UNIVERSIDADE D
COIMBRA

Maria Lourenço e Castro

# ROBUST METHODOLOGY FOR OCT SEGMENTATION

## SELECTION OF THE BEST ALGORITHM BASED ON CROSSED METRICS

**Dissertação no âmbito do Mestrado Integrado em Engenharia Física, ramo de Instrumentação orientada pelo Doutor Luís Mendes e pelo Professor Doutor Jorge Henriques e apresentada ao Departamento de Física da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.**

Setembro de 2020

This work was developed at:

# Resumo

A Tomografia de Coerência Ótica(TCO) é uma técnica de imagiologia que permite obter imagens volumétricas da estrutura da retina. A segmentação automática dos volumes produzidos por esta técnica é um processo complexo e, embora existam alguns algoritmos capazes de produzir bons resultados para volumes com alguns tipos de lesão, ainda não existe um algoritmo capaz de dar uma resposta adequada em todas as situações.

Este projeto de investigação propõe uma nova abordagem para a segmentação de volumes de TCO. Em vez de aplicar apenas um algoritmo de segmentação, a estratégia consiste em selecionar o melhor algoritmo de segmentação para cada situação, a partir de um conjunto pré-definido de algoritmos.

Neste projeto foi considerada uma base de dados com 123 volumes TCO, previamente segmentados por um especialista, de pacientes diabéticos. Em relação aos algoritmos de segmentação, foram considerados quatro algoritmos baseados na teoria de grafos para a segmentação de volumes TCO.

O critério de seleção toma como entrada as métricas cruzadas, que são métricas calculadas entre a segmentação produzida por cada um dos algoritmos, e devolve o melhor algoritmo para a segmentação em questão. Este critério foi desenvolvido com base em algoritmos de Machine Learning para classificação. Foram consideradas várias configurações para a seleção dos algoritmos e para a utilização das métricas na implementação dos diferentes algoritmos de classificação.

Os resultados obtidos neste trabalho evidenciam que a seleção do melhor algoritmo para a segmentação permite alcançar melhores resultados do que a implementação individual de qualquer um dos algoritmos. Os melhores resultados para o critério de seleção foram obtidos para a aplicação de uma Rede Neuronal Artificial na seleção do melhor algoritmo em cada B-scan, com a utilização de todas as métricas como entrada. Este modelo produziu uma sensibilidade de 95%, 17% e 71%, especificidade de 91%, 65% e 77% e Fscore de 93%, 27% e 74% para a classificação individual de cada algoritmo

de segmentação, quando aplicado num grupo de dados de teste.

O critério desenvolvido foi usado para produzir a segmentação dos volumes de teste, implementando a seleção do melhor algoritmo. Os resultados mostram uma redução média de $7,99\%$ do desvio absoluto médio, quando se comparam as segmentações obtidas com a seleção e as segmentações produzidas pelo algoritmo que obtém melhores resultados individualmente.

Palavras chave: Tomografia de Coerência Ótica, Segmentação, Métricas Cruzadas, Machine Learning.

# Abstract

Optical Coherence Tomography(OCT) is an imaging technique that allows the examination of the retinal structure. The automated segmentation of the volumes produced by this technique is a complex process and, although there are some algorithms that are capable of producing good results on volumes with some types of lesions, there is still no algorithm capable of performing well on all situations.

This research project proposes a new approach for the segmentation of OCT volumes. Instead of applying only one segmentation algorithm, the strategy is to select the best segmentation algorithm for each situation, from a pre-defined set of algorithms.

For this project a data base with 123 OCT volumes, previously segmented by an expert, from diabetic patients was considered. Regarding the segmentation algorithms, four algorithms based on the graph theory approach for OCT segmentation were considered.

The selection criteria takes as input the crossed metrics, which are metrics calculated between the segmentation produced by each algorithm, and outputs the segmentation produced with the best algorithm. This criterion is developed based on Machine Learning algorithms for classification. For the implementation of the different classification algorithms, several configurations were considered for the metrics and for the selection of the best algorithm.

The results obtained in this work show that the selection of the best algorithm for the segmentation leads to better results then the individual implementation of any of the algorithms. The best results for the selection criteria were obtained for the application of an Artificial Neural Network for the selection of the best segmentation algorithm for each B-scan, using all the metrics available as input. This model produced a precision of 95%, 17% e 71%, a recall of 91%, 65% e 77% and a F-score of 93%, 27% e 74% for the classification of each individual segmentation algorithm, when applied to a test data set.

The developed criteria was applied for the segmentation of the volumes on the test data set, implementing the selection of the best algorithm. The results show a mean reduction of $7,99\%$ of the mean unsigned error, when comparing the segmentation obtained with the selection and the segmentations obtained with the algorithm that presents the best results.

Key words: Optical Coherence Tomography, Segmentation, Crossed Metrics, Machine Learning.

# Nomenclature

| | |
|---|---|
| $E_{95}$ | Ninety-Fifth Percentile |
| AD | Alzheimer's Disease |
| AMD | Age-related Macular Degeneration |
| ANN | Artificial Neural Networks |
| DR | Diabetic Retinopathy |
| DT | Decision Tree |
| ELM | External Limiting Membrane |
| FA | Fluorescein Angiography |
| FD-OCT | Fourier Domain Optical Coherence Tomography |
| FN | False Negatives |
| FP | False Positives |
| GCL | Ganglion Cell Layer |
| ILM | Internal Limiting Membrane |
| INL | Inner Nuclear Layer |
| INL | Inner Plexiform Layer |
| LTD | Layer Thickens Difference |
| ML | Machine Learning |
| MS | Multiple Sclerosis |

MSE Mean Signed Error

MUE Mean Unsigned Error

NFL Nerve Fiber Layer

OCT-A Optical Coherence Tomography Angiography

OLM Outer Limiting Membrane

ONL Outer Nuclear Layer

OPL Outer Plexiform Layer

RF Random Forest

RPE Retinal Pigment Epithelium

RTD Retinal Thickness Difference

SD-OCT Spectral Domain Optical Coherence Tomography

SE Signed Error

SS-OCT Swept-Source Optical Coherence Tomography

SSE Standard deviation of the Signed Error

SUE Standard deviation of the Unsigned Error

TD-OCT Time Domain Optical Coherence Tomography

TN True Negatives

TP True Positives

UE Unsigned Error

# List of Figures

# List of Tables

# Contents

# Chapter 1

# Introduction

## 1.1 Context

OCT is an imaging technique used for obtaining cross-sectional volumetric images of the retina. It is a growing technique used to assist the diagnosis of pathologies such as Age-related Macular Degeneration(AMD) [40], Diabetic Retinopathy(DR) [22], Glaucoma [6], Multiple Sclerosis(MS) [15] and Alzheimer's Disease(AD) [26].

The segmentation of OCT volumes consists of the identification and delineation of the retinal structures. Based on the segmentation, several metrics can be computed, allowing the extraction of quantitative values to assist the diagnosis, find new biomarkers, evaluate the progression of the disease and adjust the treatment plan accordingly.

The manual segmentation is not viable for daily exam analysis since it is a very time-consuming process. This way, the only way to introduce the segmentation of OCT scans on daily clinical practice is by implementing automatic segmentation algorithms. However, although the retinal segmentation (segmentation of the first and last layers of the retina) is already available in clinical practice, the retinal layer segmentation represents a complex problem and there is still no algorithm that can perform well on all types of images.

Nowadays, several OCT segmentation algorithms produce good results for OCT volumes from healthy subjects and some are well suited for the segmentation of OCT volumes from patients with certain pathologies. However, there is not yet an algorithm that is versatile and can deliver the required accuracy to guarantee the necessary clinical confidence.

## 1.2 Aim of this project

The approach to the segmentation problem presented in this project consists of developing a method for the selection of the best segmentation algorithm in each situation. This approach considers a set of predefined algorithms and takes advantage of the quality of each one for solving specific problems, providing a more versatile solution.

To evaluate the performance of a segmentation algorithm, usually, a set of metrics are implemented between the segmentation produced by that algorithm and a reference segmentation, manually traced by an expert. These metrics are denominated reference metrics. In this project, the goal is to use this set of metrics but apply them between the segmentation produced by different segmentation algorithms. These metrics, calculated between the segmentation of each algorithm, are denominated crossed metrics and will be used to select the best algorithm.

This project aims to:

1. Prove the benefit of selecting the best algorithm on each situation, instead of considering only the same algorithm for all situations.

2. Develop a method to implement the selection of the best algorithm

## 1.3 Outline

This thesis starts by giving a theoretical approach to the topics that involve the research project. In chapter 2, the eye anatomy and the OCT imaging technique are introduced, as well as their interconnection, this is, the use of the OCT imaging in clinical practice. In chapter 3, the algorithms used for the automated OCT segmentation are presented, as well as the metrics used to evaluate them. Still in this chapter, the machine learning algorithms that are used on this project are presented, as well as the metrics used to evaluate their performance.

After introducing the topics needed for the project, the methods and materials used are explained in chapter 4. Then, in chapter 5, the results produced are presented and discussed.

Finally, in chapter 6, the conclusions that resulted from this research project are presented, and future work is proposed.

# Chapter 2

# Background

## 2.1 Visual system

### 2.1.1 Overview

The visual system enables the analysis of the surrounding environment by processing the information of the light that reaches the eyes on the visible part of the electromagnetic spectrum. The eye is responsible for converting the light signal into a neural signal that is then transmitted to the brain to be interpreted.

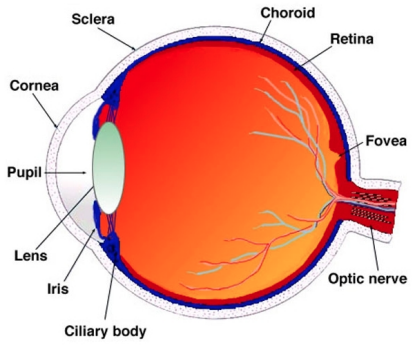


Figure 2.1: Eye anatomy. Reproduced from [27]

A scheme of the eye anatomy is presented in Figure 2.1. In very general terms, two main activities take place on the eye. First, the light that reaches the eye goes through an optic process that produces an image with a focus on the retina. After this, the light signal that reaches the retina is converted

into a neural signal to be transmitted to the brain.

Three main structures act on the light that enters the eye: the iris, the cornea and the crystalline (lens). The iris is responsible for controlling the amount of light that enters the eye through the pupil, varying the aperture diameter given the light conditions of the surrounding ambient. The cornea and the crystalline lens are responsible for placing the focus of the images that reach the eye on the retina.

The retina covers most of the interior of the eye and is responsible for converting the image that is formed on its surface into a neural signal. The neural signals are then transmitted to different parts of the brain, through the Optic Nerve, to be processed.

### 2.1.2 Retina anatomy and physiology

The retina is composed by ten layers, as presented in Figure 2.2.

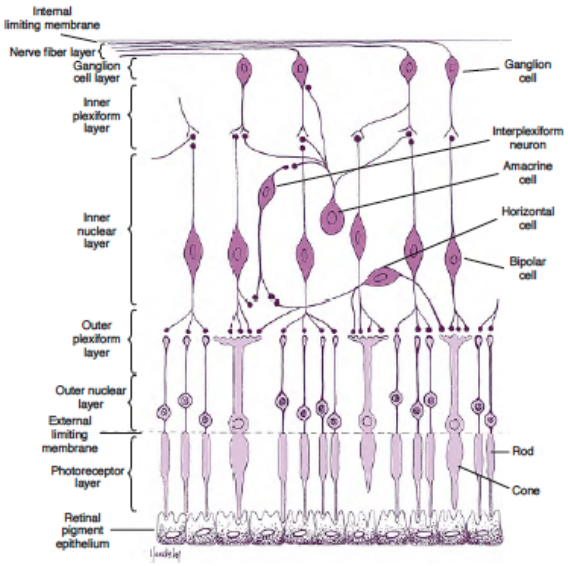


Figure 2.2: Retinal structure. Reproduced from [40].

Starting by the outer layer, the ten layers of the retina are [40]:

1. Retinal Pigment Epithelium (RPE)
2. Photoreceptor layer
3. External Limiting Membrane(ELM) or Outer Limiting Membrane(OLM)

4. Outer Nuclear Layer(ONL)
5. Outer Plexiform Layer(OPL)
6. Inner Nuclear Layer(INL)
7. Inner Plexiform Layer(IPL)
8. Ganglion Cell Layer(GCL)
9. Nerve Fiber Layer(NFL)
10. Internal Limiting Membrane(ILM)

The RPE is the outermost layer of the retina and consists of a single layer of hexagonal pigmented cells. It is connected to the choroid's Brunch membrane and controls the movement of nutrients from the choroid to the retina and the discharge of waste products from the retina.

The RPE is connected to photodetector cells, as shown in Figure 2.2. The photoreceptor cells are the Rods and Cones, which are photosensitive and convert photons into neural signals. The rods are more sensitive under low illumination conditions while the cones are more active in environments with better illumination conditions. The distribution of both cells is also not equal on all the Retina surface.

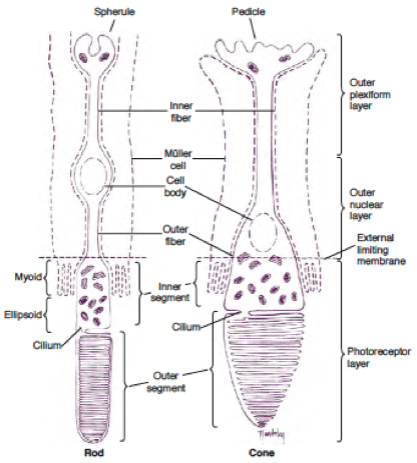


Figure 2.3: Anatomy of Rods and Cones. Reproduced from [40].

The structure of the photoreceptor cells is divided into four major parts, as shown in Figure 2.3. The outer segment is connected to the RPE in

some of the cells, as shown in Figure 2.2, and it is where the absorption of photons initiates a biochemical reaction that then leads to the neural signal. The inner segment is connected to the outer segment and is where the accumulation of energy and protein production takes place. The outer fibre connects the inner segment to the cell body, where the nucleus of the cell is. The inner fibre connects the cell body to structures that are specialized for the synapse, the Spherule and the Pedicle.

The Rodes and Cones are divided into different layers of the retina structure. The Photoreceptor layer contains the outer and inner segments of the Rods and Cones. The ELM is not a structure, but the connection between the inner segment and the outer fibre. The ONL extends from the ELM to the cell body of the cones, containing the outer fibre and the cell body of both cells. Finally, the OPL contains the inner fibre and the synapse structures that are responsible for the junction between the photoreceptor cells and the cells from the INL, bipolar cells and horizontal cells.

The INL contains the cell bodies of the Bipolar cells, Horizontal cells, Amacrine cells and Interplexiform neurons. The properties of the cells and the synapses that occur in this layer enables the process of motion detection, brightness changes, recognition of contrast and tone.

The GCL is a single-cell thick layer composed of ganglion cells that are connected to the cells of the INL.

The NFL consists of the ganglion cells axons from the GCL. The fibres are parallel to the retina surface once they are all connected to the Optic Disc. The thickness of this layer is higher near the Optic Disc due to the accumulation of the fibres from all the Retina.

The process of generating a neural signal from the light that reaches the retina surface starts at the more outer layer of the retina, the photodetector layer. This means that the light passes through all the layers of the retina before being detected. After this, the neural signal travels in the inverse direction, passing through all the layers to be processed and sent to the Optic Disc on ganglion cell axons.

## 2.2 Optical Coherence Tomography

Optical Coherence Tomography is an imaging technique that produces a volumetric image of the retinal structure.

### 2.2.1 Time Domain OCT

The first OCT technique was based on the Michelson Interferometer. It has a low-coherence light source, a 50/50 beamsplitter, a reference mirror and a photodetector.

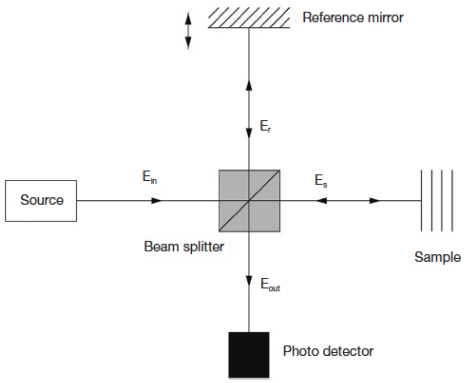


Figure 2.4: Michelson Interferometer scheme. Reproduced from [5]

As shown in Figure 2.4, the light arrives at the beamsplitter and two beams are produced with half the intensity of the incident beam. One of the beams hits the reference mirror, where it is reflected back to the beamsplitter. The other goes to the eye that is being examined, interacts with the tissue and is reflected back to the beamsplitter. On the beamsplitter, the two beams interact, producing interferometric fringes. Then, the beam resulting from that interaction goes to the photodetector [10].

The optical properties of the tissue are obtained taking into consideration the interferometric fringes that arrive at the photodetector and the information on the reference optical path. Being the reference optical path related to the distance between the reference mirror and the beamsplitter. This way, varying the distance between the reference mirror and the beamsplitter allows the extraction of the optical properties of the tissue as a function of the depth of the sample. This signal of the optical properties as a function of depth is called an A-scan.

Obtaining the A-scans on a certain axis allows the construction of a 2D cross-sectional image of the retina structure. This image is called a B-scan. Repeating the acquisition of B-scans along an axis that is perpendicular to that of the B-scan allows the construction of a 3D image of the retina.

The presented technique, based on the Michelson interferometer, is the Time Domain Optical Coherence Tomography(TD-OCT) and it was the first technique for the extraction of OCT scans. However, it is a slow process for obtaining 3D-OCT scans once it involves moving parts.

### 2.2.2 Fourier Domain OCT

The alternative, to avoid the need for moving parts, is the Fourier Domain Optical Coherence Tomography(FD-OCT) .

For the FD-OCT the position of the reference mirror is fixed. This way, the difference on the optic path between the reference path and the path on the sample is obtained with the information on the interferometric fringes as a function of the spectrum of the light source.

To obtain the interferometric fringes as a function of the spectrum of the source, there are two possible configurations: the Spectral Domain Optical Coherence Tomography(SD-OCT) and the Swept-Source Optical Coherence Tomography(SS-OCT) .

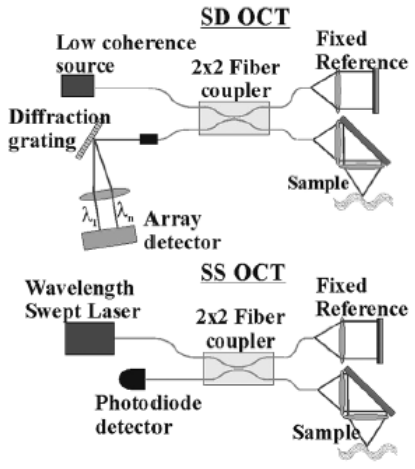


Figure 2.5: SD-OCT and SS-OCT working scheme. Reproduced from [24]

The SD-OCT uses a grating to disperse the spectrum across an array of light detectors while the SS-OCT uses a narrow band laser that allows the

source to change the wavelength, considering a time interval for the wavelength change that is compatible with the sample time of the photodetector. This way, the SD-OCT analyses the fringes as a function of position on the array, once each detector of the array will detect a certain wavelength, and the SS-OCT analyses the fringes as a function of time, since the wavelength of the sample is a function of time. A scheme of the operation of each technique is presented in Figure 2.5.

For comparing the speed of the TD-OCT and the SD-OCT lets consider the Stratus OCT and the Cirrus HD-OCT, both commercialised by Carl Zeiss Meditec. The first is a TD-OCT machine and can produce 400 A-scans/sec while the second is an SD-OCT machine and produces 27000 A-scans/sec [24]. Because of the abnormalities caused by the movement of the eye, the acquisition of the image is limited in time. This way, the speed of the acquisition process is very important for image quality. The TD-OCT here presented produces typically images of 6x128x1024 voxels in 1.92 seconds while the SD-OCT presented produces images of 200x200x1024 voxels in 1.48 seconds [24]. This data shows the difference in the image quality that can be obtained with the SD-OCT.

### 2.2.3 OCT Angiography

The gold standard for the analysis of the vascular structure of the eye is to preform a Fluorescein Angiography(FA). This process consists of administrating a fluorescent dye to the patient, through an intravenous injection, and obtaining a colour fundus photography of the eye. To obtain the FA, the wavelength that is used to illuminate the retina is the excitement wavelength of the fluorescein dye that is used. The image is then obtained with a filter to allow the analysis of the wavelength emitted by the dye, after being excited. This allows the construction of an image of the vascular structure by extracting the path of the dye on the eye. However, there are several complications associated with this procedure that include nausea, vomiting and pruritus.

Optical Coherence Tomography Angiography(OCT-A) is a new OCT technique that allows the construction of a blood flow map without the administration of any intravenous contrast, avoiding, this way, the complications associated with FA.

The operation principle of the OCT-A image is to use the moving red blood cells as an intrinsic contrast agent. The OCT signal generated by static tissue stays approximately the same over time, while the signal coming from blood vessel changes as the cells inside them are constantly moving.

This way, calculating the differences on the OCT signals acquired at the same location at different time points, it is possible to identify what parts of the image represent a steady tissue and what parts represent a blood vessel and, this way, construct the vessel structure of the eye [7].

### 2.2.4 Limitations

Besides the noise generated by the signal acquisition system, that is associated with the electronic components involved in the process, OCT images have an intrinsic speckle noise. As was said above, the signal used to generate the retinal cross-sectional images is based on the absorption and scattering properties of the tissues that are being analysed. The light signal that is used to measure these properties is scattered and absorbed by the medium. Since these interactions have always a random contribution, there is always a noise associated to them on the image. This noise is denominate speckle noise and is a characteristic of this imaging technique, due to the physical process used to obtain the images.

Besides the speckle noise, the light that arrives at the bottom layers of the retina is affected by the interference that happens on the top layers. This way, the intensity of the signal that interacts with the bottom layers is lower than the intensity of the signal that interacted with the top layers of the retina. This causes a reduction in the intensity of the signal as a function of depth that doesn't reflect the variation of the tissue properties. This property is known as amplitude-decorrelation. This is a limitation of the OCT imaging technique that affects the implementation of automated segmentation algorithms once tissues with the same optical properties don't appear on the image with the same intensity, as would be desired.

## 2.3 Clinical overview

The OCT imaging technique has proven to be useful on the diagnosis of several ophthalmic diseases such as DR, AMD and glaucoma, but also on neurological pathologies such as MS and AD.

Glaucoma causes loss of retinal ganglion cell, reducing the retinal NFL thickness [6]. This allows the detection of glaucoma with the OCT imaging technique. This pathology is mostly associated with high Internal Ocular Pressure values. This factor can affect the retinal NFL either by the increase of the pressure on that area or by the limitation of the blood perfusion. There are cases, however, where the NFL is damaged but the Internal Optic

Pressure is normal. On these cases, the cause is associated with a low perfusion pressure that leads to cell death [40]

AMD is associated with the transition of nutrients and waste products from the Choroid to the retina through the Bruch's Membrane. With age, there is an accumulation of lipids that inhibits the passage of water. This accumulation of lipids is a Drusen and causes accumulation of water between the RPE and the Brunch's Membrane, leading to displacement and, sometimes, detachment between the two structures. This process leads to a lack of nutrient supply for the retina that can cause atrophy of the RPE, followed by the death of photoreceptors, or generation of a neovascular membrane to compensate for the loss of nutrients. Both of these situations cause vision loss that is a symptom of AMD [40]. The Drusen volume is indicative of the AMD diagnosis and severity [1] and, since the OCT imaging technique allows the cross-sectional visualisation of the full drusen and subsequent calculation of it's volume, it is an indicated method for the diagnosis of AMD.

The high glucose levels on the blood of diabetic patients cause vascular disruption that is associated with retinopathy. The features of this pathology include venous dilatation, microaneurysms, retinal haemorrhages, retinal edema and hard exudates [9]. DR patients have been reported to have an increase in the retina thickness that can be detected by OCT imaging [18]. Besides this, the new OCT-A imaging allows a better visualisation of the changes in the vascular structure of the retina, caused by the progression of the disease [22].

The AD is a progressive neurodegenerative disorder that causes cognitive impairment and a decline in learning and executive functions. Recent studies have pointed out the relation between Alzheimer's disease diagnosis and the reduction of the thickness of the retinal NFL [26]. This hew biomarker shows the potential of the OCT imaging technique on the early diagnosis of the AD.

MS is a condition that affects the brain and spinal cord causing symptoms such as vision problems, arm or leg movement restrictions or balance problems. Recent studies have shown the relation between the retinal NFL atrophy and the multiple sclerosis diagnosis [15], proving the potential of the OCT imaging technique on the follow up of patients with MS.

These applications of the OCT imaging techniques show the importance of the integration of OCT segmentation on a clinical environment since all the thickness measurements are done based on the segmentation. The integration of automated segmentation algorithms would allow the clinicians to easily extract features such as retinal NFL thickness enabling the early

diagnosis of several conditions. This would also provide quantitative measurements to the physician to evaluate the progression of the diseases and adjust the treatment plan accordingly.

# Chapter 3

# State-of-art

## 3.1 Introduction

This chapter aims to present a review of the main topics of this thesis namely, OCT segmentation and Machine Learning(ML).

The OCT segmentation section is composed of two subsections: OCT segmentation algorithms and metrics used for evaluating the OCT segmentation. On the subsection of the OCT segmentation algorithms, a broad view of the different approaches that have been proposed to solve the automated OCT segmentation problem is presented. Besides the broad view, the graph theory algorithms are more profoundly addressed once the algorithms used in this research project are based on this approach. On the subsection of the metrics used for evaluating the OCT segmentation, a review of the metrics used to evaluate the segmentation is given. The metrics presented on this subsection will be used to evaluate the segmentation algorithms, by comparing the segmentation produced by the algorithms to a reference segmentation, and to provide the input of the model to select the best algorithm, by computing the crossed metrics.

On the ML section, a broad view of what is ML is presented with a focus on the ML classification models, which are the models that are used on this research project. After this review of the models, the metrics used to evaluate the performance of classification models is presented.

## 3.2 OCT segmentation

### 3.2.1 OCT segmentation algorithms

The development of an automated OCT segmentation algorithm faces several challenges associated with the limitations of the OCT imaging techniques, presented at 2.2.4, the shadows caused by the blood vessels, motion artefacts and the anatomical changes that some pathologies cause on the structure of the retina.

Facing this limitations, there are several approaches used for automatic OCT segmentation. These approaches can be split into five major categories namely, methods based on intensity variation, active contour, pattern recognition, graph theory and ML.

The first approaches on the problem of OCT segmentation consider the intensity variation or gradient information and apply peak search techniques on individual A-scans to find the boundaries between each retina layer [2, 4, 12, 19–21, 23, 28, 43, 45]. The disadvantages associated to these approaches include the need for advanced denoising methods, that are time-consuming, and the lack of ability to incorporate the 3D information.

Another approach on OCT segmentation consists on implementing active contour algorithms [17, 34–36, 49]. These algorithms act by deforming an initial curve towards the boundary to be segmented. This can be done by minimising an energy function that describes the contour. The algorithms based on this approach surpass the performance of the intensity-based algorithms in terms of accuracy and noise resistance.

Along with the active contour algorithms, the pattern recognition algorithms also surpass the intensity-based ones. These algorithms apply ML techniques such as clustering [33] and support vector machine algorithms [13] to find the best retinal layer segmentation.

The graph theory approach for segmentation of OCT data was first introduced by Garvin [16] and became the most commonly used once it can produce better accuracy results than the previous ones [24]. After the emergence of this approach, several approaches used the graph theory combined with other techniques. One of those upgrades consists of using ML algorithms to generate probability maps that are then segmented with a graph theory approach [11, 29].

Recently, with the emergence of deep learning, some studies propose segmentation algorithms that are based on the implementation of Deep Learning algorithms such as Convolutional Neural Networks [38, 41, 42]. These methods imply a big data set and computational power for the training pro-

cess but, once they are trained, the developed models are fast and can adapt to very distinct volumes, reporting accuracy results that are favourable, when compared with the previous approaches.

The algorithms used in this project are based on a 2D/3D graph theory approach. This is done because the algorithms that were considered needed to be ready to use and these algorithms were available and because, excluding most recent approaches that apply ML, these are the ones that present the best results. This way, in this section, there is an introduction to the graph theory followed by its implementation for OCT segmentation.

#### 3.2.1.1 Graph Theory for segmentation

Graph theory is a mathematical field that studies graphs which are diagrams composed by nodes and edges. The nodes, or vertex, are points and the edges, or arches, are lines that connect the nodes according to a neighbourhood system. Graphs can be presented as $G = (V, E)$ with $V$ being the set of nodes on the graph and $E$ being the set of edges [32]. A graph representation is shown on Figure 3.1.

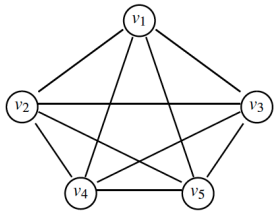


Figure 3.1: Graph structure. Reproduced from [31].

A graph can be weighted or unweighted. A weighted graph has a number, called weight or a cost, associated with each edge that represents one characteristic of the relation between the two nodes that are being connected. On unweighted graphs, the edges don't have costs associated representing only a relationship between two nodes [31].

A graph can also be directed or undirected. In directed graphs, the edges $< v_i.v_j >$ and $< v_j.v_i >$, that connect the nodes $v_i$ and $v_j$, have a preferential direction a and can have different costs. On undirected graphs $< v_i.v_j >$ and $< v_j.v_i >$ are equal and there is no preferential direction on the connection of the nodes. When presenting a graph with a diagram, the edges on the directed graphs are represented as arrows, while the edges on the undirected graphs are presented as lines [32]. Figure 3.2 Shows a

weighted and directed graph.

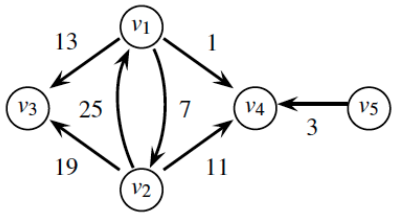


Figure 3.2: Graph diagram of a weighted and directed graph. Reproduced from [31].

The choice of using different types of graphs depends on the application for which it is being applied.

On 2006 Li [32] proposed the application of graph theory for the segmentation of 3D surfaces on medical imaging. This approach considers pixels as the nodes of the graph and the arches define the paths that the surface can take between pixels, as shown on Figure 3.3.

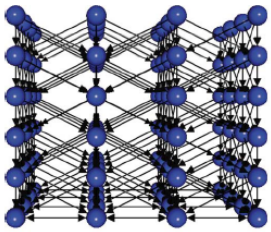

Figure 3.3: Graph presentation of the 3D image. Reproduced from [16].

The cost of a node is associated with the inverse of the probability of that node belonging to the surface that is being segmented. This means that the cost gets higher as the probability of the vertex belonging to the surface gets lower.

Since the purpose of this algorithm is to segment medical images, the surfaces are not expected to vary abruptly on a short distance because that is not the way that the tissues are distributed. This way, the authors implement smoothness constraints that limit the variation of the $z$ component of the surface along the $x$ and $y$ axes.

To establish the smoothness constraint it is convenient to present the graph as a set of columns on $(x, y)$ positions. The authors consider a four

neighbourhood system which means that each column will have four neighbours that are the columns closer to the one that is being considered. Each column will then have two neighbours on the $x$ axis and two on the $y$ axis. Considering the variation along the $x$ axis, the smoothness constrain $\Delta_x$ implies that, moving from the column $(x, y)$ to $(x + 1, y)$, the variation of the $z$ position of the surface is limited by the condition presented on equation (3.1).

$$|z - z'| \leq \Delta_x \tag{3.1}$$

Here, $z'$ is the position of the surface on the column $(x, y)$ and $z$ is the position of the surface on the column $(x+1, y)$. This constraint is presented in Figure 3.4. The same applies for the $y$ axis, being $\Delta_y$ the smoothness constraint on that case.

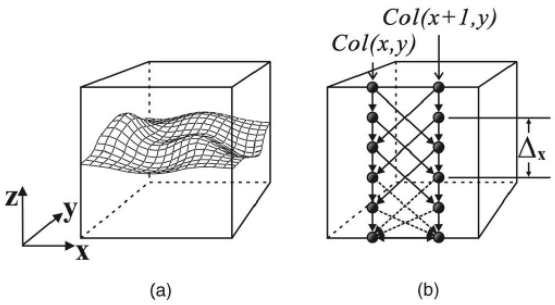


Figure 3.4: Graph presentation of the smoothness constraint. (a) Representation of the surface segmentation. (b) Representation of two adjacent columns along the $x$ axis and the correspondent smoothness constraint. Reproduced from [32].

Having the costs and the smoothness constraints, it is possible to apply an s-t cut algorithm and obtain the best surface for segmenting the image in two subsets.

The s-t cut algorithm consists of adding two columns on the opposite sides of the graph, the source $s$ and the sink $t$. Using this structure, an algorithm for the search of feasible surfaces, given the constraints that are established, is used. Having the feasible surfaces, the resulting surface is the one that has the minimum cost.

The approach presented is used for the segmentation of one surface. However, for the segmentation of $k$ surfaces, we need to have information on the relation between them. This way, the interaction constraints are implemented. These constraints limit the minimum and maximum distance

between each surface.

For the segmentation of $k$ surfaces, there are $k$ graphs representing the volume, with the costs associated with each surface. To implement the interaction constraint there are arches between each $(x, y)$ columns, on each sub-graph of the $k$ surfaces. These arches limit the distance between the surfaces that are being segmented on each sub-graph, and are represented in Figure 3.5.

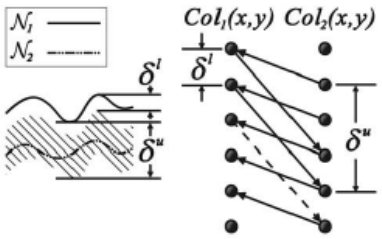


Figure 3.5: Graph presentation of the interaction constraints. Reproduced from [32].

The interaction constraints, represented in Figure 3.5, limit the separation between the surfaces $N_1$ and $N_2$. The arches connect $column_1$ and $column_2$, that are the columns on the same $(x, y)$ position on the sub-graphs for the segmentation of the two surfaces. The constraints represent the minimum, $\delta^l$, and maximum, $\delta^u$, distance between the surfaces $N_1$ and $N_2$.

This way, a 4D graph that is composed by the $k$ sub-graphs is obtained, with proper costs and edges for the segmentation of each surface and the arches between these sub-graphs represent the interaction constraints between each surface.

After establishing the graph structure, the process of segmenting the volume consists on generating the $k$ feasible surfaces, given the constraints that are established, with an s-t cut approach, and finding the set of surfaces that produce the minimum surface cost.

This approach presents the basis of the application of the graph theory for segmentation. On the following sections, the algorithms for OCT segmentation that will be used during this research work are explained. The first section presents an evolution of the 3D graph theory approach for OCT segmentation. The second section presents an approach that is based on this one but applies a 2D analysis, this is, performs the segmentation on each B-scan individually.

#### 3.2.1.2 3D graph theory for OCT segmentation

On 2009 Garvin [16] proposed a graph-based algorithm for the segmentation of SD-OCT scans based on the approach presented by Li [32]. The author proposed an improved version of the original segmentation approach that incorporates varying feasibility constraints and true regional information on the construction of the graph.

On the first approach, the vertex cost considers only an on-surface cost that represents the inverse of the probability of the vertex belonging to a certain surface. On this approach, it is implemented an in-region cost that represents the inverse of the probability of the vertex belong to a certain region. Here the regions represent the volumes below each surface. This way, the total cost of the set of surfaces that are being segmented is given by equation(3.2).

$$C_{f_1(x,y),f_2(x,y),\ldots,f_n(x,y)} = \sum_{i=1}^{n} C_{f_i(x,y)} + \sum_{i=1}^{n} C_{R_i} \tag{3.2}$$

This approach also implements varying smoothness and interactions constraints, opposingly to the Li approach that considers them as constants for all the image. This means that the constraints are defined for each $(x, y)$ position, based on training from previously segmented volumes.

The segmentation happens in tow steps. First, the author implements the segmentation of the surfaces ILM-NFL, ONL-Photodetector and Photodetector-RPE using an optimal 3D graph search approach. Using the location of the surfaces ILM-NFL and Photodetector-RPE, the author performs the flattening of the image. This is done aligning these two surfaces and adjusting the volume to this alignment. After this, the segmentation of the layers ILM-NFL, ONL-Photodetector and Photodetector-RPE are performed simultaneously. In the end, the algorithm does the segmentation of the surfaces NFL-GCL, GCL-INL, INL-OPL and OPL-Photodetector using a minimum s-t cut algorithm.

#### 3.2.1.3 2D graph theory for OCT segmentation

On 2010 Chiu [8] developed a graph theory approach for the segmentation of SD-OCT. The approach is based on the approach presented by Li [32] but applied for the segmentation of 2D images, instead of 3D.

To find the minimum cost path on the image with an automated endpoint selection, the author adds two columns on both sides of the image with arbitrary node cost and minimum weight associated with the vertex. This

allows for the cut to move freely on the vertical direction of these columns, as shown in Figure 3.6. Using Dijkstra's algorithm, the end columns will be connected by the path that contains the minimum cost vertex and, this way, the minimum cost path is found without the need for manual selection of the endpoint.

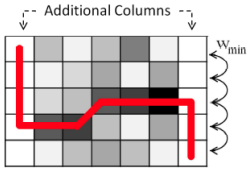


Figure 3.6: Representation of the automatic endpoint selection for finding the minimum s-t cut path.Reproduced from [8].

On this algorithm, the first step is to do the flattening of the images. This is done by estimating the RPE layer based on pixel intensity and shifting the pixels to make this layer flat.

After flattening, the weights are assigned and the NFL and ELM are segmented. Using the segmentation of these two layers, the algorithm searches for limiting the search space for the next layers to be segmented. Then, the segmentation of NFL-GCL surface is performed and then the segmentation of the surfaces between IPL and ONL. In the end, the segmentation of the ELM to the choroid is established and the image is unflattened with the respective segmentation.

Comparing Chiu and Garvin's approach, this is, the 2D and 3D graph theory approaches for OCT segmentation, the second is expected to produce better results for the once it considers the constraints across the volume, instead of just the B-scan information. The Chiu's approach is expected to show some discontinuities between each B-scan once there is no constraint to limit the variation between consecutive B-scans.

### 3.2.2 Metrics used for evaluating the OCT segmentation

This section presents a review of the metrics used for comparing the segmentation of OCT volumes. These metrics are used to validate the accuracy of automatic segmentation algorithms in comparison with a reference segmentation that is usually produced by an expert. On this project, besides evaluating the segmentation by implementing the metrics with a reference segmentation, the metrics will be implemented between the segmentation

produced by two different algorithms, the crossed metrics, that will be used as input of the selection algorithm.

Each segmented layer produced by an algorithm is represented by a vector with the position of all the pixels that are on the surface. On this analysis the representation presented by Tian [47] is considered, this is, the vector resulting from the automatic segmentation is presented as $B_{m,n}$ and the one from the reference segmentation is presented as $\bar{B}_{m,n}$, being $m, n$ the dimensions of the matrix.

The most common metric to evaluate the difference between the automated segmentation and the reference segmentation is the Unsigned Error(UE) [3, 14, 16, 29, 30, 37, 39, 44, 47, 48], or Unsigned Border Position. It is defined by equation 3.3.

$$UE = MUE \pm SUE \tag{3.3}$$

Here, the Mean Unsigned Error(MUE) is defined by 3.4 and the Standard deviation of the Unsigned Error(SUE) is defined by equation 3.5.

$$MUE = \mu(|B^i_{m,n} - \bar{B}^i_{m,n}|) \tag{3.4}$$

$$SUE = \sigma(|B^i_{m,n} - \bar{B}^i_{m,n}|) \tag{3.5}$$

In equation 3.4, $\mu$ represents the mean and $i$ represents each position on the $B_{m,n}$ and $\bar{B}_{m,n}$ matrix. The MUE is the mean of the absolute distance between the two segmentation on each point. In equation 3.5, $\sigma$ represent the standard deviation and $i$ represents each position on the $B_{m,n}$ and $\bar{B}_{m,n}$ matrix. This way, the SUE is the standard deviation of the distances between the two segmentation on each position.

The UE is a common metric because it represents the average of the distances between the two segmentation in the analysis.

Other metric that is also common is the Signed Error(SE) [14, 16, 29, 30, 47, 48], or the Signed Border Position. It is defined as presented in equation 3.6.

$$SE = MSE \pm SSE \tag{3.6}$$

Here the Mean signed Error(MSE) is defined by equation 3.7 and the Standard deviation of the Signed Error(SSE) is defined by equation 3.8.

$$MSE = \mu(B^i_{m,n} - \bar{B}^i_{m,n}) \tag{3.7}$$

$$SSE = \sigma(B^i_{m,n} - \bar{B}^i_{m,n}) \tag{3.8}$$

In equation 3.7, $\mu$ represents the mean and $i$ represents each position on the $B_{m,n}$ and $\bar{B}_{m,n}$ matrix. The MSE is the mean of the distance between the two segmentation on each point. In equation 3.8, $\sigma$ represent the standard deviation and $i$ represents each position on the $B_{m,n}$ and $\bar{B}_{m,n}$ matrix. This way, the SSE is the standard deviation of the distances between the two segmentation on each position.

This metric is similar to the UE, but the SE allows for negative values of distance when the position of the reference segmentation is above the automated segmentation. This leads to lower values than when comparing the MSE to the MUE.

Other metric that is used by Srinivasan [44] is the median of the Unsigned Border Position. On his analysis, Srinivasan calculates the absolute distance between the automated segmentation and the reference on each position and, besides obtaining the UE, analyses the median value of the distribution of the distances between the two segmentations.

The thickness of the retina is also a measure that can be used as a metric for comparing the segmentation. Mayer [34] used the thickness of retina calculated between the ONL and ILM. Equation 3.9 describes the calculus of the thickness.

$$RT^i = (ONL^i - ILM^i) \tag{3.9}$$

Here $ONL^i$ and $ILM^i$ are the border positions of the layers that limit the region on each position $i$. After having the distance on each position $i$, the difference between the thickness of the automated segmentation and the reference segmentation is calculated as presented on equation 3.10.

$$RTD^i = \bar{RT}^i - RT^i \tag{3.10}$$

In equation 3.10, $\bar{RT}^i$ is the thickness of the retina on the position $i$ for the reference segmentation and $RT^i$ is the thickness of the retina on the position $i$ for the automated segmentation. In the end, the mean of the absolute value, $|RTD^i|$, is calculated. This metric is titled Retinal Thickness Difference(RTD) .

Using an approach similar to Mayer, Chiu [8] used the average difference on the thickness of each layer. The author started by calculating the thickness of each layer for the automated and reference segmentation using the equation presented in equation 3.11.

$$L_1^i = (S_a^i - S_b^i) \tag{3.11}$$

In equation 3.11, $L_1^i$ is the thickness of the layer being considered on each position $i$, $S_a^i$ is the position of the surface above the layer on the position $i$ and $S_b^i$ is the position of the surface below the layer on the position $i$. Having the distance between surfaces on each position $i$, the difference between the thickness of the automated segmentation and the reference segmentation is calculated by equation 3.12.

$$LTD_1^i = \bar{L_1}^i - L_1^i \tag{3.12}$$

In equation 3.12, $LTD_1^i$ is the difference between the thickness of the layer on the reference segmentation, $\bar{L_1}^i$, and on the automated segmentation, $L_1^i$. Using this difference on each position $i$, the author calculates the mean and standard deviation. This metric is titled Layer Thickness Difference(LTD) . Fang [11] and Gao [14] used this procedure to evaluate the segmentation.

A metric also used by Tian [47] is the Ninety-Fifth Percentile Unsigned Errors($E_{95}$) which consists of extracting the highest value of the absolute distance between the reference and automated segmentation, after removing the highest 5% of the values.

## 3.3 Machine Learning

Machine Learning is a subset of Artificial Intelligence dedicated to the development of data-driven algorithms. This means that the ML algorithms can learn how to perform a certain task by looking at given data sets. These algorithms keep getting better as more data is given for them to learn. ML is widely used today on services such as recommender systems and speech recognition once it can perform much better than traditional algorithms when solving complex problems.

The learning process of an ML algorithm can be done in two distinct ways that are categorised as supervised and unsupervised learning. The algorithms that are based on supervised learning techniques rely on labelled data while the algorithms based on unsupervised learning techniques don't need the label to learn. The label is the expected output of the model, this way, the supervised learning procedure can be seen as an optimisation problem where the model fits its parameters to produce an output equal to the label. The unsupervised learning procedure doesn't rely on the label to learn, this way, it learns by grouping the data points with similar properties and find the possible outputs that way. These techniques are called clustering techniques and they allow the classification of the data points when labelling is not available. On this research project, the implemented models

are based on supervised learning techniques. This way, this analysis will focus on algorithms based on this learning procedure.

The training of an ML model consists of the process of optimising the model parameters so that the models produce an output equal to the label for each data point. The point of training, however, is for the model to produce good results when dealing with new data. This way, the model must learn how to correctly classify the data but not be too adapted to the training data set. This is important once, if the data gets too adapted to a specific data set, it loses the ability to adapt to new data. This problem is known as overfitting. On the other hand, when the model is too simple to capture the complexity of the training data set, it won't be able to correctly identify the new data eider. This situation is known as underfitting. These two concepts are very important aspects during the training procedure because if one of the two occurs it implies that the model won't be able to fit its purpose.

Overfitting happens because the model is too complex and is capturing the noise of the training data set. This way, possible ways to deal with overfitting include: restricting the model's complexity by adjusting the hyperparameters, getting more data for the model to train or reduce the noise on the training data by fixing data errors or removing the outliers. On the other hand, underfitting can happen when the model is too simple to capture the structure of the data. This way, possible ways to deal with underfitting include: reducing the restraints of the model by adjusting the hyperparameters, selecting a more powerful model or applying feature engineering techniques on the data to allow the model to find relevant information more easily.

To ensure that underfitting and overfitting don't occur, guaranteeing the capacity of the model to generalise for other data sets, it is a common practice to split the data into a training, validation and test sets. The training data set is used to train the model while the validation data set is used to evaluate the performance of the model, allowing the adjustment of the model's hyperparameters accordingly. After tuning the hyperparameters to ensure the best results on the validation data set, the test set is used as an independent data set to evaluate the final performance of the model on a new data set.

There are models that can be used for classification tasks, regression or both. Classification models deal with discreet or categorical outputs while regression models deal with continuous outputs. On this analysis, the classification ability of the models will be explored once the goal of this project is to solve a classification problem.

### 3.3.1 Decision Tree

The Decision Tree(DT) is a very intuitive decision-making process that works as presented in Figure 3.7. The process starts at the top of the tree, on the root node, and goes from node to node answering the questions according to the input value that is being classified. When there are no more nodes to answer the decision-making process faces a leaf node. The leaf nodes represent the end of the decision process once they contain the classification for the input data that is being evaluated.

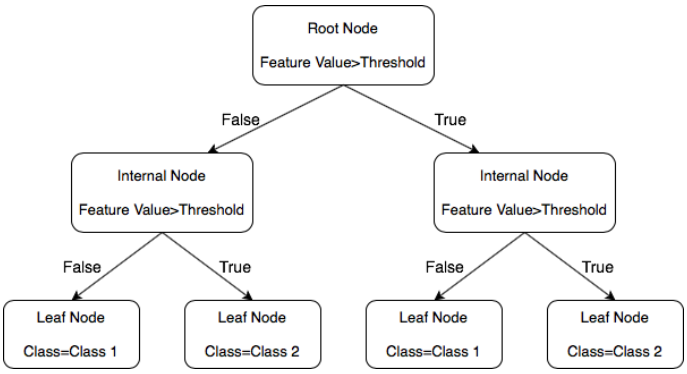


Figure 3.7: Representation of a decision tree.

The DT learning algorithm produces these trees by exploring the best decision to implement on each node so that the leaf nodes correctly identify the classification of the input data.

The process starts at the root node with all the data. The algorithm analyses each feature of the input data and looks for the threshold that better splits the data, according to the labels. When the best feature and threshold value are found, the data is split according to that rule and the process continues for the two nodes that have been created.

When all the data points on that node have the same label, a leaf node is obtained and is called a pure leaf. On the other hand, when a node contains data points with different labels it is said to be impure. This way, the goal of each splitting is to obtain two nodes that are as close to being pure as possible and the feature and threshold are selected for the nodes to according to this goal.

However, to make the selection of the best feature and threshold, it is necessary to establish a quantitative measure of the level of impunity of each node so that the algorithm makes the selection that minimises this values. One way to quantify the impurity of a node is with the Gini impurity, given

by equation 3.13.

$$G_i = 1 - \sum_{k=1}^{n} P_{i,k}^2 \tag{3.13}$$

In equation 3.13, $P$ represents the probability of classification of each class. This way, the Gini impurity is calculated for each node by subtracting to one the probability of the classification of each class squared. It is important to notice that when there is only one class on the node, the probability of that class is equal to one and the probability of the other classes is equal to zero, giving a Gini impurity equal to zero. This means that, for a pure leaf, the Gini impurity is zero.

After assessing the Gini impunity of each node, a weighted average is performed and that is the value used to compare the feature and thresholds. The pair of feature and threshold that produces the minimum weighted average of the Gini impurity is selected and the node is defined with the rule defined by that pair.

This procedure is repeated for all the nodes until all the nodes are pure or until some limitation for the growth of the tree is reached.

Another way to calculate the level of impurity of a node is by considering the entropy. This method is inspired by the thermodynamics measure of the entropy and actuates on the same way as Gini impurity. The entropy is given by the equation 3.14.

$$H_i = - \sum_{\substack{k=1 \\ P_{i,k} \neq 0}}^{n} P_{i,k} \log(P_{i,k}) \tag{3.14}$$

As for the Gini impurity, when a node is pure, the entropy of that node is equal to zero.

### 3.3.2 Random Forest

A Random Forest(RF) is an ensemble method, composed of individual DT. An ensemble method consists of considering a group of predictors that are trained and produce individual outputs. Having the outputs of each predictor, a voting system selects the output that is chosen the most as the output of the ensemble method.

The goal of an ensemble method is to have different estimators. One way to have different estimators is to consider different classification algorithms. The other way is to consider the same algorithm to generate the models but

to train them with different data. This is called bootstrap aggregating and is done by randomly selecting data from the original data set for the training set of each model. It is important to mention that bootstrap aggregating allows the repetition of data points on a certain training set.

Besides applying a bootstrap aggregating technique for selecting the training set for each model, the RF algorithm also randomly selects the features that each DT considers for the splitting. On the DT algorithm, the splitting on each node considers all the features of the data set and selects the best feature and the best threshold to generate nodes with less impurity. To ensure, once more, that the estimators are different from each other, the features that are considered for each splitting of a DT are selected randomly from the total set of features.

### 3.3.3 Artificial Neural Network

The Artificial Neural Networks(ANN) are composed by an input layer, an output layer and a group of hidden layers, as shown in Figure 3.8, and the number of hidden layers determines the depth of the ANN.

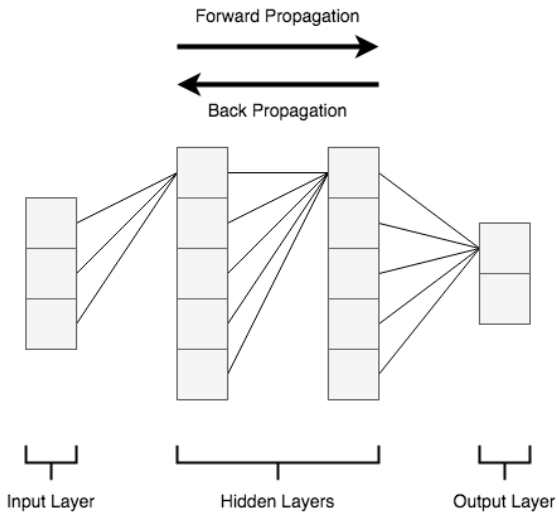


Figure 3.8: Representation of an ANN.

The hidden layers are composed by neurons, which actuate on the input data according to the diagram presented on Figure 3.9.

To obtain the value of each neuron, first, a linear combination of the

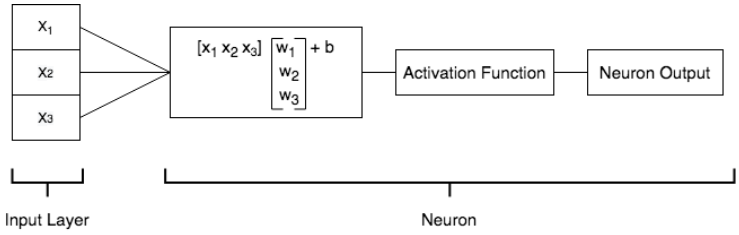


Figure 3.9: Representation of the workflow of a neuron.

input data is performed with certain weights and biases that are specific of each neuron. Then, an activation function is applied to the output of this linear combination.

After obtaining the output of each neuron of one hidden layer, other hidden layers can be added to the network by considering the neuron output values of the previous layers as input values.

#### 3.3.3.1 Activation Functions

The activation functions are non-linear, continuous and differentiable functions that are needed to allow the stack of new layers. without the activation function, the ANN could only produce linear models.

The most common activation functions are the sigmoid function(3.15), the hyperbolic tangent function(3.16), or tanh, the Rectified Linear Unit function(3.17), or ReLu, and the softmax function(3.18).

$$Sigmoid(x) = \frac{1}{1 + e^{-x}} \tag{3.15}$$

$$tanh(x) = \frac{2}{1 + e^{-2x}} - 1 \tag{3.16}$$

$$ReLu(x) = max(0, x) \tag{3.17}$$

$$Softmax = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}} \tag{3.18}$$

The sigmoid and tanh functions are directly applied by the presented equations. The ReLu function's output consists of the $x$ value, if $x$ is positive, and zero, otherwise.

The softmax function is calculated by the exponential of the linear regression of the neuron in question divided by the sum of these values for the other neurons of the layer. This activation function differs from the other

ones because it produces values that are normalised, this is, for a certain layer, the sum of the values of the neurons equals one. This is done because it considers the values of each neuron of the layer, instead of considering only the inputs of the neuron in question. The normalisation of this function is very important because it allows the representation of probabilities and this is why this function is used on the output layer for classification problems.

#### 3.3.3.2 Backpropagation

The learning process of an ANN consists of adjusting the weights and biases of the layers to minimise a loss function and this is done with backpropagation. The first step of the training consists of propagating forward and calculating the outputs with a random set of weights and biases. With the output values, a loss function is used to compare the output values to the labelled data. These loss function values are then backpropagated to update the weights and biases.

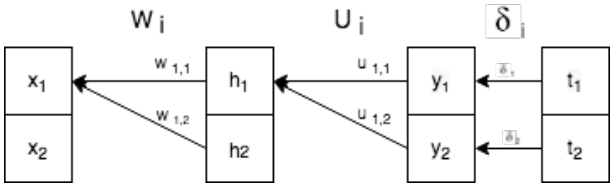


Figure 3.10: Representation of the backpropagation process used to update the weights and biases of the ANN.

Figure 3.10 shows a hypothetical network with one hidden layer with two neurons. $u_i$ and $w_i$ are the weights for the output and hidden layer, respectively, and $\delta_i$ are the loss function values for each output. To obtain the updated weights $u_{i+1}$ the equation 3.19.

$$u_{i+1} = u_i - \eta \nabla_u \delta_i \tag{3.19}$$

In equation 3.19, $\nabla$ is the derivative of the activation function of the layer hose weights are being updated and $\eta$ is the learning rate. This technique for the update of the weights is called Stochastic Gradient Descent and the main goal is to find the weights that minimise the loss function.

The learning rate defines how much the weights will change on each iteration. It is a hyperparameter of the ANN and is defined between zero and one. If the value is too small, the network can take too long to converge, however, if the value is too large, the network can miss the point where the

minimum loss values are achieved. This way, it is important to have a larger learning rate when the weight is far from the the point with minimum loss value, and a smaller learning rate for weights near that point, so that the network converges faster and doesn't miss the weight value that presents the lowest value for the loss function. One way to do this is by using Adam algorithm [25], which is an adaptive learning rate algorithm. This algorithm presents an evolution of the Stochastic Gradient Descent by adapting the learning rate according to the rhythm at which the network is converging towards the point with minimum loss value.

After updating all the weights and biases, the forward propagation is repeated, and the process continues until the minimum cost function is achieved or another pre-established limiting criterion is reached.

### 3.3.4 Metrics to evaluate the models

After training the model, it is necessary to evaluate the performance of a classification model, this is, to assess if the prediction the model is producing agrees with the true labels of each data point. To do this, there are several metrics to answer different questions about the performance of the model.

To evaluate a classification model, for each label, the predicted values can be: True Positives(TP), True Negatives(TN), False Positives(FP) or False Negatives(FN) . To explain each definition two classes, A and B, are considered. Using the evaluation from the standpoint of the class A, a TP is a data point where the predicted value is A and the true label is also equal to A. A TN is a data point where the predicted value equals to B and the true label is also equal to B. A FP is a case where the predicted value equals to A but the true label equals to B. Finally, a FN is a situation where the predicted value equals to B but the true label is A.

One common way to represent the TP, FP, TN and FN values is with a confusion matrix. This is a representation widely used to evaluate classification models once it gives information about the performance of the model on the classification o each class. A representation of a confusion matrix, for the example given above, is presented in Figure 3.11.

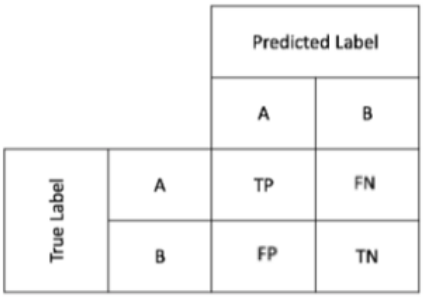


Figure 3.11: Representation of a confusion matrix. The TP, TN, FP and FN classification are given on respect to the label A.

#### 3.3.4.1 Accuracy

The accuracy metric consists of counting the total number of predicted labels that agree with the real labels and divide this value by the total number of data points.

This metric answers the question "What percentage of the total data-set is being predicted correctly?" and is calculated with equation 3.20.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{3.20}$$

It is an important metric for general purpose, but can be misleading. For example, on an imbalanced data set, where most of the data points have the same label, the accuracy score can by high, even if the model only selects the label of the most selected label. This means that this model isn't able to distinguish between the different labels but, since it has a lot of data points from the only label that is selected, the accuracy is high.

#### 3.3.4.2 Precision

The precision gives information about the percentage of the data points predicted for one class that are TP. This metric can be calculated for each label by the equation 3.21.

$$Precision = \frac{TP}{TP + FP} \tag{3.21}$$

The precision is an important metric when the goal is to reduce the number of FP.

#### 3.3.4.3 Recall

The recall informs about the fraction of labels that were correctly predicted, this is, the number of TP among all the points whose true label belongs to a certain class. This metric can be calculated for each label by the equation 3.22.

$$Recall = \frac{TP}{TP + FN} \tag{3.22}$$

This metric is especially important when the goal is to reduce the number of FN.

#### 3.3.4.4 F1-score

Finally, the $F_1$ Score is the harmonic mean of precision and recall. This metric is calculated with equation 3.23.

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3.23}$$

It is used when the goal is to find a balance between the number of FP and FN.

# Chapter 4

# Methods and materials

## 4.1 Introduction

This project aims to develop a strategy for the selection of the best algorithm for the segmentation of an OCT volume. The strategy will take as input the crossed metrics, which are the metrics calculated between the segmentation of each algorithm. Using these metrics, the strategy selects the best algorithm for the segmentation.

A representation of the proposed procedure for the segmentation, using the selection of the best algorithm, is presented in Figure 4.1, with a representative example of using three algorithms.

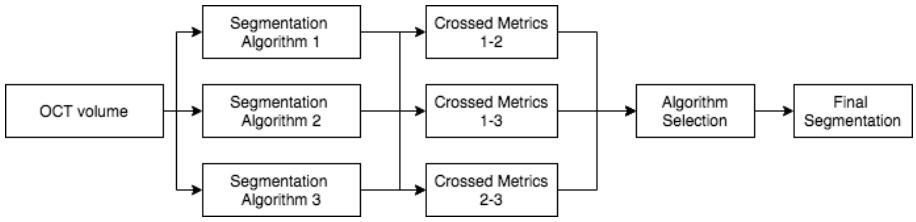


Figure 4.1: Representation of the proposed approach on OCT segmentation.

Since this is a new idea on the field of OCT segmentation, it is important first to evaluate the benefit that this approach can bring to the field and then show how it can be implemented. This way, this research project is divided into three parts, presented in the project outline in Figure 4.2.

The first part of the project aims to show the benefit of selecting the best algorithm for the segmentation, instead of using only one algorithm for all situations. This part is composed of two separate sections. Firstly,

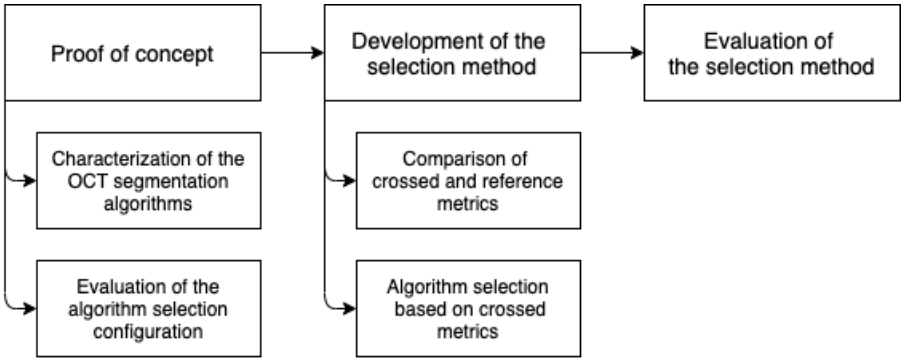


Figure 4.2: Representation of the project outline.

the algorithms to be considered on this project are evaluated for a better understanding of their response. After this, the selection configurations are evaluated and compared to the use of individual algorithms to assess the benefit of using the algorithm selection. The selection configuration consists of the way the selection will be implemented, which can be done by selecting the best algorithm for the total volume or for specific regions and can also include the selection of the best algorithm for the segmentation of individual layers.

The second part of the project consists of the development of the selection strategy. The first section of this part consists of evaluating the use of the cross metrics. This way, in this section, crossed metrics are compared to the reference metrics to assess the information that can be extracted. After this, the ML models for the development of the selection criteria are applied, with different configurations of metrics as input, to evaluate the best method to implement the selection.

The third part of the project aims to evaluate the results produced by the selection strategy that was developed. This way, on this part of the project the developed method is applied to the segmentation of a test data set to evaluate the benefits it can bring when compared with the implementation of the individual algorithms.

On this chapter, the OCT segmentation algorithms considered in this study are presented, as well as the data set that is considered for all the steps of the project. After this, the methodology used to evaluate the points presented above is explained.

All the processing of the segmentation, metrics extraction and OCT data visualization presented in this chapter was performed using MATLAB

(version $R2019b$), while the data processing of the metrics and the model implementation was done using the Python programming language (version 3.8.5) on a Jupyter Notebook environment (version 6.0.3).

## 4.2 OCT segmentation algorithms

On this research project, four segmentation algorithms are considered. These algorithms are based on the graph theory approach for segmentation, introduced in section 3.3, and were selected because they produce good overall results and because they were already implemented and available to be used in this research project.

Three of the algorithms consider the 3D graph theory approach, presented by Garvin, that is explained in section 3.3. One of these algorithms was implemented by AIBILI (Association for Innovation and Biomedical research on Light and Image) and will be referenced from now on as AIBILI algorithms. The second algorithm was implemented by Carl Zeiss Meditech and will be referenced from now on as the Zeiss algorithm. The third algorithm is the IOWA Reference Algorithm OCT-Explorer(version 3.8, Retinal Image Analysis Lab, Iowa Institute for Biomedical Imaging, Iowa City) that will be referenced from now on as Explorer algorithm.

The fourth algorithm consists of the 2D graph theory approach proposed by Chiu [8] that is also presented in section 3.3. This algorithm was implemented by the CASEREL project [46] and will be referenced from now on as the 2D algorithm.

The four algorithms can segment different layers. Once the goal is to compare the algorithms, only the seven surfaces that were common to all the four algorithms were considered. These surfaces are: ILM, NFL-GCL, IPL-INL, INL-OPL, OPL-ONL, IS-OS and RPE-CH. Most of these interfaces have a direct connection between the layers presented on 2.1.2. One of the exceptions is the IS-OS that represents the junction of the inner and outer segments of the photoreceptor, located on the photoreceptor layer. The other layer that doesn't have a direct interpretation is the the RPE-CH that represents the surface that separates the RPE layer from the choroid.

## 4.3 Dataset

The data set considered for this project was extracted from the Progress (NCT03010397) study, developed by AIBILI. This study contains OCT-A scans collected over a period of five years that record the progression of patients with type 2 diabetes mellitus.

For this project, a total of 123 OCT-A scans, from the fifth visit, were considered, once they were previously manually segmented by an expert. From the 123 patients, 38 were female and 85 were male and the characterisation of the patients is presented in Tables 4.1 and 4.2. In Table 4.1, the age and the HbA1c are presented, being the HbA1c a measure of the long term control of diabetes. In Table 4.2 the number of patients in each level of the EDTRS grading system is presented. This grading system is used to assess the severity of the DR pathology.

Table 4.1: Characterisation of the patients considered for the study.

| | Mean ± Standard Deviation |
|---|---|
| Age (years) | 64.0 ± 6.7 |
| HbA1c at v1 | 8.30 ± 1.3 |

Table 4.2: Characterisation of EDTRS level of the patients at the first visit of the study.

| EDTRS level | Number of patients |
|---|---|
| 10 | 3 |
| 20 | 37 |
| 35 | 83 |

The OCT-A volumes were acquired with the Zeiss Angioplex (Carl Zeiss Meditec, Dublin, CA) system. The 6x6mm scan protocol was used for the acquisition which considers 350 A-scans per B-scan and a total of 350 B-scans.

**Semi-automatic segmentation**

All the volumes contain the correspondent reference segmentation. To obtain this reference segmentation, the AIBILI segmentation algorithm is applied and the resulting segmentation is then corrected by an expert on the points where the segmentation doesn't fit the image.

## 4.4 Proof of concept

### 4.4.1 Characterisation of the OCT segmentation algorithms

The main goal of this section is to characterise the output of each algorithm and evaluate the relation between the four algorithms. This section also aims to verify the assumption that when different algorithms fail, on certain circumstances, at least one of the others can provide a proper segmentation.

#### 4.4.1.1 OCT segmentation and metrics extraction

The first step consists of applying the segmentation algorithms to the OCT volumes on the data set and implementing the reference metrics.

The Zeiss and AIBILI algorithm were directly applied. However, the segmentations produced by the 2D and Explorer algorithms required some processing.

Due to the nature of the 2D algorithm, a median filter was applied to smooth the discontinuities between B-scans. The median filter is applied to the $350 \times 350$ matrix that represents each surface. Each row of the matrix represents the position of the interface on a one A-scan and each column represents one B-scan. This way, each column will contain the position of the surface on each A-scan of one B-scan. The dimensions of the median filter were defined to minimise the MUE value and the best results were obtained with a filter with dimensions $1 \times 9$. This is, the surface position for each A-scan is defined by the median over that individual A-scan positions along the nine subsequent B-scans, as presented in Figure 4.3. The units considered for the matrix and vector representation are voxels.

The implementation of the Explorer algorithm is affected by an intrinsic bias between the reference segmentation and the segmentation produced by the algorithm. This bias occurs because the convenient location of the layers, for the Explorer algorithm, slightly deviates from the ones of the reference segmentation. This leads to a consistent dislocation of each layer in one certain direction. To correct this bias, a pixel shift is applied to each segmented layer of the Explorer algorithm [11]. This shift implies that a constant value is summed to each layer vector, to compensate for the constant bias. To determine the best values to apply to the shift on each layer, the mean of the MSE for the data set was calculated and the values obtained were adopted for the shift. The resultant shifts to be applied are presented in Table 4.3.

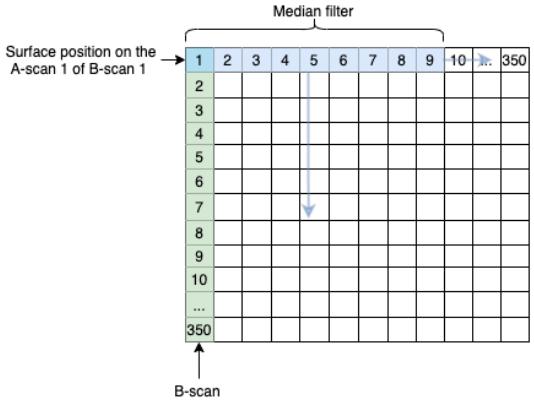


Figure 4.3: Scheme of the median filter application on each surface matrix.

Table 4.3: Pixel shift applied to the segmentation of the Explorer algorithm.

| Layer | Shift |
|---|---|
| ILM | +2 |
| NFL-GCL | +3 |
| IPL-INL | +3 |
| INL-OPL | +3 |
| OPL-ONL | +3 |
| IS-OS | -2 |
| RPE-CH | +3 |

After completing this processing step, the metrics presented on section 3.2.2 were applied for each layer, according to the distribution presented on Table 4.4. This way, each segmentation is characterised by a total of 35 metrics.

Table 4.4: Overview of the extracted metrics and the layers involved on the calculus of each metric.

| Metrics | Layers Involved | Metrics | Layers Involved | Metrics | Layers Involved |
|---|---|---|---|---|---|
| MUE | ILM<br>NFL-GCL<br>IPL-INL<br>INL-OPL<br>OPL-ONL<br>IS-OS<br>RPE-CH | MSE | ILM<br>NFL-GCL<br>IPL-INL<br>INL-OPL<br>OPL-ONL<br>IS-OS<br>RPE-CH | Median | ILM<br>NFL-GCL<br>IPL-INL<br>INL-OPL<br>OPL-ONL<br>IS-OS<br>RPE-CH |
| $E_{95}$ | ILM<br>NFL-GCL<br>IPL-INL<br>INL-OPL<br>OPL-ONL<br>IS-OS<br>RPE-CH | LTD | ILM,NFL-GCL<br>NFL-GCL,IPL-INL<br>IPL-INL, INL-OPL<br>INL-OPL, OPL-ONL<br>OPL-ONL, IS-OS<br>IS-OS, RPW-CH | RTD | ILM, OPL-ONL |

#### 4.4.1.2 Characterisation of the OCT algorithms

To characterise the algorithm's response, the volumes with the five best and worst segmentation were selected for each algorithm to be individually analysed. To perform this selection, the average of the MUE over all the layers was computed and the five volumes that presented segmentations with the lowest MUE were selected as the best ones for each algorithm. For the selection of the five worst segmentation, the same process was done but this time the five volumes with the highest average MUE were selected. It is important to notice that the MUE was the only metric considered once it is representative of the performance of the algorithm on the segmentation of the volume.

After the selection, the analysis of the segmentation consists of evaluating the plot of the segmentation of each algorithm against the reference segmentation and comparing the metrics of the four algorithms that are being considered.

### 4.4.2 Evaluation of the algorithm selection configuration

This section has two main goals. The first is to verify that the use of the best segmentation algorithm for each situation leads to better segmentation results than implementing one individual algorithm for all situations. The second goal is to evaluate which selection configuration can produce the best results.

The first configuration considered consists of selecting the best algorithm for the OCT volume and is represented in Figure 4.4 a). This is the simplest implementation and can avoid the big deviation that happens on certain volumes for each algorithm. However, even the best segmentation algorithm for one volume can have deviations from the reference segmentation on certain regions that could be better identified by one of the other algorithms.

The second configuration consists of dividing the volume into three regions, as shown in Figure 4.4 b), and selecting the best algorithm for each region. The middle region contains the fovea and the other two represent the remaining volume. In the fovea region, the layers of the retina present a different dynamic from the rest of the volume. Then, it is possible that an algorithm, that performs well on the rest of the volume, fails in that region and this is why it is interesting to study if this selection will have an impact.

Finally, the third configuration consists of choosing the best algorithm for each individual B-scan, as represented in Figure 4.4 c). Although it is more complex once involves the extraction of a higher amount of data,

this configuration aims at avoiding small local deviation. The selection of the best algorithm for each individual B-scan is not expected to show a big difference when compared to the implementation of the selection for the volume but can avoid some small deviation on the specific local deviation. For example, when there is a lesion that is not detected by the best algorithm for the volume, but is detected by one of the others, this selection allows the correction of that specific region.

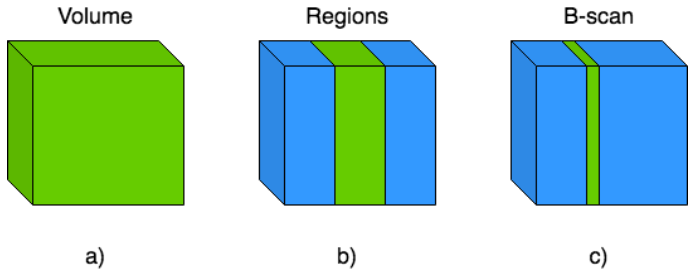


Figure 4.4: Representation of the three selection configurations. a) Selection of the segmentation algorithm for each volume. b) Selection of the segmentation algorithm for each region. c) Selection of the segmentation algorithm for each B-scan.

Besides the presented configurations, the benefit of choosing the best algorithm for the segmentation of each layer will also be evaluated. For the first configuration, this is done by selecting the best algorithm for each layer of the volume. For the second configuration, this is done by selecting the best algorithm for each layer of each region. The same is applied for the third configuration, for each B-scan.

For the first configuration, the selection of the algorithms is done by considering the mean over all the reference metrics and selecting the algorithm which presents the segmentation with the lowest mean value. The same is done for the second configuration but, in this situation, the metrics are calculated over the segmentation corresponding to each region. For the third configuration, the selection is also done by considering the mean over all metrics but, in this case, the metrics are calculated for the segmentation of each B-scan.

For the selection of each layer, the same principle is applied as for the previous selection configurations. The only difference is that the mean is performed over the metrics related to each layer, instead of all the metrics. The assignment of the MUE, MSE, median, and $E_{95}$ metrics for each layer

is direct, once they are directly extracted from each layer. Besides this, the RTD metric is used for all the layers, once it reflects all the retina. Finally, since the LTD metric involves the use of two subsequent surfaces, once it is related to the thickness of each layer, the metrics considered for each layer are the ones that involve that layer on the calculus, this is, the one above and below the layer in question. Once the two surfaces on the external borders only have one value, the value was considered twice, so that the dimension of the vectors of each surface are equal. The distribution of the LTD metrics is presented on Table 4.5.

Table 4.5: LTD Metrics considered for the selection of each layer. The layers involved on the calculus of the LTD are identified between brackets.

| Layer | LTD Metrics |
|---|---|
| ILM | LTD(ILM,NFL-GCL) |
| NFL-GCL | LTD(ILM,NFL-GCL) LTD(NFL-GCL,IPL-INL) |
| IPL-INL | LTD(NFL-GCL,IPL-INL) LTD(IPL-INL,INL-OPL) |
| INL-OPL | LTD(IPL-INL,INL-OPL) LTD(INL-OPL,OPL-ONL) |
| OPL-ONL | LTD(INL-OPL,OPL-ONL) LTD(OPL-ONL,IS-OS) |
| IS-OS | LTD(IOPL-ONL,IS-OS) LTD(IS-OS,RPE-CH) |
| RPE-CH | LTD(IS-OS,RPE-CH) |

To evaluate the results of each configuration, after the selection of the best algorithms for each configuration, the segmentation of the volumes are generated, considering the algorithms selected for each configuration. Having the segmentation for each configuration, the metrics are calculated and the mean of each metric over all volumes is measured. To compare the results of the selection configurations and the individual algorithms, the mean results are plotted.

The plots of the mean metrics are then analysed to see if there is a decrease of the metrics mean values with the algorithms selection and if there is a decrease as the regions for the selection get more specific. This is, if the metrics mean gets smaller as the selection goes from the selection of the best algorithm for the volume to the selection of the best algorithm for each B-scan.

## 4.5 Development of the selection method

### 4.5.1 Comparison of crossed and reference metrics

The aim of this section is to explore the relationship between the crossed metrics and the reference metrics. This step is important to better understand the behaviour of the crossed metrics and how they can be used for the selection of the best algorithm.

As it was said above, the reference metrics are the ones calculated between the segmentation produced by an algorithm and the reference segmentation while the crossed metrics are the metrics calculated between the segmentation produced by two different algorithms, as shown in Figure 4.5. For this reason, the reference metrics are the metrics used to evaluate each segmentation and generate the label data, while the crossed metrics are the metrics that will be used for the algorithm selection. This way, it is important to see how the two variables are related to better understand the models that will be implemented.

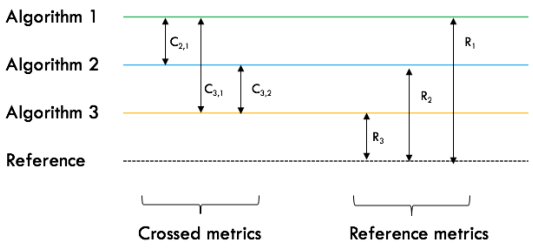


Figure 4.5: Schematic representation of the reference and crossed metrics.

To see how the information on the reference metrics is reproduced by the crossed metrics, it is convenient to compare the same variables. This is done by considering one of the algorithms as the reference and comparing the crossed metrics between the segmentation of one of the reference algorithm and the other two. To reproduce this quantity with the reference metrics, the reference metrics of the segmentation of the reference algorithm are subtracted to the reference metrics of the segmentation of the other two algorithms, as shown on Figure 4.6.

Having the crossed metrics and the reference metrics for each algorithm, with the procedure indicated above, the variables indicated above were plotted and evaluated for each volume of the five best and worst, selected at

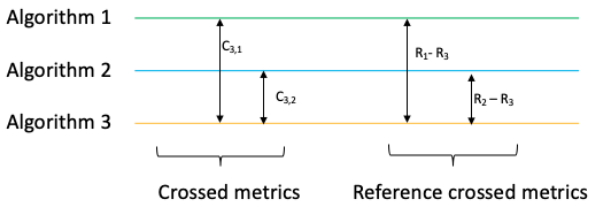


Figure 4.6: Schematic representation of the reference and crossed metrics comparison.

Section 4.4.1.

It is important to notice that the crossed metrics won't be able to reproduce the value of the reference metrics. The crossed metrics only reflect the agreement between the segmentation of each algorithm. This way, they won't be able to identify which algorithm has the lowest reference metric. The goal is to use the pattern of the crossed metrics to identify the one that presents the best metrics and not to reproduce the reference metrics with the crossed metrics.

### 4.5.2 Algorithm selection based on the crossed metrics

The goal of this project is to develop a method to select the best algorithm to segment a certain OCT volume and the possible configurations for this selection were presented above.

On this section, the methods for the selection of the best algorithm are evaluated. Firstly, the three data configurations that are considered are presented with its advantages and disadvantages. Then, the implementation of the ML models is explained.

#### 4.5.2.1 Data configuration

Each segmentation is characterised by 35 reference metrics that were identified in section 4.4.1.1. When comparing three algorithms, a total of 105 crossed metrics will be available for each volume. This is due to the fact that the three algorithms form three pairs between each other and there are 35 crossed metrics extracted from each pair.

This information can be arranged in different configurations and on this project three data configurations are considered. These three data configurations are presented on Figure 4.7.

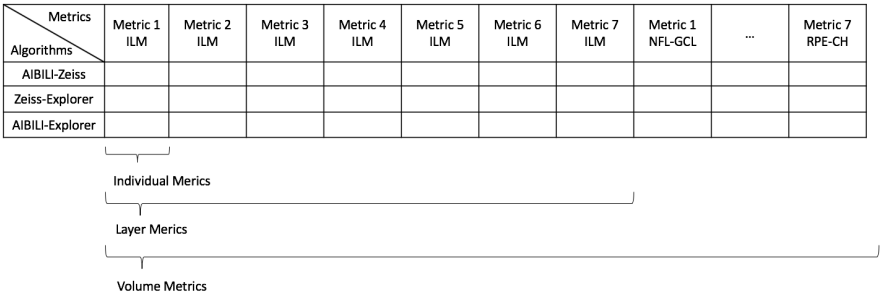


Figure 4.7: Representation of the data configurations that were considered.The seven metrics presented are the metrics associated with each layer, specified on section 4.4.2

The first data configuration consists of assuming that each metric has the same behaviour and that it is possible to the selection of the best algorithm for each metric independently. This is, considering as input the 3 crossed metrics available for each metric that characterises the volume. In this situation, the label for each metric is the algorithm that presents the lower reference metric on that specific metric. The selection of the best algorithm for the segmentation of each volume/region/B-scan can then be done by considering a voting system where the algorithm that is selected the most, out of the 35 metrics, is the selected algorithms for the segmentation of that volume/region/B-scan. This can also be done for the selection of the best algorithm for each layer, on any configuration, by considering the same voting system for the layer metrics presented in section 4.4.2.

The second data configuration consists of using the layer metrics presented at section 4.4.2 as input. The labels, in this situation, are the algorithms that present the lower mean of the reference metrics considered on each layer, as was considered for the layer selection on section 4.4.2. This data configuration can also be used to select the best algorithm for the selection of each volume by considering a voting system. This voting system selects the algorithm that is selected the most, for the seven layers, as the chosen algorithm to segment the volume.

The third data configuration consists of considering all the available crossed metrics for one volume/region/B-scan as input. The label, on this situation, is the algorithm that presents the lower mean over all the reference metrics, as was presented on section 4.4.2. This data configuration doesn't allow the selection of the best algorithm for the segmentation of each layer, only the selection of the best algorithm for the volume/region/B-scan.

A summary of the use of the different data configurations is presented on Figure 4.8.

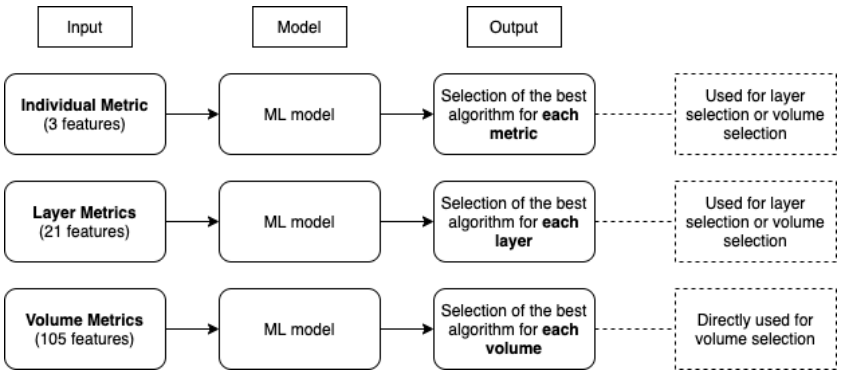


Figure 4.8: Representation of the model input and output for the different data configuration. Here is presented the outline for the volume selection but it is the same for the region and B-scan selection.

#### 4.5.2.2 Selection criteria implementation

In this section, the ML model's implementation is explained. The models that are considered on this research project are the machine learning classifiers presented in section 3.3, this is, the DT, RF and ANN. These models were implemented for each selection configuration (volume, regions and B-scan) and for each data configuration (individual metrics, layer metrics and volume metrics) to evaluate which presents the best solution for the selection problem.

The first step of the analysis consists of defining the test set. The selection of the volumes for the test set consists of applying a stratified split to separate 20% of the volumes according to the volume classification. The stratified split ensures that there is an equal percentage of each classification on both the training as test sets. To apply the stratified split, the $StratifiedShuffleSplit$ function of the $model\_selection$ module from the Python's $Scikitslearn$ library was applied. It is important to notice that the volumes on the test set were the same during all the procedure.

### Pre-processing

Due to the way the reference segmentation is produced, correcting the segmentation produced by the AIBILI algorithm, there is a bias towards the selection of the AIBILI algorithm. For this reason, the data set is imbalanced, this is, the number of data points where the true label is AIBILI is much higher than the number of data points assigned to the other labels.

Imbalanced data sets are a problem for training models once the models learn to always select the most frequent label instead of learning how to differentiate between the different ones. This leads to a high accuracy value, due to the high prevalence of the most frequent label, but the model isn't able to identify when not to choose that class. One approach to avoid this situation is to sample the data set. Sampling the data set consists of transforming the imbalanced data set into a balanced data set. This can be done by undersampling or oversampling. Undersampling consists of reducing the number of data points with the most frequent label while oversampling consists of increasing the number of data points with the labels that are less represented.

One way to implement the undersampling approach is to consider the distribution of the data points on the different classes and randomly delete data points from the dominant classes until the dimensions are equal, keeping the data distribution. For the oversampling it is possible to apply a similar approach but by adding copies of the instances of under-represented classes, keeping the initial distribution, until the dimensions are equal.

To avoid losing data, the oversampling approach was implemented. This was done by using the *resamle* function from the *utils* module of the *Scikitslearn* library. This function applies the oversampling approach described above. This way, the data points with Zeiss and Explorer label was expanded to match the dimension of the data with the AIBILI label.

The metrics have different ranges, for example, the $E_95$ tends to be always higher than the other metrics. This factor doesn't affect the DT and RF models but affects the ANN, that can give more importance to the features with higher values. To avoid this, the features are normalised. This normalisation is applied to each feature and the result is that the values appear on a scale from zero to one. This way, all the features have the same scale. To apply this normalisation the *normalize* function from *utils* module of *keras* library from the $TensorFlow$ platform.

### Cross-validation and parameter tuning

The models have several hyperparameters that need to be tuned to ensure that the maximum potential of the model is reached. To do this, it is necessary to evaluate for underfitting or overfitting of the model. This is why the validation set is needed.

Ideally, there is a train, validation and test data set where the train data set is used to train the model, the validation is used to evaluate the model and tune the hyperparameters and the test data set is used in the end, to check the performance of the model on an independent data set. However, when few data is available, the cross-validation approach is widely used to evaluate the model.

One cross-validation technique is the K-fold cross-validation. This technique consists of splitting the train and validation data into k subsets. Then, k-1 subsets are used for training the model and the remaining is used for validation. After this, the validation fold changes and the process is repeated until all the subsets are used as validation.

Since there is not much data available, especially for the volume and region selection configuration, the k-fold cross-validation was used on this project for the evaluation of each model and selection of the hyperparameters. The k-fold cross-validation was implemented with 3 fold. This way, the data is split into 3 subsets using $KFold$ function from $model_selection$ module of the *Scikitslearn* library. The cross-validation then occurs in three iterations, where the validation data is changing between the three available subsets and the model is trained on the remaining subsets.

Inside each fold, the training data is pre-processed with the treatment described above and then the model is trained. After the training is completed, the model is used to predict the training and validation data set. With the true label and predicted label of the training and validation, the precision, recall and $F_1$ Score are calculated, for each class, with the $precision\_recall\_fscore\_support$ function from the *metrics* module of the *Scikitslearn* library. In the end, the true and predicted labels are registered as well as the metrics.

After all the folds are completed, the true and predicted labels are used to generate a confusion matrix and the mean and standard deviation of the metrics are calculated. The confusion matrix is generated with the $confusion\_matrix$ function from the *metrics* module of the *Scikitslearn* library.

To select the best hyperparameters for each model, the cross-validation was applied for a set of hypothetical hyperparameters and, for each hyper-

parameter, the metrics related to each classification were extracted. These metrics were then plotted for each value of the hyperparameter and these plots were then used to select the hyperparameter that produced the best overall results for the three classes.

#### ML Models

The implementation of the DT model was done with the $DecisionTree\ Classifier$ from *tree* module of *Scikitslearn* library. The hyperparameter that was evaluated during the training was the maximum depth of the tree that is defined as $max\_depth$ on the $DecisionTreeClassifier$ function.

The implementation of the RF was done with $RandomForest\ Classifier$ function from *endemble* module of the *Scikitslearn* library. The hyperparameters that were evaluated during the training were the maximum depth of the trees, defined as $max_depth$, and the number of estimators to consider, defined as $n\_estimators$, on the $RandomForestClassifier$ function.

The implementation of the ANN was done with the *keras* module of the $TensorFlow$ platform. The model was defined with two hidden layers. The hyperparameters that were evaluated were the activation function of the hidden layers and the number of neurons on each hidden layer. The method for the update of the weights was the Adam optimiser.

#### Training and testing

After finding the best hyperparameters for each model, the models are trained on the full train and validation data set and the test set is used to evaluate the model on an independent data set.

## 4.6 Evaluation of the developed selection method

This section aims to evaluate the selection criteria developed on this project. After the evaluation of all the models, presented in the previous section, the model that presents the best results, in terms of the classification metrics, is selected as the best one.

The developed selection criteria is used to generate the segmentation of the volumes of the test data set. The reference metrics are then extracted to allow the comparison of the obtained segmentations with the segmentation produced by the individual algorithms. This way, while on the previous section the models were evaluated in terms of the classification metrics, on this section the selection criteria is evaluated in terms of the segmentation

results. This is done to see if the developed strategy can bring benefit to the segmentation or if there is a need for other approaches to implement the selection.

To implement the selection criteria, first, the selection of the algorithms is performed on the volumes of the test data set. After this, the selection is used to generate the segmentation. This includes analysing the selected algorithm on each situation and generating the segmentation with the segmentation of the correspondent algorithm.

The segmentations obtained, with the selection, are then evaluated with the reference metrics and the results are compared to the results of the independent algorithms.

# Chapter 5

# Results and discussion

## 5.1 Proof of concept

### 5.1.1 Characterisation of the OCT segmentation algorithms

The MUE results of the segmentation produced by each algorithm can be consulted in Appendix A. Given these results, the volumes selected for the manual analysis are presented in Table 5.1. Representations of the obtaied segmentations for the volumes presented in Table 5.1 can also be consulted in Appendix B.

Table 5.1: ID of the volumes with the best and worst segmentation results for each segmentation algorithm.

| Algorithm | Class | Files ID |
|---|---|---|
| 2D | best | 67, 57, 137, 54, 3 |
| | worst | 82, 90, 80, 159, 95 |
| AIBILI | best | 89, 127, 137, 67, 3 |
| | worst | 82, 17, 121, 90, 91 |
| Zeiss | best | 69, 105, 170, 36, 74 |
| | worst | 10, 31, 63, 154, 1 |
| Explorer | best | 3, 99, 85, 36, 34 |
| | worst | 82, 83, 90, 37, 91 |

### 2D

The volumes selected as the best ones for the 2D algorithm present smooth surfaces, without big structural changes, as it was expected for all the algorithms. On these volumes, the surfaces identified by the algorithm are close to the reference ones on most of the B-scans, however, there are several B-scans that present results far from the reference, as shown on Figure 5.1. On the segmentations with the best results, these changes are mostly seen on the interior layers and not so much on the exterior ones and the metrics confirm this result.

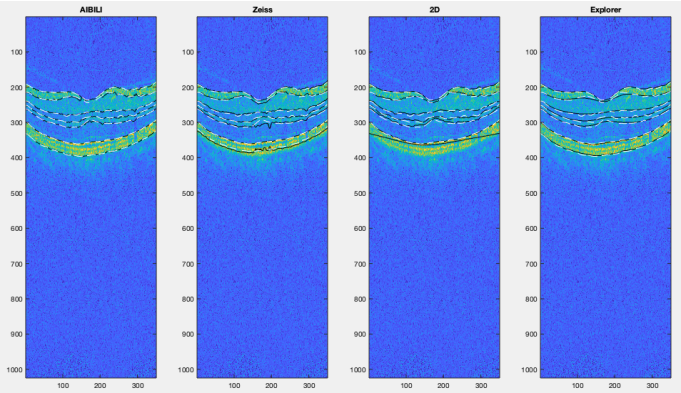


Figure 5.1: Representation of the segmentation of each algorithm for volume 137. The dashed white line represents the reference segmentation.

The individual changes on B-scans are expected on this algorithm because it performs the segmentation on each B-scan individually, unlike the other three algorithms that perform the segmentation on the full volume. To compensate for this expected behaviour, a median filter was implemented, as explained in section 4.4.1.1. However, even with the filter, some of the changes are still present, especially on the internal layers, on the RPE-CH layer and on areas with irregularities, such as the fovea.

In the worst volumes, there are few B-scans that present an agreement with the reference segmentation. In these volumes, the segmented layers appear most of the times out of the range of the retina structures, as shown in Figure 5.2.

In terms of metrics, the 2D algorithm shows generally big distances from the other three, even for the volumes where the performance of the algorithm was good. On the worst volumes, the distances from the metrics of the other three algorithms are very significant for almost all the metrics. This

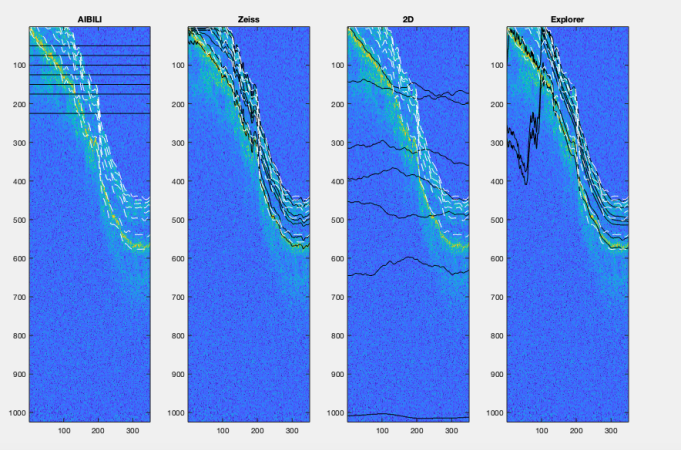


Figure 5.2: Representation of the segmentation of each algorithm for volume 82. The dashed white line represents the reference segmentation.

is important because it shows that it might be possible to identify this significant deviation, based on the crossed metrics.

It's important to point out that, even on the best volumes, the 2D algorithm doesn't produce the best results, when compared with the other three algorithms.

### AIBILI

Although the segmentations produced by the Zeiss and AIBILI algorithms are similar, on the volumes selected as the best ones for the AIBILI algorithm, This one is the one that presents the best results in terms of the metrics.

In one volume that presented a big slope, presented in Figure 5.2, the algorithm fails on the segmentation. This is associated with an error on the algorithm initialisation, due to the slope of the layers on the first B-scans, and not on the algorithm itself. The other volumes where the algorithm seems to fail are associated with the suspension of the segmentation of three internal layers on certain regions of the image or with regions where the segmentation shows spikes.

On most of the segmentations that were analysed, the AIBILI algorithm is the one that shows the best results, although the Zeiss and Explorer algorithms also present segmentations that are close to the reference one. This fact is related to the way the reference segmentation is obtained. As ex-

plained in section 4.3, the reference segmentation is performed by an expert that uses the AIBILI segmentation and corrects it. This way, the regions where the algorithm is correct are perfectly aligned with the reference segmentation, while the other algorithms may appear slightly deviated. This semi-automatic approach introduces a bias that has to be taken into consideration on the analysis once it explains why the metrics of the AIBILI algorithm are almost always better than the others, even when the visual analysis shows that three have similar performances.

The algorithm seems to deviate more from the reference segmentation near the fovea. On Figure 5.4 it is possible to see an example of these deviations. On this example, it is also possible to see how the algorithm produces some spikes on certain regions.

### Zeiss

For the volumes selected as the best ones for the Zeiss algorithm, the metrics results of the AIBILI algorithm are better. This factor can be associated with the bias for reference segmentation presented above. It is also important to state that these volumes show mostly regular surfaces, as it would be expected. That way, each of the algorithms performs well, under those circumstances.

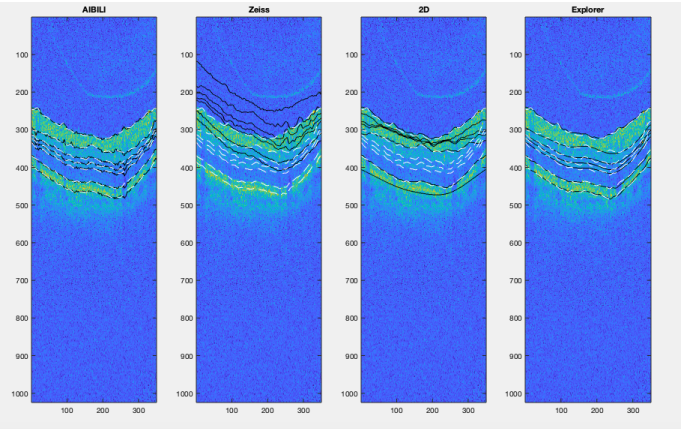


Figure 5.3: Representation of the segmentation of each algorithm for volume 1. The dashed white line represents the reference segmentation.

For the volumes selected as the worst ones for the Zeiss algorithm, the segmentation appears dislocated from the range of the retina structures, as shown in Figure 5.3. This is an error on the segmentation exportation and

not on the segmentation itself. It is possible to see that by examining metrics like the difference of the thickness of the retina and the difference of the thickness of each layer, that consider the difference between the thickness of the layers, and not metrics like the MUE, that consider the distance between the layers and the reference segmentation. On these metrics, the values obtained for the Zeiss segmentation are on the range of the ones obtained for the other segmentations. This implies that the layers are simply dislocated and not wrongly segmented.

### Explorer

The Explorer algorithm produces very smooth results, sometimes missing small structures, as shown in Figure 5.3 and Figure 5.4.

The metrics for the Explorer algorithm don't fluctuate a lot for all the segmented volumes, which is indicative that it is a consistent algorithm. Unlike the other algorithms, that have big changes on each metrics from the five best results and the five worst, the Explorer algorithm only presents significantly higher metrics for one volume, presented on Figure 5.2. This is a problematic volume, that also presents the worst results for the AIBILI and 2D algorithm. This way, although the algorithm misses some small structures that are detected by the AIBILI and Zeiss, it provides very stable results.

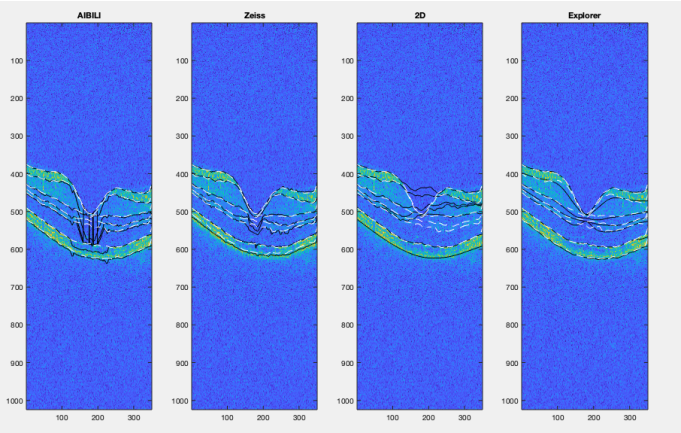


Figure 5.4: Representation of the segmentation of each algorithm for volume 69. The dashed white line represents the reference segmentation.

**Final remarks**

The algorithms show a different aptitude for identifying different layers. In other words, there are algorithms that are more efficient at identifying certain layers than others. For example, the AIBILI algorithm is the one that shows the best results in terms of the metrics, however, in several volumes, that is not the case for the IS-OS layer.

The three algorithms that are based on the 3D graph theory are the ones that produce the best results. However, these three present different behaviours. The AIBILI algorithm seems to spike more easily, while the Explorer algorithm gives very smooth results. The Zeiss algorithm performance is closer to the AIBILI algorithm but varies a bit less. These characteristics are associated with different smoothness constraints of each algorithm and are a property that can indicate that the algorithms are suited for different types of volumes.

Once the 2D algorithm isn't able to produce consistent results, it was excluded from the analysis from this section on.

The metrics analysis allowed the conclusion that, when one of the algorithms fails, the distance between the metrics of that algorithm and the other one's increases. This way, if the metrics from two algorithms are close and there is one that if significantly further, then it is possible to infer that that algorithm failed on the segmentation.

It is also important to notice that in all the analysed volumes, every time one of the algorithms failed, at least one of the others was able to produce a proper segmentation. This suggests that the algorithm selection is a viable solution for the lack of ability of the algorithms to adapt to all situations. Besides this, even the algorithm that shows the best results (AIBILI algorithm) fails on the segmentation of some volumes, which suggests that the algorithm selection has the potential to produce better results than the implementation of the individual algorithms.

### 5.1.2 Evaluation of the algorithm selection configuration

Figure 5.5 presents the comparison of the mean of each metric for the different selection configurations and each individual algorithm. It is possible to see that the mean of the metrics is lower for the algorithm's selection configuration than for the individual algorithms.

The Explorer algorithm is the one that presents the lower metrics mean value, being closer to the results presented by the algorithm's selection. This can be associated with the fact that this algorithm doesn't fail completely

on any volume, as said above, while the other two fail on some volumes. However, even for this algorithm that doesn't fail, it is preferable to have the selection.

In Figure 5.6, the selection configurations are compared. There isn't a significant difference between each model, however, it is possible to see that the models with layer selection are below the models that don't consider the selection of each layer.

For better interpretation of the relation between each selection configuration, the MUE mean and standard deviation of each layer is presented for each selection configuration and individual algorithms on Tables 5.2 and 5.3.

Table 5.2: Results of the mean and standard deviation of the MUE of each layer for the segmentation considering each selection configuration(part 1).

| | ILM | NFL_GCL | IPL_INL | INL_OPL |
|---|---|---|---|---|
| AIBILI | 3.97±32.32 | 5.18±32.32 | 4.68±32.13 | 4.76±32.02 |
| Zeiss | 7.59±24.22 | 8.77±23.89 | 8.91±23.76 | 8.14±24.02 |
| Explorer | 1.66±1.55 | 2.72±2.08 | 2.62±2.65 | 2.56±2.23 |
| Volume | 1.13±0.51 | 2.24±0.83 | 1.82±1.44 | 1.85±1.19 |
| Region | 1.15±0.53 | 2.20±0.82 | 1.79±1.48 | 1.81±1.19 |
| B-scan | 1.15±0.48 | 2.22±0.83 | 1.77±1.48 | 1.78±1.16 |
| Volume + Layer | 1.10±0.51 | 2.17±0.77 | 1.75±1.41 | 1.76±1.11 |
| Fovea + Layer | 1.09±0.49 | 2.06±0.75 | 1.70±1.37 | 1.72±1.10 |
| B-scan + Layer | 1.08±0.46 | 1.99±0.74 | 1.73±1.37 | 1.67±1.08 |

Table 5.3: Results of the mean and standard deviation of the MUE of each layer for the segmentation considering each selection configuration(part 2).

| | OPL_ONL | IS_OS | RPE_CH |
|---|---|---|---|
| AIBILI | 4.75±32.15 | 5.31±33.76 | 4.39±32.93 |
| Zeiss | 9.08±23.80 | 7.46±24.15 | 13.77±22.33 |
| Explorer | 3.11±1.86 | 1.70±3.91 | 1.94±4.36 |
| Volume | 1.87±1.09 | 1.76±1.50 | 1.58±1.84 |
| Region | 1.87±1.01 | 1.62±1.30 | 1.57±1.60 |
| B-scan | 1.88±0.99 | 1.50±1.11 | 1.52±1.47 |
| Volume + Layer | 1.68±0.87 | 1.31±0.99 | 1.51±1.78 |
| Region + Layer | 1.66±0.85 | 1.28±0.99 | 1.44±1.42 |
| B-scan + Layer | 1.66±0.82 | 1.22±0.90 | 1.38±1.38 |

In tables 5.2 and 5.3 it is possible to confirm the decrease of the metrics as the algorithm selection goes from volume selection to B-scan selection

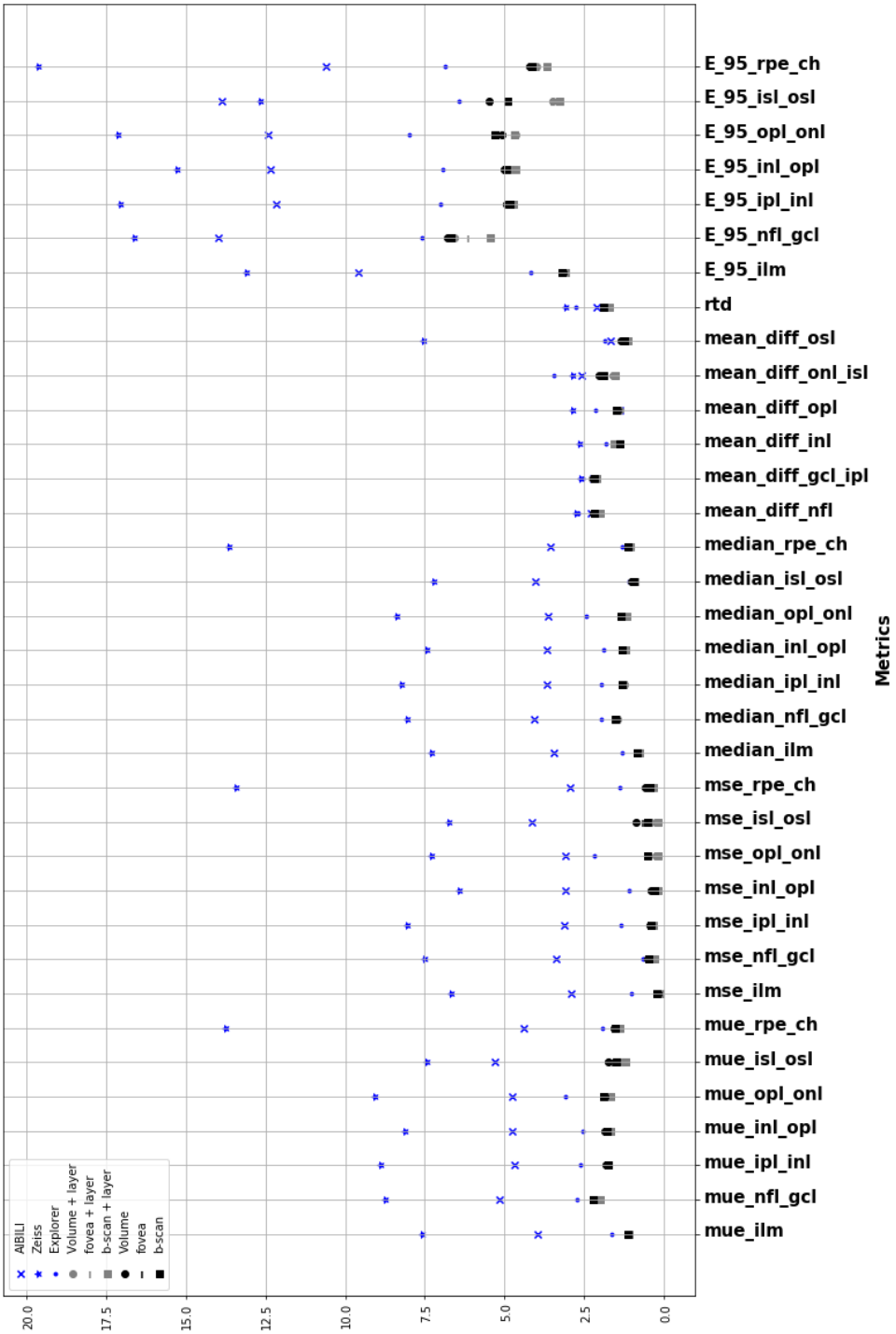


Figure 5.5: Comparison of the mean the metrics for each algorithm selection configuration and each individual algorithm.

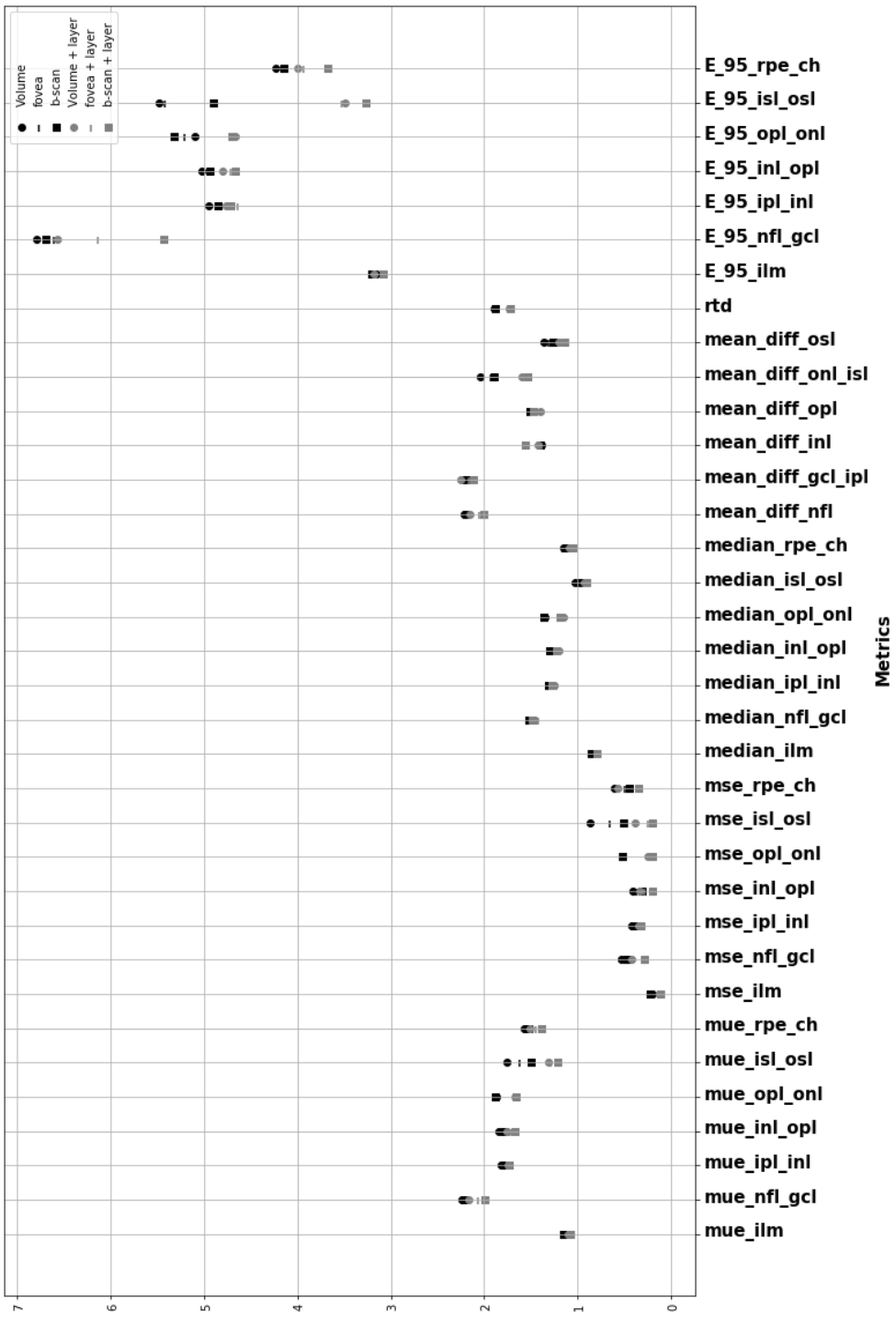


Figure 5.6: Comparison of the mean of each metric for each algorithm selection configuration.

and the clear reduction of the metrics, when comparing any selection configuration with the implementation of an individual algorithm.

The AIBILI and Zeiss algorithm presents a standard deviation that is one order of magnitude higher than the mean value. This is because the algorithms fail completely on the segmentation of some volumes, as was presented in the previous section. For the MUE of the ILM, for example, most of the volumes present metrics around one but, on the few volumes where the segmentation fails, these values are much larger. For the AIBILI algorithm, there is only one volume where this happens (the volume 82) but on this volume, the MUE of the ILM layer is 359.48 and this is what causes a high standard deviation. More volumes fail for the Zeiss algorithm, but the MUE values aren't as large as for the AIBILI, reaching a maximum value of 162.07 at volume 10, and this is why the mean and standard deviation are closer for the Zeiss algorithm than for the AIBILI.

On the algorithm selection, without layer selection, there is an increase in the metrics of certain layers as the algorithm selection progresses on region specificity. For example, the MUE of the ILM layer is higher for the region selection than for the volume selection and the MUE of the IS-OS layer is higher for the volume selection than for the Explorer algorithm. These differences can be associated with situations where the selected algorithm wasn't the best algorithm for that specific layer, but for the layers that showed higher differences, when compared with the reference. The ILM layer, for example, is usually well-segmented so the algorithm selection must have occurred to benefit the other layers that are more problematic and can show a larger difference between the algorithms. For the internal layers, for example, the difference in metrics is kept as expected.

The expectation about the results was that there was a significant reduction between the use of individual algorithms and the selection of the best one and that there is a metric reduction as the regions for the selection keep getting more specific. Even with the small deviation pointed, it is possible to see that results meet this expectation. This way, the algorithm's selection is beneficial, when compared with the implementation of only one algorithm, and the more specific region selection implementations are beneficial, when compared with only selecting the best algorithm for the total volume.

## 5.2 Development of the selection method

### 5.2.1 Comparison of crossed and reference metrics

The comparison of the crossed metrics and reference metrics was done for the volumes presented on Table 5.1 and were evaluated with plots such as the ones presented on Figures 5.7 and 5.8.

Figure 5.7 presents the results of a volume where the Zeiss algorithm fails. This is easily detected by the resulting metrics once all the metrics, excluding the thickness metrics, are very distant from the ones from Explorer. This is expected once when the Zeiss algorithm fails, it's position is far from the other two, that doesn't fail.

On the other hand, in Figure 5.8, it is presented a volume where any of the algorithms fail. It is possible to see that the pattern between the Explorer and Zeiss is kept from the reference crossed metrics to the crossed metrics, but with smaller values on the reference ones. This is, the reference crossed metrics appear dislocated down from the crossed metrics. This factor can be associated with situations where the segmentation of the AIBILI algorithm is not the one that is closer to the reference, as presented in Figure 4.6. On these situations, the subtraction that is applied results on smaller values than the real distances between the two segmentation. However, since the pattern of the distances between the three algorithms remains the same, only with a different scale, this situation doesn't affect the prediction.

For the volumes were one of the algorithms is very distant from the others, it is easy to detect that that algorithm failed and that is possible to do even with the values of only one metric. However,the selection of the best algorithm from the remaining two might be a difficult challenge that might imply an analysis of the relation between the distances of the algorithms on a group of metrics, instead of just one.

Some direct rules could be applied to the crossed metrics to directly select the best algorithm. For example, excluding the algorithm that is farther from the other two and then, from the remaining two, selects the one that is selected more often with the reference metrics. Approaches with similar direct rules were tried, but the results weren't very good, implying the need for a more complex model for the selection.

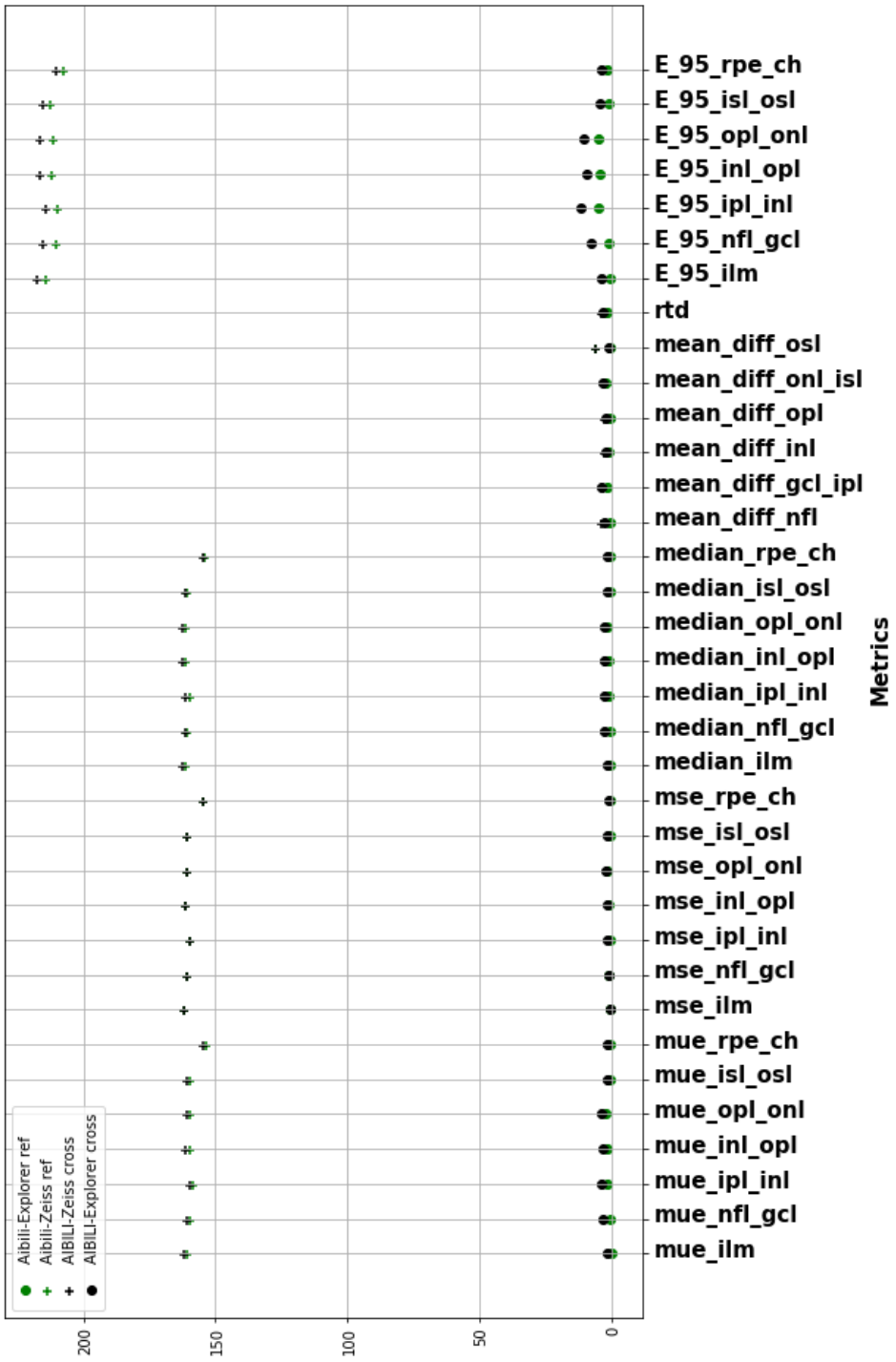


Figure 5.7: Results of the crossed metrics and reference metrics analysis of a volume where the Zeiss algorithm fails (volume 10).

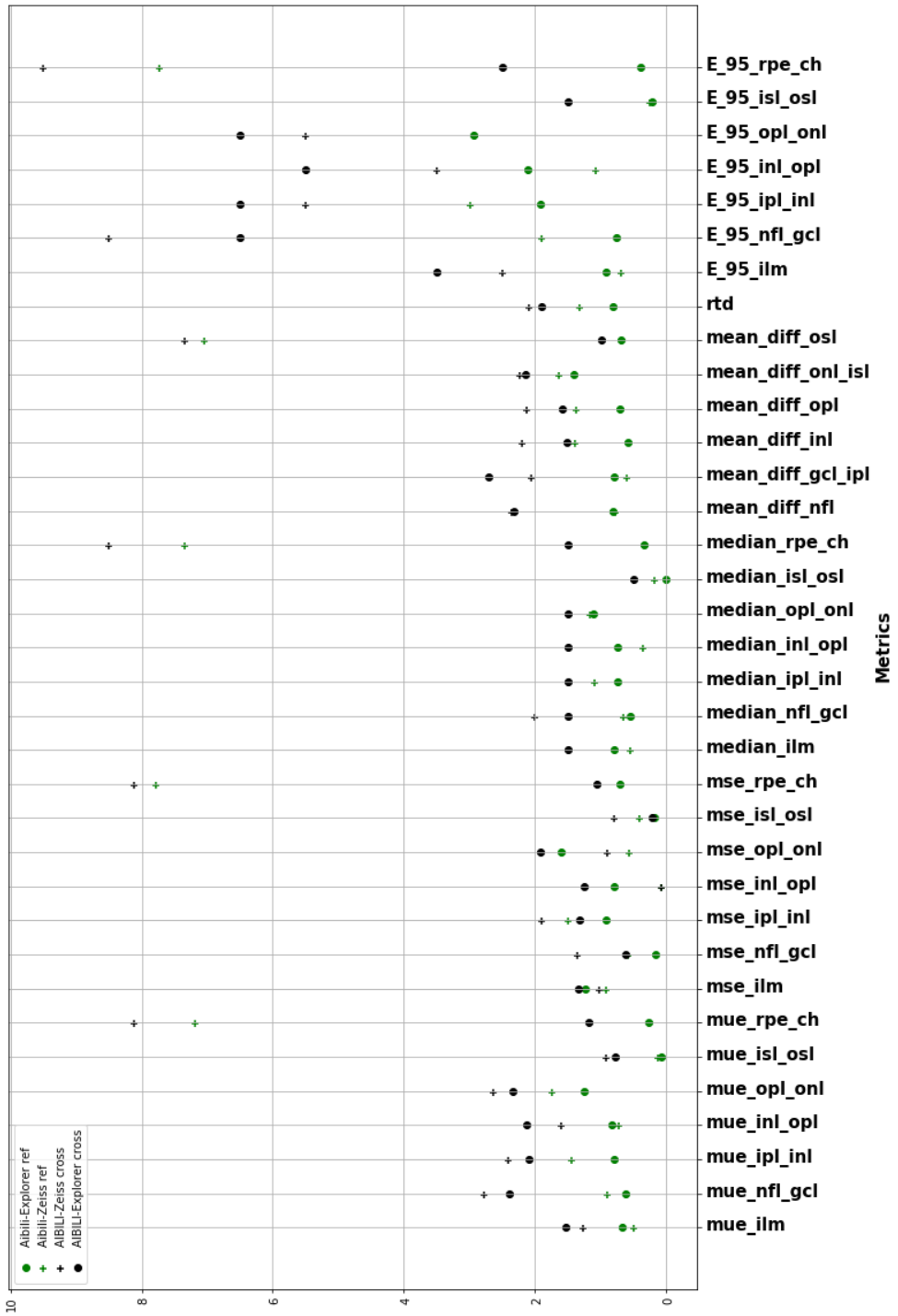


Figure 5.8: Results of the crossed metrics and reference metrics analysis of a volume where any algorithm fails (volume 127).

### 5.2.2 Algorithms selection based on crossed metrics

The results presented here are obtained from the implementation of the RF, DT and ANN models, as presented in section 4.5.2.2. This section is divided into three subsections, for the results of the models of each selection configuration. On each subsection, a small discussion of the results is presented. In the end, there is a discussion of the results presented for all the selection configuration.

#### 5.2.2.1 Volume

In this subsection, the results for the model implementation on the selection of the best algorithm for each volume are presented according to each the data configuration.

**Individual metrics**

Table 5.4 presents the results of the cross-validation on each algorithm for the selection of the best algorithm for the volume segmentation, considering the individual metrics as input. Table 5.5 presents the results of prediction performed by each algorithm on an independent test set.

Table 5.4: Cross-validation results of the algorithms for volume selection considering individual metrics as input. $n$ is the number of data points for each class.

| Model | Class | Precision | Recall | F1 Score | n |
|---|---|---|---|---|---|
| DT | AIBILI | 0.86 ± 0.01 | 0.71 ± 0.04 | 0.78 ± 0.02 | 2606 |
| | Zeiss | 0.26 ± 0.05 | 0.45 ± 0.05 | 0.33 ± 0.05 | 287 |
| | Explorer | 0.32± 0.04 | 0.47± 0.05 | 0.38 ± 0.03 | 537 |
| RF | AIBILI | 0.88 ± 0.02 | 0.75 ± 0.03 | 0.81± 0.02 | 2606 |
| | Zeiss | 0.37 ± 0.06 | 0.57 ± 0.02 | 0.45 ± 0.05 | 287 |
| | Explorer | 0.38 ± 0.03 | 0.54 ± 0.02 | 0.44 ± 0.03 | 537 |
| ANN | AIBILI | 0.88 ± 0.02 | 0.16 ± 0.07 | 0.27 ± 0.11 | 2606 |
| | Zeiss | 0.16 ± 0.02 | 0.82 ± 0.05 | 0.27 ± 0.03 | 287 |
| | Explorer | 0.23 ± 0.01 | 0.65 ± 0.05 | 0.34 ± 0.01 | 537 |

The results presented on table 5.5 are not very good for the Zeiss and Explorer classification. It is also possible to see that the results of the ANN are worst than the results of the other three methods for any of the classification.

Table 5.5: Test results of the algorithms for volume selection considering individual metrics as input. $n$ is the number of data points for each class.

| Model | Class | Precision | Recall | F1 Score | n |
|---|---|---|---|---|---|
| DT | AIBILI | 0.86 | 0.71 | 0.78 | 677 |
| | Zeiss | 0.17 | 0.44 | 0.25 | 52 |
| | Explorer | 0.41 | 0.52 | 0.46 | 146 |
| RF | AIBILI | 0.86 | 0.78 | 0.82 | 677 |
| | Zeiss | 0.23 | 0.37 | 0.28 | 52 |
| | Explorer | 0.41 | 0.50 | 0.45 | 146 |
| ANN | AIBILI | 0.76 | 0.25 | 0.38 | 677 |
| | Zeiss | 0.11 | 0.67 | 0.18 | 52 |
| | Explorer | 0.22 | 0.47 | 0.30 | 146 |

### Layer metrics

Table 5.6 presents the results of the cross-validation on each algorithm for the selection of the best algorithm for the volume segmentation, considering the layer metrics as input. Table 5.7 presents the results of prediction performed by each algorithm on an independent test set.

Table 5.6: Cross-validation results of the algorithms for volume selection considering layer metrics as input. $n$ is the number of data points for each class.

| Model | Class | Precision | Recall | F1 Score | n |
|---|---|---|---|---|---|
| DT | AIBILI | 0.92±0.02 | 0.80±0.05 | 0.85±0.03 | 573 |
| | Zeiss | 0.50±0.24 | 0.32±0.13 | 0.38±0.17 | 24 |
| | Explorer | 0.27±0.02 | 0.50±0.13 | 0.35±0.05 | 89 |
| RF | AIBILI | 0.91±0.03 | 0.91±0.03 | 0.91±0.02 | 573 |
| | Zeiss | 0.51±0.24 | 0.36±0.13 | 0.42±0.17 | 24 |
| | Explorer | 0.40±0.02 | 0.44±0.15 | 0.40±0.06 | 89 |
| ANN | AIBILI | 0.95±0.03 | 0.70±0.06 | 0.80±0.04 | 573 |
| | Zeiss | 0.49±0.23 | 0.74±0.17 | 0.52±0.09 | 24 |
| | Explorer | 0.28±0.06 | 0.67±0.15 | 0.38±0.04 | 89 |

Table 5.7: Test results of the algorithms for volume selection considering layer metrics as input. $n$ is the number of data points for each class.

| Model | Class | Precision | Recall | F1 Score | n |
|---|---|---|---|---|---|
| DT | AIBILI | 0.91 | 0.88 | 0.90 | 146 |
| | Zeiss | 0.00 | 0.00 | 0.00 | 6 |
| | Explorer | 0.33 | 0.43 | 0.38 | 23 |
| RF | AIBILI | 0.90 | 0.95 | 0.92 | 146 |
| | Zeiss | 0.00 | 0.00 | 0.00 | 6 |
| | Explorer | 0.56 | 0.43 | 0.49 | 23 |
| ANN | AIBILI | 0.95 | 0.85 | 0.90 | 146 |
| | Zeiss | 0.17 | 0.17 | 0.17 | 6 |
| | Explorer | 0.34 | 0.57 | 0.43 | 23 |

The metrics with zero values on the Zeiss classification, in table 5.7, are associated with the lack of training and test data. In the test data set, there were only six data points for the Zeiss algorithm and the DT and RF algorithm predicted them wrongly. The ANN, however, was able to identify correctly some of the data points.

### Volume metrics

Table 5.8 presents the results of the cross-validation on each algorithm for the selection of the best algorithm for the volume segmentation, considering the volume metrics as input. Table 5.9 presents the results of prediction performed by each algorithm on an independent test set.

Table 5.8: Cross-validation results of the algorithms for volume selection considering volume metrics as input. $n$ is the number of data points for each class.

| Model | Class | Precision | Recall | F1 Score | n |
|---|---|---|---|---|---|
| DT | AIBILI | 0.87±0.04 | 0.82±0.07 | 0.84±0.03 | 80 |
| | Zeiss | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 | 1 |
| | Explorer | 0.31±0.15 | 0.40±0.07 | 0.33±0.10 | 17 |
| RF | AIBILI | 0.91±0.04 | 0.78±0.07 | 0.84±0.05 | 80 |
| | Zeiss | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 | 1 |
| | Explorer | 0.36±0.17 | 0.64±0.10 | 0.45±0.15 | 17 |
| ANN | AIBILI | 0.95±0.07 | 0.79±0.10 | 0.85±0.04 | 80 |
| | Zeiss | 0.00±0.00 | 0.00±0.00 | 0.00±0.00 | 1 |
| | Explorer | 0.43±0.10 | 0.83±0.24 | 0.53±0.03 | 17 |

Table 5.9: Test results of the algorithms for volume selection considering volume metrics as input. $n$ is the number of data points for each class.

| Model | Class | Precision | Recall | F1 Score | n |
|---|---|---|---|---|---|
| DT | AIBILI | 0.86 | 0.95 | 0.90 | 20 |
| | Zeiss | 0.00 | 0.00 | 0.00 | 1 |
| | Explorer | 0.67 | 0.50 | 0.57 | 4 |
| RF | AIBILI | 0.94 | 0.85 | 0.89 | 20 |
| | Zeiss | 0.00 | 0.00 | 0.00 | 1 |
| | Explorer | 0.43 | 0.75 | 0.55 | 4 |
| ANN | AIBILI | 1.00 | 0.95 | 0.97 | 20 |
| | Zeiss | 0.00 | 0.00 | 0.00 | 1 |
| | Explorer | 0.67 | 1.00 | 0.80 | 4 |

There is only one data point labelled with the Zeiss classification on the training data set. This is why the metrics for the cross-validation, presented on table 5.8 are always zero. The one data point can only be on the training or on the validation data set. This way, when it is on the training, there is no data point to validate the training of the Zeiss class and when it is on the validation, there is no data point for the algorithm to learn how to identify the Zeiss class. This is why the metrics are always equal to zero for this class.

Even though the metrics for the Zeiss classification are always equal to zero, the algorithm that seems to produce the best results for the volume selection is the ANN applied using the volume metrics. However, to validate this statement, the lack of data on the test set needs to be taken into consideration, specially for the Zeiss and Explorer classification. It is possible that the algorithm has performed well on those four cases, but can fail when faced with different cases.

#### 5.2.2.2 Regions

In this subsection, the results for the model implementation on the selection of the best algorithm for each region are presented according to each the data configuration.

**Individual metrics**

Table 5.10 presents the results of the cross-validation on each algorithm for the selection of the best algorithm for the region segmentation, considering the individual metrics as input. Table 5.11 presents the results of prediction performed by each algorithm on an independent test set.

Table 5.10: Cross-validation results of the algorithms for region selection considering individual metrics as input. $n$ is the number of data points for each class.

| Model | Class | Precision | Recall | F1 Score | n |
|---|---|---|---|---|---|
| DT | AIBILI | 0.84±0.01 | 0.73±0.01 | 0.78±0.01 | 7681 |
| | Zeiss | 0.24±0.03 | 0.42±0.05 | 0.30±0.04 | 841 |
| | Explorer | 0.37±0.02 | 0.45±0.02 | 0.41±0.02 | 1768 |
| RF | AIBILI | 0.86±0.01 | 0.73±0.03 | 0.79±0.02 | 7681 |
| | Zeiss | 0.28±0.02 | 0.50±0.04 | 0.36±0.03 | 841 |
| | Explorer | 0.39±0.01 | 0.51±0.01 | 0.44±0.01 | 1768 |
| ANN | AIBILI | 0.85±0.01 | 0.32±0.04 | 0.46±0.04 | 7681 |
| | Zeiss | 0.16±0.03 | 0.67±0.02 | 0.25±0.04 | 841 |
| | Explorer | 0.275±0.002 | 0.59±0.03 | 0.38±0.01 | 1768 |

Table 5.11: Test results of the algorithms for region selection considering individual metrics as input. $n$ is the number of data points for each class.

| Model | Class | Precision | Recall | F1 Score | n |
|---|---|---|---|---|---|
| DT | AIBILI | 0.86 | 0.72 | 0.78 | 2000 |
| | Zeiss | 0.18 | 0.36 | 0.24 | 168 |
| | Explorer | 0.38 | 0.51 | 0.43 | 457 |
| RF | AIBILI | 0.86 | 0.74 | 0.80 | 2000 |
| | Zeiss | 0.22 | 0.40 | 0.29 | 168 |
| | Explorer | 0.40 | 0.54 | 0.46 | 457 |
| ANN | AIBILI | 0.85 | 0.32 | 0.46 | 2000 |
| | Zeiss | 0.16 | 0.67 | 0.25 | 168 |
| | Explorer | 0.28 | 0.59 | 0.38 | 457 |

The results presented on table 5.11 are similar to the ones presented at table 5.5, for the volume selection. This applies both in terms of the low metrics for the classification of the Zeiss and Explorer and for the results of the ANN that are worst than for the other methods.

### Layer metrics

Table 5.12 presents the results of the cross-validation on each algorithm for the selection of the best algorithm for the region segmentation, considering the layer metrics as input. Table 5.13 presents the results of prediction performed by each algorithm on an independent test set.

Table 5.12: Cross-validation results of the algorithms for region selection considering layer metrics as input. $n$ is the number of data points for each class.

| Model | Class | Precision | Recall | F1 Score | n |
|---|---|---|---|---|---|
| DT | AIBILI | 0.91±0.01 | 0.80±0.04 | 0.86±0.02 | 1700 |
| | Zeiss | 0.25±0.06 | 0.33±0.09 | 0.26±0.02 | 90 |
| | Explorer | 0.32±0.03 | 0.51±0.08 | 0.39±0.01 | 268 |
| RF | AIBILI | 0.94±0.01 | 0.81±0.02 | 0.87±0.01 | 1700 |
| | Zeiss | 0.20±0.06 | 0.46±0.08 | 0.27±0.07 | 90 |
| | Explorer | 0.37±0.01 | 0.53±0.06 | 0.43±0.03 | 268 |
| ANN | AIBILI | 0.93±0.01 | 0.71±0.07 | 0.81±0.04 | 1700 |
| | Zeiss | 0.13±0.03 | 0.46±0.14 | 0.19±0.03 | 90 |
| | Explorer | 0.32±0.05 | 0.50 ± 0.08 | 0.39±0.05 | 268 |

Table 5.13: Test results of the algorithms for region selection considering layer metrics as input. $n$ is the number of data points for each class.

| Model | Class | Precision | Recall | F1 Score | n |
|---|---|---|---|---|---|
| DT | AIBILI | 0.92 | 0.82 | 0.87 | 449 |
| | Zeiss | 0.19 | 0.32 | 0.24 | 19 |
| | Explorer | 0.28 | 0.46 | 0.35 | 57 |
| RF | AIBILI | 0.94 | 0.79 | 0.86 | 449 |
| | Zeiss | 0.14 | 0.53 | 0.22 | 19 |
| | Explorer | 0.39 | 0.53 | 0.45 | 57 |
| ANN | AIBILI | 0.93 | 0.71 | 0.81 | 449 |
| | Zeiss | 0.12 | 0.44 | 0.19 | 19 |
| | Explorer | 0.32 | 0.50 | 0.39 | 57 |

The results presented on table 5.13 show the same tendency for the classification of the Zeiss and Explorer that was presented with the individual metrics. The major difference between this results and the ones from the individual metrics is that with the layer metrics input, the ANN produces results comparables to the ones obtained by the remaining models while on the individual metrics analysis it preformed worst.

### Volume metrics

Table 5.14 presents the results of the cross-validation on each algorithm for the selection of the best algorithm for the region segmentation, considering the volume metrics as input. Table 5.15 presents the results of prediction performed by each algorithm on an independent test set.

Table 5.14: Cross-validation results of the algorithms for region selection considering volume metrics as input. $n$ is the number of data points for each class.

| Model | Class | Precision | Recall | F1 Score | n |
|---|---|---|---|---|---|
| DT | AIBILI | 0.91±0.04 | 0.85±0.06 | 0.88±0.02 | 228 |
| | Zeiss | 0.83±0.17 | 0.75±0.25 | 0.79±0.21 | 5 |
| | Explorer | 0.56±0.07 | 0.66±0.14 | 0.59±0.02 | 61 |
| RF | AIBILI | 0.95±0.03 | 0.78±0.05 | 0.85±0.02 | 228 |
| | Zeiss | 0.33±0.47 | 0.33±0.47 | 0.33±0.5 | 5 |
| | Explorer | 0.48±0.04 | 0.82±0.13 | 0.60±0.04 | 61 |
| ANN | AIBILI | 0.94±0.01 | 0.89±0.01 | 0.917±0.004 | 228 |
| | Zeiss | 0.17±0.17 | 0.5±0.5 | 0.25±0.25 | 5 |
| | Explorer | 0.64±0.05 | 0.76±0.07 | 0.69±0.02 | 61 |

Table 5.15: Test results of the algorithms for region selection considering volume metrics as input. $n$ is the number of data points for each class.

| Model | Class | Precision | Recall | F1 Score | n |
|---|---|---|---|---|---|
| DT | AIBILI | 0.92 | 0.92 | 0.92 | 63 |
| | Zeiss | 0.00 | 0.00 | 0.00 | 1 |
| | Explorer | 0.58 | 0.64 | 0.61 | 11 |
| RF | AIBILI | 0.89 | 0.81 | 0.85 | 63 |
| | Zeiss | 0.00 | 0.00 | 0.00 | 1 |
| | Explorer | 0.28 | 0.45 | 0.34 | 11 |
| ANN | AIBILI | 0.96 | 0.78 | 0.86 | 63 |
| | Zeiss | 0.00 | 0.00 | 0.00 | 1 |
| | Explorer | 0.38 | 0.82 | 0.51 | 11 |

The zero values of the metrics on the Zeiss label in Table 5.15 are associated to the fact that there is only five data points on the training data set and only one on the test data set. This means that the algorithms have few data points to learn and only one opportunity to classify the Zeiss algorithm, where they all fail.

It is important to notice that the models with the volume metrics as input are the ones that produce the best results for the AIBILI and Explorer classification.

#### 5.2.2.3 B-scan

In this subsection, the results for the model implementation on the selection of the best algorithm for each B-scan are presented according to each the data configuration.

**Individual metrics**

Table 5.16 presents the results of the cross-validation on each algorithm for the selection of the best algorithm for the B-scan segmentation, considering the individual metrics as input. Table 5.17 presents the results of prediction performed by each algorithm on an independent test set.

Table 5.16: Cross-validation results of the algorithms for B-scan selection considering individual metrics as input. $n$ is the number of data points for each class.

| Model | Class | Precision | Recall | F1 Score | n |
|---|---|---|---|---|---|
| DT | AIBILI | 0.84±0.01 | 0.55±0.05 | 0.67±0.04 | 858490 |
| | Zeiss | 0.23±0.01 | 0.64±0.02 | 0.34±0.01 | 109230 |
| | Explorer | 0.36±0.01 | 0.51±0.01 | 0.42±0.01 | 232496 |
| RF | AIBILI | 0.85±0.01 | 0.58±0.03 | 0.68±0.03 | 858490 |
| | Zeiss | 0.233±0.002 | 0.61±0.04 | 0.34±0.01 | 109230 |
| | Explorer | 0.37±0.01 | 0.53±0.01 | 0.433±0.004 | 232496 |
| ANN | AIBILI | 0.80±0.02 | 0.27±0.04 | 0.40±0.05 | 858490 |
| | Zeiss | 0.17±0.02 | 0.66±0.03 | 0.27±0.02 | 109230 |
| | Explorer | 0.276±0.001 | 0.58±0.05 | 0.37±0.01 | 232496 |

Table 5.17: Test results of the algorithms for B-scan selection considering individual metrics as input. $n$ is the number of data points for each class.

| Model | Class | Precision | Recall | F1 Score | n |
|---|---|---|---|---|---|
| DT | AIBILI | 0.85 | 0.58 | 0.69 | 226996 |
| | Zeiss | 0.18 | 0.55 | 0.27 | 21937 |
| | Explorer | 0.35 | 0.51 | 0.42 | 57235 |
| RF | AIBILI | 0.85 | 0.59 | 0.70 | 226996 |
| | Zeiss | 0.19 | 0.55 | 0.28 | 21937 |
| | Explorer | 0.36 | 0.52 | 0.43 | 57235 |
| ANN | AIBILI | 0.82 | 0.20 | 0.32 | 226996 |
| | Zeiss | 0.13 | 0.62 | 0.21 | 21937 |
| | Explorer | 0.26 | 0.64 | 0.37 | 57235 |

Although there is more data available, the results for the Zeiss and Explorer labels, shown on Table 5.17, are still worst than the ones for the AIBILI label. Besides this, the results from the ANN are also worst than the ones from the other algorithms.

**Layer metrics**

Table 5.18 presents the results of the cross-validation on each algorithm for the selection of the best algorithm for the B-scan segmentation, considering the layer metrics as input. Table 5.19 presents the results of prediction performed by each algorithm on an independent test set.

Table 5.18: Cross-validation results of the algorithms for B-scan selection considering layer metrics as input. $n$ is the number of data points for each class.

| Model | Class | Precision | Recall | F1 Score | n |
|---|---|---|---|---|---|
| DT | AIBILI | 0.89±0.01 | 0.72±0.02 | 0.80±0.02 | 178461 |
| | Zeiss | 0.21±0.01 | 0.50±0.02 | 0.30±0.01 | 13032 |
| | Explorer | 0.424±0.003 | 0.56±0.03 | 0.48±0.01 | 48607 |
| RF | AIBILI | 0.89± 0.01 | 0.84±0.02 | 0.87±0.01 | 178461 |
| | Zeiss | 0.33±0.02 | 0.38±0.04 | 0.35±0.01 | 13032 |
| | Explorer | 0.53±0.02 | 0.61±0.02 | 0.56±0.01 | 48607 |
| ANN | AIBILI | 0.91±0.01 | 0.73±0.04 | 0.81±0.03 | 178461 |
| | Zeiss | 0.22±0.02 | 0.63±0.04 | 0.32±0.03 | 13032 |
| | Explorer | 0.46±0.01 | 0.56±0.01 | 0.51±0.01 | 48607 |

Table 5.19: Test results of the algorithms for B-scan selection considering layer metrics as input. $n$ is the number of data points for each class.

| Model | Class | Precision | Recall | F1 Score | n |
|---|---|---|---|---|---|
| DT | AIBILI | 0.90 | 0.70 | 0.79 | 47663 |
| | Zeiss | 0.14 | 0.44 | 0.21 | 2340 |
| | Explorer | 0.40 | 0.59 | 0.48 | 11247 |
| RF | AIBILI | 0.90 | 0.84 | 0.87 | 47663 |
| | Zeiss | 0.21 | 0.35 | 0.26 | 2340 |
| | Explorer | 0.53 | 0.60 | 0.56 | 11247 |
| ANN | AIBILI | 0.92 | 0.76 | 0.83 | 47663 |
| | Zeiss | 0.17 | 0.64 | 0.26 | 2340 |
| | Explorer | 0.50 | 0.57 | 0.53 | 11247 |

The results presented on table 5.19 show better metrics for the Explorer classification than the ones obtained with the layer metrics for the other selection configurations.

On this results it is also possible to see that the best models for the layer selection are the RF and ANN.

### Volume metrics

Table 5.20 presents the results of the cross-validation on each algorithm for the selection of the best algorithm for the B-scan segmentation, considering the volume metrics as input. Table 5.21 presents the results of prediction performed by each algorithm on an independent test set.

Table 5.20: Cross-validation results of the algorithms for B-scan selection considering volume metrics as input. $n$ is the number of data points for each class.

| Model | Class | Precision | Recall | F1 Score | n |
|---|---|---|---|---|---|
| DT | AIBILI | 0.96±0.03 | 0.81±0.02 | 0.88±0.02 | 25631 |
| | Zeiss | 0.13±0.08 | 0.43±0.19 | 0.20±0.12 | 505 |
| | Explorer | 0.612±0.005 | 0.81±0.08 | 0.70±0.03 | 8164 |
| RF | AIBILI | 0.97±0.02 | 0.826±0.004 | 0.70±0.01 | 25631 |
| | Zeiss | 0.27±0.04 | 0.60±0.06 | 0.37±0.04 | 505 |
| | Explorer | 0.63±0.04 | 0.86±0.08 | 0.73±0.02 | 8164 |
| ANN | AIBILI | 0.95±0.02 | 0.88±0.01 | 0.92±0.02 | 25631 |
| | Zeiss | 0.26±0.12 | 0.48±0.13 | 0.31±0.09 | 505 |
| | Explorer | 0.72±0.03 | 0.86±0.03 | 0.78±0.03 | 8164 |

Table 5.21: Test results of the algorithms for B-scan selection considering volume metrics as input. $n$ is the number of data points for each class.

| Model | Class | Precision | Recall | F1 Score | n |
|---|---|---|---|---|---|
| DT | AIBILI | 0.91 | 0.93 | 0.92 | 7111 |
| | Zeiss | 0.42 | 0.26 | 0.32 | 62 |
| | Explorer | 0.65 | 0.58 | 0.61 | 1577 |
| RF | AIBILI | 0.96 | 0.85 | 0.90 | 7111 |
| | Zeiss | 0.14 | 0.77 | 0.24 | 62 |
| | Explorer | 0.59 | 0.79 | 0.68 | 1577 |
| ANN | AIBILI | 0.95 | 0.91 | 0.93 | 7111 |
| | Zeiss | 0.17 | 0.65 | 0.27 | 62 |
| | Explorer | 0.71 | 0.77 | 0.74 | 1577 |

The results presented on table 5.21 show best metrics for the Explorer classification than any of the ones previously presented. However, the Zeiss classification still presents bad results in terms of the presented metrics.

The change on the Explorer classification is associated with the increase of data that happened from the volume and region selection to the B-scan selection and shows the advantage of using the volume metrics as the input of the model, when compared with the layer metrics or individual metrics.

#### 5.2.2.4 Discussion

In general, the best results were obtained with the ANN models. However, this is not true for the models that take the individual metrics as input. When the inputs are the individual metrics, the RF and DT methods produce better results than the ANN. This fact can be associated with the overfitting of the model, since the less complex models produce better results than the most complex one.

Overall, there is an improvement of the metrics as the models go from considering the individual metrics to the volume metrics as inputs. The models considering the individual metrics as inputs don't produce great results in any of the selection configuration, suggesting that this isn't the best data configuration and it is preferable to use the layer or volume metrics.

Any of the models could fit well the Zeiss classification. This fact is associated with the lack of volumes/regions/B-scans were the Zeiss segmentation was considered the best. It is expected that the results would be more balanced if a more diverse data set was available.

It is also important to mention that, in terms of the data extracted for each of the selection configurations, the main difference between each model is the amount of data extracted. The data set that was considered has only 123 volumes, which is a very small data set for training and testing the model, especially volume selection. However, each volume has 350 B-scans, which provides a much larger data set for training and testing the models considering the B-scan selection.

Finally, from the presented analysis, it is possible to conclude that the best results were obtained for the ANN with the volume metrics as input, for the selection of each B-scan segmentation algorithm.

## 5.3 Evaluation of the selection method

Tables 5.22 and 5.23 presents the results of the mean of the MUE of the segmentation produced by the model selected above and the segmentation of the individual algorithms for the volumes considered on the test set.

Table 5.22: Results of the MUE of the segmentation produced by the best model and the individual algorithms for the test data set (part 1).

| | ILM | NFL-GCL | IPL-INL | INL-OPL |
|---|---|---|---|---|
| AIBILI | 1.066±0.57 | 2.17±0.78 | 1.60±0.74 | 1.67±0.57 |
| Zeiss | 5.56±12.54 | 6.46±12.20 | 6.54±11.83 | 5.97±12.45 |
| Explorer | 1.47±0.41 | 2.48±0.76 | 2.42±0.78 | 2.44±0.65 |
| Model | 1.19±0.55 | 2.05±0.50 | 1.54±0.58 | 1.61±0.49 |

Table 5.23: Results of the MUE of the segmentation produced by the best model and the individual algorithms for the test data set (part 2).

| | OPL-ONL | IS-OS | RPE-CH |
|---|---|---|---|
| AIBILI | 1.58±0.61 | 1.91±1.64 | 1.36±0.64 |
| Zeiss | 6.83±12.27 | 5.35±12.48 | 10.63±9.20 |
| Explorer | 2.94±0.77 | 1.19±0.42 | 1.47±0.50 |
| Model | 1.67±0.63 | 1.38±0.94 | 1.32±0.63 |

The mean MUE of the selection model is smaller than the mean MUE of the individual algorithms for most of the layers except. However, for the ILM and OPL-ONL, the MUE of the model is slightly higher than the MUE of the AIBILI algorithm and for the IS-OS layer, the MUE of the model is higher than the MUE of the segmentation of the Explorer algorithm.

The presented differences can be associated with two reasons. The first is the justification given on the discussion of the selection configurations. This is, the selection happens for the benefit of the group of layers and sometimes the selected algorithm isn't the algorithm that produces the best results on all the layers. The second reason is related to the wrong selection of the model that happens due to situations that the selection criteria can't identify correctly.

To quantify the benefit of using the selection criteria, the average decrease of the MUE, comparing the segmentation with the algorithm selection and the segmentation from each algorithm, is calculated for each layer. Making an average of the values for each layer, the selection criteria achieved an average reduction of the MUE of 7.99% from the AIBILI algorithm, 20.60%

from the Explorer algorithm and 76.24% from the Zeiss algorithm. This way, even with the limitation on the selection model, the selected model produces better results than the direct application of only the best algorithm that, in this case, is considered the AIBILI algorithm.

# Chapter 6

# Conclusion

## 6.1 Conclusion

The presented results allow the conclusion that the selection of the best algorithm has the potential to produce best results than the implementation of individual segmentation algorithms.

It was also shown that the best method to implement the selection consists on applying an ANN(when compared with a RF and a DT). This ANN takes as input all the crossed metrics that represent a B-scan and selects the best algorithm for the segmentation of that B-scan.

Although the results of the implementation of the developed selection criteria on a test set are encouraging, the classification metrics related to the selection of the Zeiss algorithm are still lower than for the selection of the other two algorithms. This is associated with the lack of data for this class on the data set that was considered on this research project.

Finally, the method proposed has proven to produce better results than the implementation of the individual segmentation algorithm, leading to the conclusion that this could be a good approach towards the implementation of OCT layer segmentation on daily clinical practice.

## 6.2 Future work

Since one of the limitations of this research work was the tendency towards the selection of the AIBILI algorithm, due to the way the reference segmentation was produced, it would be interesting to try the presented approach without this bias to see if there is an improvement of the results, especially for the selection of the Zeiss algorithm. Besides this, it would be interesting

to try this approach on a larger and more diverse data set, with more difficult volumes and more diverse type of lesions, to see if this leads to a more balanced data set and more consistent results.

Furthermore, it is important to explore different approaches for the development of the selection criteria. One of the possibilities is to use unsupervised learning models, namely clustering techniques, to group the volumes by the characteristics of the crossed metrics and attribute the selected algorithm for each of the groups. Another possibility is to implement a Convolutional Neural Network on the volumes to directly select the algorithm based on the volume properties.

# Bibliography


[1] N. S. Abdelfattah, H. Zhang, D. S. Boyer, P. J. Rosenfeld, W. J. Feuer, G. Gregori, and S. R. Sadda. Drusen volume as a predictor of disease progression in patients with late age-related macular degeneration in the fellow eye. *Investigative Ophthalmology & Visual Science*, 57(4):1839–1846, 2016.

[2] C. Ahlers, C. Simader, W. Geitzenauer, G. Stock, P. Stetson, S. Dastmalchi, and U. Schmidt-Erfurth. Automatic segmentation in three-dimensional analysis of fibrovascular pigmentepithelial detachment using high-definition optical coherence tomography. *British Journal of Ophthalmology*, 92(2):197–203, 2008.

[3] B. J. Antony, M. D. Abràmoff, M. M. Harper, W. Jeong, E. H. Sohn, Y. H. Kwon, R. Kardon, and M. K. Garvin. A combined machine-learning and graph-based framework for the segmentation of retinal surfaces in sd-oct volumes. *Biomedical optics express*, 4(12):2712–2728, 2013.

[4] A. M. Bagci, M. Shahidi, R. Ansari, M. Blair, N. P. Blair, and R. Zelkha. Thickness profiles of retinal layers by optical coherence tomography image segmentation. *American journal of ophthalmology*, 146(5):679–687, 2008.

[5] A. Baghaie, Z. Yu, and R. M. D'Souza. State-of-the-art in retinal optical coherence tomography image analysis. *Quantitative imaging in medicine and surgery*, 5(4):603, 2015.

[6] I. I. Bussel, G. Wollstein, and J. S. Schuman. Oct for glaucoma diagnosis, screening and detection of glaucoma progression. *British Journal of Ophthalmology*, 98(Suppl 2):ii15–ii19, 2014.

[7] C.-L. Chen and R. K. Wang. Optical coherence tomography based angiography. *Biomedical optics express*, 8(2):1056–1082, 2017.

[8] S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, and S. Farsiu. Automatic segmentation of seven retinal layers in sdoct images congruent with expert manual segmentation. *Optics express*, 18(18):19413–19428, 2010.

[9] T. A. Ciulla, A. G. Amador, and B. Zinman. Diabetic retinopathy and diabetic macular edema: pathophysiology, screening, and novel therapies. *Diabetes care*, 26(9):2653–2664, 2003.

[10] D. C. DeBuc. A review of algorithms for segmentation of retinal image data using optical coherence tomography. *Image Segmentation*, 1:15–54, 2011.

[11] L. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. Li, and S. Farsiu. Automatic segmentation of nine retinal layer boundaries in oct images of non-exudative amd patients using deep learning and graph search. *Biomedical optics express*, 8(5):2732–2744, 2017.

[12] D. C. Fernández, H. M. Salinas, and C. A. Puliafito. Automated detection of retinal layer structures on optical coherence tomography images. *Optics express*, 13(25):10200–10216, 2005.

[13] A. Fuller, R. Zawadzki, S. Choi, D. Wiley, J. Werner, and B. Hamann. Segmentation of three-dimensional retinal image data. *IEEE transactions on visualization and computer graphics*, 13(6):1719–1726, 2007.

[14] Z. Gao, W. Bu, Y. Zheng, and X. Wu. Automated layer segmentation of macular oct images via graph-based slic superpixels and manifold ranking approach. *Computerized Medical Imaging and Graphics*, 55:42–53, 2017.

[15] E. Garcia-Martin, V. Pueyo, I. Pinilla, J.-R. Ara, J. Martin, and J. Fernandez. Fourier-domain oct in multiple sclerosis patients: reproducibility and ability to detect retinal nerve fiber layer atrophy. *Investigative ophthalmology & visual science*, 52(7):4124–4131, 2011.

[16] M. K. Garvin, M. D. Abramoff, X. Wu, S. R. Russell, T. L. Burns, and M. Sonka. Automated 3-d intraretinal layer segmentation of macular spectral-domain optical coherence tomography images. *IEEE transactions on medical imaging*, 28(9):1436–1447, 2009.

[17] I. Ghorbel, F. Rossant, I. Bloch, S. Tick, and M. Paques. Automated segmentation of macular layers in oct images and quantitative evaluation of performances. *Pattern Recognition*, 44(8):1590–1603, 2011.

[18] W. Goebel and T. Kretzchmar-Gross. Retinal thickness in diabetic retinopathy: a study using optical coherence tomography (oct). *Retina*, 22(6):759–767, 2002.

[19] M. R. Hee. *Optical coherence tomography of the eye*. PhD thesis, Massachusetts Institute of Technology, 1997.

[20] A. Herzog, K. L. Boyer, and C. Roberts. Robust extraction of the optic nerve head in optical coherence tomography. In *Computer vision and mathematical methods in medical and biomedical image analysis*, pages 395–407. Springer, 2004.

[21] Y. Huang, A. V. Cideciyan, G. I. Papastergiou, E. Banin, S. L. Semple-Rowland, A. H. Milam, and S. G. Jacobson. Relation of optical coherence tomography to microanatomy in normal and rd chickens. *Investigative ophthalmology & visual science*, 39(12):2405–2416, 1998.

[22] T. S. Hwang, Y. Jia, S. S. Gao, S. T. Bailey, A. K. Lauer, C. J. Flaxel, D. J. Wilson, and D. Huang. Optical coherence tomography angiography features of diabetic retinopathy. *Retina (Philadelphia, Pa.)*, 35(11):2371, 2015.

[23] H. Ishikawa, D. M. Stein, G. Wollstein, S. Beaton, J. G. Fujimoto, and J. S. Schuman. Macular segmentation with optical coherence tomography. *Investigative ophthalmology & visual science*, 46(6):2012–2017, 2005.

[24] R. Kafieh, H. Rabbani, and S. Kermani. A review of algorithms for segmentation of optical coherence tomography from retina. *Journal of medical signals and sensors*, 3(1):45, 2013.

[25] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[26] S. Kirbas, K. Turkyilmaz, O. Anlar, A. Tufekci, and M. Durmus. Retinal nerve fiber layer thickness in patients with alzheimer disease. *Journal of Neuro-Ophthalmology*, 33(1):58–61, 2013.

[27] H. Kolb. Gross anatomy of the eye. In *Webvision: The Organization of the Retina and Visual System [Internet]*. University of Utah Health Sciences Center, 2007.

[28] D. Koozekanani, K. Boyer, and C. Roberts. Retinal thickness measurements from optical coherence tomography using a markov boundary model. *IEEE transactions on medical imaging*, 20(9):900–916, 2001.

[29] A. Lang, A. Carass, M. Hauser, E. S. Sotirchos, P. A. Calabresi, H. S. Ying, and J. L. Prince. Retinal layer segmentation of macular oct images using boundary classification. *Biomedical optics express*, 4(7):1133–1152, 2013.

[30] K. Lee, M. Niemeijer, M. K. Garvin, Y. H. Kwon, M. Sonka, and M. D. Abramoff. Segmentation of the optic disc in 3-d oct scans of the optic nerve head. *IEEE transactions on medical imaging*, 29(1):159–168, 2009.

[31] O. Lézoray, L. Grady, et al. Graph theory concepts and definitions used in image processing and analysis. *Image processing and analysing with graphs: theory and practice. CRC Press, Boca Raton, FL*, pages 1–24, 2012.

[32] K. Li, X. Wu, D. Z. Chen, and M. Sonka. Optimal surface segmentation in volumetric images-a graph-theoretic approach. *IEEE transactions on pattern analysis and machine intelligence*, 28(1):119–134, 2005.

[33] M. Mayer, R. Tornow, R. Bock, J. Hornegger, and F. Kruse. Automatic nerve fiber layer segmentation and geometry correction on spectral domain oct images using fuzzy c-means clustering. *Investigative Ophthalmology & Visual Science*, 49(13):1880–1880, 2008.

[34] M. A. Mayer, J. Hornegger, C. Y. Mardin, and R. P. Tornow. Retinal nerve fiber layer segmentation on fd-oct scans of normal subjects and glaucoma patients. *Biomedical optics express*, 1(5):1358–1383, 2010.

[35] A. Mishra, A. Wong, K. Bizheva, and D. A. Clausi. Intra-retinal layer segmentation in optical coherence tomography images. *Optics express*, 17(26):23719–23728, 2009.

[36] M. Mujat, R. C. Chan, B. Cense, B. H. Park, C. Joo, T. Akkin, T. C. Chen, and J. F. De Boer. Retinal nerve fiber layer thickness map determined from optical coherence tomography images. *Optics Express*, 13(23):9480–9491, 2005.

[37] J. Oliveira, S. Pereira, L. Gonçalves, M. Ferreira, and C. A. Silva. Multi-surface segmentation of oct images with amd using sparse high order potentials. *Biomedical optics express*, 8(1):281–297, 2017.

[38] M. Pekala, N. Joshi, T. A. Liu, N. M. Bressler, D. C. DeBuc, and P. Burlina. Deep learning based retinal oct segmentation. *Computers in Biology and Medicine*, 114:103445, 2019.

[39] G. Quellec, K. Lee, M. Dolejsi, M. K. Garvin, M. D. Abramoff, and M. Sonka. Three-dimensional analysis of retinal layer texture: identification of fluid-filled regions in sd-oct of the macula. *IEEE transactions on medical imaging*, 29(6):1321–1330, 2010.

[40] L. A. Remington and D. Goodwin. *Clinical anatomy of the visual system E-Book*. Elsevier Health Sciences, 2011.

[41] A. G. Roy, S. Conjeti, S. P. K. Karri, D. Sheet, A. Katouzian, C. Wachinger, and N. Navab. Relaynet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks. *Biomedical optics express*, 8(8):3627–3642, 2017.

[42] A. Shah, L. Zhou, M. D. Abramoff, and X. Wu. Multiple surface segmentation using convolution neural nets: application to retinal layer segmentation in oct images. *Biomedical optics express*, 9(9):4509–4526, 2018.

[43] M. Shahidi, Z. Wang, and R. Zelkha. Quantitative thickness measurement of retinal layers imaged by optical coherence tomography. *American journal of ophthalmology*, 139(6):1056–1061, 2005.

[44] P. P. Srinivasan, S. J. Heflin, J. A. Izatt, V. Y. Arshavsky, and S. Farsiu. Automatic segmentation of up to ten layer boundaries in sd-oct images of the mouse retina with and without missing layers due to pathology. *Biomedical optics express*, 5(2):348–365, 2014.

[45] V. J. Srinivasan, B. K. Monson, M. Wojtkowski, R. A. Bilonick, I. Gorczynska, R. Chen, J. S. Duker, J. S. Schuman, and J. G. Fujimoto. Characterization of outer retinal morphology with high-speed, ultrahigh-resolution optical coherence tomography. *Investigative ophthalmology & visual science*, 49(4):1571–1579, 2008.

[46] P.-y. Teng et al. Caserel–an open source software for computer-aided segmentation of retinal layers in optical coherence tomography images. *Zenodo, DOI*, 10, 2013.

[47] J. Tian, B. Varga, G. M. Somfai, W.-H. Lee, W. E. Smiddy, and D. C. DeBuc. Real-time automatic segmentation of optical coherence tomog-

raphy volume data of the macular region. *PloS one*, 10(8):e0133908, 2015.

[48] Q. Yang, C. A. Reisman, Z. Wang, Y. Fukuma, M. Hangai, N. Yoshimura, A. Tomidokoro, M. Araie, A. S. Raza, D. C. Hood, et al. Automated layer segmentation of macular oct images using dual-scale gradient information. *Optics express*, 18(20):21293–21307, 2010.

[49] A. Yazdanpanah, G. Hamarneh, B. Smith, and M. Sarunic. Intra-retinal layer segmentation in optical coherence tomography using an active contour approach. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 649–656. Springer, 2009.

# Appendix A

# Individual algorithm's MUE

In this appendix, the MUE of the segmentations obtained with the individual algorithms is presented.

Table A.1: MUE data for the segmentations obtained with the 2D algorithm.

| *ID* | *ILM* | *NFL_GCL* | *IPL_INL* | *INL_OPL* | *OPL_ONL* | *ISL_OSL* | *RPE_CH* |
|---|---|---|---|---|---|---|---|
| 1 | 27.188 | 7.134 | 20.304 | 27.378 | 24.535 | 27.998 | 5.837 |
| 2 | 84.765 | 6.127 | 11.095 | 14.849 | 14.442 | 86.515 | 3.594 |
| 3 | 1.386 | 3.199 | 6.953 | 9.314 | 9.016 | 10.178 | 15.759 |
| 4 | 2.097 | 5.147 | 14.644 | 20.066 | 17.266 | 1.014 | 2.541 |
| 5 | 109.090 | 10.655 | 25.123 | 27.974 | 26.480 | 104.470 | 11.259 |
| 8 | 1.668 | 4.779 | 7.542 | 9.746 | 9.571 | 1.409 | 2.559 |
| 9 | 22.411 | 5.459 | 11.834 | 15.972 | 14.162 | 22.718 | 2.857 |
| 10 | 71.514 | 3.737 | 13.987 | 19.495 | 16.330 | 72.228 | 3.988 |
| 11 | 27.195 | 5.536 | 10.715 | 14.539 | 13.706 | 27.790 | 2.355 |
| 12 | 16.308 | 5.259 | 16.133 | 22.285 | 17.503 | 16.193 | 2.874 |
| 13 | 91.596 | 6.737 | 18.618 | 25.700 | 22.252 | 93.033 | 13.776 |
| 14 | 18.215 | 4.068 | 13.665 | 18.247 | 14.781 | 18.211 | 4.566 |
| 16 | 1.492 | 26.544 | 42.962 | 51.643 | 49.544 | 84.641 | 62.951 |
| 17 | 27.788 | 13.226 | 17.515 | 21.823 | 19.607 | 30.201 | 3.816 |
| 18 | 119.920 | 5.454 | 20.178 | 27.886 | 26.130 | 111.801 | 4.313 |
| 19 | 4.752 | 8.489 | 10.827 | 13.259 | 11.795 | 4.432 | 4.415 |
| 20 | 8.647 | 3.727 | 17.174 | 25.688 | 21.156 | 11.224 | 2.667 |
| 23 | 5.814 | 5.093 | 14.258 | 18.813 | 14.964 | 7.801 | 3.652 |
| 24 | 97.744 | 8.317 | 12.531 | 16.112 | 15.619 | 96.059 | 3.610 |
| 25 | 65.331 | 9.533 | 12.281 | 14.967 | 13.553 | 66.061 | 2.454 |
| 28 | 2.325 | 8.909 | 18.358 | 24.245 | 20.312 | 7.956 | 4.707 |
| 30 | 1.835 | 4.836 | 10.598 | 15.418 | 13.975 | 1.940 | 5.836 |
| 31 | 25.357 | 24.295 | 33.770 | 40.017 | 34.075 | 60.115 | 8.208 |
| 32 | 3.235 | 6.032 | 14.218 | 18.112 | 13.074 | 10.401 | 2.903 |
| 34 | 2.128 | 4.903 | 7.868 | 9.638 | 8.598 | 1.852 | 2.266 |
| 36 | 54.038 | 5.076 | 8.069 | 10.442 | 10.033 | 54.403 | 2.086 |
| 37 | 124.230 | 13.016 | 33.080 | 37.151 | 34.408 | 120.412 | 10.473 |
| 38 | 22.509 | 5.001 | 11.695 | 15.826 | 13.234 | 23.258 | 3.404 |
| 39 | 82.449 | 15.171 | 30.413 | 38.275 | 33.266 | 76.661 | 2.863 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 40 | 1.733 | 3.420 | 7.834 | 9.997 | 7.929 | 1.717 | 2.885 |
| 41 | 126.148 | 6.438 | 25.739 | 36.635 | 33.327 | 127.163 | 3.259 |
| 42 | 73.227 | 6.479 | 21.135 | 28.818 | 26.782 | 73.433 | 4.817 |
| 44 | 2.447 | 4.582 | 15.662 | 21.229 | 18.198 | 4.125 | 3.804 |
| 45 | 25.442 | 6.029 | 17.382 | 23.090 | 19.246 | 25.246 | 2.627 |
| 46 | 3.526 | 4.317 | 16.947 | 23.029 | 19.998 | 3.060 | 4.389 |
| 47 | 118.690 | 10.176 | 16.031 | 21.392 | 22.089 | 116.252 | 4.095 |
| 49 | 114.009 | 5.716 | 10.394 | 13.578 | 13.658 | 114.774 | 9.742 |
| 53 | 107.836 | 4.187 | 18.019 | 25.510 | 23.735 | 108.559 | 4.571 |
| 54 | 1.348 | 2.791 | 8.694 | 11.959 | 11.696 | 0.857 | 1.985 |
| 56 | 58.451 | 3.430 | 12.226 | 17.675 | 15.284 | 60.197 | 3.058 |
| 57 | 20.469 | 3.230 | 4.378 | 5.306 | 5.458 | 20.454 | 1.854 |
| 59 | 22.818 | 5.708 | 13.402 | 18.158 | 17.535 | 24.510 | 3.250 |
| 61 | 58.955 | 27.032 | 37.100 | 43.994 | 38.707 | 81.829 | 15.339 |
| 62 | 62.192 | 6.388 | 20.982 | 28.130 | 23.559 | 65.028 | 3.611 |
| 63 | 2.141 | 4.820 | 11.174 | 14.210 | 11.939 | 1.460 | 3.721 |
| 65 | 4.704 | 19.775 | 34.910 | 43.888 | 36.854 | 39.354 | 25.991 |
| 66 | 1.519 | 7.693 | 25.898 | 33.721 | 30.292 | 0.975 | 2.239 |
| 67 | 1.440 | 2.835 | 4.524 | 5.630 | 5.413 | 1.061 | 6.354 |
| 69 | 92.012 | 7.032 | 21.565 | 29.782 | 24.210 | 93.197 | 1.989 |
| 70 | 108.056 | 3.367 | 6.512 | 8.270 | 8.844 | 108.774 | 14.502 |
| 71 | 33.882 | 4.829 | 11.250 | 14.846 | 10.756 | 34.599 | 2.487 |
| 72 | 44.713 | 5.102 | 17.097 | 23.204 | 18.620 | 45.468 | 3.107 |
| 73 | 1.924 | 7.581 | 12.197 | 15.748 | 15.825 | 2.797 | 2.864 |
| 74 | 123.593 | 4.089 | 11.909 | 16.383 | 14.271 | 124.294 | 7.588 |
| 75 | 2.350 | 3.369 | 8.589 | 11.427 | 11.538 | 1.933 | 3.096 |
| 77 | 109.248 | 4.769 | 16.200 | 22.969 | 18.894 | 109.854 | 16.244 |
| 78 | 55.080 | 10.104 | 14.303 | 17.122 | 16.735 | 54.171 | 3.700 |
| 79 | 23.018 | 4.049 | 8.895 | 11.915 | 9.250 | 22.586 | 10.118 |
| 80 | 64.837 | 98.510 | 110.139 | 104.525 | 96.384 | 182.225 | 22.423 |
| 82 | 290.222 | 310.372 | 316.627 | 325.422 | 329.236 | 397.087 | 393.556 |
| 83 | 68.507 | 15.450 | 17.511 | 18.939 | 17.598 | 69.684 | 7.072 |
| 85 | 102.909 | 4.837 | 11.182 | 15.066 | 11.718 | 104.110 | 10.575 |
| 87 | 64.614 | 11.643 | 22.440 | 29.117 | 25.470 | 73.230 | 24.261 |
| 88 | 16.508 | 4.253 | 11.480 | 16.505 | 15.739 | 16.566 | 8.682 |
| 89 | 50.970 | 6.014 | 9.400 | 11.939 | 11.045 | 51.630 | 8.198 |
| 90 | 95.748 | 81.488 | 93.532 | 90.458 | 91.766 | 242.877 | 157.308 |
| 91 | 128.614 | 42.216 | 49.296 | 34.489 | 33.690 | 142.811 | 11.296 |
| 92 | 114.852 | 7.511 | 16.066 | 21.520 | 18.112 | 114.668 | 16.590 |
| 93 | 43.685 | 11.011 | 15.328 | 19.885 | 17.661 | 47.002 | 5.554 |
| 94 | 1.648 | 6.420 | 17.148 | 22.742 | 16.166 | 2.154 | 2.495 |
| 95 | 1.643 | 47.350 | 61.400 | 69.391 | 70.318 | 191.083 | 176.059 |
| 96 | 67.009 | 5.464 | 20.646 | 30.816 | 26.259 | 68.201 | 3.443 |
| 97 | 92.212 | 11.728 | 20.125 | 26.220 | 22.953 | 90.792 | 4.332 |
| 98 | 8.560 | 4.795 | 18.761 | 25.190 | 19.570 | 8.709 | 3.556 |
| 99 | 106.307 | 2.877 | 10.573 | 14.609 | 13.123 | 107.042 | 6.354 |
| 100 | 2.031 | 5.324 | 14.716 | 20.678 | 17.005 | 1.494 | 10.412 |
| 101 | 12.887 | 6.496 | 19.656 | 25.666 | 20.096 | 17.727 | 5.340 |
| 104 | 2.947 | 7.306 | 17.296 | 21.838 | 18.485 | 2.382 | 2.391 |
| 105 | 87.048 | 3.003 | 6.336 | 8.114 | 8.171 | 87.616 | 2.938 |
| 106 | 94.590 | 7.528 | 20.668 | 26.397 | 22.234 | 95.108 | 6.520 |
| 107 | 73.393 | 5.296 | 14.661 | 20.453 | 13.852 | 73.239 | 8.497 |
| 108 | 8.858 | 11.760 | 23.232 | 31.622 | 29.608 | 22.000 | 5.365 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 109 | 73.965 | 8.541 | 23.073 | 29.633 | 25.255 | 74.688 | 2.314 |
| 110 | 42.317 | 4.679 | 11.411 | 15.351 | 12.273 | 43.515 | 2.487 |
| 113 | 8.887 | 35.787 | 45.524 | 48.960 | 46.604 | 66.336 | 19.789 |
| 115 | 118.356 | 3.759 | 10.786 | 14.224 | 11.670 | 120.183 | 4.969 |
| 116 | 50.201 | 21.875 | 34.511 | 40.448 | 42.105 | 142.600 | 120.667 |
| 119 | 34.321 | 3.711 | 9.304 | 13.060 | 12.967 | 34.490 | 2.921 |
| 120 | 2.358 | 5.452 | 11.022 | 13.071 | 9.328 | 1.921 | 2.387 |
| 121 | 84.053 | 6.663 | 17.529 | 24.805 | 22.160 | 86.709 | 11.209 |
| 122 | 4.611 | 3.523 | 10.009 | 13.180 | 9.266 | 3.831 | 7.887 |
| 123 | 1.832 | 3.794 | 11.396 | 15.672 | 13.219 | 2.563 | 2.961 |
| 126 | 2.732 | 17.826 | 33.110 | 41.457 | 35.616 | 20.255 | 15.454 |
| 127 | 22.310 | 4.463 | 8.536 | 10.859 | 8.920 | 23.091 | 3.979 |
| 128 | 15.513 | 5.094 | 13.174 | 17.982 | 16.999 | 15.094 | 2.226 |
| 130 | 14.555 | 6.265 | 23.020 | 30.956 | 27.721 | 15.226 | 4.296 |
| 131 | 22.086 | 15.609 | 18.456 | 21.567 | 19.667 | 34.928 | 5.492 |
| 132 | 1.431 | 16.296 | 29.194 | 35.696 | 31.083 | 31.888 | 16.855 |
| 133 | 1.836 | 6.645 | 13.513 | 17.449 | 16.194 | 1.314 | 3.206 |
| 134 | 19.108 | 5.071 | 24.070 | 33.154 | 26.966 | 18.909 | 2.433 |
| 135 | 19.525 | 5.836 | 21.259 | 27.578 | 20.868 | 19.680 | 2.642 |
| 137 | 42.366 | 3.131 | 5.216 | 6.167 | 4.906 | 42.567 | 13.947 |
| 139 | 76.285 | 6.982 | 25.007 | 32.396 | 31.339 | 73.862 | 11.834 |
| 140 | 6.936 | 4.892 | 12.311 | 17.191 | 14.284 | 6.864 | 2.316 |
| 145 | 26.053 | 4.718 | 9.591 | 12.461 | 12.440 | 27.167 | 10.585 |
| 146 | 55.960 | 7.759 | 19.045 | 25.102 | 22.629 | 56.991 | 3.138 |
| 149 | 2.363 | 11.963 | 27.417 | 34.563 | 29.795 | 3.847 | 5.215 |
| 151 | 1.979 | 6.475 | 25.447 | 33.979 | 29.694 | 1.725 | 4.195 |
| 152 | 32.878 | 4.372 | 12.093 | 15.962 | 14.665 | 33.651 | 2.207 |
| 153 | 111.793 | 24.185 | 43.216 | 46.484 | 42.800 | 137.440 | 24.361 |
| 154 | 70.024 | 8.068 | 25.609 | 35.066 | 29.388 | 71.518 | 4.476 |
| 155 | 7.173 | 3.527 | 11.729 | 15.925 | 10.544 | 6.588 | 2.616 |
| 156 | 3.193 | 4.070 | 9.738 | 12.990 | 9.774 | 2.898 | 5.350 |
| 157 | 84.524 | 6.251 | 18.876 | 25.678 | 22.059 | 85.585 | 3.138 |
| 159 | 13.762 | 38.588 | 63.143 | 77.768 | 99.929 | 403.327 | 439.558 |
| 160 | 58.778 | 3.968 | 5.647 | 6.511 | 7.346 | 59.415 | 3.641 |
| 163 | 60.462 | 6.297 | 24.540 | 31.807 | 27.749 | 61.362 | 3.967 |
| 165 | 17.663 | 26.659 | 44.297 | 53.470 | 56.773 | 133.280 | 150.651 |
| 166 | 6.484 | 11.072 | 30.632 | 38.184 | 36.115 | 75.319 | 87.013 |
| 169 | 1.853 | 3.931 | 12.658 | 18.052 | 13.937 | 1.042 | 2.908 |
| 170 | 1.564 | 29.670 | 32.132 | 33.397 | 32.398 | 61.071 | 24.412 |
| 179 | 1.481 | 9.672 | 18.693 | 23.473 | 18.437 | 1.520 | 2.396 |
| 180 | 2.253 | 14.936 | 41.529 | 18.526 | 20.089 | 13.975 | 4.329 |

Table A.2: MUE data for the segmentations obtained with the AIBILI algorithm.

| *ID* | *ILM* | *NFL_GCL* | *IPL_INL* | *INL_OPL* | *OPL_ONL* | *ISL_OSL* | *RPE_CH* |
|---|---|---|---|---|---|---|---|
| 1 | 1.004 | 1.799 | 1.399 | 1.520 | 1.472 | 1.155 | 1.294 |
| 2 | 1.127 | 1.897 | 1.451 | 1.489 | 1.455 | 1.205 | 1.199 |
| 3 | 0.686 | 1.466 | 1.099 | 1.125 | 1.092 | 0.825 | 0.823 |
| 4 | 0.745 | 1.334 | 1.185 | 1.277 | 1.159 | 0.729 | 0.803 |
| 5 | 2.034 | 2.929 | 2.461 | 2.532 | 2.354 | 9.000 | 2.577 |
| 8 | 0.958 | 1.835 | 1.227 | 1.202 | 1.306 | 0.960 | 1.119 |
| 9 | 1.025 | 2.058 | 2.363 | 2.094 | 1.718 | 1.890 | 1.652 |
| 10 | 1.303 | 2.170 | 1.746 | 1.673 | 1.624 | 1.415 | 1.405 |
| 11 | 0.895 | 1.950 | 1.276 | 1.346 | 1.252 | 1.083 | 1.707 |
| 12 | 0.940 | 2.322 | 1.476 | 1.563 | 1.446 | 1.176 | 1.275 |
| 13 | 0.892 | 2.037 | 1.236 | 1.326 | 1.195 | 0.938 | 0.967 |
| 14 | 0.945 | 2.101 | 1.888 | 1.964 | 1.718 | 1.386 | 1.154 |
| 16 | 0.883 | 2.267 | 1.753 | 1.839 | 1.934 | 3.161 | 1.817 |
| 17 | 1.517 | 10.696 | 15.646 | 19.009 | 22.859 | 2.724 | 1.767 |
| 18 | 1.959 | 3.215 | 2.498 | 2.365 | 2.431 | 6.920 | 2.664 |
| 19 | 0.855 | 2.536 | 1.500 | 1.736 | 2.688 | 0.887 | 0.944 |
| 20 | 1.159 | 3.629 | 2.378 | 2.207 | 2.976 | 2.581 | 2.409 |
| 23 | 1.203 | 2.171 | 1.781 | 1.900 | 2.000 | 4.113 | 1.955 |
| 24 | 1.133 | 2.080 | 1.492 | 1.535 | 1.472 | 5.689 | 1.652 |
| 25 | 0.852 | 1.574 | 1.200 | 1.290 | 1.292 | 0.802 | 0.892 |
| 28 | 1.172 | 2.853 | 2.318 | 2.240 | 2.013 | 4.124 | 1.694 |
| 30 | 0.875 | 2.653 | 1.832 | 1.844 | 1.622 | 1.817 | 1.487 |
| 31 | 1.034 | 2.411 | 1.777 | 1.695 | 1.561 | 1.622 | 1.492 |
| 32 | 1.070 | 2.712 | 1.986 | 2.127 | 1.949 | 6.381 | 2.295 |
| 34 | 0.886 | 1.726 | 1.105 | 1.188 | 1.146 | 0.928 | 0.971 |
| 36 | 0.785 | 1.816 | 1.029 | 1.096 | 1.085 | 1.024 | 0.993 |
| 37 | 3.447 | 3.146 | 2.628 | 2.614 | 2.472 | 5.311 | 2.665 |
| 38 | 0.786 | 2.179 | 1.373 | 1.525 | 1.410 | 1.193 | 1.224 |
| 39 | 1.264 | 1.912 | 1.504 | 1.514 | 1.451 | 5.976 | 1.502 |
| 40 | 0.836 | 1.885 | 1.569 | 1.423 | 1.243 | 1.184 | 1.098 |
| 41 | 1.616 | 1.905 | 1.949 | 2.104 | 1.940 | 1.580 | 1.773 |
| 42 | 1.091 | 2.002 | 1.546 | 1.667 | 1.541 | 1.230 | 1.284 |
| 44 | 0.993 | 2.302 | 1.816 | 1.959 | 1.950 | 2.678 | 1.994 |
| 45 | 1.027 | 1.987 | 1.443 | 1.480 | 1.431 | 1.187 | 1.212 |
| 46 | 0.790 | 1.511 | 1.670 | 1.650 | 1.381 | 0.949 | 0.907 |
| 47 | 1.535 | 2.698 | 1.976 | 2.167 | 2.158 | 4.070 | 2.149 |
| 49 | 1.249 | 2.248 | 1.687 | 1.855 | 1.935 | 1.897 | 1.688 |
| 53 | 0.949 | 1.747 | 1.135 | 1.217 | 1.188 | 0.866 | 0.884 |
| 54 | 0.787 | 1.767 | 1.227 | 1.309 | 1.609 | 0.945 | 0.985 |
| 56 | 0.985 | 2.332 | 1.576 | 2.475 | 1.712 | 3.170 | 1.103 |
| 57 | 0.839 | 1.688 | 1.028 | 1.036 | 1.037 | 0.718 | 0.801 |
| 59 | 0.902 | 2.469 | 1.853 | 1.937 | 1.942 | 3.240 | 1.889 |
| 61 | 1.405 | 2.073 | 1.700 | 1.808 | 1.838 | 2.755 | 1.782 |
| 62 | 1.016 | 1.989 | 1.413 | 1.475 | 1.386 | 2.141 | 1.710 |
| 63 | 0.776 | 1.740 | 1.153 | 1.084 | 1.016 | 0.815 | 0.857 |
| 65 | 0.943 | 2.112 | 1.374 | 1.367 | 1.281 | 1.084 | 1.106 |
| 66 | 0.846 | 1.461 | 1.083 | 1.416 | 1.268 | 0.991 | 1.026 |
| 67 | 0.797 | 1.525 | 1.107 | 1.108 | 1.028 | 0.740 | 0.808 |
| 69 | 1.218 | 3.599 | 3.365 | 2.736 | 3.731 | 2.393 | 2.845 |
| 70 | 0.932 | 1.521 | 1.214 | 1.421 | 1.603 | 0.898 | 0.954 |
| 71 | 0.875 | 1.807 | 1.198 | 1.221 | 1.079 | 0.874 | 0.884 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 72 | 1.119 | 2.626 | 1.709 | 1.537 | 1.395 | 1.065 | 1.132 |
| 73 | 1.175 | 2.319 | 1.627 | 1.722 | 1.844 | 1.862 | 1.838 |
| 74 | 0.886 | 1.367 | 1.154 | 1.425 | 1.139 | 0.824 | 0.905 |
| 75 | 0.786 | 1.820 | 1.086 | 1.123 | 1.331 | 0.997 | 0.898 |
| 77 | 0.906 | 2.008 | 1.230 | 1.367 | 1.278 | 0.867 | 0.934 |
| 78 | 1.992 | 2.939 | 2.500 | 2.560 | 2.485 | 5.286 | 2.683 |
| 79 | 0.822 | 1.817 | 1.193 | 1.254 | 1.316 | 1.038 | 1.249 |
| 80 | 1.129 | 2.291 | 1.895 | 2.036 | 1.585 | 4.281 | 1.475 |
| 82 | 359.476 | 360.479 | 357.745 | 356.447 | 357.562 | 375.887 | 366.562 |
| 83 | 2.033 | 2.755 | 2.158 | 2.170 | 2.054 | 2.330 | 1.719 |
| 85 | 0.762 | 1.715 | 1.161 | 1.151 | 1.108 | 0.870 | 0.843 |
| 87 | 0.843 | 1.564 | 1.117 | 1.250 | 1.180 | 2.218 | 1.279 |
| 88 | 0.928 | 2.188 | 1.441 | 1.596 | 1.671 | 1.649 | 1.470 |
| 89 | 0.732 | 1.570 | 0.962 | 1.011 | 1.046 | 0.687 | 0.746 |
| 90 | 1.823 | 4.009 | 4.424 | 3.700 | 3.413 | 9.149 | 2.581 |
| 91 | 1.563 | 3.095 | 1.645 | 1.620 | 1.757 | 15.159 | 1.947 |
| 92 | 1.533 | 2.082 | 1.807 | 1.768 | 1.653 | 1.262 | 1.323 |
| 93 | 1.269 | 2.303 | 1.688 | 1.929 | 2.027 | 7.978 | 2.451 |
| 94 | 1.000 | 2.424 | 1.701 | 1.734 | 1.780 | 2.055 | 1.698 |
| 95 | 0.867 | 2.450 | 2.225 | 2.337 | 2.152 | 2.187 | 1.674 |
| 96 | 1.337 | 3.386 | 2.846 | 1.941 | 3.203 | 1.708 | 2.931 |
| 97 | 1.359 | 2.800 | 2.789 | 2.602 | 2.434 | 4.308 | 2.006 |
| 98 | 0.759 | 1.501 | 1.197 | 1.185 | 1.277 | 1.066 | 1.241 |
| 99 | 0.927 | 1.728 | 1.061 | 1.142 | 1.086 | 0.764 | 0.849 |
| 100 | 0.905 | 2.157 | 1.822 | 1.994 | 1.926 | 1.678 | 1.253 |
| 101 | 1.426 | 2.850 | 2.038 | 2.055 | 2.041 | 3.928 | 2.215 |
| 104 | 1.124 | 2.280 | 1.604 | 1.590 | 1.604 | 2.555 | 1.487 |
| 105 | 0.817 | 1.734 | 0.988 | 1.002 | 1.048 | 0.758 | 0.837 |
| 106 | 0.923 | 2.109 | 1.691 | 1.991 | 1.761 | 1.160 | 1.155 |
| 107 | 0.882 | 1.940 | 1.367 | 1.361 | 1.346 | 0.996 | 1.049 |
| 108 | 0.812 | 1.720 | 1.228 | 1.411 | 1.344 | 0.990 | 1.018 |
| 109 | 1.004 | 2.155 | 1.544 | 1.532 | 1.680 | 1.705 | 1.604 |
| 110 | 0.802 | 1.688 | 1.059 | 1.048 | 1.031 | 0.781 | 0.825 |
| 113 | 0.809 | 1.915 | 1.320 | 1.477 | 1.380 | 3.507 | 1.300 |
| 115 | 0.823 | 2.208 | 1.799 | 2.250 | 1.684 | 1.011 | 0.881 |
| 116 | 0.804 | 1.829 | 1.153 | 1.132 | 1.051 | 0.825 | 0.889 |
| 119 | 0.953 | 2.578 | 1.708 | 1.777 | 1.904 | 1.868 | 1.517 |
| 120 | 0.722 | 1.815 | 1.199 | 1.122 | 1.097 | 1.177 | 1.007 |
| 121 | 1.343 | 6.449 | 9.347 | 11.196 | 12.462 | 2.169 | 1.631 |
| 122 | 0.896 | 1.653 | 1.249 | 1.216 | 1.257 | 0.948 | 1.033 |
| 123 | 0.799 | 1.992 | 1.339 | 1.495 | 1.422 | 1.511 | 1.356 |
| 126 | 0.950 | 2.171 | 1.532 | 1.511 | 1.538 | 3.486 | 1.748 |
| 127 | 0.735 | 1.453 | 1.092 | 1.091 | 0.984 | 0.714 | 0.753 |
| 128 | 0.865 | 2.081 | 1.291 | 1.365 | 1.334 | 1.651 | 1.183 |
| 130 | 0.933 | 1.800 | 1.402 | 1.613 | 1.412 | 1.962 | 1.335 |
| 131 | 1.361 | 4.735 | 3.822 | 3.172 | 2.203 | 4.610 | 2.504 |
| 132 | 0.812 | 2.331 | 1.750 | 1.732 | 1.673 | 4.695 | 1.875 |
| 133 | 0.929 | 1.640 | 1.337 | 1.378 | 1.402 | 1.070 | 1.058 |
| 134 | 0.855 | 1.877 | 1.317 | 1.419 | 1.192 | 1.067 | 1.049 |
| 135 | 0.977 | 2.204 | 1.469 | 1.400 | 1.409 | 1.558 | 1.350 |
| 137 | 0.685 | 1.764 | 0.998 | 1.069 | 0.985 | 0.787 | 0.812 |
| 139 | 1.808 | 2.317 | 2.114 | 2.095 | 2.170 | 2.789 | 2.003 |
| 140 | 0.873 | 1.846 | 1.169 | 1.330 | 1.192 | 1.012 | 1.059 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 145 | 0.757 | 1.730 | 1.566 | 1.946 | 1.625 | 1.418 | 1.207 |
| 146 | 0.903 | 1.780 | 1.185 | 1.384 | 1.133 | 0.799 | 0.887 |
| 149 | 1.421 | 2.402 | 1.562 | 1.714 | 1.545 | 2.918 | 1.591 |
| 151 | 0.800 | 1.790 | 1.326 | 1.533 | 1.298 | 1.034 | 0.987 |
| 152 | 0.901 | 1.601 | 1.124 | 1.208 | 1.198 | 0.837 | 0.874 |
| 153 | 1.446 | 2.973 | 1.893 | 1.960 | 1.719 | 2.218 | 1.524 |
| 154 | 1.861 | 2.594 | 2.117 | 1.789 | 1.688 | 1.489 | 1.356 |
| 155 | 0.912 | 1.880 | 1.170 | 1.240 | 1.169 | 0.873 | 0.939 |
| 156 | 0.942 | 2.396 | 1.613 | 1.635 | 1.610 | 1.449 | 1.363 |
| 157 | 0.910 | 1.815 | 1.461 | 1.511 | 1.237 | 0.999 | 1.064 |
| 159 | 1.033 | 2.270 | 1.869 | 1.962 | 1.723 | 4.810 | 2.068 |
| 160 | 0.914 | 1.736 | 1.317 | 1.476 | 1.423 | 0.959 | 1.008 |
| 163 | 1.249 | 2.302 | 1.743 | 1.928 | 1.711 | 1.631 | 1.644 |
| 165 | 0.876 | 2.090 | 1.587 | 1.870 | 1.728 | 5.046 | 1.837 |
| 166 | 1.124 | 3.226 | 2.501 | 2.119 | 1.876 | 3.802 | 1.885 |
| 169 | 0.816 | 1.817 | 1.292 | 1.375 | 1.085 | 0.820 | 0.868 |
| 170 | 0.843 | 2.092 | 1.206 | 1.259 | 1.240 | 1.180 | 1.173 |
| 179 | 0.868 | 1.840 | 1.492 | 1.468 | 1.538 | 1.327 | 1.309 |
| 180 | 0.732 | 2.068 | 1.326 | 1.908 | 1.391 | 7.197 | 1.472 |

Table A.3: MUE data for the segmentations obtained with the Zeiss algorithm.

| *ID* | *ILM* | *NFL_GCL* | *IPL_INL* | *INL_OPL* | *OPL_ONL* | *ISL_OSL* | *RPE_CH* |
|---|---|---|---|---|---|---|---|
| 1 | 60.210 | 61.394 | 61.556 | 59.531 | 60.707 | 60.822 | 67.487 |
| 2 | 1.467 | 3.178 | 2.965 | 1.840 | 2.841 | 1.574 | 8.226 |
| 3 | 52.518 | 52.121 | 51.566 | 52.799 | 52.806 | 52.352 | 47.633 |
| 4 | 1.614 | 2.314 | 2.332 | 1.847 | 2.772 | 1.003 | 8.243 |
| 5 | 2.253 | 4.230 | 4.190 | 2.566 | 3.810 | 2.100 | 10.712 |
| 8 | 1.313 | 2.360 | 2.509 | 1.566 | 2.671 | 1.269 | 8.424 |
| 9 | 1.487 | 2.681 | 2.561 | 2.046 | 2.787 | 1.097 | 9.428 |
| 10 | 162.071 | 161.912 | 160.076 | 161.609 | 161.141 | 161.095 | 154.929 |
| 11 | 1.565 | 2.471 | 2.809 | 1.821 | 2.838 | 1.500 | 7.750 |
| 12 | 1.359 | 2.471 | 2.861 | 2.090 | 3.005 | 1.317 | 7.823 |
| 13 | 1.668 | 2.727 | 2.568 | 1.771 | 2.647 | 1.043 | 8.106 |
| 14 | 1.536 | 2.464 | 3.037 | 1.691 | 2.857 | 1.085 | 8.282 |
| 16 | 1.529 | 2.606 | 3.311 | 1.976 | 3.548 | 1.906 | 8.494 |
| 17 | 2.051 | 3.616 | 3.507 | 3.339 | 3.488 | 1.634 | 8.669 |
| 18 | 2.102 | 3.320 | 3.607 | 2.748 | 3.620 | 1.818 | 8.345 |
| 19 | 1.343 | 2.400 | 2.440 | 1.788 | 3.725 | 0.995 | 8.180 |
| 20 | 1.767 | 2.210 | 2.341 | 2.655 | 3.390 | 2.712 | 7.519 |
| 23 | 1.677 | 3.291 | 2.799 | 2.062 | 3.410 | 1.355 | 9.949 |
| 24 | 1.623 | 3.813 | 3.140 | 2.446 | 3.374 | 1.277 | 8.505 |
| 25 | 38.324 | 36.981 | 36.536 | 38.074 | 39.032 | 37.529 | 30.624 |
| 28 | 2.146 | 3.823 | 3.442 | 2.688 | 3.522 | 2.823 | 9.719 |
| 30 | 1.444 | 2.568 | 2.608 | 1.946 | 3.136 | 1.402 | 9.018 |
| 31 | 145.004 | 144.425 | 144.931 | 142.876 | 144.811 | 143.557 | 151.663 |
| 32 | 1.755 | 2.722 | 2.795 | 2.200 | 3.196 | 1.479 | 8.456 |
| 34 | 1.333 | 2.322 | 2.374 | 1.508 | 2.436 | 1.294 | 8.375 |
| 36 | 1.261 | 2.257 | 2.153 | 1.601 | 2.610 | 0.833 | 8.060 |
| 37 | 3.098 | 5.134 | 5.032 | 4.399 | 4.762 | 2.650 | 7.377 |
| 38 | 1.141 | 2.273 | 2.648 | 1.989 | 2.962 | 0.948 | 8.219 |
| 39 | 2.042 | 3.894 | 6.188 | 6.078 | 7.651 | 1.524 | 8.689 |
| 40 | 1.590 | 2.300 | 2.498 | 1.459 | 2.551 | 1.486 | 8.611 |
| 41 | 1.834 | 3.452 | 4.368 | 2.680 | 3.074 | 1.487 | 8.338 |
| 42 | 1.625 | 3.490 | 4.471 | 2.617 | 2.973 | 1.143 | 8.044 |
| 44 | 1.837 | 2.724 | 2.452 | 1.792 | 3.143 | 2.400 | 8.028 |
| 45 | 1.458 | 3.292 | 3.142 | 1.976 | 3.202 | 1.234 | 8.060 |
| 46 | 48.679 | 49.422 | 49.718 | 48.364 | 48.918 | 48.773 | 53.779 |
| 47 | 2.040 | 2.550 | 3.102 | 2.885 | 3.750 | 1.778 | 8.012 |
| 49 | 1.801 | 3.503 | 3.524 | 2.651 | 3.528 | 1.128 | 7.981 |
| 53 | 1.351 | 2.291 | 2.614 | 1.748 | 2.661 | 1.038 | 8.574 |
| 54 | 1.271 | 1.925 | 2.443 | 1.758 | 3.472 | 0.917 | 9.025 |
| 56 | 1.675 | 2.726 | 3.371 | 2.576 | 3.322 | 1.552 | 9.553 |
| 57 | 1.815 | 1.838 | 2.222 | 1.407 | 2.377 | 1.192 | 8.281 |
| 59 | 1.345 | 4.001 | 3.207 | 2.106 | 3.012 | 1.272 | 9.182 |
| 61 | 1.880 | 3.588 | 4.027 | 2.127 | 3.446 | 1.363 | 8.692 |
| 62 | 1.669 | 3.981 | 3.359 | 2.209 | 2.757 | 1.572 | 8.232 |
| 63 | 109.582 | 109.452 | 108.734 | 110.998 | 109.183 | 109.599 | 102.574 |
| 65 | 1.579 | 2.347 | 2.262 | 1.438 | 2.635 | 1.152 | 8.540 |
| 66 | 1.404 | 2.628 | 2.373 | 1.725 | 2.960 | 1.163 | 8.420 |
| 67 | 26.867 | 26.787 | 26.638 | 27.151 | 27.622 | 27.007 | 25.653 |
| 69 | 1.795 | 1.859 | 2.091 | 1.720 | 2.286 | 1.443 | 6.362 |
| 70 | 1.454 | 2.336 | 2.389 | 1.642 | 3.358 | 0.890 | 7.946 |
| 71 | 1.913 | 1.943 | 2.319 | 1.626 | 2.486 | 1.403 | 7.719 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 72 | 1.520 | 2.580 | 2.823 | 2.068 | 3.370 | 1.229 | 8.557 |
| 73 | 1.646 | 3.686 | 6.140 | 7.474 | 9.282 | 1.427 | 7.862 |
| 74 | 1.286 | 2.200 | 2.181 | 1.796 | 2.410 | 1.028 | 7.942 |
| 75 | 1.339 | 2.062 | 2.340 | 1.500 | 2.915 | 1.306 | 7.760 |
| 77 | 1.384 | 2.198 | 2.438 | 1.666 | 2.620 | 0.829 | 8.421 |
| 78 | 2.380 | 4.104 | 4.143 | 3.059 | 3.517 | 2.049 | 8.845 |
| 79 | 1.161 | 2.586 | 2.555 | 1.611 | 3.145 | 0.856 | 8.810 |
| 80 | 1.761 | 3.381 | 3.523 | 2.306 | 3.215 | 1.827 | 8.841 |
| 82 | 4.650 | 8.699 | 15.778 | 12.810 | 9.458 | 11.277 | 20.356 |
| 83 | 2.041 | 9.081 | 5.116 | 3.684 | 4.081 | 6.441 | 10.609 |
| 85 | 2.168 | 2.662 | 3.097 | 1.784 | 3.393 | 1.503 | 9.085 |
| 87 | 1.211 | 2.892 | 2.654 | 1.741 | 2.584 | 0.851 | 8.573 |
| 88 | 1.363 | 2.187 | 2.504 | 1.704 | 2.671 | 1.033 | 8.903 |
| 89 | 8.399 | 8.223 | 7.833 | 9.218 | 8.485 | 8.601 | 3.652 |
| 90 | 2.115 | 4.515 | 4.196 | 3.293 | 4.013 | 2.621 | 9.513 |
| 91 | 1.867 | 4.843 | 4.960 | 4.391 | 5.912 | 2.192 | 10.058 |
| 92 | 1.842 | 2.764 | 2.946 | 1.764 | 2.630 | 1.075 | 8.660 |
| 93 | 1.706 | 3.319 | 3.261 | 2.384 | 3.315 | 1.665 | 9.144 |
| 94 | 1.421 | 2.632 | 2.429 | 1.779 | 3.293 | 1.534 | 8.360 |
| 95 | 1.319 | 2.412 | 2.703 | 2.399 | 3.102 | 0.979 | 8.379 |
| 96 | 2.121 | 3.475 | 3.501 | 2.884 | 2.397 | 2.309 | 6.546 |
| 97 | 2.249 | 3.676 | 4.200 | 3.205 | 4.691 | 2.538 | 9.820 |
| 98 | 1.410 | 2.591 | 2.944 | 1.636 | 3.292 | 1.154 | 9.049 |
| 99 | 1.317 | 1.893 | 2.221 | 1.518 | 2.432 | 0.932 | 9.066 |
| 100 | 1.994 | 2.789 | 3.211 | 2.042 | 3.894 | 1.296 | 10.338 |
| 101 | 1.867 | 3.072 | 3.415 | 2.338 | 3.687 | 1.737 | 8.705 |
| 104 | 1.616 | 2.650 | 2.796 | 1.921 | 3.069 | 1.542 | 8.097 |
| 105 | 1.190 | 1.956 | 2.233 | 1.457 | 2.346 | 1.154 | 8.002 |
| 106 | 1.292 | 2.556 | 2.818 | 1.689 | 2.748 | 1.077 | 7.865 |
| 107 | 1.260 | 2.764 | 2.674 | 1.779 | 2.781 | 0.895 | 9.181 |
| 108 | 1.506 | 2.662 | 2.703 | 1.734 | 2.880 | 1.412 | 8.136 |
| 109 | 1.517 | 2.560 | 2.866 | 1.838 | 2.951 | 1.153 | 8.062 |
| 110 | 1.504 | 1.898 | 2.381 | 1.455 | 2.487 | 1.180 | 8.151 |
| 113 | 1.519 | 2.633 | 2.782 | 1.933 | 3.190 | 1.257 | 8.597 |
| 115 | 1.283 | 2.387 | 2.621 | 1.613 | 3.067 | 0.777 | 8.468 |
| 116 | 1.364 | 2.659 | 2.425 | 1.504 | 3.029 | 0.900 | 8.190 |
| 119 | 1.570 | 3.034 | 2.716 | 2.059 | 3.399 | 1.386 | 8.621 |
| 120 | 1.747 | 2.240 | 2.732 | 1.406 | 2.906 | 1.174 | 8.294 |
| 121 | 1.740 | 5.640 | 4.833 | 3.097 | 3.207 | 1.245 | 8.733 |
| 122 | 1.531 | 2.327 | 2.700 | 1.579 | 3.244 | 1.029 | 8.488 |
| 123 | 1.776 | 1.943 | 2.627 | 1.485 | 3.025 | 1.652 | 8.190 |
| 126 | 1.799 | 2.763 | 2.670 | 2.000 | 3.482 | 1.488 | 8.957 |
| 127 | 1.228 | 2.351 | 2.531 | 1.816 | 2.724 | 0.841 | 7.945 |
| 128 | 1.284 | 2.521 | 2.620 | 1.780 | 3.109 | 0.939 | 8.701 |
| 130 | 1.564 | 3.208 | 3.267 | 2.159 | 3.298 | 1.402 | 7.933 |
| 131 | 1.760 | 4.373 | 3.790 | 2.531 | 3.642 | 2.238 | 9.288 |
| 132 | 1.236 | 2.417 | 2.570 | 1.879 | 3.422 | 1.569 | 7.950 |
| 133 | 2.254 | 2.953 | 2.994 | 1.861 | 3.200 | 2.057 | 8.654 |
| 134 | 1.347 | 2.435 | 2.656 | 1.873 | 2.652 | 1.061 | 7.996 |
| 135 | 1.631 | 2.332 | 3.166 | 1.876 | 3.225 | 1.590 | 8.242 |
| 137 | 1.359 | 2.118 | 2.377 | 1.464 | 2.711 | 0.801 | 8.880 |
| 139 | 2.428 | 2.570 | 3.424 | 2.332 | 3.248 | 1.733 | 8.330 |
| 140 | 1.296 | 2.384 | 2.202 | 1.653 | 2.381 | 1.395 | 7.969 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 145 | 1.387 | 2.193 | 2.405 | 1.728 | 2.494 | 1.172 | 8.356 |
| 146 | 1.638 | 2.564 | 2.550 | 1.760 | 2.255 | 1.190 | 7.989 |
| 149 | 1.654 | 5.164 | 5.808 | 4.349 | 5.369 | 1.346 | 8.659 |
| 151 | 1.446 | 1.915 | 2.846 | 2.254 | 3.501 | 1.158 | 8.053 |
| 152 | 1.572 | 2.495 | 2.629 | 1.753 | 2.914 | 1.454 | 8.144 |
| 153 | 2.008 | 3.079 | 2.727 | 2.094 | 3.220 | 1.837 | 8.476 |
| 154 | 89.132 | 88.288 | 87.941 | 88.955 | 88.192 | 89.225 | 81.922 |
| 155 | 7.844 | 8.236 | 8.660 | 7.884 | 8.402 | 7.794 | 11.590 |
| 156 | 1.445 | 2.261 | 2.526 | 1.761 | 3.207 | 1.127 | 9.068 |
| 157 | 1.514 | 3.458 | 3.415 | 2.027 | 2.652 | 1.122 | 7.871 |
| 159 | 1.518 | 2.984 | 3.003 | 2.379 | 3.130 | 1.765 | 8.689 |
| 160 | 1.671 | 2.345 | 2.650 | 1.756 | 2.970 | 1.381 | 8.215 |
| 163 | 1.842 | 3.590 | 3.819 | 2.496 | 3.351 | 1.534 | 7.530 |
| 165 | 1.470 | 2.390 | 2.317 | 2.047 | 3.011 | 1.254 | 8.539 |
| 166 | 1.408 | 3.174 | 2.793 | 2.415 | 3.482 | 1.458 | 8.563 |
| 169 | 1.347 | 2.644 | 2.749 | 1.815 | 2.523 | 1.167 | 8.441 |
| 170 | 1.315 | 2.034 | 2.094 | 1.458 | 2.635 | 1.154 | 8.004 |
| 179 | 1.437 | 2.763 | 2.894 | 1.803 | 2.949 | 1.062 | 8.138 |
| 180 | 1.759 | 3.479 | 2.964 | 2.310 | 3.108 | 1.355 | 9.050 |

Table A.4: MUE data for the segmentations obtained with the Explorer algorithm.

| *ID* | *ILM* | *NFL_GCL* | *IPL_INL* | *INL_OPL* | *OPL_ONL* | *ISL_OSL* | *RPE_CH* |
|---|---|---|---|---|---|---|---|
| 1 | 1.716 | 2.466 | 2.182 | 2.187 | 2.641 | 1.500 | 1.284 |
| 2 | 1.249 | 2.934 | 3.197 | 2.727 | 3.669 | 1.631 | 2.034 |
| 3 | 1.259 | 1.668 | 1.759 | 1.844 | 2.378 | 0.845 | 0.789 |
| 4 | 1.763 | 1.911 | 1.781 | 2.091 | 2.300 | 0.760 | 1.445 |
| 5 | 1.873 | 3.061 | 3.288 | 2.764 | 3.816 | 1.698 | 2.672 |
| 8 | 1.281 | 2.300 | 1.816 | 2.033 | 2.686 | 1.024 | 1.528 |
| 9 | 1.537 | 2.053 | 2.110 | 2.572 | 2.962 | 1.214 | 1.663 |
| 10 | 1.341 | 2.523 | 3.076 | 2.931 | 3.516 | 1.338 | 1.087 |
| 11 | 1.647 | 2.321 | 2.791 | 2.532 | 3.726 | 1.208 | 2.153 |
| 12 | 1.393 | 2.842 | 2.481 | 2.249 | 2.772 | 1.725 | 1.314 |
| 13 | 1.845 | 2.445 | 2.192 | 2.282 | 2.425 | 0.824 | 0.942 |
| 14 | 1.490 | 2.559 | 2.702 | 2.225 | 3.225 | 0.973 | 1.182 |
| 16 | 1.376 | 2.332 | 2.694 | 2.395 | 3.521 | 1.894 | 1.720 |
| 17 | 1.848 | 4.798 | 4.324 | 4.212 | 4.252 | 1.155 | 1.203 |
| 18 | 1.846 | 2.934 | 3.755 | 3.425 | 4.185 | 1.433 | 2.540 |
| 19 | 1.501 | 2.782 | 2.356 | 2.962 | 3.779 | 0.937 | 1.127 |
| 20 | 1.378 | 2.482 | 1.847 | 2.607 | 2.666 | 1.622 | 1.352 |
| 23 | 1.413 | 2.670 | 1.920 | 2.156 | 2.804 | 1.447 | 2.929 |
| 24 | 1.382 | 2.865 | 2.342 | 2.464 | 3.246 | 1.740 | 1.846 |
| 25 | 1.234 | 2.275 | 2.006 | 2.591 | 2.576 | 0.749 | 1.110 |
| 28 | 1.832 | 3.377 | 2.566 | 2.742 | 3.100 | 2.260 | 2.683 |
| 30 | 1.524 | 2.395 | 2.222 | 2.180 | 2.783 | 1.491 | 1.900 |
| 31 | 1.175 | 2.415 | 2.052 | 1.932 | 2.625 | 1.123 | 1.262 |
| 32 | 1.403 | 2.391 | 2.246 | 2.256 | 3.025 | 1.784 | 1.451 |
| 34 | 1.268 | 1.895 | 1.797 | 1.960 | 2.227 | 0.922 | 1.022 |
| 36 | 1.273 | 1.925 | 1.760 | 2.003 | 2.274 | 1.051 | 0.777 |
| 37 | 3.087 | 4.165 | 4.006 | 3.618 | 4.093 | 2.445 | 2.246 |
| 38 | 1.115 | 4.168 | 3.925 | 4.022 | 4.243 | 0.774 | 0.997 |
| 39 | 1.900 | 2.511 | 2.578 | 2.155 | 2.979 | 1.776 | 1.310 |
| 40 | 1.724 | 1.866 | 1.728 | 2.112 | 2.647 | 1.076 | 1.074 |
| 41 | 1.499 | 2.673 | 2.895 | 2.489 | 3.191 | 1.166 | 2.020 |
| 42 | 1.283 | 2.740 | 2.798 | 2.555 | 3.181 | 1.951 | 1.318 |
| 44 | 1.776 | 2.495 | 2.156 | 1.961 | 3.091 | 1.541 | 1.654 |
| 45 | 1.405 | 2.610 | 2.215 | 2.103 | 2.679 | 0.909 | 1.285 |
| 46 | 1.500 | 2.062 | 1.975 | 2.142 | 2.402 | 3.533 | 1.007 |
| 47 | 1.766 | 2.780 | 2.493 | 2.511 | 3.344 | 1.607 | 1.968 |
| 49 | 1.644 | 3.424 | 3.556 | 3.432 | 4.696 | 1.183 | 1.647 |
| 53 | 1.230 | 2.349 | 2.198 | 2.011 | 2.976 | 0.925 | 0.961 |
| 54 | 1.172 | 1.826 | 1.822 | 2.208 | 2.627 | 0.808 | 0.946 |
| 56 | 1.078 | 2.706 | 2.300 | 2.889 | 2.401 | 0.846 | 1.340 |
| 57 | 1.926 | 2.271 | 1.871 | 1.823 | 1.991 | 0.820 | 0.847 |
| 59 | 1.182 | 2.663 | 2.350 | 2.535 | 3.154 | 1.422 | 2.859 |
| 61 | 1.487 | 2.597 | 2.579 | 2.111 | 3.086 | 1.197 | 1.301 |
| 62 | 1.594 | 2.476 | 2.413 | 2.466 | 2.617 | 1.112 | 1.507 |
| 63 | 1.676 | 2.286 | 2.104 | 2.011 | 2.492 | 0.841 | 1.163 |
| 65 | 1.571 | 2.419 | 2.002 | 1.860 | 2.441 | 1.161 | 1.234 |
| 66 | 1.511 | 1.972 | 1.857 | 2.124 | 2.586 | 0.974 | 1.319 |
| 67 | 1.281 | 1.711 | 1.812 | 2.035 | 2.344 | 1.229 | 0.945 |
| 69 | 1.593 | 3.425 | 2.738 | 2.454 | 2.374 | 1.610 | 2.084 |
| 70 | 1.388 | 2.026 | 1.819 | 1.876 | 3.181 | 1.172 | 0.992 |
| 71 | 1.995 | 1.940 | 2.274 | 2.147 | 2.476 | 1.084 | 1.005 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 72 | 1.403 | 2.697 | 1.795 | 2.209 | 2.785 | 0.946 | 1.412 |
| 73 | 1.594 | 2.226 | 2.142 | 2.409 | 3.096 | 1.096 | 1.422 |
| 74 | 1.272 | 1.904 | 1.918 | 2.056 | 2.132 | 0.811 | 1.171 |
| 75 | 1.397 | 1.833 | 2.013 | 2.023 | 2.820 | 1.165 | 0.848 |
| 77 | 1.420 | 2.260 | 2.052 | 1.915 | 2.230 | 0.935 | 0.962 |
| 78 | 2.197 | 3.218 | 3.148 | 3.264 | 3.806 | 1.871 | 2.077 |
| 79 | 1.053 | 2.406 | 2.011 | 2.161 | 2.458 | 0.964 | 1.371 |
| 80 | 1.362 | 2.709 | 2.468 | 2.059 | 2.886 | 1.244 | 2.003 |
| 82 | 18.422 | 24.056 | 30.774 | 26.236 | 22.176 | 44.033 | 48.769 |
| 83 | 2.033 | 7.206 | 5.395 | 4.386 | 4.431 | 5.504 | 10.351 |
| 85 | 1.476 | 1.831 | 1.877 | 1.881 | 2.147 | 0.773 | 0.847 |
| 87 | 1.039 | 2.364 | 2.051 | 2.322 | 2.441 | 1.081 | 0.970 |
| 88 | 1.335 | 1.934 | 2.047 | 2.141 | 3.219 | 1.024 | 1.203 |
| 89 | 1.591 | 2.255 | 1.900 | 2.065 | 2.263 | 1.164 | 1.279 |
| 90 | 1.801 | 4.695 | 4.360 | 4.560 | 4.615 | 3.308 | 3.080 |
| 91 | 1.469 | 3.924 | 2.972 | 2.634 | 3.566 | 4.504 | 3.553 |
| 92 | 1.804 | 2.611 | 3.771 | 3.436 | 4.563 | 1.615 | 1.487 |
| 93 | 1.492 | 4.241 | 3.789 | 3.354 | 4.301 | 1.599 | 1.802 |
| 94 | 1.290 | 2.387 | 1.935 | 1.883 | 3.086 | 1.191 | 1.329 |
| 95 | 1.093 | 2.013 | 2.567 | 2.617 | 3.156 | 0.968 | 1.611 |
| 96 | 1.487 | 2.709 | 2.504 | 2.661 | 2.442 | 1.761 | 1.630 |
| 97 | 1.773 | 3.402 | 3.050 | 3.260 | 4.356 | 1.447 | 1.609 |
| 98 | 1.525 | 2.261 | 2.321 | 2.101 | 3.072 | 0.904 | 1.571 |
| 99 | 1.229 | 1.613 | 1.837 | 1.869 | 2.306 | 0.711 | 0.996 |
| 100 | 1.291 | 1.943 | 1.979 | 2.108 | 2.766 | 1.137 | 1.279 |
| 101 | 1.374 | 2.916 | 2.269 | 2.214 | 3.180 | 1.401 | 1.690 |
| 104 | 1.376 | 2.416 | 2.039 | 1.949 | 2.769 | 1.893 | 1.494 |
| 105 | 1.035 | 1.663 | 1.740 | 1.946 | 2.180 | 1.193 | 1.485 |
| 106 | 1.291 | 1.957 | 2.059 | 2.013 | 2.435 | 1.485 | 1.331 |
| 107 | 1.093 | 2.686 | 2.308 | 2.142 | 2.604 | 1.288 | 0.886 |
| 108 | 1.232 | 2.022 | 1.856 | 1.909 | 2.568 | 1.036 | 1.114 |
| 109 | 1.456 | 2.149 | 2.238 | 2.028 | 3.251 | 0.946 | 1.245 |
| 110 | 1.634 | 2.062 | 1.940 | 1.819 | 2.008 | 2.346 | 0.946 |
| 113 | 1.206 | 2.387 | 1.619 | 1.704 | 2.506 | 1.076 | 1.439 |
| 115 | 1.326 | 2.031 | 1.929 | 1.543 | 2.328 | 1.146 | 1.695 |
| 116 | 1.423 | 2.458 | 2.040 | 1.803 | 2.528 | 1.361 | 1.410 |
| 119 | 1.387 | 2.382 | 2.221 | 2.403 | 3.647 | 1.111 | 1.899 |
| 120 | 2.149 | 2.222 | 2.130 | 1.772 | 2.494 | 1.100 | 1.100 |
| 121 | 1.662 | 3.353 | 4.315 | 3.928 | 4.557 | 1.076 | 2.065 |
| 122 | 1.644 | 2.201 | 1.973 | 1.882 | 2.455 | 0.811 | 1.318 |
| 123 | 1.537 | 1.583 | 1.823 | 1.664 | 2.827 | 1.113 | 1.027 |
| 126 | 1.587 | 2.630 | 2.057 | 1.922 | 2.933 | 1.266 | 2.084 |
| 127 | 1.406 | 2.064 | 1.878 | 1.913 | 2.229 | 0.645 | 1.025 |
| 128 | 1.394 | 2.475 | 2.292 | 2.363 | 2.836 | 0.871 | 1.250 |
| 130 | 1.494 | 2.218 | 1.834 | 1.862 | 2.781 | 1.587 | 1.503 |
| 131 | 1.429 | 3.456 | 3.473 | 2.986 | 4.061 | 2.011 | 2.606 |
| 132 | 1.223 | 2.332 | 2.210 | 2.218 | 3.309 | 1.289 | 2.556 |
| 133 | 1.862 | 2.215 | 1.911 | 2.022 | 2.945 | 1.166 | 1.072 |
| 134 | 1.395 | 2.189 | 1.931 | 1.782 | 2.267 | 0.783 | 0.996 |
| 135 | 1.423 | 2.265 | 2.144 | 1.955 | 2.437 | 1.010 | 1.223 |
| 137 | 1.526 | 1.900 | 1.702 | 1.798 | 1.945 | 1.135 | 1.875 |
| 139 | 2.260 | 3.204 | 3.625 | 3.281 | 4.827 | 1.706 | 1.826 |
| 140 | 1.146 | 2.031 | 1.814 | 2.051 | 2.252 | 0.961 | 1.029 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 145 | 1.334 | 1.830 | 1.944 | 2.192 | 3.001 | 0.938 | 1.111 |
| 146 | 1.703 | 2.476 | 2.229 | 2.434 | 2.290 | 0.774 | 1.186 |
| 149 | 2.072 | 2.600 | 2.028 | 1.891 | 2.664 | 1.989 | 1.905 |
| 151 | 1.483 | 1.702 | 2.086 | 1.914 | 2.387 | 0.932 | 1.482 |
| 152 | 1.431 | 2.164 | 1.980 | 2.036 | 2.553 | 0.924 | 0.995 |
| 153 | 1.646 | 3.383 | 2.407 | 2.273 | 3.150 | 1.341 | 2.228 |
| 154 | 2.084 | 3.087 | 3.201 | 2.813 | 3.368 | 1.244 | 1.513 |
| 155 | 1.756 | 1.869 | 2.047 | 2.002 | 2.746 | 1.149 | 0.854 |
| 156 | 1.331 | 2.051 | 2.021 | 2.085 | 2.946 | 0.999 | 1.815 |
| 157 | 1.411 | 2.303 | 2.567 | 2.333 | 2.555 | 1.757 | 1.508 |
| 159 | 1.458 | 2.802 | 2.268 | 2.400 | 2.865 | 1.393 | 1.539 |
| 160 | 1.583 | 1.913 | 2.024 | 2.030 | 2.636 | 1.059 | 1.058 |
| 163 | 1.952 | 2.639 | 2.714 | 2.586 | 3.430 | 1.768 | 1.414 |
| 165 | 2.437 | 3.446 | 2.684 | 2.678 | 3.146 | 2.814 | 2.700 |
| 166 | 1.317 | 4.264 | 3.791 | 3.927 | 3.818 | 1.061 | 1.563 |
| 169 | 1.512 | 2.329 | 2.005 | 1.963 | 2.176 | 0.770 | 0.903 |
| 170 | 1.423 | 2.048 | 1.583 | 1.613 | 2.064 | 1.128 | 1.295 |
| 179 | 1.542 | 2.419 | 2.538 | 2.543 | 3.151 | 1.935 | 0.916 |
| 180 | 1.671 | 2.614 | 2.485 | 2.382 | 2.968 | 1.654 | 1.877 |

# Appendix B

# Plots of the selected volumes

This appendix aims to present the volumes selected at Table 5.1 as the best and worst for each algorithm. The Figures present on this appendix represent one B-scan form each volume.

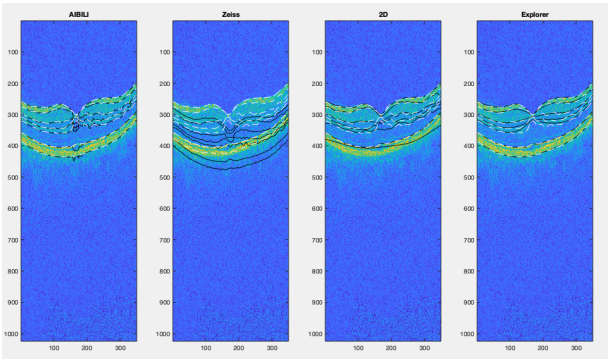


Figure B.1: Representation of the segmentation of each algorithm for volume 3. The dashed white line represents the reference segmentation.

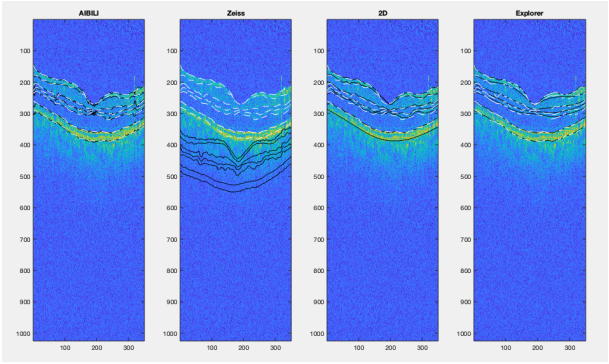


Figure B.2: Representation of the segmentation of each algorithm for volume 10. The dashed white line represents the reference segmentation.

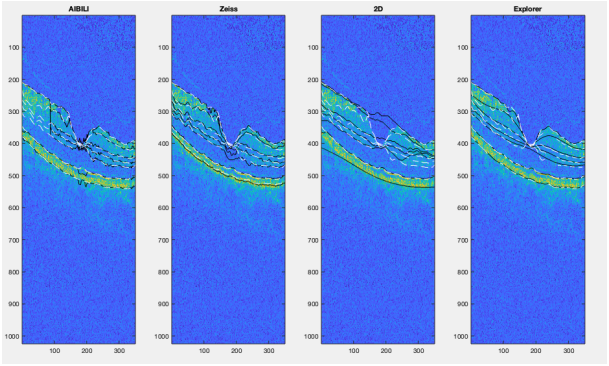


Figure B.3: Representation of the segmentation of each algorithm for volume 17. The dashed white line represents the reference segmentation.

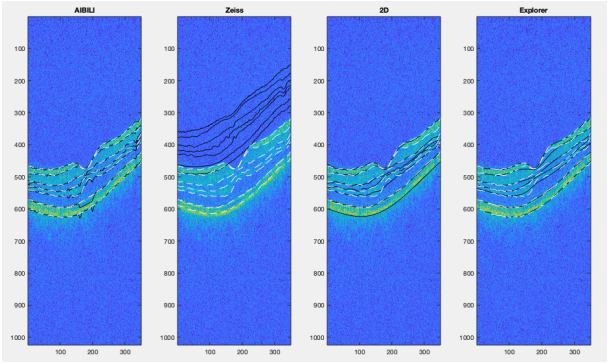


Figure B.4: Representation of the segmentation of each algorithm for volume 31. The dashed white line represents the reference segmentation.

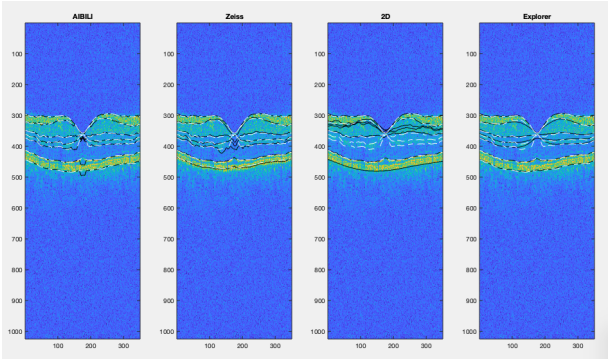


Figure B.5: Representation of the segmentation of each algorithm for volume 34. The dashed white line represents the reference segmentation.

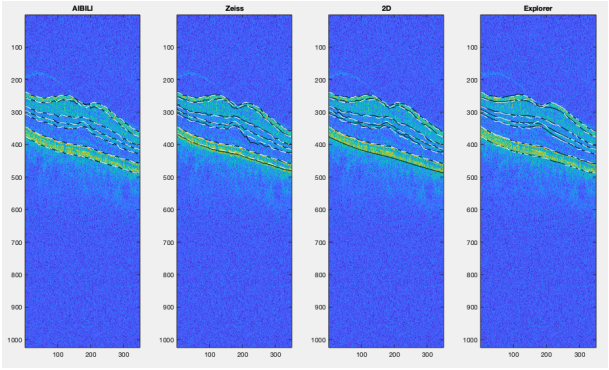


Figure B.6: Representation of the segmentation of each algorithm for volume 36. The dashed white line represents the reference segmentation.

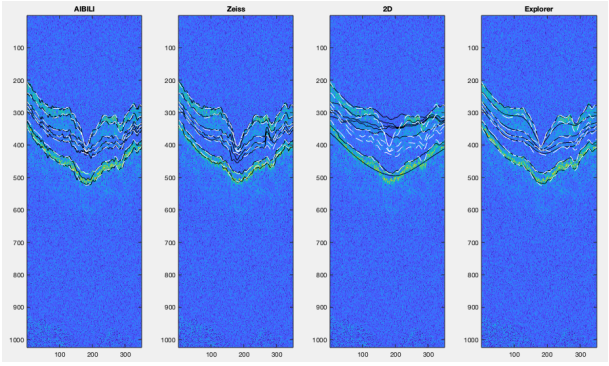


Figure B.7: Representation of the segmentation of each algorithm for volume 37. The dashed white line represents the reference segmentation.

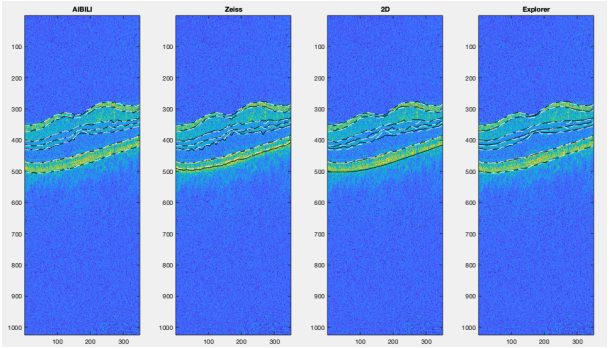


Figure B.8: Representation of the segmentation of each algorithm for volume 54. The dashed white line represents the reference segmentation.

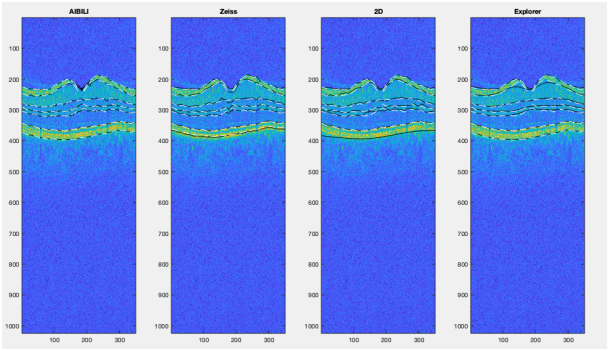


Figure B.9: Representation of the segmentation of each algorithm for volume 57. The dashed white line represents the reference segmentation.

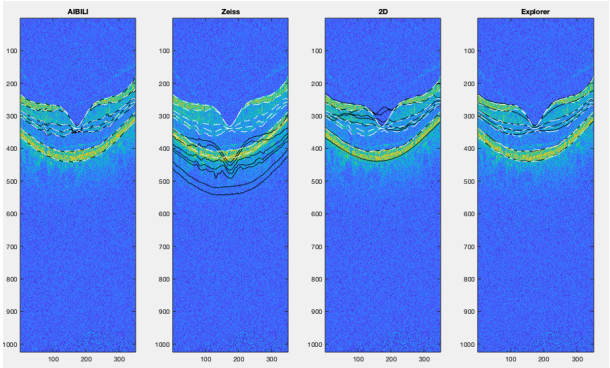


Figure B.10: Representation of the segmentation of each algorithm for volume 63. The dashed white line represents the reference segmentation.

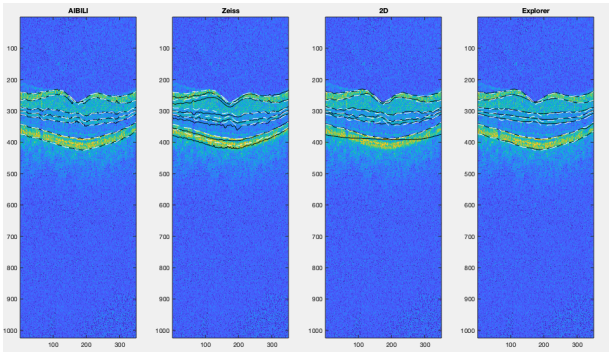


Figure B.11: Representation of the segmentation of each algorithm for volume 67. The dashed white line represents the reference segmentation.

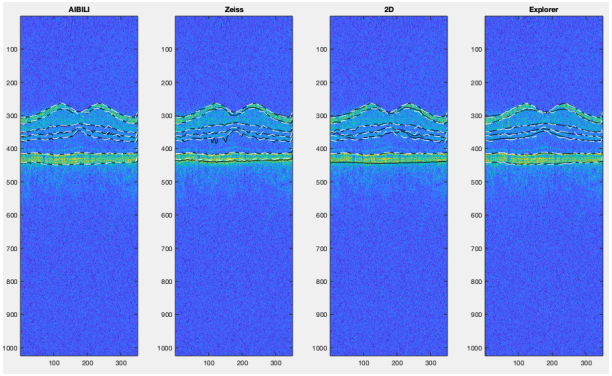


Figure B.12: Representation of the segmentation of each algorithm for volume 74. The dashed white line represents the reference segmentation.

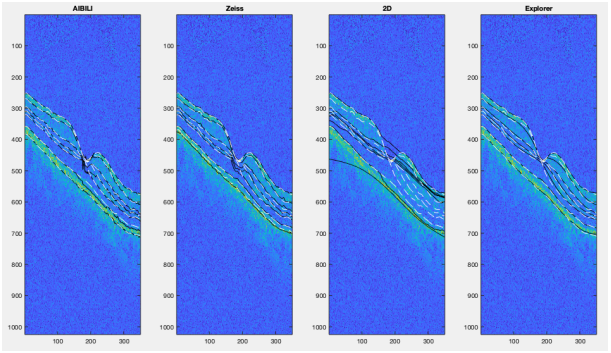


Figure B.13: Representation of the segmentation of each algorithm for volume 80. The dashed white line represents the reference segmentation.

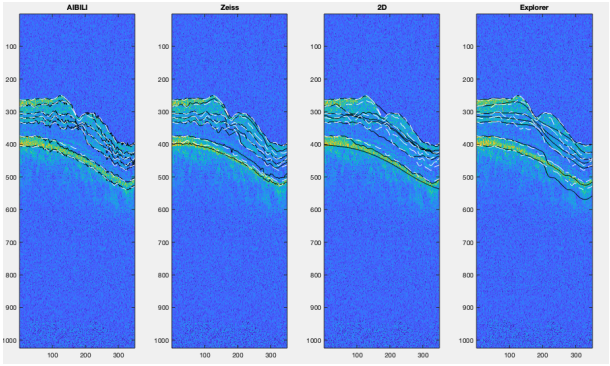


Figure B.14: Representation of the segmentation of each algorithm for volume 83. The dashed white line represents the reference segmentation.

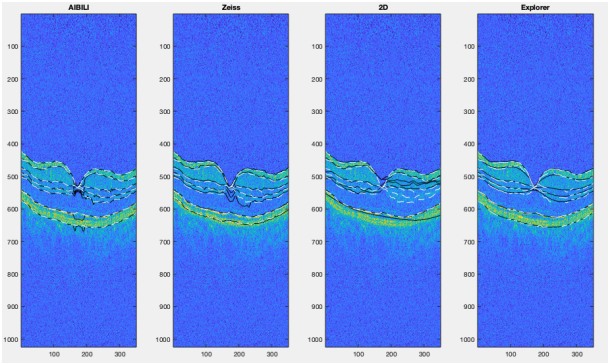


Figure B.15: Representation of the segmentation of each algorithm for volume 85. The dashed white line represents the reference segmentation.

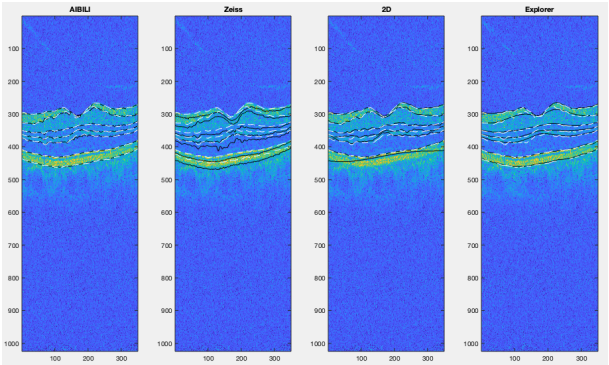


Figure B.16: Representation of the segmentation of each algorithm for volume 89. The dashed white line represents the reference segmentation.

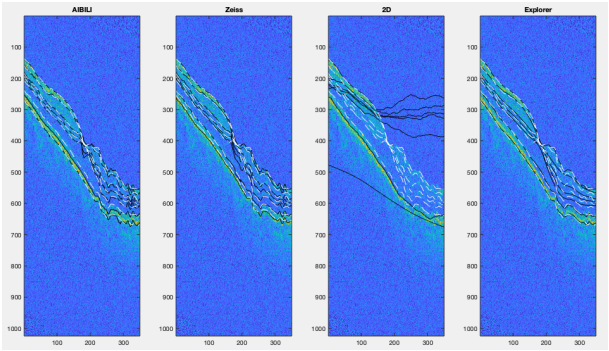


Figure B.17: Representation of the segmentation of each algorithm for volume 90. The dashed white line represents the reference segmentation.

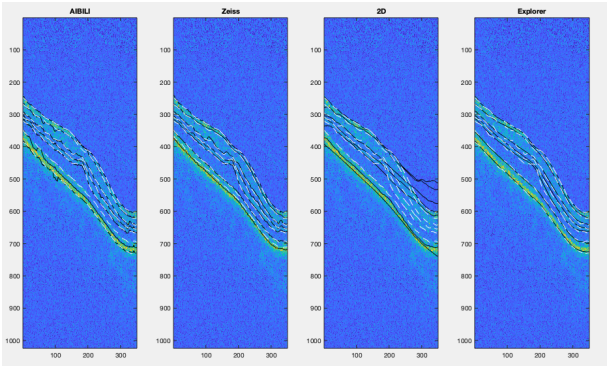


Figure B.18: Representation of the segmentation of each algorithm for volume 91. The dashed white line represents the reference segmentation.

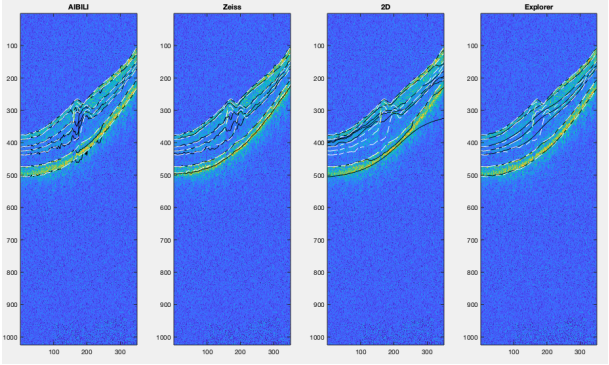


Figure B.19: Representation of the segmentation of each algorithm for volume 95. The dashed white line represents the reference segmentation.

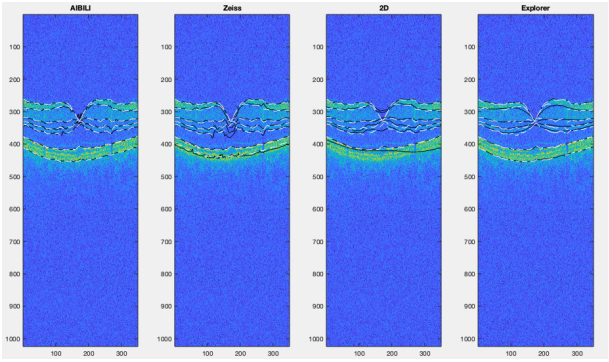


Figure B.20: Representation of the segmentation of each algorithm for volume 99. The dashed white line represents the reference segmentation.

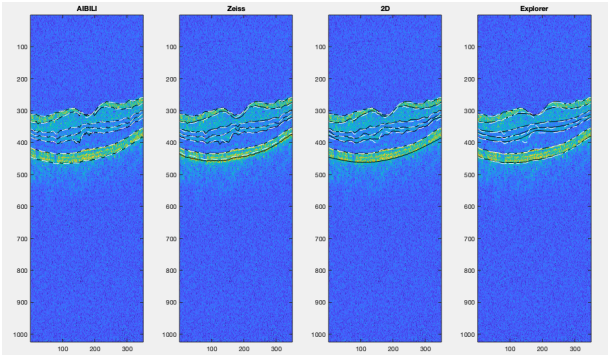


Figure B.21: Representation of the segmentation of each algorithm for volume 105. The dashed white line represents the reference segmentation.

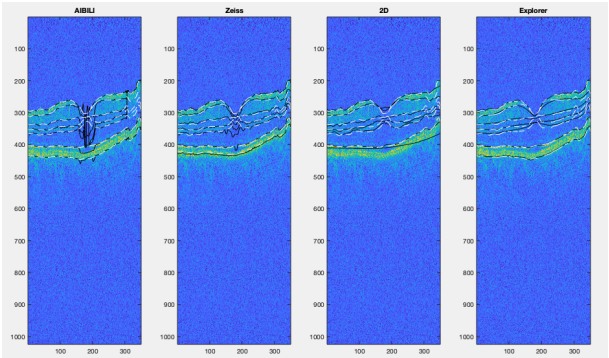


Figure B.22: Representation of the segmentation of each algorithm for volume 121. The dashed white line represents the reference segmentation.

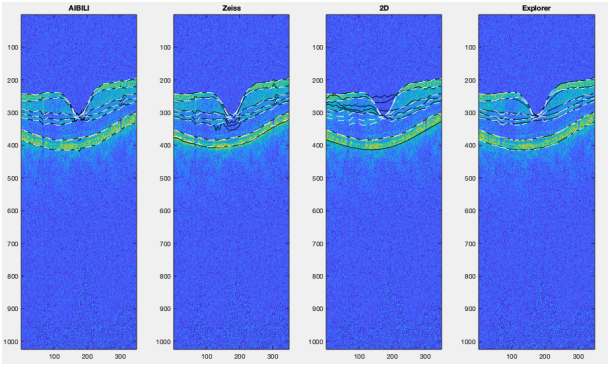


Figure B.23: Representation of the segmentation of each algorithm for volume 127. The dashed white line represents the reference segmentation.

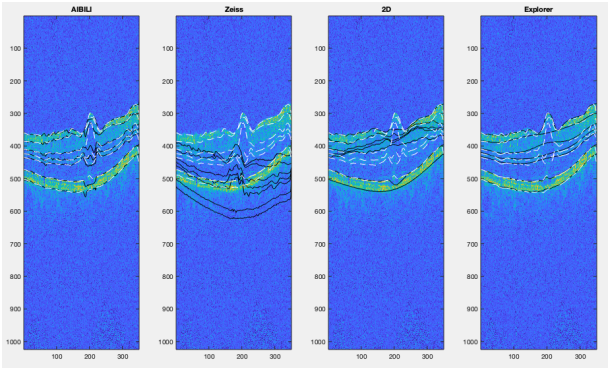


Figure B.24: Representation of the segmentation of each algorithm for volume 154. The dashed white line represents the reference segmentation.

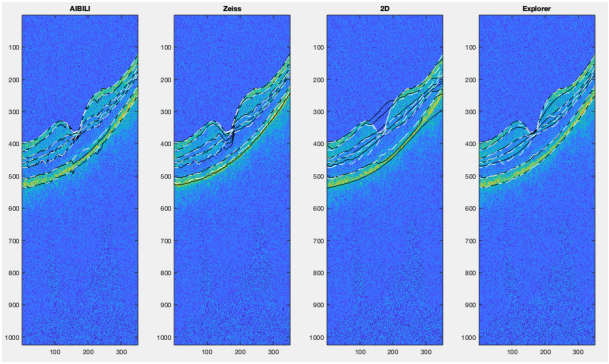


Figure B.25: Representation of the segmentation of each algorithm for volume 159. The dashed white line represents the reference segmentation.

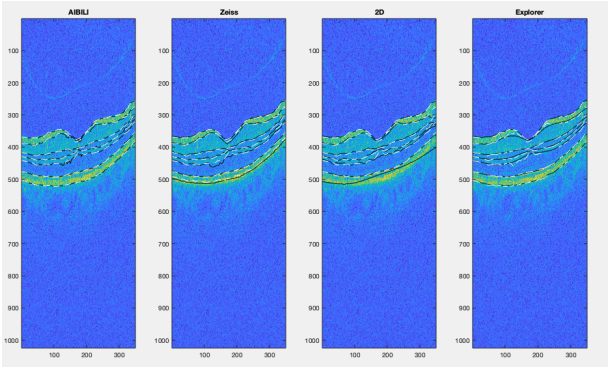


Figure B.26: Representation of the segmentation of each algorithm for volume 170. The dashed white line represents the reference segmentation.