84]	$3{\cdot}10^{-9}$*
Rim height (μm)	-0.79 [-2.34, 0.75]	-0.23 [-0.67, 0.21]	0.31
Rim radius (mm)	-0.01 [-0.04, 0.01]	-1.26 [-3.23, 0.7]	0.2
Mean slope (°)	-0.45 [-0.65, -0.25]	-6.57 [-9.51, -3.64]	$5{\cdot}10^{-5}$*

* statistically significant.

Table 4.11: Bias introduced by the standard macular raster protocol. The difference in the estimation obtained by the standard and the high-density protocol (i.e., relative bias) is below 1% for all macular thicknesses. However, this bias is not negligible for foveal pit parameters. CFT: central foveal thickness. From [183].

Upon the results, it can be concluded that using only 25 B-scans undersamples the central region, resulting in a systematic overestimation of central thicknesses and an underestimation of foveal slope measurements. However, the introduced bias is relatively small in most cases, with less than a 5% deviation from the true value. In conclusion, we considered the standard protocol to be a appropriate because it introduces a small bias and requires a shorter scanning time, which is particularly important when imaging subjects with neurodegenerative diseases.

5 | Retinal morphology in a healthy population

Previous research conduced with PD patients has revealed retinal changes associated with the disease. However, demographical factors such as age and sex may also influence structural aspects of the macula. It is therefore crucial to accurately determine how the retina evolves on healthy subjects to be able to discern disease-related changes. In this regard, literature investigating the effect of age and sex on healthy cohorts is limited for several reasons. Firstly, most studies examining macular thickness have relied on the standard ETDRS sectorization, and there is a lack of comprehensive investigation into the impact of sex and age on the geometry of the foveal pit.

This chapter aims to bridge this knowledge gap by investigating age-related changes and sex differences on the retinal morphology with high spatial detail. As a key factor, we studied macular thickness with a 20 x 20 square grid sectorization and the foveal pit morphology using 24 angular directions. This was performed to extend existing literature by providing a more detailed normative database compared to the existing literature. All the results presented here were published by the author in [183].

5.1 Subjects and OCT features

A total of 444 healthy subjects from the BHRI dataset were included in the analysis (see Table 5.1). All the images were acquired following a macula raster acquisition previously described in Section 3.3.

Group	Subjects	Eyes	Age (years)*
All	444	855	54.9 ± 12.7 [21,88]
Female	281	543	54.3 ± 12.6 [22,88]
Male	163	312	56.0 ± 12.8 [21,87]

* mean $\pm$ σ [range].

Table 5.1: Subjects included in the healthy population study. Adapted from [183].

The investigated OCT features are illustrated in Figure 5.1 and were:

- *Sectorized thickness*: the investigated layers are depicted in Figure 5.1a and were: TRT, NFL, GCIPL, INL, ONPL, and ELM-BM. For each layer, sectorized thickness values derived from the whole macula as well as a 20 x 20 square grid were analyzed (see Figure 5.1c). The latter sectorization excluded areas outside the fovea-centered circle of $3\,mm$. Furthermore, the thickness measurements of the NFL, GCIPL, and INL layers were not analyzed within the central sectors (a region of 1.2 x $1.2\,mm$ at the center). This decision was made because the inner layers in the central foveal region have almost zero thickness, and including these sectors could introduce significant bias due to segmentation errors.

- *Foveal pit morphology*: four features were analyzed: the CFT, rim height, rim radius, and mean slope. These four were derived from the TRT profile were analyzed for the entire fovea and 24 angular directions separately (see Figure 5.1d).

The concrete procedure followed to compute the aforementioned features is described in detail in Section 4.5. The features were computed for both eyes of each subject.

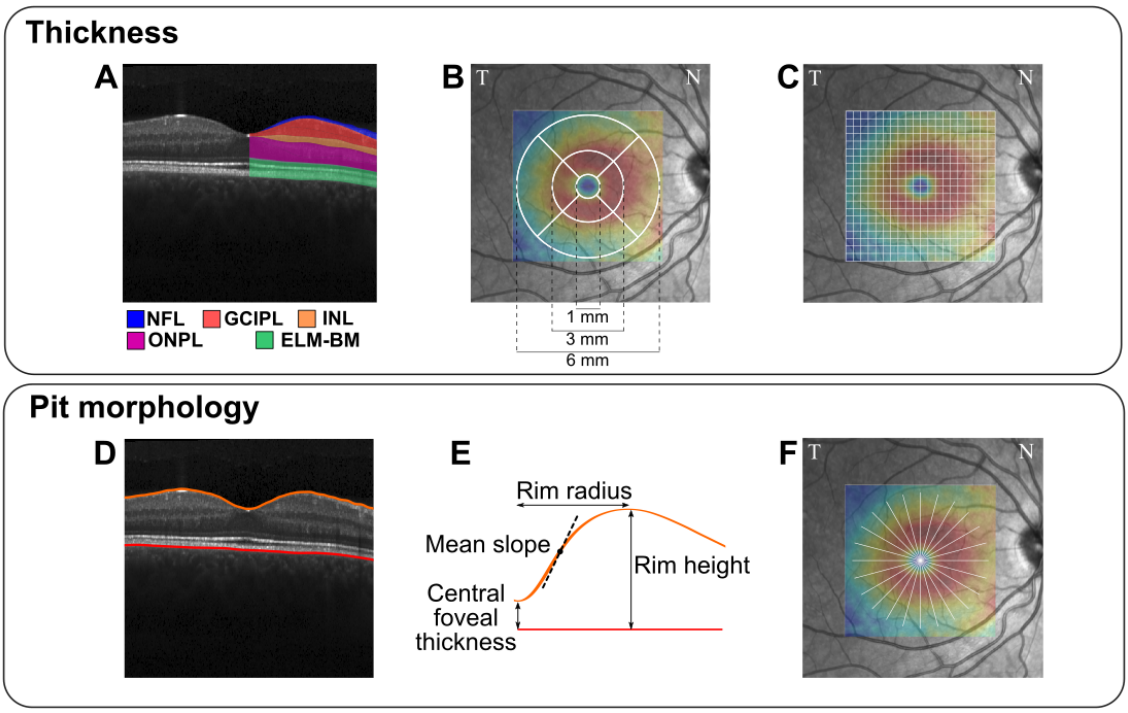


Figure 5.1: Summary of parameters extracted from macular OCT images. A) Retinal layers. **B)** ETDRS sectorization. **C)** 20 x 20 square grid sectorization. **D)** Segmentation of top and bottom boundaries of the retina. **E)** Studied foveal pit geometrical features. **F)** Foveal pit radial analysis. From [183].

5.2 Data analysis

The impact of sex and age on thickness maps and foveal pit parameters was analyzed in both absolute and percentage terms. A mixed-effects multivariate regression analysis was employed, incorporating fixed terms for age and sex. Females were set as the reference category in the latter. A fixed term for the Spectralis scan focus variable was also included to address variations in ocular shape. Determined by the scanner during image focusing, this variable accounts for the refractive error specific to each eye, which can influence retinal measurements. To consider the correlation between eyes, a random intercept ($\gamma_{subject}$) was added for each subject. As a preliminary step in model selection, two models were fitted with linear (Equation 5.1) and quadratic (Equation 5.2) age effects. From these, the model with the lowest Akaike information criterion (AIC) value was chosen.

$$y = \beta_0 + \beta_{sex} isMale + \beta_{age} age + \beta_{sf} scanFocus + \gamma_{subject} \tag{5.1}$$

$$y = \beta_0 + \beta_{sex} isMale + \beta_{age} age + \beta_{age2} age^2 + \beta_{sf} scanFocus + \gamma_{subject} \tag{5.2}$$

To provide comparable results between linear and quadratic age models, a combined age effect coefficient was estimated, representing the average yearly change between the ages of 40 and 80. A 95% CI and a corresponding p-value were calculated for each coefficient. In cases where the selected model was quadratic, a single p-value was computed to assess the combined linear and quadratic age effect. This was done by means of an F-test, which compares a reference model without any age term to the quadratic model.

To express the age coefficients in percentage values, absolute coefficients were divided by the average parameter value in the youngest age group (age < 40), which included 51 subjects out of the total 444. A similar procedure was carried out to obtain percentages for the sex coefficients, i.e., these were divided by the estimate for females in the youngest age group (age < 40, n $= 37$).

The significance level was set at 0.05 and the Holm-Bonferroni correction was used to adjust for multiple comparisons for the macular region analysis [194]. For both the ETDRS and high-resolution sectorizations, a correction based on the false discovery rate (FDR) [195] was applied due to the large number of sectors and potential statistical dependence between tests. The marginal R^2 was used to evaluate the model fit. This method measures the proportion of variance explained by the fixed terms alone. Since changes measured on an annual basis were found to be minimal, we report the age coefficient as changes per 10 year period.

5.3 Results

Thickness analysis

Figure 5.2 plots the average thickness of each retinal layer in the macular region with respect to age. The corresponding regression coefficient estimates, p-values, and the R^2 of the model are presented in Table 5.2.

The results indicate that all layers showed a significant decrease in thickness with age. The exception to this trend was the NFL, which exhibited non-significant thickening and high dispersion . The thinning effect was most pronounced for the TRT, GCIPL, and INL layers. The GCIPL reported the highest percentage loss (-2.41 [-2.90, -1.90]% per 10 years), contributing significantly to the overall reduction in TRT (-1.05 [-1.32, -0.77]% per 10 years).

It was observed that male TRT was 4.14 [1.78, 6.50] μm greater than that of females. This difference results from variations in all individual retinal layers, although statistically significant differences between males and females were only found in the INL, ONPL, and ELM-BM layers.

The results of the analysis using the 20 x 20 grid are shown in Figure 5.3. The age-related thinning of the TRT, GCIPL and ONPL appeared to be relatively consistent across the macular region, indicating a homogeneous pattern. However, the age-related reduction in thickness for the INL was more evident in the outer ring of the macula. The NFL maps reported mild thickening, particularly in the temporal sector. Conversely, the changes in the ELM-BM layer were minor and not statistically significant, except for the central foveal region where a slight to moderate thinning was observed. The regions where a quadratic model was selected are illustrated in Figure 5.4.

Layer	Age dependence	R^2 (%)	Age			Sex		
			β [95% CI]		p-value	β [95% CI]		p-value
			μm/ 10 years	% / 10 years		μm (male)	% (male)	
TRT	Quadratic	8.2	-3.25 [-4.09, -2.39]	-1.05 [-1.32, -0.77]	$9{\cdot}10^{-13*}$	4.14 [1.78, 6.50]	1.34 [0.58, 2.11]	$5{\cdot}10^{-4*}$
NFL	Linear	0.2	0.22 [-0.04, 0.48]	0.69 [-0.14, 1.52]	0.1	-0.03 [-0.71, 0.65]	-0.10 [-2.22, 2.03]	0.9
GCIPL	Quadratic	11.9	-1.77 [-2.13, -1.40]	-2.41 [-2.90, -1.90]	$3{\cdot}10^{-22*}$	0.57 [-0.41, 1.56]	0.78 [-0.56, 2.13]	0.3
INL	Quadratic	7.1	-0.45 [-0.60, -0.30]	-1.31 [-1.74, -0.88]	$9{\cdot}10^{-9*}$	0.87 [0.46, 1.28]	2.57 [1.36, 3.77]	$3{\cdot}10^{-5*}$
ONPL	Linear	2.6	-0.66 [-1.12, -0.19]	-0.73 [-1.24, -0.21]	0.006*	1.80 [0.59, 3.02]	2.01 [0.66, 3.36]	0.004*
ELM-BM	Quadratic	6.0	-0.45 [-0.63, -0.26]	-0.56 [-0.78, -0.33]	$2{\cdot}10^{-4*}$	0.93 [0.48, 1.39]	1.17 [0.60, 1.74]	$7{\cdot}10^{-5*}$

* statistically significant after Holm-Bonferroni correction (number of tests: 12, $\alpha = 0.05$).

Table 5.2: Regression results of mean macular layer thickness. Each row reports age and sex regression coefficients along with the obtained R^2. Sex coefficients are reported as the difference of male minus female. From [183].

Parameter	Age dependence	R^2 (%)	Age			Sex		
			β [95% CI]		p-value	β [95% CI]		p-value
			X / 10 years	% / 10 years		X (male)	% (male)	
CFT (μm)	Quadratic	4.8	-1.44 [-2.93, 0.03]	-0.62 [-1.25, 0.01]	0.08	7.99 [4.22, 11.77]	3.47 [1.83, 5.11]	$3{\cdot}10^{-5*}$
Rim height (μm)	Quadratic	11.9	-3.42 [-4.34, -2.51]	-0.97 [-1.22, -0.71]	10^{-11*}	8.63 [5.94, 11.33]	2.46 [1.69, 3.23]	$5{\cdot}10^{-10*}$
Rim radius (μm)	Linear	8.4	-7.66 [-14.99, -0.33]	-0.69 [-1.36, -0.03]	0.04	-59.4 [-78.34, -40.46]	-5.24 [-6.91, -3.57]	10^{-9*}
Mean slope (°)	Linear	4.1	-0.06 [-0.15, 0.02]	-0.99 [-2.27, 0.29]	0.15	0.39 [0.18, 0.6]	6.23 [2.84, 9.63]	$3{\cdot}10^{-4*}$

* statistically significant after Holm-Bonferroni correction (number of tests: 8, $\alpha = 0.05$).

Table 5.3: Regression results for foveal pit morphology. Each row reports age and sex regression coefficients along with the obtained R^2. Sex coefficients are reported as the difference of male minus female. CFT: central foveal thickness. From [183].

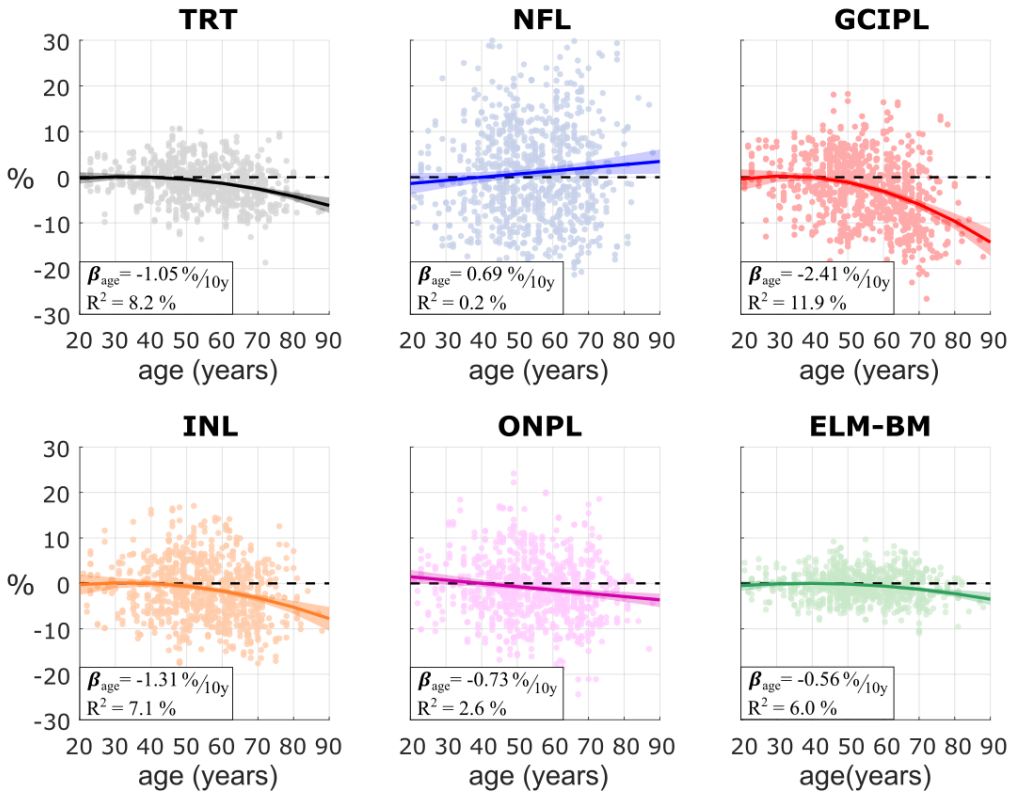


Figure 5.2: Percentual change in macular thicknesses as a function of age. Thickness values were computed for the $3\,mm$ radius circular macular region. Individual absolute thickness values were transformed into percentages as the relative difference with respect to the average thickness in the youngest group (age < 40). From [183].

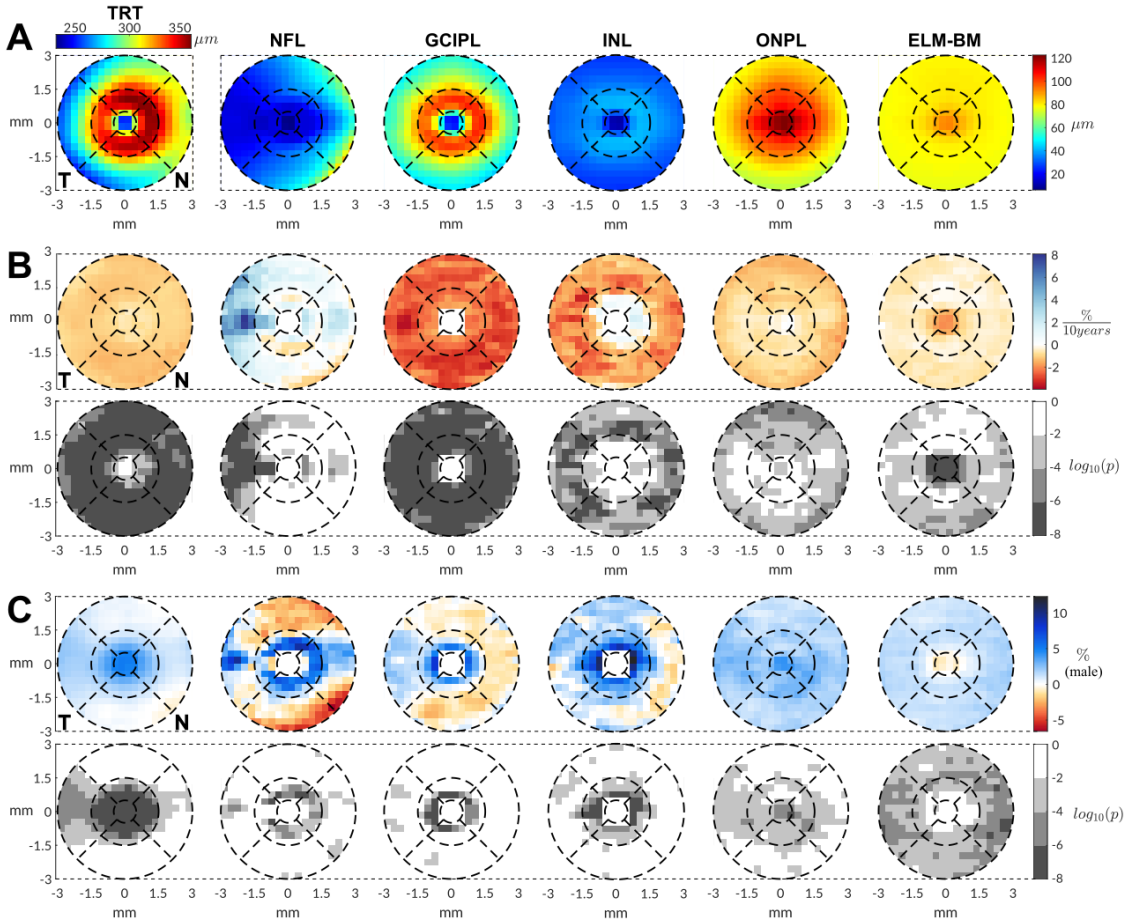


Figure 5.3: Thickness analysis results for the 20 x 20 regular grid sectorization. A) Population mean thicknesses. **B)** Age-related changes of retinal layer thicknesses measured as % of change per 10 years (top) and corresponding p-values (bottom). **C)** Sex differences in percentual units for males (top) and associated p-values (bottom). p-values are reported in logarithmic scale after FDR correction. Thickness analysis results for the 20 x 20 regular grid sectorization. From [183].

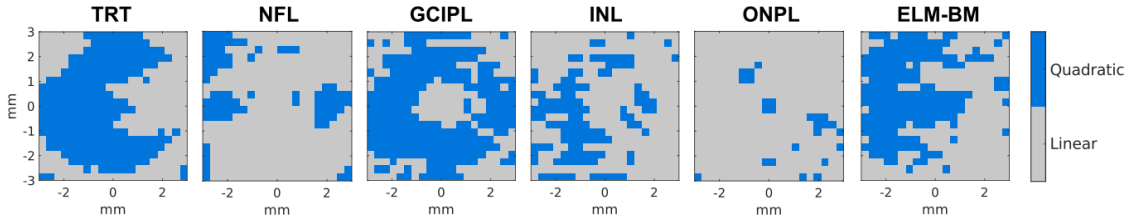


Figure 5.4: Selected age model for each 20 x 20 grid sector. Adapted from [183].

In terms of sex differences, significant variations in the TRT were observed primarily in the central region (radius $< 1.5\,mm$), with males exhibiting up to 4% thicker retina than females. The differences between sexes in the NFL, GCIPL, and INL layers were more pronounced in the inner ring ($0.5\,mm \leq$ radius $\leq 1.5\,mm$), and males presented significantly higher thickness values than females. However, as the radius increased, these differences reduced significantly and even reversed. Differences between sexes in the ONPL and ELM-BM layers were less evident and presented a relatively homogeneous pattern in percentage terms. Nevertheless, considerable differences were observed in the outer ring (perifovea) for the ELM-BM and in the inner ring (parafovea) for the ONPL.

Foveal pit morphology analysis

Both sex and age have an impact on the morphology of the foveal pit, as evidenced by the results in Table 5.3. Overall, foveal parameters were more affected by sex than age. The rim height presented a statistically significant decrease of 0.97 [0.71, 1.22]% per 10 years, indicating an age-related effect. However, considerable inter-subject variability was found in the CFT, rim radius, and mean slope, and thus no obvious age-related trend could be detected.

In all parameters there were clear variations between the sexes. The CFT was larger in males (+7.99 [4.22, 11.77] μm) , the rim height was greater (+8.63 [5.94, 11.33] μm), the rim radius shorter (-59.40 [-78.34, -40.46] μm), and the mean slope steeper(+0.39 [0.18, 0.60] °). The percentage differences were most pronounced for the mean slope (6.23 [2.84, 9.63]%).

Radial examination of the foveal pit morphology also revealed age and sex differences as depicted in Figure 5.5. Specifically, the rim height was found to reduce in all directions, ranging from 0.6% to 0.8% per 10 year periods. The rim radius also decreases in all directions (1% decrease per 10 years), with a more pronounced effect in the inferior and nasal sectors. In contrast, sex-related differences were evident in every angular direction and evenly distributed across sectors. Somewhat larger variations were observed in the superior and inferior sectors of the rim radius, however.

5.4 Discussion

In this work, we examined the effect of age and sex on the structure of the retina. The study was conducted using finely sectorized thickness maps of macular layers and foveal pit morphology metrics obtained from OCT images of 444 healthy subjects (855 eyes). Our findings revealed a consistent thinning of the TRT of 1.1% per 10 years, the primary cause of which is a decrease in GCIPL thickness (-2.4%), followed by INL thinning (-1.3%), and ONPL thinning (-0.7%). The foveal rim height also exhibited a significant decline of 1.0% every 10 years.

Furthermore, we observed that, male retinas were thicker than female retinas ($4.1\,\mu m$ on average), with more pronounced differences in the central region (radius $< 1.5\,mm$). A larger CFT, higher rim height, shorter rim radius, and steeper mean slope were also found to be characteristic of male retinas.

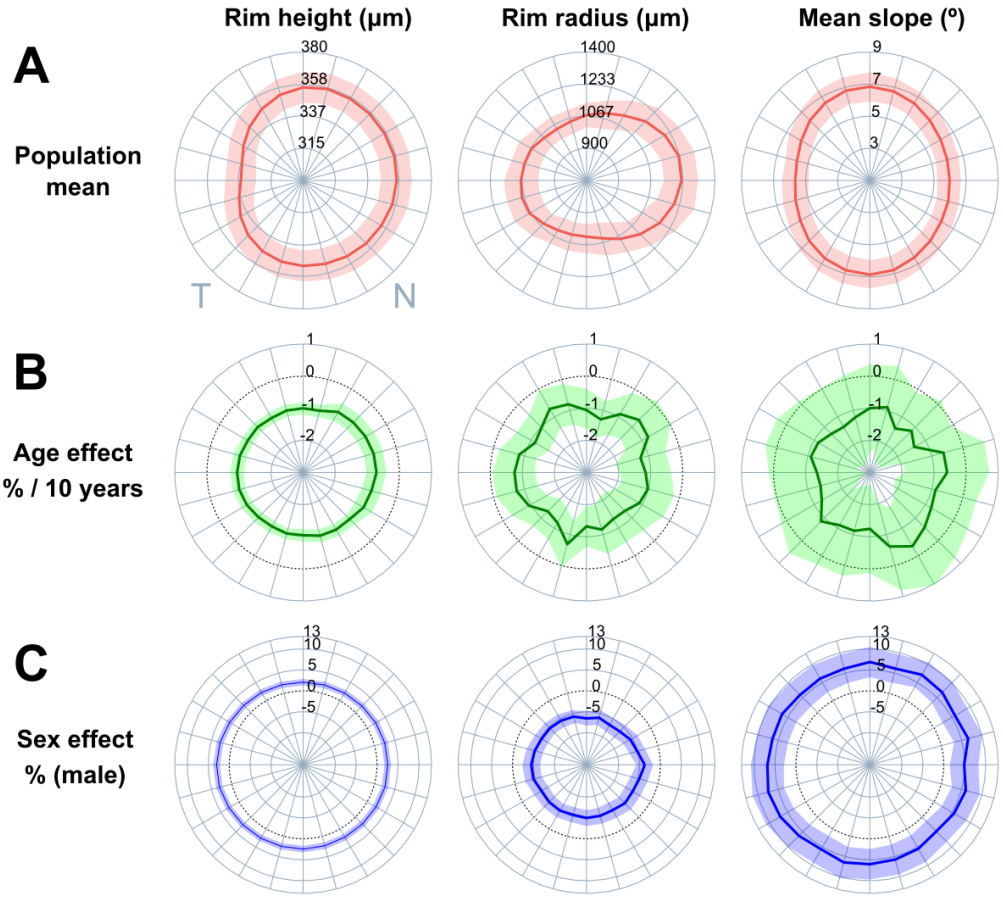


Figure 5.5: Radial analysis of rim height, rim radius and mean slope. A) Population mean (central colored line indicates the mean, shaded region depicts the 2.5 and 97.5 percentiles). Percentual effect of age **B)** and sex **C)** are shown as the normalized regression coefficients (β_{age} and β_{sex}) for each of the 24 angular directions. The shaded region illustrates the 95% CI while the dashed black circle locates the origin (coefficients equal to zero). From [183].

The age-related decline in TRT that we observed indicates that the retina undergoes structural changes over time. These results corroborate much of the published data [96–101]. Notably, a previous literature review described the age-related thinning pattern as spatially dependent, with a preserved or increased thickness in the central retina and the maximum thinning occurring in the parafoveal region [96]. In terms of the loss of TRT in percentage terms, we found a relatively consistent thinning effect for eccentricities larger than $0.5\,mm$. This would seem to indicate that variations in the rate of thinning, except in the central macula, can be attributed to differences in the initial thickness of the TRT.

More critical than analysis of the TRT, however, is identifying the specific layers that contribute to the thinning phenomenon. The published data lacks agreement, and various studies have reported thinning [104, 106], no significant effect [99, 100], or even thickening [98, 107]. A decline in the number of fibers comprising the optic nerve and the NFL as a consequence of aging has also been found in some histological investigations [102, 103]. Nonetheless, in the present study we observed that the NFL either remained unaltered or exhibited thickening in certain temporal regions. Such inconsistent outcomes could be attributed to the relatively thin nature of the NFL in relation to the axial resolution, which increases its susceptibility to segmentation errors. Thus, it can be concluded that peripapillary OCT might serve as a more appropriate tool for evaluating the NFL than macular OCT.

Our findings in relation to the GCIPL are consistent with histological evidence suggesting a decline in ganglion cell density with age [103]. Both the GCL [104, 106, 108] and the IPL have

presented thinning in several OCT studies [98, 101, 105, 116]. In line with this, the GCIPL was found to be the layer most affected by aging in our investigation, which would seem to suggest its particular sensitivity to age-related changes. We also observed a consistent percentage loss in GCIPL thickness across the macula, except in the central region, which is in line with previous research [113, 114]. Hence, regional differences in absolute thickness loss, such as the more pronounced thinning in the parafoveal area [116], could be a result of disparities in the initial thickness of the GCL rather than a spatially-dependent susceptibility to age-related decline. As for the INL, previous studies have also reported thinning [98, 99, 101, 116], and our results support these findings. We observed that the INL exhibited the second most significant age-related decline, primarily characterized by thinning in the outer regions.

In the case of the ONPL and the ELM-BM, a thinning trend was observed, which is in agreement with recent studies [101, 107, 116]. Two notable thinning patterns were detected: a prevalent and uniform thinning of the ONPL in the outer regions and a highly localized thinning of the central region of the ELM-BM. These distinct patterns may be attributed to differences in the cellular configuration of these layers.

The Spectralis device has been reported to exhibit excellent repeatability [196], and it is worth noting that the observed yearly changes are generally small in comparison to the coefficient of variation. As an example, the coefficient of variation (COV) for Spectralis TRT measurements using eye-tracking mode (0.86%) [197] is equivalent to the expected age change over a decade. From this we can conclude that: 1) natural age-related changes are likely to have a limited impact on longitudinal studies with regular follow-ups, particularly those conducted within a span of fewer than five years, and 2) despite the high repeatability of OCT measurements, detecting small changes in the retina requires large sample sizes and groupwise statistical analysis.

Previous research has reported a thicker retina in males compared to females, both in adults [97, 99, 117, 118] and children [198]. These studies consistently found that the differences in thickness were more pronounced in the inner macular ring and decreased in the outer ring, which corresponds to the pattern we observed. Moreover, we measured spatially localized differences in all retinal layers, indicating that sex influences the entire retina. Layer-specific differences have also been reported in the literature, some studies have found a thicker NFL in males [99, 101, 105, 119], while others have observed it to be thicker in females [100, 104, 106]. Similarly, males present greater thickness in males in the GCL [99, 101, 112, 119], IPL [99, 101, 105], and INL [100]. Sex differences have also been found in the outer retinal layers [100, 105].

This the thicker retina observed in males could be explained by systematic macroscopic difference. In fact, MRI studies have consistently shown that males have larger ocular globes [199] and it could thus be hypothesized that larger eyes are associated with a thicker retina. Conversely, other studies have found a negative relationship between axial length and retinal thickness has been reported [121], although it is possible this is due to the influence of ocular biometry on lateral image scaling [13, 14]. In other words, in larger eyes, the same field of view covers a larger region, which could impact the relationship between axial length and retinal thickness.

In a similar vein, the fact that sex differences in inner layer thicknesses are localized in the inner ring could be a result of a sex bias in the estimation of lateral image scale. While adjustments for differences in axial length are typically made in lateral scaling [101, 133], imaging devices often adopt a nominal corneal curvature value for both males and females (e.g., Spectralis uses a $7.7\,mm$ [122]). Considering that corneal curvature is positively related to lateral scaling and tends to be smaller in females [127], this assumption could result in an overestimation of lateral scaling in females. Consequently, this discrepancy could shift the peaks of TRT, GCIPL, and INL thickness profiles in females to larger eccentricities, contributing to the observed pattern.

Turning to the influence of age on the foveal pit, no clear effect on the CFT has been reported in the literature [128, 131] nor was evident in our own results (refer to Table 5.4). This could be attributed to the absence of inner retinal layers in the central region, which are more susceptible to thinning with age [96]. As for rim height, we observed a thinning of the rim, consistent with

the known age-related TRT decline. Smaller previous studies, however, did not find a statistically significant effect [128,131]. A decrease in the rim radius was also detected in our results, which may be a result of the reduction in rim heigh and subsequent rim flattening. However, this finding did not retain statistical significance after multiplicity correction. The mean slope exhibited high inter-subject variability, leading to considerable uncertainty in the estimates. This likely contributes to the lack of consistency in the published data [129, 130, 132].

To date various studies have investigated sex differences in the foveal pit. Although these have employed different parameter definitions and mathematical models, one finding common to all and the present study is a broader and shallower pit in females [125, 127, 132–134] (Table 5.5). Similar to the observed thickness differences, the bias introduced by ocular magnification in lateral scale estimation could contribute to an overestimation of lateral scaling in females and thereby explain the differences observed in slope and radius. Our study also validates previous findings of greater rim height in males [125]. It would seem that these differences are a result of a higher overall TRT in males, given that foveal pit metrics are effectively thickness measurements.

Beyond its global morphology, the foveal pit is a radially asymmetric structure, with a broader horizontal plane than the vertical directions [125]. Our analysis of the foveal pit in 24 individual angular directions revealed relatively uniform percentage age and sex effects across all directions. This suggests that the fovea undergoes homogeneous changes and evolves consistently, despite its structural asymmetry.

The results of this study are subject to certain limitations. First, the number of used B-scans is relatively low (25). However, as showed in Section 4.6, this choice has minimal impact on thickness measurements. Secondly, we employed raster scans and interpolation to analyze the fovea radially instead of using a radial acquisition pattern. While this approach may limit the accurate reconstruction of the TRT profile in the vertical direction, we chose it to avoid potential biases introduced by irregular sampling density in radial patterns, especially when measuring thicknesses far from the central region or correcting fixation errors.

Display distortion was not corrected due to the unavailability of biometric and scanner optical information, which are required for accurate correction [14]. Although such distortion has minimal impact on thickness measurements and small fields of view, it can affect slope metrics [76]. For this reason, we computed the slope after flattening the retina using the TRT, even though it does not directly measure the slope observed in OCT images. This approach is commonly employed and assists in mitigating the effects of retinal curvature and display distortion.

Axial length also has an impact on retinal measurements. To account for this, all regression models were adjusted using the scan focus parameter, which is exported by the scanner and incorporates the refractive error of each eye during image focusing [122]. Considering the strong correlation between refractive error and axial length ($R^2 > 0.72$) [200], we assumed the scan focus parameter to be a reasonable proxy for axial length. Additionally, we relied on the lateral image scale estimation performed by the Spectralis scanner to address the issue of ocular magnification. However, it is important to note that this procedure may have limitations when default corneal curvature values are used [122]. On another note, we considered the horizontal direction as the temporal-nasal axis instead of using the fovea-optic disc axis. This was because the limited field of view of the images prevented from recovering the optic disc position precisely. Lastly, we did not include an interaction term between sex and age in our analysis due to the high inter-subject variability. This could reduce the statistical power to detect potentially very small effects.

5.5 Conclusions

Thinning of most retinal layers occurs over time, particularly in the GCIPL. The percentage changes in both the TRT and GCIPL are consistent, including in small sector analyses. Males generally tend to have thicker retinal layers compared to females, with more pronounced differences

Study	Cohort	N	Age	CFT (μm / 10y)	Rim height (μm / 10y)	Pit depth (μm / 10y)	Rim radius (μm / 10y)	Slope† (° / 10y)
Present study		444	21-88	-1.44	-3.42*	-	-7.66*	-0.06
Tick et al. [128]		57	18-45	NAF	NAF	NAF	NAF	-
Nesmith et al. [129]		390	13-97	-	-	-	-	Increase*
Gella et al. [130]		668	≥ 40	-	-	-	-	Decrease*
Sepulveda et al. [131]	Young	20	24-33	NAF	NAF	-	Decrease	-
	Old	10	62-76					
Zouache et al. [132]	Ghanaian	84	45-82	-	-	-4.8	-46*	-0.054*
	Caucasian	37	41-85	-	-	-1.5	+30	-0.012
Olvera-Barrios et al. [134]		63939	40-69	-	-	-	-	Increase in females*

* $p < 0.05$

† In the present study the mean slope was studied instead of the maximum slope. Olvera-Barrios et al. measured foveal curvature instead of slope.

Table 5.4: Comparison of studies analyzing the effect of age on the foveal pit. CFT: central foveal thickness. NAF: no association found, estimations not reported. Adapted from [183].

Study	Cohort	N_{male}	N_{female}	Age	CFT (μm)	Rim height (μm)	Pit depth (μm)	Rim radius (μm)	Slope† (°)
Present study		163	281	21-88	+7.99	+8.63*	-	-59.4*	+0.39*
Wagner-Schuman et al. [133]		47	43	27.8 ± 9.0	-	-	+1	-15	+0.4
Dubis et al. [127]		26	16	18-67	-	-	+13	+79	+0.5
Scheibe et al. [125]		109	111	21-77	+4.2	+6.0*	-	-27.5	+ 0.5
Zouache et al. [132]	Ghanaian	30	54	45-82	-	-	+7	-145	+2.26
	Caucasian	9	28	41-85	-	-	-10	-180	+1.05
Olvera-Barrios et al. [134]		28842	35097	40-69	+6.8*	-	-	-	+0.87*

* $p < 0.05$

† In the present study the mean slope was studied instead of the maximum slope. Olvera-Barrios et al. measured foveal curvature instead of slope.

Table 5.5: Comparison of studies analyzing the effect of sex on the foveal pit. CFT: central foveal thickness. Adapted from [183].

in the inner ring. The primary age-related effect on the foveal pit is a reduction in rim height. Notably, significant differences between males and females are observed in the fovea, with females presenting a shallower and broader pit. Both sex and age effects are evident across all angular directions. A more comprehensive description of the macula can be obtained with advanced analysis techniques, such as detailed sector-based thickness assessment and radial geometrical analysis of the foveal pit.

6 | Patient classification and clinical assessment

Previous research has demonstrated the existence of retinal changes in PD using both OCT and histology [41, 57]. However, whether the retinal changes measured by OCT are sufficiently pronounced to accurately assess a patient clinically is still an open question. In this regard, there are two main applications of OCT for evaluating the current state of a patient:

- **Diagnosis**: determining if an individual has PD.
- **Severity assessment**: determining the impairment of a patient in a certain domain (cognitive, motor, visual, etc.).

Previous research addressing both topics has largely been focused on conventional thickness features, yielding promising yet diverging results. To extend previous work, in this chapter we investigate the potential of a larger set of OCT features to tackle these two tasks. First, Section 6.1 outlines the employed dataset. Then, in Section 6.2, we explain how we used the computed features to train classifiers for PD diagnosis. Finally, Section 6.3 describes the development of regression models for clinical variables related to cognitive and motor impairment.

6.1 Subjects and OCT features

A total of 174 PD patients and 174 healthy controls from the BHRI dataset were included (Table 6.1). The controls were chosen to match patients based on age and sex. The matching process was conducted iteratively, where each patient was paired with a control of the same sex and the closest age. To prevent significant age disparities arising from a lack of male controls, in cases where no same-sex control within a 5-year age difference was available, a control of the opposite sex was selected if one with less than a 5-year age difference was present.

Group	N	Age (years)	Male %	Disease duration (years)	MoCA	UPDRS III	Hoehn Yahr
PD	174	65.0 (8.5)	61.5	6.0 (4.7)	23.3 (4.5)	23.0 (11.4)	2.0 (0.6)
Control	174*	64.9 (8.5)	59.8	-	25.7 (3.5)	-	-

* From the 174 matched controls only 93 underwent MoCA assessment.

Table 6.1: Matched dataset used in diagnostic model development. Summary statistics are given in format mean (σ).

A comprehensive set of 4909 OCT features were investigated. The features were selected based on the existing literature with the aim of covering a broad spectrum of structural properties of

the macula. Figure 6.1 shows an illustration of the retinal layers analyzed as well as the employed features.

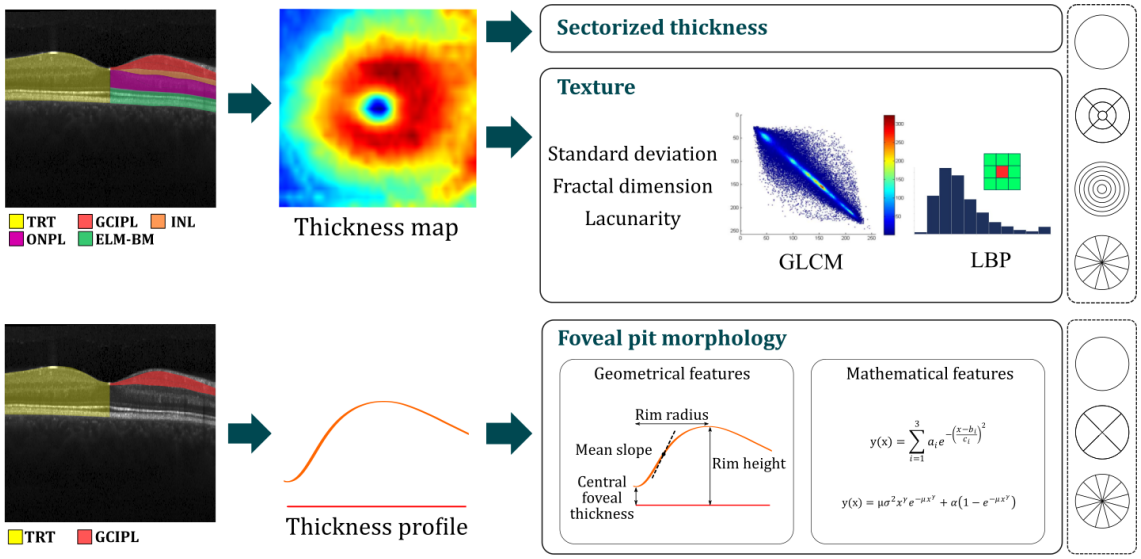


Figure 6.1: Overview of the explored features. Starting with the analyzed retinal layers in the left, the figure shows how the different features are computed. Sectorized thickness and texture features were computed from thickness maps. These were extracted for the four sectorization grids shown in the right. Additionally, multiple foveal features were derived from foveal pit morphology. These included both geometrical and mathematical features and were derived using the three sectorizations showed in the bottom-right. GLCM: gray-level co-occurrence matrix. LBP: local binary patterns.

The included features were derived from three families:

- **Sectorized thickness**: thickness features have been widely studied before with positive results in PD [41]. Here the following layers were studied: TRT, GCIPL, INL, ONPL, and ELM-BM. Average thickness measurements were computed for the aforementioned layers using 4 different sectorization grids: Macula, ETDRS, 5 rings, and 12 angular directions (all depicted in the top-right of Figure 6.1). We included multiple sectorizations as there is no conclusive evidence about the precise localization of changes in PD. Considering all the layers (5) and sectors (27), this family included a total of 135 features.

- **Texture**: these features included the fractal dimension, lacunarity, standard deviation, LBP features, and GLCM features. All the texture features were computed for the same layers and sectors investigated with sectorized thickness. The number of final computed texture features were 4320, resulting from combining 5 layers, 27 sectors, and 32 texture feature types.

  The inclusion of these features is justified by the literature. Fractal dimension has recently shown diagnostic potential in AD [160]. Similarly, lacunarity has proven to be useful in PD-related retinal image analysis [161]. In addition, recent studies have also investigated the use of LBP and GLCM features with positive results in MS, AD, and PD [157, 159].

- **Foveal pit morphology**: these features can provide complementary information about the retinal structure, with some evidence of a foveal remodeling in PD [74, 154]. Computed geometrical features included: mean slope, maximum slope, CFT, rim height, rim radius, pit depth, maximum slope height, max slope radius, pit area, rim disk perimeter, rim disk area, and maximum slope disk area [72, 77]. Additionally, features derived from four mathematical models were also included [73, 74, 76, 77]. Both geometrical and mathematical features were computed for two retinal layers: TRT and GCIPL. All radial features were averaged across the entire macula, 4 ETDRS quadrants, and 12 angles. In addition, the standard deviation

of each radial feature across all 12 directions was also computed. In total, 454 foveal pit features were computed.

The complete OCT processing pipeline along with the concrete computation procedure of each feature are described in detail in Section 4.5. In this analysis, features from left and right eyes were averaged to obtain a single value per subject. When only one of the eyes was available the features from that eye were considered as the subject average.

6.2 Diagnosis

In this section, we investigated the diagnostic potential of the aforementioned features. We began by exploring each feature individually and subsequently proceeded to train multivariate classification models.

6.2.1 Univariate feature exploration

As an initial step, we evaluated the individual ability of each feature to discriminate between PD and HC subjects. To do this, three metrics were computed for each feature: 1) the effect size measured as the standardized mean difference (i.e., Cohens' d), 2) a p-value based on a two-sample t-test, and 3) the AUC computed by using feature values as scores in a receiver operating characteristic (ROC) curve analysis.

The analysis revealed that, overall, features showed a limited ability to distinguish between the two groups. The greatest effect size was $|d| = 0.4$, which is commonly considered a small-medium effect. Accordingly, the highest AUC was 0.618, highlighting the difficulty of diagnosing PD from individual OCT features.

To better understand which features performed best, Table 6.2 displays the best 5 features for each sectorization grid and feature family. As shown, features derived from ETDRS sectors reached better performance than those from other sectorizations. The best performance was achieved by a foveal feature of the GCIPL, namely, the foveal radius at the point where the slope is maximized within the temporal ETDRS quadrant. A similar performance was obtained by some texture features (AUC = 0.613) measured on outer-inferior and outer-nasal sectors. These features included the GLCM *Difference Variance* of the TRT, the *standard deviation* of the ELM-BM thickness, and the LBP *Entropy* feature derived from the INL.

Notably, conventional sectorized thickness features did not show great diagnostic performance. The best results in this family were obtained by the GCIPL in the central ETDRS sector, which was 2.06 μm thicker in PD patients and yielded an AUC of 0.586 ($d = 0.25$, $p = 0.02$). This was the only thickness feature that showed statistically significant differences between the two groups ($p = 0.02$). Indeed, all the other thickness features showed a much smaller effect size ($|d| \leq 0.10$) that was not sufficient to discriminate between groups. A detailed description of the group differences in all ETDRS thicknesses can be found in Table 6.3.

Grid	Thickness			Texture				Foveal pit			
	Layer	Sector	AUC	Layer	Sector	Name	AUC	Layer	Sector	Name	AUC
Macula	ONPL	-	0.538	ELM-BM	-	LBP entropy	0.586	GCIPL	-	Ding ($s1$)	0.583
	TRT	-	0.518	INL	-	LBP mean	0.584	TRT	-	Scheibe (α^*)	0.582
	GCIPL	-	0.511	INL	-	LBP σ	0.583	TRT	-	Ding ($s2$)	0.581
	ELM-BM	-	0.507	INL	-	LBP range	0.583	GCIPL	-	Max Slope Radius	0.580
	INL	-	0.506	INL	-	LBP kurtosis	0.581	GCIPL	-	Max Slope Disk Area	0.580
ETDRS	GCIPL	C	0.586	TRT	OI	GLCM Dif. Variance	0.613	GCIPL	T	Max Slope Radius	0.618
	ONPL	OI	0.559	ELM-BM	OI	Standard deviation	0.613	GCIPL	T	Scheibe (μ)	0.585
	INL	IS	0.553	INL	ON	LBP Entropy	0.611	TRT	N	Max Slope Radius	0.574
	TRT	C	0.543	TRT	OI	GLCM Dif. Entropy	0.610	TRT	S	Scheibe (μ)	0.571
	ONPL	ON	0.543	TRT	IN	LBP iqr	0.608	TRT	I	Scheibe (μ)	0.568
5 rings	ONPL	$\varnothing$ 4-5 mm	0.550	INL	$\varnothing$ 4-5 mm	GLCM IMC2	0.606	-	-	-	-
	ONPL	$\varnothing$ 5-6 mm	0.543	INL	$\varnothing$ 2-3 mm	Lacunarity	0.605	-	-	-	-
	ONPL	$\varnothing$ 3-4 mm	0.542	GCIPL	$\varnothing$ 1-2 mm	LBP median	0.602	-	-	-	-
	INL	$\varnothing$ 5-6 mm	0.540	INL	$\varnothing$ 4-5 mm	GLCM correlation	0.599	-	-	-	-
	GCIPL	$\varnothing$ 1-2 mm	0.536	INL	$\varnothing$ 3-4 mm	LBP mean	0.598	-	-	-	-
12 angles	ONPL	I3	0.550	INL	S1	Standard deviation	0.599	GCIPL	T3	Max Slope Radius	0.601
	ONPL	N1	0.546	ONPL	N1	LBP mean	0.597	GCIPL	T2	Max Slope Radius	0.600
	ONPL	I2	0.542	ONPL	N1	LBP Std. Dev.	0.597	GCIPL	T3	Scheibe (μ)	0.587
	TRT	N3	0.540	GCIPL	T3	LBP median	0.593	GCIPL	T1	Scheibe (μ)	0.584
	ONPL	T1	0.539	TRT	S1	GLCM IMC1	0.591	GCIPL	T1	Max Slope Radius	0.583

* Standard deviation of the feature.

Table 6.2: Best OCT features for diagnosis. Sectors: central (C), nasal (N), superior (S), temporal (T), inferior (I), inner-nasal (IN), inner-superior (IS), inner-temporal (IT), inner-inferior (II), outer-nasal (ON), outer-superior (OS), outer-temporal (OT), outer-inferior (OI). Angular sectors are described in Figure 4.17. AUC: area under the curve. GLCM: gray-level co-occurrence matrix. LBP: local binary patterns.

Layer	Sector	Thickness (μm)		Δ (μm)	Cohen's d	p-value
		Control	Parkinson			
TRT	C	278.0 (21.0)	280.5 (20.4)	2.55	0.12	0.25
	IN	345.0 (15.3)	344.2 (15.7)	-0.77	-0.05	0.64
	IS	342.6 (14.8)	341.4 (15.2)	-1.19	-0.08	0.46
	IT	329.8 (14.8)	329.9 (14.3)	0.10	0.01	0.95
	II	338.6 (14.7)	338.3 (14.7)	-0.30	-0.02	0.85
	ON	314.2 (15.0)	312.3 (14.6)	-1.92	-0.13	0.23
	OS	297.9 (13.6)	296.4 (12.7)	-1.44	-0.11	0.31
	OT	281.4 (12.8)	281.5 (12.6)	0.13	0.01	0.92
	OI	287.4 (13.6)	286.9 (12.6)	-0.58	-0.04	0.68
GCIPL	C	36.8 (8.2)	38.8 (8.2)	2.06	0.25	0.02*
	IN	93.3 (7.5)	93.5 (9.2)	0.24	0.03	0.79
	IS	93.2 (7.2)	93.3 (8.3)	0.05	0.01	0.96
	IT	89.6 (7.6)	90.1 (8.1)	0.45	0.06	0.60
	II	91.8 (7.1)	91.7 (8.3)	-0.05	-0.01	0.95
	ON	66.8 (6.4)	66.4 (6.7)	-0.38	-0.06	0.59
	OS	63.1 (5.9)	63.3 (5.3)	0.20	0.04	0.74
	OT	66.4 (6.3)	67.0 (6.2)	0.60	0.10	0.37
	OI	60.6 (5.7)	60.7 (5.8)	0.08	0.01	0.90
INL	C	20.9 (5.7)	21.0 (5.6)	0.08	0.01	0.89
	IN	41.0 (4.1)	40.5 (4.2)	-0.54	-0.13	0.22
	IS	40.7 (3.9)	40.0 (3.7)	-0.70	-0.18	0.09
	IT	37.8 (3.7)	37.8 (3.4)	-0.09	-0.03	0.81
	II	40.5 (4.0)	40.3 (3.9)	-0.18	-0.04	0.68
	ON	34.1 (2.5)	34.4 (2.5)	0.25	0.10	0.35
	OS	30.8 (2.4)	30.9 (2.4)	0.08	0.03	0.75
	OT	32.8 (2.3)	33.1 (2.3)	0.34	0.14	0.18
	OI	30.7 (2.3)	30.7 (2.4)	0.02	0.01	0.94
ONPL	C	119.6 (9.3)	118.9 (9.1)	-0.67	-0.07	0.50
	IN	106.9 (7.5)	106.2 (7.8)	-0.69	-0.09	0.40
	IS	102.2 (7.1)	101.9 (7.5)	-0.31	-0.04	0.69
	IT	103.3 (7.1)	102.9 (7.5)	-0.34	-0.05	0.66
	II	100.8 (7.1)	100.4 (7.4)	-0.40	-0.06	0.60
	ON	86.4 (6.7)	85.4 (7.0)	-0.97	-0.14	0.19
	OS	86.4 (5.9)	85.4 (6.3)	-0.97	-0.16	0.14
	OT	84.4 (6.0)	83.7 (6.4)	-0.62	-0.10	0.35
	OI	79.4 (5.8)	78.4 (5.9)	-1.06	-0.18	0.09
ELM-BM	C	88.6 (3.7)	89.1 (3.8)	0.46	0.12	0.25
	IN	82.9 (3.1)	83.1 (3.2)	0.21	0.07	0.53
	IS	81.7 (3.0)	81.8 (2.9)	0.09	0.03	0.77
	IT	81.6 (2.9)	81.8 (2.8)	0.21	0.07	0.49
	II	80.6 (2.9)	80.9 (2.8)	0.34	0.12	0.27
	ON	78.8 (2.9)	78.8 (2.7)	-0.01	0.00	0.97
	OS	79.8 (2.9)	79.6 (2.8)	-0.18	-0.06	0.55
	OT	78.6 (2.7)	78.5 (2.5)	-0.07	-0.03	0.81
	OI	77.4 (2.8)	77.5 (2.8)	0.06	0.02	0.85

*$p \leq 0.05$

Table 6.3: Group differences in ETDRS thicknesses. Thickness estimates are given as *mean (σ)*. ETDRS sectors: central (C), inner-nasal (IN), inner-superior (IS), inner-temporal (IT), inner-inferior (II), outer-nasal (ON), outer-superior (OS), outer-temporal (OT), outer-inferior (OI).

6.2.2 Multivariate model development

After the initial univariate analysis, two machine learning models were trained using combinations of features: LR and SVM. The selection of these two models is justified by their distinct strengths. LR offers interpretability and efficiency, and was used to determine the performance of a purely linear classification model. On the other hand, SVMs are better suited to model non-linear patterns, maximizing margins for better generalization, and maintaining robustness against outliers. In total, 12 different models were trained and evaluated from all the combinations of the four sectorization grids (Macula, ETDRS, 5 rings, and 12 angles), and the three feature families (thickness, texture, and foveal pit morphology).

The training and evaluation pipeline was the same for each model and is shown in Figure 6.2. First, the initial set of features were normalized to have zero mean and unit standard deviation. To avoid a potential bias due to outliers, both the mean and standard deviation of each feature were computed after trimming the lowest 5% and highest 95% percentiles of the distribution. Next, to reduce the dimensionality, the best 50 features were selected based on the minimum-redundancy-maximum-relevance algorithm [201, 202], a method designed to select a subset of non-redundant features that maximize class separability. This was only applied to the configurations with more than 50 initial features. For instance, when evaluating sectorized thickness over the macular grid there were only 5 macular features and, therefore, this method was not necessary. The selected features were then fed into a forward stepwise feature selection procedure to determine the best feature combination. Starting with an empty model, the features yielding the highest diagnostic accuracy were iteratively added. To do this, all the potential combinations of already selected and new features were evaluated in each iteration based on a leave-one-out cross-validation. That is, models were trained on N-1 samples and used to predict the remaining test sample. Then, the feature showing the highest test accuracy was selected and added to the model. The maximum number of features in the model was set to 10. The leave-one-out cross-validation strategy was selected in view of the small sample size, as it reduces the bias due to not using the full dataset for training [203]. As an important remark, SVM model training included hyperparameter fine-tuning based on an internal 10-fold cross-validation. All the models were evaluated in terms of accuracy, sensitivity, specificity, and AUC.

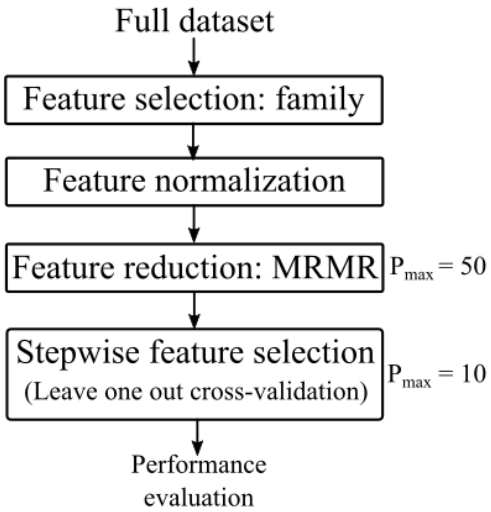


Figure 6.2: Multivariate diagnosis model training pipeline.

In addition to the 12 models described, four additional models were subsequently trained using features from all three families combined. More concretely, the features finally selected by applying the aforementioned pipeline to each family were used for this analysis. The goal of this experiment was to determine if combining the best features in each family could improve performance.

The obtained results are displayed in Table 6.4. With the exception of the models trained only on macular features, the test accuracy ranged from 60% to 67%, with a maximum AUC slightly higher than the one obtained by individual features (0.68 vs 0.62). The models using ETDRS features showed better performance than those trained on macular features only (i.e., features averaged over the entire macula). The results obtained by features computed over 5 rings and 12 angles grids were on par with those from ETDRS features.

Grid	Metric	Thickness		Texture		Foveal pit		All	
		LR	SVM	LR	SVM	LR	SVM	LR	SVM
Macula	**Accuracy**	0.57	0.60	0.56	0.61	0.58	0.59	0.57	0.61
	Sensitivity	0.55	0.61	0.54	0.64	0.61	0.57	0.59	0.65
	Specificity	0.59	0.58	0.59	0.57	0.55	0.61	0.58	0.56
	AUC	0.58	0.60	0.53	0.62	0.55	0.58	0.59	0.62
ETDRS	**Accuracy**	0.61	0.62	0.66	0.67	0.61	0.64	0.66	0.67
	Sensitivity	0.57	0.74	0.65	0.67	0.62	0.72	0.64	0.72
	Specificity	0.65	0.49	0.68	0.67	0.60	0.55	0.69	0.62
	AUC	0.59	0.61	0.66	0.66	0.60	0.64	0.67	0.69
5 rings	**Accuracy**	0.62	0.65	0.64	0.66	-	-	0.66	0.66
	Sensitivity	0.61	0.66	0.63	0.66	-	-	0.60	0.61
	Specificity	0.64	0.65	0.65	0.66	-	-	0.69	0.68
	AUC	0.61	0.66	0.62	0.68	-	-	0.65	0.66
12 angles	**Accuracy**	0.66	0.63	0.65	0.64	0.60	0.60	0.65	0.64
	Sensitivity	0.70	0.58	0.66	0.57	0.57	0.61	0.59	0.62
	Specificity	0.64	0.66	0.64	0.71	0.64	0.59	0.70	0.65
	AUC	0.68	0.62	0.66	0.65	0.61	0.60	0.66	0.66

Table 6.4: Diagnostic model performance. The performance is reported for each grid and feature family. The *All* column refers to combining the best features of each three feature families. AUC: area under the curve. LR: logistic regression. SVM: support vector machine.

Overall, texture features performed best in both Macula and ETDRS grids. The performance was slightly better for SVM models. Although combining the best features of each family did not result in a large improvement, the best performance was obtained by an SVM model using a total of 9 ETDRS features from all families, which achieved a 69% accuracy and an AUC of 0.69. The sensitivity and specificity of this model were 0.72 and 0.62, respectively. The included 9 features and are listed in Table 6.5 and included three thickness features measured in outer ring sectors, two texture features in the central region, and four foveal features.

Rank	Family	Layer	Sector	Feature
1	Thickness	GCIPL	Outer-superior	Mean
2	Thickness	INL	Outer-inferior	Mean
3	Texture	TRT	Central	GLCM cluster prominence
4	Texture	TRT	Central	GLCM joint variance
5	Fovea (model)	GCIPL	Nasal	Scheibe (α)
6	Thickness	ONPL	Outer-inferior	Mean
7	Fovea (geometry)	GCIPL	Temporal	Max slope radius (mean)
8	Fovea (model)	TRT	Temporal	Scheibe (γ)
9	Fovea (model)	GCIPL	Temporal	Scheibe (μ)

Table 6.5: Features included in the best diagnostic model. TRT: total retinal thickness. GCIPL: ganglion cell–inner plexiform layer. INL: inner nuclear layer. ONPL: outer nuclear and plexiform layer. GLCM: gray-level co-occurrence matrix.

As an additional illustration, Figure 6.3 shows the evolution of the performance of the best

configuration (i.e., *All* family and ETDRS grid) as more features are added to the model. As shown, the performance of both LR and SVM models improved as more features were included. Although the absolute maximum performance was achieved by the latter, the LR model showed a better balance between the performance of both classes (i.e., sensitivity and specificity were more similar).

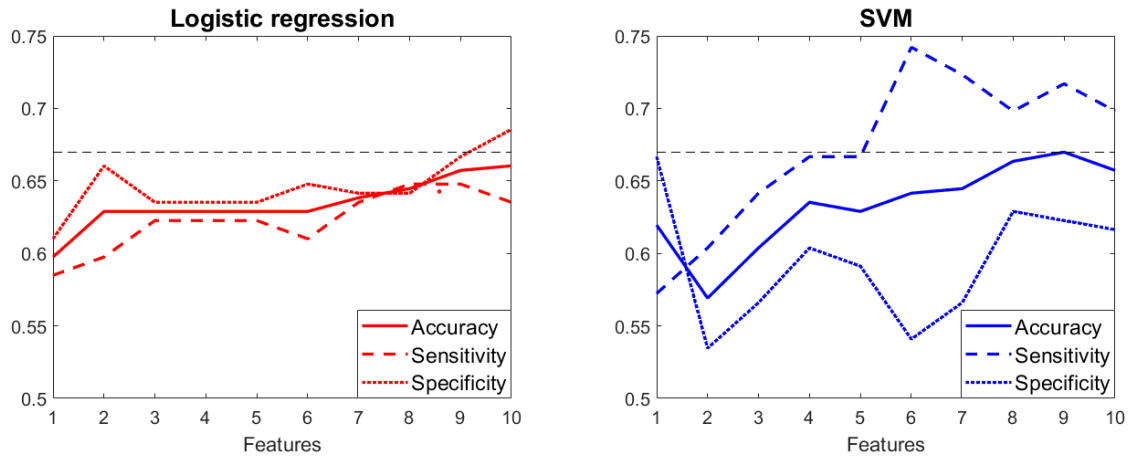


Figure 6.3: Best diagnostic model performance over feature count. The horizontal dashed black line depicts the absolute maximum accuracy obtained by the SVM model. SVM: support vector machine.

6.2.3 Alternative diagnostic strategies

In the previous analysis we made use of features computed as the average of both eyes (e.g., we computed the mean foveal slope for left and right eyes and used the average of both as the final value for each subject). However, the values derived from both eyes can also be used without averaging. By adopting this approach, a model is trained considering both eyes of each subject as different data samples and an independent prediction is made for each eye. This methodology has previously been used under the assumption that changes may occur on a single eye and averaging may hinder their detection. To explore this idea, in this section we investigated whether using features from each eye individually resulted in different performance.

Using both eyes separately, however, raises a challenging methodological question, as it entails generating two independent predictions that may not agree. Indeed, in the special case of PD diagnosis, there are four possible scenarios depending on the predictions of left (OS) and right (OD) eyes (see Table 6.6). In two of these scenarios opposite predictions are produced by each eye and, therefore, there is no obvious prediction to be made on a subject level. This problem can be omitted if model performance is reported only at an eye level. However, this disregards the real goal of a PD diagnostic model, which is to diagnose a patient and not each eye separately. Thus, a seemingly sensible alternative strategy could be not making a prediction when both eyes disagree. This means training a model using both eyes but adding a rule so that when left and right eyes produce opposite predictions, no label is assigned to the subject.

		Prediction OS	
		HC	PD
Prediction OD	HC	Agree	Disagree
	PD	Disagree	Agree

Table 6.6: Possible scenarios with individual eye prediction.

To investigate the described strategy an additional SVM model was trained using both eyes and the best features identified in the previous analysis (see Table 6.5). As before, features were normalized and fed into a forward stepwise model selection procedure. As a key difference, in this case the leave-one-out cross-validation procedure was performed on a subject level to ensure that the eyes of the each subject were used only for either training or testing. Model performance was evaluated in two ways:

- *Eye-level*: computed using all eyes as independent test samples. Although this metric disregards the cases in which the predictions of both eyes did not align, it is computed for comparison purposes.
- *Subject-level*: computed considering only those cases in which the prediction for both eyes of the same subject was the same.

The obtained results are shown in Table 6.7. At an eye-level, the model obtained an accuracy of 59.5%. Interestingly, this accuracy increased up to an 63.1% when predictions were only made for subjects for which both eyes provided the same prediction. It is important to mention however that the latter represented the 83.4% of the subjects, which means that, this approach failed to provide a prediction for 16.6% of the subjects but improved the accuracy in those cases in which both eyes were in agreement.

	Eye-level	**Subject-level**
Accuracy	59.5	63.1
Sensitivity	59.6	64.0
Specificity	59.5	62.4

Table 6.7: Diagnostic performance using both eyes. All the estimates are reported as %.

6.2.4 Patient stratification

After evaluating the overall diagnostic performance, we explored the potential of OCT features to diagnose PD in patients at different disease stages. The goal was to determine if PD diagnosis improved when comparing controls only with patients in a more advanced disease stage. To this aim, we created different subgroups of patients based on disease duration, MoCA, and UPDRS-III. For the disease duration variable, three groups were created based on distribution tertiles (3.20 and 7.04 years). As for MoCA, fixed cut-offs were established at 23 and 26 scores. These values correspond to previously used thresholds for cognitive impairment screening [166], and allowed us to split the sample into three groups of similar size. Regarding UPDRS-III, patients were also split into three subgroups according to distribution tertiles (19 and 29). In all cases, control subjects were correspondingly split into groups according to each patient-control matched pair. This ensured that patient and control subgroups had the same sex and age distribution.

As in the previous sections, the features from thickness, texture, and foveal pit morphology were used to train LR and SVM models. In this case, only features from the ETDRS sectorization grid were used as they showed the best diagnostic performance. In total 4 models were trained for each patient subgroup, one for each feature family (thickness, texture, and foveal pit morphology), and one combining the best features of the three families. Model training and evaluation was carried out following the same pipeline described in section 6.2.2. In this analysis all the features were derived as the average between both eyes.

The performance obtained by the best diagnostic model in each patient subgroup is displayed in Table 6.8. Overall, the performance improved significantly with respect to the previous diagnostic yield. For instance, a model trained on the patient subgroup with the longest disease duration reached an accuracy of 77% and an AUC of 0.79. This performance was achieved by an SVM model combining 9 texture features. The sensitivity and specificity of the model were balanced

(0.75 and 0.79, respectively). As for MoCA subgroups, a LR model trained on texture features demonstrated an accuracy of 75% and an AUC of 0.71 to discern the patients with the highest cognitive impairment. Similarly, UPDRS-III subgroups also improved the performance with an accuracy above 70%.

Clinical variable	Threshold	N	Features	P	Model	Acc.	Sen.	Spe.	AUC
Disease duration (years)	$x < 3.2$	57	All	10	SVM	0.74	0.79	0.72	0.79
	$3.2 \leq x < 7.04$	57	Texture	9	LR	0.73	0.75	0.70	0.70
	$x \geq 7.04$	57	Texture	9	SVM	0.77	0.75	0.79	0.79
MoCA	$x \geq 26$	62	Texture	2	LR	0.69	0.73	0.65	0.65
	$23 \leq x < 26$	55	All	6	LR	0.74	0.73	0.77	0.72
	$x < 23$	55	Texture	9	LR	0.75	0.69	0.80	0.71
UPDRS-III	$x \leq 19$	56	Texture	8	SVM	0.71	0.73	0.69	0.70
	$19 \leq x < 29$	58	All	10	LR	0.73	0.74	0.71	0.73
	$x \geq 29$	58	Texture	10	LR	0.73	0.76	0.67	0.75

Table 6.8: Best diagnostic model performance with patient subgroups. For each patient subgroup the performance of the best model is reported. Additionally, the number of features included in each model (P) as well as their type is reported. *All* feature family refers to combining the best 10 features of each previous category. AUC: area under the curve. LR: logistic regression. SVM: support vector machine.

The features included in the best models applied to the patients with the longest disease duration, lowest MoCA, and highest UPDRS-III are listed in Table 6.9, 6.10, and 6.11, respectively. As shown, texture features dominated these models. As a recurrent finding, most of the features included in these models corresponded to the central ETDRS sector. Regarding the specific layers of the features, most of the features were derived primarily from TRT.

Rank	Family	Layer	Sector	Feature
1		INL	Outer-superior	GLCM Sum of squares
2		TRT	Inner-inferior	GLCM Inverse difference moment
3		ELM-BM	Central	Fractal dimension
4		TRT	Inner-inferior	GLCM Inverse Difference
5	Texture	TRT	Central	GLCM Difference variance
6		TRT	Central	LBP range
7		GCIPL	Central	GLCM Energy
8		GCIPL	Central	GLCM IMC2
9		TRT	Central	LBP standard deviation

Table 6.9: Features included in the longest disease duration diagnostic model. Patients in the group diagnosed by this model had a disease duration greater or equal than 7.04 years. GLCM: gray-level co-occurrence matrix. LBP: local binary patterns.

Rank	Family	Layer	sector	Feature
1		TRT	Central	GLCM Max probability
2		TRT	Central	GLCM IMC2
3		TRT	Central	LBP mean
4		TRT	Central	LBP standard deviation
5	Texture	TRT	Central	GLCM Cluster prominence
6		GCIPL	Central	GLCM Max probability
7		TRT	Central	GLCM Correlation
8		TRT	Central	LBP entropy
9		TRT	Central	GLCM Inverse difference

Table 6.10: Features included in the lowest MoCA diagnostic model. Patients in the group diagnosed by this model had a MoCA lower than 23. GLCM: gray-level co-occurrence matrix. LBP: local binary patterns.

Rank	Family	Layer	sector	Feature
1		ELM-BM	Central	GLCM Dif. Variance
2		TRT	Inner-superior	LBP interquartile range
3		ONPL	Inner-superior	GLCM Cluster shade
4		ONPL	Inner-superior	GLCM Cluster shade
5	Texture	INL	Inner-temporal	LBP mean
6		ELM-BM	Outer-superior	GLCM Inverse difference normalized
7		ELM-BM	Central	LBP mean
8		TRT	Inner-inferior	LBP median
9		GCIPL	Inner-nasal	GLCM Sum average
10		TRT	Outer-temporal	Lacunarity

Table 6.11: Features included in the highest UPDRS-III diagnostic model. Patients in the group diagnosed by this model had a UPDRS-III higher or equal than 29.

6.3 Clinical assessment

In this second part of the chapter we developed and evaluated predictive models for MoCA, UPDRS-III, and HY scores. The final objective was to evaluate how informative OCT features are as a predictor of cognitive and motor impairment of PD patients. To assess this, we built regression models for each variable following three strategies:

- *Reference performance*: using the age, sex, and disease duration of each patient for prediction. The performance of this approach serves to establish a reference to compare against.
- *Individual OCT features*: using only individual OCT features for prediction. This was carried out to identify associations between disease severity and specific features.
- *Multivariate modeling*: combining OCT features to explore how a using more than one feature influences performance.

All the regression models described in the next sections were evaluated using the same leave one out cross-validation strategy. Model training was different depending on the followed approach and is described in each of the subsequent sections. Model evaluation was carried out by means of the R^2 as well as the RMSE obtained on the test set. Additionally, the AIC was also computed for each training set. The obtained AIC values were then averaged to obtain a measurement on the adequacy of each model.

6.3.1 Reference performance

In this first approach, we fitted linear regression models with sex, age, age^2, and disease duration to predict MoCA, UPDRS-III, and HY scores. The age values were centered at 60 years old. In addition, we also evaluated the performance of using the average value of each clinical variable in the train set to predict the test sample. This was done to obtain a baseline reference for the RMSE.

The regression results for each model are shown in Table 6.12. Age was the best predictor for MoCA, explaining up to 22% of its variance and reducing the RMSE by 0.56 points. As evidenced by the AIC and illustrated in Figure 6.4, the relationship between age and MoCA was better modeled by a quadratic function. Indeed, a model with *age* and *age*^{2} was found to be better than a constant model in both control (F-test $p = 10^{-8}$) and PD patients (F-test $p = 10^{-12}$). This model was able to predict MoCA with an RMSE of 4.10, a 12% lower than that of a constant model (RMSE = 4.61). Conversely, neither sex nor disease duration showed an association with MoCA.

On the other hand, it was not possible to accurately predict UPDRS-III scores from any of the evaluated predictors, which were not able to reduce the RMSE of 11.49 obtained by a constant model.

As for HY, although age and sex did not show any predictive performance, a larger disease duration was with a higher HY score (R^2 =0.07). This association resulted in a 3.5% reduction in the prediction error (from an RMSE of 0.629 to 0.607).

Outcome	Model	R^2	RMSE	AIC	p-value (β)
MoCA	constant*	-	4.61	979	-
	sex	-0.02	4.63	980	0.42
	age	0.20	4.10	939	$1.4{\cdot}10^{-10}$
	age + age^2	0.22	4.05	935	$3.3{\cdot}10^{-6}$, 0.013
	disease duration	-0.02	4.61	986	0.48
UPDRS-III	constant*	-	11.50	1282	-
	sex	-0.01	11.49	1282	0.12-
	age	-0.02	11.57	1284	0.83
	age + age^2	-0.03	11.59	1285	0.47, 0.27
	disease duration	-0.03	11.59	1284	0.79
Hoehn-Yahr	constant*	-	0.629	319	-
	sex	-0.03	0.633	321	0.91
	age	-0.01	0.629	319	0.19
	age + age^2	-0.02	0.632	321	0.36, 0.42
	disease duration	0.07	0.607	302	$1.1{\cdot}10^{-4}$

* Predict the test sample as the average value on the training set.

Table 6.12: Regression results for reference approaches. The best model in each clinical variable is highlighted in light blue color. p-values refer to each specific regression coefficient.

6.3.2 Individual OCT features

Here, we sought to explore which OCT features were more strongly related with clinical variables. To investigate this, we fitted simple linear regression models using clinical variables as an outcome and each individual OCT feature as a single predictor. In this regression analysis the test R^2 and RMSE were computed following the leave-one-out cross-validation procedure. Additionally, a Pearson correlation and a p-value were derived for each predictor using the entire dataset. The

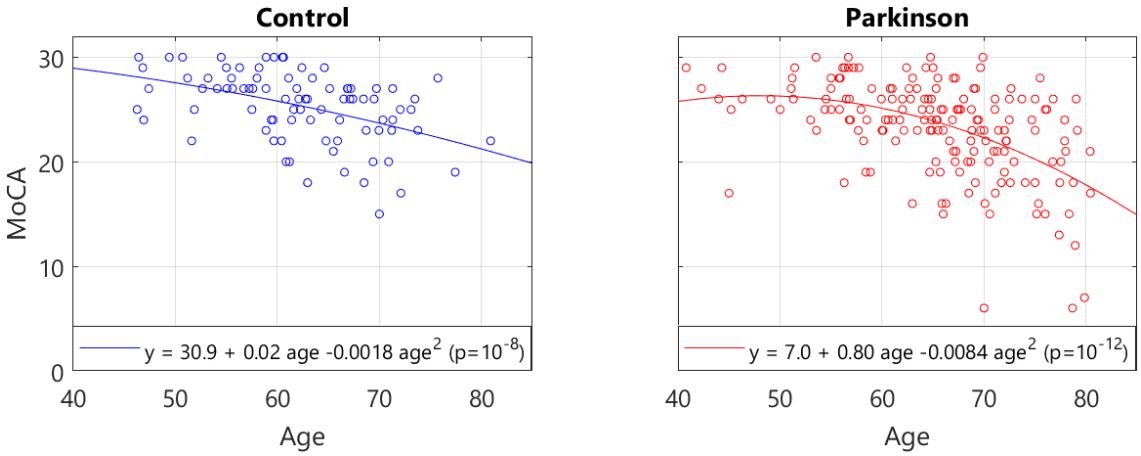


Figure 6.4: MoCA as a function of age for controls and patients. The data and regression results in controls are shown for reference purposes. Except for this figure, all the results presented in Table 6.12 and this part of the chapter were computed for PD patients only.

analysis was carried out using OCT features described previously in Section 6.1.

The 5 features showing the highest test R^2 from each feature family are listed in Tables 6.13 (MoCA), 6.14 (UPDRS-III), and 6.15 (HY).

Regarding MoCA, a positive association was found mostly with GCIPL thickness. As shown in Figure 6.5, this association was higher for the inner macular ring (maximum $R^2 = 0.09$). Related with this finding, the GCIPL rim height also showed a similarly good performance both when averaged across the entire macula and when measured separately for each angular directions. Additionally, the A_{22} coefficient of the mathematical model proposed by Ding et al. [74] also demonstrated predictive value ($R^2 = 0.08$). As for texture features, the standard deviation of GCIPL thickness across the entire macula was the best feature of its kind.

Neither thickness nor texture features showed a clear association with UPDRS-III. In fact, the only feature with a mild association with the outcome was the GCIPL CFT, that is, the GCIPL point thickness at the very foveal center ($R^2 = 0.06$).

In a similar vein, thickness and texture features did not explain more than a 5% of the variance in HY scores. This low yet non-negligible performance was mainly driven by a positive association with ELM-BM thickness (see bottom-right part of Figure 6.5) and a negative association with GCIPL standard deviation. Importantly, GCIPL foveal pit depth features outperformed the latter two and achieved the highest predictive performance ($R^2 = 0.08$).

6.3.3 Multivariate modeling

After establishing a reference performance and investigating individual OCT features, in this last section we explored if a combination of multiple features could improve prediction. This analysis encompassed two approaches: the first involved combining exclusively OCT features, while the second incorporated age as a fixed predictor in models using multiple OCT features. Age was included in the model because it showed predictive value for MoCA and is a basic predictor.

Regarding the concrete regression models, three multivariate methods were evaluated:

- *Least squares*: in each iteration the best 5 features are selected from the training set (i.e., N-1 samples) based on the highest correlation with the outcome. These features are then linearly combined to predict the outcome on the test subject.
- *Least squares + Lasso*: instead of selecting the features, in this approach a Lasso regular-

Thickness					Texture						Foveal pit					
Layer	Sector	R	R^2	RMSE	Layer	Sector	Name	R	R^2	RMSE	Layer	Sector	Name	R	R^2	RMSE
GCIPL	IT	0.32	0.09	4.34	GCIPL	Macula	Std. Dev.	0.31	0.08	4.36	GCIPL	I1	Rim height	0.34	0.10	4.31
GCIPL	II	0.29	0.07	4.38	GCIPL	I2	Std. Dev.	0.31	0.08	4.37	GCIPL	T	Rim height	0.34	0.10	4.32
GCIPL	⌀ 2-3	0.29	0.07	4.39	GCIPL	T2	Std. Dev.	0.31	0.07	4.37	GCIPL	T2	Rim height	0.34	0.09	4.33
GCIPL	IN	0.28	0.06	4.40	GCIPL	I3	Std. Dev.	0.30	0.07	4.39	GCIPL	Macula	Rim height	0.33	0.09	4.33
GCIPL	IS	0.28	0.06	4.41	GCIPL	T1	Std. Dev.	0.29	0.07	4.39	GCIPL	N3	Rim height	0.33	0.09	4.34

Table 6.13: Best OCT features for MoCA prediction. Sectors are defined in Figure 4.17. Rings are defined by their diameter range (⌀) in *mm*.

Thickness					Texture						Foveal pit					
Layer	Sector	R	R^2	RMSE	Layer	Sector	Name	R	R^2	RMSE	Layer	Sector	Name	R	R^2	RMSE
ELM-BM	ON	-0.15	0.00	11.37	ELM-BM	Macula	GLCM Max prob.	0.20	0.02	11.26	GCIPL	Macula	CFT	0.29	0.06	11.00
ELM-BM	⌀ 4-5	-0.13	-0.01	11.41	ELM-BM	Macula	GLCM Energy	0.19	0.02	11.28	TRT	S2	Scheibe (α)	-0.21	0.02	11.26
ONPL	C	-0.12	-0.01	11.41	ONPL	⌀ 1-2	Std. Dev.	-0.21	0.02	11.28	GCIPL	I2	Pit depth	-0.20	0.02	11.27
ELM-BM	N2	-0.12	-0.01	11.41	GCIPL	I2	Std. Dev.	-0.19	0.01	11.32	TRT	I3	Scheibe (α)	-0.20	0.02	11.28
ELM-BM	S1	-0.13	-0.01	11.42	ELM-BM	Macula	GLCM Sum entropy	-0.16	0.00	11.34	TRT	T1	Pit depth	-0.20	0.02	11.28

Table 6.14: Best OCT features for UPDRS-III prediction. Sectors are defined in Figure 4.17. Rings are defined by their diameter range (⌀) in *mm*.

Thickness					Texture						Foveal pit					
Layer	Sector	R	R^2	RMSE	Layer	Sector	Name	R	R^2	RMSE	Layer	Sector	Name	R	R^2	RMSE
ELM-BM	N3	-0.25	0.04	0.619	GCIPL	I2	Std. Dev.	-0.27	0.05	0.614	GCIPL	I2	Pit depth	-0.31	0.08	0.605
ELM-BM	S1	-0.24	0.04	0.621	GCIPL	N3	Std. Dev.	-0.22	0.03	0.619	GCIPL	I1	Pit depth	-0.29	0.06	0.611
ELM-BM	ON	-0.24	0.03	0.621	GCIPL	Macula	Std. Dev.	-0.21	0.02	0.620	GCIPL	N3	Pit depth	-0.29	0.06	0.611
ELM-BM	N2	-0.23	0.03	0.622	GCIPL	I1	Std. Dev.	-0.22	0.02	0.620	GCIPL	I	Pit depth	-0.28	0.06	0.612
ELM-BM	T1	-0.22	0.03	0.622	GCIPL	I3	Std. Dev.	-0.21	0.02	0.623	GCIPL	Macula	Pit depth	-0.27	0.05	0.613

Table 6.15: Best OCT features for Hoehn-Yahr score prediction. Sectors are defined in Figure 4.17.

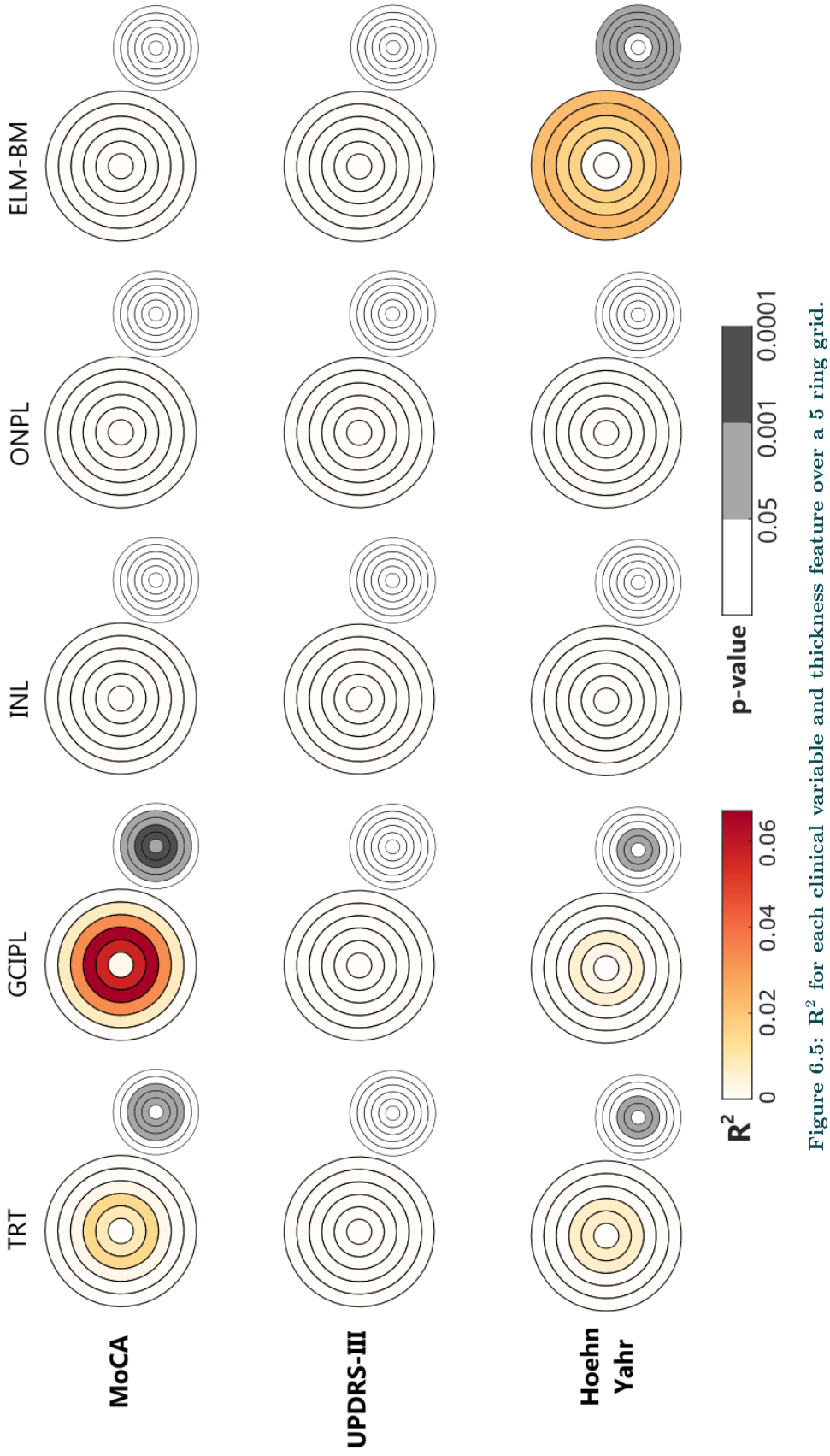


Figure 6.5: R^2 for each clinical variable and thickness feature over a 5 ring grid.

ization term is added to the least squares regression. The amount of regularization (λ) is selected by means of an inner 10-fold cross-validation.

- *SVM + Lasso*: same procedure as in the second method but using an SVM model instead of least squares. The hyperparameters of the SVM are tuned using an inner 10-fold cross-validation.

All the results are summarized in Table 6.16. Previously derived reference performance obtained by either a constant model, age, and disease duration alone is also displayed for comparison.

Overall, combining multiple OCT features did not show a clear performance improvement. Regarding MoCA, models using OCT features obtained a maximum R^2 of 0.09, notably below the performance obtained by an age-only model ($R^2 = 0.22$). This latter performance was only marginally improved when combining OCT features with age in a single least squares model ($R^2 = 0.23$).

As for UPDRS-III, OCT features showed low predictive power with a maximum R^2 of 0.06. Age and sex were not related with UPDRS-III either and, consequently, combining OCT features with age did not yield better results.

Finally, HY score could not be predicted by age. However, OCT features showed to be somewhat informative with an R^2 of 0.08. As with UPDRS-III, combining age with OCT features did not provide improvements in the predictive performance.

Strategy	**Model**	**MoCA**		**UPDRS-III**		**Hoehn-Yahr**	
		R^2	RMSE	R^2	RMSE	R^2	RMSE
Reference	constant*	-	4.58	-	11.5	-	0.629
	age + age^2	0.22	4.00	-0.03	11.5	-0.02	0.635
	disease duration	-0.02	4.61	-0.03	11.6	0.07	0.607
OCT	Least squares	0.09	4.34	0.06	11.0	0.08	0.605
	Least squares + Lasso	0.09	4.34	-0.02	11.5	0.06	0.610
	SVM + Lasso	0.07	4.39	0.02	11.3	0.03	0.620
OCT + Age†	Least squares	0.23	3.99	0.05	11.1	0.06	0.608
	Least squares + Lasso	0.21	4.03	-0.02	11.5	0.06	0.610
	SVM + Lasso	0.21	4.04	0.02	11.3	0.03	0.620

* Predict test value as the average value on the training set.

† Age was forced to be in all models.

Table 6.16: Regression results for different outcomes and modeling strategies.

6.4 Discussion

In this chapter we investigated whether conventional or advanced OCT features could provide valuable information for PD diagnosis or severity assessment. First, we focused on PD diagnosis and found that individual features showed a limited class separability (maximum AUC = 0.62). This likely indicates that OCT features do contain some information about PD changes yet not enough to reach an accurate diagnosis individually.

From the explored individual features those derived from the ETDRS grid worked better than those computed from the entire macular region. This suggests that averaging thickness points over a region as large as the entire macula hinders the detection of PD-related changes. Similarly, ETDRS features showed equal or similar performance than those derived from multiple rings or angles. This could indicate that PD-related changes are not better captured by splitting the macular area into more rings and angular sectors than the conventional ETDRS does.

Combining multiple OCT features improved the diagnostic performance up to an AUC of 0.67. This possibly reflects that, as no individual feature is able to accurately diagnose PD, the models benefit from combining multiple features.

As an alternative approach, we also explored a diagnostic methodology using both eyes (without averaging). This approach reached a lower accuracy than the one obtained averaging both eyes, which may signify that using both eyes introduces higher variability into the model and averaging is indeed beneficial. Nevertheless, we observed how the accuracy can be improved by only providing a diagnosis when the prediction of both eyes are in agreement. This result suggests that inter-eye agreement can be used as measurement of model uncertainty, which could be of value in settings where data from both eyes is available.

On another note, the diagnostic performance improved further when patients were split into subgroups based on disease duration, MoCA, and UPDRS-III. In fact, the obtained accuracy was higher for patients with cognitive impairment (i.e., MoCA below 26). A potential explanation to this effect is that as the cognition of patients worsens the retinal changes become more accentuated, which in turn, results in a more accurate diagnosis.

PD diagnosis has been addressed by previous work with diverging results (refer to Table 2.6 in Chapter 2 for a detailed list of studies). Notably, most studies included less than 55 patients [74,147–150,154,155,157], and showed a great variability in the reported performance. In fact, there were studies reporting low [147, 148, 154], medium [74, 149, 155], and high diagnostic performance [157]. This variability can be largely influenced by the small sample size. In addition, two larger studies have investigated PD diagnosis from OCT features. Garcia-Martin et al. trained a LDA model on ETDRS thickness features, achieving an AUC of 0.90 in a cohort with 200 controls and 111 patients [146]. In a second study, Pinkhardt et al. investigated foveal pit features in a cohort of 176 controls and 184 patients. The reported results were not as good (62% accuracy and AUC = 0.60) [156].

Compared to these previous studies, we explored a much wider range of OCT features in a sample of similar or larger size. However, the achieved performance was not on par with the best results reported in the literature (AUC = 0.90). A first explanation for this discrepancy are the existing methodological differences between the studies regarding OCT image processing and diagnostic model development. These differences complicate direct comparisons and may have impacted the results. Alternatively, the heterogeneity of the BHRI dataset could have had an influence as well. In fact, the individuals included in this dataset were imaged by different operators over a period of several years. Thus, certain differences between acquisitions are to be expected, which may have have impacted the results negatively. This limitation can also be seen as a more severe test to the hypothesis that OCT features can be reliably used to diagnose PD. That is, any potential clinical use of an OCT-based diagnostic model would inevitably involve different operators and acquisition protocols. Therefore, if such differences were to affect the performance of the model so dramatically, this would be an indicator of low model stability that would make it unsuitable for the task.

Beyond mere model performance, it is interesting to reflect on the features that worked best. In the literature, thickness features have been by far the most investigated with somewhat consistent results pointing towards an inner retinal thinning in PD [41]. Contrary to that, conventional thickness features did not yield great performance in our dataset. In fact, the best features were often derived from the central ETDRS sector, a macular area in which inner retinal layers are practically non-existent. This may indicate that the central sector is capturing an effect related with fixation problems in PD patients rather than a true biological effect.

In the second part of the chapter, we evaluated if OCT features could be used for an accurate clinical severity assessment. As the main conclusion, OCT features convey some information about cognitive status and motor impairment of a patient. However, the predictive performance is clearly insufficient to be considered for clinical use. Indeed, age alone was a better predictor for MoCA than any OCT feature. The age dependency of MoCA reflects the well-known cognitive deterioration

associated with aging and has been extensively described in the literature [204]. Importantly, this pattern is not confined to PD patients and is present also healthy subjects.

Nevertheless, it is important to further reflect on the features showing a highest association with MoCA. In this case, we found a clear positive association between MoCA and GCIPL thickness, a finding in line with previous literature investigating cognition in PD [145, 171, 174]. Interestingly, Sung et al. investigated GCIPL thickness with different angular sectors and found all to be positively correlated with MoCA [174]. Altogether, the evidence suggest that the GCIPL thinning associated with MoCA decline is radially homogeneous, but largely confined to the 0.5 to 2 mm radius region.

In any case, the association between cognitive decline and GCIPL does not necessarily need to be a causal relationship. Indeed, as we described in Chapter 5, GCIPL thickness decreases with age, and, therefore, one could hypothesize that aging is the underlying phenomenon driving both the GCIPL thinning as well as cognitive decline.

On the predictive modeling side, combining multiple OCT features either alone or with age did not significantly improve predictive performance. This can indicate that the existing relationship between MoCA and OCT features stems from a single effect well captured by GCIPL thickness measurements and, consequently, adding other features does not effectively increase the amount of information.

Regarding UPDRS-III, the best performance was obtained by the GCIPL CFT feature. It is important to note that the CFT measures the thickness of a layer at the very center of the fovea. Considering that the GCIPL thickness in that region is practically non-existent, it is possible that the CFT is capturing some kind of centering artifact correlated with motor impairment rather than a biological change. In this sense, it could be hypothesized that patients with worse motor impairment may have higher difficulties maintaining a stable fixation, which could lead to poorer centering. Nevertheless, all the images were manually reviewed to discard centering problems, so the effect may be subtle. This effect shares similarities with previous diagnostic results, where central thickness features showed the highest discriminative power.

Surprisingly, none of the conventional thickness features showed any clear association with UPDRS-III. This is not in line with previous literature reporting a negative correlation with the TRT [167] and GCIPL [145]. This difference may stem from dataset variability because, as already mentioned, patients were assessed across various research projects, dates, and clinicians, potentially impacting UPDRS-III measurements.

Finally, we found HY scores to be negatively associated primarily with GCIPL pit depth, CFT, and more weakly with ELM-BM and GCIPL thickness. The former parallel the CFT-UPDRS-III relationship described before and may be capturing a similar phenomenon related to motor impairment and image fixation. Regarding the former, previous studies reported a negative association with HY and the TRT [142, 170, 171], GCIPL [138, 145, 170–172], INL [171], and ONPL [171], but no study has investigated ELM-BM thickness. As a consequence, it is difficult to postulate an underlying biological mechanism and additional research is required to validate this finding.

In conclusion, the explored OCT features capture certain aspects of PD but are not sufficient to be considered for disease diagnosis or severity assessment in a real clinical setting. Nevertheless, although we explored a comprehensive set of features, other features and approaches not investigated here might be able to capture PD-related changes more effectively. For instance, deep-learning models trained on fundus images, thickness maps and even OCT volumes are starting to be explored and may boost the clinical utility of retinal imaging for PD [205].

7 | Patient evolution and monitoring

Previous cross-sectional studies have reported thinner inner retinal layers in PD [41]. However, as observed in the literature and highlighted by Chapter 6, the differences between patients and controls are small, and can be challenging to detect. This difficulty is partly a consequence of the high inter-subject variability of retinal thickness, which may mask small changes at a patient level. Therefore, it is of great interest to investigate how the retina evolves over time for the same eye and subject.

With this aim, in this Chapter we leverage two longitudinal cohorts to examine the evolution of the retina in PD patients and its relationship with disease progression. After describing data curation and image processing, four separate analyses are presented. First, Section 7.2 focuses on determining the structural aspects of the retina which change more prominently in PD. In Section 7.3 the estimated clinical evolution of the patients is described, and the link between retinal changes and clinical evolution is then explored in Section 7.4. Finally, Section 7.5 investigates whether OCT features measured in a single visit can be used to predict cognitive and motor evolution of a patient.

7.1 Subjects and OCT features

Subjects from both the BHRI and AlzEye datasets were used (Table 7.1). From the BHRI dataset, a total of 158 patients and 72 controls were included. As previously described in Section 3.3, all included BHRI subjects underwent a screening protocol to exclude those with potential confounding factors which could affect OCT measurements and clinical outcomes. For every PD patient in BHRI the following clinical variables were measured: MoCA, HY score, and UPDRS-III.

Dataset	Group	N	Age (years)*	Male (%)	Follow-up* (years)	Visits*
BBHRI	Control	72	61.3 (7.6)	42.3	2.8 (1.7)	2.4 (0.5)
	PD	158	64.9 (8.6)	65.2	2.6 (1.7)	2.4 (0.5)
AlzEye	Control	873	75.7 (9.4)	60.6	1.4 (1.6)	4.4 (5.6)
	PD	167	76.7 (8.6)	55.7	1.3 (1.4)	4.7 (6.0)

*mean (σ)

Table 7.1: Subjects included in the longitudinal analysis.

Two groups were initially formed with subjects from the AlzEye database who possessed macular OCT images ready for analysis: the PD group ($N = 1283$) and the control group ($N = 125110$). The PD group was defined as individuals with at least one hospital admission labeled as PD (i.e., an entry in the HES database with an ICD code for PD). The data curation process is outlined in Figure 7.1.

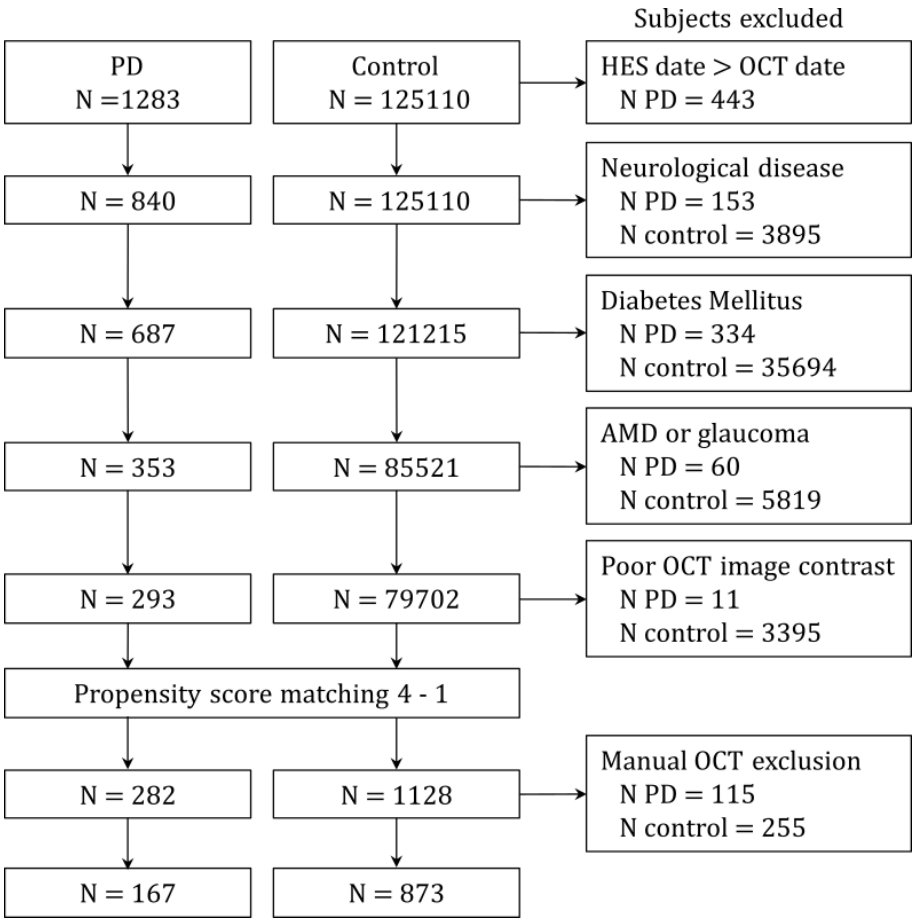


Figure 7.1: AlzEye data curation process. The figure shows how many subjects remain and how many are excluded in each step of the data curation process.

First, all PD images captured before the first HES date with a PD code were excluded. This was to ensure that only images obtained after the onset of the disease were included for PD patients. Subjects with Diabetes Mellitus were also excluded to align with the exclusion criteria used in the BHRI dataset. Then, individuals with ICD codes indicating the neurological conditions of AD, cerebral amyloid angiopathy, Huntington's disease, motor neuron disease, MS, progressive supranuclear palsy, schizophrenia, subarachnoid hemorrhage, and vascular dementia were removed. Ophthalmological labels for AMD and glaucoma were considered for exclusion and OCT volumes with poor contrast were subsequently eliminated based on the image quality metric provided by the Topcon device metadata. The threshold for removal (set at < 29) was determined by maximizing the Youden Index using a separate dataset that included manually graded Topcon OCT images from the UKBB dataset [206]. Propensity score matching based on age, sex, ethnicity, and hypertension was employed to select four control subjects for each PD subject. In the final step, the included images were manually inspected by two experienced graders (i.e. a biomedical engineer and an optometrist) to identify and remove cases with image artifacts, poor image centration, segmentation errors, and ocular lesions. This inspection process relied on summary reports, an example of which is presented in Figure 7.2. The final AlzEye cohort consisted of 167 PD patients and 873 controls.

It should be noted that the two final datasets were significantly different. BHRI subjects had between 1 and 3 visits at fixed time points of around 1, 2, 3, and 5 years. Conversely, there was a higher heterogeneity in AlzEye subjects, who had more visits but were unevenly distributed with a shorter follow-up time. Additionally, the AlzEye dataset did not include any clinical variables related to disease duration, cognitive status, and motor impairment. For this reason, the BHRI was employed as the primary dataset and AlzEye was only used as a replication dataset for analyses solely involving OCT measurements (Section 7.2).

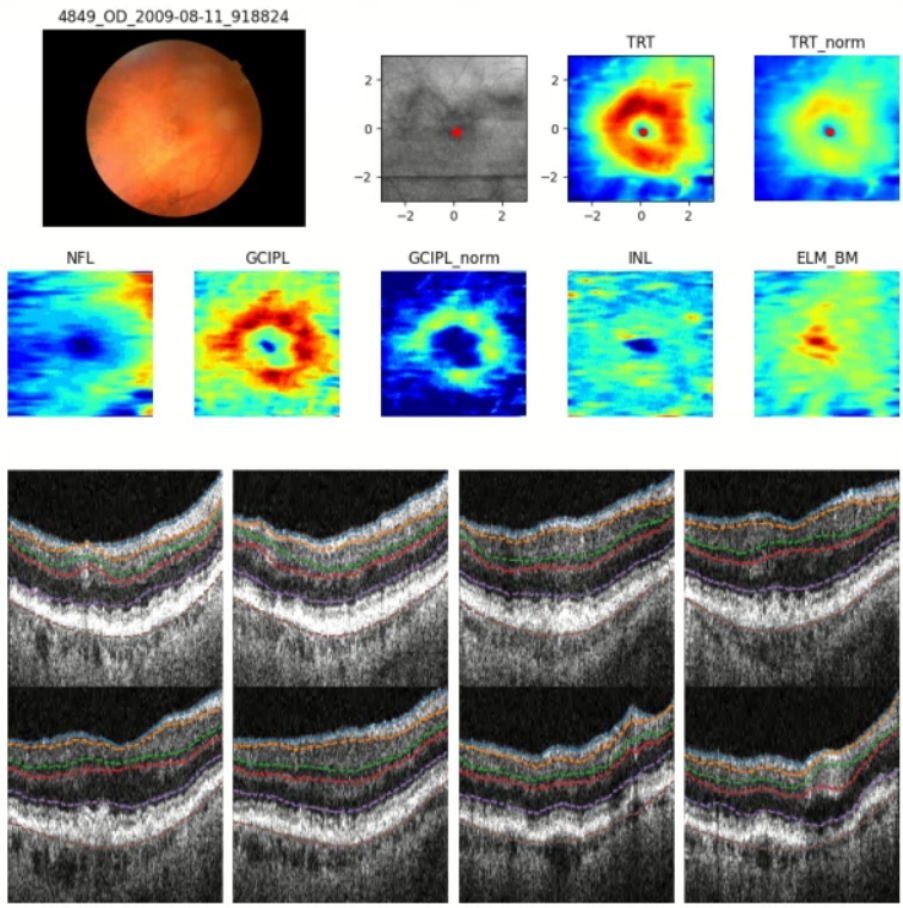


Figure 7.2: Example summary figure used for AlzEye image exclusion. A color fundus image acquired along with OCT images is shown in the top-left corner to identify pathologies. Top-center grayscale reflectance map is displayed to identify blinking artifacts. Then, multiple thickness maps are shown. These were used to detect incorrect centration as well as ocular pathologies. Finally, 8 B-scans uniformly selected are also displayed to further evaluate segmentation errors and quality problems.

Both datasets underwent similar image processing as described in Section 4.5. The analyzed OCT features were the same in both datasets and included:

- *Sectorized thickness*: analyzed retinal layers were the TRT, GCIPL, INL, ONPL, and ELM-BM. We used three sectorizations: macula, ETDRS rings, and 5 rings.
- *Foveal pit morphology*: the following features were calculated for both the TRT and GCIPL: *rim height*, *rim radius*, *pit depth*, and *mean slope*. A single average value for the entire fovea was analyzed.

All the computed features were averaged between left and right eyes to obtain a single value per subject and visit. In cases where only one eye was included, the values from that particular eye were used as the subject average. Texture features were not investigated in this chapter, as the primary focus was on discerning structural alterations within the retina and changes in textural attributes lack a direct structural correlate.

7.2 Longitudinal retinal changes

In a first analysis, longitudinal changes in retinal thicknesses and foveal pit morphology were investigated to determine if any of the features evolved differently in PD patients. For each retinal feature a mixed-effects linear regression model with an interaction term between follow-up time and having PD was fitted (Equation 7.1). Models were adjusted for sex (β_{sex}) and age at baseline (β_{age}). The latter was centered at 60 years old to obtain a meaningful intercept. The statistical dependence across visits was modeled by a random intercept for each subject (γ_{sub}). No random slope was used because many subjects had only 1 or 2 visits, which was insufficient to accurately fit such a complex model.

$$y = \beta_0 + \beta_{sex} isMale + \beta_{age} age60 + \beta_{hc} time + \beta_{pd} PD + \beta_{\Delta pd} timePD + \gamma_{sub} \tag{7.1}$$

From the fitted model, two coefficients were studied: the annualized rate of change in controls (β_{hc}), and the difference in that change between controls and patients ($\beta_{\Delta pd}$) (e.g., additional annualized thinning of a certain layer due to having PD). For each coefficient a p-value and a 95% confidence interval were reported. This analysis was carried out first for BHRI subjects and then replicated in AlzEye.

The results for ETDRS sectors in the BHRI dataset are illustrated in Figure 7.3. Correspondingly, Table 7.2 displays the estimated annualized rates of change for foveal pit morphology parameters and retinal thicknesses in the entire macular region, inner ring, and outer ring. Estimates are given in μm/year for all features except the *mean slope*, which is measured in °/year.

The TRT was found to decrease an additional 0.56 μm/year ($p = 0.0001$) in PD patients from the BHRI dataset. The equivalent estimation for AlzEye was 0.72 μm/year ($p = 0.012$). The GCIPL also showed an increased thinning in PD. While in the BHRI dataset the differences were statistically significant in all three regions, in AlzEye the effect was more evident in the parafoveal region. Importantly, in this region the increased thinning estimate from both datasets was very similar: 0.31 μm/year ($p = 0.0005$) and 0.27 μm/year ($p = 0.0076$) for BHRI and AlzEye, respectively. Results from AlzEye did not reveal any other effect on either the INL, ONPL, or ELM-BM layers. Conversely, results from the BHRI dataset did present a thinning effect of the ONPL of up to 0.30 μm/year ($p = 0.006$) in the parafoveal region.

Interestingly, a thinning effect was also observed in controls. For instance, macular GCIPL was found to decrease at 0.21 μm/year in BHRI. This phenomenon was accentuated in AlzEye, where the TRT, GCIPL, INL, and ELM-BM were all found to become thinner in healthy subjects for some of the sectors.

In addition to thickness differences, a few foveal pit measurements also showed significant differences between controls and patients. The most consistent finding between both datasets was a decrease in the *rim height* for both TRT and GCIPL. Specifically, TRT had an additional decrease in PD patients of 0.72 μm/year ($p = 0.0001$) and 1.00 μm/year ($p = 0.011$) in BHRI and AlzEye, respectively. As for the GCIPL, the increased annualized thinning in patients was 0.28 μm/year ($p = 0.0053$) and 0.35 μm/year ($p = 0.0012$) in each dataset. Although no clear effect of any of the remaining foveal features was observed in AlzEye, the GCIPL *pit depth* was found to significantly decrease in BHRI PD patients.

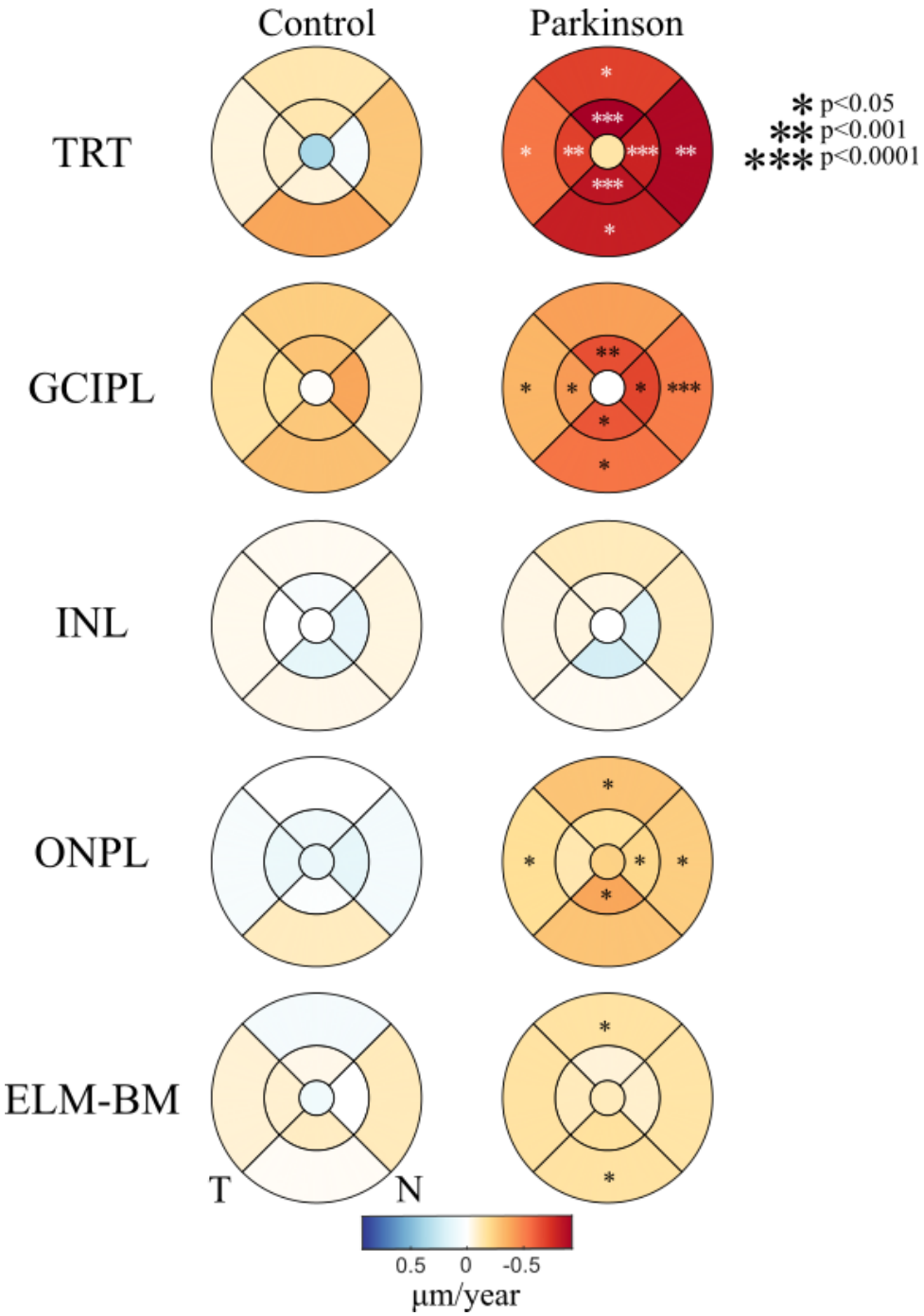


Figure 7.3: Longitudinal changes in ETDRS thicknesses in the BHRI dataset. Each column shows the annualized rate of thickness change in control subjects (β_{hc}) and PD patients ($\beta_{hc}+\beta_{\Delta pd}$), respectively. Sectors in the second column are marked with asterisks depending on the p-value of the interaction ($\beta_{\Delta pd}$).

Category	Layer	Feature	BBHRI			AlzEye		
			β_{hc} (X / year)	$\beta_{\Delta pd}$ (X / year)	p-value	β_{hc} (X / year)	$\beta_{\Delta pd}$ (X / year)	p-value
Thickness	TRT	Macula	-0.15 [-0.39, 0.08]	-0.56 [-0.85, -0.28]	0.0001*	-0.54 [-0.76, -0.32]	-0.72 [-1.27, -0.16]	0.012*
		Inner ring	-0.06 [-0.33, 0.22]	-0.75 [-1.09, -0.42]	$1.1{\cdot}10^{-5}$*	-0.63 [-0.88, -0.38]	-0.82 [-1.46, -0.18]	0.012*
		Outer ring	-0.23 [-0.47, 0.01]	-0.51 [-0.81, -0.22]	0.0007*	-0.55 [-0.77, -0.33]	-0.65 [-1.21, -0.09]	0.023*
	GCIPL	Macula	-0.21 [-0.31, -0.11]	-0.25 [-0.37, -0.13]	$5.7{\cdot}10^{-5}$*	-0.24 [-0.30, -0.17]	-0.16 [-0.34, 0.01]	0.064
		Inner ring	-0.28 [-0.42, -0.13]	-0.31 [-0.49, -0.14]	0.0005*	-0.38 [-0.46, -0.30]	-0.27 [-0.47, -0.07]	0.0076*
		Outer ring	-0.22 [-0.33, -0.10]	-0.24 [-0.38, -0.11]	0.0006*	-0.20 [-0.28, -0.13]	-0.11 [-0.30, 0.07]	0.24
	INL	Macula	-0.00 [-0.06, 0.06]	-0.03 [-0.10, 0.04]	0.45	-0.12 [-0.15, -0.07]	-0.02 [-0.12, 0.09]	0.74
		Inner ring	0.08 [-0.04, 0.20]	-0.02 [-0.16, 0.13]	0.81	-0.17 [-0.22, -0.12]	-0.10 [-0.24, 0.03]	0.13
		Outer ring	-0.04 [-0.09, 0.01]	-0.03 [-0.10, 0.03]	0.29	-0.12 [-0.16, -0.08]	0.02 [-0.09, 0.13]	0.71
	ONPL	Macula	0.03 [-0.11, 0.16]	-0.26 [-0.43, -0.09]	0.003*	0.05 [-0.07, 0.16]	-0.19 [-0.48, 0.10]	0.21
		Inner ring	0.08 [-0.08, 0.25]	-0.30 [-0.50, -0.10]	0.003*	-0.09 [-0.21, 0.04]	-0.17 [-0.48, 0.15]	0.30
		Outer ring	-0.01 [-0.16, 0.13]	-0.25 [-0.42, -0.07]	0.006*	0.09 [-0.03, 0.21]	-0.17 [-0.46, 0.13]	0.28
	ELM-BM	Macula	-0.04 [-0.14, 0.06]	-0.10 [-0.22, 0.02]	0.11	-0.26 [-0.38, -0.15]	-0.21 [-0.50, 0.07]	0.15
		Inner ring	-0.05 [-0.18, 0.08]	-0.05 [-0.20, 0.11]	0.53	0.02 [-0.12, 0.16]	-0.26 [-0.63, 0.10]	0.15
		Outer ring	-0.04 [-0.14, 0.06]	-0.12 [-0.24, 0.00]	0.057	-0.35 [-0.47, -0.24]	-0.22 [-0.51, 0.08]	0.15
Foveal pit morphology	TRT	Rim height	-0.09 [-0.39, 0.20]	-0.72 [-1.08, -0.36]	0.0001*	-0.59 [-0.89, -0.28]	-1.00 [-1.77, -0.23]	0.011*
		Pit depth	-0.46 [-1.04, 0.12]	-0.09 [-0.79, 0.62]	0.81	-0.67 [-1.30, -0.04]	-0.22 [-1.83, 1.40]	0.79
		Rim radius	-3.37 [-6.08, -0.65]	-0.51 [-3.81, 2.80]	0.76	-2.01 [-4.31, 0.28]	-2.96 [-8.83, 2.91]	0.32
		Mean slope	-0.00 [-0.04, 0.03]	-0.00 [-0.05, 0.04]	0.88	-0.02 [-0.05, 0.01]	0.01 [-0.07, 0.08]	0.87
	GCIPL	Rim height	-0.33 [-0.49, -0.17]	-0.28 [-0.47, -0.08]	0.0053*	-0.42 [-0.50, -0.34]	-0.35 [-0.56, -0.14]	0.0012*
		Pit depth	0.10 [-0.19, 0.39]	-0.51 [-0.87, -0.16]	0.0047*	-0.50 [-0.70, -0.29]	0.07 [-0.45, 0.60]	0.78
		Rim radius	-0.54 [-3.25, 2.17]	-3.20 [-6.50, 0.10]	0.057	0.44 [-1.44, 2.31]	-1.81 [-6.59, 2.96]	0.46
		Mean slope	0.00 [-0.02, 0.03]	-0.01 [-0.03, 0.02]	0.55	-0.03 [-0.04, -0.01]	0.01 [-0.02, 0.04]	0.53

* $p < 0.05$

Table 7.2: Annualized rates of change of retinal thicknesses and foveal parameters. β_{hc}: annualized change rate in controls. $\beta_{\Delta pd}$: additional change in PD patients per year. The coefficients are measured in μm/year except for the mean slope (°/year). Each coefficient is shown with a corresponding 95% confidence interval.

As an illustration of the differences between layers, Figure 7.4 depicts the annualized thinning rate for the GCIPL in each dataset.

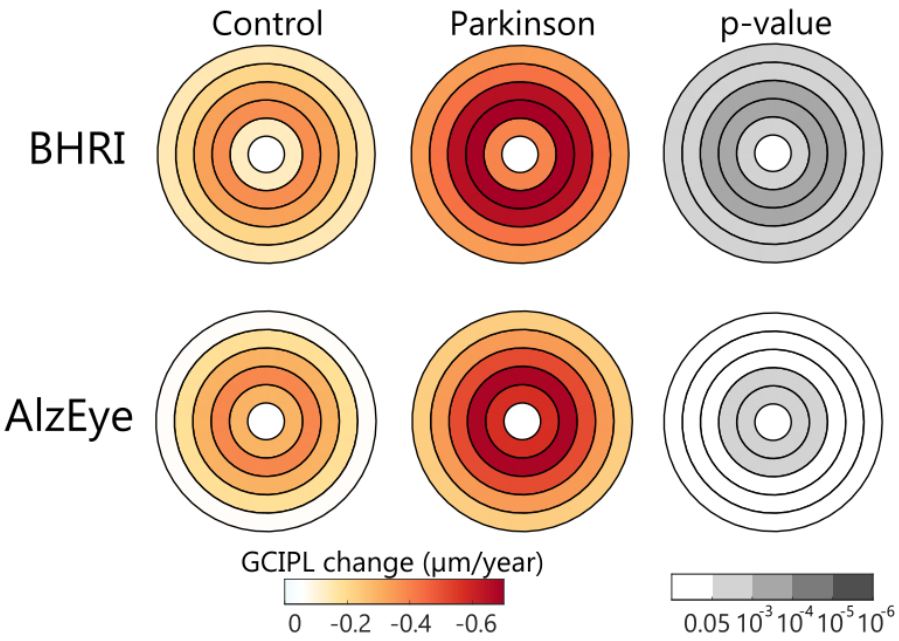


Figure 7.4: Longitudinal GCIPL thickness changes. First two columns show the annualized rate of change of GCIPL thickness in control subjects and PD patients, respectively. Each row shows equivalent results for each dataset.

Corresponding numerical results are displayed in Table 7.3. As observed, the highest differences in the BHRI dataset correspond to the 1-2 mm radius ring. On the other hand, although the pattern is similar, in AlzEye the strictly parafoveal ring (i.e., 0.5-1.5 mm radius ring) shows the strongest effect.

Ring	**BBHRI**			**AlzEye**		
	β_{hc}	$\beta_{\Delta pd}$	p-value	β_{hc}	$\beta_{\Delta pd}$	p-value
⌀ 1-2 mm	-0.13 [-0.32, 0.06]	-0.27 [-0.49, -0.04]	0.0217*	-0.32 [-0.42, -0.23]	-0.27 [-0.51, -0.03]	0.028*
⌀ 2-3 mm	-0.37 [-0.52, -0.22]	-0.34 [-0.52, -0.16]	$3.1 \cdot 10^{-4}$*	-0.42 [-0.50, -0.34]	-0.27 [-0.48, -0.07]	0.0096*
⌀ 3-4 mm	-0.33 [-0.48, -0.19]	-0.33 [-0.50, -0.15]	$3.3 \cdot 10^{-4}$*	-0.33 [-0.41, -0.25]	-0.19 [-0.39, 0.02]	0.0074*
⌀ 4-5 mm	-0.21 [-0.33, -0.08]	-0.23 [-0.39, -0.08]	0.0033*	-0.22 [-0.29, -0.14]	-0.14 [-0.34, 0.05]	0.15
⌀ 5-6 mm	-0.15 [-0.26, -0.04]	-0.20 [-0.33, -0.07]	0.0024*	-0.09 [-0.17, -0.02]	-0.16 [-0.35, 0.04]	0.11

* $p < 0.05$

Table 7.3: Annualized rates of change of GCIPL thickness β_{hc}: annualized change rate in controls. $\beta_{\Delta pd}$: additional change in PD patients per year. Both coefficients are measured in μm/year. Each coefficient is shown with a corresponding 95% confidence interval.

To further illustrate the effect on the GCIPL, Figure 7.5 shows the evolution of parafoveal GCIPL (pfGCIPL) for each subject in the BHRI dataset. We highlight this feature as an example feature with one of the highest increased thinning in patients in both datasets. Despite non-negligible variation across measurements, the overall effect is visually apparent.

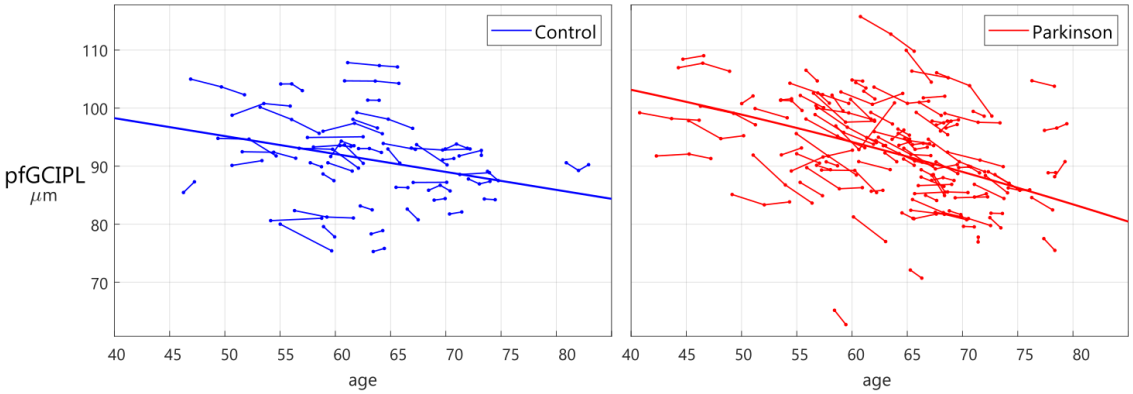


Figure 7.5: Evolution of parafoveal GCIPL thickness in each subject. Each individual line represents a single subject from the BHRI dataset. The thicker line represents the average effect of age.

In addition to the numerical results it is of interest to visualize retinal changes directly in a B-scan to contextualize the size of the changes. Figure 7.6 shows the longitudinal evolution of the retina of a PD patient who had a clear decrease in GCIPL thickness (i.e., pfGCIPL values were 105.7, 104.1, and 97.9 μm at baseline, year 3 and year 5 visits). More concretely, the en-face thickness map is depicted along with 5 B-scans covering the central part of the macula. As shown, thickness changes can be slightly intuited in the thickness maps but are hard to pinpoint when looking directly into B-scan segmentation.

In the previous analyses longitudinal changes have been analyzed in absolute units (i.e., using μm for thickness changes). Nevertheless, it is known that the initial thickness varies for each layer, subject, and sector. Therefore, it is of interest as well to investigate relative changes in thickness values. To this aim, thickness values of each feature were transformed into % changes with respect to the baseline visit. Then, the same regression analysis described before was carried out. The results of this analysis are shown in Figure 7.7 (BHRI dataset) and Table 7.4 (both datasets).

When measuring the changes in percentage terms, the GCIPL layer showed the highest changes, with an annualized decrease of 0.32% and 0.66% in macular thickness in controls and patients, respectively. The decrease in the thickness of the other layers in patients was: TRT (0.23%), ONPL (0.26%), and ELM-BM (0.15%). As observed when analyzing absolute thickness changes, the INL did not reveal a clear thinning effect.

Respect to regional differences, normalized changes in GCIPL thickness were higher for the nasal sector compared to the temporal (-0.79%/year and -0.40%/year for outer-nasal and outer-temporal, respectively).

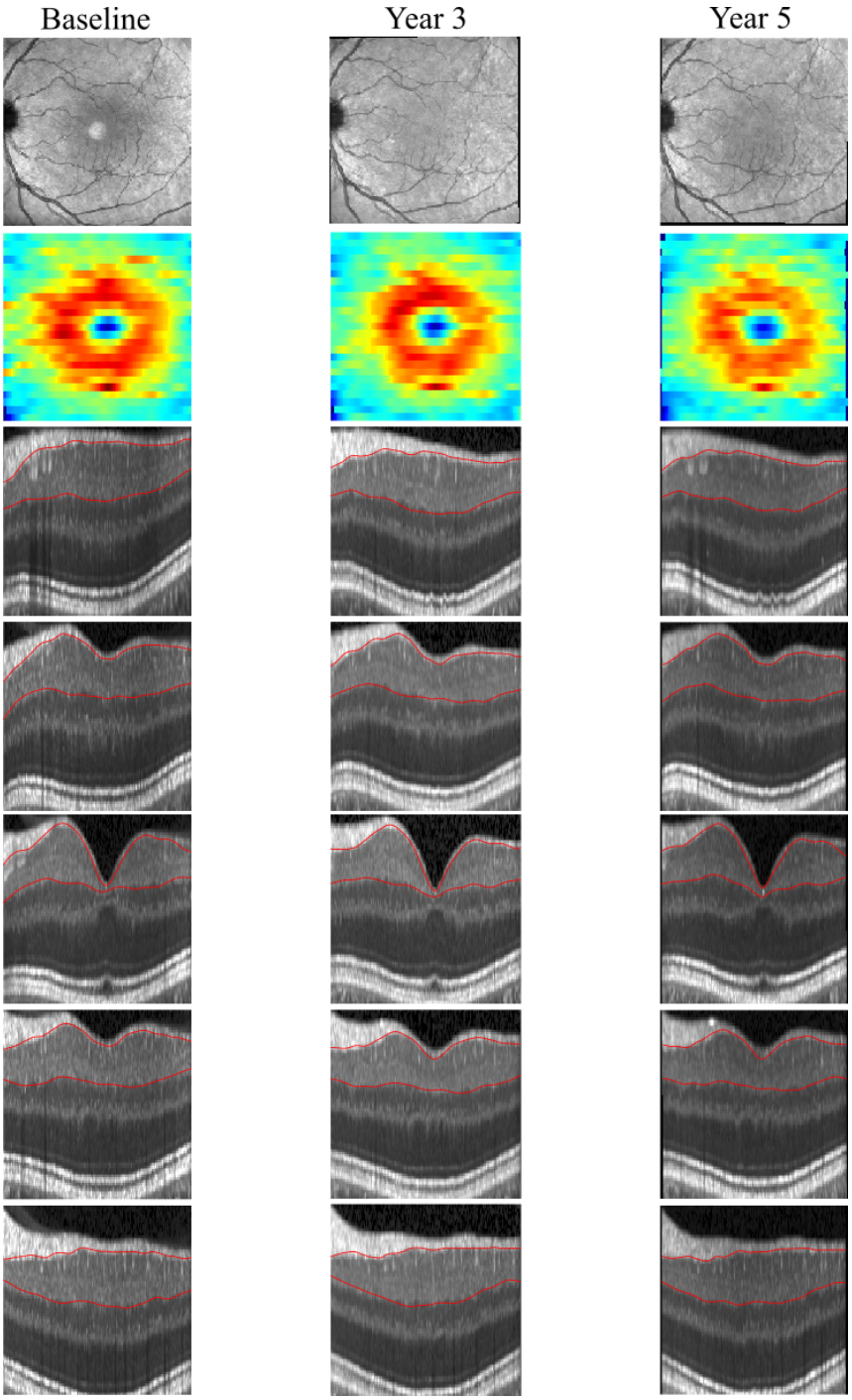


Figure 7.6: Longitudinal evolution of the retina of a PD patient. Each column represents a single visit. First and second rows depict the retinal fundus and the GCIPL thickness map, respectively. Subsequent rows show centermost B-scans with the segmentation of the GCIPL. The aspect ratio has been modified for visualization purposes but is the same for each image.

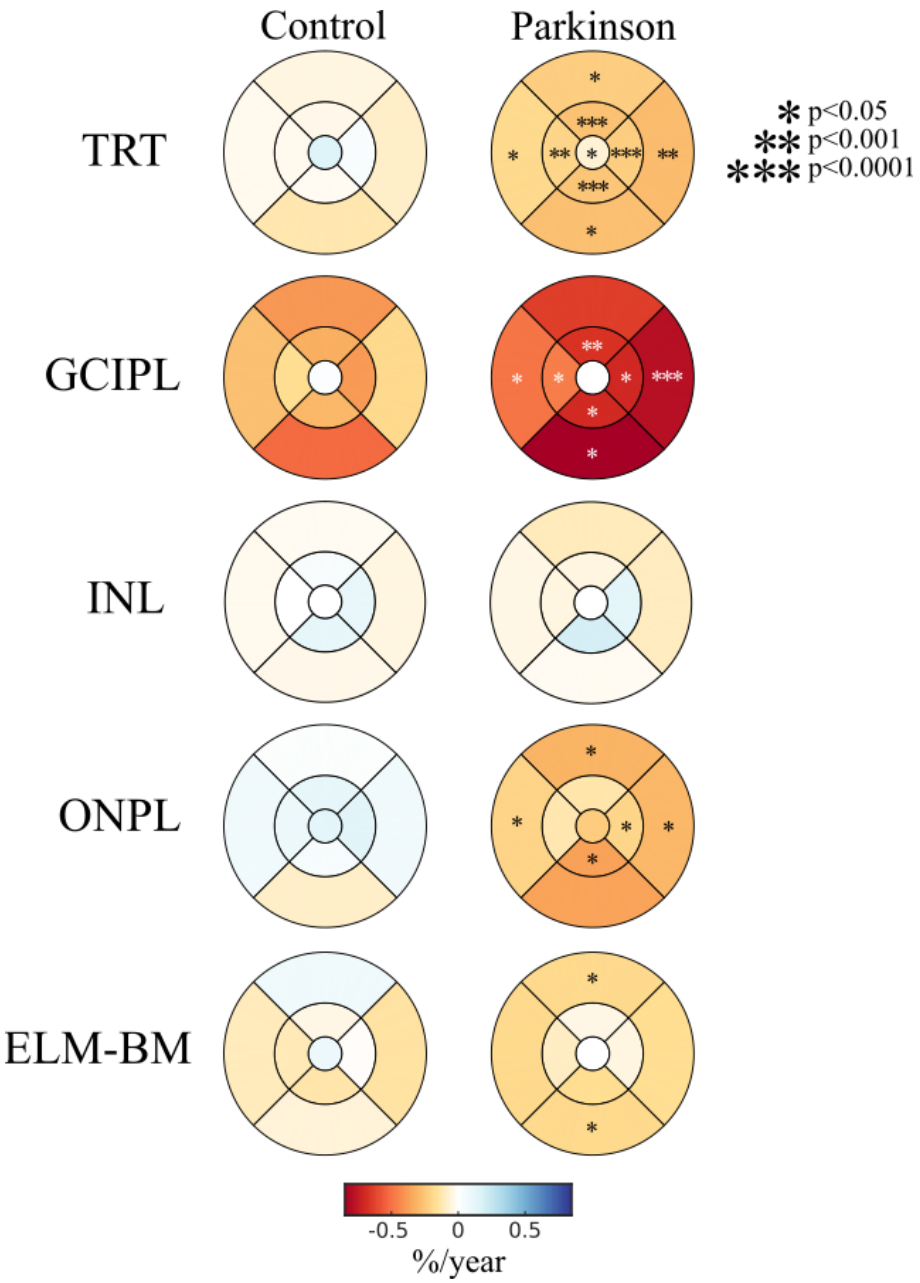


Figure 7.7: Normalized longitudinal changes in ETDRS thicknesses. Each column shows the annualized rate of thickness change in control subjects (β_{hc}) and PD patients ($\beta_{hc} + \beta_{\Delta pd}$), respectively. Sectors in the second column are marked with asterisks depending on the p-value of the interaction ($\beta_{\Delta pd}$).

Category	Layer	Feature	BBHRI			AlzEye		
			β_{hc} (% / year)	$\beta_{\Delta pd}$ (% / year)	p-value	β_{hc} (% / year)	$\beta_{\Delta pd}$ (% / year)	p-value
Thickness	TRT	Macula	-0.05 [-0.11, 0.01]	-0.18 [-0.26, -0.11]	$1.3{\cdot}10^{-6*}$	-0.17 [-0.23 -0.12]	-0.21 [-0.35 -0.06]	0.005*
		Inner ring	-0.01 [-0.07, 0.05]	-0.22 [-0.30, -0.15]	$1.4{\cdot}10^{-8*}$	-0.20 [-0.26 -0.14]	-0.21 [-0.36 -0.06]	0.007*
		Outer ring	-0.22 [-0.30, -0.15]	-0.17 [-0.25, -0.09]	$1.8{\cdot}10^{-5*}$	-0.18 [-0.24 -0.12]	-0.19 [-0.34 -0.04]	0.012*
	GCIPL	Macula	-0.32 [-0.43, -0.21]	-0.34 [-0.47, -0.21]	$6.8{\cdot}10^{-7*}$	-0.32 [-0.39 -0.24]	-0.26 [-0.47 -0.06]	0.012*
		Inner ring	-0.30 [-0.41, -0.18]	-0.33 [-0.48, -0.19]	$5.2{\cdot}10^{-6*}$	-0.43 [-0.50 -0.36]	-0.30 [-0.50 -0.11]	0.0019*
		Outer ring	-0.35 [-0.49, -0.22]	-0.36 [-0.53, -0.20]	$2.2{\cdot}10^{-5*}$	-0.28 [-0.37 -0.19]	-0.21 [-0.45 0.02]	0.073
	INL	Macula	0.03 [-0.10, 0.16]	-0.07 [-0.22, 0.09]	0.41	-0.32 [-0.41 -0.22]	-0.11 [-0.36 0.13]	0.37
		Inner ring	0.28 [0.05, 0.51]	-0.03 [-0.31, 0.24]	0.8	-0.38 [-0.48 -0.28]	-0.26 [-0.52 0.01]	0.056
		Outer ring	-0.09 [-0.22, 0.03]	-0.10 [-0.25, 0.05]	0.21	-0.35 [-0.45 -0.25]	-0.02 [-0.29 0.25]	0.88
	ONPL	Macula	0.05 [-0.07, 0.17]	-0.31 [-0.45, -0.16]	$4.9{\cdot}10^{-5*}$	0.10 [-0.01, 0.21]	-0.20 [-0.48, 0.08]	0.16
		Inner ring	0.10 [-0.03, 0.22]	-0.30 [-0.45, -0.15]	0.0001*	-0.07 [-0.17 0.03]	-0.14 [-0.40 0.12]	0.29
		Outer ring	0.01 [-0.13, 0.14]	-0.31 [-0.47, -0.15]	0.00019*	0.16 [0.04 0.27]	-0.19 [-0.50 0.12]	0.23
	ELM-BM	Macula	-0.05 [-0.15, 0.04]	-0.10 [-0.21, 0.02]	0.1	-0.35 [-0.45 -0.24]	-0.10 [-0.38 0.17]	0.46
		Inner ring	-0.07 [-0.19, 0.05]	-0.02 [-0.16, 0.13]	0.84	-0.05 [-0.18 0.08]	-0.08 [-0.42 0.27]	0.67
		Outer ring	-0.02 [-0.16, 0.13]	-0.12 [-0.24, -0.01]	0.037*	-0.45 [-0.56 -0.39]	-0.13 [-0.42 0.16]	0.39
Foveal pit morphology	TRT	Rim height	-0.02 [-0.09, 0.04]	-0.20 [-0.28, -0.12]	8.3e-07	-0.179 [-0.244 -0.114]	-0.25 [-0.422 -0.078]	0.004*
		Pit depth	-0.45 [-0.89, -0.02]	0.15 [-0.37, 0.68]	0.56	-0.94 [-1.34 -0.54]	0.33 [-0.74 1.39]	0.55
		Rim radius	-0.30 [-0.50, -0.10]	-0.04 [-0.28, 0.20]	0.72	-0.21 [-0.37 -0.04]	-0.27 [-0.71 0.17]	0.23
		Mean slope	-0.08 [-0.54, 0.39]	0.20 [-0.37, 0.76]	0.49	-0.48 [-0.83 -0.13]	0.83 [-0.11 1.77]	0.08
	GCIPL	Rim height	-0.32 [-0.44, -0.20]	-0.27 [-0.41, -0.12]	0.00027*	-0.43 [-0.50 -0.37]	-0.34 [-0.52 -0.16]	0.00018*
		Pit depth	0.21 [-0.05, 0.46]	-0.64 [-0.95, -0.34]	$4.1{\cdot}10^{-5*}$	-0.582 [-0.79 -0.38]	0.38 [-0.16 0.91]	0.17
		Rim radius	-0.01 [-0.20, 0.17]	-0.29 [-0.51, -0.06]	0.012*	0.01 [-0.11 0.13]	-0.09 [-0.41 0.23]	0.58
		Mean slope	0.22 [-0.12, 0.56]	-0.27 [-0.68, 0.15]	0.2	-0.44 [-0.68 -0.22]	0.48 [-0.14 1.10]	0.13

* $p < 0.05$

Table 7.4: Annualized relative ates of change of retinal thicknesses and foveal parameters. β_{hc}: annualized change rate in controls. $\beta_{\Delta pd}$: additional change in PD patients per year. Each coefficient is shown with a corresponding 95% confidence interval.

After exploring overall longitudinal retinal changes, we further explored if the thinning observed was homogeneous across subgroups of subjects. In this sense, one could hypothesize that the thinning rate does not necessarily need to be similar for all the patients and may change across lifespan and disease progression. To explore this idea we investigated if patients with a thinner baseline pfGCIPL had a different thinning rate. We focused on the pfGCIPL thickness feature because it showed the highest consistency between the two datasets (see Figure 7.4) and has also been previously identified as an important target of PD in a previous work [176]. In addition, this retinal feature reported the highest correlation with MoCA in the cross-sectional analysis presented in Figure 6.5.

First, we built a normative distribution of pfGCIPL thickness in healthy controls. From this distribution we defined the 25 percentile as the cut-off threshold and considered retinas below that value to have low pfGCIPL. The computed cut-off values were 89.8 μm and 78.7 μm for BHRI and AlzEye datasets, respectively. The differences in the cut-off value can be explained by the differences in scanning devices and the older population in AlzEye. Regarding the number of subjects, in BHRI the entire pool of healthy controls used in Chapter 5 was used. In AlzEye only the controls described in Table 7.1 were considered becauase this was the subset of the dataset that underwent careful quality assurance before inclusion.

Based on the cut-off value, PD patients and controls were separately split into two groups with a *low* and *high* baseline pfGCIPL thicknesses. Then, a regression analysis was carried out to estimate the annualized rate of change of pfGCIPL thickness. The same model was applied to each subgroup (Equation 7.2). The model contained a fixed term for the effect of interest (follow-up time (β_t)) and was adjusted for sex (β_{sex}) and age at baseline (β_{age}). As in the previous section, the model included a random intercept for each subject to model the statistical dependency across visits of the same subject (γ_{sub}). This model was fitted to the four combinations of PD, HC, and *low*/*high* pfGCIPL at baseline.

$$y = \beta_0 + \beta_{sex} isMale + \beta_{age} age60 + \beta_t time + \gamma_{sub} \tag{7.2}$$

To formally compare both *low* and *high* groups in each category, an interaction term between follow-up time and ocurrance of a low pfGCIPL was added to the model (β_{tLow}) (see Equation 7.3). This model was fitted to PD and HC groups separately.

$$y = \beta_0 + \beta_{sex} isMale + \beta_{age} age60 + \beta_t time + \beta_{low} low + \beta_{tLow} time \cdot low + \gamma_{sub} \tag{7.3}$$

The results of the analysis are reported in Table 7.5. The seventh column displays the results obtained with Equation 7.2, while the last column shows the estimated interaction term in Equation 7.3.

When studying only the control group, the annualized pfGCIPL loss was found to be very similar between *low* and *high* subgroups. In fact, the estimated interaction term was close to zero in both BHRI (0.01 μm/year, $p = 0.93$), and AlzEye (-0.05 μm/year, $p = 0.65$). In contrast, results in PD clearly showed that PD patients with a *high* baseline pfGCIPL thickness had a more accentuated loss. For instance, BHRI patients belonging to the *high* group were estimated to experience a faster thinning of 0.25 μm/year ($p = 0.036$). AlzEye results followed a similar trend with an estimation of a 0.32 μm/year ($p = 0.11$) increased thinning in patients with *high* pfGCIPL. However, in the latter case, statistical significance was not achieved.

In light of these results, we further investigated potential differences between PD patients in *low* and *high* subgroups. As the main differentiating factor, patients in the *low* group were older on average in both BHRI and AlzEye (Table 7.5). Regarding sex, both *low* and *high* groups contained a different proportion of males. However, the differences in male proportion pointed in different directions for BHRI and AlzEye, suggesting that sex may not influence the results.

Dataset	Class	Group	N	Age (years)	Male (%)	β_t (μm/year)	β_{tLow} (μm/year)
BHRI	**HC**	High	45	59.6 (7.2)	46.7	-0.29 [-0.42, -0.15]	0.01 [-0.26, 0.29]
		Low	27	64.3 (7.3)	34.7	-0.26 [-0.48, -0.04]	$p = 0.93$
	PD	High	92	62.0 (8.1)	72.8	-0.66 [-0.79, -0.53]	0.25 [0.02, 0.49]
		Low	66	68.9 (7.8)	54.5	-0.40 [-0.58, -0.21]	$p = 0.036^*$
AlzEye	**HC**	High	665	74.7 (9.6)	60.3	-0.38 [-0.47, -0.29]	-0.05 [-0.28, 0.18]
		Low	218	78.7 (8.4)	61.5	-0.41 [-0.60, -0.21]	$p = 0.65$
	PD	High	115	75.3 (8.8)	53.0	-0.73 [-0.93, -0.53]	0.32 [-0.07, 0.72]
		Low	52	79.6 (7.2)	61.5	-0.40 [-0.71, -0.09]	$p = 0.11$

$^*p < 0.05$

Table 7.5: Annualized pfGCIPL rates of change in *low* and *high* pfGCIPL groups. β_t: annualized pfGCIPL loss, β_{tLow}: interaction term . Both coefficients are reported as β [95% CI].

7.3 Longitudinal clinical changes

In this analysis we explored the clinical evolution of PD patients by describing longitudinal changes in MoCA, UPDRS-III, and HY scores. The aim was to determine the rate of change of each variable. As a first visual illustration, Figure 7.8 depicts the raw clinical measurements and their evolution for each PD patient in the BHRI dataset.

To obtain a numerical estimate of the evolution, a regression analysis was carried out to model these measurements as a function of follow-up time (β_t). All three models were adjusted for age at baseline (β_{age}) and sex (β_{sex}), and a random intercept was included for each subject.

The results are shown in Table 7.6. The MoCA was estimated to decrease by 0.18 points/year and, as shown in Chapter 6 was highly related with age. On the other hand, both the UPDRS-III (1.28 points/year) and HY score (0.06 points/year) increased longitudinally.

Variable	N	β_{age}	β_{sex}	β_t
MoCA	132	-4.24 [-5.73, -2.76] ($p = 10^{-93}$)	0.48 [-0.95, 1.91] ($p = 0.51$)	-0.18 [-0.37, 0.02] ($p = 0.07$)
UPDRS-III	157	2.84 [-0.32, 6.00] ($p = 0.08$)	1.31 [-1.68, 4.31] ($p = 0.39$)	1.28 [0.63, 1.92] ($p = 10^{-4}$)
Hoehn-Yahr	156	0.15 [-0.02, 0.32] ($p = 0.09$)	-0.11 [-0.28, 0.05] ($p = 0.17$)	0.06 [0.02, 0.09] ($p = 8{\cdot}10^{-4}$)

Table 7.6: Regression estimates for each clinical variable. Each estimated regression coefficient is reported as β [95% CI] (p-value).

Overall, the results highlight the expected worsening in the cognitive and motor function of the patients as the disease progresses. Beyond that fact, it is important to reflect on the variability in measurements illustrated in Figure 7.8. As shown, the evolution of subjects with more than 2 visits does not always follow a monotonic trend (e.g., the condition of a patient can worsen from baseline to the second visit and improve from the second to the third visit). This variability likely reflects the varying nature of patient status and the difficulty of obtaining stable measurements of cognitive and motor function. Without undermining the conclusions obtained on a groupwise level, this variability should be considered when evaluating patients individually.

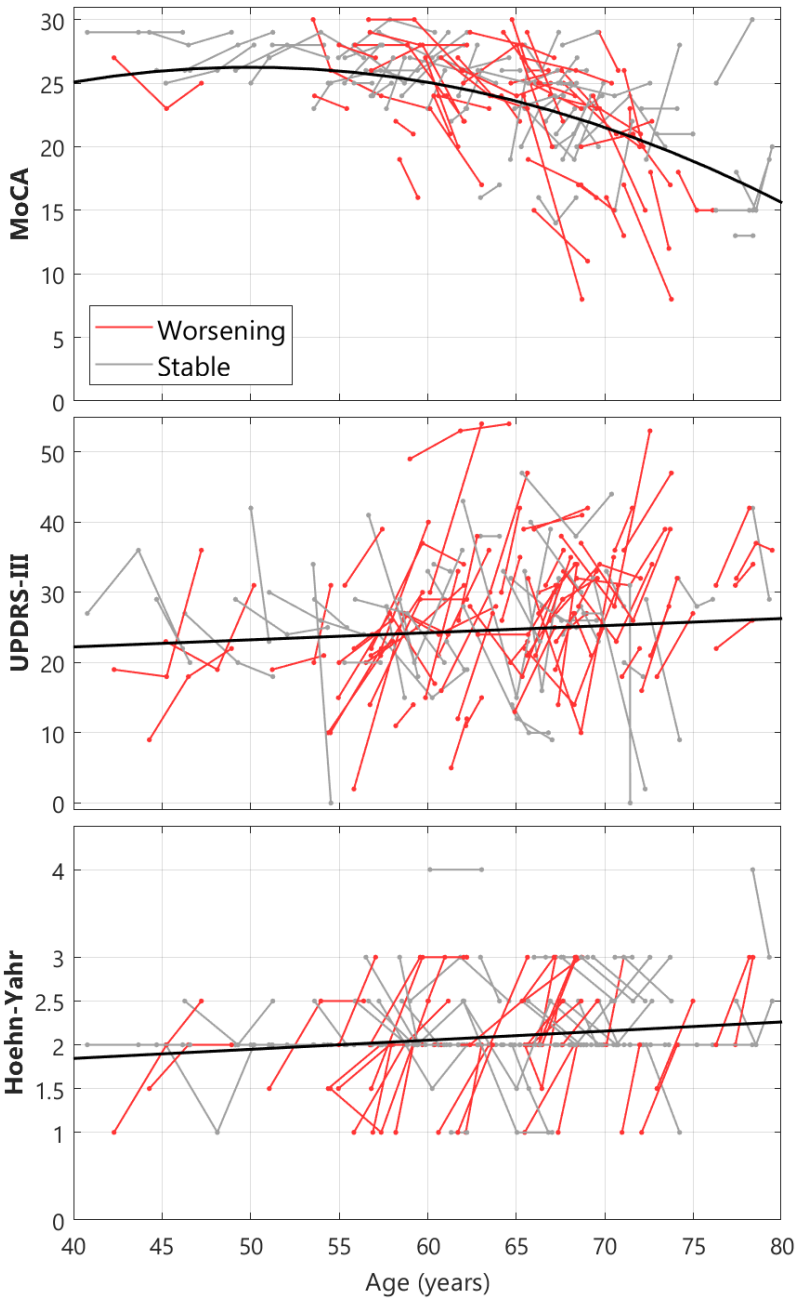


Figure 7.8: Longitudinal evolution of MoCA, UPDRS-III, and HY. Each PD patient is represented as a single line with dots corresponding to visits. Those patients whose last clinical measurement was worse than the that of the initial visit are shown in red (e.g., the MoCA scores at the first and last visits are 27 and 23, respectively). Remaining patients are shown in gray. The black lines depicts the evolution as a function of age.

7.4 Patient monitoring

In the previous analyses we showed that certain retinal features change more prominently in PD. In addition, we confirmed the expected worsening in the clinical status of a patient associated with disease progression.

Building upon those results, in this analysis we looked for a link between changes in both domains. The objective was to determine whether OCT features can be used to monitor disease progression at a patient level under the hypothesis that a change in the severity of a patient may have a corresponding and synchronized change in the retina.

To this end, we first converted visit measurements into changes by subtracting the values between subsequent visits. This procedure was applied both to retinal features and clinical variables (MoCA, UPDRS-III, and HY). Then, we fitted regression models with the clinical change as the outcome and retinal change as a single predictor. This model was fitted to all combinations of the three clinical variables and high-level retinal features. These retinal features included the sectorized thickness measured on three sectors (macula, parafovea, and perifovea) as well as foveal pit morphology features.

$$\Delta_{clinical} = \beta_0 + \beta_\Delta \Delta_{retinalFeature} + \gamma_{sub} \tag{7.4}$$

By adopting this approach, we are in effect testing whether a linear relationship exists between clinical progression and retinal evolution. As in previous analyses, the data included more than one value for several subjects and, hence, a mixed-effects model variant was used.

The results of this analysis are shown in Table 7.7. Overall, changes in retinal features presented weak or no association with clinical changes. In the case of MoCA, a marginally positive relationship was observed with ELM-BM thickness ($R^2 = 3.5\%$). Similarly, a worsening in UPDRS-III was weakly associated with a decrease in outer INL thickness and an increase in ONPL thickness ($R^2 = 5.5\%$). Finally, a worsening in motor status measured by the HY score was also weakly associated with an increase in inner ONPL thickness, a decrease in TRT *pit depth*, a decrease in both TRT and GCIPL *mean slope*, and an increase in GCIPL *rim radius*.

This analysis was carried out using absolute values (i.e., measuring thickness changes in μm). Nevertheless, it could be argued that relative changes may better describe longitudinal evolution (i.e., % change between subsequent visits). Thus, we repeated the previous analysis using percentage values to ensure that the use of absolute values did not preclude the detection of a stronger association. The results of this additional analysis are displayed in Table 7.8 and did not reveal a stronger relationship than those measured in absolute terms.

After the previous analyses, we focused again on patient subgroups with *low* and *high* pfGCIPL and studied if the two groups of patients differed in either the baseline value or evolution of any other clinical variable. This analysis was performed only on BHRI patients, as the AlzEye dataset did not include clinical data. First, the baseline clinical characteristics of the two groups were compared by means of a two-sample t-test. Then, a mixed-effects model regression was used to estimate the longitudinal changes in clinical variables. As in the previous analysis, individual models were first fitted for *low* and *high* groups and then a combined model with an interaction term was employed to formally compute a p-value for the effect under test.

Category	Layer	Feature	MoCA			UPDRS-III			Hoehn-Yahr		
			R^2 (%)	β_Δ	p-value	R^2 (%)	β_Δ	p-value	R^2 (%)	β_Δ	p-value
Thickness	TRT	Macula	1.7	0.16	0.12	0.1	0.10	0.77	0.1	0.01	0.77
		Inner ring	1.1	0.10	0.21	0.3	0.18	0.53	0.0	0.00	0.89
		Outer ring	1.6	0.15	0.14	0.0	0.07	0.85	0.1	0.01	0.69
	GCIPL	Macula	0.1	0.11	0.66	0.6	-0.78	0.35	0.0	0.01	0.84
		Inner ring	0.2	-0.08	0.60	1.3	-0.73	0.18	0.6	-0.03	0.34
		Outer ring	0.3	0.14	0.55	0.1	-0.33	0.67	0.5	0.03	0.41
	INL	Macula	0.0	-0.05	0.90	2.1	-2.44	0.082	0.0	-0.02	0.81
		Inner ring	0.1	-0.05	0.78	0.0	-0.07	0.91	0.1	0.01	0.75
		Outer ring	0.0	-0.01	0.99	4.0	-3.78	0.016*	0.1	-0.03	0.73
	ONPL	Macula	0.1	0.06	0.74	4.9	1.52	0.0079*	2.7	0.06	0.052
		Inner ring	0.3	0.09	0.52	3.5	1.10	0.025*	2.1	0.04	0.0088*
		Outer ring	0.0	0.01	0.93	5.5	1.53	0.0050*	2.6	0.05	0.053
	ELM-BM	Macula	2.8	0.46	0.047*	0.0	0.11	0.89	0.0	-0.00	0.96
		Inner ring	3.5	0.40	0.024*	1.1	0.75	0.21	0.0	0.01	0.84
		Outer ring	1.9	0.38	0.10	0.2	-0.44	0.58	0.0	-0.01	0.84
Foveal pit morphology	TRT	Rim height	1.0	0.09	0.23	0.3	0.17	0.51	0.0	0.00	0.81
		Pit depth	1.5	-0.07	0.14	0.1	-0.04	0.78	3.1	-0.02	0.037*
		Rim radius	0.6	0.0075	0.37	0.1	0.0105	0.72	0.3	0.0010	0.51
		Mean slope	1.4	-1.07	0.16	0.0	-0.19	0.94	3.0	-0.27	0.038*
	GCIPL	Rim height	0.2	-0.08	0.60	0.4	-0.37	0.44	0.4	-0.02	0.44
		Pit depth	2.2	-0.15	0.075	0.6	-0.27	0.35	1.6	-0.02	0.14
		Rim radius	0.1	-0.0030	0.74	0.7	0.0309	0.32	2.8	0.0032	0.047*
		Mean slope	0.7	-1.16	0.33	2.0	-6.76	0.091	3.1	-0.43	0.036*

* $p < 0.05$

Table 7.7: Longitudinal relationship between retinal changes and clinical changes. All regression coefficients (β) were measured as points/μm except for the mean slope (points/°).

Category	Layer	Feature	MoCA			UPDRS-III			Hoehn-Yahr		
			R^2 (%)	β_Δ	p-value	R^2 (%)	β_Δ	p-value	R^2 (%)	β_Δ	p-value
Thickness	TRT	Macula	1.56	2.14	0.14	1.26	-8.76	0.42	0.27	-2.05	0.57
		Inner ring	0.94	1.53	0.25	0.95	-4.54	0.65	0.38	-2.32	0.48
		Outer ring	1.43	1.97	0.15	1.33	-9.02	0.39	0.16	-1.48	0.67
	GCIPL	Macula	0.37	0.59	0.47	1.38	-5.57	0.37	0.19	-0.97	0.64
		Inner ring	0.08	-0.23	0.74	1.91	-6.53	0.21	0.82	-1.81	0.29
		Outer ring	0.46	0.55	0.42	0.95	-2.35	0.65	0.03	0.01	0.99
	INL	Macula	0.15	-0.30	0.65	4.42	-11.20	0.02	0.37	-1.12	0.49
		Inner ring	0.34	-0.24	0.49	2.52	-4.14	0.11	0.22	-0.44	0.61
		Outer ring	0.03	-0.14	0.84	3.64	-10.73	0.04*	0.13	-0.67	0.71
	ONPL	Macula	0.00	-0.00	0.99	3.72	10.86	0.04*	0.85	1.89	0.28
		Inner ring	0.27	0.43	0.54	3.15	9.70	0.07	0.84	1.87	0.29
		Outer ring	0.07	-0.20	0.75	3.81	9.77	0.04*	0.75	1.57	0.31
	ELM-BM	Macula	2.24	1.53	0.07	1.73	-7.45	0.25	0.07	-0.48	0.82
		Inner ring	3.22	1.46	0.03*	1.04	3.00	0.56	0.04	-0.15	0.93
		Outer ring	1.44	1.20	0.15	3.24	-11.86	0.06	0.08	-0.52	0.81
Foveal pit morphology	TRT	Rim height	0.95	1.48	0.25	0.92	-3.98	0.66	0.23	-1.66	0.60
		Pit depth	1.26	-0.29	0.18	1.05	0.96	0.56	2.53	-1.01	0.06
		Rim radius	0.58	0.38	0.36	0.87	0.96	0.76	0.06	0.19	0.85
		Mean slope	1.43	-0.29	0.15	1.11	0.99	0.51	2.21	-0.87	0.08
	GCIPL	Rim height	0.01	-0.08	0.91	1.10	-3.35	0.53	0.83	-1.80	0.29
		Pit depth	1.87	-0.55	0.10	1.03	-1.43	0.57	1.38	-1.16	0.17
		Rim radius	0.03	-0.09	0.85	1.45	3.44	0.34	1.31	1.59	0.18
		Mean slope	1.00	-0.31	0.23	1.79	-2.34	0.24	1.78	-1.02	0.13

* $p < 0.05$

Table 7.8: Longitudinal relationship between relative retinal changes and relative clinical changes. All regression coefficients (β) were measured as %/%.

The results of the comparisons are set out in Table 7.9. The analysis revealed that patients with *low* pfGCIPL had a longer disease duration, lower MoCA, and a more severe motor impairment at baseline. On the other hand, no conclusive evidence was found regarding differences in the rate of change for any of the clinical variables.

Category	Variable	High†	Low†	p-value
Baseline	Disease duration	5.2 (4.1)	7.1 (5.2)	0.012
	MoCA	24.3 (4.0)	21.3 (4.7)	$9{\cdot}10^{-5}$
	UPDRS-III	23.2 (10.8)	24.8 (11.7)	0.37
	Hoehn-Yahr	1.9 (0.6)	2.2 (0.7)	0.0045
Longitudinal	MoCA	-0.12 [-0.34, 0.09]	-0.39 [-0.79, 0.001]	0.28
	UPDRS-III	1.30 [0.60, 2.01]	1.16 [-0.25, 2.57]	0.85
	Hoehn-Yahr	0.08 [0.04, 0.12]	0.03 [-0.04, 0.10]	0.22

† Baseline estimates given as mean (σ). Longitudinal values as β [95% CI] (X/year).

Table 7.9: Differences in clinical variables for *low* and *high* pfGCIPL PD groups.

7.5 Disease prognosis

A very relevant clinical application of any biomarker is disease prognosis. That is, predicting if the condition of a patient will worsen or improve over time based only on the information available at a single visit.

In this last analysis, we investigated whether subject characteristics (age, sex, etc.) and OCT features measured at baseline hold any relationship with cognitive and motor evolution. To this aim, we restricted the analysis to the 46 PD patients from the BHRI MJFF database. These patients had a second visit around 3 years after the first, which allowed us to evaluate the change between the two time points. Based on these two visits, we first computed the change in both MoCA and UPDRS-III for every subject. These changes are illustrated in Figure 7.9 as a function of age. Notably, five patients did not complete follow-up. These patients were all on the older group (> 70 years old in baseline).

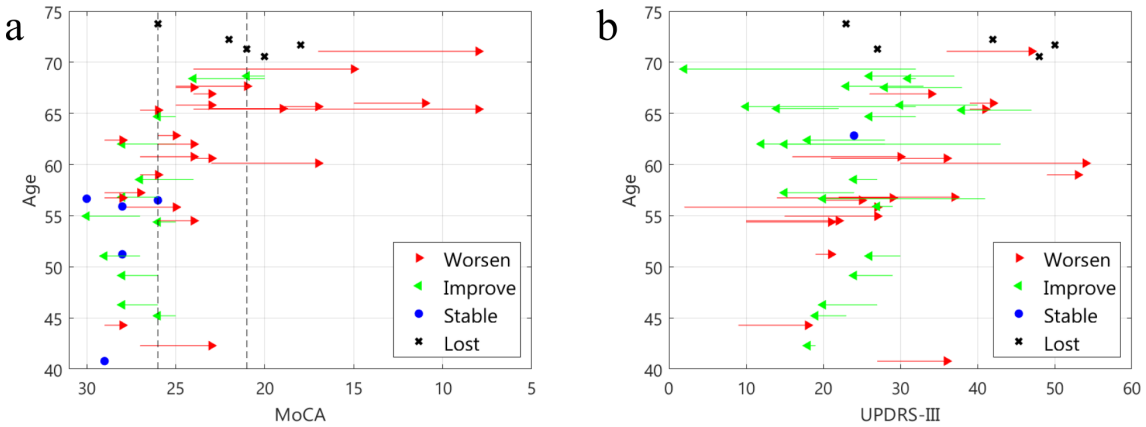


Figure 7.9: 3-year evolution as a function of age. a) MoCA and b) UPDRS-III. Each PD patient is represented horizontally. This graphic only includes patients from the MJFF project.

The computed 3-year change in both MoCA and UPDRS-III was used as the outcome in a regression analysis. In this analysis several features were used as individual predictors including age, sex, disease duration, MoCA, baseline UPDRS-III, and baseline OCT features. The latter

included two types of features: 1) thickness values measured across the whole macula, inner ring, and outer ring for the TRT, GCIPL, INL, ONPL, and ELM-BM layers, and 2) foveal features including the rim height, rim radius, pit depth, and mean slope for both the TRT and the GCIPL.

The results of the statistical analysis can be found in Table 7.10. The evolution in the cognitive status of a patient was only significantly correlated with age (i.e., being older was associated with having a larger decrease in MoCA, $p = 0.028$) and OCT features did not show any clear relationship with the outcome. Similarly, changes in UPDRS-III could not be predicted by OCT features at baseline. Indeed, the only significant predictor was the UPDRS-III value at baseline, which showed a negative relationship with the evolution.

Category	**Layer**	**Feature**	**MoCA**		**UPDRS-III**	
			β	p-value	β	p-value
Reference	-	Age	-0.16	0.028*	-0.45	0.079
		Sex	0.2	0.87	-1.3	0.77
		Disease duration	-0.13	0.43	-0.91	0.10
		MoCA (baseline)	0.28	0.12	0.25	0.69
		UPDRS-III (baseline)	-0.064	0.25	-0.7	0.00008*
Thickness	TRT	Macula	-0.052	0.30	0.23	0.19
		Inner ring	-0.027	0.53	0.25	0.096
		Outer ring	-0.057	0.27	0.19	0.27
	GCIPL	Macula	0.03	0.76	0.22	0.53
		Inner ring	0.04	0.60	0.31	0.24
		Outer ring	0.025	0.80	0.13	0.71
	INL	Macula	0.13	0.64	0.94	0.33
		Inner ring	0.1	0.57	0.78	0.21
		Outer ring	0.12	0.68	0.64	0.53
	ONPL	Macula	-0.16	0.12	0.48	0.20
		Inner ring	-0.18	0.056	0.41	0.21
		Outer ring	-0.15	0.17	0.45	0.22
	ELM-BM	Macula	-0.43	0.11	0.56	0.55
		Inner ring	-0.45	0.082	-0.22	0.80
		Outer ring	-0.38	0.14	0.73	0.41
Foveal pit morphology	TRT	Rim height	-0.019	0.65	0.21	0.15
		Pit depth	-0.00012	0.99	0.053	0.56
		Rim radius	0.00089	0.89	-0.02	0.35
		Mean slope	-0.15	0.76	2	0.23
	GCIPL	Rim height	0.068	0.38	0.26	0.33
		Pit depth	0.079	0.25	0.31	0.19
		Rim radius	-0.0015	0.80	-0.023	0.27
		Mean slope	0.66	0.47	5.3	0.085

* $p < 0.05$

Table 7.10: Regression results for 3-year change prognosis.

7.6 Discussion

In this chapter, we have investigated longitudinal changes in the retina and their link with disease progression. In a first analysis, we demonstrated that the thickness of both the TRT and GCIPL

decreases with time at a higher rate in PD patients. For instance, average GCIPL thickness was measured to decrease at 0.21 μm/year and 0.46 μm/year in controls and patients, respectively. This difference corresponds to a loss of around twice as fast in patients, which highlights the importance of the effect.

Notably, the analysis confirmed that the thinning effect is not limited to patients and is also present in healthy controls, a finding that is in line with the cross-sectional results presented in Chapter 5. Together, both analyses suggest that two factors may be mediating the retinal thinning: normal aging and PD-related neurodegeneration.

Importantly, we were able to replicate the observed GCIPL thinning pattern in a separate cohort. Considering the notable dissimilarities between the two datasets, this successful replication supports the idea that the effect is generalizable to a broader population. Nevertheless, important differences between the results in both datasets were also observed. First, thinning estimates in the control group were higher for AlzEye. This could be explained by the nature of the subjects included in AlzEye, who were older on average and eminently presented a considerably worse state of health, as they were imaged as part of regular clinical practice. Although we applied strict data inclusion criteria to mitigate this potential confounder, some of the AlzEye subjects finally included in the control group may still have conditions affecting the retinal structure.

In line with our results, previous studies investigating longitudinal changes in PD all found a thinning effect of the TRT [142–144]. As for the GCIPL, a study using a smaller subset of BHRI dataset found similar GCIPL changes [176]. It is important to note that all previous work relied on either average measurements of the macula or the ETDRS sectorization. In contrast, we employed a more detailed ring sectorization that enabled us to determine that the retinal area most highly affected by PD was the GCIPL in the ring with 0.5 mm to 2 mm radii. This area mostly overlaps with the commonly studied parafoveal ring, which has been identified as an important indicator of changes in PD [176].

On another note, we found an ONPL thinning effect in the BHRI dataset that we could not replicate in AlzEye. A possible explanation for this difference could be the lower statistical power of AlzEye. In fact, although the total number of individuals included in AlzEye was greater, the average follow-up period was relatively short ($\leq$ 2 years), a fact that may have hindered our capacity to accurately detect longitudinal effects. Nonetheless, to the best of our knowledge no previous longitudinal study has reported changes in the ONPL. Likewise, cross-sectional studies have largely focused on inner retinal layers. Further research should therefore explore the ONPL to confirm our findings.

In addition to the thinning effect, we also explored changes in the foveal pit morphology. This idea was largely motivated by previous works suggesting a foveal remodeling effect in PD [74, 151, 154] and the lack of longitudinal studies investigating this hypothesis. Overall, the results from both datasets showed a clear reduction of the *rim height* in PD patients. This effect was present for both the TRT and GCIPL, and it is likely a consequence of the predominant thinning effect of both layers in the inner ring of the macula. In addition, we found a slight reduction of the GCIPL *pit depth* in BHRI, a result that is consistent with the reduction in *rim height*. However, this finding was not evident in AlzEye. Again, the noisier nature of AlzEye as well as the lower statistical power could explain these discrepancies.

Of perhaps more importance than the *rim height*—which conveys similar information to thickness features—is the analysis of features that describe other aspects of the fovea. In this regard, we did not find any changes in the *mean slope*, and only observed a slight increase of the GCIPL *rim radius* in the BHRI dataset. As a conclusion, the results only support a remodeling of the fovea in the vertical axis directly related to a thinning effect. It should be mentioned, however, that slope and width measurements are inherently less reproducible and, as discussed in Section 4.3, detecting changes in these features is more challenging.

In the clinical domain, as expected, the condition of the PD patients worsened with time in

terms of cognition (assessed by MoCA), and motor symptoms (measured by UPDRS-III and HY scores). In a subsequent analysis we thus sought links between these clinical changes and previously described retinal changes in PD.

As a first hypothesis we tested whether a change in a certain clinical variable could be predicted by a change in retinal features. This analysis only revealed a few minor associations (maximum R^2 of 5.5%) with no clear link between anatomical and clinical realms. A possible explanation for these results could be that changes in the retina do not parallel disease progression. This idea has recently been posited [207] and would suggest that retinal changes may precede or follow clinical changes without a clear association generalizable to all patients.

An alternative explanation is measurement error. When measuring a relationship between changes in two variables, measurement error can easily distort that relationship and render an association zero. Critically, the test-retest repeatability of individual OCT thickness measurements is in the order of a few μm. This value is considered very high for absolute measurements as it represents a small % of the value to be measured. However, longitudinal changes are much smaller (typically below $1\,\mu m$), and therefore, measurement error has a greater impact on these measurements.

To quantitatively appraise the latter idea, we performed a simulation to estimate the COV of thickness changes. The COV is defined as the standard deviation to mean ratio and is used to determine the dispersion of a measurement. Oberwahrenbrock et al. calculated that GCIPL thickness measurements with Spectralis follow-up function have at best a COV of 0.36% [196]. Based on the values presented in that work we simulated the actual COV for longitudinal GCIPL thickness differences. The results are shown in Figure 7.10 and demonstrate that the effective COV of thickness changes is much higher than that of single measurements.

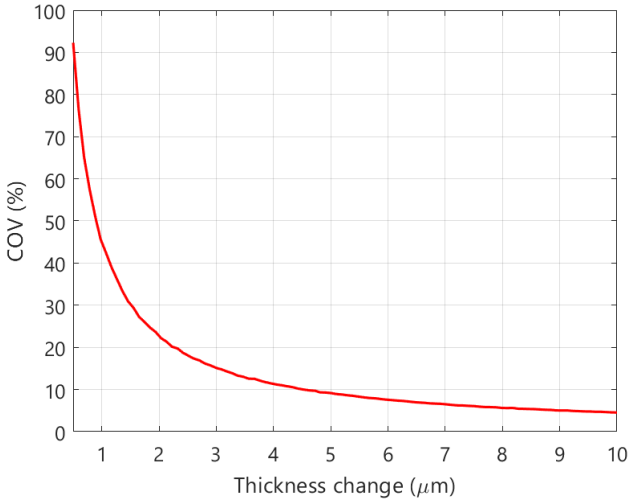


Figure 7.10: Simulated coefficient of variation for GCIPL thickness changes. The plot shows the real coefficient of variation (%) as a function of the thickness change intended to be measured.

Altogether, it can be concluded that the repeatability of OCT measurements is likely insufficient to measure very small changes accurately at a patient level. This, in turn, limits the potential of OCT to monitor the disease for a single patient when the expected changes are small.

Considering the challenge of directly correlating retinal and clinical changes, we performed a group-level analysis by creating subgroups of subjects with *low* and *high* baseline pfGCIPL thickness. As the principal conclusion, PD subjects in the *high* group were younger on average and had less deteriorated cognition and motor function. Annualized pfGCIPL thickness loss was also higher in this group. Importantly, this effect was not present in the control group, which supports the hypothesis of an effect circumscribed to patients.

These results suggest that the effect of PD on the retina is not homogeneous across patients and varies at different stages of the disease. In particular, retinal neurodegeneration may occur primarily at early stages when patients have relatively mild clinical symptoms. As the disease progresses, the retinal thinning rate may stabilize and clinical symptoms may worsen significantly. Interestingly, the idea of an early accentuated GCIPL loss followed by stabilization differs notably from the results presented in Chapter 5, which showed an accelerated thinning as a function of age in healthy individuals. This could highlight the presence of two effects: a slowly progressing thinning due to age, and an early neurodegeneration caused by PD. To confirm this hypothesis, however, more longitudinal studies are needed with longer follow-up times and more visits.

Finally, these analyses are not without their limitations. First, we did not consider the effect of axial length in any of the analyses. Although this fact is known to have an influence on OCT measurements, its effect is not likely to critically impact longitudinal measurements as the eye under comparison is the same. On the other hand, we chose not to adjust p-values for multiple comparisons because of the exploratory nature of the analyses. Given the limited published literature on the topic, rather than confirming pre-existing hypotheses, we approached the matter with the aim of uncovering new ideas to be more formally tested by subsequent studies.

8 | Conclusions

This last chapter summarizes the main contributions and findings of this doctoral study. First, Section 8.1 lists the developments and analyses in which we have extended previous work. Then, the main research conclusions are described in Section 8.2. Finally, Sections 8.3 and 8.4 list the publications derived from the thesis and future research lines, respectively.

8.1 Main contributions

The purpose of this research work was twofold: 1) to improve existing methods for OCT image analysis and build a robust feature extraction pipeline and 2) to explore the potential of the computed OCT features as a biomarker for PD. The first set of contributions refer to the former and involve the technical developments carried out to extend OCT analysis methods:

- *Automatic foveal location algorithm*: we developed a new foveal location algorithm based on thickness maps and a flooding procedure. The algorithm performed well, and unlike existing deep-learning models, could be directly generalized to other scanning devices.
- *Image anomaly detection algorithm*: we trained a classification model to flag OCT image volumes to be excluded in the presence of imaging artifacts, poor image contrast, and ocular pathologies affecting retinal measurements. The algorithm reported an accuracy of above 90% and can be used as a first screening step to detect problems in a dataset.
- *RETIMAT toolbox*: we created the first general purpose open-source toolbox for OCT image analysis. The software includes functionalities for multiple file format reading, generating reports, preprocessing, and feature extraction. RETIMAT has already been presented to the community in conferences [187] and is freely accessible online.
- *Large dataset processing*: we used RETIMAT and previously developed methods to process all macular images from the BHRI and AlzEye datasets. The computed retinal features provided the basis for extensive research investigating multiple clinical conditions, part of which has been published [161, 178, 208].

In the second part of the study, the OCT features derived previously were used to investigate the retinal structure in both healthy subjects and PD patients.

- *Healthy population study*: we investigated the effect of age and sex on both retinal thickness and foveal pit morphology. Our analysis was more spatially detailed than previously published studies, in that we used small macular sectors and multiple angular directions.
- *Normative database*: the retinal measurements obtained in the healthy population study were publicly released, and thus can be used as a reference population value (N=444).
- *Diagnostic model evaluation*: we evaluated the performance of multiple retinal features for PD diagnosis. We extended previous work by exploring both conventional and novel retinal features.

- *Severity assessment model evaluation*: as with the diagnostic models, we made use of a comprehensive OCT feature set to investigate the relationship between retinal features and cognitive function and motor impairment.
- *Longitudinal analysis*: we made use of two of the largest longitudinal datasets compiled to date to investigate retinal and clinical changes associated with PD progression.

8.2 Research conclusions

The research conclusions obtained in the course of this research project are as follows:

- *Effect of age:* most retinal layers become thinner with age, and the layer most affected is the GCIPL. The thinning effect is homogenous when measured in % terms. The height of the fovea changes with age in all angular directions.
- *Sex differences*: the male retina is thicker especially for inner sectors. The male fovea is also deeper and sharper across every radial direction.
- *PD diagnosis*: our findings indicate that an accurate PD diagnosis from explored OCT features is not possible. As illustrated by the literature and confirmed in this work, the existing differences between control and PD groups are not large enough for it to be considered for clinical use. Using advanced features does not significantly improve the performance.
- *PD severity assessment*: the retina holds some information about disease severity. Inner GCIPL thickness is associated with cognitive status. However, the strength of the association does not appear to be sufficient for reliable clinical application at an individual patient level.
- *Longitudinal retinal progression*: the retina becomes thinner with age in both healthy and PD individuals. This thinning rate is notably faster in PD patients. The biggest differences were found for inner GCIPL and foveal pit rim height.
- *PD monitoring*: using OCT to monitor the disease at a patient level is not supported by the evidence. Even if a relationship existed, the longitudinal retinal changes are smaller than the measurement error and challenging to determine accurately. On a subgroup level, however, the results point towards a faster decline in GCIPL in early disease stages followed by stabilization and then a clinical worsening.

8.3 Publications

Research articles

- Foveal Pit Morphology Characterization: A Quantitative Analysis of the Key Methodological Steps. *Entropy*. 2021.
  Romero-Bascones D, Barrenechea M, Murueta-Goyena A, Galdós M, Gómez-Esteban JC, Gabilondo I, Ayala U.

- Spatial characterization of the effect of age and sex on macular layer thicknesses and foveal pit morphology. *PLoS ONE*. 2022.
  Romero-Bascones D, Ayala U, Alberdi A, Erramuzpe A, Galdós M, Gómez-Esteban JC, Murueta-Goyena A, Teijeira S, Gabilondo I, Barrenechea M.

- Association of retinal thinning rate with the progression of cognitive decline in Parkinson's disease. *Submitted*.

Murueta-Goyena A, **Romero-Bascones D**, Teijeira-Portas S, Aritz Urcola J, Ruiz-Martínez J, Petzold A, Wagner SK, Keane PA, Ayala U, Barrenechea M, Gómez-Esteban JC, Gabilondo I.

Conferences

- Caracterización de la morfología foveal: parametrización, diferencias de sexo y efectos de la edad. *CASEIB* 2020.
  Romero-Bascones D, Gabilondo I, Barrenechea M, Ayala U.

- RETIMAT: un Toolbox de MATLAB para el análisis de imágenes OCT de retina. *CASEIB* 2022.
  Romero-Bascones D, Gabilondo I, Barrenechea M, Ayala U.

- Evaluación del potencial de las imágenes OCT maculares para el diagnóstico de la enfermedad de Parkinson. *CASEIB* 2022.
  Romero-Bascones D, Gabilondo I, Barrenechea M, Ayala U.

- RETIMAT: an open-source software for retinal OCT image analysis. *ARVO* 2023.
  Romero-Bascones D, Murueta-Goyena A, Wagner SK, Struyven RR, Williamson DJ, Keane PA, Barrenechea M, Gabilondo I, Ayala U.

Related publications

Research and conference papers to which this research work has contributed on a smaller scale and which are not described in the thesis.

- Foveal Remodeling of Retinal Microvasculature in Parkinson's Disease. *Front Neurosci.* 2021.
  Murueta-Goyena A, Barrenechea M, Erramuzpe A, Teijeira-Portas S, Pengo M, Ayala U, **Romero-Bascones D**, Acera M, Del Pino R, Gómez-Esteban JC, Gabilondo I.

- Association Between Retinal Features From Multimodal Imaging and Schizophrenia. *JAMA Psychiatry.* 2023.
  Wagner SK, Cortina-Borja M, Silverstein SM, Zhou Y, **Romero-Bascones D**, Struyven RR, Trucco E, Mookiah MRK, MacGillivray T, Hogg S, Liu T, Williamson DJ, Pontikos N, Patel PJ, Balaskas K, Alexander DC, Stuart KV, Khawaja AP, Denniston AK, Rahi JS, Petzold A, Keane PA.

- Retinal thickness as a biomarker of cognitive impairment in manifest Huntington's disease. *Journal of Neurology.* 2023.
  Murueta-Goyena A, Del Pino R, Acera M, Teijeira-Portas S, **Romero-Bascones D**, Ayala U, Fernández-Valle T, Tijero B, Gabilondo I, Gómez-Esteban JC.

- Retinal optical coherence tomography features associated with incident and prevalent Parkinson disease. *Neurology.* 2023.
  Wagner SK, **Romero-Bascones D**, Cortina-Borja M, Williamson DJ, Struyven RR, Zhou Y, Salil Patel, Weil RS, Antoniades CA, Topol EJ, Korot E, Foster PJ, Balaskas K, Ayala U, Barrenechea M, Gabilondo I, Schapira A, Khawaja AP, Patel PJ, Rahi JS, Denniston AK, Petzold A, Keane PA.

- Deep learning applied to retinal OCT images to differentiate Parkinson's disease patients from healthy controls. *MDS* 2022.
  Barrenechea M, **Romero-Bascones D**, Murueta-Goyena A, Gomez-Esteban JC, Alberdi A, Erramuzpe A, Ayala U, Gabilondo I.

- Deep-learning fusion of OCT imaging and traditional risk factors to improve dementia detection in AlzEye. *ARVO* 2023.
  Struyven RR, Williamson DJ, Wagner SK, **Romero-Bascones D**, Zhou Y, Liu T, Wu Y, Balaskas K, Cortina Borja M, Rahi J, Petzold A, Lee AY, Lee CS, Denniston AK, Alexander D, Keane PA.

- Bidirectional retinal oculomics in Parkinson's disease: A cross-sectional analysis of two cohorts. *ARVO* 2023.
  Wagner SK, **Romero-Bascones D**, Cortina-Borja M, Williamson DJ, Struyven RR, Zhou Y, Foster P, Ayala U, Barrenechea M, Gabilondo I, Khawaja AP, Patel PJ, Rahi JS, Denniston AK, Petzold A, Keane PA.

- Retinal biomarkers for systemic diseases: an oculome-wide association study in 164,784 individuals. *ARVO* 2023.
  Liu T, Wagner, Wagner SK, Struyven RR, Zhou Y, Williamson DJ, **Romero-Bascones D**, Gende M, Pontikos N, Patel PJ, Cortina-Borja M, Rahi JS, Petzold A, Khawaja AP. Denniston AK, Balaskas K, Keane PA.

8.4 Future lines of work

Building upon the findings and contributions of this work, four main avenues for future research are identified:

- *Full OCT quality control*: the quality control model developed in this work approaches all quality problems combined. However, an ideal solution would take an OCT volume as input and first detect quality problems such cropping or shades at an A-scan level. In a second step, entire B-scans could be analyzed for poor contrast issues. Then, problems affecting the entire volume such as motion artifacts and incorrect alignment would be detected. Finally, disease specific models would then be applied to identify ocular pathologies affecting the retina. This framework would be able to handle every aspect affecting OCT at a level not presently achieved. One drawback is that developing such a method would require a large dataset with highly granular labeling as well as multiple methods to detect each problem.

- *Extension to other neurodegenerative diseases*: the methods developed in this work have been mainly applied to PD only. Nevertheless, there is evidence of retinal changes in other neurodegenerative diseases such as AD, MS, and Huntington's disease. Therefore, a natural continuation of the research would be applying similar methods to these diseases to determine specific changes in each instance.

- *Extension of the longitudinal analysis*: as shown in this work, longitudinal datasets have helped uncover an accelerated thinning rate of certain retinal layers in PD. However, both our work and that of previous researchers has been limited in terms of the number of visits and the follow-up period. Thus, a clear improvement could be attained by collecting a larger database. Increasing the number of visits would help reduce the inherent variability of OCT measurements and estimate longitudinal changes more accurately at a subject level.

- *Multimodal imaging in PD*: a key point not addressed in this study is the use of more than one imaging modality to investigate PD. First, research has shown that OCTA images can provide valuable information about the disease. It would therefore be of great interest to combine both OCT and OCTA techniques to understand simultaneously structural and vascular changes in PD. Second, PD is eminently a neurological disease and thus the main research focus is centered around neuroimaging. This opens up an interesting line of research linking OCT findings with measurements from MRI under the assumption that this could help establish temporal relationships between changes in the retina and the brain. The biggest challenge in this case lies in the difficulty of acquiring such a dataset.

In addition to the aforementioned research lines, a largely overlooked line work involves the development of software for OCT image analysis. Indeed, despite the remarkable need for automated tools in the field, existing open software solutions are quite limited and not widely used as a standard. During this work we developed RETIMAT, which proved useful for facilitating research. However, MATLAB has limited traction as a programming language and developing an alternative in language like python would reach a larger user group.

Bibliography

[1] K. Nishino and S. K. Nayar, "Extraction of Visual Information from Images of Eyes", in *Passive Eye Monitoring: Algorithms, Applications and Experiments*, R. I. Hammoud, Ed. Springer Berlin, Heidelberg, 2008, ch. 7, p. 165. Available: https://doi.org/10.1007/978-3-540-75412-1

[2] S. Anstis, "A chart demonstrating variations in acuity with retinal position", *Vision Research*, vol. 14, no. 7, pp. 589–592, 1974. Available: https://doi.org/10.1016/0042-6989(74)90049-2

[3] H. Kolb, "Photoreceptors", in *Webvision. The Organization of the Retina and Visual System*, 2013. Available: https://webvision.med.utah.edu/book/part-ii-anatomy-and-physiology-of-the-retina/photoreceptors

[4] C. W. Oyster, *The Human Eye: Structure and Function*. Sunderland, MA: Sinauer, 1999.

[5] H. Kolb, "How the Retina Works", *American Scientist*, vol. 91, no. 1, pp. 28–35, 2003. Available: https://www.americanscientist.org/article/how-the-retina-works

[6] H. Kolb, "Part XIII: Facts and Figures concerning the human retina by Helga Kolb", in *Webvision. The Organization of the Retina and Visual System*, 2011. Available: https://webvision.med.utah.edu/book/part-xiii-facts-and-figures-concerning-the-human-retina

[7] B. M. Koeppen and B. A. Stanton, Eds., *Berne & Levy Physiology*. Elsevier, 2017.

[8] S. Aumann *et al.*, "Optical Coherence Tomography (OCT): Principle and Technical realization", in *High Resolution Imaging in Microscopy and Ophthalmology*, J. F. Bille, Ed. Springer Nature Switzerland AG, 2019, ch. 3, pp. 59–86. Available: https://doi.org/10.1007/978-3-030-16638-0_3

[9] D. Huang *et al.*, "Optical coherence tomography", *Science.*, vol. 254, no. 5035, pp. 1178–1181, 1991. Available: https://doi.org/10.1126/science.1957169

[10] W. Drexler and J. G. Fujimoto, "State-of-the-art retinal optical coherence tomography", *Progress in Retinal and Eye Research*, vol. 27, no. 1, pp. 45–88, 2008. Available: https://doi.org/10.1016/j.preteyeres.2007.07.005

[11] U. Morgner *et al.*, "Spectroscopic optical coherence tomography", *Optics Letters*, vol. 25, no. 2, p. 111, 2000. Available: https://doi.org/10.1364/OL.25.000111

[12] M. Wojtkowski *et al.*, "In vivo human retinal imaging by Fourier domain optical coherence tomography", *Journal of Biomedical Optics*, vol. 7, no. 3, p. 457, 2002. Available: https://doi.org/10.1117/1.1482379

[13] A. N. Kuo *et al.*, "Correction of ocular shape in retinal optical coherence tomography and effect on current clinical measures", *American Journal of Ophthalmology*, vol. 156, no. 2, pp. 304–311, 2013. Available: https://doi.org/10.1016/j.ajo.2013.03.012

[14] J. Straub and M. Steidle, "Estimating the shape of the human eye using widefield optical coherence tomography (OCT)", in *Proc. SPIE 10685, Biophotonics: Photonic Solutions for Better Health Care VI, 106851V*, 2018. Available: https://doi.org/10.1117/12.2306604

[15] R. P. McNabb *et al.*, "Wide-field whole eye OCT system with demonstration of quantitative retinal curvature estimation", *Biomedical Optics Express*, vol. 10, no. 1, p. 338, 2019. Available: https://doi.org/10.1364/BOE.10.000338

[16] J. M. Schmitt *et al.*, "Speckle in Optical Coherence Tomography", *Journal of Biomedical Optics*, vol. 4, no. 1, p. 95, 1999. Available: https://doi.org/10.1117/1.429925

[17] M. Li *et al.*, "Statistical model for OCT image denoising", *Biomedical Optics Express*, vol. 8, no. 9, p. 3903, 2017. Available: https://doi.org/10.1364/BOE.8.003903

[18] Z. Amini and H. Rabbani, "Optical coherence tomography image denoising using Gaussianization transform", *Journal of Biomedical Optics*, vol. 22, no. 08, p. 1, 2017. Available: https://doi.org/10.1117/1.JBO.22.8.086011

[19] S. Apostolopoulos *et al.*, "Automatically Enhanced OCT Scans of the Retina: A proof of concept study", *Scientific Reports*, vol. 10, no. 1, p. 7819, 2020. Available: https://doi.org/10.1038/s41598-020-64724-8

[20] Z. Dong *et al.*, "Optical coherence tomography image denoising using a generative adversarial network with speckle modulation", *Journal of Biophotonics*, vol. 13, no. 4, 2020. Available: https://doi.org/10.1002/jbio.201960135

[21] M. Niemeijer *et al.*, "Registration of 3D spectral OCT volumes using 3D SIFT feature point matching", *Medical Imaging 2009: Image Processing*, vol. 7259, p. 72591I, 2009. Available: https://doi.org/10.1117/12.811906

[22] L. Pan *et al.*, "Segmentation Guided Registration for 3D Spectral-Domain Optical Coherence Tomography Images", *IEEE Access*, vol. 7, pp. 138 833–138 845, 2019. Available: https://doi.org/10.1109/ACCESS.2019.2943172

[23] T. Fiore *et al.*, "Repeatability of Retinal Macular Thickness Measurements in Healthy Subjects and Diabetic Patients with Clinically Significant Macular Edema: Evaluation of the Follow-Up System of Spectralis Optical Coherence Tomography", *Ophthalmologica*, vol. 234, no. 3, pp. 167–171, 2015. Available: https://doi.org/10.1159/000435841

[24] M. M. Khansari *et al.*, "Automated Deformation-Based Analysis of 3D Optical Coherence Tomography in Diabetic Retinopathy", *IEEE Transactions on Medical Imaging*, vol. 39, no. 1, pp. 236–245, 2020. Available: https://doi.org/10.1109/TMI.2019.2924452

[25] L. Pan *et al.*, "OCTRexpert: A Feature-Based 3D Registration Method for Retinal OCT Images", *IEEE Transactions on Image Processing*, vol. 29, pp. 3885–3897, 2020. Available: https://doi.org/10.1109/TIP.2020.2967589

[26] M. Garvin *et al.*, "Automated 3-D Intraretinal Layer Segmentation of Macular Spectral-Domain Optical Coherence Tomography Images", *IEEE Transactions on Medical Imaging*, vol. 28, no. 9, pp. 1436–1447, 2009. Available: https://doi.org/10.1109/TMI.2009.2016958

[27] S. J. Chiu *et al.*, "Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation", *Optics Express*, vol. 18, no. 18, p. 19413, 2010. Available: https://doi.org/10.1364/OE.18.019413

[28] Q. Yang *et al.*, "Automated layer segmentation of macular OCT images using dual-scale gradient information", *Optics Express*, vol. 18, no. 20, p. 21293, 2010. Available: https://doi.org/10.1364/OE.18.021293

[29] A. Lang *et al.*, "Retinal layer segmentation of macular OCT images using boundary classification", *Biomedical Optics Express*, vol. 4, no. 7, pp. 1133–1152, 2013. Available: https://doi.org/10.1364/BOE.4.001133

[30] A. G. Roy *et al.*, "ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks", *Biomedical Optics Express*, vol. 8, no. 8, p. 3627, 2017. Available: https://doi.org/10.1364/BOE.8.003627

[31] M. Pekala *et al.*, "Deep learning based retinal OCT segmentation", *Computers in Biology and Medicine*, vol. 114, p. 103445, 2019. Available: https://doi.org/10.1016/j.compbiomed.2019.103445

[32] J. Kugelman *et al.*, "A comparison of deep learning U-Net architectures for posterior segment OCT retinal layer segmentation", *Scientific Reports*, vol. 12, no. 1, p. 14888, 2022. Available: https://doi.org/10.1038/s41598-022-18646-2

[33] N. Man *et al.*, "Multi-layer segmentation of retina OCT images via advanced U-net architecture", *Neurocomputing*, vol. 515, pp. 185–200, 2023. Available: https://doi.org/10.1016/j.neucom.2022.10.001

[34] O. Ronneberger *et al.*, "U-Net: Convolutional Networks for Biomedical Image Segmentation", *arXiv*, 2015. Available: https://doi.org/10.48550/arXiv.1505.04597

[35] J. Tian *et al.*, "Performance evaluation of automated segmentation software on optical coherence tomography volume data", *Journal of Biophotonics*, vol. 9, no. 5, pp. 478–489, 2016. Available: https://doi.org/10.1002/jbio.201500239

[36] "HEYEX 2 Image Management and Device Integration Platform". Available: https://business-lounge.heidelbergengineering.com/ca/en/products/heidelberg-eye-explorer/heyex-2/#product-details

[37] P. A. Keane *et al.*, "Optical Coherence Tomography in the UK Biobank Study – Rapid Automated Analysis of Retinal Thickness for Large Population-Based Studies", *PLOS ONE*, vol. 11, no. 10, p. e0164095, 2016. Available: https://doi.org/10.1371/journal.pone.0164095

[38] J. Ho *et al.*, "Clinical assessment of mirror artifacts in spectral-domain optical coherence tomography." *Investigative ophthalmology and visual science*, vol. 51, no. 7, pp. 3714–3720, 2010. Available: https://doi.org/10.1167/iovs.09-4057

[39] J. Wang *et al.*, "Deep learning for quality assessment of retinal OCT images", *Biomedical Optics Express*, vol. 10, no. 12, p. 6057, 2019. Available: https://doi.org/10.1364/BOE.10.006057

[40] J. Kauer-Bonin *et al.*, "Modular deep neural networks for automatic quality control of retinal optical coherence tomography scans", *Computers in Biology and Medicine*, vol. 141, p. 104822, 2022. Available: https://doi.org/10.1016/j.compbiomed.2021.104822

[41] A. Chrysou *et al.*, "Retinal layers in Parkinson's disease: A meta-analysis of spectral-domain optical coherence tomography studies", *Parkinsonism and Related Disorders*, vol. 64, pp. 40–49, 2019. Available: https://doi.org/10.1016/j.parkreldis.2019.04.023

[42] S. Sadigh *et al.*, "Abnormal Thickening as well as Thinning of the Photoreceptor Layer in Intermediate Age-Related Macular Degeneration", *Investigative Opthalmology and Visual Science*, vol. 54, no. 3, p. 1603, 2013. Available: https://doi.org/10.1167/iovs.12-11286

[43] A. Samii *et al.*, "Parkinson's disease", *The Lancet*, vol. 363, no. 9423, pp. 1783–1793, 2004. Available: https://doi.org/10.1016/S0140-6736(04)16305-8

[44] B. R. Bloem *et al.*, "Parkinson's disease", *The Lancet*, vol. 397, no. 10291, pp. 2284–2303, 2021. Available: https://doi.org/10.1016/S0140-6736(21)00218-X

[45] R. S. Weil *et al.*, "Visual dysfunction in Parkinson's disease", *Brain*, vol. 139, no. 11, pp. 2827–2843, 2016. Available: https://doi.org/10.1093/brain/aww175

[46] L. J. Dommershuijsen *et al.*, "Life expectancy of parkinsonism patients in the general population", *Parkinsonism and Related Disorders*, vol. 77, pp. 94–99, 2020. Available: https://doi.org/10.1016/j.parkreldis.2020.06.018

[47] M. G. Erkkinen *et al.*, "Clinical Neurology and Epidemiology of the Major Neurodegenerative Diseases", *Cold Spring Harbor Perspectives in Biology*, vol. 10, no. 4, p. a033118, 2018. Available: https://doi.org/10.1101/cshperspect.a033118

[48] E. Ray Dorsey *et al.*, "Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016", *The Lancet Neurology*, vol. 17, no. 11, pp. 939–953, 2018. Available: https://doi.org/10.1016/S1474-4422(18)30295-3

[49] E. R. Dorsey *et al.*, "The emerging evidence of the Parkinson pandemic", *Journal of Parkinson's Disease*, vol. 8, no. s1, pp. S3–S8, 2018. Available: https://doi.org/10.3233/jpd-181474

[50] A. David Smith and J. Paul Bolam, "The neural network of the basal ganglia as revealed by the study of synaptic connections of identified neurones", *Trends in Neurosciences*, vol. 13, no. 7, pp. 259–265, 1990. Available: https://doi.org/10.1016/0166-2236(90)90106-K

[51] The Parkinson Study Group, "Levodopa and the Progression of Parkinson's Disease", *New England Journal of Medicine*, vol. 351, no. 24, pp. 2498–2508, 2004. Available: https://doi.org/10.1056/NEJMoa033447

[52] A. L. Benabid *et al.*, "Deep brain stimulation of the subthalamic nucleus for the treatment of Parkinson's disease", *The Lancet Neurology*, vol. 8, no. 1, pp. 67–81, 2009. Available: https://doi.org/10.1016/S1474-4422(08)70291-6

[53] L. Marsili *et al.*, "Diagnostic criteria for Parkinson's disease: From James Parkinson to the concept of prodromal disease", *Frontiers in Neurology*, vol. 9, 2018. Available: https://doi.org/10.3389/fneur.2018.00156

[54] N. Pyatigorskaya *et al.*, "A review of the use of magnetic resonance imaging in Parkinson's disease", *Therapeutic Advances in Neurological Disorders*, vol. 7, no. 4, pp. 206–220, 2014. Available: https://doi.org/10.1177/1756285613511507

[55] U. Saeed *et al.*, "Imaging biomarkers in Parkinson's disease and Parkinsonian syndromes: Current and emerging concepts", *Translational Neurodegeneration*, vol. 6, no. 1, pp. 1–25, 2017. Available: https://doi.org/10.1186/s40035-017-0076-6

[56] N. S. Bidesi *et al.*, "The role of neuroimaging in Parkinson's disease", *Journal of Neurochemistry*, vol. 159, no. 4, pp. 660–689, 2021. Available: https://doi.org/10.1111%2Fjnc.15516

[57] J.-Y. Lee *et al.*, "Multimodal brain and retinal imaging of dopaminergic degeneration in Parkinson disease", *Nature Reviews Neurology*, vol. 18, no. 4, pp. 203–220, 2022. Available: https://doi.org/10.1038/s41582-022-00618-9

[58] I. Bodis-Wollner and M. D. Yahr, "Measurements of Visual Evoked Potentials in Parkinson's Disease", *Brain*, vol. 101, no. 4, pp. 661–671, 1978. Available: https://doi.org/10.1093/brain/101.4.661

[59] B.-S. Jeon *et al.*, "Flash ERG Findings in Parkinson‘s Disease", *J. Korean Neurol. Assoc.*, vol. 5, no. 1, pp. 6–12, 1987.

[60] I. Bodis-Wollner *et al.*, "Visual dysfuntion in Parkinson's disease. Loss in spatiotemporal contrast sensitivity", *Brain*, vol. 110, no. 6, pp. 1675–1698, 1987. Available: https://doi.org/10.1093/brain/110.6.1675

[61] C. Harnois and T. Di Paolo, "Decreased dopamine in the retinas of patients with Parkinson's disease." *Investigative ophthalmology and visual science*, vol. 31, no. 11, pp. 2473–2475, 1990. Available: http://www.ncbi.nlm.nih.gov/pubmed/2243012

[62] I. Ortuño-Lizarán *et al.*, "Dopaminergic Retinal Cell Loss and Visual Dysfunction in Parkinson Disease", *Annals of Neurology*, vol. 88, no. 5, pp. 893–906, 2020. Available: https://doi.org/10.1002/ana.25897

[63] I. Ortuño-Lizarán *et al.*, "Phosphorylated α-synuclein in the retina is a biomarker of Parkinson's disease pathology severity", *Movement Disorders*, vol. 33, no. 8, pp. 1315–1324, 2018. Available: https://doi.org/10.1002/mds.27392

[64] F. Wang *et al.*, "Automated Detection of the Foveal Center Improves SD-OCT Measurements of Central Retinal Thickness", *Ophthalmic Surgery, Lasers and Imaging Retina*, vol. 43, no. 6, pp. S32–S37, 2012. Available: https://doi.org/10.3928/15428877-20121001-06

[65] S. Niu *et al.*, "Automated detection of foveal center in SD-OCT images using the saliency of retinal thickness maps", *Medical Physics*, vol. 44, no. 12, pp. 6390–6403, 2017. Available: https://doi.org/10.1002/mp.12614

[66] B. Liefers *et al.*, "Automatic detection of the foveal center in optical coherence tomography", *Biomedical Optics Express*, vol. 8, no. 11, p. 5160, 2017. Available: https://doi.org/10.1364/BOE.8.005160

[67] S. Schurer-Waldheim *et al.*, "Robust Fovea Detection in Retinal OCT Imaging Using Deep Learning", *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 8, pp. 3927–3937, 2022. Available: https://doi.org/10.1109/jbhi.2022.3166068

[68] P. Tewarie *et al.*, "The OSCAR-IB consensus criteria for retinal OCT quality assessment", *PLoS ONE*, vol. 7, no. 4, pp. 1–7, 2012. Available: https://doi.org/10.1371/journal.pone.0034823

[69] D. M. Stein *et al.*, "A new quality assessment parameter for optical coherence tomography", *British Journal of Ophthalmology*, vol. 90, no. 2, pp. 186–190, 2006. Available: http://dx.doi.org/10.1136/bjo.2004.059824

[70] Y. Huang *et al.*, "Signal Quality Assessment of Retinal Optical Coherence Tomography Images", *Investigative Opthalmology and Visual Science*, vol. 53, no. 4, p. 2133, 2012. Available: https://doi.org/10.1167/iovs.11-8755

[71] "Early Treatment Diabetic Retinopathy Study Design and Baseline Patient Characteristics", *Ophthalmology*, vol. 98, no. 5, pp. 741–756, 1991. Available: https://doi.org/10.1016/S0161-6420(13)38009-9

[72] S. K. Yadav *et al.*, "CuBe: parametric modeling of 3D foveal shape using cubic Bézier", *Biomedical Optics Express*, vol. 8, no. 9, pp. 4181–4199, 2017. Available: https://doi.org/10.1364/BOE.8.004181

[73] A. M. Dubis *et al.*, "Reconstructing foveal pit morphology from optical coherence tomography imaging", *British Journal of Ophthalmology*, vol. 93, no. 9, pp. 1223–1227, 2009. Available: https://doi.org/10.1136/bjo.2008.150110

[74] Y. Ding *et al.*, "Application of an OCT data-based mathematical model of the foveal pit in Parkinson disease", *Journal of Neural Transmission*, vol. 121, no. 11, pp. 1367–1376, 2014. Available: https://doi.org/10.1007/s00702-014-1214-2

[75] L. Liu *et al.*, "A sloped piecemeal Gaussian model for characterising foveal pit shape", *Ophthalmic and Physiological Optics*, vol. 36, no. 6, pp. 615–631, 2016. Available: https://doi.org/10.1111/opo.12321

[76] K. Breher *et al.*, "Direct modeling of foveal pit morphology from distortion-corrected OCT images", *Biomedical Optics Express*, vol. 10, no. 9, pp. 4815–4824, 2019. Available: https://doi.org/10.1364/boe.10.004815

[77] P. Scheibe *et al.*, "Parametric model for the 3D reconstruction of individual fovea shape from OCT data", *Experimental Eye Research*, vol. 119, pp. 19–26, 2014. Available: https://doi.org/10.1016/j.exer.2013.11.008

[78] R. M. Haralick *et al.*, "Textural Features for Image Classification", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973. Available: https://doi.org/10.1109/TSMC.1973.4309314

[79] T. Ojala *et al.*, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002. Available: https://doi.org/10.1109/TPAMI.2002.1017623

[80] K. Huang and S. Aviyente, "Wavelet Feature Selection for Image Classification", *IEEE Transactions on Image Processing*, vol. 17, no. 9, pp. 1709–1720, 2008. Available: https://doi.org/10.1109/TIP.2008.2001050

[81] M. Long and F. Peng, "A Box-Counting Method with Adaptable Box Height for Measuring the Fractal Feature of Images", *Radioengineering*, vol. 22, no. 1, pp. 208–213, 2013. Available: https://www.radioeng.cz/fulltexts/2013/13_01_0208_0213.pdf

[82] A. Roy and E. Perfect, "Lacunarity analyses of multifractal and natural grayscale patterns", *Fractals*, vol. 22, no. 03, p. 1440003, 2014. Available: https://doi.org/10.1142/S0218348X14400039

[83] G. Quellec *et al.*, "Three-Dimensional Analysis of Retinal Layer Texture: Identification of Fluid-Filled Regions in SD-OCT of the Macula", *IEEE Transactions on Medical Imaging*, vol. 29, no. 6, pp. 1321–1330, 2010. Available: https://doi.org/10.1109/TMI.2010.2047023

[84] A. G. Roy, "RelayNet". Available: https://github.com/ai-med/ReLayNet

[85] P. A. Dufour, "OCT Segmentation Application". Available: https://pascaldufour.net/Research/software_data.html

[86] P. A. Dufour *et al.*, "Graph-based multi-surface segmentation of OCT data using trained hard and soft constraints", *IEEE Transactions on Medical Imaging*, vol. 32, no. 3, pp. 531–543, 2013. Available: https://doi.org/10.1109/TMI.2012.2225152

[87] "OCTExplorer 3.8.0". Available: https://iibi.uiowa.edu/oct-reference

[88] M. D. Abràmoff *et al.*, "Retinal imaging and image analysis", *IEEE Reviews in Biomedical Engineering*, vol. 3, no. March, pp. 169–208, 2010. Available: https://doi.org/10.1109/RBME.2010.2084567

[89] Kang Li *et al.*, "Optimal Surface Segmentation in Volumetric Images-A Graph-Theoretic Approach", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 119–134, 2006. Available: https://doi.org/10.1109/TPAMI.2006.19

[90] R. Kafieh *et al.*, "Automatic Multifaceted Matlab Package for Analysis of Ocular Images (AMPAO)", *SoftwareX*, vol. 10, p. 100339, 2019. Available: https://doi.org/10.1016/j.softx.2019.100339

[91] P.-y. Teng, "Caserel: Computer-aided Segmentation of Retinal Layers in Optical Coherence Tomography Images", 2013. Available: https://zenodo.org/record/17893#.ZDQtOXZByzV

[92] M. Mayer, "OCTSEG (Optical Coherence Tomography Segmentation and Evaluation GUI)", 2016. Available: http://www5.cs.fau.de/research/software/octseg

[93] M. Montazerin, "Livelayer-Software", 2021. Available: https://github.com/MansoorehMontazerin/Livelayer-Software

[94] M. Montazerin *et al.*, "Livelayer: a semi-automatic software program for segmentation of layers and diabetic macular edema in optical coherence tomography images", *Scientific Reports*, vol. 11, no. 1, p. 13794, 2021. Available: https://doi.org/10.1038/s41598-021-92713-y

[95] "AURA tools: AUtomated Retinal Analysis tools.". Available: https://www.nitrc.org/projects/aura_tools

[96] Y. Subhi *et al.*, "Macular thickness and volume in the elderly: A systematic review", *Ageing Research Reviews*, vol. 29, pp. 42–49, 2016. Available: https://doi.org/10.1016/j.arr.2016.05.013

[97] P. J. Patel *et al.*, "Spectral-Domain Optical Coherence Tomography Imaging in 67 321 Adults: Associations with Macular thickness in the UK Biobank Study", *Ophthalmology*, vol. 123, no. 4, pp. 829–840, 2015. Available: https://doi.org/10.1016/j.ophtha.2015.11.009

[98] Q. Xu *et al.*, "Assessment of the effect of age on macular layer thickness in a healthy Chinese cohort using spectral-domain optical coherence tomography", *BMC Ophthalmology*, vol. 18, no. 1, pp. 1–9, 2018. Available: https://doi.org/10.1186/s12886-018-0842-y

[99] N. Hashmani *et al.*, "Assessing reproducibility and the effects of demographic variables on the normal macular layers using the spectralis SD-OCT", *Clinical Ophthalmology*, vol. 12, pp. 1433–1440, 2018. Available: https://doi.org/10.2147/OPTH.S172109

[100] M. Nieves-Moreno *et al.*, "Impacts of age and sex on retinal layer thicknesses measured by spectral domain optical coherence tomography with Spectralis", *PLOS ONE*, vol. 13, no. 3, p. e0194169, 2018. Available: https://doi.org/10.1371/journal.pone.0194169

[101] J. Chua *et al.*, "Age-related changes of individual macular retinal layers among Asians", *Scientific Reports*, vol. 9, no. 1, pp. 1–11, 2019. Available: https://doi.org/10.1038/s41598-019-56996-6

[102] J. B. Jonas *et al.*, "Human optic nerve fiber count and optic disc size", *Investigative Ophthalmology and Visual Science*, vol. 33, no. 6, pp. 2012–2018, 1992. Available: https://iovs.arvojournals.org/article.aspx?articleid=2161180

[103] H. Gao and J. G. Hollyfield, "Aging of the human retina: Differential loss of neurons and retinal pigment epithelial cells", *Investigative Ophthalmology and Visual Science*, vol. 33, no. 1, pp. 1–17, 1992. Available: https://iovs.arvojournals.org/article.aspx?articleid=2160563

[104] S. Ooto *et al.*, "Effects of age, sex, and axial length on the three-dimensional profile of normal macular layer structures", *Investigative Ophthalmology and Visual Science*, vol. 52, no. 12, pp. 8769–8779, 2011. Available: https://doi.org/10.1167/iovs.11-8388

[105] J. Y. Won *et al.*, "Effect of age and sex on retinal layer thickness and volume in normal eyes", *Medicine (United States)*, vol. 95, no. 46, p. e5441, 2016. Available: https://doi.org/10.1097/MD.0000000000005441

[106] A. P. Khawaja *et al.*, "Comparison of Associations with Different Macular Inner Retinal Thickness Parameters in a Large Cohort: The UK Biobank", *Ophthalmology*, vol. 127, no. 1, pp. 62–71, 2020. Available: https://doi.org/10.1016/j.ophtha.2019.08.015

[107] M. M. Mauschitz *et al.*, "Determinants of Macular Layers and Optic Disc Characteristics on SD-OCT: The Rhineland Study", *Translational Vision Science & Technology*, vol. 8, no. 3, p. 34, 2019. Available: https://doi.org/10.1167/tvst.8.3.34

[108] N. Demirkaya *et al.*, "Effect of age on individual retinal layer thickness in normal eyes as measured with spectral-domain optical coherence tomography", *Investigative Ophthalmology and Visual Science*, vol. 54, no. 7, pp. 4934–4940, 2013. Available: https://doi.org/10.1167/iovs.13-11913

[109] C. A. Girkin *et al.*, "Variation in Optic Nerve and Macular Structure with Age and Race with Spectral-Domain Optical Coherence Tomography", *Ophthalmology*, vol. 118, no. 12, pp. 2403–2408, 2011. Available: https://doi.org/10.1016/j.ophtha.2011.06.013

[110] N. Hammel *et al.*, "Comparing the Rates of Retinal Nerve Fiber Layer and Ganglion Cell–Inner Plexiform Layer Loss in Healthy Eyes and in Glaucoma Eyes", *American Journal of Ophthalmology*, vol. 178, pp. 38–50, 2017. Available: https://doi.org/10.1016/j.ajo.2017.03.008

[111] M. C. Savastano *et al.*, "Differential vulnerability of retinal layers to early age-related macular degeneration: Evidence by SD-OCT segmentation analysis", *Investigative Ophthalmology and Visual Science*, vol. 55, no. 1, pp. 560–566, 2014. Available: https://doi.org/10.1167/iovs.13-12172

[112] J. Wang *et al.*, "Swept-source optical coherence tomography imaging of macular retinal and choroidal structures in healthy eyes", *BMC Ophthalmology*, vol. 15, no. 1, pp. 1–10, 2015. Available: https://doi.org/10.1186/s12886-015-0110-3

[113] N. Yoshioka *et al.*, "Pattern recognition analysis of age-related retinal ganglion cell signatures in the human eye", *Investigative Ophthalmology and Visual Science*, vol. 58, no. 7, pp. 3086–3099, 2017. Available: https://doi.org/10.1167/iovs.17-21450

[114] J. Tong *et al.*, "Development of a Spatial Model of Age-Related Change in the Macular Ganglion Cell Layer to Predict Function From Structural Changes", *American Journal of Ophthalmology*, vol. 208, pp. 166–177, 2019. Available: https://doi.org/10.1016/j.ajo.2019.04.020

[115] Y. J. Yoo *et al.*, "Inner macular layer thickness by spectral domain optical coherence tomography in children and adults: A hospital-based study", *British Journal of Ophthalmology*, vol. 103, no. 11, pp. 1576–1583, 2018. Available: https://doi.org/10.1136/bjophthalmol-2018-312349

[116] B. C. Chauhan *et al.*, "Differential Effects of Aging in the Macular Retinal Layers, Neuroretinal Rim, and Peripapillary Retinal Nerve Fiber Layer", *Ophthalmology*, vol. 127, no. 2, pp. 177–185, 2020. Available: https://doi.org/10.1016/j.ophtha.2019.09.013

[117] W. K. Song *et al.*, "Macular thickness variations with sex, age, and axial length in healthy subjects: A spectral domain-optical coherence tomography study", *Investigative Ophthalmology and Visual Science*, vol. 51, no. 8, pp. 3913–3918, 2010. Available: https://doi.org/10.1167/iovs.09-4189

[118] A. H. Kashani *et al.*, "Retinal Thickness Analysis by Race, Gender, and Age Using Stratus OCT", *American Journal of Ophthalmology*, vol. 149, no. 3, pp. 496–512, 2010. Available: https://doi.org/10.1016/j.ajo.2009.09.025

[119] A. Palazon-Cabanes *et al.*, "Normative Database for All Retinal Layer Thicknesses Using SD-OCT Posterior Pole Algorithm and the Effects of Age, Gender and Axial Lenght", *Journal of Clinical Medicine*, vol. 9, no. 10, p. 3317, 2020. Available: https://doi.org/10.3390/jcm9103317

[120] A. Nunes *et al.*, "Sexual Dimorphism of the Adult Human Retina Assessed by Optical Coherence Tomography", in *XV Mediterranean Conference on Medical and Biological Engineering and Computing – MEDICON 2019*, J. Henriques *et al.*, Eds. Springer International Publishing, 2020, pp. 1830–1834. Available: https://doi.org/10.1007/978-3-030-31635-8_222

[121] A. Szigeti *et al.*, "The Effect of Axial Length on the Thickness of Intraretinal Layers of the Macula", *PLOS ONE*, vol. 10, no. 11, p. e0142383, 2015. Available: https://doi.org/10.1371/journal.pone.0142383

[122] I. Ctori *et al.*, "The effects of ocular magnification on Spectralis spectral domain optical coherence tomography scan length", *Graefe's Archive for Clinical and Experimental Ophthalmology*, vol. 253, no. 5, pp. 733–738, 2015. Available: https://doi.org/10.1007/s00417-014-2915-9

[123] H. Chen *et al.*, "Distribution of axial length, anterior chamber depth, and corneal curvature in an aged population in South China", *BMC Ophthalmology*, vol. 16, no. 1, pp. 1–8, 2016. Available: https://doi.org/10.1186/s12886-016-0221-5

[124] M. Trinh *et al.*, "Modelling normal age-related changes in individual retinal layers using location-specific OCT analysis", *Scientific Reports*, vol. 11, no. 1, p. 558, 2021. Available: https://doi.org/10.1038/s41598-020-79424-6

[125] P. Scheibe *et al.*, "Analysis of foveal characteristics and their asymmetries in the normal population", *Experimental Eye Research*, vol. 148, pp. 1–11, 2016. Available: https://doi.org/10.1016/j.exer.2016.05.013

[126] Y. Barak *et al.*, "Mathematical analysis of specific anatomic foveal configurations predisposing to the formation of macular holes", *Investigative Ophthalmology and Visual Science*, vol. 52, no. 11, pp. 8266–8270, 2011. Available: https://doi.org/10.1167/iovs.11-8191

[127] A. M. Dubis *et al.*, "Relationship between the foveal avascular zone and foveal pit morphology", *Investigative Ophthalmology and Visual Science*, vol. 53, no. 3, pp. 1628–1636, 2012. Available: https://doi.org/10.1167/iovs.11-8488

[128] S. Tick *et al.*, "Foveal shape and structure in a normal population", *Investigative Ophthalmology and Visual Science*, vol. 52, no. 8, pp. 5105–5110, 2011. Available: https://doi.org/10.1167/iovs.10-7005

[129] B. Nesmith *et al.*, "Mathematical analysis of the normal anatomy of the aging fovea", *Investigative Ophthalmology and Visual Science*, vol. 55, no. 9, pp. 5962–5966, 2014. Available: https://doi.org/10.1167/iovs.14-15278

[130] L. Gella *et al.*, "Foveal slope measurements in diabetic retinopathy: Can it predict development of sight-threatening retinopathy? Sankara Nethralaya Diabetic Retinopathy Epidemiology and Molecular Genetics Study (SN-DREAMS II, Report no 8)", *Indian Journal of Ophthalmology*, vol. 63, no. 6, pp. 478–481, 2015. Available: https://doi.org/10.4103/0301-4738.162578

[131] J. A. Sepulveda *et al.*, "Individual differences in Foveal shape: Feasibility of individual maps between structure and function within the macular region", *Investigative Ophthalmology and Visual Science*, vol. 57, no. 11, pp. 4772–4778, 2016. Available: https://doi.org/10.1167/iovs.16-19288

[132] M. A. Zouache *et al.*, "Comparison of the Morphology of the Foveal Pit Between African and Caucasian Populations", *Translational Vision Science & Technology*, vol. 9, no. 5, p. 24, 2020. Available: https://doi.org/10.1167/tvst.9.5.24

[133] M. Wagner-Schuman *et al.*, "Race- and sex-related differences in retinal thickness and foveal pit morphology", *Investigative Ophthalmology and Visual Science*, vol. 52, no. 1, pp. 625–634, 2011. Available: https://doi.org/10.1167/iovs.10-5886

[134] A. Olvera-Barrios *et al.*, "Foveal Curvature and Its Associations in UK Biobank Participants", *Investigative ophthalmology and visual science*, vol. 63, no. 8, p. 26, 2022. Available: https://doi.org/10.1167/iovs.63.8.26

[135] R. Inzelberg *et al.*, "Retinal nerve fiber layer thinning in Parkinson disease", *Vision Research*, vol. 44, no. 24, pp. 2793–2797, 2004. Available: https://doi.org/10.1016/j.visres.2004.06.009

[136] J. G. Yu *et al.*, "Retinal nerve fiber layer thickness changes in Parkinson disease: A meta-analysis", *PLoS ONE*, vol. 9, no. 1, p. e85718, 2014. Available: https://doi.org/10.1371/journal.pone.0085718

[137] L. Huang *et al.*, "Central retina changes in Parkinson's disease: a systematic review and meta-analysis", *Journal of Neurology*, vol. 268, pp. 4646–4654, 2020. Available: https://doi.org/10.1007/s00415-020-10304-9

[138] W. Zhou *et al.*, "Optical coherence tomography measurements as potential imaging biomarkers for Parkinson's disease: A systematic review and meta-analysis", *European Journal of Neurology*, vol. 28, no. 3, pp. 763–774, 2021. Available: https://doi.org/10.1111/ene.14613

[139] Y. Deng *et al.*, "Evaluation of retina and microvascular changes in the patient with Parkinson's disease: A systematic review and meta-analysis", *Frontiers in Medicine*, vol. 9, 2022. Available: https://doi.org/10.3389/fmed.2022.957700

[140] P. Mailankody *et al.*, "The role of Optical Coherence Tomography in Parkinsonism: A critical review", *Journal of the Neurological Sciences*, vol. 403, pp. 67–74, 2019. Available: https://doi.org/10.1016/j.jns.2019.06.009

[141] J. N. Alves *et al.*, "Structural and functional changes in the retina in Parkinson's disease", *Journal of Neurology, Neurosurgery and Psychiatry*, vol. 94, pp. 448–456, 2023. Available: https://doi.org/10.1136/jnnp-2022-329342

[142] M. Satue *et al.*, "Evaluation of progressive visual dysfunction and retinal degeneration in patients with parkinson's disease", *Investigative Ophthalmology and Visual Science*, vol. 58, no. 2, pp. 1151–1157, 2017. Available: https://doi.org/10.1167/iovs.16-20460

[143] L.-J. Ma *et al.*, "Progressive Changes in the Retinal Structure of Patients with Parkinson's Disease", *Journal of Parkinson's Disease*, vol. 8, no. 1, pp. 85–92, 2018. Available: https://doi.org/10.3233/JPD-171184

[144] S. Hasanov *et al.*, "Functional and morphological assessment of ocular structures and follow-up of patients with early-stage Parkinson's disease", *International Ophthalmology*, vol. 39, no. 6, pp. 1255–1262, 2019. Available: https://doi.org/10.1007/s10792-018-0934-y

[145] A. Murueta-Goyena *et al.*, "Retinal thickness predicts the risk of cognitive decline in Parkinson's disease", *Annals of Neurology*, vol. 89, no. 1, pp. 165–176, 2020. Available: https://doi.org/10.1002/ana.25944

[146] E. Garcia-Martin *et al.*, "Retina measurements for diagnosis of Parkinson disease", *Retina*, vol. 34, no. 5, pp. 971–980, 2014. Available: https://doi.org/10.1097/IAE.0000000000000028

[147] S. Miri *et al.*, "A combination of retinal morphology and visual electrophysiology testing increases diagnostic yield in Parkinson's disease", *Parkinsonism and Related Disorders*, vol. 22, pp. S134–S137, 2016. Available: https://doi.org/10.1016/j.parkreldis.2015.09.015

[148] J. Huang *et al.*, "Combination of Multifocal Electroretinogram and Spectral-Domain OCT Can Increase Diagnostic Efficacy of Parkinson's Disease", *Parkinson's Disease*, vol. 2018, pp. 1–7, 2018. Available: https://doi.org/10.1155/2018/4163239

[149] J. Zou *et al.*, "Combination of optical coherence tomography (OCT) and OCT angiography increases diagnostic efficacy of Parkinson's disease", *Quantitative Imaging in Medicine and Surgery*, vol. 10, no. 10, pp. 1930–1939, 2020. Available: https://doi.org/10.21037/qims-20-460

[150] M. Satue *et al.*, "Ability of Swept-source OCT and OCT-angiography to detect neuroretinal and vasculature changes in patients with Parkinson disease and essential tremor", *Eye*, vol. 37, no. 7, pp. 1314–1319, 2023. Available: https://doi.org/10.1038/s41433-022-02112-4

[151] B. Spund *et al.*, "Remodeling of the fovea in Parkinson disease", *Journal of Neural Transmission*, vol. 120, no. 5, pp. 745–753, 2013. Available: https://doi.org/10.1007/s00702-012-0909-5

[152] A. Pilat *et al.*, "In vivo morphology of the optic nerve and retina in patients with Parkinson's disease", *Investigative Ophthalmology and Visual Science*, vol. 57, no. 10, pp. 4420–4427, 2016. Available: https://doi.org/10.1167/iovs.16-20020

[153] I. Bodis-Wollner, "Foveal vision is impaired in Parkinson's disease", *Parkinsonism and Related Disorders*, vol. 19, no. 1, pp. 1–14, 2013. Available: https://doi.org/10.1016/j.parkreldis.2012.07.012

[154] S. Slotnick *et al.*, "A novel retinal biomarker for Parkinson's disease: Quantifying the foveal pit with optical coherence tomography", *Movement Disorders*, vol. 30, no. 12, pp. 1692–1695, 2015. Available: https://doi.org/10.1002/mds.26411

[155] J. B. Young *et al.*, "Assessing Retinal Structure in Patients with Parkinson's Disease", *Journal of Neurology & Neurophysiology*, vol. 10, no. 1, pp. 1–7, 2019. Available: htttps://doi.org/10.4172/2155-9562.1000485

[156] E. H. Pinkhardt *et al.*, "The intrinsically restructured fovea is correlated with contrast sensitivity loss in Parkinson's disease", *Journal of Neural Transmission*, vol. 127, no. 9, pp. 1275–1283, 2020. Available: https://doi.org/10.1007/s00702-020-02224-9

[157] A. Nunes *et al.*, "Retinal texture biomarkers may help to discriminate between Alzheimer's, Parkinson's, and healthy controls", *PLoS ONE*, vol. 14, no. 6, p. e0218826, 2019. Available: https://doi.org/10.1371/journal.pone.0218826

[158] J. Zhang *et al.*, "Wavelet Features of the Thickness Map of Retinal Ganglion Cell-Inner Plexiform Layer Best Discriminate Prior Optic Neuritis in Patients With Multiple Sclerosis", *IEEE Access*, vol. 8, pp. 221 590–221 598, 2020. Available: https://doi.org/10.1109/ACCESS.2020.3041291

[159] H. D. Tazarjani *et al.*, "Retinal OCT Texture Analysis for Differentiating Healthy Controls from Multiple Sclerosis (MS) with/without Optic Neuritis", *BioMed Research International*, vol. 2021, pp. 1–13, 2021. Available: https://doi.org/10.1155/2021/5579018

[160] L. Jáñez-García *et al.*, "Roughness of retinal layers in Alzheimer's disease", *Scientific Reports*, vol. 11, no. 1, p. 11804, 2021. Available: https://doi.org/10.1038/s41598-021-91097-3

[161] A. Murueta-Goyena *et al.*, "Foveal Remodeling of Retinal Microvasculature in Parkinson's Disease", *Frontiers in Neuroscience*, vol. 15, 2021. Available: https://doi.org/10.3389/fnins.2021.708700

[162] C. G. Goetz *et al.*, "Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results", *Movement Disorders*, vol. 23, no. 15, pp. 2129–2170, 2008. Available: https://doi.org/10.1002/mds.22340

[163] M. M. Hoehn and M. D. Yahr, "Parkinsonism: Onset, progression and mortality", *Neurology*, vol. 57, no. 10, pp. S11–S26, 2001. Available: https://doi.org/https://doi.org/10.1212/wnl.17.5.427

[164] D. J. Gill *et al.*, "The montreal cognitive assessment as a screening tool for cognitive impairment in Parkinson's disease", *Movement Disorders*, vol. 23, no. 7, pp. 1043–1046, 2008. Available: https://doi.org/10.1002/mds.22017

[165] E. A. Pugh *et al.*, "Effects of Normative Adjustments to the Montreal Cognitive Assessment", *The American Journal of Geriatric Psychiatry*, vol. 26, no. 12, pp. 1258–1267, 2018. Available: https://doi.org/10.1016/j.jagp.2018.09.009

[166] N. Carson *et al.*, "A re-examination of Montreal Cognitive Assessment (MoCA) cutoff scores", *International Journal of Geriatric Psychiatry*, vol. 33, no. 2, pp. 379–388, 2018. Available: https://doi.org/10.1002/gps.4756

[167] Ö. Altintaş *et al.*, "Correlation between retinal morphological and functional findings and clinical severity in Parkinson's disease", *Documenta Ophthalmologica*, vol. 116, no. 2, pp. 137–146, 2008. Available: https://doi.org/10.1007/s10633-007-9091-8

[168] B. Jiménez *et al.*, "Development of a prediction formula of Parkinson disease severity by optical coherence tomography", *Movement Disorders*, vol. 29, no. 1, pp. 68–74, 2014. Available: https://doi.org/10.1002/mds.25747

[169] M. Satue *et al.*, "Retinal thinning and correlation with functional disability in patients with Parkinson's disease", *British Journal of Ophthalmology*, vol. 98, no. 3, pp. 350–355, 2014. Available: https://doi.org/10.1136/bjophthalmol-2013-304152

[170] M. M. El-Kattan *et al.*, "Optical coherence tomography in patients with Parkinson's disease", *The Egyptian Journal of Neurology, Psychiatry and Neurosurgery*, vol. 58, no. 1, p. 21, 2022. Available: https://doi.org/10.1186/s41983-021-00421-1

[171] X. Wang *et al.*, "The macular inner plexiform layer thickness as an early diagnostic indicator for Parkinson's disease", *npj Parkinson's Disease*, vol. 8, no. 1, p. 63, 2022. Available: https://doi.org/10.1038/s41531-022-00325-8

[172] E. S. Sari *et al.*, "Ganglion Cell–Inner Plexiform Layer Thickness in Patients With Parkinson Disease and Association With Disease Severity and Duration", *Journal of Neuro-Ophthalmology*, vol. 35, no. 2, pp. 117–121, 2015. Available: https:/doi.org/10.1097/WNO.0000000000000203

[173] E. Cubo *et al.*, "Lack of association of morphologic and functional retinal changes with motor and non-motor symptoms severity in Parkinson's disease", *Journal of Neural Transmission*, vol. 121, no. 2, pp. 139–145, 2014. Available: https://doi.org/10.1007/s00702-013-1093-y

[174] M. S. Sung *et al.*, "Inner retinal thinning as a biomarker for cognitive impairment in de novo Parkinson's disease", *Scientific Reports*, vol. 9, no. 1, p. 11832, 2019. Available: https://doi.org/10.1038/s41598-019-48388-7

[175] Z. Chang *et al.*, "Retinal Nerve Fiber Layer Thickness and Associations With Cognitive Impairment in Parkinson's Disease", *Frontiers in Aging Neuroscience*, vol. 14, 2022. Available: https://doi.org/10.3389/fnagi.2022.832768

[176] A. Murueta-Goyena *et al.*, "Parafoveal thinning of inner retina is associated with visual dysfunction in Lewy body diseases", *Movement Disorders*, vol. 34, no. 9, pp. 1315–1324, 2019. Available: https://doi.org/10.1002/mds.27728

[177] V. T. Chan *et al.*, "Spectral-Domain OCT Measurements in Alzheimer's Disease: A Systematic Review and Meta-analysis", *Ophthalmology*, vol. 126, no. 4, pp. 497–510, 2019. Available: https://doi.org/10.1016/j.ophtha.2018.08.009

[178] S. K. Wagner *et al.*, "AlzEye: longitudinal record-level linkage of ophthalmic imaging and hospital admissions of 353 157 patients in London, UK", *BMJ Open*, vol. 12, no. 3, p. e058552, 2022. Available: https://doi.org/10.1136/bmjopen-2021-058552

[179] G. Liew *et al.*, "The Retinal Vasculature as a Fractal: Methodology, Reliability, and Relationship to Blood Pressure", *Ophthalmology*, vol. 115, no. 11, pp. 1951–1956.e1, 2008. Available: https://doi.org/10.1016/j.ophtha.2008.05.029

[180] A. Perez-Rovira *et al.*, "VAMPIRE: Vessel assessment and measurement platform for images of the REtina", in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2011, pp. 3391–3394. Available: https://doi.org/10.1109/IEMBS.2011.6090918

[181] Y. Zhou *et al.*, "AutoMorph: Automated Retinal Vascular Morphology Quantification Via a Deep Learning Pipeline", *Translational Vision Science & Technology*, vol. 11, no. 7, p. 12, 2022. Available: https://doi.org/10.1167/tvst.11.7.12

[182] D. Romero-Bascones *et al.*, "Foveal Pit Morphology Characterization: A Quantitative Analysis of the Key Methodological Steps", *Entropy*, vol. 23, no. 6, p. 699, 2021. Available: https://doi.org/10.3390/e23060699

[183] D. Romero-Bascones *et al.*, "Spatial characterization of the effect of age and sex on macular layer thicknesses and foveal pit morphology", *PLOS ONE*, vol. 17, no. 12, p. e0278925, 2022. Available: https://doi.org/10.1371/journal.pone.0278925

[184] I.-K. Yeo, "A new family of power transformations to improve normality or symmetry", *Biometrika*, vol. 87, no. 4, pp. 954–959, 2000. Available: https://doi.org/10.1093/biomet/87.4.954

[185] M. L. Kirby *et al.*, "Foveal anatomic associations with the secondary peak and the slope of the macular pigment spatial profile", *Investigative Ophthalmology and Visual Science*, vol. 50, no. 3, pp. 1383–1391, 2009. Available: https://doi.org/10.1167/iovs.08-2494

[186] T. K. Koo and M. Y. Li, "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research", *Journal of Chiropractic Medicine*, vol. 15, no. 2, pp. 155–163, 2016. Available: https://doi.org/10.1167/iovs.08-2494

[187] D. Romero-Bascones *et al.*, "RETIMAT: an open-source software for OCT image analysis", in *ARVO*, 2023. Available: https://iovs.arvojournals.org/article.aspx?articleid=2790292

[188] "Unified OCT explorer". Available: https://bitbucket.org/uocte/

[189] M. Graham, "OCT-Converter (v0.5.0)". Available: https://pypi.org/project/oct-converter/

[190] S. K. Wagner *et al.*, "Association Between Retinal Features From Multimodal Imaging and Schizophrenia", *JAMA Psychiatry*, vol. 80, no. 5, p. 478, 2023. Available: https://doi.org/10.1001/jamapsychiatry.2023.0171

[191] L. K. Soh and C. Tsatsoulis, "Texture analysis of sar sea ice imagery using gray level co-occurrence matrices", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 2 I, pp. 780–795, 1999. Available: https://doi.org/10.1109/36.752194

[192] J. J. Van Griethuysen *et al.*, "Computational radiomics system to decode the radiographic phenotype", *Cancer Research*, vol. 77, no. 21, pp. e104–e107, 2017. Available: https://doi.org/10.1158/0008-5472.CAN-17-0339

[193] S. K. Wagner *et al.*, "Retinal Optical Coherence Tomography Features Associated With Incident and Prevalent Parkinson Disease", *Neurology*, 2023. Available: https://doi.org/10.1212/WNL.0000000000207727

[194] S. Holm, "A simple sequentially rejective multiple test procedure." *Scandinavian journal of statistics*, vol. 6, no. 2, pp. 65–70, 1979. Available: https://www.jstor.org/stable/4615733

[195] Y. Benjamini and Y. Hochberg, "Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing", *Journal of the Royal Statistical Society. Series B*, vol. 57, no. 1, pp. 289–300, 1995. Available: https://doi.org/10.2307/2346101

[196] T. Oberwahrenbrock *et al.*, "Reliability of Intra-Retinal Layer Thickness Estimates." *PloS one*, vol. 10, no. 9, p. e0137316, 2015. Available: https://doi.org/10.1371/journal.pone.0137316

[197] M. N. Menke *et al.*, "Reproducibility of Retinal Thickness Measurements in Healthy Subjects Using Spectralis Optical Coherence Tomography", *American Journal of Ophthalmology*, vol. 147, no. 3, pp. 467–472, 2009. Available: https://doi.org/10.1016/j.ajo.2008.09.005

[198] J. Barrio-Barrio *et al.*, "Multicenter Spanish study of spectral-domain optical coherence tomography in normal children", *Acta Ophthalmologica*, vol. 91, no. 1, pp. 56–63, 2013. Available: https://doi.org/10.1111/j.1755-3768.2012.02562.x

[199] J. M. Pope *et al.*, "Three-dimensional MRI study of the relationship between eye dimensions, retinal shape and myopia", *Biomedical Optics Express*, vol. 8, no. 5, p. 2386, 2017. Available: https://doi.org/10.1364/boe.8.002386

[200] J. Denniss *et al.*, "Individualized structure-function mapping for glaucoma: Practical constraints on map resolution for clinical and research applications", *Investigative Ophthalmology and Visual Science*, vol. 55, no. 3, pp. 1985–1993, 2014. Available: https://doi.org/10.1167/iovs.13-13758

[201] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data", *Journal of Bioinformatics and Computational Biology*, vol. 03, no. 02, pp. 185–205, 2005. Available: https://doi.org/10.1142/S0219720005001004

[202] H. Peng *et al.*, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005. Available: https://doi.org/10.1109/TPAMI.2005.159

[203] T. Hastie *et al.*, *The Elements of Statistical Learning*, ser. Springer Series in Statistics. New York, NY: Springer New York, 2001. Available: https://doi.org/10.1007/978-0-387-21606-5

[204] H. C. Rossetti *et al.*, "Normative data for the Montreal Cognitive Assessment (MoCA) in a population-based sample", *Neurology*, vol. 77, no. 13, pp. 1272–1275, 2011. Available: https://doi.org/10.1212/WNL.0b013e318230208a

[205] M. Barrenechea *et al.*, "Deep learning applied to retinal OCT images to differentiate Parkinson's disease patients from healthy controls", in *Movement Disorders*. Wiley, 2022.

[206] A. N. Warwick *et al.*, "UK Biobank retinal imaging grading: methodology, baseline characteristics and findings for common ocular diseases", *Eye*, vol. 37, no. 10, pp. 2109–2116, 2023. Available: https://doi.org/10.1038/s41433-022-02298-7

[207] N. Hannaway *et al.*, "Visual dysfunction is a better predictor than retinal thickness for dementia in Parkinson's disease", *Journal of Neurology, Neurosurgery and Psychiatry*, vol. 94, pp. 742–750, 2023. Available: https://doi.org/10.1136/jnnp-2023-331083

[208] A. Murueta-Goyena *et al.*, "Retinal thickness as a biomarker of cognitive impairment in manifest Huntington's disease", *Journal of Neurology*, vol. 270, pp. 3821–3829, 2023. Available: https://link.springer.com/10.1007/s00415-023-11720-3