UNISINOS

Programa de Pós-Graduação em

Computação Aplicada

Mestrado Acadêmico

GABRIEL CASTRO FERNANDES

EXPLORING ARTIFICIAL INTELLIGENCE METHODS FOR THE AUTOMATIC MEASUREMENT OF A NEW BIOMARKER AIMING AT GLAUCOMA DIAGNOSIS

São Leopoldo
2024

Gabriel Castro Fernandes

EXPLORING ARTIFICIAL INTELLIGENCE METHODS FOR THE AUTOMATIC MEASUREMENT OF A NEW BIOMARKER AIMING AT GLAUCOMA DIAGNOSIS

A dissertation presented as a partial requirement for obtaining a Master's degree from the Postgraduate Program in Applied Computing at University of Vale do Rio dos Sinos — UNISINOS

Advisor:
Prof. PhD. Sandro José Rigo

Co-advisor:
Prof. PhD. Luiz José Schirmer Silva

São Leopoldo
2024

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)

F363e Fernandes, Gabriel Castro

Exploring Artificial Intelligence methods for the automatic measurement of a new biomarker aiming at glaucoma diagnosis / Gabriel Castro Fernandes — 2024.

67 f.: il.; 30 cm.

Dissertação (mestrado) — Universidade do Vale do Rio dos Sinos, Graduate Program in Applied Computing, São Leopoldo, 2024.

"Advisor: Prof. PhD. Sandro José Rigo; Co-advisor: Prof. PhD. Luiz José Schirmer Silva."

1. Glaucoma. 2. Artificial Intelligence. 3. Diagnosis. 4. Biomarker. I. Título.

CDU 004.8

Responsible Librarian: Bruna Sant'Anna — CRB 10/2360

ABSTRACT

Analyzing retina structure in high-resolution images, such as those obtained in optical coherence tomography, is one of the most widespread ways of identifying structural changes that may indicate the onset or progression of visual impairment. During the diagnosis process, the specialist performs several manual analyses of the data generated by imaging equipment. There is a consensus regarding the benefits of having support from automated approaches to help in this diagnosis process. Nevertheless, automated glaucoma detection using optical coherence tomography is still considered an area needing further research. This work presents an approach to foster automatic glaucoma evaluation considering convolutional neural networks for semantic segmentation of retinal layers through optical coherence tomography images and image processing for measuring the cup region in the optic nerve head portion. We provide a quantitative evaluation comparing the results obtained by a specialist physician. The work's main contribution is presenting the first approach supporting the automation of a new biomarker for diagnosing glaucoma.

Keywords: Glaucoma. Artificial Intelligence. Diagnosis. Biomarker.

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AI	Artificial Intelligence
BM	Bruch's Membrane
BMO	Bruch's Membrane Opening
CDR	Cup-to-Disc Ratio
CNN	Convolutional Neural Network
CSF	Retrobulbar Cerebrospinal Fluid
CV	Computer Vision
DAB	Dilated Attention Block
DAC	Dense Atrous Convolution
DL	Deep Learning
EDI-OCT	Enhanced Depth Imaging OCT
EDI	Enhanced Depth Imaging
F1	F1-Score
FCN	Fully Convolutional Networks
GELU	Gaussian Error Linear Unit
ILM	Internal Limiting Membrane
INL	Inner Nuclear Layer
IOP	Intraocular Pressure
IoU	Intersection Over Union
IPL	Inner Plexiform Layer
LC	Lamina Cribosa
ML	Machine Learning
NFL	Nerve Fiber Layer
NN	Neural Network
OCT	Optical Coherence Tomography
OD	Optic Disc
OLM	Outer Limiting Membrane
ON	Optic Nerve
ONH	Optic Nerve Head
ONL	Outer Nuclear Layer
OPL	Outer Plexiform Layer
PLNT	Prelaminar Neural Tissue
PR	Photoreceptor Layer

ReLU	Rectified Linear Unit
RGC	Retinal Ganglion Cells
RNFL	Retinal Nerve Fiber Layers
RNN	Recurrent Neural Networks
RPE	Retinal Pigment Epithelium
SD-OCT	Spectral-Domain OCT
SS-OCT	Swept-Source OCT
TD-OCT	Time-Domain OCT
TLP	Transient Translaminar Pressure
VF	Visual Field

CONTENTS

1 INTRODUCTION

1.1 Context and motivation

With the demographic growth of society and the increase in life expectancy, many countries in the world have shown an increase in the average age of their population and also in the rise of diseases considered non-communicable and causing disabilities, such as cataracts, glaucoma, refractive error insufficiently repaired, age-related macular degeneration, and diabetic retinopathy. Visual impairments and blindness can arise at any stage of life, but people over the age of 50 years are more likely to have some of these disabilities (BOURNE et al., 2021). Follow-up through clinical examinations allows physicians to have an initial assessment, accompanied by intraocular pressure (IOP) measurements, field-of-view testing, and structural imaging parameters. In cases of glaucoma, the disease may begin with a minimal disorder in the nerve fibers, almost invisible, and the loss of retinal ganglion cells (RGC). If it progresses to the intermediate stage, changes already become noticeable in the retinal nerve fiber layers (RNFL), along with peripheral vision loss. Finally, when fully installed, it presents serious damage and great vision loss (KHALIL et al., 2018).

Glaucoma is a chronic and progressive optic neuropathy characterized by the death of RGC associated with increased cupping of the optic nerve head (ONH) and a corresponding loss of visual field (VF). Several studies have shown that the lamina cribrosa (LC) has a role in the pathophysiology of glaucoma. The LC is the major area related to mechanical damage in glaucoma, and the detailed assessment of disease-related changes at that site is relevant in understanding factors concerning the development and progression of glaucoma. Furthermore, Andrade and coworkers have also proposed that LC structural differences could be responsible for different effects of IOP on the tissue, contributing to individual susceptibility to glaucomatous damage mediated by the IOP (ANDRADE et al., 2022).

Strong evidence also suggests that the biomechanics of the ONH connective tissues (LC and sclera) may dictate the IOP level that can be safely sustained. Unfortunately, the entire LC is rarely visible in the routine ophthalmoscopic examination due to the prelaminar neuroretinal tissue. Even in some far-advanced glaucomatous eyes where little remnant prelaminar tissue exists, only the anterior surface of the LC is visible with deeper portions of the LC observable by histologic sectioning. Optical coherence tomography (OCT) can provide high-resolution cross-sectional images of the ONH *in vivo*. Improvements in imaging technologies, including enhanced depth imaging (EDI) and sophisticated analyzing techniques, enable the unveiling of the LC *in vivo* features (KIM et al., 2013).

Still, Andrade and coworkers (ANDRADE et al., 2022) suggested that the LC is constantly submitted to transient alterations of transient translaminar pressure (TLP) (acute and chronic), which generates a constant remodeling in response to physiologic and pathophysiologic stimuli. This may result in a pathologic remodeling of the structure, with focal laminar defects, changes

in pore morphology, thinning, and posterior displacement.

The relationship between structural and functional damage is relevant for diagnosing glaucoma. Structural evaluation of the ONH, RNFL, and the macula, when analyzed by OCT, can provide important clinical information for the diagnosis of glaucoma (LAVINSKY et al., 2018). Therefore, optical coherence tomography, as a technology capable of providing high-resolution images of the structure of the retina, is one of the key resources used to inspect it, allowing access to the visualization of the layers to follow changes in thickness and possible deformations (LI et al., 2021).

Detecting progress with visual field tests or optic disc imaging methods is limited due to the retrospective nature and high variability. Currently, functional and structural clinical tests suffer from the need for analysis and the significant time delay required to establish disease progression and response to treatment. So does the progress of the ganglion cell loss over that time. There is, therefore, a significant unmet need for glaucoma-related biomarkers to improve clinical trials, both for early diagnosis and detection of disease progression. Biomarkers then present a possibility to provide information that will eventually affect the decisions to be taken clinically (BEYKIN et al., 2021).

Combining new technologies, such as imaging methods (OCT) and clinical evaluation based on parameters, is one way to assess glaucoma's criticality continuously. Optical coherence tomography provides relevant information such as cup-to-disc ratio (CDR), cup volume, rim area, and rim volume. Evaluations carried out by OCT demonstrated that regions such as Bruch's membrane opening (BMO) for the internal limiting membrane (ILM) are correlated with the thickness of the RFNL. As well as the posterior LC, a histopathological mark of glaucoma deformation was demonstrated in the deep part of the ONH structure. This information, made available by the imaging method, allowed the correlation of both regions, where Lavinsky and coworkers formulated the hypothesis that it would be helpful as a quantitative evaluation parameter of ONH remodeling in glaucoma (LAVINSKY et al., 2018). BMO and LC represent the legs of a right triangle and result in the evaluation by discovering the hypotenuse of the vertical region of the ONH cup. The results are presented as a morphometric biomarker.

Currently, there are several studies related to the structure of the retina, and many contain some computational uses related to the detection of regions, segmentation, and other analyses (SCHMIDT-ERFURTH et al., 2018). These studies are supported by the rapid advancement in scanning technologies, which allow more and more physicians to obtain images with a high amount of information, an important factor for any research that aims to analyze and extract data (SANDER et al., 2005). Thus, as mentioned earlier, methods such as using biomarkers support the diagnosis at different stages of a disease. Moreover, the use of several areas of computing allows for achieving new knowledge and analyses that would previously be difficult to carry out manually. Therefore, several studies (e.g., (PRANANDA et al., 2023), (BERENGUER-VIDAL et al., 2022), and (XUE et al., 2023)) foster new ways to expand knowledge and analysis regarding the support of glaucoma diagnosis.

1.2 Research question

As cited and shown previously, there are many fields of study using computer methods related to Computer Vision (CV). The main interest in these initiatives is to leverage the knowledge and the possibilities in obtaining ways to analyze some illnesses accurately. Biomarkers, when combined with different tools, could facilitate how a physician will make some decisions about the problem analyzed, but as shown by Beykin and coworkers, some biomarkers are carried out manually. The problem of doing tests as mentioned is time-consuming, difficult to obtain results faster, or even compare with other resources if possible.

Considering the presented context, the research question intends to answer the question: How can provide glaucoma diagnosis information from a new and specific biomarker using Artificial Intelligence techniques, based on Deep Learning (DL) and Computer Vision (CV)? With the computer methods available is possible to reproduce the findings in this biomarker and compare the results with those obtained by the physician?

1.3 Hypothesis

The specific biomarker was proposed by Lavinsky and coworkers, in a previous study (LAVINSKY et al., 2018) to measure the remodeling of ONH using the Pythagoras Theorem. The experiments were performed manually, showing the importance of the find and how could be important to analyze the remodeling of the ONH region.

Based on previous studies from other authors, computational methods could help to reach the results from the specific biomarker. Recently Deep Learning and Computer Vision have shown great performance in tasks of extracting information from patterns. The proposed study considered in use of some existent convolution neural network to perform the segmentation of an external and inner layer of the retina (Inner Nuclear Layer and Retinal Pigment Epithelium on the level of Bruch's Membrane Opening). The findings are used to track the extension and depth of the cup region portion using pixel-level calculations. Lastly, the Pythagoras Theorem is performed to obtain the results, based on the rules applied in the previous study (LAVINSKY et al., 2018).

1.4 Objective

The proposed work looks to do experiments with the real images shared by the physician measuring a specific region from the retina to support glaucoma diagnosis. We have the intent to recognize the portion of the cup region and after that, the length of remodeling of the region, based on rules available by Lavinsky and coworkers (LAVINSKY et al., 2018).

1.5 Specific objectives

For the proposed work we split our research and some specific goals. To reach the measurement of the cup region and find the hypotenuse, we proposed first to find two main regions of the retina in the OCT images, and second, to perform the calculation on the region. To do this we list some objectives:

- Identify the state of the art related to studies on glaucoma and retina.
- Identify computational methods for segmentation.
- Identify methods used in similar problems.
- Apply semantic segmentation on retinal regions.
- Evaluate results based on computational metrics.
- Define strategies to implement the proposed method in (LAVINSKY et al., 2018).
- Compare computational results with those presented by the specialist physician.

The objectives above were built in a pipeline, composed of different steps until the final results, to be compared with the present by the specialist. This flow considered methods already used in different research presented during the step of reading and finding related methods. It's important to emphasize that this pipeline and the biomarker analyzed are new too, and we don't consider a comparison with other research.

1.6 Contributions

As contributions, we highlight the following:

- Creation of an annotation dataset using patient images
  - Data annotation
  - Separation of layers
  - Organization in the folder structure to use in the neural network
- Definition of a model and computational flow to meet the necessities to support the automation of biomarker identification.
- Submission of an article in the Brazilian Congress of Ophthalmology (CBO[1]).
- Presentation of one article in the Conference on Intelligent Systems - BRACIS[2].

[1] https://www.cbo2023.com.br/cbo2023
[2] https://www.bracis.dcc.ufmg.br

1.7 Dissertation structure

The following sections will present the main theoretical foundations related to the research (Section 2), relevant related work (Section 3), materials, and methods that describe what was done during the research period (Section 4). In section five, the results are then shown, and finally, in section six, a conclusion is presented.

2 BACKGROUND

In this section, the supporting themes for this research will be presented. Section 2.1 presents concepts related to ocular structure, disease, and diagnostic support devices. Section 2.2 presents the main concepts of the computational methods used in the research.

2.1 Ophthalmology

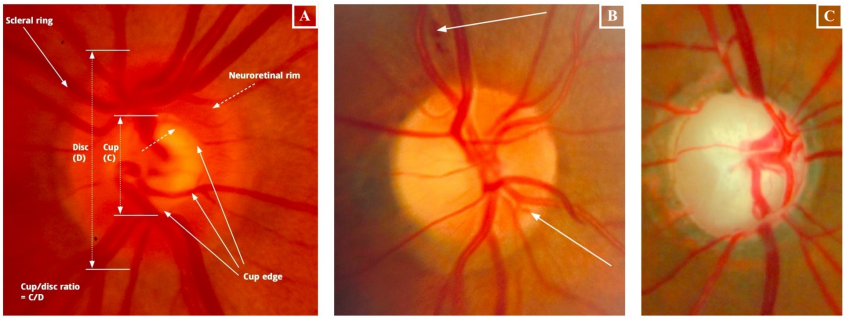
Ophthalmology is the branch of medicine dealing with the eye's health. This section presents some concepts related to the goals of research showing concepts of the anatomy of the retina, components from the eye, and the relationship with glaucoma. Further, an explanation of OCT also is presented.

2.1.1 Optic nerve head

The ONH can be examined using a direct ophthalmoscope, an indirect ophthalmoscope, or a posterior pole lens with a slit lamp. The ONH, or optic disc (OD), is a round/oval 'plughole' down which more than a million retinal nerve fibers descend through a sieve-like sheet known as the lamina cribrosa. The retinal nerve fibers are then bundled together behind the eye to form the optic nerve which then continues towards the brain. As the nerve fibers approach the edge of the disc they pour over the scleral ring (which marks the edge of the disc) and then down its inner surface. The dense packing of nerve fibers just inside the scleral ring is visualized as the neuroretinal rim. The cup is the area central to the neuroretinal rim. The cup edge (where it meets the neuroretinal rim) is best seen by the bend in small and medium-sized blood vessels as they leave, or descend into, the cup (BOURNE; KHATIB, 2023). Still, according to Bourne and Khatib, some of the characteristics of a glaucomatous ONH are generalized/focal enlargement of the cup. (Note that the cup always appears smaller when viewed monoscopically than in stereo); Disc hemorrhage (within one disc diameter of ONH); Thinning of the neuroretinal rim (usually at the superior and inferior poles). The figure 1 shows the difference between non-glaucomatous and glaucomatous eyes.

The optic nerve head is the structure in the posterior ocular fundus that allows the exit of the retinal ganglion cell axons and the entry and exit of the retinal blood vessels (WANG; PANDA-JONAS; JONAS, 2021). The canal consists anatomically of three layers: Bruch's membrane opening, the choroidal layer, and the layer of the peripapillary scleral flange, which is covered by the lamina cribrosa (figure 2). The peripapillary scleral flange serves as the biomechanical anchor of the lamina cribrosa to the sclera and is separated from the lamina cribrosa by the intertwined peripapillary border tissue of Elschnig. The choroidal lamella is separated from the intrapapillary compartment by the peripapillary border tissue of Jacoby, which ends at Bruch's membrane (BM). The Jacoby tissue continues into the Elschnig tissue and further on into the

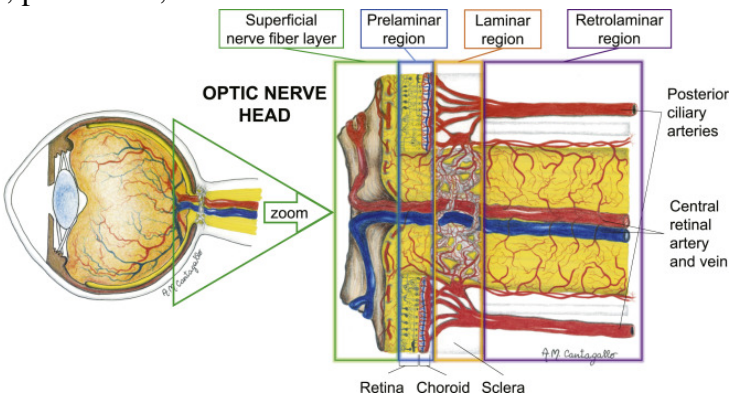
Figure 1 – Normal optic nerve head (A), Glaucomatous optic neuropathy: splinter hemorrhages (B) and Advanced glaucoma (C).



Source: (BOURNE; KHATIB, 2023).

optic nerve pia mater (ZHANG et al., 2019). The limits of BM are shown in the figure 3. Still, according to Wang and coworkers, the peripapillary scleral flange that is covered by the lamina cribrosa allows the entry and exit of the retinal ganglion cell axons and retinal blood vessels and simultaneously mostly preventing the exchange of fluid between the vitreous cavity on its inner side and the optic nerve and the retrobulbar cerebrospinal fluid (CSF) space on its outer side. The lamina cribrosa also serves as a pressure barrier between the intravitreal compartment with the pulsating intraocular pressure (better termed “transcorneal pressure difference”) and the retrobulbar compartment with the pulsating CSF pressure.

Figure 2 – From an anatomical point of view, the ONH canal can be regarded to consist of three regions: superficial, prelaminar, and laminar.

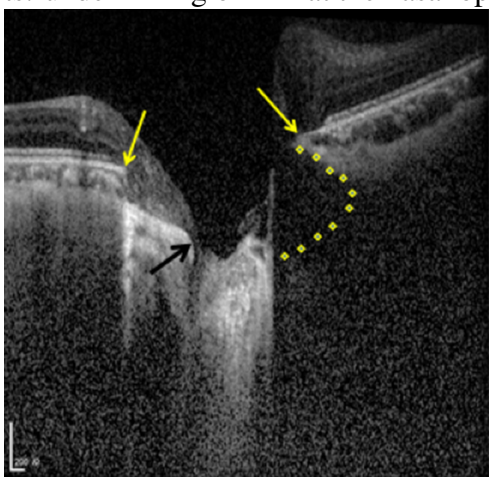


Source: (PRADA et al., 2016).

2.1.1.1 Bruch’s membrane opening

Bruch’s membrane opening, which remains stable with the progression of glaucoma, is an important biomarker in the optic nerve head in Spectral-Domain OCT (SD-OCT). It is defined as the termination of the retinal pigment epithelium (RPE) layer. The BMO points in SD-OCT represent the true position of the optic disc. Thus, the disc margin can also be identified by align-

Figure 3 – Yellow arrows: ends of BM; between left yellow arrow and black arrow: parapapillary gamma zone; yellow dots: undermining of BM at the nasal optic disc border.



Source: (ZHANG et al., 2019).

ing the color fundus image and the SD-OCT volumes. Moreover, BMO plays an indispensable role in lamina cribrosa surface depth measurement, which is also an important parameter for measuring the progression of glaucoma. Hence, an accurate and robust detection method of BMO is needed to measure those structural parameters automatically (CHEN et al., 2019).

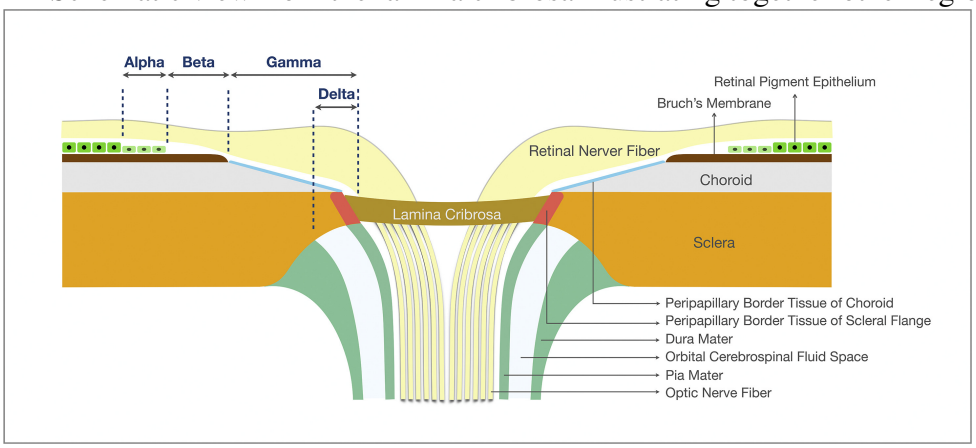
2.1.1.2 Lamina Cribrosa

Lamina cribrosa allows the entry and exit of the retinal ganglion cell axons and retinal blood vessels and simultaneously mostly prevents the exchange of fluid between the vitreous cavity on its inner side and the optic nerve and the retrobulbar cerebrospinal fluid (CSF) space on its outer side. The lamina cribrosa also serves as a pressure barrier between the intravitreal compartment with the pulsating intraocular pressure (IOP). The figure 4 shows a schematic view from the region, together with other components.

Lamina cribrosa, also called lamina cribriform, is a porous structure with a sieve-like appearance, located in the scleral foramen, formed by a weft of collagen fiber bundles from the inner layers of the sclera. This has the least structural integrity of any region of the sclera, being crossed by the axons of the RGC and the central retinal artery and vein to form the optic nerve (ON), and has been considered as an anatomic site of glaucomatous ON injury. The analysis of the LC anatomy *in vivo* is of great interest in the elucidation of factors related to the development and progression of glaucoma. Finally, with the advent of enhanced depth imaging OCT (EDI-OCT) and swept-source OCT (SS-OCT), it became possible to evaluate LC details *in vivo* in both normal and glaucomatous eyes.

The LC is a complex three-dimensional fenestrated mesh formed by a weft of bundles of collagen fibers from the innermost layers of the sclera and vascular capillaries that provide structural and nourishing support to the laminar region cells and the RGC axons that cross

Figure 4 – Schematic view from the lamina cribrosa illustrating together other regions too.



Source: (WANG; PANDA-JONAS; JONAS, 2021).

the structure. Located in the scleral foramen, the LC presents approximately one-third of the thickness of the sclera in the scleral canal (1–1.5 mm thick).

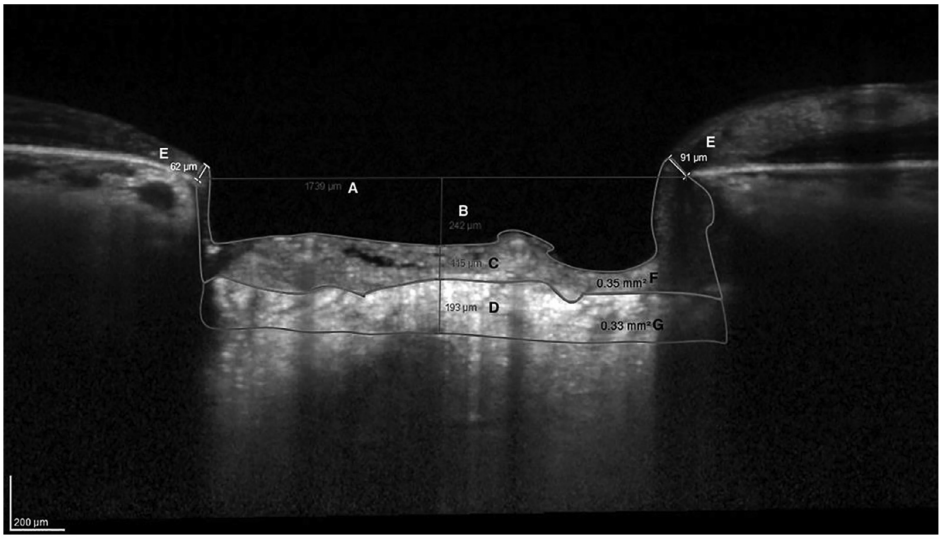
With the uniqueness of its anatomy and location, the LC helps to preserve the pressure gradient between the intra and extra-ocular spaces. From the biomechanical point of view, the structure represents the region of maximum scleral fragility, thus configuring the site of greatest vulnerability of the axons in their path from the retina to the cerebral cortex. That leads to a persistent remodeling in response to physiologic and pathophysiologic stimulus, which can manifest as focal laminar defects and a more generalized survey of ophthalmology changes in the thickness of the laminar bundle and the pores morphology.

Also, it is believed that the LC structural changes compromise the integrity of the axons directly by axonal compression or indirectly as a consequence of impairment of the mechanical or nutritional support of LC glial and capillary cells (ANDRADE et al., 2022). The LC parameters analyzed with the EDI-OCT are listed in the figure 5.

2.1.2 Structure of the retina

The retina is a thin (over 0.25 mm) layer of neural tissue that lines the back of the eye. It is a true part of the brain displaced into the eye during development. In addition to the light-sensitive photoreceptor cells. The retina is a relatively well-understood part of the central nervous system. Although it has been recognized to consist of five major classes of neurons, it is evident that each class of neurons consists of many subtypes of cells. The neurons are organized into three cellular (nuclear) layers, which are separated by two synaptic (plexiform) layers. Virtually all the junctions (synapses) between the retinal neurons are made in the two synaptic layers, and all visual information passes across at least two synapses, one in the outer plexiform layer and another in the inner plexiform layer (DOWLING, 2002). Clear images pass through the cornea, aqueous humor, lens, and vitreous humor, reaching the retina in focus,

Figure 5 – ED-OCT B-scan showing each optic nerve head parameter evaluated. A: Bruch's membrane opening (BMO); B: cup depth; C: prelaminar neural tissue (PLNT) thickness; D: lamina cribrosa (LC) thickness; E: BMO-minimum rim width; F: PLNT area; and G: LC area.



Source: (ANDRADE et al., 2022).

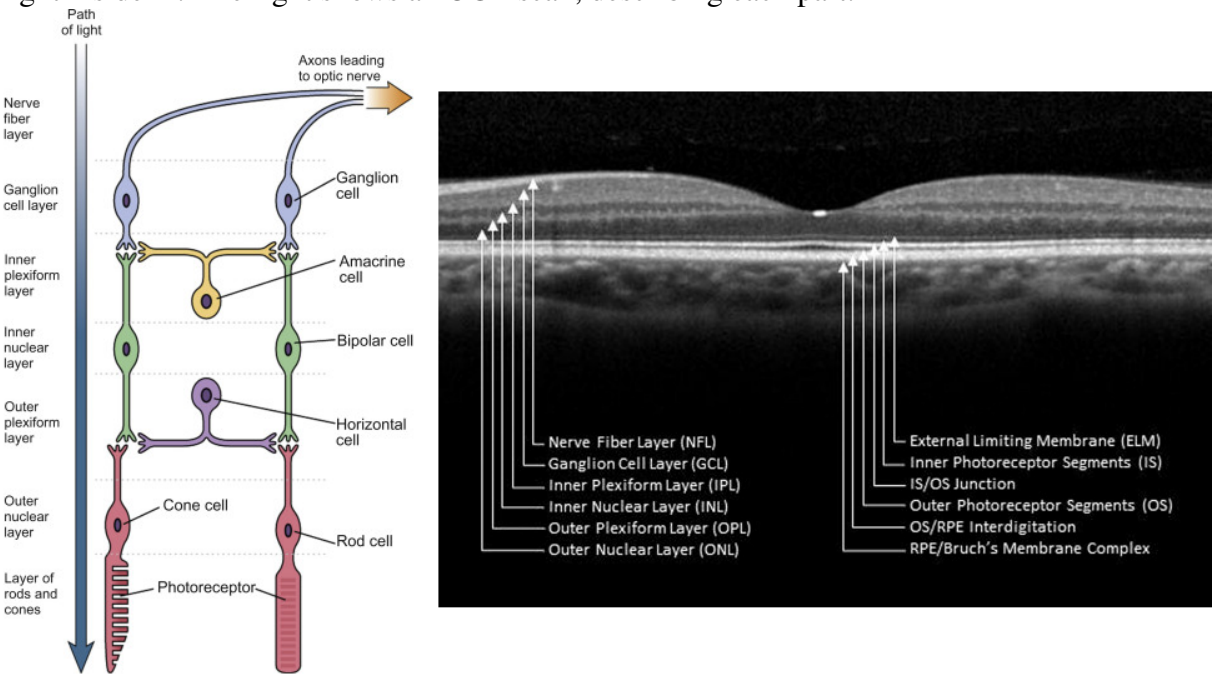
which is prepared to receive information through retinal photoreceptors (Figure 6, left). The neural signals must then pass in the reverse direction to the inner layer of the retina before being gathered and flowing into the optic nerve, through which they leave the eye (CARLSON, 2019).

The general organization of the neural retina is complex, with layers and membranes as described in the figure 6, right. The neural retina consists of the internal limiting membrane, nerve fiber layer (NFL), ganglion cell layer, inner plexiform layer (IPL; synaptic processes between bipolar cells and ganglion cells), inner nuclear layer (INL; bipolar nuclei, horizontal, amacrine, and Muller cells), outer plexiform layer (OPL; synaptic processes between bipolar cells and photoreceptors), outer nuclear layer (ONL; rod and cone nuclei), the outer limiting membrane (OLM; formed by cell junctions between photoreceptors and terminal optical processes of Muller cells) and photoreceptor layer (PR) (rod and cone cells) (SCHWARTZ et al., 2014).

2.1.3 Glaucoma

Glaucoma has been called the "silent thief of sight" because it can cause vision loss without any noticeable symptoms until the disease has progressed quite far (PATEL, 2023). The illness involves the progressive loss of retinal ganglion cells and characteristic changes in the tissue at the neuroretinal edge of the optic nerve head, accompanied by constriction of the visual field (MCMONNIES, 2017). It is also considered that the increase in intraocular pressure is a modifiable risk factor (MANTRAVADI; VADHAR, 2015), as shown in figure 7. The factors

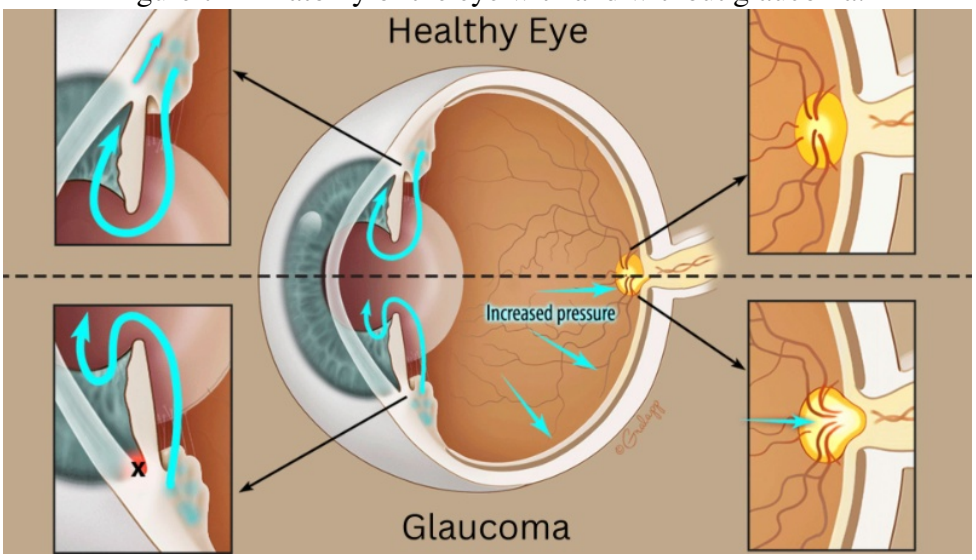
Figure 6 – Anatomy of the retina - On the left is a schema of layers, showing the direction of light inside in. The right shows an OCT scan, describing each part.



Source: Left: (CARLSON, 2019). Right: (SCHWARTZ et al., 2014).

that cause the disease are most often linked to advancing age, high intraocular pressure, high myopia, or a family history of the disease (SCHUSTER et al., 2020).

Figure 7 – Anatomy of the eye with and without glaucoma.



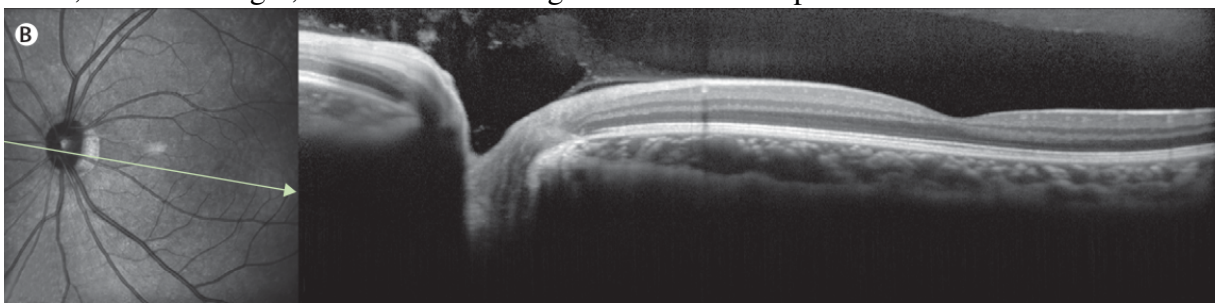
Source: (PATEL, 2023).

This combination of damage should be sufficiently characteristic to indicate the death of a substantial number of retinal ganglion cells in the inner retina and the loss of their axons in the optic nerve. At the optic disc, nerve fibers from the retinal ganglion cells pass out of the eye, most often leaving a central depression (known as a cup) that is lighter than the edge containing

these nerve fibers. Physicians can compare this cup to the overall size of the disc to establish the cup-to-disc ratio. As more retinal ganglion cells and their axons are affected by glaucoma, the ratio of the cup-to-disc progressively increases. The structural alteration most frequently recognized clinically by ophthalmoscopy and imaging devices is the cup's topographic deepening and widening (cupping), as shown in figure 8. This factor is used in optical coherence tomography to show individual layers and measure the thickness loss in the retinal ganglion cell layer and retinal nerve fiber layer in the zona (indicated by the green line in the left retinal image). This excavation consists of the loss of the axons of the retinal ganglion cells and the deformation of the connective tissues that support the optic disc. Structural loss of axons can also be detected by thinning the nerve fiber layer surrounding the disc consisting of retinal ganglion cell axons, by clinical examination, or by imaging methods such as optical coherence tomography or laser scanning polarimetry (QUIGLEY, 2011).

Also, according to Quigley and coworkers, the characteristic functional loss is assessed by measuring the sensitivity to light in the central 30° of vision. The international consensus panel definition specifies that glaucoma is present when three or more field test sites in a particular pattern are notably outside the bounds of normal variability and when, in the same eye, the cup-to-disc ratio is more significant than that seen in 97.5% of the general population. These criteria ensure that the structural finding is unlikely to be simply a typical variation in healthy individuals and that structural and functional damage has occurred.

Figure 8 – The structural alteration shown by OCT scan. The left image shows a C-Scan from ONH, and on the right, a B-Scan from the green arrow was captured.



Source: (QUIGLEY, 2011).

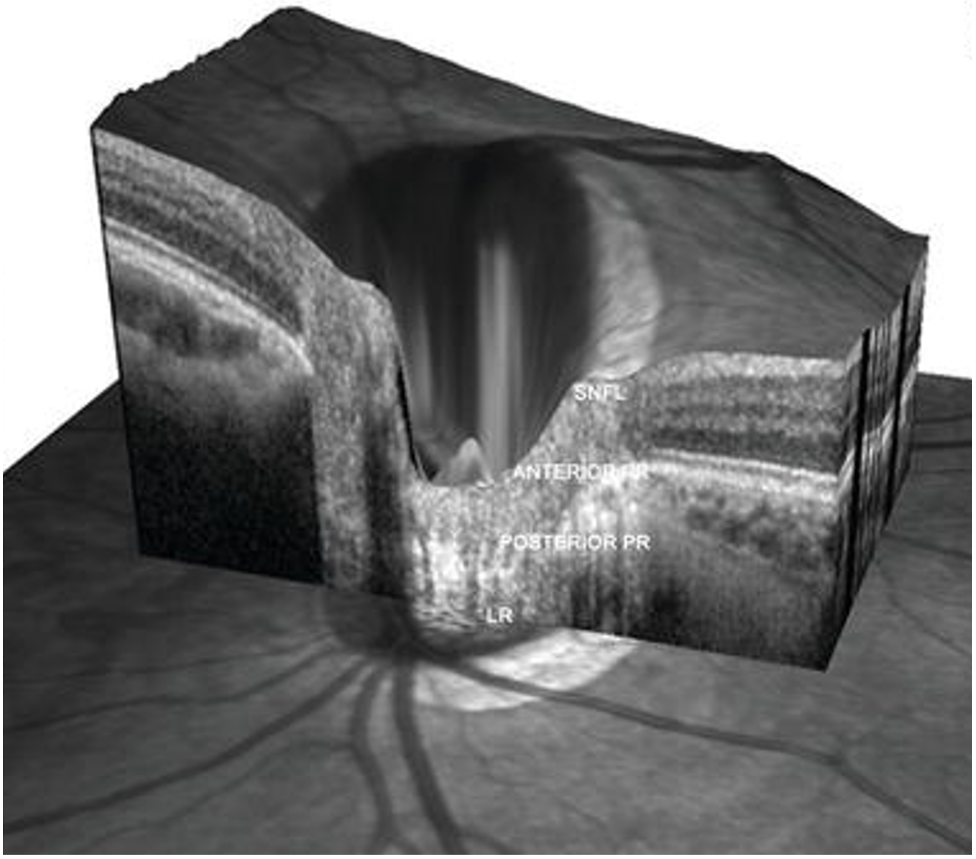
2.1.4 Optical Coherence Tomography

Optical coherence tomography is a reliable, fast, non-invasive, and easy-to-use device providing high-resolution images of retinal microstructures. In many cases, OCT-based images are referred to as biomarkers, helping to identify and study the disease process in various ocular disorders (PHADIKAR et al., 2017).

OCT can produce a real-time cross-sectional image of the object: a two-dimensional image in space (lateral coordinate, axial coordinate) and in some cases a three-dimensional map (Figure 9). When applied to the image of the human eye's retina, confocal microscopy is limited

by the compound effect of the low numerical aperture of the eye and any aberrations that may exist. In confocal microscopy, the lateral and axial resolutions are determined by the numerical aperture of the microscope objective. In OCT, axial resolution is mainly determined by the optical source. Therefore the retina of the human eye can be visualized with axial resolution at least 100 times better than that obtained using confocal microscopy. This high-depth resolution is achievable even when imaging eyes with aberrations and using small diameter beams (PODOLEANU, 2012).

Figure 9 – 3D map of ONH with spectral domain OCT.



Source: (SALAZAR et al., 2018).

There are two main OCT methods, time domain (TD) and spectral domain (SD) (DREXLER; FUJIMOTO, 2008). According to (PODOLEANU, 2012), SD-OCT eliminates the need for depth scanning, which in TD-OCT is usually performed mechanically. SD-OCT methods can be implemented in two formats: (i) spectrometer-based (SB) - based on a spectrometer - or (ii) using a tunable laser or a scanning source, known as a scan source (SS). Each format, SB or SS, has its own merits and shortcomings. The depth resolution achieved depends on the optical source bandwidth in TD-OCT and SD-OCT and the adjustment bandwidth in SB or SS.

There is a fundamental difference between the TD and SD methods. TD-OCT methods provide the reflectivity of a single point at depth when the light path difference is equal to zero (OPD = 0), while low-to-SD coherence interferometry and SD-OCT return the reflectivity

values for all points along the axial range at once. These configurations give SD methods superiority in terms of the acquisition rate (PODOLEANU, 2012).

Spectral-domain OCT uses light to capture high-resolution optical cross-sections of the retina, which can be assembled into three-dimensional volume images of living retinal tissue. OCT imaging is now a standard of care to guide the diagnosis and treatment of some of the leading causes of blindness worldwide (KERMANY et al., 2018). For this reason, the OCT has received several changes over the years, enabling much more accurate analyses today. In the first two versions, it was possible to observe general changes related to retinal thickness and the detection of fluids inside and behind the neurosensory retina. In the third version, intermediate layers of the sensorineural retina are only vaguely discernible (SANDER et al., 2005). The test speed and resolution of SD-OCT have improved when compared with TD-OCT. Scan speeds between 29,000 and 70,000 Hz and an axial resolution of up to 2 μm can be demonstrated. Other features used to improve image quality include image averaging, noise reduction, and eye-tracking systems could be included in some commercial devices. Lastly, swept-source OCT (SS-OCT) is the newest technology that can reach higher speed rates (up to 400 KHz) (LAVINSKY et al., 2018).

2.2 Computational Methods

2.2.1 Deep Learning

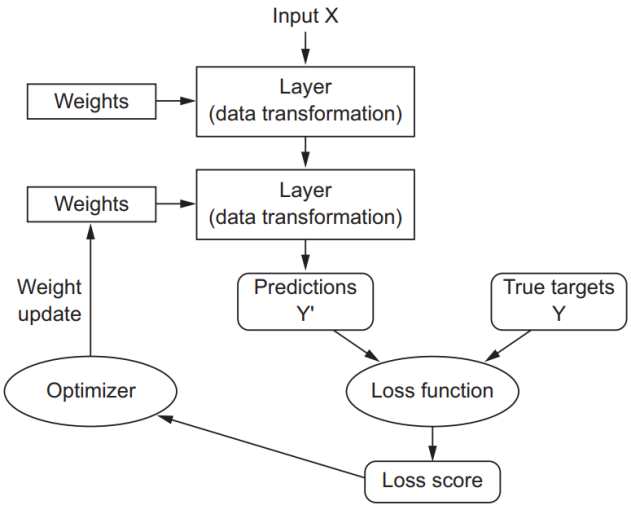
For decades, building a pattern recognition system or a learning system required multiple domains and expertise to design features that transform raw data into an adequate internal representation, from which the learning subsystem can detect or classify patterns in the input data. However, this became different with the introduction of several sets of methods, which allowed a computational model to receive real data (within the context of the model) and perform necessary representations to perform a task, such as classifying or detecting objects (LECUN; BENGIO; HINTON, 2015).

Artificial Intelligence (AI) comprises areas dedicated to tasks such as perception, logic, and learning. Within this context and encapsulation, there are very important and widely used subsets, such as Machine Learning (ML), which is a computational technique capable of processing raw data (LECUN; BENGIO; HINTON, 2015) and endowing applications with improvements in their performance for future tasks based on experiences and observations (RUSSEL; NORVIG, 2020). It is also important to mention Deep Learning, a Machine Learning method based mainly on neural networks (NN), in which several layers of NN are used to create models that can find the best way to perform a task on their own, based on a set of data (ALZUBAIDI et al., 2021).

The amount of layers contributing to the model is called the model depth. In modern applications, involving more than hundreds of successive layers of representation is common – and

they can learn when exposed to training data. Therefore, it can be said that DL is a framework for learning representations through data, and this structure works with a training model in repetition, which repeats a certain number of times, in search of prediction on the available data (Figure 10). This process is possible with the use of related components (CHOLLET, 2018). Finally, it is essential to mention that there are different types of neural networks, such as recurrent neural networks (RNN) or convolutional neural networks (CNN) (ALZUBAIDI et al., 2021).

Figure 10 – Deep Learning flow.

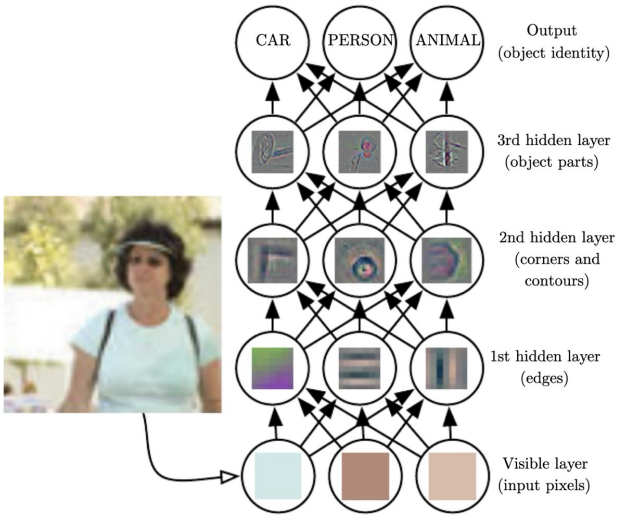


Source: (CHOLLET, 2018).

2.2.1.1 Layers

From the point of view of a computer, it is difficult to understand the meaning of a raw input image, such as this image represented as a collection of pixel values. The function mapping from a set of pixels to an object identity is very complicated. Deep learning resolves this difficulty by breaking the desired complicated mapping into a series of nested simple mappings, each described by a different layer of the model. The input is presented at the visible layer, so named because it contains the variables that we can observe. Then a series of hidden layers extracts increasingly abstract features from the image. These layers are called “hidden” because their values are not given in the data; instead, the model must determine which concepts are useful for explaining the relationships in the observed data (GOODFELLOW; BENGIO; COURVILLE, 2016). Given the pixels, a respective layer can extract some information, and after that also transmit it to another layer. Finally, this description of the image in terms of the object parts it contains can be used to recognize the objects present in the image.

Figure 11 – Illustration of a deep learning model.



Source: (GOODFELLOW; BENGIO; COURVILLE, 2016).

2.2.1.2 Optimizer

Optimizers are algorithms or methods used to minimize an error function (loss function) or to maximize the efficiency of production. It depends on the model's learnable parameters i.e. Weights and Biases. Optimizers help to know how to change weights and learning rate of neural networks to reduce the losses.

In Machine Learning it's impossible to calculate the perfect weights for a neural network; there are too many unknowns. Instead, the problem of learning is cast as a search or optimization problem and an algorithm is used to navigate the space of possible sets of weights the model may use to make good or good enough predictions.

Typically, is trained using the stochastic gradient descent optimization algorithm, and weights are updated using the back-propagation of error algorithm.

The "gradient" in gradient descent refers to an error gradient. The model with a given set of weights is used to make predictions and the error for those predictions is calculated.

The gradient descent algorithm seeks to change the weights so that the next evaluation reduces the error, meaning the optimization algorithm is navigating down the gradient (or slope) of the error (GOODFELLOW; BENGIO; COURVILLE, 2016).

2.2.1.3 Loss Function

Loss, in Machine Learning, helps to understand the difference between the predicted value and the actual value. The Function used to quantify this loss during the training phase in the form of a single real number is known as the "Loss Function".

The terms cost function and loss function could considered analogous, but normally loss

function is used when we refer to the error for a single training example, already cost function is used to refer to an average of the loss functions over an entire training data.

Neural networks are trained using stochastic gradient descent and require a loss function when designing and configuring your model.

There are many loss functions to choose from and it can be challenging to know what to choose, or even what a loss function is and the role it plays when training a neural network.

2.2.2 Convolutional Neural Network

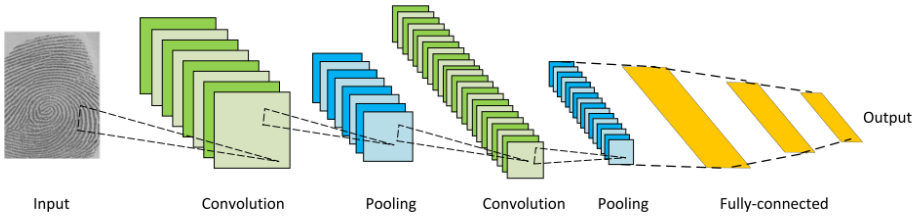
Convolutional Neural Networks are among the most successful and widely used architectures in the Deep Learning community, especially for computer vision tasks. They were initially proposed by Fukushima in his seminal article on the "Neocognitron" (FUKUSHIMA, 1980). Later, CNNs with shared weights between temporal receptive fields and backpropagation training for phoneme recognition were introduced (WAIBEL et al., 1989). Years later, a CNN architecture for document recognition was developed (LECUN et al., 1998). This architecture is designed to process various types of data, especially two-dimensional images. It is directly inspired by the brain's visual cortex, which consists of a hierarchy of two basic cell types: single cells, which react to primitive patterns in sub-regions of visual stimuli, and complex cells which synthesize information from simple cells to identify more intricate shapes (MARY et al., 2020). CNNs are applied to mimic three key ideas: local connectivity, location invariance, and transition invariance (MIN; LEE; YOON, 2017).

The basic structure of CNNs (Figure 12) consists of convolution layers, where a kernel (or filter) of weights is used to extract features; non-linear layers, which apply an activation function to feature maps (usually on elements) to allow modeling of non-linear functions across the network; and clustering layers, which replace a small neighborhood of a feature map with some statistical information (e.g., average, maximum) about the neighborhood and reduce the spatial resolution. The main computational advantage of CNN is that all receptive fields in a layer share weights, resulting in significantly fewer parameters than fully connected neural networks (MARY et al., 2020).

To be specific, to build a CNN model, four components are typically needed. Convolution is a pivotal step for feature extraction. The outputs of convolution can be called feature maps. When setting a convolution kernel with a certain size, we will lose information in the border. Hence, padding is introduced to enlarge the input with zero value, which can adjust the size indirectly. Besides, for the sake of controlling the density of convolving, stride is employed. The larger the stride, the lower the density. After convolution, feature maps consist of a large number of features that are prone to causing overfitting problems. As a result, pooling (a.k.a. down-sampling) is proposed to obviate redundancy, including max pooling and average pooling. The procedure of a CNN is shown in figure 13. Moreover, there exist a variety of awesome convolutions, such as Separable convolutions, group convolutions, multi-dimensional convo-

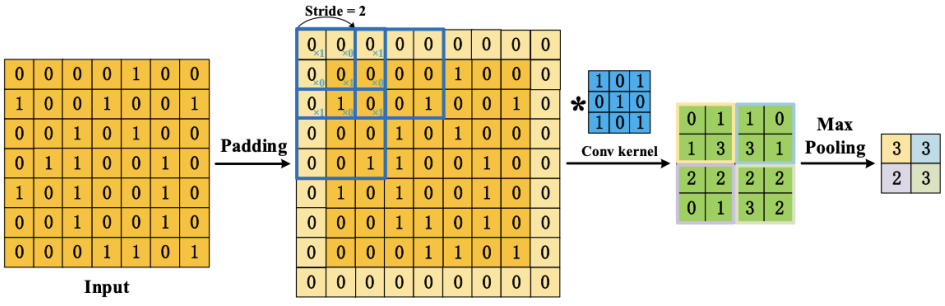
lutions, deformable convolutions, and convolution kernels to perceive larger areas, as dilated convolutions (LI et al., 2022).

Figure 12 – CNN architecture.



Source: (LECUN et al., 1998).

Figure 13 – Two-dimensional CNN.



Source: (LI et al., 2022).

2.2.3 Image Segmentation

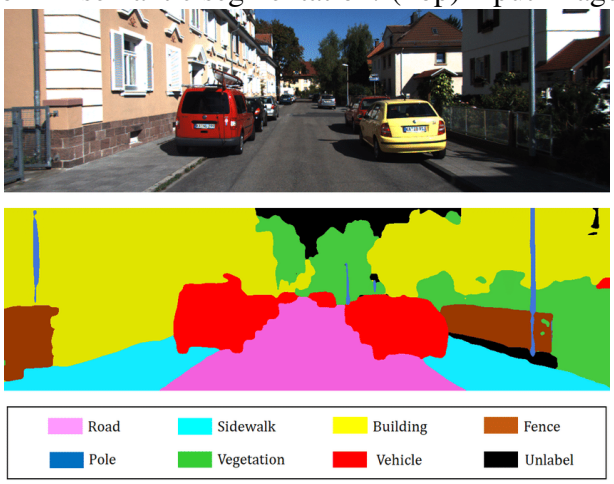
Image segmentation is one of the oldest and most widely studied problems in computer vision. Early techniques used region splitting or merging, corresponding to splitting and clustering algorithms in the clustering literature. Newer algorithms often optimize global criteria such as intra-region consistency and boundary lengths across regions or dissimilarity (SZELISKI, 2010).

The segmentation can be seen as a problem of classifying N elements in K regions, such that K $\leq$ N; $U_k^{i=1} k_i$ = image; $k_i \cap k_j = \emptyset$, if i $\neq$ j; and elements in the same K region have properties similar to each other and different from the properties of elements in other regions. Thus, segmentation can also be modeled as a combinatorial optimization problem, in which an optimal region is sought according to some similarity criterion between the elements of the same region (RONCERA, 2005).

Image segmentation is an essential component in many visual comprehension systems. It involves partitioning images (or video frames) into multiple segments or objects. It plays a central role in various applications, including medical image analysis (e.g., extracting tumor

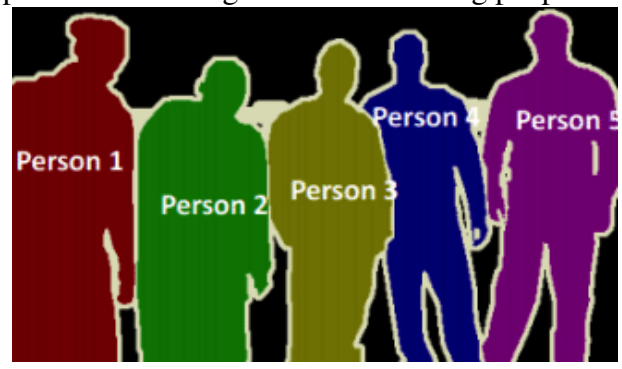
boundaries and measuring tissue volumes), autonomous vehicles (e.g., surface navigable and pedestrian detection), video surveillance, augmented reality, and other possible applications. Image segmentation can be formulated as a pixel classification problem with semantic labels (semantic segmentation), where it performs pixel-level labeling with a set of object categories (e.g., human, car, tree, sky) for all the image pixels, as shown in figure 14. Alternatively, it is possible to extend to an instance segmentation (figure 15), detecting and delineating each object of interest in the image (e.g., selecting people as a single individual) (MARY et al., 2020).

Figure 14 – Example of 2D semantic segmentation: (Top) input image (Bottom) prediction.



Source: (JEONG; YOON; PARK, 2018).

Figure 15 – Example of instance segmentation selecting people as single individuals.



Source: (VARATHARASAN et al., 2020).

In addition to the class of classic segmentation techniques, such as thresholds, clustering methods, edge detection, histograms, and others, there is a class based on AI techniques, such as some works based on DL. These are Fully Convolutional Neural Networks, models based on Encoder-Decoder, convolutional models with graphical models, pyramid and multi-scale networks, and R-CNN for segmentation (MARY et al., 2020).

3 RELATED WORK

This section presents approaches using deep learning or image processing applied to problems related to glaucoma and retinal regions problems.

To identify relevant articles, keywords were applied to relate the computational topics with the medical approach. Subsequently, refinements were made to extract the studies that focused on using deep learning to segment the layers of the retina without focusing on the studied deformities (glaucoma or not). Finally, more specific searches were performed to relate the desired Machine Learning approach to the prospective medical approach.

Therefore, some databases were selected to search for articles. Among them are Google Scholar, IEEE, Scopus, and PubMed. Searches were carried out using keywords such as: "optic disc"; "optic nerve head"; "optical coherence tomography"; "segmentation of retinal layers"; "glaucoma diagnosis"; "retinal layer cup area"; "circumpapillary retinal nerve fiber layer (cpRNFL)"; "machine learning"; "deep learning"; and "convolutional neural networks". The combinations also varied, depending on the databases accessed and the results obtained. Publication dates were also considered when the search related to computing and clinical study, prioritizing articles up to eight years old to be used as study references[1]. The open database *Papers with Code*[2] was also used as a sequence of research carried out in secure databases. The Paper With Code database differs from the previous ones due to the possibility of making codes and databases available freely for use in other studies.

3.1 Clinical investigations and support on diagnosis

Eye diseases are still very vast fields of study, and the diagnosis often differs depending on the stage of the disease, as well as the structural conditions of the patient. The auxiliary equipment provides a vast of relevant information to medical specialists. However, factors and problems can still be studied for a more accurate diagnosis.

Clinical data, observations, and diagnoses are often relevant in treating a patient. In addition, with technological advances, it seeks to expand knowledge and provide new means. Diagnosis through OCT images has been the focus of a wide range of researchers for many years, but in recent years with the improvement of the device's technology, it has been possible to have results more accurate. In the last three decades, OCT has revolutionized ophthalmic diagnosis, therapy monitoring, and guidance (LEITGEB et al., 2021). Motion contrast-based angiography, cellular level retinal visualization, visible light OCT for oximetry and unprecedented retinal layer detection, functional and contrast-enhanced extensions, and artificial intelligence (AI)-enhanced performance contributed to this success. Most of these superb technological developments can be directly translated to the original motivation and idea of OCT: to enable optical biopsy. Ac-

[1] Clinical investigations and theory do not consider publication date.
[2] https://paperswithcode.com/

cording to Leitgeb and coworkers, aside from OCT angiography, no other functional or contrast-enhancing OCT extension has accomplished comparable clinical and commercial impacts. But some more recently developed extensions, e.g., optical coherence elastography, dynamic contrast OCT, optoretinography, and artificial intelligence-enhanced OCT are also considered with high potential for the future.

Side by the options available using the old or new technologies, extracting information from the devices it's crucial to extend studies and develop new ways of analyzing and have directions on treatments of patients with eye disease. Following the study presented in (BEYKIN et al., 2021), biomarkers are defined as a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention. The relevance of a biomarker's ability to provide information that will eventually affect clinical decisions and endpoints. For glaucoma, the currently most used clinical biomarker outside of IOP screening, i.e., visual field testing, has been extensively reviewed elsewhere with important advances in instrumentation, testing protocols, and analyses. The introduction of validated biomarkers would allow for prompt intervention in the clinic to help with prognosis prediction and treatment response monitoring, as well as to conduct early-phase clinical trials more quickly, for example, to test potential medicines (FEA et al., 2023).

Studies using OCT-generated are vital and a field of study over the years. Many methods were developed to analyze or to improve some identification of the structure of the retina region. Since the initial devices, two kinds of scans have been produced, the B-scan and the C-scan. The B-Scan allows a view of layers of the retina, making it possible to analyze other components, such as the lamina cribrosa. But also, it's possible to relate with the C-Scan, where the view is about the optic nerve head with cup and disc area, blood vessels, and other components.

Several types of research focus on parts like the inner retina, cup area, deformation of layers or displacement from lamina cribrosa, and other issues that can represent structural disorganization. Most of the methods presented were using the OCT technology, to gather retinal imaging. [(SANDER et al., 2005), (XU et al., 2010), (PODOLEANU, 2012)]. Since early 2000, the studies presented some kind of observation of different aspects. Sander and coworkers present observations to improve tomograms and allow the identification of structural disorganization and attenuation of the inner retina. Xu and coworkers proposed a comparison between different types of imaging extractions to automate the volumetric measurements of the ONH using a photogrammetric algorithm. In (SHIN et al., 2014) presented an investigative approach to the effects of different types of optic disc damage on diagnostic performance by comparing the pattern of visual field defect and parameters extracted from the OCT in early cases of glaucoma. Already (KIM et al., 2013) shows the importance of monitoring the changes of the structures about acute and chronic elevation of intraocular pressure and the relationship with lamina cribrosa, using OCT images.

In order of OCT-based imaging, other studies proposed biomarkers to detect and classify the disease in early stages or measure the severity (qualitatively and quantitatively). Phadikar and

coworkers (PHADIKAR et al., 2017) reviewed several biomarkers based on images from OCT equipment and the potential to be considered as a substitute outcome for diabetic retinopathy, age-related macular degeneration, retinitis pigmentosa, and vitreomacular interface. In (KHALIL et al., 2018) was described a process for extracting and contouring the inner limiting membrane layer. In calculating the disc diameter, the endpoints of the RPE layer were used to define the disc margin. Other studies could be mentioned as presented by Fu and coworkers (FU et al., 2015), through a detection method via low-level reconstruction to perform the identification of the optic disc in OCT slices. The method was applied to discover the boundary of the optic disc that appears at the end of the retinal pigment epithelium. Finally, MacCormick and coworkers (MACCORMICK et al., 2019) show how the detection of glaucoma analyzing color photographs to quantify optic nerve rim consistency around the whole disc at 15-degree intervals, providing a profile of the cup/disc ratio, in contrast to the vertical cup/disc ratio in common use.

In the scenario of clinical practice, ophthalmologists suspect glaucoma by analyzing optic nerve head anatomy, cup-to-disc ratio, retinal nerve fiber layer thinning, presence of disc hemorrhages, or presence of parapapillary atrophy. The diagnostic process could be challenging considering the extreme variance of these parameters. For this, AI algorithms can extract various optic disc features (FEA et al., 2023). Combining studies that try to describe or even predict new regions, based on OCT-imaging, the use of artificial intelligence is expanding rapidly and allowing a more sophisticated and self-programming way to use machines in automatic data analysis. The advanced network architectures developed over the previous years also allowed to use of more complex data sources. AlexNet, VGGNet, and ResNet could be considered as networks that opened some doors to other research that coming after in different fields of study. [(KRIZHEVSKY; SUTSKEVER; HINTON, 2012), (SIMONYAN; ZISSERMAN, 2015), (HE et al., 2015)]. For example, the last one, ResNet, makes use of shortcut connections to solve the vanishing gradient problem. We can cite Inception (SZEGEDY et al., 2014), which increases the network space from which the best network is to be chosen via training. Each inception module can capture salient features at different levels.

The popularization of these works allowed many others, not only used for detection tasks but also for segmentation tasks. Fully Convolutional Networks (FCN) (SHELHAMER; LONG; DARRELL, 2015) was one of them, mainly for semantic segmentation. They employ solely locally connected layers, such as convolution, pooling, and upsampling. Avoiding the use of dense layers, which means fewer parameters (making the networks faster to train). It also means that a FCN can work for variable image sizes given all connections are local. The network consists of a downsampling path, used to extract and interpret the context, and an upsampling path, which allows for localization. FCNs also employ skip connections to recover the fine-grained spatial information lost in the downsampling path. Almost together, U-Net (RONNEBERGER; FISCHER; BROX, 2015) was initially designed for biomedical image segmentation but has applications in various domains due to its flexibility and efficacy, summarized as a super pow-

erful tool. U-Net is primarily designed to work with image data. In the context of biomedical applications, U-Net has been a back-holding for complex microscopy images, CT scans, and MRI images. Furthermore, U-Net can be adapted to handle multi-channel images or even 3D volumes. Design as an U-shape architecture, which consists of a contracting path, followed by an expanding path. The contracting path extracts features from the input image and reduces its resolution, while the expanding path restores the original size of the image and produces the segmentation map. The combination of the two paths enables U-net to learn both global and local features and to achieve high accuracy in segmentation tasks.

The early detection of glaucoma could eventually preserve vision in affected people. However, due to its clinical history of being symptomatic only in advanced stages and when most of the RGCs are already compromised, it is crucial to introduce a tool to detect glaucoma in clinical practice in pre-symptomatic form automatically. Artificial intelligence algorithms such as U-Net and FCN can extract various optic disc features and automatically detect patterns of glaucoma from OCT imaging.

During the years following the development of U-Net, variations were being created, trying to solve some medical problems that the base architecture could not. Extensions like the one presented by Azad and coworkers used densely connected convolutions (AZAD et al., 2019). In Zhuang et al., it was possible to demonstrate the use of multiple U-Nets connected in the form of a chain for the segmentation of blood vessels of the optic disc (ZHUANG, 2019). Sedai and coworkers (SEDAI et al., 2021) went further, presenting a partially supervised form of retinal segmentation with OCT images, which consists of a so-called "student-teacher" approach applied to two datasets: the first being small and annotated; the second much more extensive and without annotations. A new multi-prediction guided attention network (MPG-Net) for automated segmentation of the retinal layer in OCT images was proposed too in (FU et al., 2021). The method consists of two main steps to strengthen the discriminate power of an FCN network in the form of U and perform the segmentation. The work presented by (GOPINATH; RANGREJ; SIVASWAMY, 2018) described a framework for performing segmentation by applying it to OCT images of normal and pathological eyes. In this case, a CNN was combined with Bidirectional Long Short-Term Memory (BLSTM).

More robust frameworks were also built, as presented by Roy and coworkers (ROY et al., 2017), featuring multilayer retinal segmentation and fluid pocket delineation in OCT imaging. The Fully Convolutional Neural Network architecture was used and adapted for semantic segmentation. Training and validation were performed using images from 10 patients and validated through a comparison with five state-of-the-art methods on segmentation of the same studied area. Still, it is possible to highlight the framework created for 3D segmentation (DEVALLA et al., 2020). Specifically, two sets of networks were developed: one to improve and harmonize the quality of images from three devices and another, named by the authors as *ONH-Net*, to perform 3D segmentation of 6 optic nerve head tissues.

Li and coworkers present a form of retinal segmentation (LI et al., 2021), focused on im-

ages of patients with glaucoma. For this, an architecture of convolutional neural networks was developed in which it was possible to identify the nine layers of the retina and the optic disc for the final segmentation. The following were applied: an optic disc detection neural network, a neural network for the segmentation of retinal layers, and a fusion module. A score of 82% and 83% was achieved in terms of the average data score and pixel accuracy, respectively. Zang and coworkers proposed the development of two separate neural networks to perform automated segmentation of the optic disc boundary and segmentation of the peripapillary retinal layer (ZANG et al., 2019), where the final boundary of the peripapillary retina was calculated based on the prediction and gradient map, using an algorithm of multiple weights graph search. The study carried out was based on the use of a classic neural network architecture, the U-Net. After the composition of some variations in the architecture, an accuracy for images of the healthy optic disc and with glaucoma of 84.11% and 83.53%, respectively, was reached.

Focus on a specific area from the retina also it's proposed by some authors. Chen and coworkers (CHEN et al., 2019) used a strategy for BMO detection using region segmentation with a fixed detection stage based on the U-Net. A qualitative and quantitative assessment of calcified drusen was studied in (LU et al., 2023), reporting an automated algorithm to segment and quantify calcified drusen on SS-OCT images, together with a correction method for the segmentation of the retina pigment epithelium. Support for the process of identification of blood vessels was proposed in (ALSAYAT et al., 2023) with a multi-layer preprocessing stage and a subsequent segmentation stage employing a U-Net with a multi-residual attention block.

3.2 Final discussion

Making decisions guided by some evaluation using computer methods is a vast field of study and improvements, but that already could be used as support in many cases. Since the first research, a big potential of powerful methods was shown, presenting good results and analysis from specific cases. However, it's important to mention that the last decision made must be by the physician who will monitor the evolution of the disease. Eye illness is an open field of study still, whereas the whole structure of the eye is passive of limitations due to the retrospective nature and high variability.

Discussions are presented by structural disorganization, i.e. deformation of layers. They address how could be possible to analyze these areas showing methods for collecting information from eyes with and without the disease, correlations between data provided by the equipment (not only OCT, but vision field tests and others), automatic tests to perform faster and accurate results allowing even a better treatment decision, and still open relationships between the structure and cells.

The application of AI together allows open new ways of analyzing the data imaging collected, since the equipment also has evolved and makes available images with more quality and metrics. Still difficult for physicians to identify micro problems in the structure during daily

routine, as fluids on inner layers, deformations, and more. The new version of OCT as the SD or SS allows faster captures and results but not all analyses are available during the process, which needs a further step for different studies. Many studies nowadays look for support using AI, which allows faster results and even search for new problems that before weren't possible.

Directing the discussion to the field of AI, it's clear that during the last years, an increase in research was done. The process of collecting data and transforming the information got easier in many cases. However, it's possible to notice that working with scans is a challenging task, due to the quality of the images that improve a lot but not enough to see with clarity the whole components (i.e. division of layers, micro problems, enhanced details, and more). The SD-OCT and SS-OCT capture slices from ONH with high quality in terms of complexity of anatomy, but still present some problems like enhanced quality in contrast, noise reduction, and understanding of the structure. As presented by (KHALIL et al., 2018), a workflow was created to remove noise before trying to identify the regions of interest in the OCT images used. Another important point mentioned is that segmenting medical images is a delicate process where the lines most of the time are thin, which could impact the process of training a model, resulting in a high loss or low accuracy. Some of the works presented some methods to improve the performance of neural networks. Recently, a study focused in to present a method for denoising medical image segmentation (CHEN et al., 2023) proposing a novel edge loss and a plug-and-play fine-tuning denoising module. Chen and coworkers proposed an approach to ensure that the models learn a sufficient number of features before mitigating the impact of noise. Incorporating this module ensures better performance as a preprocessing technique. Jang and coworkers also tried enhanced images in self-supervised denoising (JANG et al., 2023) expanding receptive fields and using dilated attention blocks (DABs) for global context incorporation.

Despite this problem, which impacts the performance of model training, building a network model with fewer parameters and low weights has been one of the points of interest in computational terms. Together, the use of small datasets also is a field of study, as presented by Chen and coworkers. The collection of data in some cases could be more challenging, due to the restrictions of study or terms.

Most of these topics and methods are used in the vast field of biomedicine. For eye illness, segmentation tasks are very helpful because open the opportunity to separate every component in the retina and identify some different behavior in the area. The use of this approach also helps limit regions, like cup/disc area. Further, the studies could apply some approach to analyze the area or perform identifications that could be transformed into biomarkers. The use of this task helps reach a post-processing goal, as presented by Rahamn and coworkers (RAHMAN et al., 2021), which finds the anterior LC curve boundary information and quantifies it from BMO and LC curve information using an image processing algorithm.

4 PROPOSED APPROACH

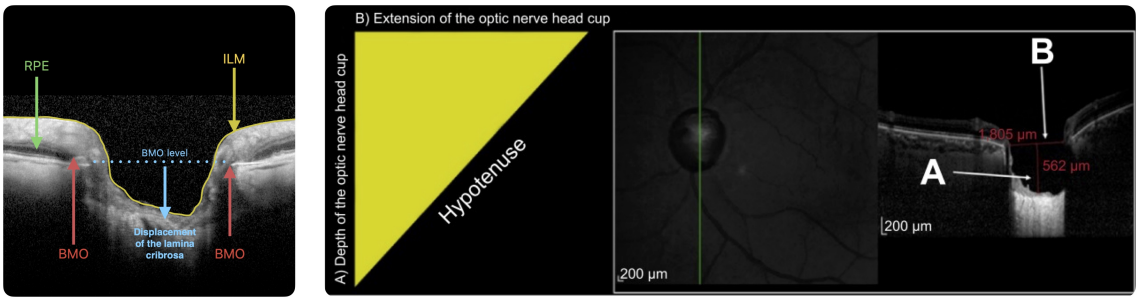
The research sought to contribute by exploring computational techniques to foster extracting information related to a new biomarker on the progress of glaucoma. In this chapter, the research procedure performed is described. Throughout this section, elements related to the understanding and functioning of a set of actions used for the preparation of data and computational architecture before the execution of experiments will be addressed.

Section 4.1 presents the main steps adopted in the proposed solution. The dataset developed is described in section 4.2.1 and section 4.2.2. In section 4.3.1 the CNN Architecture developed to perform image segmentation is described. The section 4.4 describes the biomarker measurement procedures.

4.1 Overview

Some studies demonstrate the importance of obtaining evaluation parameters (biomarkers). One of the ways observed was using the Pythagorean theorem (LAVINSKY et al., 2018) to assist in evaluating the vertical optic nerve head using SD-OCT enhanced depth imaging. The vertical B-scan with the largest cup seen in infrared is used for hypotenuse measurement. The depth of posterior displacement of the lamina cribrosa (LC), measured from the opening of Bruch's membrane level, and the length of the excavation between the BMO boundary and the ILM layer, form the sides of a right triangle, used in the calculation (Fig. 16 right). The discovery of the hypotenuse is a helpful parameter for the quantitative assessment of optic nerve head remodeling in patients with glaucoma. Therefore, the motivation of the work is associated with supporting the identification of the new biomarker, applying computational techniques for the measurement, and quantitative analysis of the identified method.

Figure 16 – Left figure shows each part of retina, in a B-Scan. Right figure shown the hypotenuse measurement (LAVINSKY et al., 2018).

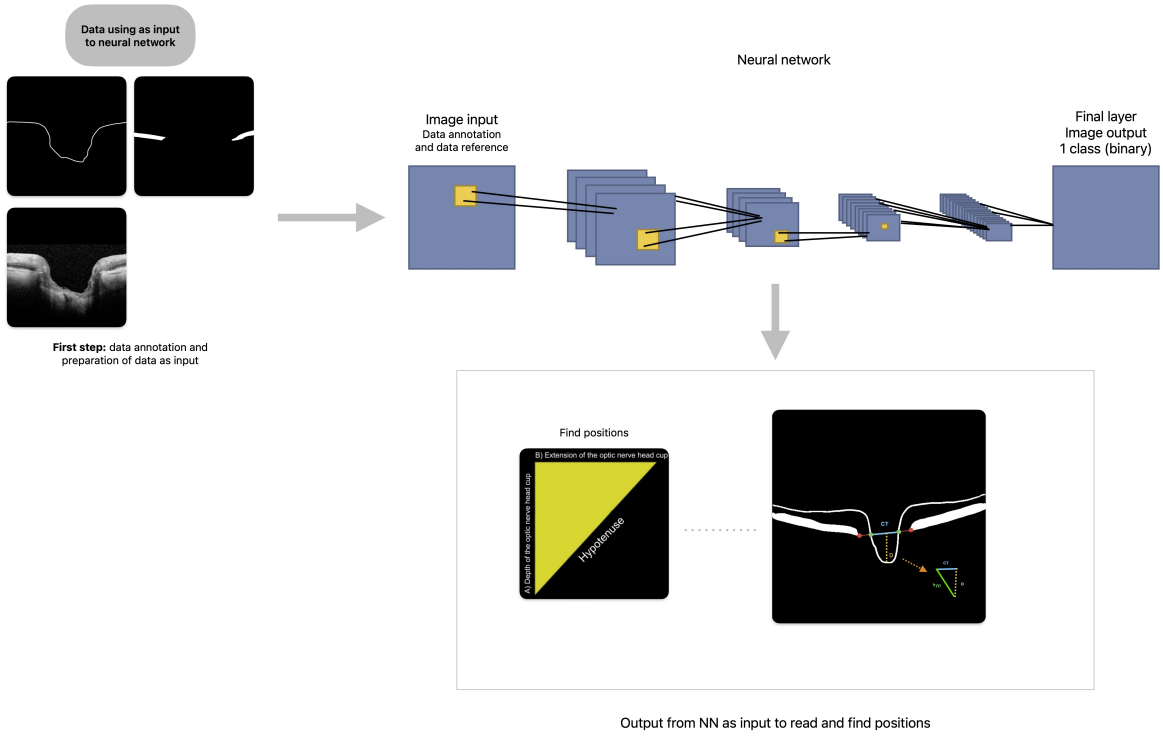


Source: (LAVINSKY et al., 2018).

To obtain analysis on the mentioned biomarker, the research was developed in three stages. The first stage includes collecting the OCT images with the physician to take notes and select

the regions of interest. In the second stage, these images are used as input data in a convolutional neural network for semantic segmentation. In the last stage the output of the neural network, containing the segmentation images, a computer vision algorithm is applied to obtain coordinates that will result in the identification of the region of interest. This last stage is performed by implementing the biomarker approach defined by (LAVINSKY et al., 2018). The results are compared with those analyzed by the specialist in the manual experiments in a previous study. The figure 17 shows an overview of each step.

Figure 17 – General overview



Source: Elaborated by the author.

4.2 Data management

4.2.1 Data annotation

This research developed a new annotated dataset to support the experiments. The overall context and main procedures are commented on in this section.

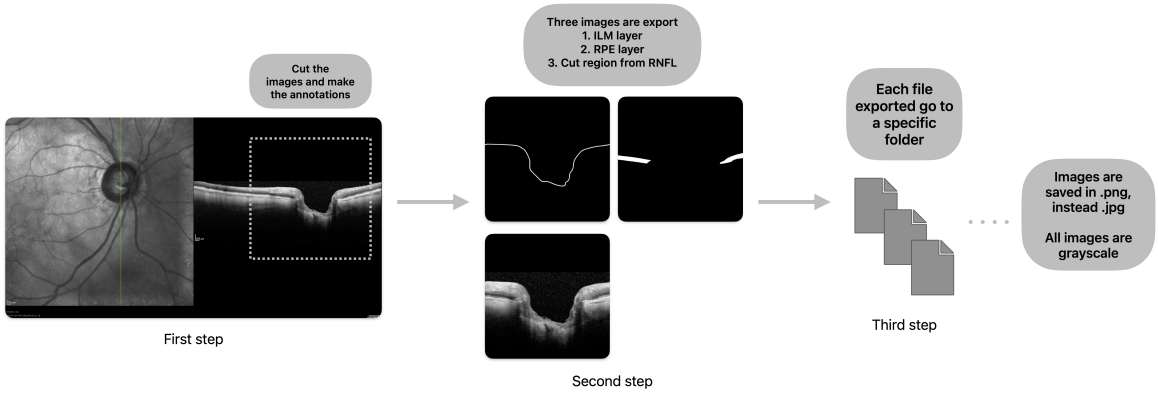
All original research participants underwent SD-OCT (Spectralis OCT; Heidelberg Engineering GmbH, Dossenheim, Germany)[1]. Moreover, the scans (B-Scan and C-Scan) were exported in JPEG format files for the current research. Each image represents only one eye (right

[1]The ethics committee of the Hospital de Clinicas de Porto Alegre approved the study and it was conducted by the Declaration of Helsinki. Informed consent was obtained from all participants.

or left), with the union of the two scans in the same image. Finally, Photoshop image editing software was used to support the cuts, annotations, and export to the desired sizes. The table 1 summarizes the received content.

As shown in figure 18, the first step refers to loading the images and selecting the region of interest to perform the cut. All images had the same size after cropping. Cropping all images with the same size is essential for input data in the neural network. In supervised neural networks, the image task must have the same pattern as the input data. The second step, shown in the exact figure, is the result of annotating the regions of interest after making the cuts. This step was conducted with a medical student to obtain the layers of the retina as accurately as possible, where a manual segmentation was obtained. Each layer was exported separately after the annotation process: ILM layer annotation, RPE layer annotation, and the gray-scale image with the selected region of the retinal layer. The exported clippings were turned into PNG files because the format allows lossless compression and less loss of quality. The files are organized in different folders in the third step according to each export performed. This distinction is because the neural network requires actual (or ground truth) and annotated data as input. Therefore, the annotations images are to be used with the image cropped under the same region.

Figure 18 – Annotated data flow.



Source: Elaborated by the author.

The set of images obtained contains 80 folders that represent images of both eyes. However, some exclusions were made. Folders that did not contain one or both images, corrupted files, and folders with the same images were exclusion criteria. From the total of 80 folders which should be 160 images, 134 images remained for use.

Throughout the experiments, the number of images was also modified. Tests were carried out with the entire set of images and new exclusions and separations. Tests without images with high contrast or noise were also applied. The same equipment can generate important differences in terms of noise presence or overall illumination in the image. This context fostered some experiments with image subsets, according to these main aspects.

Table 1 – Summarize of content.

Total of folders	80
Content from folders	Two files: left and right eye scan
Type of scans	B and C scan together, to each file
Format of content	JPEG

We applied some different tests with the images obtained during the research period: test using all data; changing the process for choosing the images, where the images that seemed to present greater similarity in overall aspects were chosen manually; the way of annotating the RPE layer where outlines have been replaced by fills (this aspect was possible because to us it's essential the limit of RPE layer on level of BMO as shown on figure 16 left, and not the separation from each inner layer).

4.2.2 Dataset split

During the process of implementation and execution of the pipeline, some experiments were applied to create the dataset used for model segmentation. We can reference five versions: (i) the images were split in 70% for training, and 15% for both, test and validation. Together, a Keras function called *ImageDataGenerator* was used to load the images and also generate a set of data augmentations; (ii) the split was decreased to the training set, to increase the number of images in the test set. Also, was modified in the code how to read the images from the directory, because the function *ImageDataGenerator* was deprecated. From this point, *from_tensor_slice* was the option to use and transform in a dataset (See attachment A). As mentioned, the deprecated function allowed together generate augmentations during the execution of the model. However, with the new augmented class, from Keras preprocessing layers, it was noticed that all operations applied modified the current image, not creating a new instance from the original with the augmentation to each operation. This was then created an algorithm using OpenCV, generating new images from the original, but each one with different modifications (Table 2); (iii) the third version was almost the same but used fewer images, selecting only the images with the best quality (manually). From 80 images, 70%, 15%, and 15% was the size of the split to train, test, and validate; (iv) At this version, the number of images decreased to 77, but the most important were two modifications. The first one was to export all images again but change the size from 512px square to 640px square (See section 4.2.1). The second was re-annotated the RPE layer (Figure 19). We could identify that for the inner layer, the neural networks don't perform well in segmenting a thin line where there are similar regions above and below (other layers). The importance of the segment RPE layer is based in after find it the limit at the level of BMO. To help in this process we thought to fill the annotation, and using not as a thin line more; (v) The last modification in the set was together with the code. Here we don't have more of a validation and test folder with images. Both, the validation and test images were put together

Table 2 – Data augmentation operations.

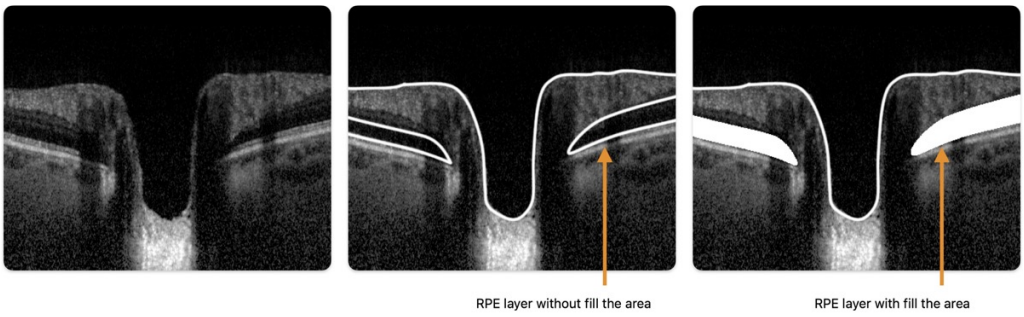
Operations
Horizontal flip
Rotation + zoom in (both sides)
Zoom in
Horizontal flip + zoom in
Shift (40px on each side)
Shift + horizontal flip (40px on each side)

Table 3 – Summarize of data split.

Version	**Total of images**	**Split**
(i)	134 images	70% for training, and 15% for both, test and validation
(ii)	134 images	55% for training, 15% for validation, and 30% to test
(iii)	80 images	70% for training, and 15% for both, test and validation
(iv)	77 images	70% for training, and 15% for both, test and validation
(v)	77 images	85% for training (with cross-validation), 15% for test

with train images - creating a unique place with the whole set of images - because were used the model with *Cross Validation* technique, where there are N folds that repeat to a value of X epochs. The table 3 summarizes the content explained in this section.

Figure 19 – Demonstration of annotations in the RPE layer.



Source: Elaborated by the author.

4.2.3 Data augmentation

As mentioned before, the use of data augmentation technique was applied in this work, since the set of images is small. The main goal is to enrich training dataset diversity to improve model generalization and performance.

In the case of this work, applying several tasks could be a problem, so we needed to be careful with the results. To generalize this context the operations can't result in new images drastically different, i.g. vertical flip, rotations with high angles, shifts that could cut parts of

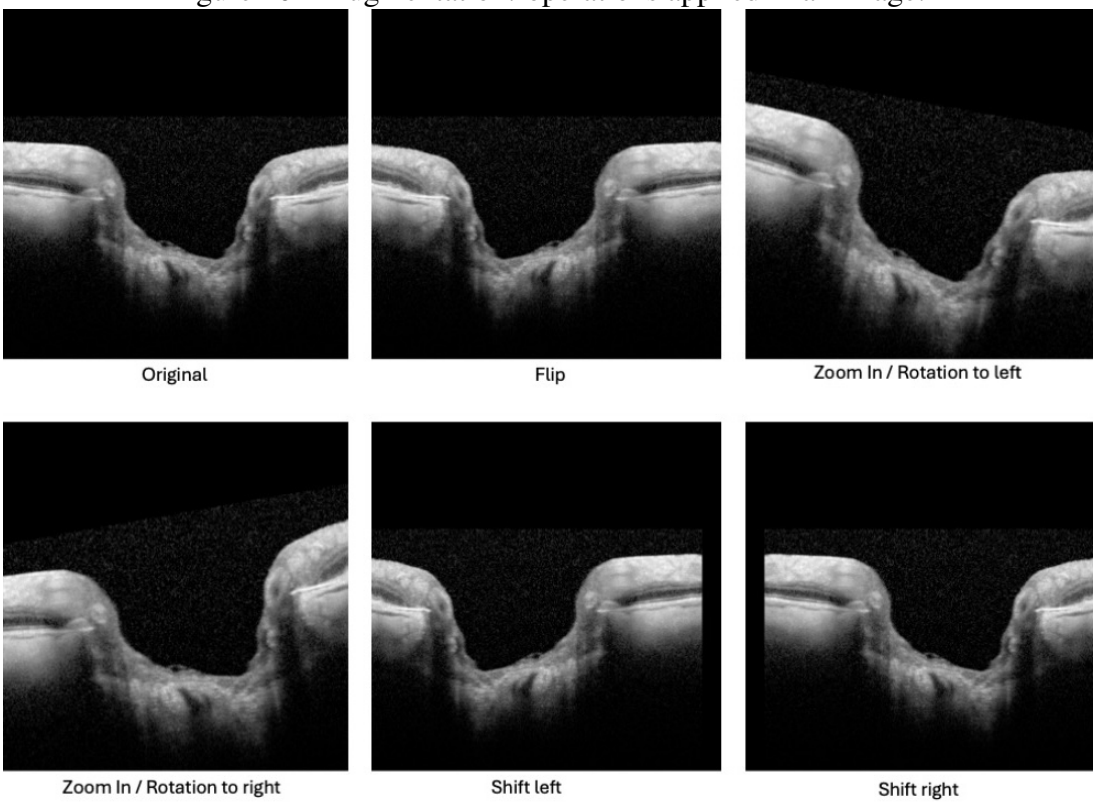
important regions.

The content used was small as mentioned in the table 3, and to increase the number shown, the operations showed in the table 2 were used manually with the OpenCV methods. Every image from the set separated to train generated a similar image but with a slight final difference, which generates nine new images. For all operations, the images were saved in a different folder, which is consumed by the process when is set to use the augmentation data. We evaluate the CNN model using both, with and without the use of image augmentation, to see if an increase in the number of images could affect the results. The structure of the folder is shown below, and how was distributed the set of images. Note that in this structure, folder test and validation were present. Only in version V (Table 3), they aren't present, since we needed only one folder to get all images and forward, during the loop of k-folds, split the set in train and test.

```
/
└── dataset folder
    ├── L1
    │   ├── train
    │   │   ├── images
    │   │   ├── images aug
    │   │   ├── masks
    │   │   └── masks aug
    │   ├── test
    │   │   ├── images
    │   │   └── masks
    │   └── validation
    │       ├── images
    │       ├── images aug
    │       ├── masks
    │       └── masks aug
    └── L2
        ├── train
        │   ├── images
        │   ├── images aug
        │   ├── masks
        │   └── masks aug
        ├── test
        │   ├── images
        │   └── masks
        └── validation
            ├── images
            ├── images aug
            ├── masks
            └── masks aug
```

Lastly, the figure 20 shows some operations applied in one unique image.

Figure 20 – Augmentation: operations applied in an image.



Source: Elaborated by the author.

4.2.4 Cross-validation

Cross-validation is a statistical method used to estimate the skill of machine learning models. Is a resampling procedure used to evaluate machine learning models on a limited data sample.

Normally, splitting the data into training and test sets is done once. As a result the possibility of evaluating the model only once isn't good. To be more confident about the results, it's necessary to evaluate the model multiple times.

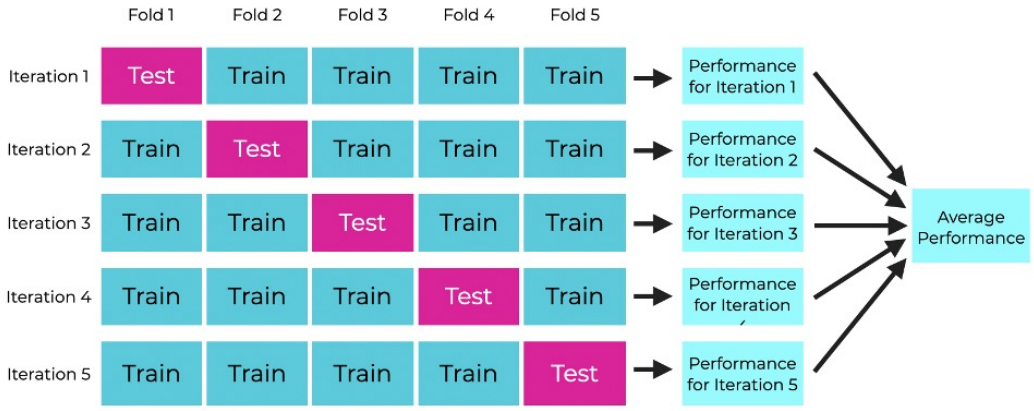
The procedure has a single parameter called *k* that refers to the number of groups that a given data sample is to be split into. As such, the procedure is often called k-fold cross-validation. When a specific value for k is chosen, it may be used in place of k in the reference to the model, such as k=10 becoming a 10-fold cross-validation.

Cross-validation is primarily used in applied machine learning to estimate the skill of a machine learning model on unseen data. That is, to use a limited sample to estimate how the model is expected to perform in general when used to make predictions on data not used during the training of the model (Figure 21).

To our work, we applied the following rules:

- Split the dataset into *k* groups.

Figure 21 – Cross-validation.



Source: Elaborated by the author.

Table 4 – Definition of split data and k-fold.

		Set to test (15%)	**k-fold**	**Set to val**	**Set to train**
Images collected	77	11	7	10	56
Sum of collected and augmented	770	110	7	100	560

- For each unique group:
  - Take the group as a holdout or test data set.
  - Take the remaining groups as a training data set.
  - Fit a model on the training set and evaluate it on the test set.
  - Retain the evaluation score and discard the model.
- Summarize the model's skill using the sample of model evaluation scores.

As mentioned before (Table 3) the number of images used with Cross-Validation was from a total of 77. From the whole set, every iteration was separated 15% to evaluate. From the 85% remained, was separated 15% to validate (Data on which to evaluate the loss and the model metrics at the end of each epoch).

To have an integer number of iterations (k-folds) and a small k number was considered k lesser than 10. Both when had or not the presence of augmentation, the number chosen was k=7. To reach this number, the total of images is divided by the number of test images. The numbers are shown in the table 4.

4.3 Neural network to segmentation

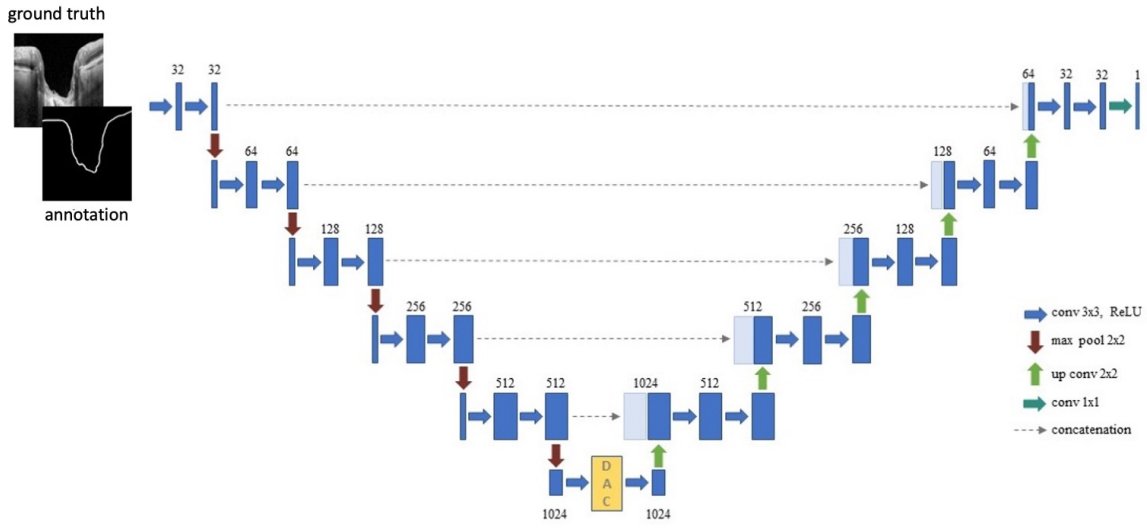
4.3.1 CNN Architecture

The present work used a neural network to perform tasks related to image segmentation. The use of the U-Net architecture (RONNEBERGER; FISCHER; BROX, 2015) was defined as a starting point. Modifications were then included, based on other works, to foster the results.

The U-Net architecture consists of a contraction path to capture context and a symmetric expansion path for localization. The contraction path consists of a typical convolutional network, with two convolutions followed by a ReLU (rectified linear unit) activation function and a max polling operation. The use of the regularization dropout technique to reduce overfitting was also considered. The number of resource channels doubles each reduction step (downsampling). The expansion path consists of an upsampling of the feature map, followed by an up-convolution that halves the number of feature channels, a concatenation with the corresponding clipped feature map from the contracting path, and two convolutions followed each by the ReLU function. Still, according to the authors, clipping is necessary due to the loss of edge pixels in each convolution. In the final layer, a 1x1 convolution maps each feature vector to the desired number of classes.

This is a network with no fully connected layers and which uses only the valid part of each convolution, i.e., the segmentation map only contains the pixels for which the entire context is available in the input image. It also allows the continuous segmentation of arbitrarily large images by a block overlay strategy, predicting the pixels in the region of the edge of the image that, due to the missing context, are extrapolated, mirroring the input image.

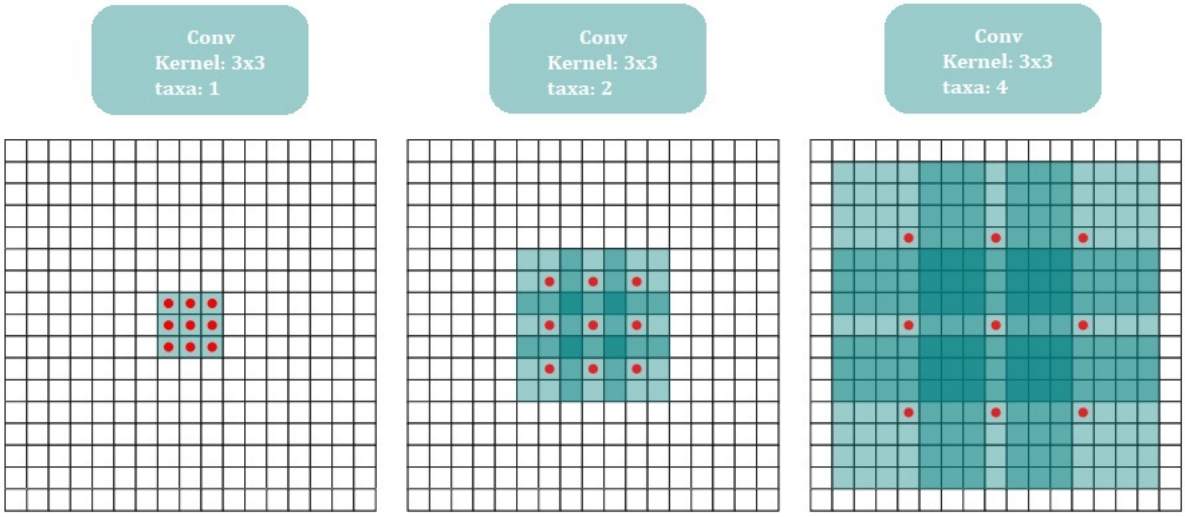
Figure 22 – U-Net Architecture.



Source: Elaborated by the author.

Changes were made to the architecture (Fig. 22) based on two articles, (GU et al., 2019), (YU; KOLTUN, 2016). In both, there is the presence of additional blocks between the path of contraction and expansion. These modules allowed the extraction of information from the semantic context and the generation of high-level resource maps. The dense atrous convolution (DAC) module was used for the present work. DAC was designed to extract features from objects with different sizes using a set of dilated convolutions.

Figure 23 – Dilated convolution.

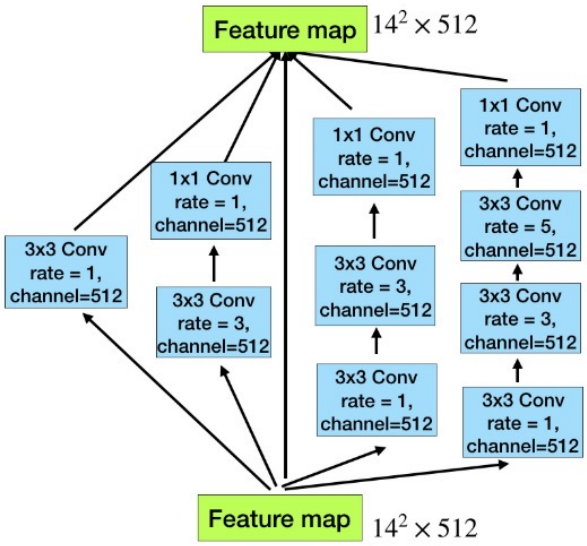


Source: (YU; KOLTUN, 2016).

Dilated convolution was initially proposed for computing the wavelet transform. Allows insertion of zeros between consecutive values along the spatial dimension. Standard convolutions have a rate of change equal to 1, while atrous convolution allows changing this value. This type of operation supports the exponential expansion of receptive fields without loss of resolution or coverage (YU; KOLTUN, 2016) (Fig. 23). Thus, a DAC block was inserted between the contraction and extension paths to extract the features at different scales. Based on Gu and coworkers, the same branch structure was used within the DAC module (Fig. 24), only changing the feature map from 512 to 1024.

Finally, the cross entropy loss function was initially used for model training due to its ability to calculate the difference between two probability distributions, which in this case are the pixels for the background and foreground. However, objects in medical images, such as the optic disc and retinal layers, occupy small regions compared to the whole. According to (GU et al., 2019), cross-entropy is not performant in these cases and can be replaced by the loss function known as the Dice coefficient, which looks for the correlation between the prediction points and the annotations.

Figure 24 – Dense atrous convolution (DAC).



Source: (GU et al., 2019).

4.3.2 Train metrics

Over the training process, training metrics are measurements that help evaluate the effectiveness of training programs by assessing how well employees learn and apply new skills. A metric is a function that is used to judge the performance of the model. Metric functions are similar to loss functions, except that the results from evaluating a metric are not used when training the model.

The chosen metric was the Dice Coefficient since it is a similarity metric commonly used in image segmentation. The Dice coefficient is a measure of the similarity between two sets, A and B. The coefficient ranges from 0 to 1, where 1 indicates that the two sets are identical, and 0 indicates that the two sets have no overlap.

The implementation for the dice coefficient used is shown below.

```
def dice_coef(y_true, y_pred):
    y_true = Flatten()(y_true)
    y_pred = Flatten()(y_pred)
    intersection = tf.reduce_sum(y_true * y_pred)

    numerator = 2. * intersection + smooth
    denominator = tf.reduce_sum(y_true) + tf.reduce_sum(y_pred) + smooth

    result = numerator / denominator

    return result
```

Listing 4.1 – Dice Coefficient function.

To make it a loss, was needs to be made into a function to minimize. This can be accom-

plished by making it negative.

```
def dice_loss(y_true, y_pred):
    return 1.0 - dice_coef(y_true, y_pred)
```

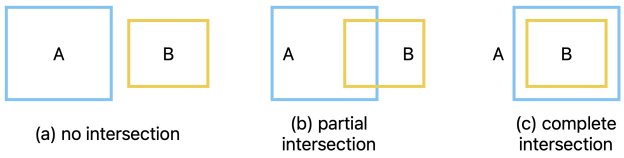
Listing 4.2 – Dice Coefficient loss.

4.3.3 Metrics of evaluation over test set

For the model evaluation, we decided to use mainly one metric for segmentation tasks. The evaluation metrics were the Intersection Over Union (IoU). As mentioned in the work of Rezatofighi and coworkers (REZATOFIGHI et al., 2019), IoU, or the Jaccard index, is the most commonly used metric for comparing the similarity between two arbitrary shapes. IoU encodes the shape properties of the objects under comparison, e.g., the widths, heights, and locations of two bounding boxes, into the region property and then calculates a normalized measure that focuses on their areas. Also, according to another work (TAHA; HANBURY, 2015), both metrics (IoU and F1) measure the same aspects and provide the same system classification. Therefore, the use of each one may depend on the case.

IoU, mathematically can be related to quantifying in an objective manner several types of similarity between two or more mathematical structures such as scalars, sets, vectors, matrices, functions, densities, graphs, etc. The basic Jaccard index can be simply expressed where set A and set B are any two sets to be compared (Figure 25). (F. COSTA, 2021).

Figure 25 – Demonstration of intersection between two sets.



Source: Elaborated by the author, based on (F. COSTA, 2021).

In our work, we employ the prediction of a set of images and the binary similarity. The code below shows a part of the code.

```
color_img = cv2.imread(img_name, cv2.IMREAD_COLOR)
normalize_img = color_img_bw / 255.0
expand_dim_img = np.expand_dims(normalize_img, axis=0)
float_img = expand_dim_img.astype(np.float32)

mask = cv2.imread(mask_name, cv2.IMREAD_GRAYSCALE)

y_pred = model.predict(float_img, verbose=2)[0]
y_pred = np.squeeze(y_pred, axis=-1)
```

```
y_pred = y_pred >= 0.5
y_pred = y_pred.astype(np.int32)

mask = mask / 255.0
mask = (mask > 0.5).astype(np.int32).flatten()
y_pred = y_pred.flatten()

iou_value = jaccard_score(mask, y_pred, labels=[0, 1], average="binary")
```

Listing 4.3 – Evaluation code.

4.4 Measurement of cup portion on ONH structure

The capture, annotation, and prediction process allows a flow to obtaining segmented images by a neural network. This result is then used as a final step to obtain values in the retinal excavation region. In this way, the calculation presented (LAVINSKY et al., 2018) can be carried out and evaluated computationally (Figure 16 right).

The first necessary step is to obtain the coordinates and excavation length value between the BMO boundary and the ILM layer. Next, the segmented images of the two layers are joined to be able to search the depth. The depth of posterior displacement of the lamina cribrosa, measured from the line created at the level of opening of Bruch's membrane level, is then used as a source for obtaining the measurement.

The values obtained make it possible to assume that they are the legs in the Pythagorean equation. Moreover, in this way, obtaining the hypotenuse is discovered in pixel value. The last step is the conversion from pixel to micron, the unit of measurement used by the scanning equipment. This entire process is performed using the Python programming language.

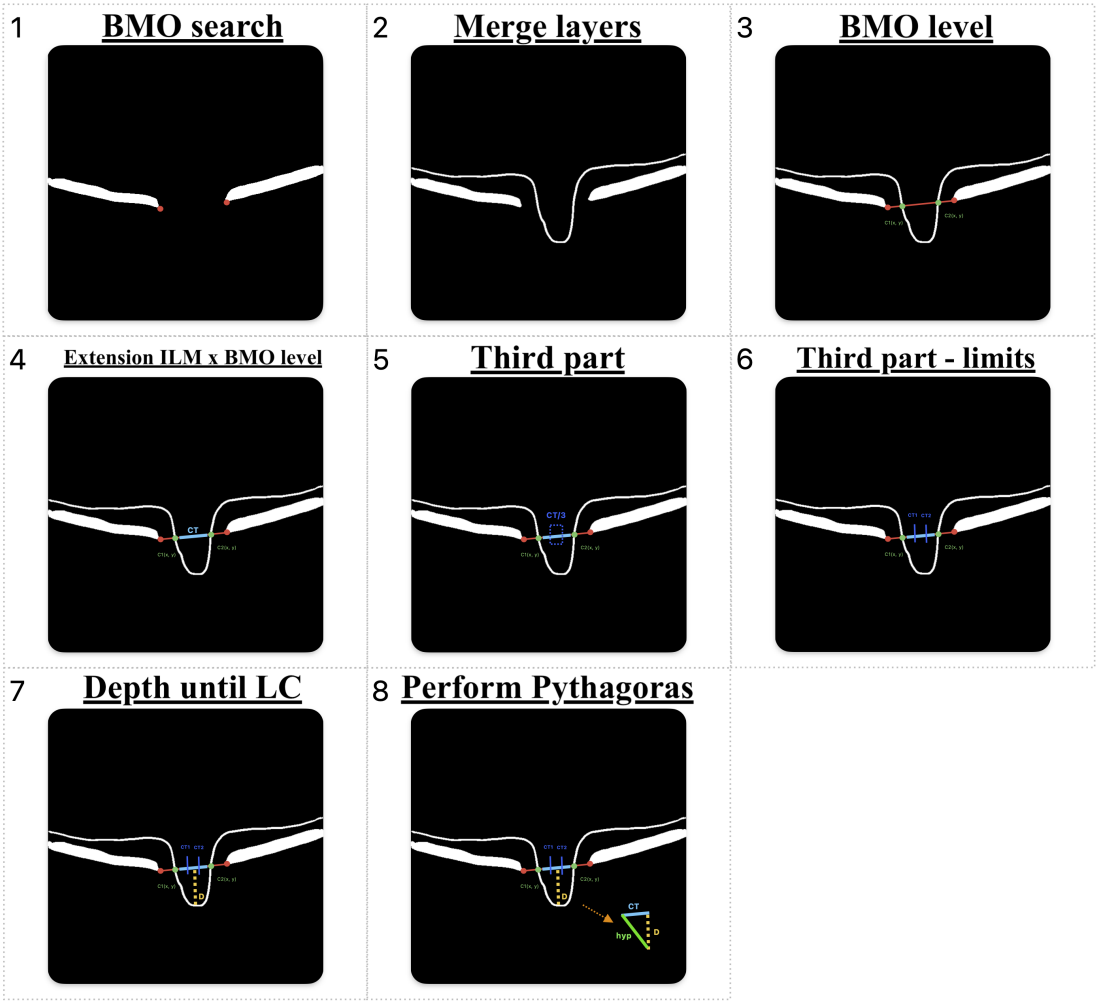
A flow was then created with steps to be followed, described below[2] and shown in the figure 26.

```
In L2, read the inner ends.
Merge L1 and L2.
Search for coordinate [X, Y] in layer L1 based on L2.
Find an area named CT, using: CT = c2(x) - c1(x).
Find the third part of the region CT: CT/3.
Search for CT1 and CT2.
Find greater depth (D).
Convert to the unit micron meter.
Use CT(x) and D(y), converted, to calculate the hypotenuse.
    hyp = sqr(CT(x)^2 + D(y)^2)
```

Listing 4.4 – Steps to find the hypotenuse.

[2]L1 represents the external annotation (ILM), and L2 the internal annotation (RPE).

Figure 26 – Organization of steps to find the hypotenuse.



Source: Elaborated by the author.

4.4.1 Conversion to micron meter

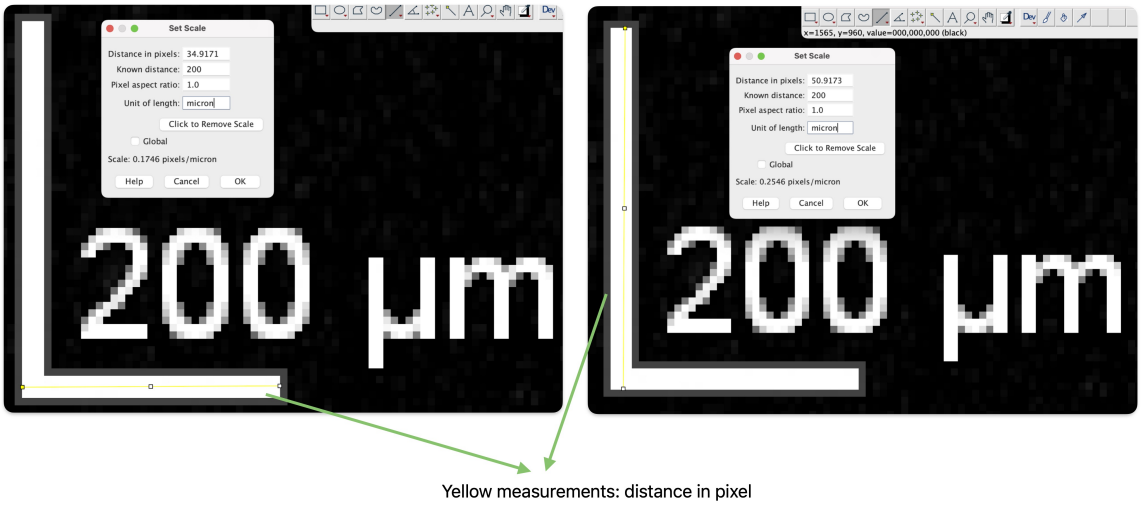
The values found are still in pixels at the end of the search step for regions connected to the cup region. Therefore, there is a need for conversion to the measurement unit used by physicians.

The conversion process is currently manual since we use a scale available in the image obtained by EDI SD-OCT. This scale is always found in the lower part of the images and in an "L" shape, where the vertical bar represents a scale, and the horizontal bar represents another scale.

Using the ImageJ software, it is possible to convert the pixel distance of each orientation to the displayed reference (200 μm), as shown in the figure 27. The returned values are then used as variables in the flow presented in the previous section, to calculate and represent the final

values.

Figure 27 – Measurements using ImageJ software.



Source: Elaborated by the author.

5 EXPERIMENTS AND RESULTS

This chapter describes the results of the segmentation of retinal layers, and also the results of obtaining the cup portion-based methodology described, where the discovery of the hypotenuse over the disc region of the cup is performed and compared with the physicians' measurements. For the neural network, the implementation is based on the Keras API running under TensorFlow. All training was performed using Google Colab connected to GPUs.

5.1 Results over-segmentation of layers

The tests were performed with different numbers of images, changes in hyper-parameters, re-annotation of images, and excluding some images as mentioned in the chapter 4. Mainly, was considered for both L1 and L2, the use of only the set of images pre-processed (77 images), and with augmentation (increase to 770 images). Lastly, both lengths from the sets were used in two neural network models: U-Net and U-Net modified. The test shown is about the versions of datasets iv and v (Table 3) where the structure of folders was organized to use or not cross-validation. The table 5 shows two datasets: DSV4 and DSV5 (V4 and V5, represent versions) with the number of images, and the number of epochs used to train.

From this point, we define eight tests, where: four are with the DSV4 and four with DSV5. Both of the groups used U-Net and U-Net with DAC, but DSV5 was conducted with Cross-Validation. The whole process of eight tests also was done for L1 and L2, which resulted in sixteen tests. Table 6 described the mentioned tests and the figure 28 and figure 29 show the results of learning curves.

5.1.1 Learning Curves and model results

Learning curves (LC) are widely used for the diagnostic of the train and validation model, over the epochs. It can be used to diagnose problems with learning, such as underfitting or overfitting the model. Diagnostic the training set gives the idea of how well the model is learning, and diagnostic the validation set gives the idea of how well the model is generalized. Also, is important to mention that both figures of the LC plot the optimization curve (loss) and performance curve (dice).

There are three common dynamics observed in LCs. They are: underfit; overfit; and good fit. The underfit curve refers to a model that cannot learn the training dataset. The overfit curve

Table 5 – Specifications from datasets used.

	Epochs	**DS-size**	**Cross-Validation**	**k-fold**
DSV4	80	77	No	-
DSV5	80	770	Yes	7

Table 6 – Script of tests.

Name of test
UNET-DSV4
UNET-DSV4-aug
UNET-DAC-DSV4
UNET-DAC-DSV4-aug
UNET-DSV5-CV
UNET-DSV5-CV-aug
UNET-DAC-DSV5-CV
UNET-DAC-DSV5-CV-aug

refers to a model that has learned the training dataset too well, including the statistic noise or random fluctuations in the training dataset. Lastly, a good fit curve is when it's possible to identify a training and validation loss that decreases to a point of stability with a minimal gap between the two final loss values. Normally, the loss of the model will almost always be lower on the training set than the validation set, and this means that there could be a gap between the train and validation loss learning curves. This gap is referred to as the generalization. Table 7 summarizes each dynamic over the aspect of loss.

Both figures (Figures 28 and 29) show the separation of three groups (blue, green, and yellow). The first one, considers only the set of images with or without augmentation, using U-Net or U-Net with DAC. The other two groups apply cross-validation in both kinds of NN used. From the perspective of the results indicated, we can notice some aspects.

About the blue box, the use of pure U-Net with L1 and L2 showed good performance with generalized learning when data augmentation was not employed. However, when data augmentation was used, it was not possible to complete the training. This resulted in an underfit curve, indicating that the small dataset combined with a deep network led to poor results. For models with or without data augmentation, but with augmentation, notable results were achievable in the training curve. However, there was a tendency for L1 to show signs of overfitting.

In the green box, we have the tests with CV and without DAC. We can observe that for both layers of data, L1 and L2, there is a tendency towards a good fit in most cases or folds, except for one fold of L1. In this case, there is a set of data where some images do not fit well into the learning process.

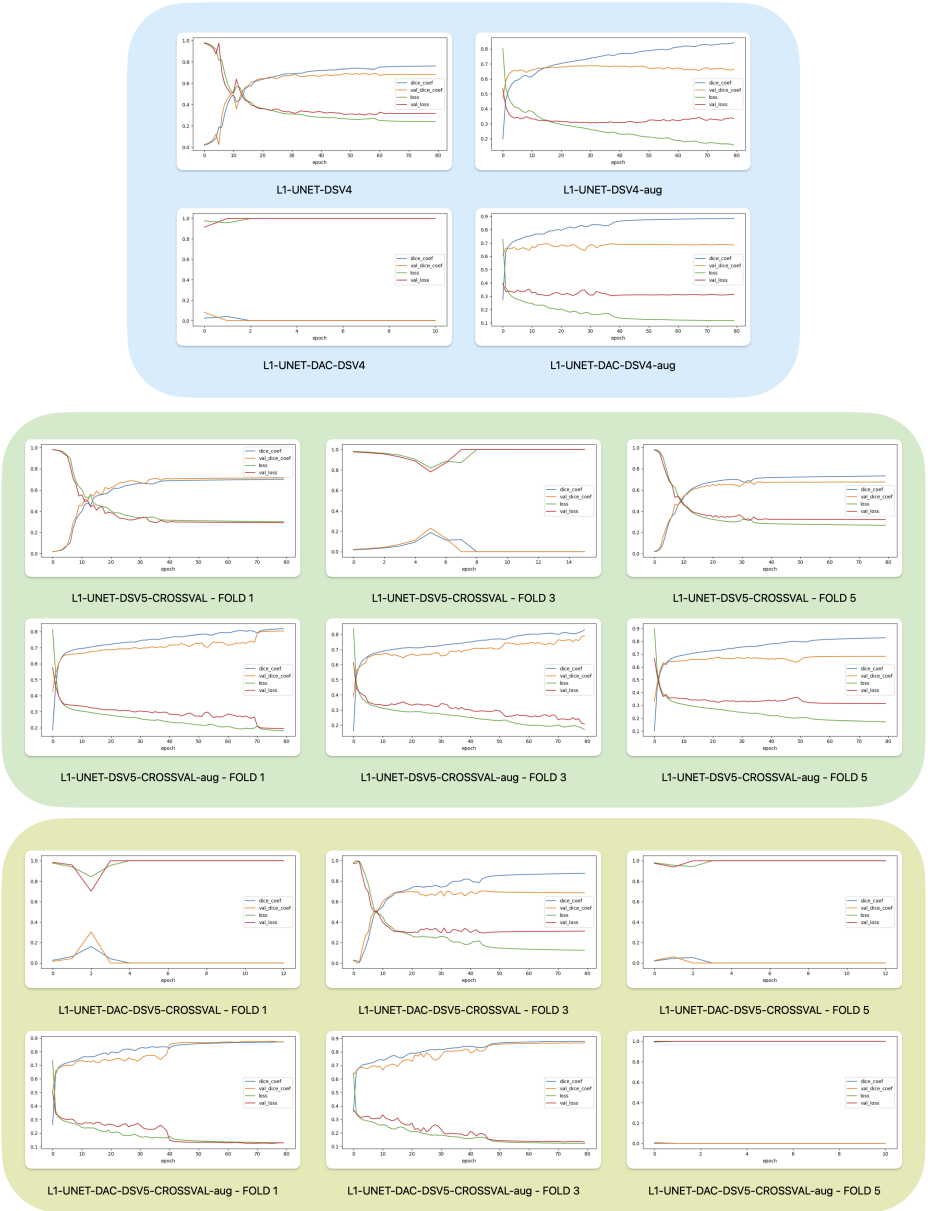
The yellow items were enlarged using DAC. The items remained consistent, whether they were augmented or not in the L2 set. For L1, some folds do not generalize well and do not facilitate effective learning, leading to underfitting in multiple instances.

Increasing the amount of data often does not lead to significant improvements in learning curves. Instead, it can result in overfitting or underfitting. Expanding the neural network also yields mixed results, sometimes leading to underfitting where the model is unable to learn from the data presented.

Table 7 – Learning curves behaviors.

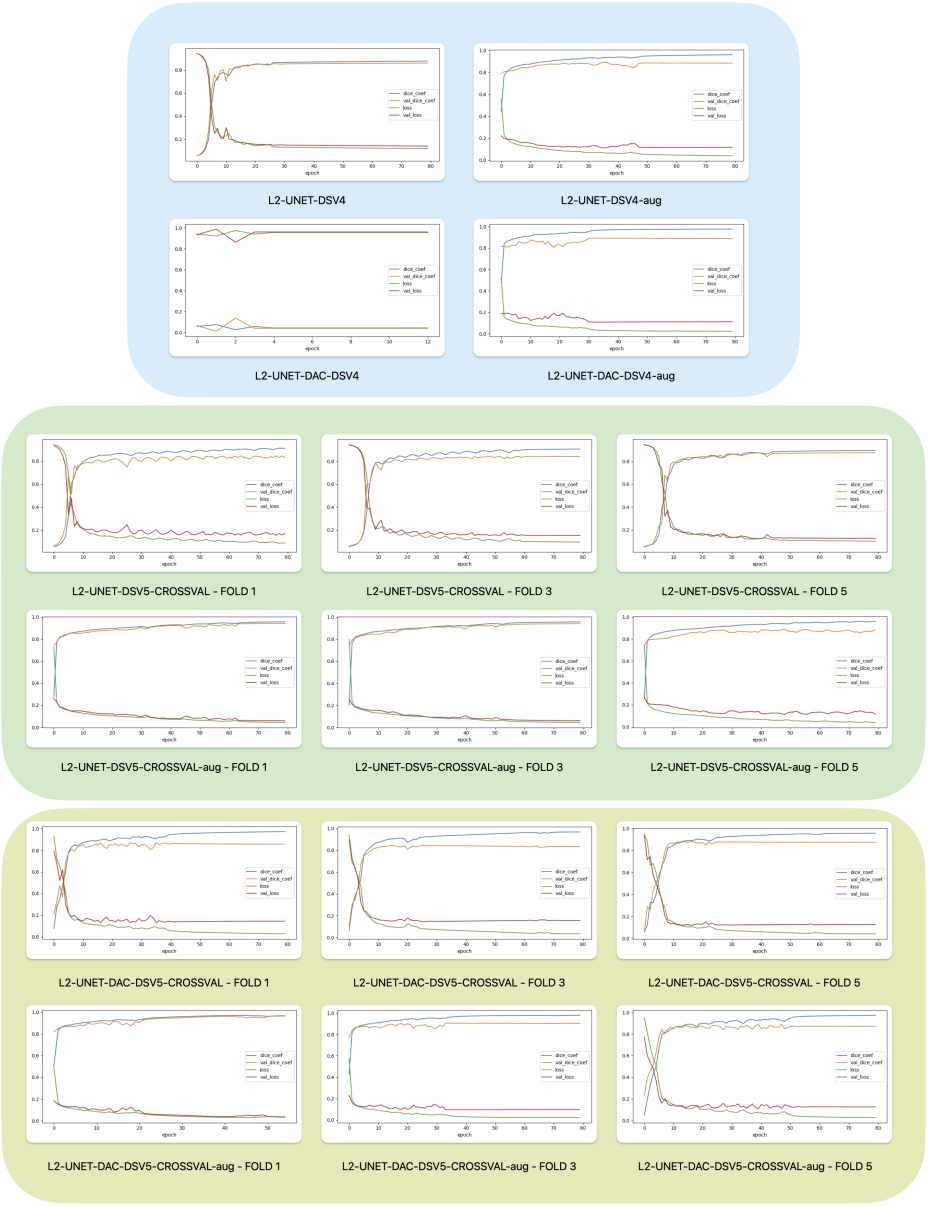
Curve kind	**Curve behavior**
Underfit	Train and validation loss remain flat constantly.
Overfit	Training loss decreases normally, and validation loss decreases, but then starts to increase again.
Good fit	Training and validation loss decrease with stability, with a small gap between the two curves.

Figure 28 – Learning curves over each test, to L1.



Source: Elaborated by the author.

Figure 29 – Learning curves over each test, to L2.



Source: Elaborated by the author.

5.1.2 Results over evaluation in test set

After creating the models, we evaluated the test set. For the models using version four (DSV4), we had previously separated the images, while for version five (DSV5), we separated the images during the execution of each fold. As mentioned in section 4.3.3, the evaluation was based on the IoU metric. The results [1] for each test for L1 and L2 are shown in Table 8 and Table 9.

To the outermost layer (ILM), it was possible to notice that there was a slight variation in

[1] To the CV was an employee the mean of folds.

Table 8 – IoU results obtained for the ILM layer.

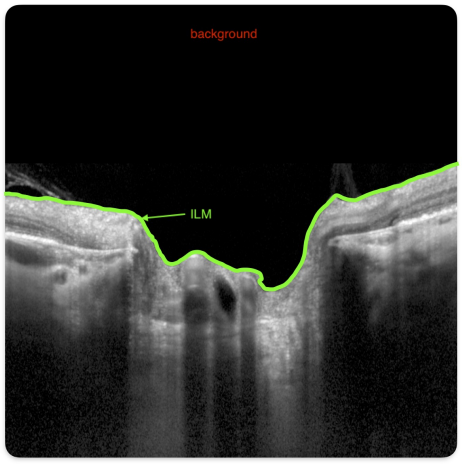
Test name	**IoU**
L1-UNET-DSV4	0.57539
L1-UNET-DSV4-aug	0.55978
L1-UNETDAC-DSV4	0.00095
L1-UNETDAC-DSV4-aug	0.59040
L1-UNET-DSV5-CROSSVAL	0.56843
L1-UNET-DSV5-AUG-CROSSVAL	0.57965
L1-UNETDAC-DSV5-CROSSVAL	0.33452
L1-UNETDAC-DSV5-AUG-CROSSVAL	0.47646

Table 9 – IoU results obtained for the RPE layer.

Test name	**IoU**
L2-UNET-DSV4	0.7612
L2-UNET-DSV4-aug	0.7883
L2-UNETDAC-DSV4	0.0235
L2-UNETDAC-DSV4-aug	0.7767
L2-UNET-DSV5-CROSSVAL	0.7659
L2-UNET-DSV5-AUG-CROSSVAL	0.7836
L2-UNETDAC-DSV5-CROSSVAL	0.7553
L2-UNETDAC-DSV5-AUG-CROSSVAL	0.7619

the results obtained where most of the collected images there was little difference in contrast and variation in quality besides being a border region between the background and the scanned retina (green line on figure 30). In this way, it was possible to obtain results in this layer with less variation.

Figure 30 – Demonstration of annotation to ILM layer.



Source: Elaborated by the author.

In contrast, the RPE layer initially obtained lower values than the other region. We noticed

that the neural network did not perform well when detecting a more internal part with a thin line (annotation). Architectural changes were made but without success. The annotations for the last two versions of the dataset were then modified, containing thicker annotations, focusing more on the boundary region between the RPE and the BMO since it is vital for the final calculation (detailed in the section 4.2.2).

The results also show that adjusting the neural network and defining it as an isolated neural network for each image context is necessary. Two of these adjustments were based on epochs and depth of layers. A plateau was possible to observe earlier for the RPE layer, unlike the ILM layer which took longer. Build blocks also show differences between the two layers. The ILM layer even with DAC on the bottleneck performs a little better when goes deeper, already the RPE layer performs the same with more or fewer blocks.

As shown, the results for the ILM layer don't change drastically from one case to another. Only it's possible to note that the result is slightly better with more blocks. Another observation is about the use of Cross-Validation, which helps in performing a model that can see the whole dataset (except the test set). For the RPE layer, we had better results since the first execution. The possible cause of this behavior could be related to the size of annotations, where it's thicker than presented on the ILM layer.

5.2 Discovery of the hypotenuse in the excavation region

Experiments were carried out at this stage to understand the feasibility of using computer vision libraries, such as OpenCV in Python, to identify measurements in the excavation region.

The algorithm created to identify what was described in the section 4.4 obtained results that were compared with the values obtained by the physician in his study. As shown in the table 10, columns: length, depth, and hypo are the values obtained by the physician. Columns with the same name but with the suffix "-alg" it is a reference to the algorithm created and described in the section 4.4. The column "diff: hypo" shows the difference between the results obtained by the physician and the algorithm.

Also, in table 10, it is possible to see the comparisons between each measured part and the final result, the hypotenuse. We can see there are differences between both measurements. The final differences between the measurements taken by the doctor and those found by the algorithm can still be caused by manual measurement and the erroneous reading of some pixels in the final step.

For the first case, was possible to see how it obtained the values when the doctor analyzed the OCT-imaging by the scan. The depth of the posterior LC displacement, measured from the BMO level, and the length of the vertical cup. Both are measured by eye, where the results obtained are through manual tracing which could imply differences if the same analysis is performed more than once.

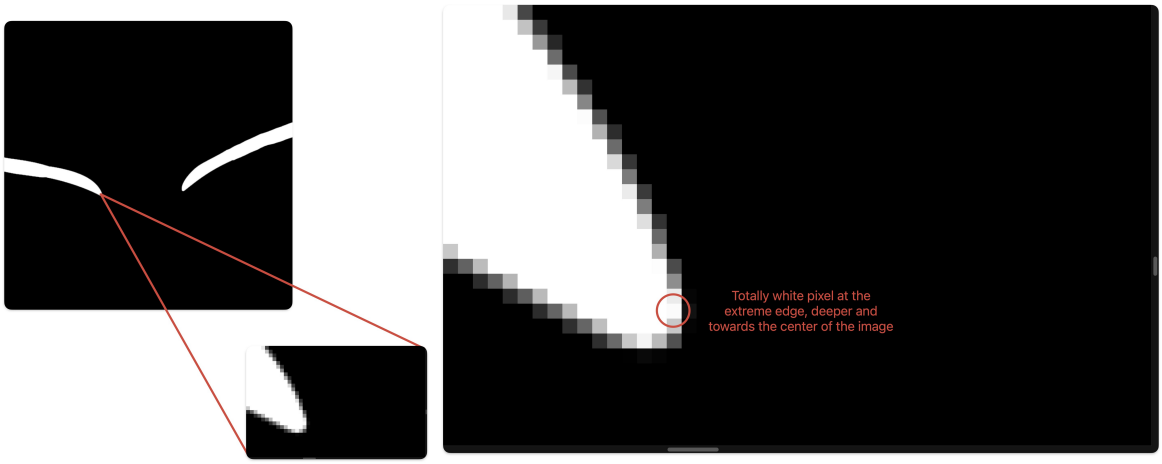
The second case is about the algorithm created and used to analyze the segmentation gener-

Table 10 – Results obtained for hypotenuse findings (µm)

length	length-alg	depth	depth-alg	hyp	hyp-alg	diff: hyp
978	1062.5	288	384	1019.52	1129.762	110.239
517	662.5	150	212	538.320	695.593	157.273
699	887.5	313	424	765.878	983.581	**217.703**
1559	1725	385	476	1605.84	1789.469	183.634
985	1162.5	481	592	1096.17	1304.557	208.389
267	275	132	124	297.847	301.664	**3.816**
1377	1550	327	364	1415.29	1592.167	176.873
1338	1487.5	502	584	1429.07	1598.034	168.961
924	1037.5	486	572	1044.02	1184.732	140.715

ated by the NN. As described in the section 4.4, was performed the reading of pixels and when the limits between white pixel and black pixel are found, in the regions of interest, these are used for the calculations. Some of these values could be slightly different from those obtained by the physician. The figure 31 shows an example of interpretation from the algorithm to choose the coordinate where was considered the limit between RPE and BMO. We can interpret that this value read (as the others) could be different from where the physician marked causing the differences shown in the table 10.

Figure 31 – Demonstration of pixel value selected in the region from RPE with BMO.



Source: Elaborated by the author.

The values in bold from the table 10 show the closer and further values obtained. In the perspective of only a pure comparison, the values aren't so high. However, on the micro-scale, these could represent a high difference in terms of distance and evaluation of cup area.

Moreover, finally, to calculate the hypotenuse, there is still a need to investigate how to convert the pixel aspect ratio to the micron unit (µm) without using secondary software. Currently, it is necessary to open some images through the ImageJ software and to read information from the scale bar, where it is possible to know the distances of microns in pixels. This is necessary

to convert the results from hypotenuse findings (section 4.4.1).

6 CONCLUSION

Optical coherence tomography is a modern and sophisticated image extraction technology that can provide excellent results when scanning the retina and inner layers. However, several studies still enable new ways for more accurate disease analysis using ML methods for automated glaucoma detection based on trained structural and non-structural features, with OCT images and other ways such as fundus photography and VF images. Still, it is essential to mention that glaucoma is a complex disease and that the more tests are available, the more possibility that the diagnoses don't have failures.

The retina structure presents susceptible characteristics in which detailed analyses are necessary. The study had the support of a specialist physician and a medical student who participated, enabling the availability of OCT images and further helping in taking notes. As a first study, we focus efforts on using deep learning and semantic segmentation to evaluate the possible results and benefits of the technique, aiming to perform a mathematical calculation to reproduce the hypotenuse measurement over the optic nerve's head. Regarding these aspects, the main contribution of this work is the creation of a methodology for collecting the data, and performing the pre-processing, to further predict the necessary areas and perform the calculation to use as a new biomarker to diagnose the remodeling of the cup area in patients with glaucoma.

Further, the research had the opportunity to acquire authentic images from patients with glaucoma. These images were used in a previous study by Lavinsky and coworkers and granted by the physician Dr. Lavinsky for the present study. We needed to understand the scans and decide how to organize them because all images are composite by two scans (B-Scan and C-Scan). Both show information from ONH but with different components. The images in B-Scan show the layers from the retina, where, for now, the study focused on collecting information through a (i) manual annotation from ILM and RPE layers for (ii) input in an NN and predicted semantic segmentation from these layers. These steps were important because we could create a new dataset to perform the predictions and tests (the set of images has also passed through some modifications while testing them with the NN model until the last version presented in this study is found). From this point of view, it is necessary to emphasize the contribution of creating a new dataset. Working with new and real data presents challenges and difficulties, from cleaning and organizing to use in the following steps.

A part of the acquired images was left to be used in the test of NN and collect results based on segmentation metrics, too. Finally, another set of images was also used in the last part: identify the cup region and limits between ILM at the BMO level and the cup's depth at the BMO level until LC. The process allowed us to collect results and calculate a "path" that represents the remodeling of the optic nerve head cup in patients with glaucoma. To do this, we implemented an algorithm using libraries of computer vision (OpenCV) where the aim was to receive the output from NN and collect information from pixels based on the limited areas mentioned to perform the final calculation.

From the point of view of the pipeline, after having the set of images to use in the study, we focus on the DL model. For this moment, the architecture used as a base was the U-Net. There is strong evidence in the literature showing studies applying to many different areas of biomedicine. Our aim in this step was to identify the layers of the retina and do their segmentation. To do this, we train a model with a pure U-Net, and after a while, we add new blocks based on other studies from the literature. We generate some models to compare the results - from each layer - and can analyze the learning curve of the model. We noticed that U-Net performs well without any adjustment when the set input has characteristics such as thicker or more detailed notes in the image. Notes as lines thinner do not present difficulty to segment, but it is challenging to have detail and precision. In an image where almost the whole area is black and only one part has information, a poorly outlined segmentation may represent a decrease in results. Another highlight point is the different shades of gray, implied in poor segmentation when darker areas are present. We also noticed the need to separate the neural network in the future, applying different architectures for each layer of the annotated retina because it was possible to see that both the model and hyperparameter had different behaviors for other patterns (or different areas from the image in our case).

The results obtained were used to measure the region of the cup in the optic nerve head and calculate the hypotenuse using Pythagoras Theorem. It was possible to test only a few sets of images from the total granted; therefore, we were likely to notice that as we applied some rules to identify the areas in the images, some of them turned out to be different when compared with the values obtained by the physician. One of the drawbacks in this comparison is that on one side are rules, and on the other is knowledge and a "free hand" to collect the information. However, most of the values presented here are not distant values, even if a refinement is necessary to get closer results.

One limiting point that follows since the beginning of the research is the number of images. In which, as highlighted in section 4.2.1, images were excluded by quality criteria. There is still a significant challenge in working with variations in images, such as contrast or the presence of noise. Another highlight is that we use images in large sizes compared to other datasets from the same theme and different themes. The drawback here is the dataset size, which can increase, and the time it takes to learn and predict the areas by the NN model since we use large sizes. In contrast, there is the advantage of using images with large sizes; they provide the information clearer than images resized to small size, and for OCT scans, the decrease in quality could imply a loss of information. Compared with other datasets, which are available and open to download, we need to mention again that our dataset is part of the contributions of this work, not only the pipeline until we find the final values because the images have the best quality in most cases. We tried to use the original size without some compression.

The use of images from only one OCT device can be a factor that negatively contributes to the creation of a bias since we do not know the limitations and the pattern of the generated images. Still, there is more than one type of glaucoma, which can also affect results generation

if the region does not have a similar pathology.

It will be possible to obtain even better results from future adjustments in the neural network and the extraction algorithm since the hypotenuse measurement over the excavation region is simple and reproducible. New studies are considered since we have a range of opportunities for studies and architectures to explore, looking for better results in extracting the ONH region. Following the line of segmentation, new studies and developments described by the academic community could help perform the results obtained. The cup region also needs further research to see if other methods could be better than the actual ones created, looking for more accurate results and a possible comparison with the present results. Also, perform a way that the results of the whole pipeline turn closer to what the doctor could have done manually. Finally, the step of micro conversion could be reproduced using an NN, selecting the region of interest, and training a new model to see if the identification could be performed inside the pipeline without a third party, which is still manual in this study.

Moreover, the research contributes to the possibility of making this approach viable as a support tool in diagnosing patients with glaucoma. In this first study, we had the opportunity to present at two different conferences. First was the Brazilian Congress of Ophthalmology, and second was the Conference on Intelligent Systems.

REFERENCES

ALSAYAT, A. et al. Multi-layer preprocessing and u-net with residual attention block for retinal blood vessel segmentation. **Diagnostics**, [S.l.], v. 13, n. 21, 2023.

ALZUBAIDI, L. et al. Review of deep learning: concepts, cnn architectures, challenges, applications, future directions. **Journal of Big Data**, [S.l.], v. 8, n. 1, p. 53, 2021.

ANDRADE, J. C. F. et al. Elucidation of the role of the lamina cribrosa in glaucoma using optical coherence tomography. **Survey of Ophthalmology**, [S.l.], v. 67, p. 197–216, 1 2022.

AZAD, R. et al. Bi-directional ConvLSTM U-net with densley connected convolutions. **Proceedings - 2019 International Conference on Computer Vision Workshop, ICCVW 2019**, [S.l.], p. 406–415, 2019.

BERENGUER-VIDAL, R. et al. Decision trees for glaucoma screening based on the asymmetry of the retinal nerve fiber layer in optical coherence tomography. **Sensors**, [S.l.], v. 22, n. 13, 2022.

BEYKIN, G. et al. Discovery and clinical translation of novel glaucoma biomarkers. **Progress in Retinal and Eye Research**, [S.l.], v. 80, 1 2021.

BOURNE, R. R. A.; KHATIB, T. The optic nerve head in glaucoma. **COMMUNITY EYE HEALTH JOURNAL**, [S.l.], 7 2023.

BOURNE, R. R. et al. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to vision 2020: the right to sight: an analysis for the global burden of disease study. **The Lancet Global Health**, [S.l.], v. 9, n. 2, p. e144–e160, 2021.

CARLSON, B. M. Chapter 7 - Special Senses—Vision and Hearing. In: CARLSON, B. M. (Ed.). **The human body**. [S.l.]: Academic Press, 2019. p. 177–207.

CHEN, H. et al. Ms-unet-v2: adaptive denoising method and training strategy for medical image segmentation with small training data. , [S.l.], 2023.

CHEN, Z. et al. Region-segmentation strategy for bruch’s membrane opening detection in spectral domain optical coherence tomography images. **Biomed. Opt. Express**, [S.l.], v. 10, n. 2, p. 526–538, Feb 2019.

CHOLLET, F. **Deep learning with python**. New York, USA: Manning Publications Co., 2018.

DEVALLA, S. K. et al. Towards label-free 3d segmentation of optical coherence tomography images of the optic nerve head using deep learning. **arXiv**, [S.l.], v. 11, n. 11, p. 6356–6378, 2020.

DOWLING, J. E. Retina. In: RAMACHANDRAN, V. S. (Ed.). **Encyclopedia of the human brain**. New York: Academic Press, 2002. p. 217–235.

DREXLER, W.; FUJIMOTO, J. G. State-of-the-art retinal optical coherence tomography. **Progress in Retinal and Eye Research**, [S.l.], v. 27, n. 1, p. 45–88, 2008.

F. COSTA, L. da. Further generalizations of the jaccard index. **CoRR**, [S.l.], v. abs/2110.09619, 2021.

FEA, A. M. et al. Precision medicine in glaucoma: artificial intelligence, biomarkers, genetics and redox state. **International Journal of Molecular Sciences**, [S.l.], v. 24, n. 3, 2023.

FU, H. et al. Automatic Optic Disc Detection in OCT Slices via Low-Rank Reconstruction. **IEEE Transactions on Biomedical Engineering**, [S.l.], v. 62, n. 4, p. 1151–1158, 2015.

FU, Z. et al. Mpg-net: multi-prediction guided network for segmentation of retinal layers in oct images. **European Signal Processing Conference**, [S.l.], v. 2021-Janua, p. 1299–1303, 2021.

FUKUSHIMA, K. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. **Biological Cybernetics**, [S.l.], v. 36, n. 4, p. 193–202, 1980.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. [S.l.]: MIT Press, 2016. <http://www.deeplearningbook.org>.

GOPINATH, K.; RANGREJ, S. B.; SIVASWAMY, J. A deep learning framework for segmentation of retinal layers from OCT Images. **Proceedings - 4th Asian Conference on Pattern Recognition, ACPR 2017**, [S.l.], p. 894–899, 2018.

GU, Z. et al. CE-Net: context encoder network for 2d medical image segmentation. **IEEE Transactions on Medical Imaging**, [S.l.], v. 38, n. 10, p. 2281–2292, 2019.

HE, K. et al. **Deep residual learning for image recognition**. 2015.

JANG, H. et al. **Puca**: patch-unshuffle and channel attention for enhanced self-supervised image denoising. 2023.

JEONG, J.; YOON, T.; PARK, J. Towards a meaningful 3d map using a 3d lidar and a camera. **Sensors**, [S.l.], v. 18, p. 2571, 08 2018.

KERMANY, D. S. et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. **Cell**, [S.l.], v. 172, n. 5, p. 1122–1131.e9, feb 2018.

KHALIL, T. et al. Detection of Glaucoma Using Cup to Disc Ratio from Spectral Domain Optical Coherence Tomography Images. **IEEE Access**, [S.l.], v. 6, p. 4560–4576, 2018.

KIM, T. W. et al. Imaging of the lamina cribrosa in glaucoma: perspectives of pathogenesis and clinical applications. **Current Eye Research**, [S.l.], v. 38, p. 903–909, 9 2013.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, 2012. **Anais...** Curran Associates: Inc., 2012. v. 25.

LAVINSKY, F. et al. Measurement of the hypotenuse of the vertical optic nerve head cup with spectral-domain optical coherence tomography for the structural diagnosis of glaucoma. **Clinical Ophthalmology**, [S.l.], v. 12, p. 215–225, 2018.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, [S.l.], v. 521, n. 7553, p. 436–444, 2015.

LECUN, Y. et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, [S.l.], v. 86, n. 11, p. 2278–2324, 1998.

LEITGEB, R. A. et al. Enhanced medical diagnosis for doctors: a perspective of optical coherence tomography. **Journal of Biomedical Optics**, [S.l.], v. 26, n. 10, p. 100601, 2021.

LI, J. et al. Multi-scale gcn-assisted two-stage network for joint segmentation of retinal layers and disc in peripapillary oct images. , [S.l.], feb 2021.

LI, Z. et al. A survey of convolutional neural networks: analysis, applications, and prospects. **IEEE Transactions on Neural Networks and Learning Systems**, [S.l.], v. 33, n. 12, p. 6999–7019, 2022.

LU, J. et al. Automated segmentation and quantification of calcified drusen in 3d swept source oct imaging. **Biomed. Opt. Express**, [S.l.], v. 14, n. 3, p. 1292–1306, Mar 2023.

MACCORMICK, I. J. C. et al. Accurate, fast, data efficient and interpretable glaucoma diagnosis with automated spatial analysis of the whole cup to disc profile. **PLOS ONE**, [S.l.], v. 14, n. 1, p. 1–20, 01 2019.

MANTRAVADI, A. V.; VADHAR, N. Glaucoma. **Primary Care - Clinics in Office Practice**, [S.l.], v. 42, n. 3, p. 437–449, 2015.

MARY, S. P. et al. A Survey on Image Segmentation Using Deep Learning. In: JOURNAL OF PHYSICS: CONFERENCE SERIES, 2020. **Anais...** [S.l.: s.n.], 2020. v. 1712, n. 1.

MCMONNIES, C. W. Historial de glaucoma y factores de riesgo. **Journal of Optometry**, [S.l.], v. 10, n. 2, p. 71–78, 2017.

MIN, S.; LEE, B.; YOON, S. **Deep learning in bioinformatics**. 2017. 851–869 p. v. 18, n. 5.

PATEL, D. **Guide for glaucoma | types, symptoms, surgery and medication**. 2023.

PHADIKAR, P. et al. The potential of spectral domain optical coherence tomography imaging based retinal biomarkers. **International Journal of Retina and Vitreous**, [S.l.], v. 3, n. 1, p. 1–10, 2017.

PODOLEANU, A. G. Optical coherence tomography. **Journal of Microscopy**, [S.l.], v. 247, n. 3, p. 209–219, 2012.

PRADA, D. et al. Autoregulation and neurovascular coupling in the optic nerve head. **Survey of Ophthalmology**, [S.l.], v. 61, n. 2, p. 164–186, 2016.

PRANANDA, A. R. et al. Retinal nerve fiber layer analysis using deep learning to improve glaucoma detection in eye disease assessment. **Applied Sciences**, [S.l.], v. 13, n. 1, 2023.

QUIGLEY, H. A. Glaucoma. **The Lancet**, [S.l.], v. 377, n. 9774, p. 1367–1377, 2011.

RAHMAN, M. H. et al. Automatic quantification of anterior lamina cribrosa structures in optical coherence tomography using a two-stage cnn framework. **Sensors**, [S.l.], v. 21, n. 16, 2021.

REZATOFIGHI, H. et al. **Generalized intersection over union**: a metric and a loss for bounding box regression. 2019.

RONCERA, V. UM ESTUDO DE SEGMENTAÇÃO DE IMAGENS BASEADO EM UM MÉTODO DE COMPUTAÇÃO EVOLUCIONÁRIA. , [S.l.], n. 1, 2005.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-Net: convolutional networks for biomedical image segmentation. **IEEE Access**, [S.l.], v. 1, p. 16591–16603, 2015.

ROY, A. G. et al. **Relaynet**: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks. 2017.

RUSSEL, S.; NORVIG, P. **Artificial intelligence**: a modern approach. New Jersey, USA: Pearson, 2020. v. Fourth Edition.

SALAZAR, J. J. et al. Anatomy of the human optic nerve: structure and function. In: FERRERI, F. M. (Ed.). **Optic nerve**. Rijeka: IntechOpen, 2018.

SANDER, B. et al. Enhanced optical coherence tomography imaging by multiple scan averaging. **British Journal of Ophthalmology**, [S.l.], v. 89, n. 2, p. 207–212, 2005.

SCHMIDT-ERFURTH, U. et al. Artificial intelligence in retina. **Progress in Retinal and Eye Research**, [S.l.], v. 67, n. August, p. 1–29, 2018.

SCHUSTER, A. K. et al. The diagnosis and treatment of glaucoma. **Deutsches Arzteblatt International**, [S.l.], v. 117, n. 13, p. 225–234, 2020.

SCHWARTZ, S. D. et al. Chapter 68 - Retinal Degeneration. In: LANZA, R.; LANGER, R.; VACANTI, J. (Ed.). **Principles of tissue engineering (fourth edition)**. Fourth Edition. ed. Boston: Academic Press, 2014. p. 1427–1440.

SEDAI, S. et al. Uncertainty guided semi-supervised segmentation of retinal layers in OCT images. , [S.l.], p. 1–9, 2021.

SHELHAMER, E.; LONG, J.; DARRELL, T. Fully convolutional networks for semantic segmentation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, [S.l.], v. 39, n. 4, p. 640–651, 2015.

SHIN, H. Y. et al. Glaucoma diagnostic accuracy of optical coherence tomography parameters in early glaucoma with different types of optic disc damage. **Ophthalmology**, [S.l.], v. 121, n. 10, p. 1990–1997, 2014.

SIMONYAN, K.; ZISSERMAN, A. **Very deep convolutional networks for large-scale image recognition**. 2015.

SZEGEDY, C. et al. **Going deeper with convolutions**. 2014.

SZELISKI, R. **Computer vision**: algorithms and applications. [S.l.]: Springer, 2010. v. First Edition.

TAHA, A. A.; HANBURY, A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. **BMC Medical Imaging**, [S.l.], v. 15, n. 1, 2015.

VARATHARASAN, V. et al. Improving learning effectiveness for object detection and classification in cluttered backgrounds. **CoRR**, [S.l.], v. abs/2002.12467, 2020.

WAIBEL, A. et al. Phoneme recognition using time-delay neural networks. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, [S.l.], v. 37, n. 3, p. 328–339, 1989.

WANG, Y. X.; PANDA-JONAS, S.; JONAS, J. B. Optic nerve head anatomy in myopia and glaucoma, including parapapillary zones alpha, beta, gamma and delta: histology and clinical features. **Progress in retinal and eye research**, [S.l.], v. 83, 2021.

XU, J. et al. Automated volumetric evaluation of stereoscopic disc photography. **Optics Express**, [S.l.], v. 18, n. 11, p. 11347, 2010.

XUE, S. et al. Cts-net: a segmentation network for glaucoma optical coherence tomography retinal layer images. **Bioengineering**, [S.l.], v. 10, n. 2, 2023.

YU, F.; KOLTUN, V. Multi-scale context aggregation by dilated convolutions. **4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings**, [S.l.], 2016.

ZANG, P. et al. Automated segmentation of peripapillary retinal boundaries in OCT combining a convolutional neural network and a multi-weights graph search. **Biomedical Optics Express**, [S.l.], v. 10, n. 8, p. 4340, 2019.

ZHANG, Q. et al. Size and Shape of Bruch's Membrane Opening in Relationship to Axial Length, Gamma Zone, and Macular Bruch's Membrane Defects. **Investigative Ophthalmology & Visual Science**, [S.l.], v. 60, n. 7, p. 2591–2598, 06 2019.

ZHUANG, J. Laddernet: multi-path networks based on u-net for medical image segmentation. , [S.l.], p. 2–5, 2019.

ANNEX A – DATASET CLASS

The creation of the dataset using Tensorflow was using the tf.data.Dataset API where supports writing descriptive and efficient input pipelines. Dataset usage follows a common pattern:

1. Create a source dataset from the input data.
2. Apply dataset transformations to preprocess the data.
3. Iterate over the dataset and process the elements

The code below shows the class created to prepare the data used in the train model.

```
import cv2
import numpy as np
import tensorflow as tf

class Dataset:
    """Create a dataset using `from_tensor_slices`"""
    def __init__(self):
        self.imgs_gt = []      # list of images path
        self.imgs_mask = []    # list of masks path

    def create_dataset(self, batch=8):
        ds = tf.data.Dataset.from_tensor_slices((self.imgs_gt, self.
    imgs_mask))
        ds = ds.map(self.preprocess)
        ds = ds.batch(batch)
        ds = ds.prefetch(tf.data.AUTOTUNE)

        return ds

    def preprocess(self, x, y):
        def f(x, y):
            x = x.decode()
            y = y.decode()

            x = self.read_image(x)
            y = self.read_mask(y)

            return x, y

        image, mask = tf.numpy_function(f, [x, y], [tf.float32, tf.float32
    ])
        image.set_shape([640, 640, 3])
        mask.set_shape([640, 640, 1])

        return image, mask
```

```
34
35      def read_mask(self, path):
36          x = cv2.imread(path, cv2.IMREAD_GRAYSCALE)
37          x = cv2.resize(x, (640, 640))
38          x = x / 255.0
39          x = np.expand_dims(x, axis=-1)
40          x = x.astype(np.float32)
41          return x
42
43      def read_image(self, path):
44          x = cv2.imread(path, cv2.IMREAD_COLOR)
45          x = cv2.resize(x, (640, 640))
46          x = x / 255.0
47          x = x.astype(np.float32)
48          return x
```

Listing A.1 – Dataset class code.